



Automatic Image Annotation and Deep Learning for Tooth CT Image Segmentation

Miao Gou¹, Yunbo Rao¹(✉), Minglu Zhang¹, Jianxun Sun², and Keyang Cheng³

¹ School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, Sichuan, People's Republic of China

1003632332@qq.com, raoyb@uestc.edu.cn,
2016220302032@std.uestc.edu.cn

² West China School of Stomatology, Sichuan University,
Chengdu 610041, Sichuan, People's Republic of China
jxsun@scu.edu.cn

³ School of Computer Science and Telecommunications Engineering, Jiangsu University,
Zhenjiang 212013, People's Republic of China
kycheng@ujs.edu.cn

Abstract. Recently, convolutional networks show great ability dealing with the problem of biomedical imaging, such as tooth image segmentation. In this paper, we propose a novel tooth-based computer tomography (CT) image segmentation approach that integrates U-Net with a level set model. Compared with a single U-Net, our method uses the level set method to build the mask for CT images. This allows automatic annotation in our model, improving the efficiency on image segmentation. Furthermore, we make some changes to the origin U-Net structure for the feasibility to images of any sizes. Using the combination of these two models, our integrated method shows its superiority dealing with problems on tooth image segmentation, outperforming the U-Net or the level set model alone.

Keywords: Image segmentation · Automation image annotation · Level set · U-Net · Convolutional networks

1 Introduction

Image segmentation has been one of the most essential tasks in biomedical imaging research. Biomedical image segmentation can be used to support medical diagnosis, displaying the real condition on patients. Here have two limitations of the traditional method. (1) due to the complexity of physical structure for human body, especially for tooth's image segmentation, building the mask for every image manually is time-consuming and tedious. (2) the number of labelled for image segmentation is also much smaller than other types on the internet. In this paper, our main aim is using the unlabeled data as the training set for tooth's image segmentation.

Some work has been done for achieving unlabeled data in image segmentation. Lin et al. [1] proposed a method using scribbles to annotate images, and train convolutional networks for semantic segmentation supervised by scribbles. That enables researchers

annotate training set in more efficient way. Pathak et al. [2] presented an approach to learn a dense pixel-wise labeling from image-level tags. So, image-level tags can be effectively used by Convolutional Neural Network (CNN) classifier to transfer them into predicted labels. Papandreou et al. [3] develop expectation-maximization methods for semantic image segmentation under these weakly supervised or semi-supervised settings. This deep convolutional network performs successfully even with significantly less annotation effort. However, these networks still need researchers to label every image more or less.

The level set method is popular in image segmentation. The traditional method represented the interface as the level-set of a higher dimensional function [4, 5]. The advantage of implicit representation of a moving front in level set method is its ability to naturally handle changes in topology. Li et al. [6] proposed distance regularized level set evolution as an edge-based active contour model. Though it can be implemented by a simpler and more efficient numerical scheme than conventional level set methods, it relies on good contour initialization. Recently, Allaire et al. [7] proposed a framework to handle geometric constraints related to local thickness. Since the initial guesses and the specific treatment of the constraints are crucial for some topological changes, the resulting shapes are strongly dependent on these. Yang et al. [8] embedded a Markov random field energy function to the conventional level set energy function. This method is robust against various kinds of noised. Morar et al. [9] proposed active contour model without edges. It has better noise immunity, and widely been spread widely now. The level set method has been widely used for tooth segmentation in biomedical imaging research due to its superiority dealing with topological changes and contour propagation, such as [10, 11]. However, all these methods need to iterate hundreds of times only for a single image segmentation.

To solve these problems, we adopt level set method for the automatic segmentation of tooth structure from CT image data, which use the curve evolution, the initial curve converges to the image boundary and the output is used as image annotation. In our work, we adopt some manual annotations as part of our training dataset to ensure the accuracy. And then, we combine these datasets and our new U-Net model to extract the feature map of the CT images, constructing the neural network, updating the weight of the model through iteration to obtain the optimal model. The final result of our method shows proved efficiency and accurate segmentation.

2 Methods

An overview of the proposed U-Net and Level-set framework is shown in Fig. 1. Automatic image annotation is pre-trained with a small dataset with labels. Then, we use the data enhancement strategy to better utilize the data and expand the dataset. A new level set method is inputted into our model for processing the dataset. Last, a new U-Net constitutes the next part of our method is proposed as training the model.

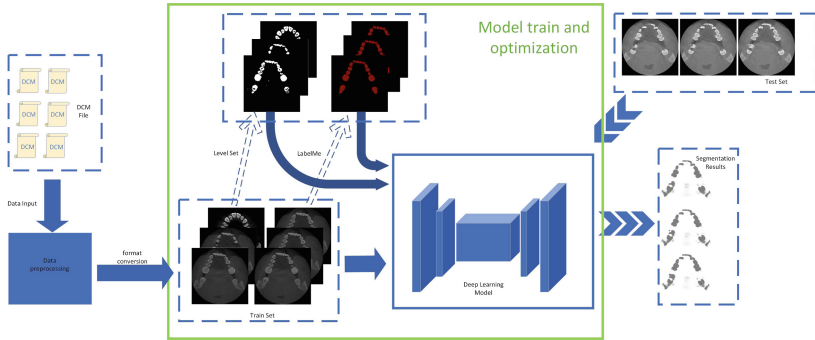


Fig. 1. The framework of our proposed method. The blue dotted box represents the input and output files, it is the pre-processing of DICOM images (Left). After conversion, the model is trained by our deep learning model, the automatic annotation using level set and manual annotation are also included in this part (Middle). The final result (Lower right) is tested by the test set (Upper right) (Color figure online)

2.1 Dataset

In this paper, we use the dataset from West China School of Stomatology. This dataset contains 5 group of complete scan results, 401 original CT images of tooth in total. Since the original DICOM image contains many other useless information for tooth segmentation, we filter out header data for communication and then transfer it into PNG images and converting the 16-bit int data to the 8-bit unit data. Besides, the Hounsfield unit (HU) of the annotated image in the dataset is $[0, 1]$, and that range of value in our model is not obvious to identify. So, we modified the total range of the HU value in PNG image into $[0, 255]$, contrasting the CT scanned tooth image and its mask. The transferred images and their masks are shown in Fig. 2.

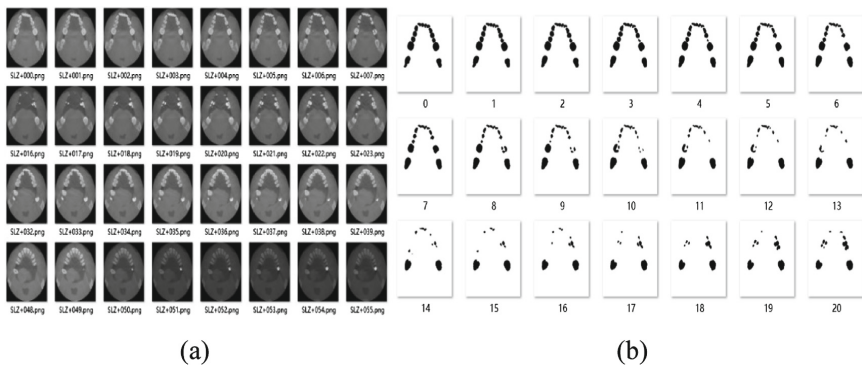


Fig. 2. The tooth images and image masks. (a) the sample of tooth images, (b) the sample of tooth masks

2.2 Automatic Image Annotation

In our work, we propose a level set algorithm to realize automatic annotation. The proposed method first is to set the dynamic parameters by using the active contour model of the following form:

$$\frac{\partial C(s, t)}{\partial t} = F * N \tag{1}$$

where F means the velocity function that controls the evolution of the curve, N is the normal vector in the curve. The formula C (s, t) is rewritten into a zero level set function $\phi(x, y, t)$, the initialization of ϕ is definition as follows:

$$\phi = \begin{cases} -d, & \text{outside the curve} \\ 0, & \text{on the curve} \\ +d, & \text{inside the curve} \end{cases} \tag{2}$$

where d defines the shortest distance from the point to the curve. The level set model has the energy function:

$$E = \mu \int_{\Omega} (|\nabla\phi| - 1)^2 dx dy + \lambda_1 \int_{inside(c)} |\mu_0(x, y) - c_1|^2 dx dy + \lambda_2 \int_{outside(c)} |\mu_0(x, y) - c_2|^2 dx dy \tag{3}$$

Ω is the whole image domain, c_1 denotes the gray mean inside the evolution curve, c_2 denotes the gray mean outside the evolution curve, $\mu, \lambda_1, \lambda_2$ are constants. Where the $\mu \int_{\Omega} (|\nabla\phi| - 1)^2 dx dy$ is the distance constraint term, it keeps the level set function consistent with the sign distance function in the evolution process. The $\lambda_1 \int_{inside(c)} |\mu_0(x, y) - c_1|^2 dx dy$ and $\lambda_2 \int_{outside(c)} |\mu_0(x, y) - c_2|^2 dx dy$ defines the external energy term, they represent the difference of gray mean value of each region inside and outside evolutionary curve respectively.

To ensure the continuity and smoothness of energy functional, Heaviside function is used as follows:

$$H(z) = \frac{1}{2} \left[1 + \frac{2}{\pi} \arctan\left(\frac{z}{\varepsilon}\right) \right] \tag{4}$$

where ε is a positive number, which approaches 0. Joint energy functional Eq. (3) with Eq. (4), and calculate the new energy functional, the segmentation result of image boundary is obtained. On the process of the automatic annotation of the level set method, we propose the new energy function as follows:

$$E = \mu \int_{\Omega} (|\nabla\phi| - 1)^2 dx dy + \lambda_1 \int_{\Omega} |\mu_0(x, y) - c_1|^2 H(\phi(x, y)) dx dy + \lambda_2 \int_{\Omega} |\mu_0(x, y) - c_2|^2 H(\phi(x, y)) dx dy \tag{5}$$

These paramaters $\lambda_1, \lambda_2, c_1, c_2$ are consistent with formula (3). The final results of automatic image annotation prove our superiority and are shown in Fig. 3.

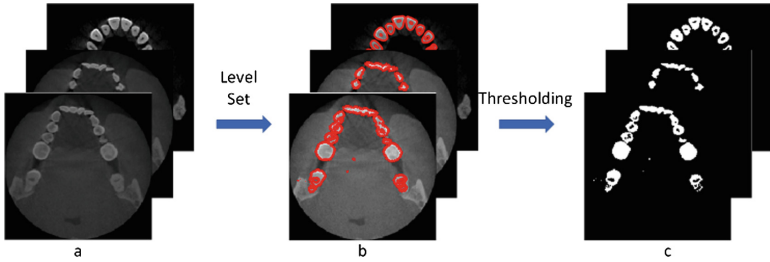


Fig. 3. The result of image boundary and image segmentation of level set method. (a) The original tooth CT images, (b) After 500 curve evolution results using level set, (c) automatic image annotation result.

2.3 Model Training and Optimization

Referring to the classical deep convolution network model U-Net [12–14], we build a new U-Net architecture for tooth segmentation in our work. The architecture consists of a contracting path for classification and an expanding path for precise localization. In this paper, the network is composed of 5 groups of nodes in the down-sampling stage, each follows two 3*3 convolutional layers and one 2*2 max pooling layer. In the up-sampling stage, each node contains two 3*3 convolutional layers and one 2*2 convolutional layer. The proposed method is shown Algorithm 1:

Algorithm 1 Model train stage method

Input : tooth images and annotation tooth CT images;

Output :

Step 1: Model weight initialization;

Step 2: Forward propagation and weighting operations using image data;

Step 3: Using loss function to compare image and mask weights;

Step 4: update weight and loss for back propagation: if the loss is small enough: save the model weight, if not: go to step 2;

Step 5: Optimal model obtained by iteration.

In this paper, our proposed new U-Net has 30 layers in total for model training. The network architecture is shown in Fig. 4.

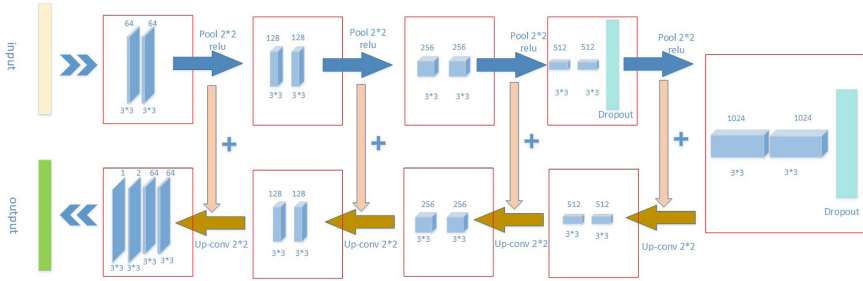


Fig. 4. The architecture of our proposed new U-Net (30 layers). The blue cuboid represents a 3*3 convolution layer, the cyan rectangle is one dropout layer, the blue arrow represents a 2*2 max pooling layer, the brown arrow represents a 2*2 deconvolution layer, pink arrows represent the clipping supplement of high-level information. (Color figure online)

3 Experiments and Analysis

Our method was evaluated with the dataset from West China School of Stomatology. After screening and pre-processing, we get 400 valid images. In this paper, these images are divided to training set and validation set according to the ratio of 3:1, training the deep network and test the prediction results, respectively. The final evolution curve is binarized, where the internal gray value of the curve is set to 255, the external gray value of the curve is set to 0. The marked pictures are obtained. The related constant parameters of the deep network are shown as follows Table 1:

Table 1. Deep network parameters (ES = evolutionary step)

Parameter	λ_1	λ_2	μ	ES	Iterations
Value (1)	1	1	0.02	0.5	500
Value (2)	0.5	2	0.02	1	500
Value (3)	1	1	0.02	0.5	800

Here, our experiment has been using three group parameters, λ_1 controls the influence of the internal energy term in the control curve for evolution results. λ_2 controls the influence of the external energy term in the control curve for evolution results. μ weakens the deviation between φ function and sign distance function while not affecting the evolutionary process. Evolutionary step determines the minimum step of curve evolution. Experiments show the inconformity between λ_1 and λ_2 can result in big error for curve convergence. And the target value of tooth size should not be too large. In our experiment, 500 iterations can complete the convergence, and 800 times of iterations can't improve the result. Ultimately, we choose Value (1) for the parameters.

We train the network under the keras framework for its wide and applicability for fast experimentation. Our experiment platform chooses NVIDIA GeForce GTX 1070 for its good performance in image processing and model train. About artificial image

annotations, LabelMe software is used to annotate tooth images for its batch convert support. In order to enough dataset training, in our work also selects several methods for data enhancement, e.g. mirroring, rotating, moving, and flip.

During training, the batch size is set 300, and the maximum number of epochs is set 1. To ensure the accuracy of the gradient descent, and diminish the effect of noise, we introduce adaptive moment estimation (Adam) [15] into our method instead of stochastic gradient descent (SGD), and the learning rate is adjusted to 0.01. The best model on the validation set was stored and used for evaluations. To determine the proximity between the actual output and the expected output, the cross-entropy loss function is used in our experiment as follows:

$$loss = - \sum_{i=1}^n \hat{y}_i \log y_i + (1 - \hat{y}_i) \log(1 - \hat{y}_i) \quad (6)$$

$$\frac{\partial loss}{\partial y} = - \sum_{i=1}^n \frac{\hat{y}_i}{y_i} - \frac{1 - \hat{y}_i}{1 - y_i} \quad (7)$$

Every model training takes about 60 s in our hardware platform, predicted results of the model are shown in Fig. 5.

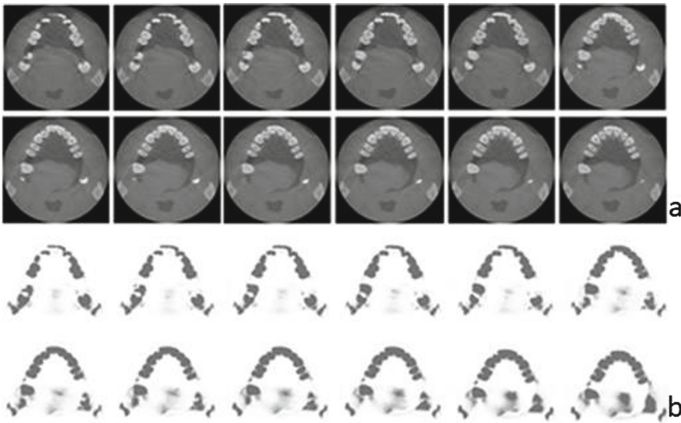


Fig. 5. The experimental results using our method. (a) the test original CT tooth image, (b) predicted results of our model.

To illustrate the precision of this method, we check the model results against real segmentation, and define the one which has more than 90% of the overlap to be successfully segmented, the result is shown in Fig. 6. On the left of the black line is the result of successful segmentation. On the right side of the black line is an example of failure. In the right image, there are redundant parts in the segmentation result, e.g. excess teeth, noise shadow.

All experiences are performed on 2D slices. To show the superiority of our integrated U-Net and level-set method, we also compare the traditional method (watershed method [13], chan-vese model [16], and graph cut method [17]) with our method, it can be shown

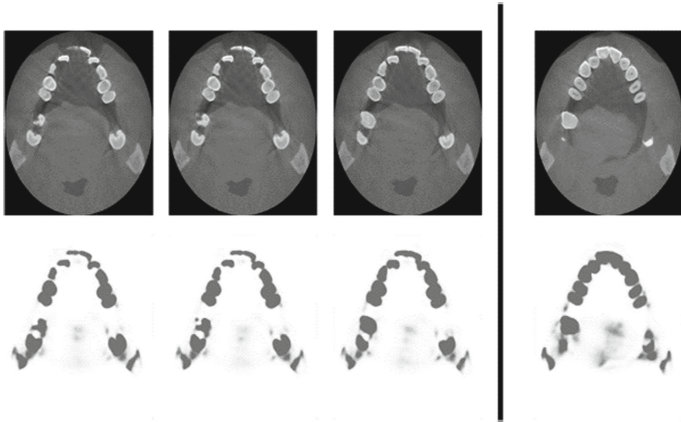


Fig. 6. Analysis of experimental results. Left: the results of successfully segmented tooth image, Right: the results of unsuccessfully segmented tooth image

in Fig. 7. From Fig. 7, we can find the graph cut method is least affected by noise and the watershed method is the most affected. Too many defects exist in the segmentation result of chan-ve-se model. Our method has clear segmentation result, no defect and little noise effect.

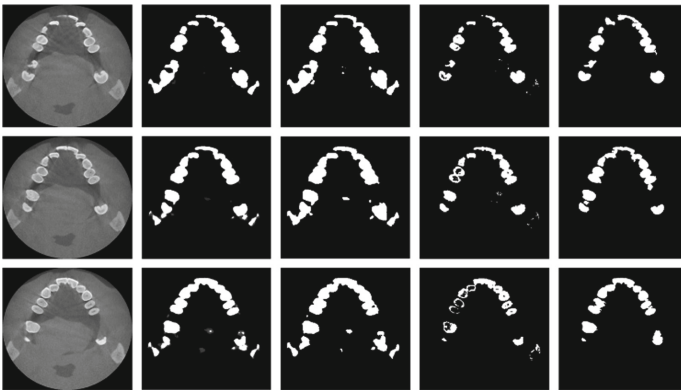


Fig. 7. Segmentation results with our method. (a) the original image, (b) our method, (c) watershed method [13], (d) chan-ve-se model [16], (e) graph cut method [17].

In order to performance automatic annotation and deep network well for tooth image segmentation, we also compare the traditional method and our method from Accuracy, artificial participation of image annotation, time-consuming. Table 2 illustrates that the graph cut method has the highest accuracy. Graph cut only processes one image at a time, and each processing requires a lot of manual delimitation of the segmentation area, it is difficult to automate, accurate segmentation of each graph requires artificial drawing of 3–4 lines, or even more. The accuracy of chan-ve-se model is close to our method,

but it takes nearly 118 s to calculate an image, our method only takes about 10 s. The numerical calculation of chan-vese model is too large and it needs many iterations, so its time-consuming problem is very serious.

Table 2. Result analysis of the traditional methods and our method (MP = manual participation, TC = time-consuming)

Comparisons method	Accuracy	MP	TC (unit: second)
Watershed [13]	50.2%	Yes	23
Chan-Vese [16]	65.3%	No	118
Graph cut [17]	70.1%	Yes	27
Our method	66.7%	No	10

4 Conclusion

In this paper, a novel technique for integrating level set with a new U-Net is presented. This technique has two significant advantages which are shown in our work. First, the proposed method has better performance than the U-Net or level set alone. Second, introducing the level set method into this network enables unlabeled data to be used in our experiment and reaches semi-supervised learning. Although mask making and model training may take some time, our method can segment the tooth in very little time after the training of the model. In future work, as the medical environment always need digital 3D modeling, we will work on the issues dealing with 3D binary segmentation.

Acknowledgments. This work was supported in part by the Science and Technology Service Industry project of Sichuan under 2019GFW126, Key R&D project of Sichuan under 2019ZDYF2790.

References

1. Lin, D., Dai, J.F., Jia, J.Y.: ScribbleSup: scribble-supervised convolutional networks for semantic segmentation. In: the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, pp. 3159–3167 (2016)
2. Pathak, D., Krahenbuhl, P., Darrell, T.: Constrained convolutional neural networks for weakly supervised segmentation. In: The IEEE International Conference on Computer Vision (ICCV), Centro Parque Convention Center in Santiago, Chile, pp. 1796–1804 (2015)
3. Papandreou, G., Chen, L.C., Murphy, K.P., Yuille, A.L.: Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In: The IEEE International Conference on Computer Vision (ICCV), Centro Parque Convention Center in Santiago, Chile, pp. 1742–1750 (2015)

4. Haslhofer, R.: Singularities of mean convex level set flow in general ambient manifolds. *Adv. Math.* **329**, 1137–1155 (2018)
5. Gibouab, F., Fedkiwc, R., Osher, S.: A review of level-set methods and some recent applications. *J. Comput. Phys.* **353**, 82–109 (2018)
6. Li, C.M., Xu, C.Y.: Distance regularized level set evolution and its application to image segmentation. *IEEE Trans. Image Process.* **19**(12), 3243–3254 (2010)
7. Allaire, G., Jouve, F., Michailidis, G.: Thickness control in structural optimization via a level set method. *Struct. Multidisc. Optim.* **53**, 1349–1382 (2016)
8. Yang, X., Gao, X.B., Tao, D.C., Li, X.L., Li, J.: An efficient MRF embedded level set method for image segmentation. *IEEE Trans. Image Process.* **24**, 9–21 (2014)
9. Morar, A., Moldoveanu, F., Gröller, E.: Image segmentation based on active contours without edges. In: *IEEE 8th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca*, pp. 213–220 (2012)
10. Gan, Y., Xia, Z., Xiong, J., Zhao, Q., Hu, Y., Zhang, J.: Toward accurate tooth segmentation from computed tomography images using a hybrid level set model. *Med. Phys.* **42**(1), 14–27 (2015)
11. Wang, L., Li, S., Chen, R., Liu, S.Y., Chen, J.C.: A segmentation and classification scheme for single tooth in micro CT images based on 3D level set and K-means++. *Comput. Med. Imaging Graph.* **57**, 19–28 (2017)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Lu, S., Wang, S., Zhang, Y.: A note on the marker-based watershed method for X-ray image segmentation. *Comput. Methods Program. Biomed.* **141**, 1–2 (2017)
14. Li, H.W., Andrii, Z.G., Bjoern, M.: Automatic brain structures segmentation using deep residual dilated U-net. *arXiv preprint arXiv:1811.04312* (2018)
15. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
16. Jung, M.Y., Chan, T.F., Vese, L.A.: Nonlocal Mumford-Shah regularizes for color image restoration. *IEEE Trans. Image Process.* **20**, 1583–1598 (2011)
17. Ju, W., Xiang, D., Zhang, B.: Random walk and graph cut for co-segmentation of lung tumor on PET-CT Images. *IEEE Trans. Image Process.* **24**, 5854–5867 (2015)