International Journal of Internet, Broadcasting and Communication Vol.16 No.1 280-286 (2024)
http://dx.doi.org/10.7236/IJIBC.2024.16.1.280

IJIBC 24-1-33

Improving Accuracy of Instance Segmentation of Teeth

Jongjin Park

Professor, Department of Computer Engineering, Chungwoon University, Incheon, Korea
jjpark@chungwoon.ac.kr

Abstract

In this paper, layered UNet with warmup and dropout tricks was used to segment teeth instantly by using data labeled for each individual tooth and increase performance of the result. The layered UNet proposed before showed very good performance in tooth segmentation without distinguishing tooth number. To do instance segmentation of teeth, we labeled teeth CBCT data according to tooth numbering system which is devised by FDI World Dental Federation notation. Colors for labeled teeth are like AI-Hub teeth dataset. Simulation results show that layered UNet does also segment very well for each tooth distinguishing tooth number by color. Layered UNet model using warmup trick was the best with IoU values of 0.80 and 0.77 for training, validation data. To increase the performance of instance segmentation of teeth, we need more labeled data later. The results of this paper can be used to develop medical software that requires tooth recognition, such as orthodontic treatment, wisdom tooth extraction, and implant surgery.

Keywords: *Instance Segmentation, Layered UNet, Warm-up trick, Dropout, Deep Learning*

1. Introduction

Recent advances in artificial intelligence have achieved remarkable performance, especially in image processing. In particular, the introduction of deep learning, like CNN, opened a new chapter in image processing and many new architectures have been proposed since then, including VGG, Inception, ResNet, and DenseNet. At the same time, we have seen a steady trend of model accuracy improvement. For example, the top-1 validation accuracy on ImageNet has been raised from 62.5% (AlexNet) to 82.7% (NASNet-A)[1]. However, these advancements did not solely come from improved model architecture. Training procedure refinements, including changes in loss functions, data preprocessing, and optimization methods also played a major role. Tong He et. al. applied a collection of training procedure and model architecture refinements to improve model accuracy. They are minor “tricks” like modifying the stride size of a particular convolution layer or adjusting learning rate schedule. However, they make a big difference of model accuracy. The tricks raises ResNet50’s top-1 validation accuracy from 75.3% to 79.29% on ImageNet[1].

Image recognition in digital image processing largely includes object detection, object recognition, object

Manuscript Received: January. 15, 2024 / Revised: January. 29, 2024 / Accepted: February. 9, 2024
Corresponding Author: jjpark@chungwoon.ac.kr
Tel:*** - **** - ****
Author’s affiliation : Professor, Department of Computer Engineering, Chungwoon University, Incheon, Korea

tracking, object classification, and object segmentation. Object segmentation in image recognition includes semantic segmentation, instance segmentation, and panoptic segmentation. Object segmentation is used in various application areas of computer vision. Among them are autonomous driving, object recognition and tracking, medical image segmentation, augmented reality (AR) and virtual reality (VR), video surveillance, image editing and forensic analysis, and robotics.

I have proposed various UNet models and the layered UNet for teeth segmentation in dental CBCT images[2]. Recently, an artificial intelligence model capable of anatomical segmentation of dental images was developed using a dataset that are labeled according to tooth numbers provided by AI-Hub. It is not easy to recognize each tooth number and separate individual tooth from dental CBCT images. For this purpose, good performance was obtained using the proposed layered UNet[3].

In this paper, A method is applied to improve the performance of individual tooth segmentation using warmup trick in addition to the results obtained by the authors. In the warmup heuristic, we use a small learning rate at the beginning and then switch back to the initial learning rate when the training process is stable[1]. It can help layered UNet learn faster and to be more accurate segmentation model. For this purpose, labeled teeth dataset was prepared.

2. Semantic instance segmentation and layered UNet

Segmentation allows us to organize the data contained within images and videos into meaningful categories. Image classification and object detection may tell us the presence and location of certain objects. Segmentation allows us to dive deeper even in a real time. Semantic segmentation predicts the category for every pixel from the input images. This type of segmentation involves grouping each pixel under a particular label. For example, any pixel belonging to a car would be assigned under the same "car" category. Instance segmentation takes semantic segmentation to the next level by distinguishing between distinct objects belonging to the same category. Each car in the previous example would still belong to the same label but would be given different colors. In the medical area, semantic and instance segmentation are applied to various fields. Image segmentation using deep learning plays a critical role in detecting medical abnormalities that appear in clinical scans, such as CT or MRI scans. Fast and accurate computer vision algorithms allow medical personnel to manage their time. It's not necessary to analyze each scan. Relying on computer vision doctors can streamline treatment and maximize the number of patients that can be examined. In the area of dentistry, dental image segmentation also plays important role in Orthodontics, Periodontology and Oral Medicine etc.[4].

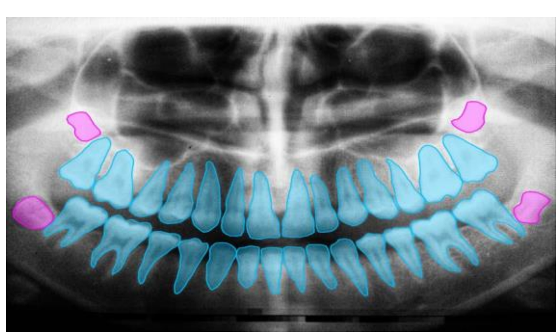


Figure 1. An example of semantic segmentation of dental image

Deep learning, a field of AI, is showing remarkable performance in medical image segmentation. In semantic segmentation, a field of medical image processing, various deep learning models have been designed, and the UNet proposed by Ronneberger et al. is attracting attention for its excellent performance. CNN (Convolution Neural Network), an early deep learning model, showed excellent performance in existing image classification problems and object detection problems. However, there was a limitation in the image segmentation problem. This is because most existing networks for image classification consist of a fully connected last layer, so they are suitable for classification problems or object detection problems, but require dense classification at the pixel level. This is because it is not suitable for the problem. To solve these limitations, FCN (Fully Convolutional Network) removes the fully connected layer and configures all layers only as convolution layers, so that the result of feature point extraction is the probability of several predefined classes for each pixel. It came out in the form of a probability map. UNet is based on this FCN and has an encoder and decoder structure, showing excellent results in semantic segmentation. Afterwards, various models such as UNet++ and UNet3+, which were improved versions of UNet, were proposed and performance was further improved.

we proposed layered UNet model to segment teeth from dental CBCT data for orthodontic treatment. The layered model uses both the inter-connection and intra-connection proposed by UNet3+ and the nested convolution block proposed by UNet++ in order to utilize the skip connection structure designed to improve performance in the existing UNet series of models. The layer UNet model shows better segmentation results than the existing UNet3+ and has excellent accuracy even though a small number of image data is used for learning[2]. The layered UNet has been used to build instance segmentation of teeth using AI-Hub dataset[3]. Figure 2 shows Layered UNet proposed by us.

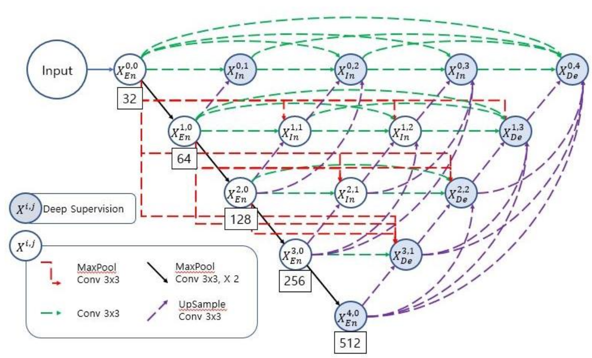
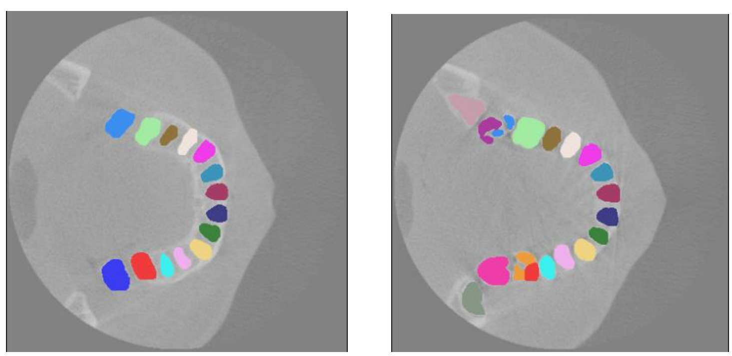


Figure 2. Layered UNet Model

3. Teeth dataset & Warmup trick

Semantic instance segmentation requires a labeled dataset. In previous studies, we used labeled datasets provided by AI-Hub. This data can only be used for learning and cannot be exported externally. Because of this, additional labeling of our own data is necessary to build various models and experiments. So, we labeled our own teeth CBCT data and did various experiments. In this paper, we learned and evaluated the semantic instance model using dental CBCT data of two patients. The tooth dataset used for learning is CBCT data from two patients (patients 1 and 2), and one CBCT consists of 280 slice images. From a total of 560 slice images, 448 were randomly used for training and the remaining 112 were used for validation. Each tooth is labeled according to tooth number like AI-Hub dataset. Fig 3 shows colors of labeled teeth according to the number of teeth. The 10th represents the right upper teeth, the 20th represents the left upper teeth, the 30th represents the left lower teeth, and the 40th represents the right lower teeth. It is by FDI World Dental Federation notation which is the world's most commonly used dental notation (tooth numbering system)[5]. Figure 3 shows colors used for teeth labeling. Fig 4 shows example slices of labeled teeth from CBCT image. It is labeled according to the colors in Figure 3. Fig 4(a) is a labeled slice 186 of teeth CBCT data. Fig 4(b) is a labeled slice 163 of teeth CBCT data.

Figure 3. Colors for teeth labeling



(a) slice 186 **(b) slice 163**

Figure 4. Examples of labeled teeth slice

Tong He et. al. applied a collection of training procedure to improve model accuracy. They are minor “tricks” like modifying the stride size of a particular convolution layer or adjusting learning rate schedule. We adopt learning rate warmup to increase accuracy of instance segmentation model. As it is known, learning rate is a hyperparameter of Neural networks. At the beginning of the training, all parameters are typically random values and therefore far away from the final solution. Using a too large learning rate may result in numerical instability. In the warmup heuristic, A small learning rate is used at the beginning and then switch back to the initial learning rate when the training process is stable[6]. Goyal et al. proposes a gradual warmup strategy that increases the learning rate from 0 to the initial learning rate linearly[7]. In other words, assume we will use the first m batches (e.g. 25 data epochs) to warm up, and the initial learning rate is η, then at batch i, $1 \leq i \leq m$, we will set the learning rate to be $i*\eta/m$. Figure 6 shows various learning rate graphs. Constant learning rate fixed to certain value. Normal learning rate decreases according to a rate of batch(epoch).

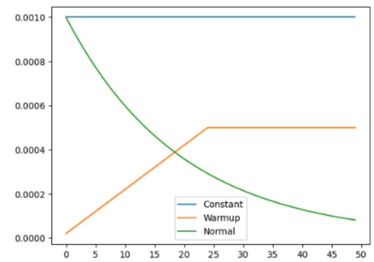


Figure 6. Various learning rate graphs

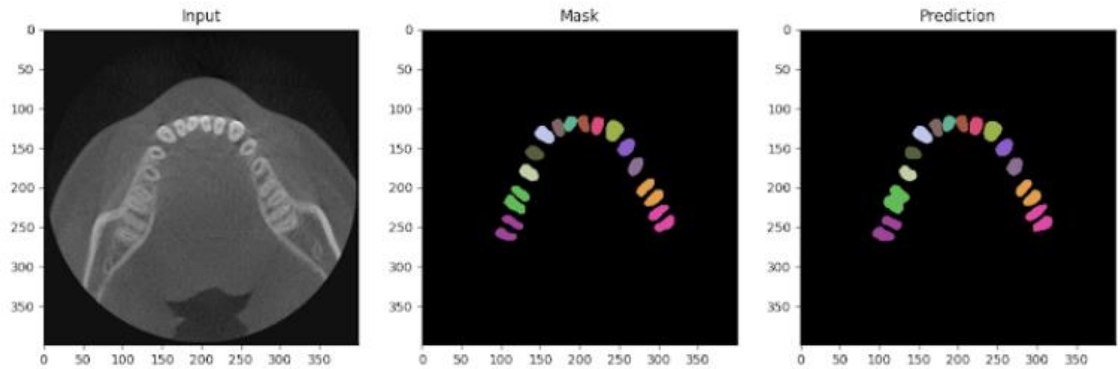
4. Simulation and Results

Using the labeled tooth 3D data set, a simulation was performed to classify gum bones and automatically segment each individual tooth using the layer UNet model. The extracted tooth represents the individual tooth intact, including the tooth root. Table 1 shows the simulation results of the learning and verification data sets using the layer UNet model. Initial LR(Learning Rate) are 1e-3 and 1e-4 and drop-out was used with LR of 1e-3. As a result of the simulation, it is seen that the indicator used for performance evaluation show excellent results. Cross-entropy was used as a loss function for learning layered UNet, and IoU was used as an indicator for evaluating the results. Result with LR of 1e-4 is best in both training and validation. Result without warmup is worst in this case.

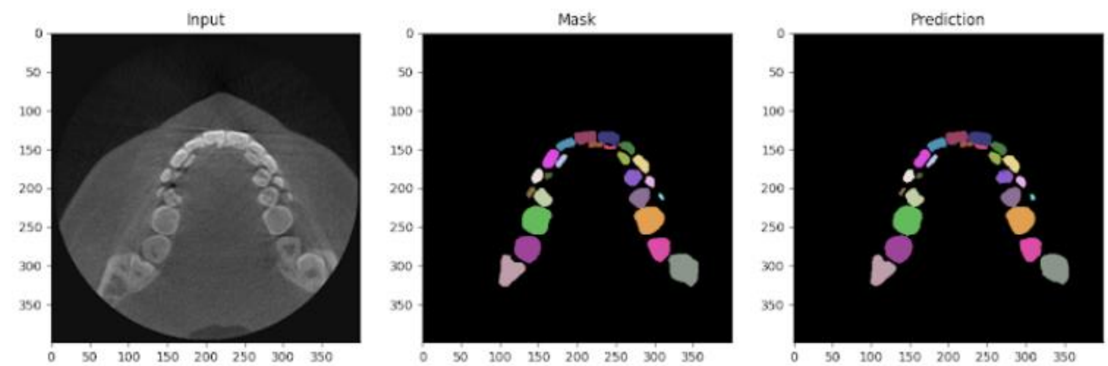
Figure 7 shows the result of tooth instance segmentation for training and validation using the teeth data annotated in the same way as AI-Hub. Figure 7(a) and 7(b) show that slices 164 and 108 from training data. Figure 7(c) shows that slice 96 from validation data. The teeth dataset used for learning is CBCT data from two patients (patients 1 and 2), and one CBCT consists of 280 slice images. The training data consists of a total of 560 slice images [2]. You can see that each tooth number, differentiated by color, is clearly identified and the instance has been divided. In the picture for each slice, the picture in the first column is the original image, and the picture in the second column is the annotation. The third column is the result predicted by layered UNet.

Table 1. Main parameters for learning

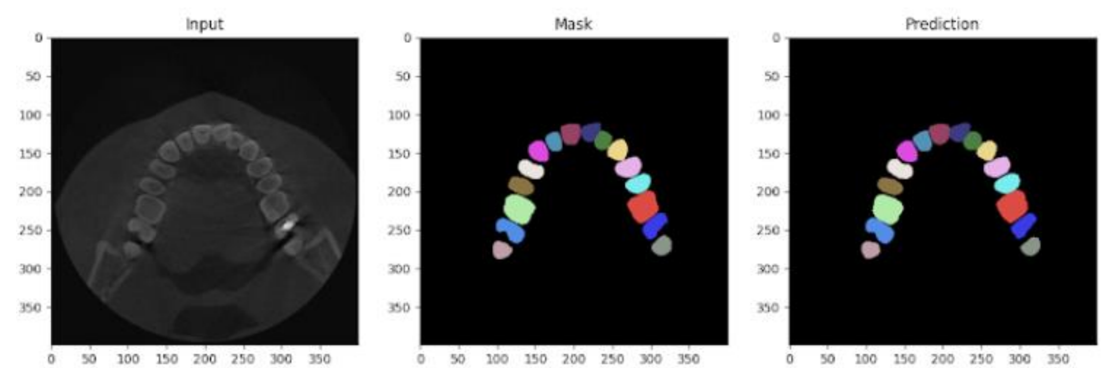
Layered UNet	LR=1e-3 (Dropout)	LR=1e-4	Without Warm-up
Train IoU	0.72	**0.80**	**0.62**
Valid IoU	**0.71**	**0.77**	**0.60**



(a) Slice 164 for training



(b) Slice 108 for training



(c) Slice 96 for validation

Figure 7. Results of instance segmentation of teeth

5. Conclusions

In this paper, layered UNet with warmup trick was used to instance-segment teeth by using data labeled for each individual tooth and increase performance of the result. The layered UNet proposed in the previous paper showed very good performance in tooth segmentation without distinguishing tooth numbers, and it is seen that it does also segment very well for each tooth distinguishing tooth number. To do instance segmentation of teeth, we labeled teeth CBCT data according to tooth numbering system which is devised by FDI World Dental Federation notation. Colors for labeled teeth are like AI-Hub teeth dataset. We simulate using Layered UNet model with different parameters of LR, and with warmup trick and without it. The simulation results show that Layered UNet model using warmup trick and with LR of 1e-4 was the best with IoU values of 0.80 and 0.77 for training, validation data. The results of this paper can be used to provide more efficient treatment in orthodontics than existing methods by considering the roots of the teeth by instance segmentation. In the future research, more data-set are needed to increase accuracy of instance segmentation of teeth and various tricks could be used to add more accuracy to the performance.

Acknowledgement

This work was conducted with support from the 2023 Chungwoon University Academic Research Fund.

References

[1] T. He, et al., "Bag of Tricks for Image Classification with Convolutional Neural Networks," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019 pp. 558-567. DOI: https://doi.org/10.48550/arXiv.1812.01187

[2] J. J. Park, et al., “Teeth Segmentation for Orthodontics based on Deep Learning,” The Transactions of the Korean Institute of Electrical Engineers, vol. 72, no. 3, pp. 440~446, 2023. DOI: https://doi.org/10.5370/KIEE.2023.72.3.440

[3] J. J. Park, et al., “Semantic Segmentation of Teeth using Layered UNet,” The Transactions of the Korean Institute of Electrical Engineers, vol. 72, no. 11, pp. 1470~1476, 2023. DOI: https://doi.org/10.5370/KIEE.2023.72.11.1470

[4] https://keymakr.com/blog/semantic-segmentation-uses-and-applications/

[5] https://en.wikipedia.org/wiki/FDI_World_Dental_Federation_notation

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[7] P. Goyal, P. Dollar, R. B. Girshick, P. Noordhuis, ´ L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, “Accurate, large minibatch SGD: training imagenet in 1 hour,” CoRR, abs/1706.02677, 2017.

[8] Yo-wei Chen, et al., “Artificial intelligence in dentistry: current applications and future perspectives,” QUINTESSENCE INTERNATIONAL, vol. 51, no. 3, March 2020. DOI: https://doi.org/10.3290/j.qi.a43952

[9] H. Ding et al., “Artificial intelligence in dentistry-A review,” vol. 4, Front. Dent. Med. Sec. Dental Materials, 20 February 2023 DOI: https://doi.org/10.3389/fdmed.2023.1085251

[10] Khanagar SB, Al-ehaideb A, Maganur PC, Vishwanathaiah S, Patil S, Baeshen HA, et al, “Developments, application, and performance of artificial intelligence in dentistry-a systematic review,” J Dent Sci. 16(1):508–22, 2021. DOI: https://doi.org/10.1016/j.jds.2020.06.019