2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Root Canal Segmentation in CBCT Images by 3D U-Net with Global and Local Combination Loss

Jian Zhang[1], Wenjun Xia[2], Jiaqi Dong[3], Zisheng Tang[4], and Qunfei Zhao[5]

Abstract—**Accurate root canal segmentation provides an important assistance for root canal therapy. The existing research such as level set method have made effective progress in tooth and root canal segmentation. In the current situation, however, doctors are required to specify an initial area for the target root canal manually. In this paper, we propose a fully automatic and high precision root canal segmentation method based on deep learning and hybrid level set constraints. We set up the global image encoder and local region decoder for global localization and local segmentation, and then combine the contour information generated by level set. Through using CLAHE algorithm and a combination loss based on dice loss, we solve the class imbalance problem and improved recognition ability. More accurate and faster root canal segmentation is implemented under the framework of multi-task learning and evaluated by experiments on 78 Cone Beam CT images. The experimental results show that the proposed 3D U-Net had higher segmentation performance than state of the art algorithms. The average dice similarity coefficient (DSC) is 0.952.**

Clinical Relevance— **We propose an end-to-end automatic root canal segmentation method, which had high accuracy and can reduce the workload of marking samples for dentists. It can also be used for root canal treatment planning and preoperative evaluation.**

I. INTRODUCTION

In the diagnosis and treatment of pulp and periapical diseases, root canal therapy is the most important treatment. Endodontics specialists need the assistance of good radiological examination. At present, with the wide application of oral Cone beam CT (CBCT) image in clinic, it provides comprehensive 3D spatial information for root canal and teeth, which makes many researchers study efficient root canal segmentation methods [1-5]. However, in most clinical root canal therapy, the root canal area is manually depicted by doctors in oral CBCT images, which is not only subjective but also time-consuming. Therefore, automatic and accurate root canal segmentation technology is of great significance for root canal therapy.

However, accurate root canal segmentation from oral CBCT images is very challenging, the main reasons are as follows. (1) CBCT image scanning covers the complex environment around the teeth, such as alveolar bone and noise; (2) The boundary is ambiguous due to the low contrast between the apical part of the root canal and its surrounding teeth. (3) The shape of the root canal is irregular, different root canals have different branches. (4) High density occlusive restorations and restorations affect the CBCT image scanning, thus affecting the image quality.

In recent years, some scholars have studied the segmentation of teeth and root canals in CBCT images. Based on the traditional level set and a series of improved methods are proposed [6]. And it has a good application in tooth segmentation. Based on the above challenges, the level set method is lack of robustness to complex fuzzy boundaries and needs manual initialization and interaction. With the development of convolutional neural network, the deep learning method shows good performance in medical image. Compared with traditional methods based on manual features, the deep learning method is more automated, simplifies the workflow, eliminates manual intervention, and can process many images quickly. Cui et al. [7] used deep convolution neural network to realize automatic and accurate segmentation and recognition of teeth from CBCT images. Ronneberger et al. [8] proposed a 3D U-shaped network for medical image segmentation, which is widely used in other fields. Generally, deep learning methods will face the following challenges in medical images: lack of shape prior, weak image intensity specificity, imbalanced class, requiring lots of training data, high requirements for GPU and CPU memory resources. To our best knowledge, the application of deep learning method in root canal CBCT image is few.

In this paper, we propose an improved 3D deep neural network called root canal segmentation network (RCS-Net) for root canal segmentation from CBCT images. An improved 3D region of interest (ROI) network is developed. In RCS-Net, the ROI is located from 3D space, and the root canal is segmented from the region of interest. RCS-Net can be divided into two stages, the global stage and the local stage. In the global stage, a shared global image encoder is used to locate the root canal center, so as to get the ROI in the original image. In the local stage, a local region decoder based on the global features is used for ROI segmentation with high accuracy. Motivated by the importance of the accurate tip of root canal in clinical surgery, we present a multitask combination dice loss function to focus on the boundary voxels of root canal, so as to further improve the accuracy. Comparisons and extensive ablation experiments show that our proposed framework is better than other related works.

[1]Jian Zhang, [3]Jiaqi Dong and [5]Qunfei Zhao (e-mail: zhaoqf@sjtu.edu.cn) are with Department of Automation, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, P.R.China.

[2]Wenjun Xia, [4]Zisheng Tang (e-mail: tangzisheng163@163.com) are with department of Endodontics, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine; College of Stomatology, Shanghai Jiao Tong University; National Center for Stomatology; National Clinical Research Center for Oral Diseases; Shanghai Key Laboratory of Stomatology.

*Research supported by Clinical Research Program of 9th People's Hospital affiliated to Shanghai Jiao Tong University School of Medicine (JYLJ201815) 3D Program (GXQ202006), and by National Natural Science Foundation of China (U2013205).

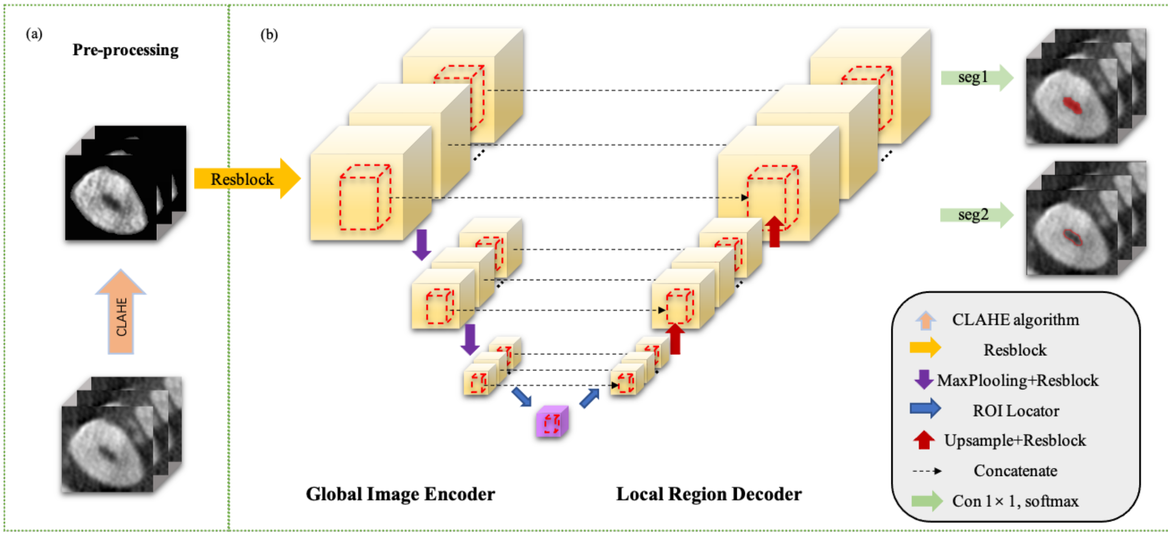

Fig. 1. Description of our framework, which can be divided into two parts: (a) In pre-processing, all images are normalized by CLAHE algorithm. (b) The network consists of the Global Image Encoder and Local Region Decoder, residual blocks are placed between neighboring downsampling or upsampling layers. Outputs are root canal prediction voxels and contours.

II. METHODS

As shown in Figure 1, this paper refers to the basic framework of 3D U-net [9] convolution neural network and constructs a segmentation model for tooth root canal in 3D CBCT image. In the global stage, we input the whole 3D image into the global image encoder for feature coding and get the spatial location of ROI. Then in the local stage, a local decoder is designed to segment the local boundary of the root canal. The local region with smaller receptive field can receive features extracted from the entire images with such operation, and the segmentation results are refined. The root canal can be segmented accurately and automatically with low GPU memory cost. The whole model is trained by end-to-end with multi-task combination loss function.

A. Network Architecture

The RCS-Net consists of an encoder for global image feature extraction and a local decoder for pixel level prediction. In the down-sampling of the encoder, the parameters of 3D convolution core are too many for general computers due to the limited GPU memory. We use the residual block in the encoder, which is conducive to accelerating network convergence and improving performance [10]. Mainly to avoid the gradient explosion and over fitting phenomenon in the training process.

Considering the small size of root canal, each residual block contains three convolution layers, the convolution layers corresponding to three normalization layers and ReLU layers, and a skip connection. Compared with the original 3D U-net network, our model size and parameters are smaller. After the process of down-sampling, the encoder could obtain the spatial information of ROI and learn the accurate bounding box. With a feature set cropped from the encoder path by applying the multi-level feature fusion mechanism, we construct a local region decoder for fine segmentation in ROI. The structure of the local decoder is partially symmetrical with that of the encoder, and the feature maps of the corresponding scales are fused through the skip connections. The decoder uses transposed convolution for image up-sampling, which branch's feature tensors is much smaller than that of the encoder. The module recovers the spatial dimension of ROI without losing local details.

B. Multi-task combination loss function

We design a multi-task combination loss function based on Dice loss to smooth the learning process, which makes the local decoder focus on the accurate contour of root canal. We add an additional softmax layer, which is considered as an auxiliary task for perceiving the root canal contour, the additional softmax layer are trained in parallel with the segmentation task. It also shares feature map at the output, as shown in Figure 1, to predict the root canal contour voxels. At the same time, this operation can solve the class imbalance problem of CBCT images.

Generally, Dice coefficient loss function is one of the most commonly used loss functions in medical image segmentation [11], this efficient automatic class balancer is expressed as:

$$L_{Dice} = 1 - 2 \times \frac{\sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + \epsilon} \qquad (1)$$

where the predicted volume $p_i \in P$ and the ground truth volume $g_i \in G$. ϵ is a small smoothness term, set to 10^{-4}. In the optimization phase, dice loss is minimized by gradient descent using the following derivatives:

$$\frac{\partial L_{Dice}}{\partial p_k} = -2 \times \frac{[\sum_{i=1}^{N}(p_i g_i) + \epsilon] - g_k[\sum_{i=1}^{N}(p_i + g_i) + \epsilon]}{[\sum_{i=1}^{N}(p_i + g_i + \epsilon)]^2} \qquad (2)$$

where $k = 1,2...N$.

Dice loss can accelerate the training process, but it is easy to cause training difficulties because the class imbalance problem of CBCT images. The contour extraction auxiliary task after the main region segmentation and the region segmentation are trained at the same time. Specifically, this additional output head is set to predict the contour, this loss function named L_{local} is denoted as follows:

$$L_{local} = L_{Dice}(P_r, G_r) + \lambda_j L_{Dice}(P_c, G_c) \tag{3}$$

where λ_j equal 0.5 is the additional task weight, it is less than 1 to ensure the regional tasks are dominant. P_r and G_r represent the predicted voxels and the ground truth voxels in ROI respectively, P_C and G_C represent the predicted voxels and the ground truth voxels of contour.

We combine the above two loss functions as a combination loss function and train the whole network end-to-end:

$$L_{total} = L_{Dice} + L_{local} + \beta\|\omega\|^2 \tag{4}$$

where β represents the balance of weight parameter and ω represent the parameters of the whole framework. Their value are obtained through experiments.

C. Details

Pre-processing: In view of the gray characteristics of the tooth image, the whole CBCT image is enhanced adaptively by using CLAHE algorithm [12], so as to enhance the local contrast, balance the gray distribution, and save the calculation consumption, as shown in Figure 2. The tooth region is roughly segmented after adaptive enhancement, which can achieve unified processing in different scanning environments. In addition, all the images should be normalized to 128×128×256.

Contour label: We use level set model [13] to generate contour label automatically, the quality of the label is certified by a senior stomatologist.

Post-processing: The output of model usually contains some false-positive areas. In practice application, root canal is an independent single connected component, we use connected component analysis to extract the maximum area of predicted voxels and remove some small false-positive voxels.

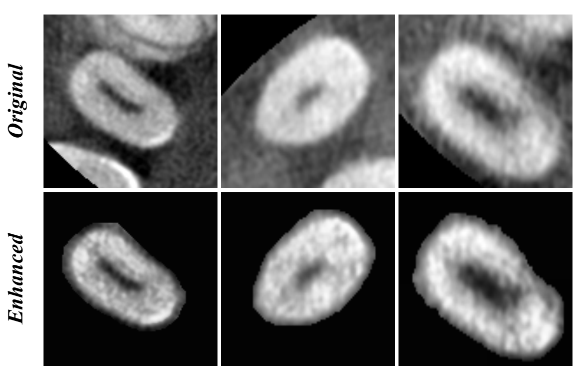

Fig 2. Comparison of the original and enhanced images by CLAHE algorithm under 2D slice.

III. EXPERIMENTS

A. Dataset and experiment settings

Dataset: The dataset used in this study included 78 tooth CBCT images, and their voxel spacing is 0.1 mm×0.1 mm× 0.1 mm, and the volume size is 128×128×(128 - 384), from Shanghai Ninth People's hospital. The dataset is divided into training set and test set, the training set includes 51 CBCT images with 9120 slices and test set includes 17 CBCT images with 4530 slices. Experienced radiologists manually drew the ground truth of root canal.

Experiment settings: We perform data augmentation when feeding training samples to avoid overfitting. The image after adaptive enhancement was randomly scaled by 0.9 to 1.1 times, rotated as a whole, and intensity jittering.

B. Network training

All the networks are implemented on PyTorch. We use Adam [14] optimizer to optimize the entire framework. The initial learning rate is 10^{-4}. The weights of convolution kernel are penalized by L2 norm. Then we first train the ROI locator with loss L_{local}, and finally train the entire framework with loss L_{total} . We train 100 epochs on NVIDIA Titan Xp GPU with 12GB GPU memory for 20 hours.

C. Evaluation metrics

We use four evaluation parameters to evaluate our method. In terms of the overall segmentation performance, we use Dice Similarity Coefficient (DSC) (5) and Volumetric Overlap Error (VOE) (6) to evaluate it quantitatively. As for root canal contour and boundary segmentation performance, we use Average Symmetric Surface Distance (ASSD) (7) and Maximum Symmetric Surface Distance (MSSD) (8) to evaluate the segmentation accuracy.

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \tag{5}$$

$$VOE = 1 - \frac{|P \cap G|}{|P| \cup |G|} \tag{6}$$

$$ASD = \frac{\sum_{a \in dP} D(a, G) + \sum_{b \in dG} D(b, P)}{|dP| + |dG|} \tag{7}$$

$$MSD = max[\min_{a \in dP} D(a, G), \min_{b \in dG} D(b, P)] \tag{8}$$

where $|\cdot|$ stands for L_1 norm, P for the predicted voxels and G for the ground truth. dP is the contour voxels sets of P, and dG the contour voxels sets of G. $D(v, V)$ is the minimum distance from the voxel v to the voxel set V.

D. Results and discussion

As shown in Figure 3, we illustrate the segmentation result of our model from three plane slices. We evaluate 3D Mask RCNN [7], 3D U-net and V-net [11] in the same way as the proposed RCS-Net, and conduct ablation experiments to validate the effects of our proposed module. The quantitative comparison is shown in Tab. 1. Through image preprocessing and adding L_{total} loss, RCS-Net has better performance than other methods. Our average DSC is 0.952 and VOE is 0.077. The ASD and MSD are also lower than other methods, which means that the segmentation effect on the root canal contour is better than other methods. Figure 4 shows the corresponding 3D reconstruction and contour prediction results. The contour of root canal can also be used as an auxiliary output for stomatologist's reference.

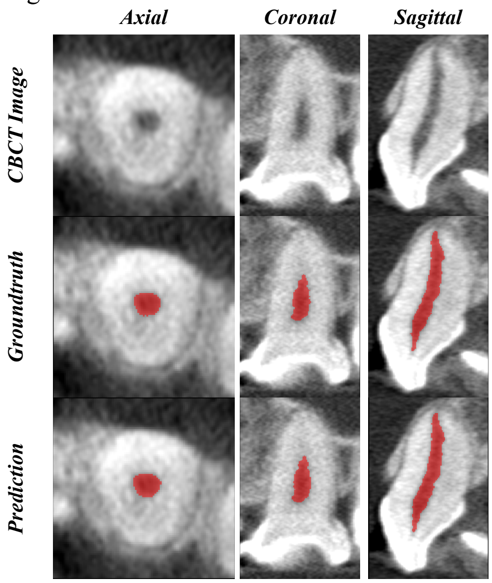

Fig 3. Results observed from different anatomical planes. The red label area mask represents ground truth of root canal, the following picture is the output of our model. Showing the robustness of our network.

RCS-Net is also suitable for uncommon root canals due to the prior condition of root canal profile. As shown in Figure 4 (c), it is also very effective for root canals with apical bifurcations. Our proposed RCS-Net is fully automatic in all stages, including image preprocessing, segmentation, level set contour acquisition and post-processing.

TABLE I.
THE QUANTITATIVE COMPARISON OF DIFFERENT METHODS AND ABLATION STUDY (MEAN ± STD)

Method	DSC	VOE	ASD(mm)	MSD(mm)
3D Mask RCNN[7]	0.897	0.135	0.192	0.534
3D U-net[9]	0.915	0.107	0.187	0.517
V-net[11]	0.929	0.095	0.161	0.472
ROI locator + Dice U-net	0.943	0.081	0.138	0.393
ROI locator + Dice U-net + Dice Local (Proposed)	**0.952**	**0.077**	**0.125**	**0.332**

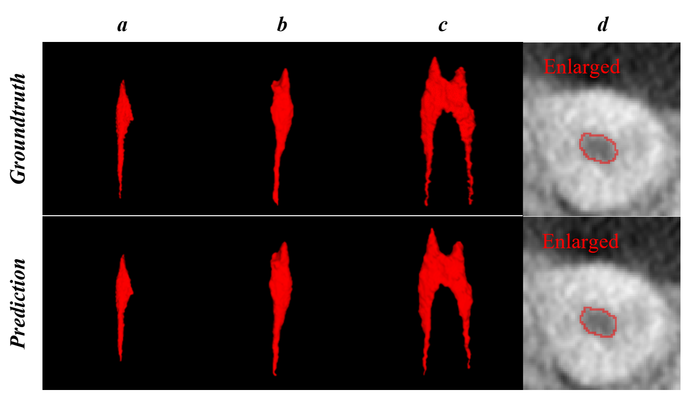

Fig 4. Segmentation results. Our segmentation results (Prediction) are highly consistent with the results of dentists (Ground truth), the contour line helps doctors accurately determine where the boundary of root canal is. (a), (b), (c) represent the visualization results of three different root canals, and (d) represents the contour line under the xoy plane.

The whole algorithm only needs about 350 ms to get the segmentation result of 3D CBCT image, this is much faster than manual labeling.

IV. CONCLUSION

This paper presents an automatic root canal segmentation network called RCS-Net based on deep learning and contour prior. In RCS-Net, we take the importance and effectiveness of global stage and local segmentation into consideration, combine with the root canal contour constraints for fast and accurate root canal segmentation. Our experimental results show that our network had higher segmentation performance than the existing segmentation algorithms. The results of segmentation are approved by our cooperated clinicians. In view of the effectiveness and practicability of RCS-Net, we hope that our method can be extended to other tubular organ segmentation tasks.

REFERENCES

[1] J. Hatvani, A. Horváth, J. Michetti, A. Basarab, D. Kouamé and M. Gyöngy, "Deep Learning-Based Super-Resolution Applied to Dental Computed Tomography," in IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 3, no. 2, pp. 120-128, March 2019.

[2] Wang L, Li JP, Ge ZP, Li G. "CBCT image based segmentation method for tooth pulp cavity region extraction," Dentomaxillofac Radiol. 2019 Feb;48(2):20180236.

[3] Seung-Min Lee, Sewoong Oh, Jinmi Kim, You-Min Kim, Youn-Kyung Choi, Hyun Ho Kwak, Yong-Il Kim, Age estimation using the maxillary canine pulp/tooth ratio in Korean adults: A CBCT buccolingual and horizontal section image analysis, Journal of Forensic Radiology and Imaging, Volume 9, 2017, Pages 1-5.

[4] Rose Sfeir, Jérôme Michetti, Bilal Chebaro, Franck Diemer, Adrian Basarab, et al.. Dental root canal segmentation from super-resolved 3D cone beam computed tomography data. IEEE Nuclear Science Symposium and Medical Imaging Conference, Oct 2017, pp.1-2.

[5] Harrison J., Chantrel S., Schmittbuhl M., de Guise J.A. (2019) Segmentation and 3D-Modelling of Single-Rooted Teeth from CBCT Data: An Automatic Strategy Based on Dental Pulp Segmentation and Surface Deformation. In: World Congress on Medical Physics and Biomedical Engineering 2018. IFMBE, vol 68/1. Springer, Singapore.

[6] Z. Xia, Y. Gan, J. Xiong, Q. Zhao and J. Chen, "Crown Segmentation From Computed Tomography Images With Metal Artifacts," in IEEE Signal Processing Letters, vol. 23, no. 5, pp. 678-682, May 2016.

[7] Z. Cui, C. Li and W. Wang, "ToothNet: Automatic Tooth Instance Segmentation and Identification From Cone Beam CT Images," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 6361-6370.

[8] Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham.

[9] O¨. C ¸ic ¸ek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ron- neberger, “3d u-net: learning dense volumetric segmentation from sparse annotation,” in International conference on medical image computing and computer-assisted intervention, pp. 424–432, 2016.

[10] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, pp. 770-778.

[11] F. Milletari, N. Navab and S. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," 2016 Fourth International Conference on 3D Vision (3DV), pp. 565-571.

[12] Zuiderveld, Karel. "Contrast limited adaptive histogram equalization." Graphics gems IV. Academic Press Professional, Inc., 1994.

[13] Wang X F, Huang D S, Xu H. An efficient local Chan-Vese model for image segmentation[J]. Pattern Recognition, 2010, 43(3): 603-618.

[14] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.