*Electronic Engineering and Informatics*
*G. Izat Rashed (Ed.)*

*doi:10.3233/ATDE240070*

# Double-Coding Skin Disease Segmentation Based on Attention Mechanism

Guoliang YANG, Ziling NIE[1], Jixiang WANG, Hao YANG, Shuaiying YU
*School of Electrical Engineering and Automation Jiangxi University of Technology, Ganzhou, China*

**Abstract.** Aiming at the problems of complex structure and lack of feature information of skin lesions, an image segmentation method based on attention mechanism and double coding network is proposed. Firstly, the dual-coded branch network is used to extract image feature information to improve the ability of network information capture. Secondly, the dual attention module is used to encode the global context information, which suppresses irrelevant features and highlights relevant features. Finally, in the coding part, depth separable convolution and cooperative attention are used to reconstruct the image. Experimental results show that on ISIC2018 data set, the accuracy, Dice similarity coefficient and Jaccard index of this method reach 96.59%, 92.98% and 82.65%, respectively. Compared with other networks, the accuracy and boundary segmentation effect are obviously improved.

**Keywords.** Image segmentation; Double coding; Context information; Depth separable convolution; Attention

## 1. Introduction

Skin cancer is a common and fatal disease, which seriously threatens people's physical and mental health [1]. Studies have shown that if melanoma can be detected in time and treated in the early stage, the probability of cure can reach 95%, which shows that skin diseases are crucial in early diagnosis and treatment [2].

With the rapid development of deep learning technology, a large number of scholars have applied convolution neural networks to image segmentation networks to improve their performance [3]. The end-to-end U-Net network proposed by Ronneberger [4] has become the basic network structure in the follow-up medical segmentation field because of its outstanding performance. Gu et al. [5] proposed a contextual coding network based on U-Net, which can preserve the spatial information of two-dimensional medical image segmentation. Wang et al. [6] introduced multiple attention into CNN architecture and proposed a comprehensive attention network based on CA-Net. Chen et al. [7] proposed the TransUNet network to obtain a global field of view by using a self-attention mechanism to divide images into small pieces, which performs well in medical image segmentation. Although the above skin lesion segmentation algorithms can achieve good

[1] Corresponding author: Ziling NIE, School of Electrical Engineering and Automation Jiangxi University of Technology, e-mail: 2420438897@qq.com

segmentation results, they can only produce limited receptive fields and cannot capture global features.

In this paper, a dual-branch coding network is proposed, which can accurately extract the image feature information through the dual-branch model at the coding end, and then re-register the features in space and channel position through the dual-attention module. Then, cooperative attention and depth separable convolution are introduced into the decoder to reduce the information loss caused by feature map dimensionality reduction, and the image reconstruction is completed to obtain the segmented image.

## 2. Network Construction

### 2.1. Network Overview

The network structure diagram of skin disease segmentation model proposed in this paper is shown in Figure 1. For the input skin lesion image, the feature is extracted by a dual-branch encoder, and then the two feature images are fused. The fused feature images are re-registered by a dual attention module composed of a spatial attention module and a channel attention module, respectively. Cooperative attention is introduced into the decoder and Depthwise Separable Convolution is used to restore the image structure. Finally, the segmented image is obtained by Sigmoid function.

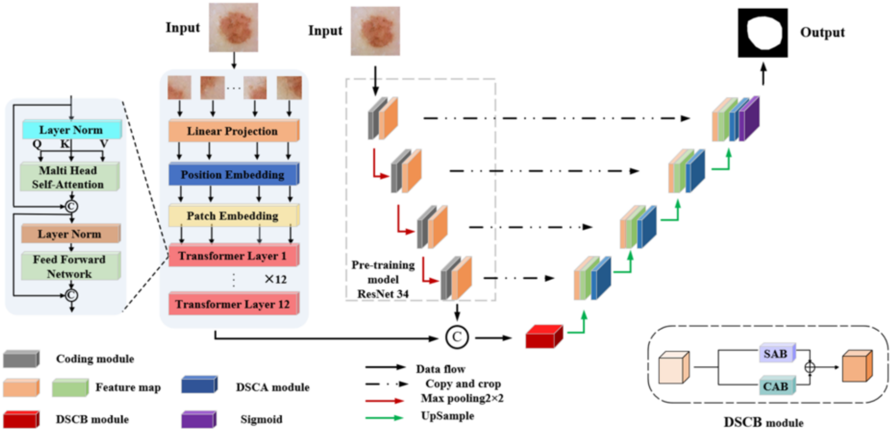


**Figure 1.** Overall network model architecture

### 2.2. Dual branch encoder module

This paper presents a dual-branch encoder structure, which consists of Transformer encoder branch and CNN encoder branch, as shown in Figure 1. In skin lesion segmentation, not only rich local features can be extracted, but also global context information can be captured. The main components of the transformer encoder branch include feature map sequence, position coding, and transformer layer [8]. As the core component of Transformer layer, MSA is mainly responsible for connecting each element in the deep-level feature diagram to get the global view. The query matrix (Q), key matrix (K), and value matrix (V) are the three inputs accepted by the Self-Attention module.

The encoder branch of CNN uses pre-trained ResNet34 to replace the encoder in U-Net as the backbone network of feature extraction. Compared with the decoder in U-Net architecture, this module subtracts the full connection layer and the average pooling layer, and retains the first four feature extraction blocks as the feature coding blocks of CNN coding branches. ResNet 34 can effectively avoid network degradation caused by repeated convolution and pooling operations, reduce operation costs and improve running speed [9].

### *2.3. Dual attention module*

The dual attention module is mainly composed of spatial attention module (SAB) and channel attention module (CAB), and its structure is shown in Figure 2. The dual-attention module mainly re-registers the input feature map in spatial dimension and channel dimension, while the SAB module establishes efficient context connection based on local features, and improves the representation ability of these features by encoding global information into local features. CAB module, using available space to improve features and highlight important features. For the input feature map $F \in R^{C\times H\times W}$, after it passes through SAB and CAB respectively, the output feature map is fused to get the final output feature map $F' \in R^{C\times H\times W}$.

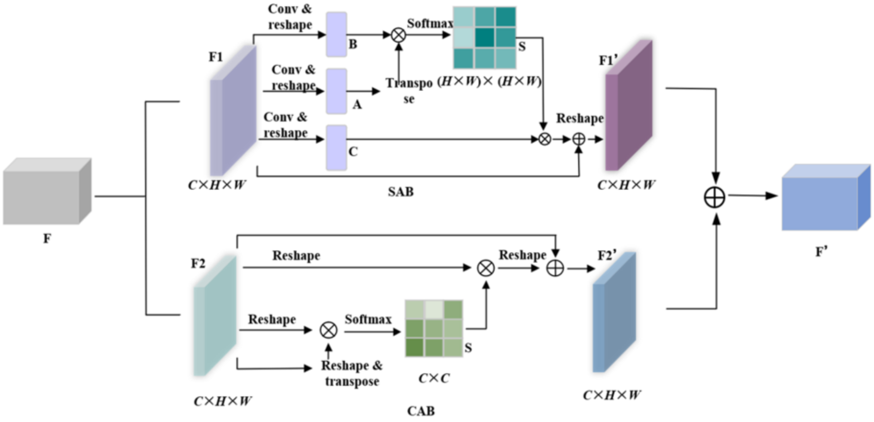
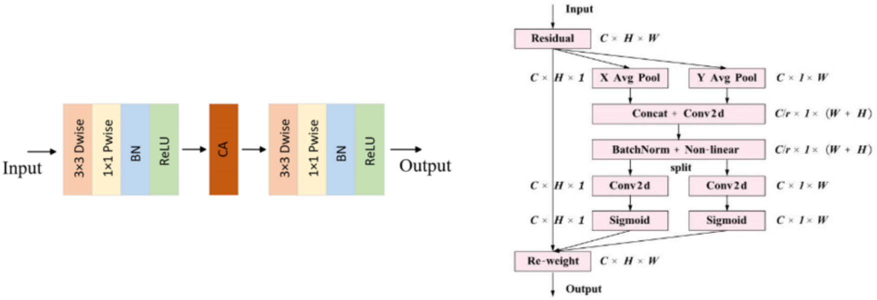


**Figure 2.** Dual attention module

### *2.4 Decoding the DSCA module*

In order to obtain more spatial feature information, the ordinary convolution of the decoded part is replaced by deep separable convolution. At the same time, the introduction of collaborative attention [10] at the decoding end enables the feature map to maintain high resolution in channel and space, pay more attention to the feature information of the segmented area. The DSCA module is shown in Figure 3(a).



(a)DSCA module (b)CA Attention Module

**Figure 3.** DSCA module and CA Attention Module

CA attention[11] is shown in Figure 3(b), which include embedding coordinate information and generating coordinate attention. Information embedding adopts formula (1), which is transformed into one-dimensional feature coding, and then uses pooled kernels with lengths and widths of (H, 1) and (1, W) to code each channel in horizontal and vertical directions, obtaining a pair of feature maps perceived in spatial directions.

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j) \tag{1}$$

where $z_c$ is the variable channel $c$ output, and the height and width of the feature map are $H$ and $W$ ; The coordinate value for $c$ channel is $X_c(i,j)$ .The second stage coordinates attention generation operation. The transformation results of the first stage are first fused, and then the transformation operation uses the 1×1 convolutional transformation function F1 to achieve.

## 3. Experimental Results and Analysis

### *3.1. Experimental environment and Dataset*

All experiments are completed by Windows operating system and Pycharm simulation platform. The computer operating system is 64 bits, the processor is Intel CoreTM i9-9900CPU. In this paper, ISBI2018 data set is used as experimental sample, including 2594 images of skin lesions, which are randomly divided into training set, test set and verification set according to the ratio of 7: 2: 1.

### *3.2. Evaluation Indicators*

In this paper, the common indicators in the field of medical imaging Accuracy (Acc), Dice Similarity Coefficient (Dice), Jaccard Index (Jac), Specificity (Spe), and Sensitivity (Sen) are used to evaluate the performance of this model and the comparison model. The above definitions are as follows

$$Acc = \frac{TP + TN}{TP + TN + FN + FP} \tag{2}$$

$$\text{Dic}=\frac{2TP}{2TP+FN+FP} \tag{3}$$

$$\text{Jac}=\frac{TP}{TP+FN+FP} \tag{4}$$

$$\text{Spe}=\frac{TP}{TN+FP} \tag{5}$$

$$\text{Sen}=\frac{TP}{TP+FN} \tag{6}$$

where, $T_P$, $T_N$, $F_p$, $F_N$ are true positives, true negatives, false positives, and false negatives, respectively.

### *3.3. Ablation experiment*

In order to evaluate the impact of each module in this algorithm on network performance, eight groups of ablation experiments were designed. Test 1 is U-Net network. Test 2 is a network with two-branch coding. Test 3 contains DCSB module. Test 4 contains DSCA module. Test 5 is the fusion of two-branch coding and DCSB. Test 6 is the fusion of two-branch coding and DSCA. Test 7 is the integration of DCSB and DSCA module. Test 8 is the complete algorithm of this paper. The experimental results are shown in Table 1.

**Table 1.** Performance comparison of different network structures

| Network | ACC | Sen | Spe | Dice | Jac |
|---|---|---|---|---|---|
| Test 1 | 94.44% | 88.59% | 93.39% | 88.66% | 77.24% |
| Test 2 | 95.63% | 88.12% | 96.44% | 92.39% | 82.12% |
| Test 3 | 95.34% | 88.27% | 93.15% | 88.82% | 77.74% |
| Test 4 | 95.55% | 89.04% | 97.14% | 92.97% | 82.65% |
| Test 5 | 96.35% | 87.17% | 97.36% | 92.94% | 82.34% |
| Test 6 | 96.42% | **89.67%** | 96.71% | 92.79% | 82.07% |
| Test 7 | 96.38% | 89.63% | 97.39% | 92.15% | 82.41% |
| Test 8 | **96.59**% | 88.63% | **97.43**% | **92.98**% | **82.65%** |

From the analysis of Table 1, it can be seen that the results of each evaluation index of U-net are low. The results of Test 2 and Test 4 are higher than that of Test 1, which shows that dual-branch coding can capture more complete feature information. The results of Test 3 show that the DCSB module can effectively suppress irrelevant features. Test 5 ~ 7 are two modules combined with each other. From the results in the table, it can be seen that the lack of one module will degrade the performance of the model. The results show that the proposed dual-branch encoder can achieve a good feature extraction, and the DSCA module can accurately restore the image after feature extraction, further improving the network performance.

### 3.4. Comparative experiments

In order to verify the reliability and effectiveness of this network, it is compared with the classical algorithm. Table 2 shows the experimental results of each network, which is best represented in bold.

**Table 2.** Segmentation results of different networks

| Network | Acc | Sen | Spe | Dice | Jac |
|---|---|---|---|---|---|
| U-Net | 94.44% | 88.59% | 93.39% | 88.66% | 77.24% |
| DeeplabV3+ | 95.87% | 88.16% | 97.37% | 92.74% | 82.43% |
| CE-Net | 95.13% | 87.47% | 97.01% | 91.42% | 81.62% |
| CA-Net | 95.05% | **88.88%** | 97.15% | 92.16% | 81.15% |
| CPF | 95.62% | 83.44% | 96.17% | 91.64% | 82.10% |
| Dense ASPP121 | 95.39% | 87.99% | **97.78**% | 90.93% | 80.92% |
| Ours | **96.59**% | 88.63% | 97.43% | **92.98**% | **82.65**% |

From Table 2, compared with U-Net network, the proposed algorithm improves by 2.15%, 4.04%, 4.32% and 5.14% on the four evaluation indexes of Acc, Spe, Dice and Jac, respectively. At the same time, compared with other models, the algorithm proposed in this paper performs best in the Acc, Dice and Jac indicators. The comparative results show that the overall index performance of the proposed algorithm is better than that of the comparison network. and the segmentation ability is better.

Figure 4 is a visualization diagram of each network division. Comparing the segmentation results, we can find that the edge processing of lesion area in this algorithm is more detailed. U-Net network can roughly segment the lesion area, but it is easy to lose edge feature information. CE-Net and DeepLabv3 + are incomplete in edge detail segmentation, and there is a phenomenon of missing lesion area. CPF-Net has insufficient ability to extract edge information, and the contour segmentation of lesion area is poor. The results show that the segmentation of lesion area by this network is more complete and accurate.

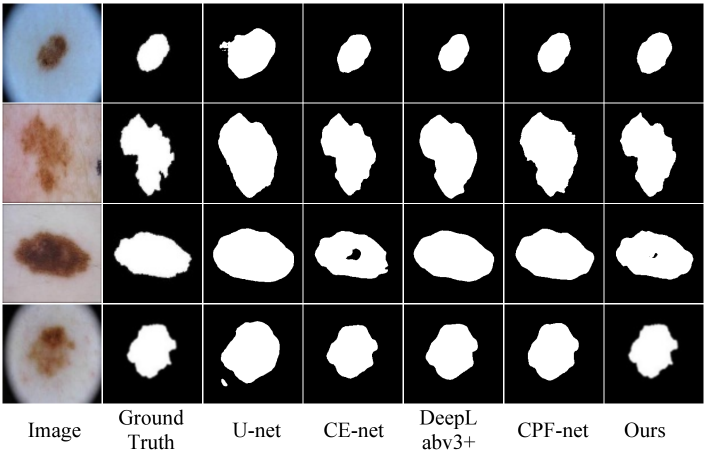


**Figure 4.** Experimental results of different networks

## 4. Conclusion

In this paper, a skin disease image segmentation method based on dual-branch coding is proposed. The feature extraction of input image is carried out by Transformer and CNN dual-branch encoders, which improves the feature extraction ability of coding network. The decoder introduces CA attention and uses depth separable convolution to decode the feature map, which alleviates the problem of information loss in the up-sampling process. Experimental results show that: The proposed algorithm for skin lesion image segmentation effect is significantly improved, edge detail processing ability has been improved, and the comprehensive index is better than other algorithms, which is helpful to further improve the accuracy and efficiency of computer-aided diagnosis of skin lesions.

## Acknowledgment

This work was supported by the Science and Technology Project of Jiangxi Education Department (GJJ190450) and the Science and Technology Project of Jiangxi Education Department (GJJ180484)

## References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. , "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries". CA Cancer J Clin, vol. 03, no. 71, pp. 209-249, 2021. https://doi.org/10.3322/caac.21660

[2] Wei. Z, Shi. F, Song. H, Ji. W, Han. G , "Attentive boundary aware network for multi-scale skin lesion segmentation with adversarial training", Multimedia Tools and Applications, vol. 37, no. 79, pp. 27115-27136, 2020. https://doi.org/10.1109/ACCESS.2019.2940794

[3] Ramadan, Rania, and Saleh Aly. . "CU-net: a new improved multi-input color U-net model for skin lesion semantic segmentation". *IEEE Access*, vol. 10, pp. 15539-15564, 2022.https://dio 10.1109 / ACCESS.2022.3148402.

[4] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox , "U-net: Convolutional networks for biomedical image segmentation", Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference. Munich, Germany, pp. 234-241,2015. https://doi.org/10.1007/978-3-319-24574-4_28.

[5] Gu. Z, Cheng. J, Fu. H, Zhou. K, Hao. H , "CE-Net: context encoder network for 2Dmedical image segmentation". IEEE Transaction on Medical Imaging, vol.38, no. 10, pp.2281-2292, 2019. https://doi.org/10.48550/arXiv.1903.02740

[6] G. Ran, W. Guotai, and S. Tao, "CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation", IEEE transactions on medical imaging, vol. 40, no. 2,pp. 699-711, 2021. doi: 10.1109/TMI.2020.3035253.

[7] Chen. J, Lu. Y, Yu. Q, Luo. X, Adeli. E, "TransUNet: Transformers make strong encoders for medical image segmentation"[EB/OL].（2021-02-08）. https://doi.org/10.48550/arXiv.2102.04306

[8] D. Yumin, W. Lixing, "Application of CNN-Transformer Network in Dermatoscope Image Segmentation ",[J/OL]. https://kns.cnki.net/kcms/detail/50.1165.n.20221207.1824.017.html

[9] He. K, Zhang. X, Ren. S, Sun. J, .Deep residual learning for image recognition[C]//Proceedings of the IEEEConference on Computer Vision and Pattern Recognition,2016: 770-778. https://doi.org/10.48550/arXiv.1512.03385

[10] Sarker, M. Mostafa Kamal, and Rashwan, "SLSNet: Skin lesion segmentation using a lightweight generative adversarial network", Expert Systems with Applications, vol.183, 2021. https://doi.org/10.48550/arXiv.1907.00856

[11] Hou, Qibin, Daquan Zhou, and Jiashi Feng,"Coordinate Attention for Efficient Mobile Network Design", 2021. https://doi.org/10.48550/arXiv.2103.0290