Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Dermoscopic Image-Based Deep Neural Network-Based Skin Lesion Diagnosis and Classification Model

Dr G Lalitha, Dr J Vijayalakshmi

Assistant Professor, MEASI Institute of Information Technology, India
Email: lalithaunom@gmail.com

The article addresses a novel technique for identifying and classifying lesions in dermoscopic images. For lesion segmentation, this method uses a deep separable convolutional network. To start, black frames and extra hair must be removed to remove artificial and natural noise that could impede lesion localization. The photos are rotated and deformed for dataset augmentation after noise reduction. A segmentation model based on deep separable convolution and void convolution forms the basis of the process. To recreate the image's detailed features, features extracted during the encoding phase are fused during the decoding step. The proposed method is effective, as evidenced by the experimental findings, which yield a remarkable segmentation accuracy of 95.24%. This outperforms alternative segmentation models such as U-Net.
Keywords: neural network, dermoscopic images.

1. Introduction

The surface of the skin is where skin disorders first appear and are very common. Repercussions in the middle and later phases of these diseases, along with intricate complications, may arise from a failure to identify and treat them early enough. As a result, the human body may sustain serious injuries or perhaps pass away. It is therefore practical to diagnose skin problems as soon as possible. Dermoscopic images' microscopic technology catches fine details such as skin surface color variations, tissue structure, and texture alterations. Skin disease morphological traits, including color, shape, and texture, can be observed by dermoscopy. Dermatology professionals have a strong basis for diagnosing skin conditions.

A great deal of study in computer-aided diagnosis based on dermoscopic images has been made possible by the integration of digital image processing and computer vision into the medical profession. Dermoscopic images are the media used by dermatology's computer-

aided diagnosis system, which uses clever computations to interpret and process data more objectively. One of the key innovations in this system that stands out is lesion area segmentation (Wencheng 2020). However, variances in the collection process—such as variations in operational processes, climatic conditions, and the distinctive qualities of patients' skin textures and illness manifestations—make it difficult to analyze dermoscopic images.

Lesions with hazy borders and asymmetrical shapes, blood vessels, hair, and bubbles in the soaked fluid obstructing lesion areas, an uneven dispersion of colors in lesion areas, and a lack of a consistent pattern in areas affected by various skin diseases are some of the ways in the challenges appear. Exact lesion segmentation is a challenging task due to the distinct characteristics of dermoscopic images. In the fields of pattern recognition and computer image processing, the segmentation problem is a crucial research topic as well as by means which the computerized detection of skin diseases is made possible.

1.1 Deep Learning in Dermatology

Deep learning is an essential subset of artificial intelligence that has been applied extensively in the domains of computational linguistics, voice understanding, and visual perception. Deep learning has two key benefits when it comes to diagnosing skin diseases. First of all, it is essential for significantly lowering the cost of diagnosing skin diseases, improving productivity, and preserving medical resources. High-quality computers can now efficiently handle tasks related to skin disease image processing that were previously undertaken such as segmentation, identification, analysis, and diagnosis. Second, the deep learning-based auxiliary diagnosis technology demonstrates objectivity and standardization, making it possible to find high-order feature information that is invisible to the human eye. Thus, this improves illness diagnostics capacities and raises diagnosis accuracy—a critical

The outline of the paper's structure is as follows: The research domain and problem were introduced in Section 1. The issue statement and a thorough evaluation of the relevant literature were presented in Section 2. The proposed methodology is described in Section 3, and the analysis of the experiment's findings is covered in Section 4. Section 5, which contains the conclusion and cited references, serves as the paper's culmination.

2. Background Study of the Problem

An increasing number of deep learning-based techniques for segmenting skin lesions have been proposed in recent years due to the rapid advancement and development of deep learning technology. A skin damage segmentation technique that enhances dense convolutional networks was presented by Qi et al. [2020]. Feature information can be fully transferred and the segmentation accuracy is enhanced over other algorithms by optimizing the network structure. However, the proposed approach necessitates extremely deep stacking. It’s quite hard to train a neural network to attain good segmentation accuracy, which is only possible with neural networks.

A skin lesion area segmentation method utilizing a U-Net model with multi-scale and multi-dimensional feature fusion was presented by Wang et al. in [2021]. The shortcomings of the conventional U-Net model's single feature extraction and loss of spatial context information

were solved by this technique. But the comparative model looks unduly complicated, which lessens its persuasiveness because it's tedious.

Jiang et al., [2021] proposed a context encoding and decoding network based on a U-shaped structure to solve the problem of poor segmentation results in skin lesion areas. By integrating efficient dual-channel attention and hole space pyramid pooling, they can obtain more spatial information and semantic information can be used to improve segmentation accuracy, and the segmentation effect is good. However, this method uses a pre-trained encoding module and later needs to be fine-tuned for the skin lesion area segmentation task.

To address the limitations observed in the aforementioned methodologies, Goyal et al., (2022) devised a refined version of the U-Net model. Their model, the Atrous Spatial Pyramid Pooling Parallel Coordinate-Attention U-Net (APC-UNet), enhances the traditional U-Net framework by integrating the Atrous Spatial Pyramid Pooling (ASPP) and ParNet modules into the encoder. This augmentation aims to boost the model's feature extraction capabilities. Moreover, they embed the Coordinate Attention (CA) module within the decoder to augment the model's positional understanding. Their proposed skin lesion area segmentation algorithm leverages the Lovász-hinge loss function to tackle the imbalanced categories in dermoscopic images, thereby enhancing segmentation accuracy. Additionally, a contrast model is introduced to amplify the contrast effect in the images, further improving the overall performance of skin lesion segmentation.

Cui et al., [2020] first performed noise removal on dermoscopic images in 2020, and then used a deep separable convolutional network method for lesion segmentation. Ordinary segmentation methods can only extract shallow feature information such as color, texture and contour of dermoscopic images, which will affect the classification effect of dermoscopic images. With the development of deep learning, convolutional neural networks have entered people's field of vision. This network can extract deep feature information of dermoscopic images, thereby improving the classification effect.

Given the characteristics of dermoscopic image data, Tian et al., [2019] used transfer learning on the ResNet 50 convolutional neural network to conduct classification research on dermoscopic images. The use of transfer learning circumvents the problem of small data volume. First, the dermoscopic images are divided into training sets and test sets. At the algorithmic level, to enhance recognition accuracy and address the issue of data imbalance, this article introduces a weighted cross-entropy loss function. This involves assigning specific weights to the transfer learning model to prioritize and emphasize learning. A small data set is used for training using a hierarchical convolutional neural network.

Wang et al., [2020] combine the two attention mechanism modules ECA-Net (Efficient Channel Attention for Deep Convolutional Neural Networks) and CBAM (Convolutional Block Attention Module). The channel attention module of this module achieves local cross-channel interaction by using one-dimensional (1D) convolution. The convolution kernel k_size is set to 5. The input features undergo 1D convolution to achieve channel interaction, and then Sigmoid is used to weight the interactive information. Allocate, and then perform a dot product between the weight and the input feature to achieve the weight allocation of the channel feature.

Esteva et al. [2019] constructed a CNN based on nearly 130,000 skin disease images (mainly clinical images, partially derived from the ISIC data set), which has reached the average classification level of dermatologists in the test of identifying MM and pigmented nevus. Receiver operating characteristic (ROC) area under the curve (The area under the curve (AUC) was 0.96. This study was published in Nature and became a milestone in AI research in the field of dermatology.

Espsito et al [2021] said that, Melasma is a common disfiguring skin disease, which mainly manifests as symmetrical brown patches involving the cheeks, forehead and mandible. It has a high incidence rate, reaching 30% among women of childbearing age. Chloasma is easy to recur and difficult to cure. It often harms patients' daily life and work. In severe cases, it can lead to mental illness. Studies have shown that the prevalence of anxiety disorders and depression among patients with chloasma is 39% and 25%, and some patients even have suicidal tendencies. Timely and effective treatment of melasma is a key measure to reduce the negative social impact of this disease. Currently, there are many clinical treatment options available, including topical drugs, systemic drugs, photoelectric therapy, traditional Chinese medicine treatment, etc. However, the overall efficacy of these methods is not good, so there is a need to develop new treatments for chloasma. ideas and methods.

Optical Coherence Tomography (OCT) can achieve high-resolution, non-invasive, and highly sensitive three-dimensional imaging of biological tissues. They believe that combining OCT with Kubelka Munk theory can quantitatively analyze the changes in optical parameters of skin at different ages. According to the situation, the research of paper by Wang et al [2020] showed that OCT technology has the potential to quantitatively assess skin surface roughness. Hyperspectral Microscopy Imaging (HMI) technology can provide image information and spectral information of samples at the same time. They used a push-broom HMI system to achieve accurate classification of three types of skin cancer tissues and an accurate classification of squamous cell carcinoma.

Schneider et al [2019] designed to solve the problem that spectral measurement technology has a small single measurement area and cannot provide heterogeneous distribution of corresponding parameters on the whole face, low-cost digital imaging technology has become a popular alternative in recent years, and a variety of equipment including dermatoscopes and technology can achieve non-destructive imaging of subcutaneous tissue to varying degrees and are currently used for skin disease screening, assisting dermatologists in quantitatively monitoring patient status and qualitatively evaluating treatment effects, and evaluating cyanosis and jaundice etc.

Wencheng et al. [2020] first performed noise removal on dermoscopic images in 2020 and then used a deep separable convolutional network method for lesion segmentation. Ordinary segmentation methods can only extract shallow feature information such as color, texture, and contour of dermoscopic images, which will affect the classification effect of dermoscopic images. With the development of deep learning, convolutional neural networks have entered people's field of vision. This network can extract deep feature information of dermoscopic images, thereby improving the classification effect.

Esteva A et al. [2019] used GoogleNet convolutional neural network to classify malignant melanoma in 2019, with an accuracy of 91%. However, the accuracy of this method is only

55% when dealing with multi-classification of dermoscopic images. They used residual neural network to extract features from dermoscopy images in 2020, and then used the support vector machine method to achieve two classifications of malignant melanoma in dermoscopy images, with an accuracy of up to 85%.

Aijaz et al. [2022] developed a CNN model capable of distinguishing various subtypes of psoriasis using clinical images, achieving an overall classification accuracy of 84.12%. Notably, the CNN exhibited an 85% sensitivity and specificity in identifying erythrodermic psoriasis, while demonstrating 73% sensitivity and 75% specificity in classifying pustular psoriasis. Meanwhile, in 2022, the Department of Dermatology at Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, introduced China's inaugural AI-driven tool for detecting rare skin diseases. This mobile-based program conducts preliminary risk assessments for GPP (generalized pustular psoriasis) using uploaded skin lesion images, boasting an identification accuracy exceeding 85%. Moreover, it offers guidance to high-risk patients, directing them to 30 specialized hospitals across 14 cities in China with standardized GPP diagnosis and treatment capabilities, ensuring timely and standardized care plans for these patients.

In recent years, with the rapid development of deep learning technology, medical image analysis technology based on deep learning, especially convolutional neural networks, has become a popular research topic in the field of computer-aided diagnosis. At present, there has been a lot of research on the diagnosis of dermatological medical imaging. The challenges of skin lesion analysis by international skin imaging organizations have also spawned many methods. The diagnostic results of these methods can almost approach the expert diagnosis results. Correspondingly, in recent years, a large number of public data sets have been used for the study of skin diseases.

2.1 Problem Statement

The segmentation of lesions by designing and using the neural network segmentation framework is a research hotspot of the current segmentation task, and the segmentation effect of this method is significantly improved compared with the traditional method. However, for the segmentation task of dermoscopic images, there are still some difficulties in the accurate segmentation of lesions due to the difficulty of accurate segmentation due to obstruction of the location of lesions by black frames and hair noise, blurred boundaries, and uneven distribution of lesions. In this paper, to overcome challenges posed by black frames and hair noise in dermoscopic images, the removal of black frames and hair is implemented. Additionally, a feature extractor is devised using depth separable convolution, based on the codec architecture, to achieve segmentation of dermoscopic images.

3. Methodology of Segmentation using Deep Convolution Model

The approach to segmentation utilizes a profound convolutional model for precise delineation. Preliminary stages involve noise elimination through pre-processing, targeting elements like black frames and unwanted hair. Feature extraction is achieved through the utilization of deep separable convolution. To handle data imbalance, a weighted cross-entropy loss function is applied. Optimization endeavours involve fine-tuning parameters

and augmenting the training Epoch to enhance overall outcomes.

The segmentation methodology employs a deep convolutional model for accurate delineation. Initial steps involve pre-processing to remove noise, such as black frames and unwanted hair. The model utilizes a deep separable convolution for feature extraction. To address data imbalance, a weighted cross-entropy loss function is implemented. Optimization strategies include tuning parameters and increasing training Epoch for improved results.

Segmentation through Deep Convolutional Models involves a two-fold process: Initial Processing (3.1) and the application of Deep Convolutional Models (3.2). The Initial Processing stage encompasses preparatory steps that often involve data cleaning, normalization, and feature extraction to ensure the input data is primed for the subsequent segmentation model. Deep Convolutional Models (DCMs), in section 3.2, form the core of this segmentation methodology.

3.1 Preprocessing

Variations in dermoscopy image quality arise from factors such as skin condition, acquisition methods, and environmental influences. These disparities introduce disturbances like black frames and hair, which obscure lesions and impact the precision of their extraction. To address this issue, the enhancement process involves eliminating black frames by identifying dim areas within the image and dealing with hair through a series of procedures. Identification of black frames hinges on their luminance in the image's HSL space, followed by their removal. Hair removal entails converting the image into LUV space, employing morphological techniques to detect hair, and managing either widespread or sparse hair patterns using a set threshold. This involves utilizing differential calculations or bilinear interpolation to refine the image and eliminate hair-related interference.

3.1.1 Noise Reduction

When collecting dermoscopy images, the quality of dermoscopy images will be uneven due to differences in skin quality, disease appearance, acquisition operation process and collection environment of different patients. The differences in the acquisition operation process and acquisition environment lead to a large number of black frames and spot noise in the image, which in turn leads to the deviation of the extraction of the lesion area. When the lesion occurs in the armpits, scalp and other hairy areas, the acquired images contain a large amount of hair, resulting in the lesion being severely occluded. Hence, to ensure the precision of lesion feature extraction, the elimination of black frame and hair noise from dermoscopic images was carried out, aiming to acquire more accurate lesion information.

3.1.2 Black Frame Removal

The black frame of the dermoscopic image refers to the band of pixels with low brightness around the image caused by improper operation during image acquisition. It can be seen that the black frame itself has the characteristic of being less brighter than other areas, so the black frame can be positioned according to the brightness. To eliminate the black frame, begin by converting the image into the HSL color space, then identify the black frame based on the luminance component, L.

$$L = \frac{\max(R+G+B) + \min(R+G+B)}{2} \quad (1)$$

Among them, R, G, and B are the three components of the RGB

3.1.3 Hair Removal

Natural noises such as hair and blood vessels cannot be avoided by image acquisition technology, but their presence has caused great interference to the accurate segmentation of lesions, so it is necessary to remove hair from the image.

The main steps of the hair removal algorithm are as follows.

Step 1: Convert the image with the black frame noise removed from the RGB space to the LUV space.

Step 2: Perform morphological closed operations on the three components of L, U, and V respectively to obtain image MR.

$$\begin{aligned} L' &= L \cdot B = (L \oplus B)!B \\ U' &= U \cdot B = (U \oplus B)!B \\ V' &= V \cdot B = (V \oplus B)!B \\ MR &= (L', U', V') \end{aligned} \quad (2)$$

Among them, $\oplus$ denotes morphological expansion operation, ! Indicates morphological corrosion operations, and B indicates structural elements.

Step 3: Perform a differential image before and after the closure operation, and perform a trough detection to obtain a hair image CR.

Step 4: Perform area connectivity processing on the CR image and count the number of hairs in the connected area N.

Step 5: If N>T (T is the threshold used to distinguish between coarse and sparse hair patterns), then go to the coarse hair processing mode, use partial differential calculation to repair the hair parts, and set the initial image to u0(i,j), using Eq. (3) to invert the image. Otherwise, go to step 6.

$$u^{t+1}(i,j) = u^t(i,j) + \frac{\lambda}{n} \sum_{p \in D} c(\nabla^t u(i,j)) \nabla u^t(i,j) \quad (3)$$

where c(x) represents the diffusion coefficient function, and the common form is

, (i,j) represents the pixel coordinates, D represents the neighborhood of the pixel (i,j), λ represents a smoothness coefficient constant, n represents the number of pixels in the neighborhood, and t represents the number of iterations.

Step 6: If N<T, then go to the sparse hair processing mode, use bilinear interpolation to repair the hair information to remove the natural hair noise, by calculating the 4 adjacent pixels in the original image, output the value of each pixel of the target image, the target image is to remove the hair image.

$$u(i,j) = u(x_1, y_1)(x_2 - i)(y_2 - j) + u(x_2, y_1)(i - x_1)(y_2 - j) + u(x_1, y_2)(x_2 - y)(j - y_1) + u(x_2, y_2)(i - x_1)(j - y_1) \quad (4)$$

where u(i,j) represents the value of each pixel in the output image, (x1,y1), (x2,y1), (x1,y2), (x2,y2) indicates the 4 pixels adjacent to each other in the original image.

3.2 Deep Convolution Models

Convolution procedures are used in image processing to pick various filtering functions and extract features from inputs. When performing convolution operations, the convolution kernel is thought of as a filter, or a filter function, that employs various filter functions to extract various properties from the input. Nevertheless, the conventional convolution operation's order of magnitude of parameters will grow exponentially with the number of network layers, adding a significant amount of computation to the deep network model and impeding its ability to become more computationally efficient.

3.2.1 Depth-separable Convolutions

The introduction of deep separable convolution addresses the challenge posed by a high volume of parameters in standard convolutions. Standard convolutions couple the image across both channel and spatial dimensions. In contrast, depth separable convolutions distinctively handle the channel and spatial dimensions separately. Initially, they conduct convolution within the channel dimension, followed by convolution within the spatial dimension.

A. Standard Convolution

In the standard convolutional shown in Figure 1, the convolutional layer has 4 filters, each filter has 3 cores, and the size of each core is 3×3, then the number of parameters in this convolutional layer is Nstd=4×3×3×3=108.

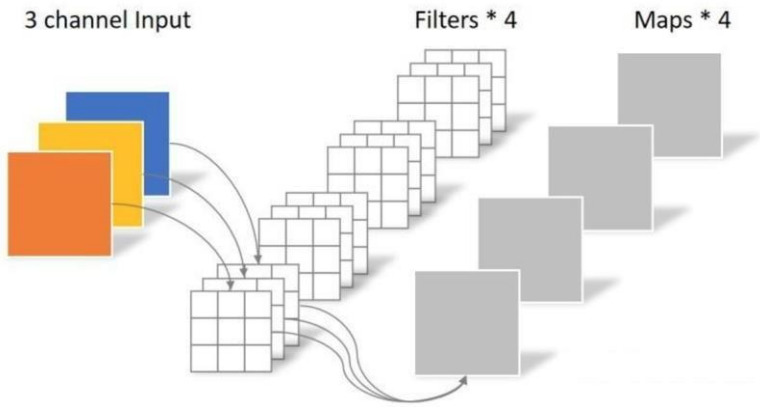


Figure 1. Standard convolution

B. Channel Dimension Convolution

In the channel dimension convolution shown in Figure 2 (a), the number of filters is the same as the depth of the input data. Therefore, the 3-channel image is passed through 3 filters to generate 3 feature maps, and the number of parameters is Ndepthwise=3×3×3=27. Therefore, it is necessary to combine the feature maps of the channel dimension to generate a new feature map by performing the convolution of the spatial dimension, and the convolution kernel is 1×1× M, and M is the depth of the input data. In this case, the number of filters determines the number of output feature maps, as shown in Figure 2 (b).

C. Spatial Dimension Convolution

In the spatial dimension convolution shown in Figure 3, the number of parameters is Npointwise=1×1×3×4=12. The final output has 4 feature maps, which is consistent with the standard convolution. The whole segregated convolution parameter consists of two parts: the convolution parameter of the channel dimension and the convolution parameter of the spatial dimension, with a total of 39 parameters.

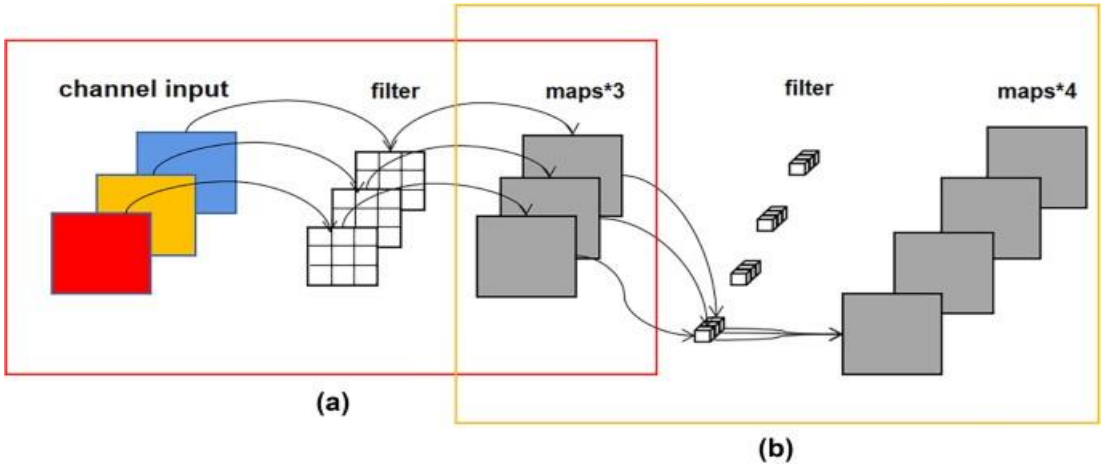


Figure 2 (a) Channel-dimensional Convolution (b) Spatial Dimension Convolution

For the same input and output, the parameter magnitude of the depth separable convolution is only 1/3 of the standard convolution, so under the same parameter magnitude, the deep separable convolution can be used to build a deeper network structure. In this paper, a feature extractor is constructed by using deep separable convolution, including two standard convolutional layers with 3× kernels and 3 deep separable convolutional layers with 20× 3 kernels, and their structures are shown in Figure 2 (b).

3.2.2 Dilated Convolution

In deep convolutional neural networks, although the standard convolution operation can extract effective features, it also faces the problem that the spatial resolution is greatly reduced with the increase of depth. In addition, multi-level convolution and pooling operations will lead to the loss of internal data structures. As a result, convolutional neural networks for image semantic segmentation have to face an embarrassing problem, that is, the shallow neural network cannot obtain enough effective features, while the deeper neural network has the problems of data structure loss and image resolution reduction. To address this issue, the paper employs void convolution during the feature extraction process. Void convolution, appropriately named, introduces voids into a standard convolution by incorporating a hyperparameter called the expansion rate. This form of dilated convolution enlarges the receptive field without compromising information loss through pooling, achieved by augmenting voids. By adjusting the expansion rate, each convolution outcome encompasses a broader scope of information, ensuring a more comprehensive range within the results.

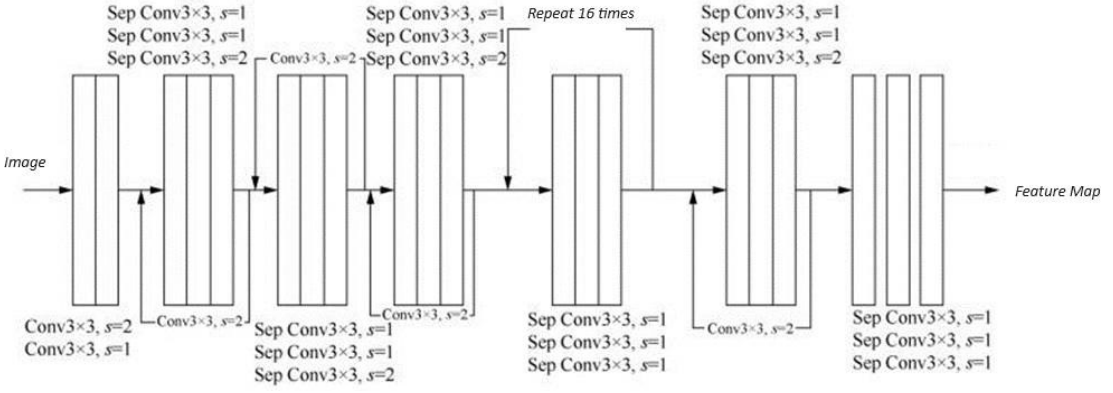
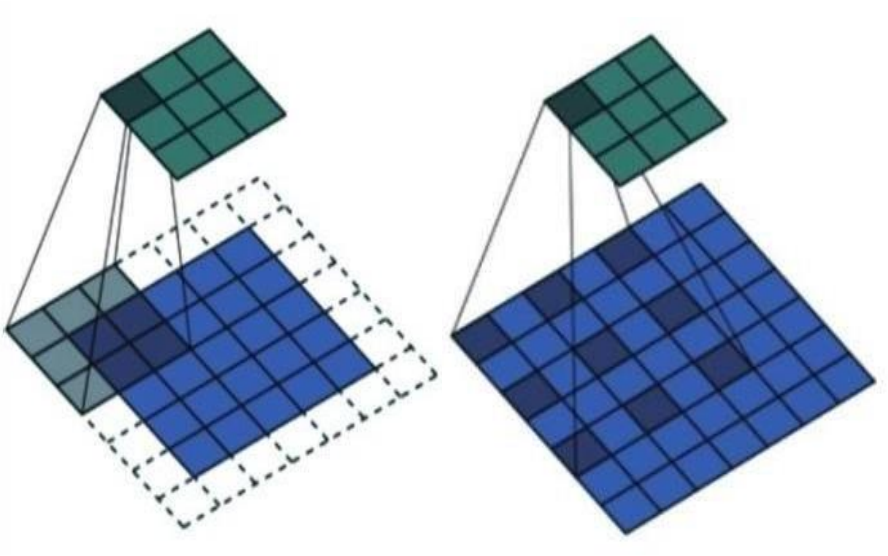


Figure 3. Feature Extraction using Depth-separable Convolution

As depicted in Figure 4, the standard convolution employs a 3x3 convolution kernel, while the void convolution also uses a 3x3 kernel with an expansion rate of 2. Contrasting with the original 3x3 convolutional kernel, the void convolution with an expansion rate of 2 generates a receptive field of 5x5 by incorporating holes. This demonstrates that adjusting the expansion rate allows for obtaining a larger receptive field compared to the standard convolution's original 3x3 kernel, thereby broadening the scope of information captured.



(a) Standard Convolution (b) Atrous Convolution

Figure 4. Dilated Convolution

The actual convolutional kernel size of a dilated convolution is:

kernel=k+(k−1)(r−1) (5)

where k is the size of the convolution kernel of the original standard convolution, and r is the expansion rate of the void convolution.

The size of the receptive field of the dilated convolution is:

$n-1$

$rn = rn-1 + (\text{kernel} - 1) \times \mathrm{G}\, si$ (6)

$i=1$

Where, rn is the receptive field of this layer of the network, rn-1 is the receptive field of the upper network, si is the step size of the layer i network convolution operation, and kernel is the actual convolution kernel size of the hole convolution, which is calculated from Eq. (5).

For dilated convolution, the middle part of the convolutional kernel contains a large number of zero elements, and when convolution operations are implemented using such a convolution kernel, the resulting output loses the continuity of information. To solve this problem, a spatial pyramid difference pool structure is used. First, the standard convolution with a kernel of 1×1 is used to reduce the number of channels, and then the multi-scale information is obtained based on the pyramid model of the dilated convolution, and the frame structure is shown in Figure 5. Here, the idea of feature stratification is used to layer the output features of the void convolution with different expansion rates, and the features of different receptive fields are fused to improve the information discontinuity caused by the hole.

The paper employs a pyramid pooling model that relies on dilated convolutions to extract more intricate features from the feature map generated by the deep separable convolution-based feature extractor. This pyramid structure comprises a standard convolution with a single 1x1 kernel and three kernels sized 1x3. Additionally, it integrates dilated convolutions using expansion rates of 3, 3, and 6 respectively, alongside a pooling layer.

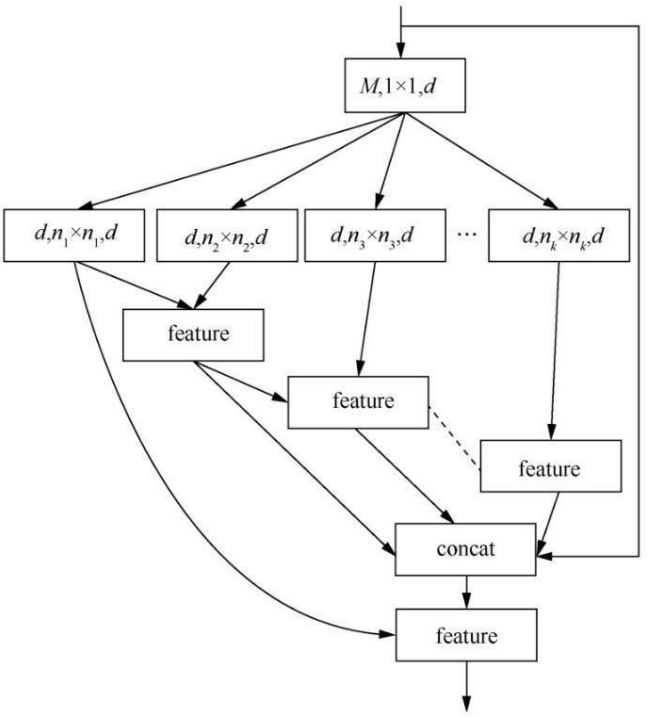


Figure 5. Framework of Pyramid Pooling Model based on Dilated Convolution

3.2.3 Codec Model

The codec model (Wang et al., (2023) is a model that uses two network modules to deal with sequence-to-sequence problems in the entire network structure, and its mathematical model is shown in Figure 6. In simple terms, it is to generate an output sequence based on the input sequence. Encoding is the sequence of inputs {x1, x2, ⋯, xT}; Convert to a vector of fixed length C； Decoding is the vector that will be encoded C Convert to an outputsequence{y1,y2,⋯,yT′}.

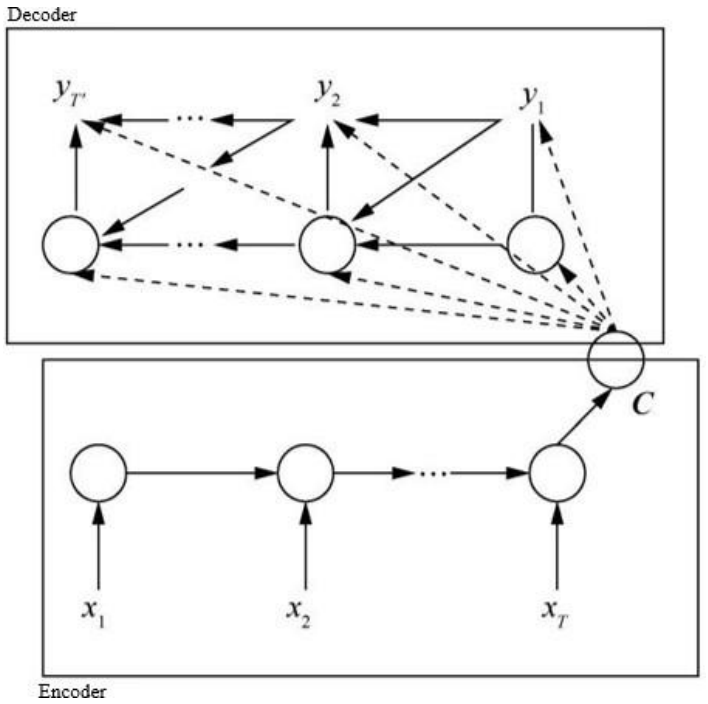


Figure 6. Codec Model

In this lesion segmentation task, the coding part is composed of two parts: a deep separable convolution module and a spatial pyramid module based on empty convolution. The original image is subjected to a low-level feature and a high-level feature by a depth-separable convolution module, the low-level features are directly output to the decoding module, and the high-level features are processed by the void space pyramid to obtain multi-scale features, and the features are subjected to a convolution operation with a kernel of 1×1 to reduce the number of channels, and then sent to the decoding part. The decoding part gradually recovers the detail information and its spatial dimension of the image through the information obtained from the encoded part, which is composed of a standard convolutional layer with a kernel of 1×1, a feature fusion layer, a standard convolutional layer with a kernel of 3×3 and an upsampling layer. The decoding module reduces the dimension of the low-level features from the depth- separable convolution through the convolution operation with a kernel of 1×1, and upsamples the received multi-scale features with bilinear difference. After the two parts of features are fused, a convolution operation with a kernel of 3×3 is performed, and the output result is oversampled.

3.2.4 DICE Loss Function

In the intelligent processing of medical imaging, it is inevitable to face the problem of unbalanced data samples. To address the challenge of data imbalance, prevalent solutions primarily involve data resampling and cost-sensitive learning. Data resampling modifies the distribution of data samples through techniques like oversampling or undersampling, while cost-sensitive learning is implemented by configuring a loss function to assign varying weights to different classes based on their prevalence.

The Dice coefficient is similar is a measure of the similarity between two sample sets and is defined as follows:

$$\text{similar} = \frac{2|X' \cap Y'|}{|X'| + |Y'|} \quad (7)$$

there into $|X' \cap Y'|$ It's a collection X' and Y' the intersection of, $|X'|$, $|Y'|$ Represents the number of elements in sets X and Y, respectively.

As can be seen from Eq. (7), the Dice coefficient is discrete. To be well-suited for the partition task, continuity is essential, leading to the definition of the Dice loss function.:

$$\text{similar} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (8)$$

where X is the prediction graph, Y is the label, and $|X \cap Y|$ is defined as the dot product between the prediction graph and the label, and the result of the dot product is added to give $|X|$ and $|Y|$ is defined as the sum of squares of the elements.

To prevent the occurrence of a special case where $|X|$ and $|Y|$ are both 0 and to mitigate the overfitting phenomenon, a smoothing operator is incorporated into the equation. (8), i.e.:

$$\text{Dice Loss} = 1 - \frac{2|X \cap Y|+1}{|X|+|Y|+1}$$

4. Experimental Results and Discussions

4.1 Datasets

In this project, the ISIC2018 dermoscopy dataset [Noel et al., (2018)] is used as the training set, and the PH2_dataset dataset is used as the validation set. The ISIC2018 dataset contains 2594 dermoscopic original images and 2594 binary label images, which include actinic keratoses and skin diseases such as intraepithelial tumors, benign keratosis, basal cell carcinoma, squamous cell carcinoma, dermatofibroma, melanoma, moles, vascular lesions, etc. The PH2_dataset holds 200 dermoscopic images alongside 200 corresponding label images, comprising 160 instances of moles and 40 cases of melanoma. These images possess unique traits that make their analysis and classification challenging. Notably, the dataset grapples with significant interference primarily caused by hair and blood vessels, often obstructing the targeted lesion. Additionally, a notable imbalance exists in the data, with some cases showing a minimal lesion area, occasionally less than 1% of the entire image, while others exhibit a maximum area of 98%. This skewed distribution complicates the training of models and requires careful consideration. Furthermore, the boundary between the lesion area and the background appears blurred, presenting a further obstacle in accurately distinguishing and delineating the target lesions from the surrounding context. Addressing these challenges is imperative for developing robust and effective algorithms for dermoscopic image analysis and diagnosis.

4.2 Lesion Segmentation Results

In the field of deep learning, the amount of data used in this paper is very small, PH2_dataset and it is time-consuming and laborious to directly use these datasets to train deep learning models from scratch, and the models obtained in this way are also prone to overfitting and affect the scalability of the model. Therefore, in the actual experimental process, this paper uses the transfer learning method to train the segmentation model, and uses the pre-trained network model weights on the large dataset to complete the model training on the dermoscopic dataset.

Therefore, this paper uses the Dice loss function as the cost function, and assigns different weight ratios to the lesion area and the background region, setting the proportion of the lesion region to 0.7 and the background region 0.3.

The network training algorithm used in the experiment is the Adaptive Momentum Estimation Algorithm (ADAM Algorithm) [Reyad et al., (2023)] with fast convergence speed, and the training process adopts 20 Epoch training, the Batchsize is set to 4, the learning rate (Lr) is 0.01, the training accuracy is stable at 86.71%, and the test accuracy is stable at 85.01%. The experimental results are shown in Figure 10.

Furthermore, to optimize the training process to the fullest extent, the model's parameters are fine-tuned and compared, accompanied by an increase in the training Epoch. The outcomes of these adjustments are presented in Table 1.

During the training process, the training speed decreases slowly with the increase of Batchsize, and when Batchsize=12, the training speed of the segmentation model decreases significantly compared with that of Batchsize=4, and memory overflow occurs. The training speed decreases because the construction of the segmentation model in this paper uses a deeper convolution operation, and the depth of the feature map output by each level gradually increases as the layers go deeper, and on this basis, increasing the Batchsize will reduce the training speed.

Table 1 Parameter Tuning

Comparison items			Epoch		
			10	20	40
Training	Batchsize=4	Lr=0.01	86.71%	90.52%	92.54%
		Lr=0.05	89.45%	92.14%	94.65%
	Batchsize=8	Lr=0.01	83.04 %	85.94%	90.81%
		Lr=0.05	84.57%	92.24%	95.24%
Testing	Batchsize=4	Lr=0.01	85.04%	88.03%	90.45%
		Lr=0.05	80.60%	84.36%	89.23%
	Batchsize=8	Lr=0.01	83.04%	90.82%	93.82%
		Lr=0.05	87.60%	91.03%	92.96%

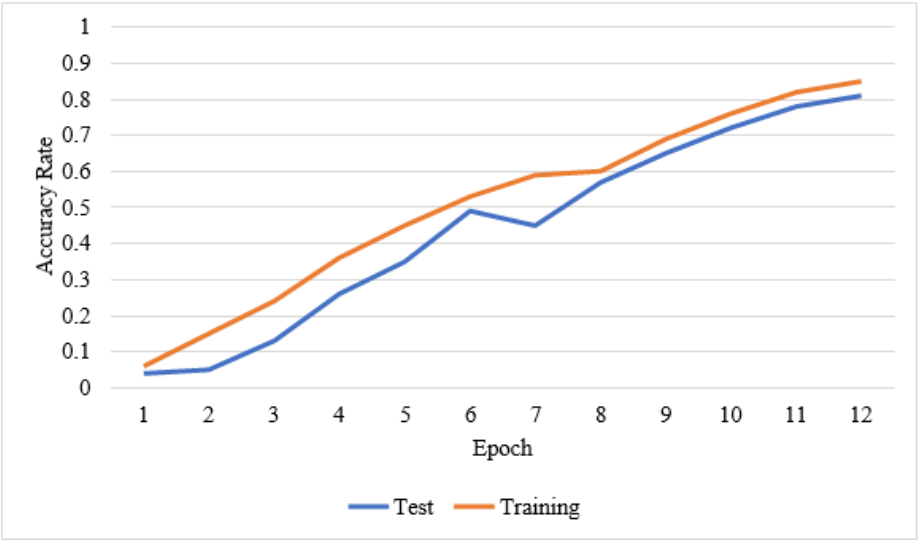


Figure 7. Training and Test Results of the Proposed Model

The study employed the U-Net segmentation model to train the research object, utilizing a training configuration of 20 Epochs, a Batchsize set to 8, and a learning rate of 0.05.

Comparative analysis of experimental results was performed against the method introduced in this paper, and the findings are visually depicted in Figure 8.

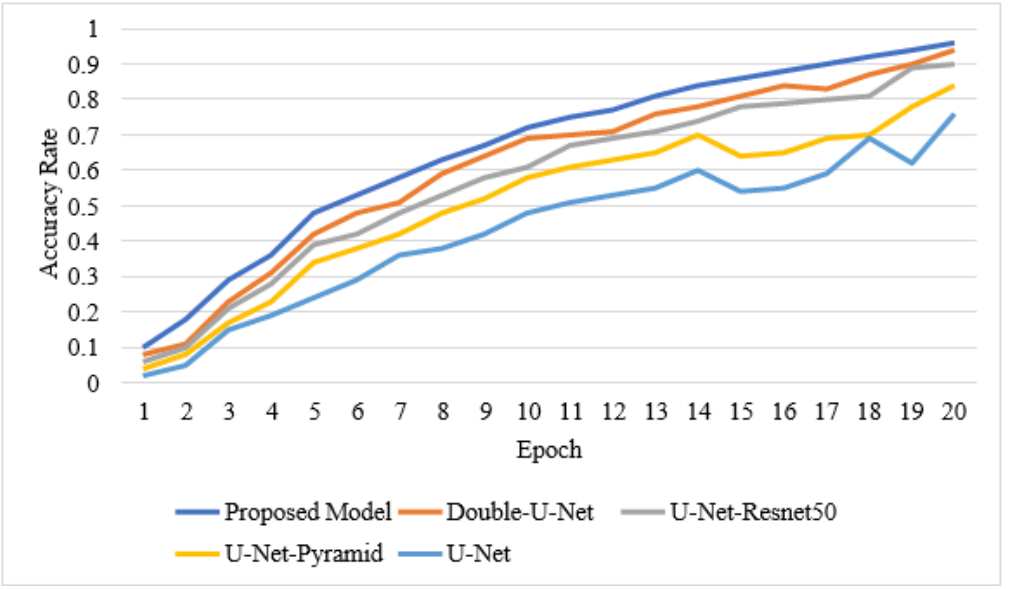


Figure 8. Training Results of the Proposed Model and other U-Net Models

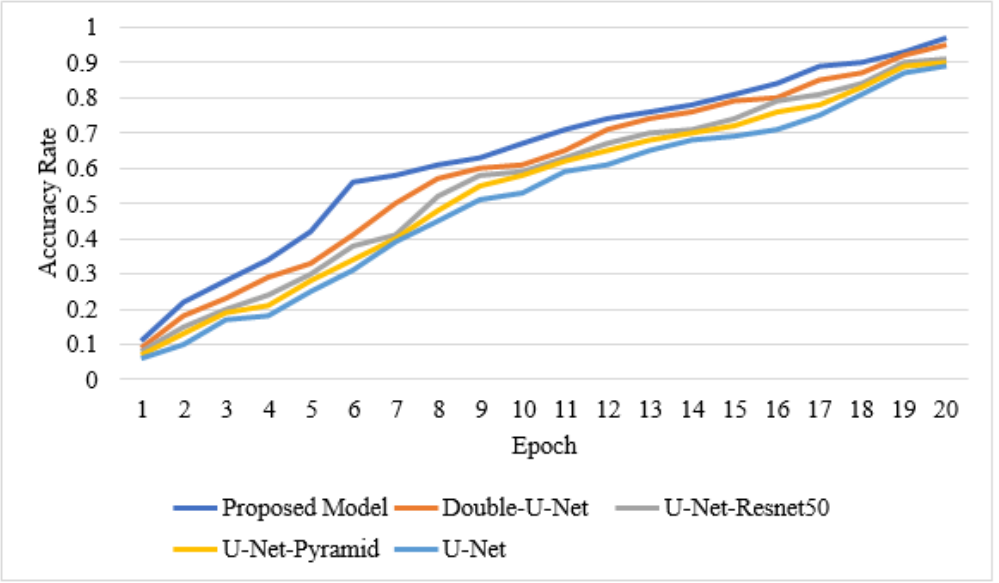


Figure 9. Test Results of the Proposed Model and other U-Net Models

The findings indicate that the codec model utilizing deep separable convolution achieves higher accuracy compared to the U-Net model. With identical parameter settings, after 20 epochs of training, the proposed method attains a training accuracy of 92.24% and a test accuracy of 90.82%. In contrast, the U-Net model achieves a training accuracy of 86.07% and a test accuracy of 84.71%. The experimental results chart in Figure 9 illustrates that, for the dermatology dataset, the proposed method can achieve a remarkable training accuracy of up to 95.24% and a test accuracy of up to 93.82%.

5. Conclusion

This paper centers on dermoscopic images, presenting an efficient noise reduction method following a thorough analysis of image characteristics. The approach tackles distinct noise elements by eliminating artificial noise like black frames and natural noise such as hairs. The segmentation model integrates depth separable convolution, dilated convolution, and a codec model to reduce parameter magnitude, acquire multi-scale features, and produce segmentation prediction maps. Empirical results showcase effective segmentation for dermoscopic images. However, challenges persist in handling noise resembling lesions, such as blood vessels, and in addressing slow training speeds for Batchsize ≥12. Future research should delve into disease identification, encompassing segmentation and identification of pathological dermatological images for a comprehensive medical auxiliary system.

References

[1]. Aijaz S F, Khan SJ, Azim F, Deep learning Application for Effective Classification of different types of psoriasis[J]. HealthcEng, 2022,2022:7541583.

[2]. Cui W C, Zhang P X, Shao H. Dermoscopic image lesion segmentation method based on deep

separable convolutional network[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 385-393.

[3]. Cui W C, Zhang P X, Shao H. Dermoscopic image lesion segmentation method based on deep separable convolutional network[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 385-393.

[4]. Espsito M C C, Espsito A C C, Jorge M F S, et al. Depression, anxiety, and self-esteem in women with facial melasma: an internet-based survey in Brazil [J]. International Journal of Dermatology, 2021, 60 (9): e346 - e347.

[5]. Esteva A, Kuprel B, Novoa R A, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. Nature, 2019, 542(7639): 115-11811.

[6]. Esteva A, Kuprel B, Novoa R A,Dermatologist level classification of Skin Cancer with Deep Neural Networks[J]. Nature,2019, 542 (7639) :115 -118.

[7]. Goyal A,Bochkovskiy A,Deng J, Non-deep networks[J].Advances in Neural Information Processing Systems, 2022,35:6789-6801.

[8]. Jiang Xinhui, Li Zhe. Research on skin lesion segmentation based on U-shaped structural context encoding and decoding network [J]. Progress in Lasers and Optoelectronics, 2021, 58(12): 122-129.

[9]. Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, Allan Halpern: "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)", 2018; https://arxiv.org/abs/1902.03368

[10]. Qi Yongfeng, Hou Lulu, Duan Youfang. Dermoscopic skin damage segmentation based on DenseNet-BC network [J]. Computer Engineering and Science, 2020, 42(6): 1060- 1067.

[11]. Reyad, M., Sarhan, A. M., & Arafa, M. (2023). A modified Adam algorithm for deep neural network optimization. Neural Computing and Applications, 1-18.

[12]. Schneider S L, Kohli I, Hamzavi I H, Emerging imaging technologies in dermatology: part I: basic principles[J]. Journal of the American Academy of Dermatology, 2019, 80(4): 1114-1120. DOI：10.3788/CJL221241.

[13]. Tian M. Image multi label classification within medical literature based on migration learning[D]. Wuhan: Wuhan University, 2019

[14]. Wang J P, Chen M H, Tan W J, Dual-modality endoscopic probe for optical coherence tomography imaging and pH sensing[J]. Chinese Journal of Lasers, 2020, 47(9): 0907001, DOI：10.3788/CJL221241.

[15]. Wang Q L WuB G ZhuP FECA-Net efficient channel attention for deep convolutional neural networks C2020IEEE CVF Conference on Computer Vision and Pattern Recognition CVPR June13-19 2020 Seattle WA USA New York IEEE Press 2020 11531-11539.

[16]. Wang Xue. Skin lesion segmentation method based on U-Net multi-scale and multi- dimensional feature fusion [J]. Journal of Jilin University (Science Edition), 2021, 59(1): 123-127.

[17]. Wencheng. Dermoscopic lesion segmentation method based on deep separable convolutional network. Chinese Journal of Science and Technology[J], 2020, 2(4): 385- 393.Wang, S., Du, H., Chen, W., & Qu, H. (2023, November). Abbreviated Weighted Graph in Multi-Agent Reinforcement Learning. In Pacific Rim International Conference on Artificial Intelligence (pp. 113-124). Singapore: Springer Nature Singapore.