Semantic Segmentation of Dental Point Cloud Based on Pointnet ++

Qi Cai[1, a], Yuan Feng[1, b]

[a]1789319416@qq.com, [b]2265346905@qq.com

[1]Software Engineering College, Sichuan University, Chuanda Road, Chengdu, China

Abstract: Using point cloud to store three-dimensional dental model and realize automatic segmentation of tooth boundary will significantly improve the measurement efficiency of arch length, which is of great significance for the measurement of dentition crowding and the formulation of subsequent corresponding correction plans. With the help of the laser scanning model of dental gypsum, it is transformed into point cloud data, and the deep learning network of local fine features and global features is carried out for the dental model to realize the accurate segmentation and extraction of each tooth boundary, so as to assist in measuring the due length of dental arch and the existing length of dental arch. This paper builds a model based on pointnet ++ network structure and tests it with the data set constructed in this paper. The eval point accuracy is 0.719, which has high accuracy and effectively realizes the accurate segmentation of teeth.

Keywords: 3D point cloud, tooth segmentation, PointNet ++

1 Introduction

For the measurement and prediction of dentition crowding, most of the existing methods are carried out on gypsum model. The operation is cumbersome and complex. It requires higher professional requirements for the assessor, and it is time-consuming and laborious.

The automatic segmentation of three-dimensional teeth is of great help to the prediction of dentition crowding and other oral diagnosis and treatment. However, for the methods of using deep learning to segment teeth, most of them can only separate the teeth, but can not realize label assignment, which is not practical in practical application. To solve the above problems, this paper uses point cloud image to save oral surface data, and uses pointnet ++ to realize semantic segmentation of teeth.

ICIDC 2022, June 17-19, Qingdao, People's Republic of China

DOI 10.4108/eai.17-6-2022.2322700

2 Related work

2.1 Automatic Ssegmentation Method Based on CBCT

Hosntalab et al. [2] obtained the tooth model by using bounding box and level set, but the result is poor because the information is obtained from two-dimensional image. Keustermans et al. [4] realized tooth segmentation through graph cut algorithm, but it requires more complex manual operation. Cui et al. [1] proposed a method to automatically and accurately segment and recognize tooth instances in cone beam CT images based on two-level convolutional neural network. Because more manual operations are still needed in the process, and affected by the unclear boundary in complex situations, it is unable to complete the fast and accurate segmentation task.

2.2 Automatic Segmentation Method Based on Point Cloud

For point cloud data, due to its unstructured and class invariance of pose transformation, the traditional CNN can not be used directly on the original point cloud. Multi view CNN (mvcnn) proposed by Su [8] projects point clouds from multiple different perspectives into two-dimensional images. However, due to the use of two-dimensional projection with limited viewing angle, the algorithm loses some of the original features of three-dimensional point cloud, which is general in complex scenes. F. J. lawin et al. [5] applied multi stream FCN to predict the pixel by pixel score on the composite image. The semantic label of each point prediction is obtained by fusing the re projected score into different views. Huang et al. [3] voxelized the disordered point cloud, input the intermediate data into the network for voxel segmentation, and assign the same label to all points in the voxel. Qi [6] proposed pointnet network structure, which directly takes the original point cloud as the input and retains the point cloud information to the greatest extent. However, pointnet ignores the feature relationship between local points and only retains the global features. Therefore, Qi et al. [7] proposed pointnet ++, which is the Network Structure applied in this paper.

3 Proposed algorithm

3.1 Algorithm Overview

The tooth point cloud model obtained by laser scanning has high-precision information of oral surface. In order to realize the feature extraction of point cloud data, the algorithm used in this paper has the following steps:

Step1. STL model data is generated through the patient's dental plaster model, the vertex information is extracted and transformed into point cloud data, and the teeth are labeled according to the doctor's clinical experience. The point cloud information and label information of teeth are used as the input data of the network.

Step2. After the data input of the model, the local and global features of the data are extracted. The farthest point sampling is used instead of random sampling to find "adjacent" points around the centroid to construct the local area set. Use the ball query method to generate local areas and enter them into the pointnet layer.

Step3. Multi scale grouping (MSG) is used to splice the input features of multiple regions, and pointnet is used to deal with MLP and pooling of the input layer, so as to obtain the point set features of the region. By using the network recursively, each point label and segmentation result are finally obtained.

3.2 Data Set Description and Preprocessing

The data set of this paper is 404 cases of oral three-dimensional models provided by Sichuan Univ.HuaXi Dental Hospital. 350 cases are selected as the training data set of this experiment, and the remaining 50 cases are used as the network test set of this experiment.

Since the original data belongs to STL format, the PCL interface is used to convert the data into point cloud format for storage, and then the tooth data is marked through cloudcompare. According to the measurement method of dentition crowding, each tooth (including the first molar) in the anterior arch of the mandibular first molar is taken as the object to be segmented, and the tooth body is classified according to the definition in the dental medicine textbook, 13 labels were used in the anterior arch of the mandibular first molar to separate each tooth of the maxillary and mandibular into one category. The result of the data annotation is demonstrated in Figure 1.

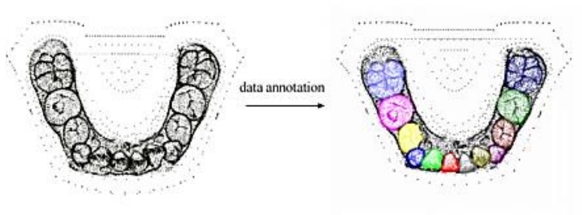

Figure 1: Data Annotation

3.3 Network Structure

Pointnet ++ uses a hierarchical structure to enable the network to provide higher-level features in larger and larger areas. Each set abstraction (SA) includes three parts: sampling layer, grouping layer and pointnet layer.

In the sample layer, the farthest point sampling method is used to sample the input points with size N×(d+C) (n is the number of input points, D is the coordinate dimension, and C is the feature dimension). The following algorithm is used:

$$\min\left(dis(P, A_1), \ldots (disP, A_n)\right) \tag{1}$$

Randomly select a point P, and then select the farthest point from A_1 to A_n to add it to the result set, and iterate this process until the number of points in the result set reaches a given value n'.

In the grouping layer, the center point extracted from the upper layer uses the ball query method to generate N 'areas. By setting the number of midpoint K in the area and the radius r of the ball, set the number in each group, and finally output n' × K×(D + C) size data to pointnet layer.

After obtaining the information of the grouping layer, the pointnet layer first translates the coordinates to a positive number, inputs the pointnet network after normalizing the data, performs MLP and max-pooling on it, and finally outputs n'× (D + C') size data, and C' is the new feature dimension.

After extracting the global features, in order to obtain the local features, it is necessary to carry out reverse interpolation and skip connection. In order to realize inverse interpolation, the following formula is required:

$$f^{(j)}(x)=\frac{\sum_{i=1}^{k}\omega_i(x)f_i^{(j)}}{\sum_{i=1}^{k}\omega_i(x)} \quad where \quad \omega_i(x)=\frac{1}{d(x,x_i)^p},\ j=1,\ldots,C \tag{2}$$

Where x is the middle point of up sampling, C is the dimension of feature space, and X_i is the k points closest to X in the global feature points in the original point cloud coordinate space. By calculating the negative correlation coefficient ω_i of the distance between Xi and X. The feature is weighted and summed with the coefficient as the weight, so as to realize the up sampling and back transmission of the feature.

After up sampling, skip connection is performed to directly concatenate the features of the corresponding layer of the previous encoder. Since the features of the corresponding layer are extracted from the larger upper layer, they represent the local features to a certain extent. Through step-by-step sampling, the local + global point wise feature is finally obtained.

4 Experimental results and analysis

The network experiment results verify the classification accuracy by using the evaluation methods of intersection over Union (IOU) and eval point accuracy. IOU is the ratio of intersection and union between calculated ground truth and predicted segmentation, which measures the accuracy of semantic segmentation, i.e

$$IoU=\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{3}$$

Among them, there are K+1 classes in the classification sample. I represents the real value, J represents the predicted value, and PIJ represents predicting I as J.

Eval point accuracy is the ratio of the correctly identified point set to the whole data point set, i.e

$$\frac{tp}{n} \tag{4}$$

Where TP is the correctly identified point set and N is the whole data point set.

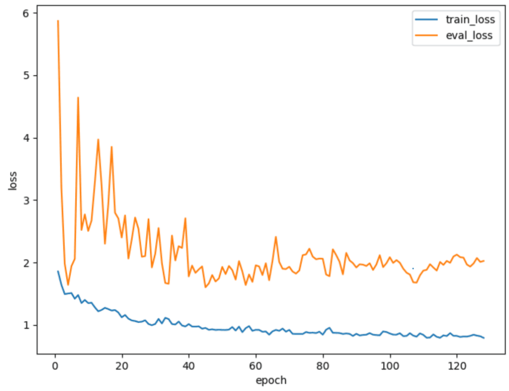

Figure 2: Development Trend of Loss Function

IoU For Each Teeth Class

Class 0	IoU: 0.527	Gingiva
Class 31	IoU: 0.335	Central Incisor(Left)
Class 32	IoU: 0.368	Second Incisor(Left)
Class 33	IoU: 0.301	Canine(Left)
Class 34	IoU: 0.239	First Premolar(Left)
Class 35	IoU: 0.395	Second Premolar(Left)
Class 36	IoU: 0.337	First Molar(Left)
Class 41	IoU: 0.333	Central Incisor(Left)
Class 42	IoU: 0.264	Second Incisor(Left)
Class 43	IoU: 0.258	Canine(Left)
Class 44	IoU: 0.416	First Premolar(Left)
Class 45	IoU: 0.425	Second Premolar(Left)
Class 46	IoU: 0.344	First Molar(Left)

Figure 3: IoU for Each Teeth Class

In this paper, the training set is input into the learning network and after 120 epochs, the results shown in Figure 2 are obtained. It can be seen that the loss function has decreased significantly and its fluctuation degree has been better improved. At the same time, the loss function of the training set and the test set tends to be close, which shows that the robustness of the model is improving. Through the calculation of Eval point accuracy and IoU, the final Eval point accuracy calculation result is 0.719, and the segmentation result is more accurate. The IoU of each label is shown in Figure 3. Through comparison, it is found that the algorithm has the best segmentation effect between gums and teeth, and the recognition of canine teeth is better than other teeth.

5 Conclusion

In this paper, the oral scanning data is stored based on point cloud, and the input data is recursively learned through pointnet + + network structure to realize the feature extraction from local to whole, so as to obtain the accurate segmentation and label matching of a single tooth. This algorithm can achieve good results in the standard tooth model, but there are some limitations in the case of tooth overlap or missing teeth, and the segmentation effect is easy to be confused. In the future work, it will further increase the learning times of the network and the amount of clinical data input, improve the complexity of tooth model input, and improve the universality of the algorithm.

References

[1] Cui, Z., Li, C., & Wang, W. (2019). ToothNet: automatic tooth instance segmentation and identification from cone beam CT images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6368-6377).

[2] Hosntalab, M., Aghaeizadeh Zoroofi, R., Abbaspour Tehrani-Fard, A., & Shirani, G. (2008). Segmentation of teeth in CT volumetric dataset by panoramic projection and variational level set. International Journal of Computer Assisted Radiology and Surgery, 3(3), 257-265.

[3] Huang, J., & You, S. (2016, December). Point cloud labeling using 3d convolutional neural network. In 2016 23rd International Conference on Pattern Recognition (ICPR) (pp. 2670-2675). IEEE.

[4] Keustermans, J., Vandermeulen, D., & Suetens, P. (2012, October). Integrating statistical shape models into a graph cut framework for tooth segmentation. In International Workshop on Machine Learning in Medical Imaging (pp. 242-249). Springer, Berlin, Heidelberg.

[5] Lawin, F. J., Danelljan, M., Tosteberg, P., Bhat, G., Khan, F. S., & Felsberg, M. (2017, August). Deep projective 3D semantic segmentation. In International Conference on Computer Analysis of Images and Patterns (pp. 95-107). Springer, Cham.

[6] Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 652-660).

[7] Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems, 30.

[8] Su, H., Maji, S., Kalogerakis, E., & Learned-Miller, E. (2015). Multi-view convolutional neural networks for 3d shape recognition. In Proceedings of the IEEE international conference on computer vision (pp. 945-953).