World Journal of Information Technology
Print ISSN: 2959-9903
Online ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3088

IMPROVED DEEP LEARNING MODEL-BASED SKIN DISEASE IMAGE SEGMENTATION ALGORITHM

Tao Song
Sichuan Technology and Business University, Chengdu 611745, Sichuan, China.

Abstract: Aiming at the problems of complex background interference, low segmentation accuracy of small targets, blurred lesion boundaries and large scale differences in skin disease image segmentation, this paper proposes two improved deep learning models, namely the CAU-Net model fused with the coordinate attention mechanism and the CSwinTransU-Net model integrated with the spatial perception module. Experimental results on the ISIC2018 skin disease dataset show that the mean Intersection over Union (mIoU), mean Dice coefficient (mDice) and average accuracy (aAcc) of the CAU-Net model reach 85.50%, 92.06% and 93.69% respectively, and the corresponding indicators of the CSwinTransU-Net model are 84.84%, 91.66% and 93.49%. Both models outperform the basic models, effectively improving the accuracy and robustness of skin disease image segmentation, and providing technical support for the early diagnosis of skin cancer.
Keywords: Deep learning; Skin disease image segmentation; Coordinate attention mechanism; Spatial perception module; Feature fusion

1 INTRODUCTION

The rapid development of artificial intelligence technology has promoted the wide application of deep learning in the field of medical image segmentation [1]. As a key link in the early diagnosis of skin cancer, skin disease image segmentation is of great significance for improving diagnostic accuracy and reducing the rates of misdiagnosis and missed diagnosis. The incidence of skin malignant tumors is rising continuously. Although melanoma accounts for only 6.8% of skin malignant tumors, its mortality rate exceeds three-quarters of all skin cancer deaths. Early diagnosis can increase the five-year survival rate of patients to more than 90%. Traditional skin disease diagnosis relies on doctors' subjective experience, which has limitations in the identification of tiny nodules and the judgment of atypia. Deep learning, with its powerful feature extraction and pattern recognition capabilities, has become an effective way to solve the problems of skin disease image segmentation [2,3].

Current skin disease image segmentation faces numerous challenges, such as diverse shapes and sizes of lesion areas, many interfering factors like hair and blood vessels in images, uneven illumination, blurred lesion edges, easy missed detection of small target lesions, and unbalanced proportion of foreground and background pixels in datasets [4,5]. Traditional image processing methods are difficult to solve these problems effectively. Existing deep learning segmentation models still have shortcomings: the U-Net model has low accuracy in small target segmentation in complex scenes, and traditional attention mechanisms have limited ability to model long-range dependencies and position information; Vision Transformer has the problem of high computational complexity, TransU-Net has insufficient ability to segment blurred boundaries, and the fusion of local and global features is inadequate [6,7].

Therefore, this paper conducts research on the optimization of deep learning architectures and proposes two improved models: the CAU-Net model is designed to address the defects of the U-Net model, which introduces the coordinate attention mechanism to enhance the perception ability of target regions; the CSwinTransU-Net model is constructed to make up for the deficiencies of the TransU-Net model, which integrates the spatial perception module to improve the segmentation accuracy of blurred boundaries. The ISIC2018 dataset is used as the experimental dataset to verify the effectiveness of the improved models, providing a new solution for skin disease image segmentation [8].

2 RELEVANT THEORIES AND TECHNOLOGIES OF IMAGE SEGMENTATION

2.1 Basic Deep Learning Models

Convolutional Neural Networks (CNNs) are the basic models for medical image segmentation, consisting of an input layer, convolutional layers, activation function layers, pooling layers, fully connected layers and an output layer. They extract local features through convolution operations, reduce dimensions and maintain translation invariance through pooling layers, and introduce nonlinear feature transformation through activation function layers to solve the problem of insufficient expression ability of linear models. Classic CNN models such as AlexNet and ResNet have been continuously optimized, providing basic architectural support for medical image segmentation [1].

Generative Adversarial Networks (GANs) are composed of a generator and a discriminator, and realize sample generation through game training between them. The generator maps random noise into samples close to the real distribution, and the discriminator distinguishes between real and generated data. In medical image segmentation, GANs can generate multi-scale lesion images to augment datasets and improve the generalization ability of models [11].

2.2 Classic Segmentation Network Architectures

U-Net++ is an improved version of U-Net, which bridges the semantic gap between encoders and decoders through dense skip connections and deep supervision, improving the segmentation accuracy of edge details; Fully Convolutional Networks (FCNs) replace fully connected layers with convolutional layers, and realize pixel-level prediction by combining upsampling and skip connections, which can process input images of any size[9]; The Transformer architecture captures global dependencies based on the self-attention mechanism, and realizes effective modeling of cross-scale features through multi-head attention, feed-forward networks and residual connections, making up for the deficiency of CNNs in modeling long-range dependencies [10].

2.3 Attention Mechanisms and Evaluation Metrics

The attention mechanism focuses on key information by dynamically assigning weights. The classic Squeeze-and-Excitation (SE) module enhances the features of important channels through channel attention, and Convolutional Block Attention Module (CBAM) fuses channel and spatial attention to further improve the feature selection ability of models [12].

The quality evaluation of medical image segmentation adopts core indicators such as average accuracy (aAcc), mean Intersection over Union (mIoU) and mean Dice coefficient (mDice). Meanwhile, mean Asymmetric Surface Distance (mASD) and maximum Asymmetric Surface Distance (mMSSD) are used to evaluate the fitting degree between the segmented boundary and the real boundary, so as to comprehensively quantify the segmentation performance of models [6,7].

3 CAU-NET MODEL FUSED WITH COORDINATE ATTENTION MECHANISM

3.1 Model Design

Aiming at the limitations of traditional attention mechanisms in modeling long-range dependencies and the low segmentation accuracy of U-Net for small targets, this paper introduces the coordinate attention mechanism into the long skip connections of U-Net to construct the CAU-Net model[2,6]. The model retains the symmetric encoder-decoder structure of U-Net: the encoder extracts contextual features and compresses the spatial dimension through convolution and max pooling; the decoder restores the spatial resolution through transposed convolution upsampling, and splices the high-resolution features of the encoder with the features of the decoder to make up for the loss of spatial information caused by downsampling [8,9].

The coordinate attention module is the core improvement of the CAU-Net model, which adopts a dual-channel axial attention modeling strategy. It performs global average pooling on the input feature map along the X-axis and Y-axis respectively to extract global features in the width and height directions. After splicing, convolution and activation, weight maps of the X-axis and Y-axis are generated, which are normalized by the Sigmoid function and then weighted on the original feature map [6,12]. This enhances the model's perception ability of the spatial position of target regions, highlights the features of small target lesions, and improves segmentation accuracy [3,5].

The model adopts a loss function combining cross-entropy loss and Dice loss. Cross-entropy loss measures the pixel-level classification error, and Dice loss measures the overlap between the predicted region and the real region, which effectively alleviates the problem of unbalanced proportion of foreground and background pixels in skin disease images [4,7].

3.2 Experimental Design and Results

Experiments were carried out on the AutoDL computing power cloud platform with hardware configuration of 15vCPU Intel(R) Xeon(R) Platinum 8474C and RTX 4090D (24GB), and software environment of Ubuntu 20.04, Python 3.8 and PyTorch 1.10.0. The experimental dataset is ISIC2018, including 2594 training images, 100 validation images and 1000 test images [8]. All images were standardized to 512×512 pixels, and processed by data augmentation such as random flipping and photometric distortion and normalization [11]. The official training set was used as the experimental training set, and the official validation set was used as the experimental test set [3,5].

The hyperparameters of the model were optimized through multiple groups of comparative experiments, and the optimal parameters were finally determined: max_iters=10000, power=0.9, lr=1e-3, min_lr=1e-4, weight_decay=7e-5, with the Adam optimizer and polynomial decay strategy adopted [4,10].

The CAU-Net model was compared with the basic U-Net model, and the results show that: the aAcc of CAU-Net is 93.69%, an increase of 0.58% compared with U-Net; the mIoU is 85.50%, an increase of 1.33%; the mDice is 92.06%, an increase of 0.81%; the mMSSD is 0.0420, lower than 0.0461 of U-Net[2,6]. The CAU-Net model outperforms the basic model in all core indicators, effectively enhancing the segmentation ability of small target lesions under complex backgrounds and improving the fitting degree of segmentation boundaries [7,12].

4 CSWINTRANSU-NET MODEL INTEGRATED WITH SPATIAL PERCEPTION MODULE

4.1 Model Design

Aiming at the problems of large scale differences of skin lesions, limited segmentation ability of TransU-Net for blurred boundaries and insufficient fusion of local and global features, this paper introduces the spatial perception module on the basis of TransU-Net to construct the CSwinTransU-Net model [7,10]. The model integrates the local feature extraction ability of CNN and the global feature capture ability of Transformer, and consists of four parts: convolutional neural network module, CSwinTransformer module, feature fusion and upsampling module, and Segmentation head module [1,4].

The convolutional neural network module extracts local texture features of images through 3×3 convolution, ReLU activation and max pooling, and gradually compresses the spatial dimension and expands the number of channels [9]; the CSwinTransformer module is the core improvement part, which fuses the Transformer module and the spatial perception module. Based on the Cross-Shaped Window Self-Attention mechanism, the spatial perception module decomposes the rectangular receptive field into axial cross computing units, calculates the self-attention in horizontal and vertical directions in parallel, dynamically assigns spatial weights, and enhances the model's sensitivity to spatial position information [7,10]. Meanwhile, position encoding is introduced to make up for the loss of spatial position information by Transformer [6,8]; the feature fusion and upsampling module adopts a hybrid upsampling strategy combining bilinear interpolation and transposed convolution, and integrates features of different resolutions through BiFPN weighted fusion and tensor splicing to realize effective fusion of local and global features [2,9]; the Segmentation head module completes pixel-level classification through convolution, upsampling and 1×1 projection convolution, and outputs the segmentation results [4,5].

4.2 Experimental Design and Results

The experimental environment, dataset and preprocessing method are consistent with those of the CAU-Net experiment [8,11]. The optimal hyperparameters of the model were determined through hyperparameter tuning: max_iters=10000, power=0.9, lr=1e-4, min_lr=1e-5, weight_decay=7e-5[4,10].

The CSwinTransU-Net model was compared with the basic TransU-Net model, and the experimental results show that: the aAcc of CSwinTransU-Net is 93.49%, an increase of 3.21% compared with TransU-Net; the mIoU is 84.84%, an increase of 6.67%; the mDice is 91.66%, an increase of 4.23%; the mMSSD is 0.0481, significantly lower than 0.0730 of TransU-Net [3,7]. The CSwinTransU-Net model effectively solves the deficiencies of TransU-Net in blurred boundary segmentation and feature fusion, and improves the segmentation accuracy and robustness for complex backgrounds and lesions of different scales [5,6].

5 CONCLUSION

5.1 Research Summary

Taking skin disease image segmentation as the research object, this paper proposes two improved models and completes experimental verification aiming at the defects of existing deep learning models [1,8]. The main research results are as follows:

The CAU-Net model is proposed, which embeds the coordinate attention mechanism into the skip connections of U-Net, enhancing the model's perception ability of the spatial position of target regions [6,12]. It solves the problems of low segmentation accuracy of U-Net for small targets and insufficient position information modeling of traditional attention mechanisms, and achieves a significant improvement in segmentation performance on the ISIC2018 dataset [2,3].

The CSwinTransU-Net model is proposed, which fuses the spatial perception module in TransU-Net and optimizes global feature modeling based on the Cross-Shaped Window Self-Attention mechanism, realizing sufficient fusion of local and global features [7,10]. It effectively improves the model's segmentation ability for blurred boundaries and lesions of different scales, and ameliorates the problems of high computational complexity and insufficient segmentation robustness of TransU-Net [4,9].

With the ISIC2018 as the experimental dataset, the effectiveness of the two improved models is verified through systematic hyperparameter tuning and comparative experiments, both of which outperform the corresponding basic models, providing new and efficient models for skin disease image segmentation [5,8].

5.2 Future Outlook

This research provides a technical reference for skin disease image segmentation, and future research will be further deepened around the following directions [1,2]:

Optimize the network architecture design, explore the fusion mechanism of Vision Transformer and neuromorphic computing, construct lightweight models, reduce the computational complexity while ensuring segmentation accuracy, and improve the real-time performance of models [10,12].

Design a data augmentation tool combining GAN with the physical principles of dermatoscopic imaging to generate diverse samples with pathological authenticity, solving the problems of high annotation cost and insufficient sample size of medical datasets [11].

Fuse deep learning with traditional image processing methods, embed modules such as structured filtering and wavelet texture analysis into deep learning networks, realize joint optimization of morphological and texture features, and further improve segmentation accuracy [4,6].

Carry out research on multi-modal data fusion, integrate multi-source data such as dermatoscopic images, optical coherence tomography and pathological sections, construct a multi-modal segmentation model, and provide more comprehensive information support for skin cancer diagnosis [3,7].

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Qiu Y, Li Q Q, Wang Y. Application of Deep Learning Technology in Medical Image Segmentation. Computer Knowledge and Technology, 2022, 18(10): 74-75.

[2] Gu M J. Research on Intelligent Segmentation and Recognition of Skin Lesions Based on Deep Learning. Nanjing University of Posts and Telecommunications, 2023.

[3] Wang A L. Research on Melanoma Image Segmentation Based on Deep Learning. Lanzhou University of Technology, 2023.

[4] Song P F. Research on Medical Image Segmentation Algorithm Based on Convolutional Neural Network and Transformer. Shandong Technology and Business University, 2023.

[5] Sun Z L. Research on Semi-Supervised Learning Based Segmentation Algorithm for Skin Disease Medical Images. Jilin University, 2023.

[6] Yin W, Zhou D M, Fan T, et al. A Skin Lesion Image Segmentation Method Based on Dense Atrous Spatial Pyramid Pooling and Attention Mechanism. Journal of Biomedical Engineering, 2022, 39(06): 1108-1116.

[7] Liang L M, Zhou L S, Yin J, et al. Skin Lesion Segmentation Algorithm Fused with Multi-Scale Transformer. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(04): 1086-1098.

[8] Chen Y X. Research on Segmentation and Classification of Skin Disease Images Based on Deep Learning. Jiangxi University of Science and Technology, 2022.

[9] Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[10] Valanarasu J M J, Patel V M. UNeXt: MLP-based Rapid Medical Image Segmentation Network. Proceedings of the 25th International Conference on Medical Image Computing and Computer Assisted Intervention– MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science, 2022, 13435: 23-33. DOI: /10.1007/978-3-031-16443-9_3.

[11] Bozorgpour A, Sadegheih Y, Kazerouni A, et al. DermoSegDiff: A Boundary-Aware Segmentation Diffusion Model for Skin Lesion Delineation. PRIME 2023. Lecture Notes in Computer Science, 2023, 14277: 146-158. DOI: 10.1007/978-3-031-46005-0_13.

[12] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. Squeeze-and-Excitation Networks, 2018: 7132-7141. DOI: 10.1109/CVPR.2018.00745.