（課程博士用）
(for Degree earned by Completing Doctoral Program)

（別記様式－博７）

千葉大学審査学位論文（要約）（Summary）

基幹工学 専攻 医工学 コース

Division Department

氏　名 YANG TINGXIAO

Name

論文題名（外国語の場合は、その和訳を併記）

Segmentation for Tongue Diagnosis Based on Deep Learning

(舌診断のためのディープラーニングを用いたセグメンテーション法の検討)

1. **General Conclusion for the Entire Thesis Work**

■ **Issues and Experience**

◆ In this thesis, the author verified the feasibility of deep learning in medical images through the segmentation of tongue and sublingual veins. During our research, the difficulties the author encountered are as below:

- Annotation precision issue
- Out of GPU Memory
- Dirty output of pixel-wise segmentation
- Some others' paper results cannot be reproduced
- All black prediction when the target is small
- Loss function cannot converge
- Limit size of the dataset
- imbalanced dataset

◆ Based on the above issues, the author summarizes the following experience:

- For labels that cannot be determined clearly, try to label the target with multiple people, and try to annotate all labels at once within a certain period of time; for unclear boundary labeling, try to label after adjusting the color saturation without destroying the original principal component information.

- When the model cannot be trained due to OM (Out of Memory), reduce the batch size; reduce the number of channels in down-sampling and up-sampling; reduce the size of the output node while down-sampling the annotation resolution (In general, 128×128 is the lower limit of the final output quality of pixel-level segmentation.)
- Try to use post-processing techniques for regional fusion or impurity removal;
- For region-base segmentation, try to use the proposed contour detection algorithm, regard the region segmentation as contour detection.
- Deep learning strongly relies on datasets. Under different datasets, performance may vary greatly. Before realizing on the custom dataset, do not blindly believe other people's conclusions. In addition, in many cases, the performance is good or bad, in fact, it has a lot to do with the depth of the model and the degree of adaptation to the current dataset.
- For small target segmentation, researchers can utilize regional multi-stage segmentation: Firstly segment the large area where the small target is located, and segment the final small target; Or implement the proposed SA U-Net: Using multiple optimizers to optimize network nodes in batches when training to capture the attention region, the small target of one-step segmentation can be achieved during the inference process.
- Check the correctness of the loss function;
- Increase the initial learning rate or use the learning rate schedule (increase learning rate at first several training steps and after that reduce the learning rate).
- Use the proposed multi-network augmentation strategy and learn the information from the
- pseudo labels or predictions, which can also be achieved by applying GAN (Generative Adversarial Network); Apply transfer-learning, pre-load the weights trained on the other large data. The pre-trained dataset preferably has a

certain correlation with the currently required dataset.

- For imbalanced dataset, resampling the data from the majority class and minority class with even alternating training; use class weights to balance the class contribution or loss weights to balance the loss component (small object segmentation sometimes are also regarded as imbalanced problem).

General Conclusion

- The author has successfully realized tongue segmentation, sublingual vein segmentation and multi-regions of tongue segmentation, which can basically cover the needs of a comprehensive tongue diagnosis. The author pointed out the drawbacks of common U-shape segmentation framework: the risk of the black output and dirty output of the segmentation. To address the problem of black output, the author progressively explored multi-label assisted two-stage segmentation networks and single-label synergistic one-stage segmentation, which successfully solved the phenomenon of all black output segmentation, so that the sublingual vein segmentation model can produce the useful outputs for the training. Meanwhile, the author puts forward the concept of BRA and mult-network augmentation strategies for the problem of few sample training dataset. On the whole, the performance of vein segmentation reached around 0.55 AIoU. For tongue segmentation, the author proposed a lightweight model PPAN, which reduces the training parameters of the classic U-Net by nearly 80% while ensuring the performance above 0.9 IoU. The pyramid predictive attention design therein contributes close to 15% parameter savings. To further solve the dirty output problem of tongue segmentation, the author proposes the I-Polar-Poly contour detection framework, which unified the segmentation and detection framework. I-Polar-Poly utilizes the advantages of both pixel-wise segmentation and instance detection to convert the problem of region segmentation into a problem of contour detection during the training. Through this framework, the author realized mAP 0.99 and around 0.90 IoU for tongue detection and masks receptively at the same time with more 90 FPS. Based on the single tongue detection and segmentation, the author further

realized multi-region segmentation of the tongue with multiple classes. The segmented regions of tongue are actually independent instance segmentation, which means that this algorithm can not only realize multi-region segmentation, but also realize the joint diagnosis of multi-region symptoms in the future. Although the method is proposed for tongue segmentation, it is applicable for other region-based tasks.

2. Future Work

■ Application Outlook

- ◆ In the current outbreak of the new coronavirus, non-contact diagnosis such as tongue diagnosis should be paid more attention. This facilitates tracking and summarizing patient information. The proposed model is a lightweight model that is suitable for home and mobile phone diagnosis application scenarios. When tongue diagnosis is successfully applied to personal convenience scenarios, coupled with cloud analysis, it is believed that in the near future, it can be effective for home tongue diagnosis and the inspecting the patient health. The cloud tongue diagnosis can provide more reliable medical advice, which can reduce the burden of treatment in hospitals.

■ Algorithm Improvement

- ◆ Although the current algorithm has achieved high detection and mask quality for contours, the fixed input size of polar coordinate conversion still causes slight artificial errors in the learning target. In the future, it is hoped that this conversion process can be re-expressed through the input form of GCN, which remove this conversion error. At the same time, the current model does not add too many complex designs due to the consideration of inference speed. Without considering the speed, more model design principles can be considered to achieve more accurate contour inference. In addition, this thesis focuses on convolutional segmentation algorithms. In this field, there is a famous framework nnU-Net research published in nature method recently, whose authors try to find a general recipe for medical image segmentation from another perspective. The provided framework performs well on a variety of

datasets, and the author of thesis will study this framework in more depth in future research to try to find its benefits and apply them in the current research topic.