Segmentation of Oral Mucosal Cell Sampling Images Based on Enhanced Deeplabv3+ Algorithm

Tianjun Zhu,[1*†] Zizheng Zhu,[2†] Zhuang Ouyang,[3] Tunglung Wu,[1]
Yuanzhi Qian,[1] Jianguo Liang,[1] and Weihao Li[1]

[1]Department of Mechanical and Automotive Engineering, Zhaoqing University, Zhaoqing 526021, China
[2]School of Automotive Studies, Tongji University, Shanghai 201804, China
[3]Guangdong Zhaoqing Institute of Quality Inspection & Metrology, Zhaoqing 526000, China

Oral mucosal cell sampling is primarily utilized to gather samples of sloughed oral epithelial cells or cells in the throat for influenza DNA testing. In the robot sampling process, the image segmentation of the oral and pharyngeal swab sampling area plays a crucial role. However, owing to the complex nature of the oral sampling area, the accuracy of image segmentation can be slightly affected. A pharyngeal swab sampling image segmentation method based on an enhanced Deeplabv3+ model in the field of machine learning is proposed in this paper. The method applied hollow convolution to capture comprehensive information of each convolution output. We utilized deep learning of machine learning methods to enhance and refine the segmentation accuracy of the sampling area and gathered 1774 oral images from 81 volunteers, including children, youth, and middle-aged and elderly individuals, for training, validation, and testing. By comparing the experimental results of U-Net, Mobilenetv2, and Xception, it has been proven that the improved Xception network model has good segmentation performance, with accuracy, recall, and precision of 92.12, 92.86, and 97.69%, respectively. The experimental results indicate that this method accurately and efficiently segments the M-region of the pharyngeal swab sampling area, overcomes boundary discontinuity or ambiguity issues common in other segmentation methods, and possesses a high segmentation accuracy.

1. Introduction

Currently, oral mucosal genetic testing is utilized as a clinical testing method for detecting the genetic structure and composition of the oral mucosa. This method enables swift screening for bacteria, fungi, and viruses, and also aids in diagnosing and treating related oral diseases such as oral mucosal inflammation and oral cancer. Typically, oral pharyngeal swabs are used for sample collection; however, there are concerns for the safety of medical personnel conducting the test, particularly in cases where patients cough and generate a significant amount of droplets

---

or aerosols. Additionally, the skill level and inconsistent collection operations of medical staff can lead to varying sample quality and misdiagnosis risks.

To mitigate infection risks and standardize oral mucosal detection sampling, a visual-guided oral mucosal sampling robot has been developed to replace manual sampling. However, the accuracy of image segmentation for the oral sampling area can be challenging owing to its complex environment. Precise identification and segmentation of the oral cavity's sampling area (i.e., M-region as depicted in Fig. 1) play a key role in the oral mucosal sampling robot's sampling process.

Image segmentation is a crucial preprocessing step for image recognition and computer vision in the field of machine learning. However, segmentation relies solely on the brightness and color of pixels within the image, which leads to various challenges when processed automatically by computers. Uneven illumination, noisy regions, unclear parts, and shadows frequently cause segmentation errors. Further exploration is necessary to advance image segmentation, which divides images into subsets called regions to isolate objects from their background.[1] Standard methods for image segmentation include threshold, clustering, and edge detection. Digital image processing comprises three stages: image processing, image analysis, and image understanding.[2] Image segmentation plays an essential role in this processing, as it extracts objects from their backgrounds,[3–7] and significantly impacts image recognition. Moreover, segmentation can transform high-resolution images into an abstract form for analysis and comprehension,[8] enabling recognition, characterization, and measurement of targets on the basis of their segmentation. In biomedical image segmentation tasks, the U-Net architecture is widely used,[9] such as for optic disc and cup image segmentation,[10] blood vessel segmentation,[11] and tumor segmentation.[12] U-Net architecture uses an encoder–decoder structure to accurately locate targets of different scales through path expansion in the prediction stage, and jump connections are used to fuse multilevel features. Such connections help restore the full spatial resolution of the network output and effectively solve issues caused by the insufficient resolution of feature maps, ensuring accuracy of the segmented boundary.

Major image segmentation strategies utilize multiple scale features, leveraging scale consciousness, context information, and robustness enhancement for enhanced segmentation precision and resilience.[13–16] An image pyramid, where subsequent images are scaled down using an average strategy, is constructed. Each pixel contains a local average that corresponds to

Fig. 1.   (Color online) Oral cavity and marked M-region in the field of machine learning.

the pixel neighborhood in the lower level of the pyramid. Once this segmentation method is completed, the pyramid is upsampled and aggregated. This method is considered reasonable as it processes coarse and fine scale information interactively on the basis of the human visual system. Multiscale methods have found widespread applications in medical image processing. Lin *et al.* proposed a multiscale level set for echo image segmentation.[17] Kumar *et al.* suggested an edge detection and segmentation of leukocyte nuclei based on K-means, which enables hematologists to quickly and clearly identify leukocytes.[18] Rezaee *et al.* combined the pyramid image with the fuzzy c-means clustering algorithm and proposed an unsupervised image segmentation method[19] that showed good results for LV cavity detection in MR images.

Taking into account the practical application requirements of developing a visually guided oral mucosal sampling robot, in this article, we focus on the problem of discontinuous or blurry boundaries in the M-region of the oral cavity, which occurs because of insufficient lighting and feature image resolution. A method for oral mucosal sampling image segmentation and recognition based on the improved Deeplabv3+model is proposed in this paper. The method includes a large amount of cavity convolution in the encoder section, while expanding the receptive domain of void convolution and adding an improved Xception module to enhance the model's adaptability to target scale migration. The experimental results reveal that the algorithm possesses good recognition abilities for the oral M-region, refines the region's boundary, and improves the detection accuracy. It is also highly portable and has important engineering application value.

The main academic contributions of this paper are summarized as follows.

1. In this study, we envisage an ingenious methodology for the tribulation and identification of oral mucosa sampling imagery, leveraging an augmented domain of machine learning by applying the advanced Deeplabv3+ model. This approach harnesses the merits of the U-Net Ed architecture and the atrous spatial pyramid pooling (ASPP) configuration, and augments the receptive field of cavity convolution to alleviate deficiencies associated with inadequate illumination and resolution, thus refining the imaging segmentation precision.

2. Traditional segmentation models can approximate the M-region in the exit cavity, but the M-region's edge segmentation precision is subpar, potentially leading to significant medical image segmentation inaccuracy. In this study, we integrate an enhanced Xception module into the machine learning domain's image segmentation recognition algorithm, thereby amplifying the model's capacity to adapt to alterations in the target scale, fortify its image feature learning prowess, and meticulously refine the region's boundary intricacies with greater precision. This enhancement substantially augments the algorithm's processing accuracy and detail management capabilities.

2. Deeplabv3+ Model

Deeplabv3+ is a top-performing semantic segmentation network. It incorporates a straightforward yet powerful decoder to enhance the segmentation results, particularly those at the boundary of the target object, by combining the spatial pyramid pool module and the codec structure. The Deeplabv3+ semantic segmentation series model builds on the pyramid structure

to expand the receptive field without altering the resolution ratio, making it ideal for multiscale segmentation. By leveraging the benefits of the U-Net Ed structure and ASPP configuration, Deeplav3+ stands out as an outstanding semantic segmentation algorithm with impressive overall performance.

2.1   Enhanced Deeplabv3+ model

The upgraded Deeplabv3+ incorporates expansion convolution residuals with varying rates to establish connections and extract enhanced features. This improves the network's receptive field and enhances its feature receptivity. The process of segmenting the M region is demonstrated in Fig. 2.

The improved Deeplabv3+ model comprises two main components: encoder and decoder. The oral image input is processed by the Xception trunk network in the encoder, and the image is sequentially extracted by the entry flow, middle flow, and exit flow in the Xception network. The data first goes through the entry flow, then through the repeated middle flow eight times, and finally through the exit flow. The image features' receptive field is then increased using atrous convolution. Subsequently, the dynamic convolution neural network (DCNN) produces two effective feature layers: the shallow effective features and the deep effective features. The latter undergoes further enhancement through the ASPP structure, followed by a new feature layer generated through 1×1 convolution.

A new feature layer is created using a 3×3 convolution with an expansion rate of 6, which is then transferred to a 3×3 convolution with expansion rates of 12 and 18 for residual connection. This process helps improve the convergence speed and performance. The deep effective feature layer is residually connected with the new feature layer generated by the 3×3 convolution with
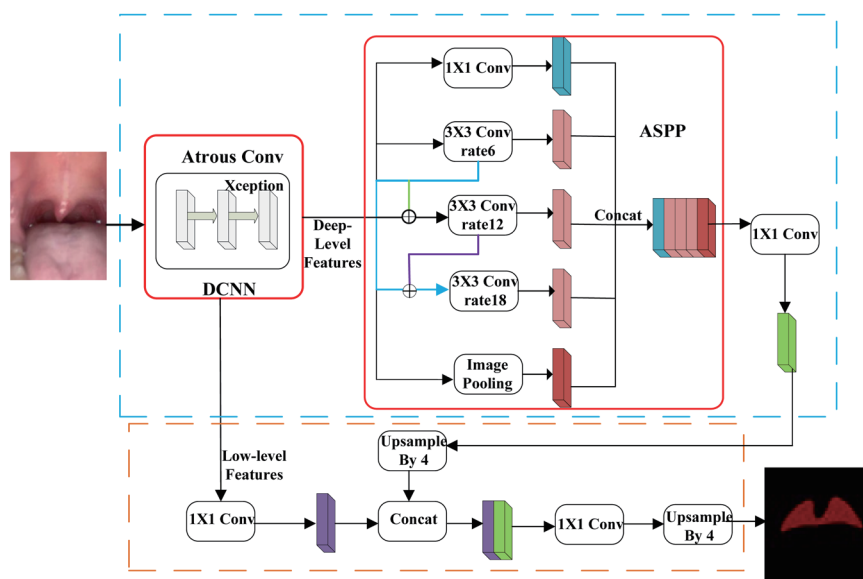

Fig. 2.   (Color online) Enhanced Deeplabv3+ model.

an expansion rate of 6. The new feature layer is then subjected to a 3 × 3 convolution with an expansion rate of 12, which is transferred to a 3 × 3 convolution with an expansion rate of 18 for residual connection. The deep effective feature layer is residually linked with two new feature layers generated by convolution with expansion rates of 6 and 12. The new feature layer is obtained by a 3 × 3 convolution with an expansion rate of 18. Finally, the image pooling technique is utilized to generate a new feature layer for the deep effective feature layer. The five new feature layers obtained from the above procedures should be concated, and the number of channels is adjusted using 1 × 1 convolution. This results in the feature layer containing relatively high semantic information, denoted by the light green feature block in Fig. 3, which is then passed on to the decoder.

In the decoder, we obtain the shallow effective feature layer from the encoder. The channel number is adjusted using 1 × 1 convolution to derive a new feature layer, denoted by the purple feature block in Fig. 3. The purple and light green feature blocks are then stacked after four rounds of upsampling, which leads to the final effective feature layer that represents the entire image. After adjusting the number of channels using 1 × 1 convolution and four more rounds of upsampling, the width and height of the final output layer match those of the input image.

2.2 Cavity convolution

Cavity convolution is an essential component of the Deeplabv3+ model and contribute to the extraction of multiscale information by controlling the receptive field size without altering the feature map's dimensions. Figure 3(a) shows a traditional convolution, and Fig. 3(b) shows a cavity convolution. Cavity convolution is the inserti a 0 in the middle of the traditional convolution kernel, which looks like there is a hole in the middle, so it is called cavity convolution. Cavity convolution has the same number of parameters as traditional convolution, and the output feature map size is the same. The receptive field refers to the area size of the input layer corresponding to an element in the output result of the convolution module. The larger the receptive field, the less likely image information is to be lost. In Fig. 3, a 3 × 3 convolution kernel can achieve a receptive field size of 5 × 5 (dilated rate = 2).

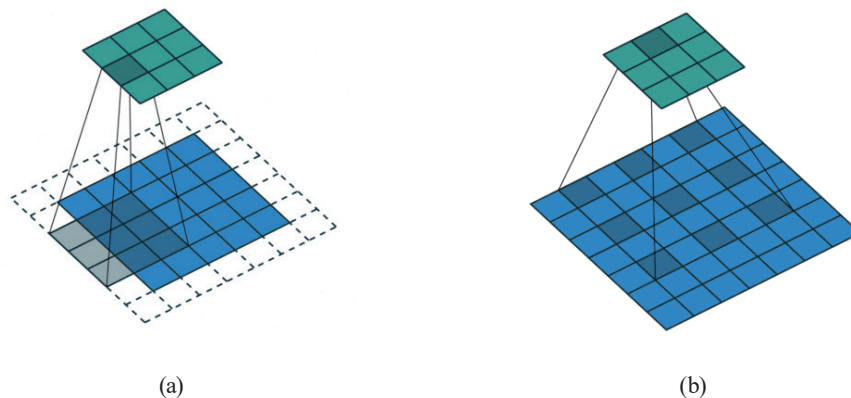(a)   (b)

Fig. 3.   (Color online) (a) Traditional convolution and (b) cavity convolution.

2.3 Major networks

Deeplabv3+ offers two main network options—Xception and Mobilenetv2. In this study, the Xception network was employed owing to its intermediate structure between traditional convolution and deep separable convolution. Xception is completely decoupled into deep separable convolution, allowing for the complete decoupling of cross channel correlation and spatial correlation in the feature mapping of convolutional neural networks. The Xception network architecture is divided into three parts: entry flow, middle flow, and exit flow. It consists of 36 convolution layers that make up the network's feature extraction library. These 36 convolution layers are divided into 14 modules, with the first and last modules being disconnected while the rest are linearly connected. First, data goes through the entry flow, then it moves to the middle flow eight times and finally to the exit flow.

2.4 Depthwise separable convolution

Depthwise separable convolution (DSC) is used to reduce network parameters and improve computational efficiency. Its core idea is to decompose a complete convolution operation into two parts, namely, depthwise convolution and pointwise convolution. The detailed structure is shown in Fig. 4. The conventional convolutional kernel needs to perform convolution operations on each channel of the input, while depthwise convolution uses a convolution kernel for each channel of the input feature map, and then concatenates the outputs of all convolution kernels to obtain its final output, as shown in Fig. 4(a). The number of feature maps after depthwise convolution is the same as the depth of the input channels. The operation of pointwise convolution is very similar to conventional convolution operations, as shown in Fig. 4(b). The main function of pointwise convolution is to change the number of channels in the feature map.
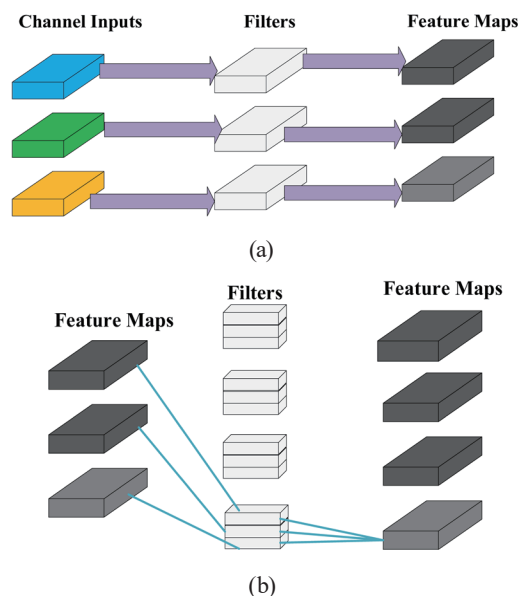

Fig. 4.   (Color online) (a) Depthwise convolution and (b) pointwise convolution.

By adjusting the number of output channels, feature dimensionality can be increased or decreased without changing the spatial dimension of the feature map.

2.5 Loss function

The loss function employed in this study comprises two components, namely, the standard cross-entropy loss function and the set similarity metric. For pixel classification, the cross-entropy loss function employs softmax, which is commonly used for minimizing losses. Specifically, Eq. (1) represents the cross-entropy loss function.

$$CE(p,q) = -\sum_{i=1}^{C} p_i \log(q_i) \tag{1}$$

Here, C represents the number of categories, $C = 1$; p_i is the true value; q_i is the predicted value; s is the Dice coefficient, which is usually used to calculate the similarity of two samples, and the value range is [0, 1]. The Dice coefficient s is determined using

$$s = \frac{2|X \cap Y|}{|X| + |Y|}. \tag{2}$$

Here, $|X \cap Y|$ is the intersection between X and Y; $|X|$ and $|Y|$ are the numbers of elements representing X and Y, respectively. The higher the Dice coefficient, the better the performance, indicating a stronger match between the predicted and actual results. Hence, when the probability of predicting the future is larger, the value of the Dice coefficient is larger.

3. Experimental Methods

3.1 Data collection and environment configuration

The robotic device for collecting oral mucosal samples incorporates a flexible arm [shown in Fig. 5(a)], a camera, a wireless transmitter, and a module for recognizing the oral M-region. Leveraging the power of AI, this system ensures accurate sampling and automated navigation, as well as automatic recognition and calibration of force control. The robotic arm employs cutting-edge motion planning technology to autonomously collect oral mucosal samples. Additionally, with the help of a computer vision system, it tracks and precisely locates the patient's mouth and throat, enabling it to retrieve oral mucosal samples with the utmost precision.

In this study, we gathered oral images from 81 participants comprising individuals of different ages. A total of 1569 oral images were collected and expanded to 7845 images through horizontal flipping and rotation. The researchers utilized annotation tools to label each image and create corresponding training sets under the professional guidance of respiratory doctors in hospitals during data collection and calibration work. The specific usage of the collected images is as follows: 80% as training data, 10% as validation data, and 10% as testing data. We strictly
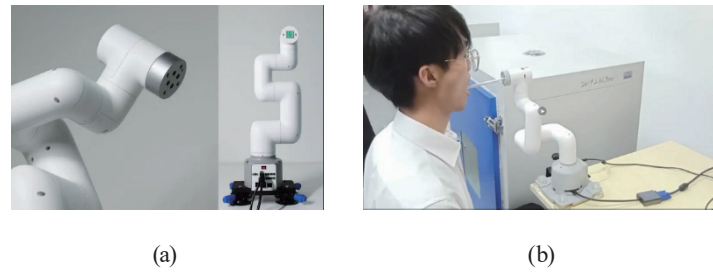
Fig. 5. (Color online) Composition of pharyngeal swab sampling system and experimental test scenario: (a) sampling mechanical arm and (b) sampling robot.

categorize the sample images of volunteers of different ages into the training dataset, validation dataset, and testing dataset in sequence. For example, if a patient's sample image is used as training set data, it cannot be placed in the validation or testing dataset. The images were also normalized for faster convergence speed in the loss function during network training.

3.2 Model training

The model training process consists of two stages: freezing and thawing. During the freezing stage, the trunk network of the model is fixed and only the feature extraction network is fine-tuned, which minimizes video memory usage. Meanwhile, during the thawing stage, the trunk network is unfrozen and the feature extraction network is modified, leading to increased video memory usage as all network parameters are changed. The Xception model is utilized as the trunk extraction network, and the pytorch GPU environment is used for model training. The training process employs a ratio of 8:1:1 for the training set, validation set, and test set, respectively. The loss function obtained from the training process is shown in Fig. 6, which depicts a gradual decrease and convergence of the loss value curve (train loss: 0.08613) for the training set and the loss value curve (val loss: 0.10026) for the validation set. As shown in Fig. 6, it can be seen that the loss value curves of the training set and the test set gradually decrease and converge, indicating that the training network is the most ideal situation. To improve the generalization ability of the model, we used validation set data (10%) to determine the hyperparameters, such as the number of network layers and the number of neurons in each layer.

3.3 Evaluating indicator

To assess the segmentation outcome of the M-region, accuracy, recall, and precision serve as evaluation metrics. The calculation formulas are provided below.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

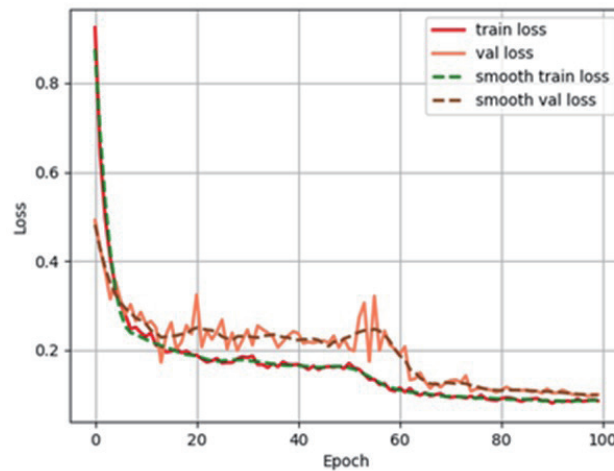$$Recall = \frac{TP}{TP + FN} \tag{4}$$

Fig. 6. (Color online) Model training curve.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

In classification analysis, true positive (*TP*) represents the accurate prediction of a positive class; false positive (*FP*) refers to the incorrect prediction of a positive class instead of the true negative (*TN*) class; *TN* indicates the correct detection of a negative class; and false negative (*FN*) signifies the misclassification of a positive class as a negative class. Researchers evaluate and categorize the segmentation results of an image and determine the appropriate values accordingly.

4. Experiment

4.1 Model training

Owing to the fact that the number of validation data samples in this study is only 10% of the total number of collected samples, to fully verify the accuracy and stability of this model, we collected 205 fresh oral images as model validation data to further verify the performance of the model. At the same time, we can also quickly adjust parameters (such as network layers, network nodes, iteration times, etc.) to obtain the current optimal model. This set comprised 182 standard oral images, as well as 23 nonstandard ones. The prediction outcomes were then subjected to review by professional oral sampling doctors to ensure that they met medical standards. The Xception test classification results are presented in Table 1 as follows: 169 *TP*s, 4 *FP*s, 19 *TN*s,

Table 1
Xception classification results.

Xception	*N*	*P*
F	13	4
T	19	169

and 13 *FN*s. As such, the model demonstrates an accuracy of 92.12%, a recall rate of 92.86%, and a precision rate of 97.69%.

4.2   Model training

To evaluate the segmentation performance of the improved Deeplabv3+ model using Xception, we compared it with U-Net and Mobilenetv2 models of the same Deeplabv3+ architecture. The detection and classification results for the U-Net model are presented in Table 2, whereas those of the Mobilenetv2 model are shown in Table 3. We evaluated the performance of the Deeplabv3+ model with Mobilenetv2, Xception networks, and U-Net on the test set, as shown in Table 4, which displays the accuracy, recall, and precision values. From the results in Table 4, it is evident that the Xception network outperforms the two other models.

The segmentation outcomes of U-Net, MobileNetv2, and Xception are shown in Fig. 7. The first two rows of the initial column include oral mucosa images of the M-region, whereas the last two rows do not contain any oral mucosa images of the M-region. The initial two rows of the second column represent the M-region contour marked by a doctor, whereas the nonstandard acquisition region is left unmarked in the last two rows. The third, fourth, and fifth columns indicate the performances of U-Net, MobileNetv2, and Xception, respectively, in separating the M-region from the background region. The experimental outcomes demonstrate that the improved Xception method can effectively distinguish the M-region from the background region and enhance the accuracy of M-region contour prediction.

To sum up, the network structure of Xception consists of a series of depthwise separable convolutional layers, which also introduce residual connections and layer-by-layer increasing methods. Residual connections can effectively solve the problems of vanishing and exploding gradients, enabling the network to train and optimize better. Compared with traditional models,

Table 2
U-Net classification results.

U-Net	*N*	*P*
F	33	14
T	9	149

Table 3
Mobilenetv2 classification results.

Mobilenetv2	*N*	*P*
F	19	13
T	10	163

Table 4
Results of different networks.

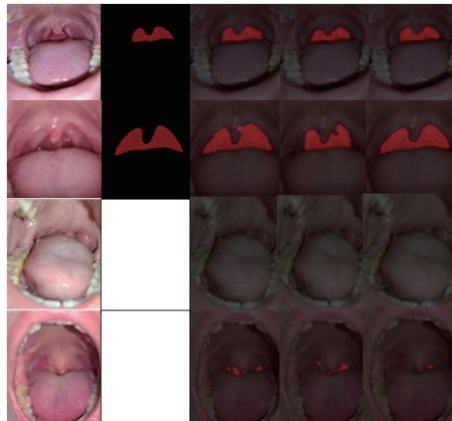|	*Accuracy* (%)	*Recall* (%)	*Precision* (%)
U-Net	84.44	89.56	92.61
Mobilenetv2	77.07	81.87	91.41
Xception	92.12	92.86	97.69

Fig. 7.   (Color online) Segmentation results of U-Net, Mobilenetv2, and Xception.

Xception has the following advantages. 1. Small number of parameters: Xception adopts DSC operation, reducing the number of parameters in the network, thereby reducing the storage and calculation load of the model. 2. Better generalization ability: Xception allows the network to better adapt to different image datasets and have better generalization ability through residual connections and layer-by-layer addition. Hence, this proposed segmentation method based on an enhanced Deeplabv3+ model accurately and efficiently segments the M-region of the pharyngeal swab.

5.   Conclusions

Accurate segmentation of the oral mucosal sampling area is crucial for guiding robots to sample the oral mucosa. However, the complex environment of the oral sampling area can negatively impact the accuracy of image segmentation. Therefore, achieving precise segmentation of the M-region is essential in the robot oral mucosal sampling process. The experimental results show that the proposed model based on the improved Xception network model demonstrates an accuracy of 92.12%, a recall rate of 92.86%, and a precision rate of 97.69%. Compared with the U-Net and Mobilenetv2 model, the improved Xception network model improves accuracy, recall rate, and precision by approximately 10%.

1. To address the above issue, we introduced improvements to the Xception network structure of Deeplabv3+. Specifically, the improved Xception network model incorporates hole convolution, multiscale feature fusion, and self-attention mechanisms to extract valuable features more effectively.

2. Experimental data were obtained from oral images of 81 volunteers of various ages. The experimental results of U-Net, Mobilenetv2, and Xception were compared. The improved Xception network model can effectively distinguish the M-region from the background region and shows excellent segmentation performance. These improvements effectively enhance the network's segmentation performance and successfully address the problem of gradient disappearance.

3. During the model training phase, *TP*, *FP*, *FP*, and *TN* were subjectively determined by researchers under the guidance of professional doctors. We are also exploring the use of other evaluation indicators in the future, such as the receiver operating characteristic curve and the PR curve, for objective evaluation.

4. The research scenario of this study is to develop an automatic oral and pharyngeal swab collection system for hospital systems. As a core part of it, the accurate segmentation of the oral mucosal sampling area is crucial for guiding robots to sample the oral mucosa. There is no publicly available oral image database for verification testing, and only volunteer data is used for model validation. In the future, it is hoped that public database resources can be established or found. At the same time, we will conduct an abortion experiment in our future research.

Acknowledgments

References

1 J. Zhao, F. Shao, Y. Xu, X. Zhang, and W. Huang: Proc. IEEE 3rd Int. Congr. Image Signal Process. Conf. (IEEE, 2010) 1317–1321.
2 V. Kilaru, M. Amin, F. Ahmad, P. Sevigny, and D. Difilippo: Electron. Imaging **24** (2015) 1. https://doi.org/10.1117/1.JEI.24.1.013028
3 Y. Cheng and B. Li: Proc. 2021 IEEE Asia-Pacific Conf. Image Process. Electron. Comput. Conf. (IEEE, 2021) 1174–1177.
4 C. Mi, Z. Zhang, X. He, Y. Huang, and W. Mi: Pol. Marit. Res. **22** (2015) 163. https://doi.org/10.1515/pomr-2015-0049
5 Y. Wang, Z. Xu, H. Zhao, J. Yang, and S. Wang: Proc. 2021 IEEE Int. Conf. Intell. Saf. Robot. Conf. (IEEE, 2021) 186–190.
6 D. Santos, P. D. Faria, B.Travenolo, and M. Z. Nascimento: Biomed. Signal Process. Control **69** (2021) 102921. https://doi.org/10.1016/j.bspc.2021.102921
7 R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi: IET Image Proc. **16** (2022) 1243. https://doi.org/10.1049/ipr2.12419
8 S. M. Raj, C. Jose, and M. H. Supriya: J. Eng. Sci. Technol. **12** (2017) 2536. https://www.researchgate.net/publication/320063513_Hardware_realization_of_canny_edge_detection_algorithm_for_underwater_image_segmentation_using_field_programmable_gate_arrays
9 O. Ronneberger, P. Fischer, and T. Brox: Proc. 2015 Int. Conf. Med. Image Comput. Computer-assisted Intervention Conf. (ICMCC, 2015) 234–241.
10 H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao: IEEE Trans. Med. Imaging **37** (2018) 1597. https://doi.org/10.1109/TMI.2018.2791488
11 K. K. Maninis, J. Tuset, P. Arbeláez, and L. Gool: Proc. 2016 Int. Conf. Med. Image Comput. Computer-assisted Intervention Conf. (ICMCC, 2016) 140–148.
12 N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, H. Kittler: Proc. 2018 IEEE 15th Int. Symp. Biomed. Imaging Conf. (ISBI, 2018) 168–172.
13 Y. Cheng, and B. Li: Proc. 2021 IEEE Asia-Pacific Conf. Image Process., Electron. Comput. Conf. (IPEC, 2021) 1174–1177.

14   V. Polge, Y. Gavet: Proc. 2021 10th Int. Symp. Signal Image Video Commun. Conf. (ISIVC, 2021) 1–6.
15   M. Vlachos and E. Dermatas: Comput. Med. Imag. Grap. **34** (2010) 213. https://doi.org/10.1016/j.compmedimag.2009.09.006
16   W. Zhou, X. Du, and S. Wang: Proc. 2021 IEEE Int. Conf. Comput. Sci., Electron. Inf. Eng. Intell. Control Technol. Conf. (ICCS, 2021) 400–403.
17   N. Lin, W. Yu, and J. Duncan: Med. Image Anal. **7** (2003) 529. https://doi.org/10.1007/3-540-45786-0-84
18   P. R. Kumar, A. Sarkar, S. N. Mohanty, and P. P. Kumar: Proc. 2020 5th Int. Conf. Comput., Commun. Secur. Conf. (ICCCS,2020) 1–4.
19   M. Rezaee, P. Zwet, B. Lelieveldt, R. J van der Geest, and J. H. Reiber: IEEE Trans. Image Process. **9** (2000) 1238. https://doi.org/10.1109/83.847836