Original Article

Tracking-based deep learning method for temporomandibular joint segmentation

Yi Liu[1], Yao Lu[2], Yubo Fan[3,4], Longxia Mao[5]

[1]School of Biological Science and Medical Engineering, Beihang University, Beijing, China; [2]School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China; [3]Beijing Advanced Innovation Centre for Biomedical Engineering, Key Laboratory for Biomechanics and Mechanobiology of Chinese Education Ministry, School of Biological Science and Medical Engineering, Beihang University, Beijing, China; [4]School of Engineering Medicine, Beihang University, Beijing, China; [5]School of Mathematics, Sun Yat-sen University, Guangzhou, China

Contributions: (I) Conception and design: Y Liu; (II) Administrative support: L Mao; (III) Provision of study materials or patients: Y Lu; (IV) Collection and assembly of data: Y Liu; (V) Data analysis and interpretation: Y Liu, Y Fan; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Longxia Mao. School of Mathematics, Sun Yat-sen University, Guangzhou 510275, China. Email: 929133682@qq.com.

Background: The shape, size, and surface information relating to the glenoid fossae and condyles in temporomandibular joints (TMJ) are essential for diagnosing and treating. Patients with TMJ disease often have surface abrasion which may cause fuzzy edges in computed tomography (CT) imaging, especially for low-dose CT, making TMJ segmentation more difficult.

Methods: In this paper, an automatic segmentation algorithm based on deep learning and post-processing was introduced. First, U-Net was used to divide images into 3 categories: glenoid fossae, condyles, and background. For structural fractures in these divided images, the internal force constraint of a snake model was used to replenish the integrity of the fracture boundary in a post-processing operation, and the initial boundary of the snake was obtained based on the basis of the tracking concept. A total of 206 cases of low-dose CT were used to verify the effectiveness of the algorithm, and such indicators as the Dice coefficient (DC) and mean surface distance (MSD) were used to evaluate the agreement between experimental results and the gold standard.

Results: The proposed method is tested on a self-collected dataset. The results demonstrate that proposed method achieves state-of-the-art performance in terms of DCs = 0.92±0.03 (condyles) and 0.90±0.04 (glenoid fossae), and MSDs =0.20±0.19 mm (condyles) and 0.19±0.08 mm (glenoid fossae).

Conclusions: This study is the first to focus on the simultaneous segmentation of TMJ glenoid fossae and condyles. The proposed U-Net + tracking-based algorithm showed a relatively high segmentation efficiency, enabling it to achieve sought-after segmentation accuracy.

Keywords: Biomedical imaging; computer-aided diagnosis; deep learning; image segmentation; low-dose computed tomography (low-dose CT); tracking.

Submitted Dec 17, 2020. Accepted for publication Mar 03, 2021.
doi: 10.21037/atm-21-319
View this article at: http://dx.doi.org/10.21037/atm-21-319

Introduction

Temporomandibular joints (TMJ) are bilateral linkage joints located on either side of the human skull near the lower part of the ears. They consist of the glenoid fossae of the temporal bone, condyles of the mandible, articular disc located between these two sites, surrounding joint capsule, and joint ligaments. Among them, the mandibular glenoid fossae are transversely ovoid, and the condyles are transversely oval. Hereinafter, the temporal glenoid fossae are referred to as the "glenoid fossae" and the mandibular condyles are referred to as the "condyles", both of which are prone to TMJ diseases and were the segmentation targets of our study.

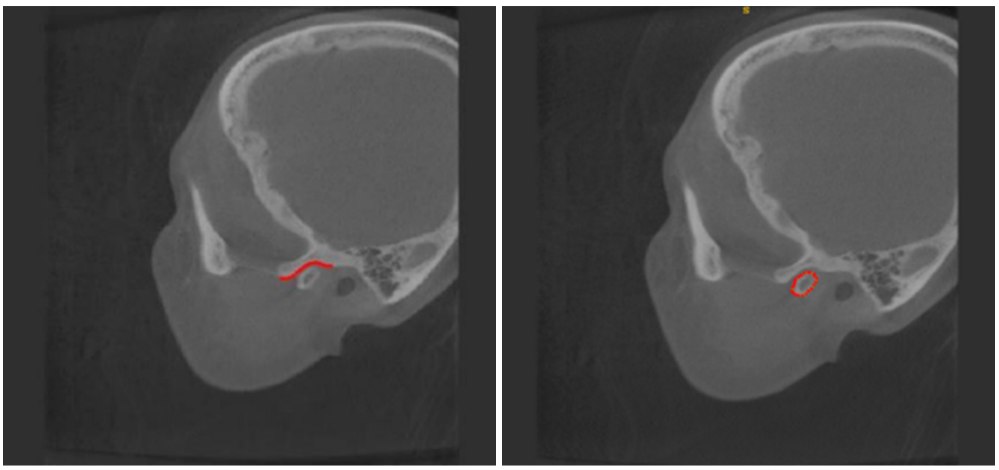

Figure 1 Segmentation targets. Left: lower glenoid fossa boundary; right: outer condyle boundary.

The health of the TMJ has a great influence on people's daily lives, with TMJ diseases affecting functions including mastication, swallowing, speech, and facial expression. In clinical practice, the TMJ medical imaging is usually required to observe the size and shape of its structure in order to analyze and diagnose disease (1). The etiology and pathogenesis of some diseases are assessed by analyzing the surface conditions of the structure and the agreement between the condyles and their corresponding glenoid fossae. To obtain this information from images accurately and efficiently, research into computer-assisted TMJ segmentation is urgently required.

This study aimed to achieve automatic segmentation of the TMJ glenoid fossae and condyles, to assist physicians in the diagnosis and treatment of TMJ pathologies. In order to reduce the damage to patient health caused by radiation, this study focused on using low-dose computed tomography (CT) imaging. However, this technique is limited by its low contrast and the disadvantage of noise interference. In response to clinical needs, the objective of this study was to achieve rapid and accurate segmentation of the boundaries of the glenoid fossae lower surfaces and the condyle external surfaces. Representative lower glenoid fossa and outer condyle boundaries were marked in red in *Figure 1*.

We present the following article in accordance with the MDAR reporting checklist (available at http://dx.doi.org/10.21037/atm-21-319).

Related work

According to the literature, the mandible has been the segmentation target of previous studies. The horseshoe-shaped mandible is the only part of the skeleton that constitutes the lower part of the human face, while the condyles segmented in this study are two protrusions on either side of the mandible against the posterior part. Existing mandibular segmentation methods are divided into three main types: the first type is manual segmentation (2), in which the user manually delineates each image containing the mandible, using delineation software. This is generally the method chosen by users with clinical experience when segmenting a small number of images. Although this is a simple method, it is time-consuming, and the segmentation accuracy is highly dependent on the user's clinical experience.

The second type is semi-automatic segmentation, which can be classified into four kinds: (I) threshold-based segmentation (3), in which the user modifies the threshold according to the grayscale level of the mandible in the CT image; (II) region-growing based segmentation (4,5), in which the seed points are given manually, or in combination with other pre-processing methods; (III) registration-based segmentation (6-8), in which some templates are accurately segmented manually, and then the segmented images are registered to the templates to produce segmentation results. This method has the limitations that manual post-processing is required after the segmentation, and that the operation is computationally expensive; and (IV) the segmentation method based on the GrowCut algorithm in open-source software (9), in which, after the prior information of the segmented target is given manually, the algorithm automatically segments the mandible.

The third type is automatic segmentation, of which there are three kinds: (I) segmentation based on statistical

shape models (9-12) and which requires a certain number of samples for training the shape models; (II) machine learning-based segmentation (13), in which features of the image such as grayscale and texture are extracted, and then a binary or multiple classification models are established to divide the pixels in the image into different categories such as target and background; and (III) segmentation based on a deep convolution network, which is suitable for cases in which there are larger datasets and gold standards (14).

None of the studies referred to above have investigated the simultaneous segmentation of the TMJ glenoid fossae and condyles. In our study, we addressed a different mandible segmentation problem and pursued different segmentation targets. The condyles are only small protrusions on the mandible, and mandible segmentation algorithms that focus on overall segmentation accuracy have not ensured that these small protrusions can be accurately segmented. In addition, there are fewer methods for segmenting the mandible based on low-dose CT, and the image quality of low-dose CT is worse than that of ordinary CT images, resulting in such problems as the easy adhesion of the glenoid fossae and condyles and easy fracture at the weak boundary of the structure during image segmentation. In summary, there is currently no accurate and efficient solution to the problem of TMJ image segmentation using low-dose CT imagery.

Our contribution

This study reported was the first to investigate the simultaneous segmentation of TMJ glenoid fossae and condyles. To resolve the problems mentioned above, we developed an automatic segmentation algorithm that uses the U-Net network for initial segmentation. Then, based on the generalized gradient vector flow (GGVF) snake model and the tracking idea, the algorithm post-processes the boundaries for the whole CT image. We have shown in our study that the new algorithm embodies high levels of accuracy and segmentation efficiency.

The organizational structure of this paper is as follows: in Section 2, the segmentation algorithm is introduced, while in Section 3, the experimentation and results achieved by algorithm application are described and reviewed. In the last section, the algorithm is summarized and discussed.

Methods

After image pre-processing, the proposed algorithm first used the U-Net network to segment the glenoid fossae and condyles in all 2-dimensional (2D) images. After image pre-processing, snake model was then used to post-process small fractures on the weak boundary of the target in the U-Net segmentation results. Increasing the internal force constraint of the snake model can prevent the boundary curve from invaginating at small faults and can help obtain an accurate target boundary. For continuous image sequences, the initial boundary required by the snake model in the first image was delineated from the U-Net, and then the initial boundaries of subsequent images were obtained by using the tracking idea to achieve automatic post-processing of sequential images.

Image pre-processing

The CT images were sourced from the patient image database of the Hospital of Stomatology at Sun Yat-sen University and consisted of 206 low-dose CT images of patients' heads. These images were stored in Digital Imaging and Communications in Medicine (DICOM) format in a variety of resolutions, including 512×512 pixels and 400×400 pixels, with voxel spaces mainly 0.3×0.3×0.3 and 0.4×0.4×0.4 mm^3. Prior to the experiment, all images were recalibrated into voxel spaces sized 0.4×0.4×0.4 mm^3. The gold standard was manually delineated by experienced physicians from the affiliated hospital at Sun Yat-sen University. All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the ethics board of the First Affiliated Hospital of Sun Yat-sen University {NO. [2019]366} and informed consent was taken from all the patients.

For a low-dose CT image sequence of the participant's head, the TMJ location was determined by the U-Net segmentation; the local 3-dimensional (3D) region containing the complete joint was then removed from the image and recorded as the region of interest (ROI). Subsequent segmentation operations were performed on the ROI to reduce interference from irrelevant information and improve segmentation efficiency.

Because the CT value ranges of the original images were relatively large and varied across different images, the grayscale value of each image had to be normalized before segmentation. This normalization operation involved linearly transforming the grayscale values of all pixels in the image to within the range of (0 to 255) according to

Eq. [1], $\tilde{g}$ represents the transformed grayscale value, and g_{max} and g_{min} represent the maximum and minimum values of the whole image before the transformation, respectively.

$$\tilde{g} = \frac{g - g_{\min}}{g_{\max} - g_{\min}} \times 255 \quad [1]$$

To prevent overfitting, we augmented the training data, applying random horizontal flipping, random rotation through an angle between ±30° and Gaussian filter blurring using randomly selected sigma values of 0.5–4.0.

U-Net network for initial segmentation

There are two types of U-Net network (15), 2D and 3D. In our study, due to particular segmentation target and data size limitations, a 2D U-Net network was selected, while a 3-class U-Net network was trained for the purpose of distinguishing between the glenoid fossa and condyle structures in the segmentation target.

The network structure is shown in *Figure 2*. The probability that each pixel belonged to a particular class was obtained for the network through symmetric contraction and expansion paths. The class into which each pixel in an image was finally allocated through the softmax function was obtained for the final layer of the network. The loss function used in the network was Dice loss, which can prevent the influence of data imbalance (16). The 3-class U-Net network divided all pixels from the ROI image extracted in the previous step into 3 classes: background, condyles, and glenoid fossae, as shown in *Figure 3*. As we were focused on whether the outer boundaries of the condyles and the lower boundaries of the glenoid fossae were accurately segmented, prior knowledge suggested that both the outer surfaces of the condyles and the lower surfaces of the glenoid fossae should be intact, without fractures. In the left diagram of *Figure 4*, the red curve marks the condyle boundary, while the green curve marks the boundary range of the glenoid fossa addressed here.

Comparing the U-Net segmentation results with the segmentation target, we could see that the boundaries of both the condyles and glenoid fossae had minor deletions. Therefore, to obtain complete and intact boundaries, the missing elements of the U-Net segmentation results needed to be re-established.

Boundary post-processing based on tracking idea

As shown in *Figure 4*, in order to replenish the fractured boundaries of the condyles and glenoid fossae in the U-Net segmentation results, the snake model (17) was used. In order to quickly complete segmentation of a sequence of images, the initial boundary of the first slice was delineated from previously U-Net segmentation, and the boundaries of subsequent slices were then automatically generated according to the boundary of the previous image, using the continuity of the target boundaries on continuous images in 3D space. The initial boundary of the current image used the boundary from the previous image, and so on, constituting application of the tracking idea to medical image segmentation.

In post-processing, the closed curve snake model was used for the condyles, while the non-closed curve snake model was used for the glenoid fossae. Because the upper boundary of the glenoid fossae was cluttered and not the focus of the segmentation problem addressed in this paper, the lower boundary of the glenoid fossae was obtained using the non-closed curve snake model to avoid accumulating errors in the tracking framework.

The snake model is represented by the parameterized curve, V (s) = [x (s), y (s)], s ∈ [0, 1], which is a set of control points connected by a straight line from the beginning to the end, and where x(s) and y(s) represent the coordinate position of each control point in the image. Variable s is an independent variable describing the boundary in Fourier transform form to minimize the following energy function:

$$E=\int_0^1 \frac{1}{2}\left[\alpha\left|x'(s)\right|^2+\beta\left|x''(s)\right|^2\right]+E_{ext}\left(x(s)\right)ds \quad [2]$$

The first term in energy function Eq. [2] represents the internal force term of the curve, and the second term is its external force term, as shown in Eq. [3]:

$$E_{int}=\frac{1}{2}\left[\alpha\left|x'(s)\right|^2+\beta\left|x''(s)\right|^2\right] \quad [3]$$

Internal force terms were used to control the magnitude of the elasticity and bending degree of the curve itself, as shown in *Figure 5*. Appropriately increasing the weight coefficient of the internal force terms, α and β, allowed us to enhance the continuity and smoothness of the curve contour in the model. When the curve was smooth enough, the smaller fracture boundary on the edge of the structure could be replenished.

Since the boundaries of the condyles and glenoid fossae had different shapes, the post-processing for each was performed separately. A condyle outer boundary is relatively continuous and smooth, and a closed initial contour was used in the snake model to obtain the real boundary.

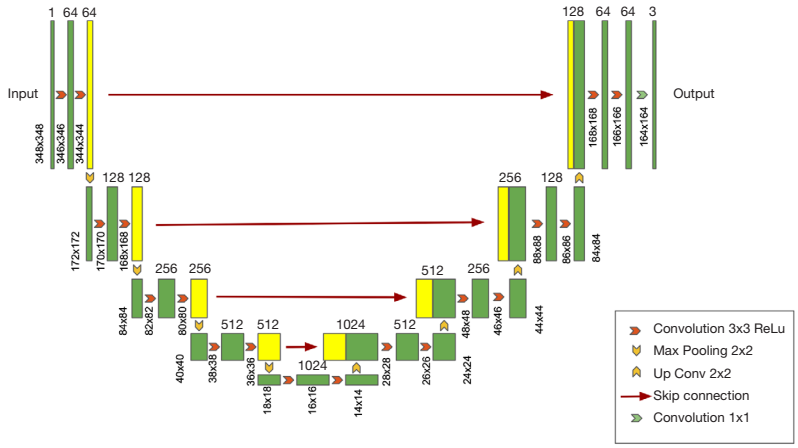

Figure 2 Three-class U-Net network structure.

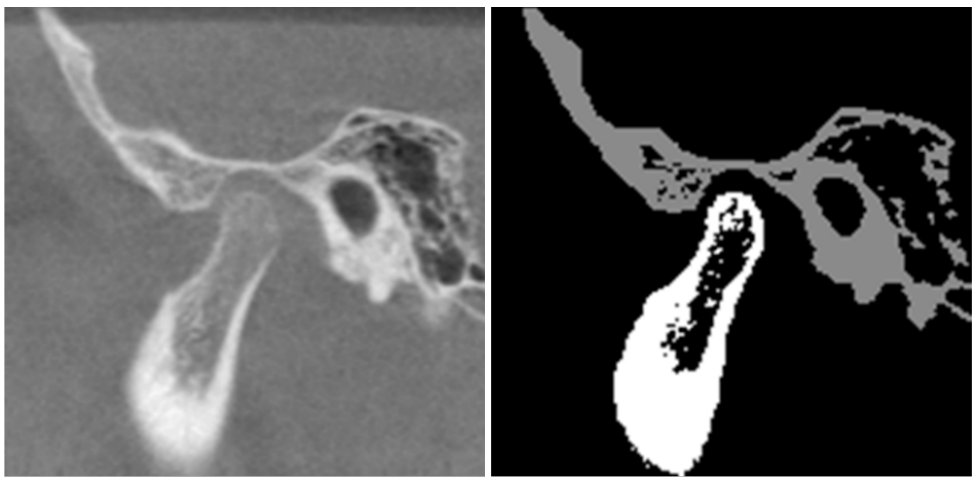

Figure 3 (Left) original, and (right) gold standard. In the gold standard, the grey target is the glenoid fossa, white target is the condyle, and black is background.

The upper boundary of the glenoid fossa is cluttered and discontinuous, and because only a part of the glenoid fossa boundary adjacent to the condylar boundary was used in our study, a non-closed initial contour was used to obtain the lower real boundary of the glenoid fossa. In order to obtain the non-closed boundary in the snake model, the value of weight coefficient β of the internal force term in Eq. [2], at the first and last two points of the contour, was set to 0, and the weight coefficient β of the internal force term for the rest of the points was set to an identical, non-0 value (18).

The GGVF (19,20) was used in our study as the external force field definition of the model. There were only two kinds of values in the U-Net segmentation results—figure, target, and background—and as there was no noise interference relative to the original image, the U-Net initial segmentation results figure was used to calculate the external force field rather than the original image.

In our work, the initial boundaries of the glenoid fossa and condyle were delineated by the U-Net for the first joint image. The snake model then evolved and converged to obtain the final boundary contour of the image. Next, the post-processing of adjacent images was performed sequentially, using the final boundary of the previous image as the initial contour of the next image. The left diagram in *Figure 6* shows a binary image of the condyles and gradient vector field calculated from the binary image, wherein the red curve is the boundary obtained from the previous image after post-processing. It can be seen in this figure that the

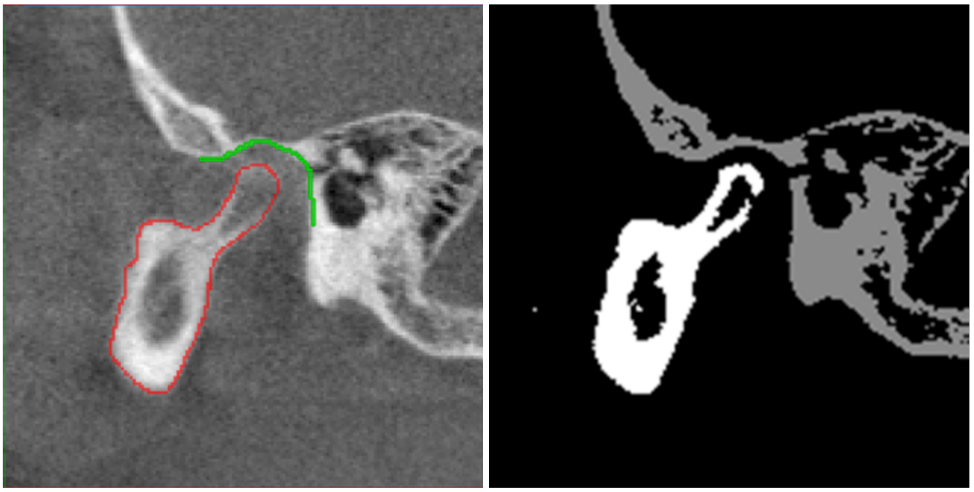

Figure 4 Left: ideal segmentation results; right: U-Net segmentation results. The left panel illustrates the ideal segmentation boundary marked on the original image; the red curve marks the outer boundary of the condyle and the green curve marks the lower boundary of the glenoid fossa.

Figure 5 Snake curve convergence results for different parameters, with the weight coefficient of internal force terms gradually increasing from left to right.

boundaries of the two adjacent images are similar, and that the red curve in the figure falls within the capture range of the current image external force field.

Since glenoid fossa post-processing led to a non-closed boundary contour, its boundary required further optimization to obtain the final results. The optimization operation procedures were as follows:

(I) Calculate the connected domain of the glenoid fossa segmented by U-Net, search the boundary of each connected domain, and record the set of boundary points as V_1, as shown in the right diagram of *Figure 7*, with different colors indicating the boundaries of different connected domains with different colors indicating.

(II) Find the points nearest to the start and end points of the initial boundary (blue curve) on the red curve in *Figure 8*, which constitutes the start point s and end point t for the optimized boundary.

(III) For each point p between s and t on the red curve, find the nearest outer boundary point $p' \in V_1$. If the distance between p' and p is less than threshold d, add p' to the set of optimized boundary points V,

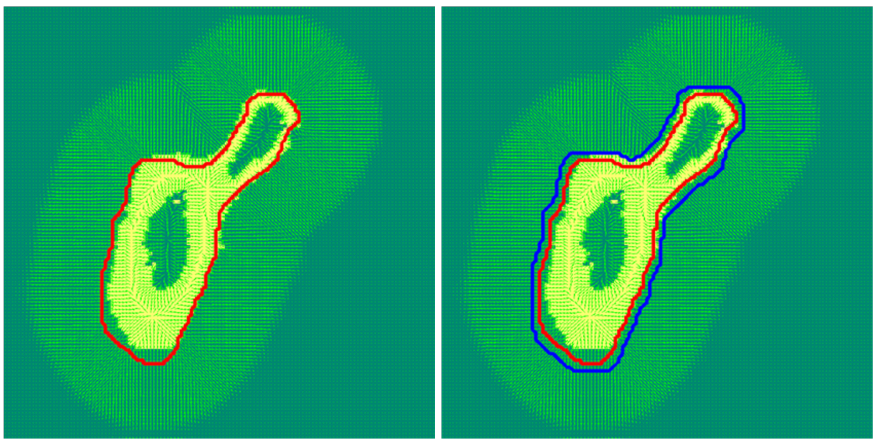

Figure 6 Initial boundary, in which both images are the superposition of the binary images of condyle and external force field. Red curve: the boundary obtained from post-processing the previous image; blue curve: obtained from expansion of the red curve.

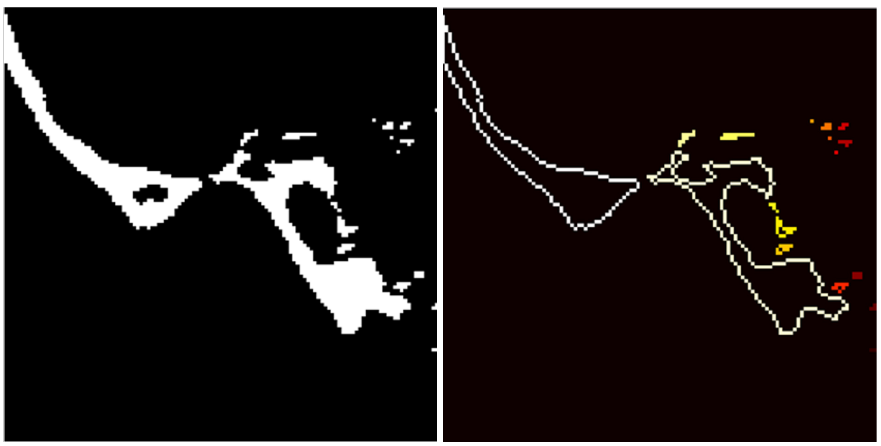

Figure 7 Left: binary Image of glenoid fossa to be post-processed; right: outer boundary of each connected domain.

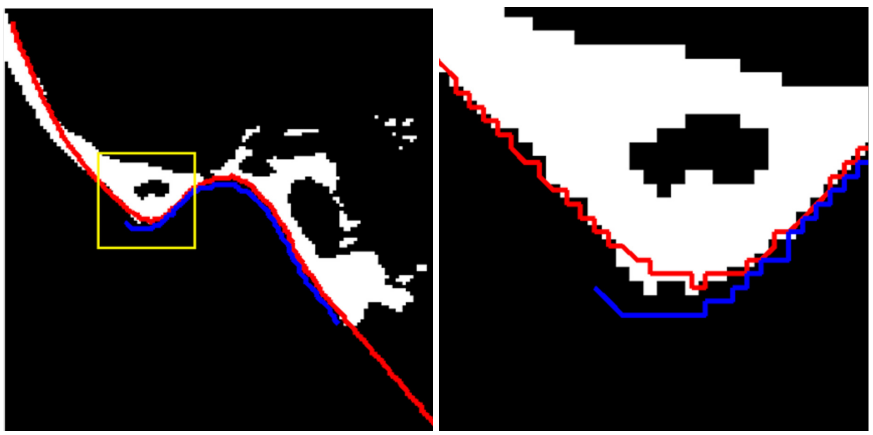

Figure 8 Lower glenoid fossa boundary, obtained by GGVF snake convergence. Blue curve: initial boundary; red curve: snake convergence results; right diagram: enlargement of the image in the yellow box in the left diagram. GGVF, generalized gradient vector flow.

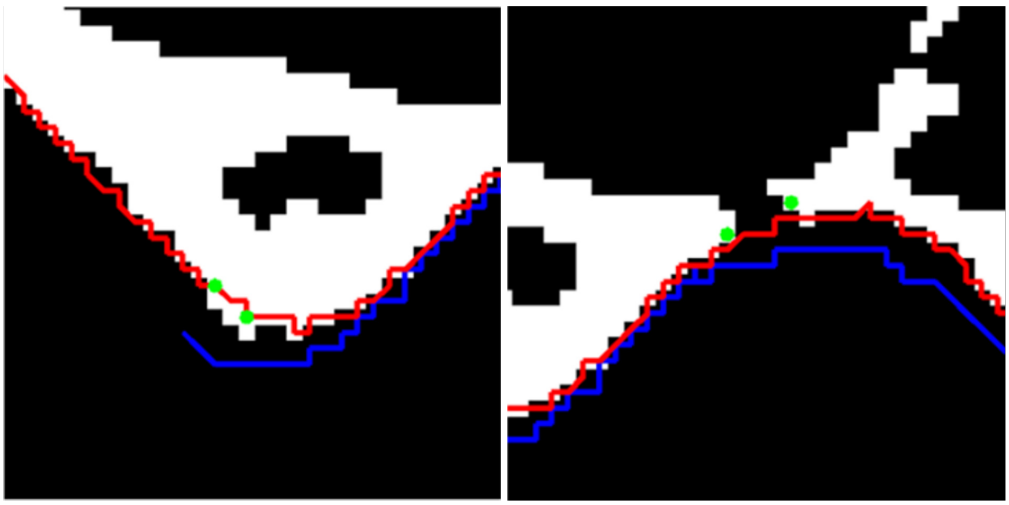

Figure 9 Two cases in which outer boundary points could not be searched.

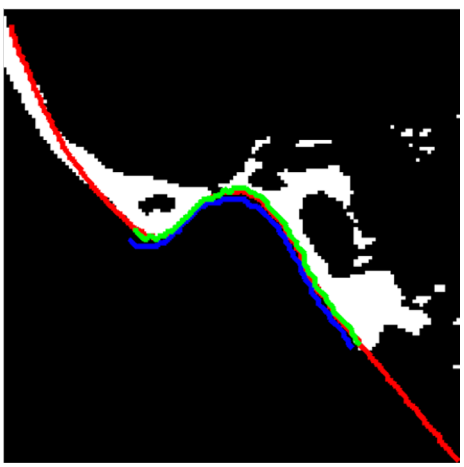

Figure 10 Post-processing results for the glenoid fossa lower boundary. Blue curve: initial boundary; red curve: GGVF snake convergence boundary; green curve: optimized boundary. GGVF, generalized gradient vector flow.

and record the order of searched outer boundary points. Otherwise, there are 2 situations. If the previous and next searched boundary points are in the same connected domain, p' is the wrong boundary point for snake convergence, as shown in the left diagram of *Figure 9*, when the corresponding boundary points on V_1 are searched as the optimized boundary points for the segment; otherwise, p are located on the fractured boundary, as shown in the right diagram of *Figure 9*, when the points on the red curve are directly replenished.

Finally, the optimized boundary is obtained as the green curve in *Figure 10*.

Results

U-Net segmentation results

After pre-processing, all images were processed into grayscale images with a resolution of 164×164 pixels, pixel size of 0.4×0.4 mm, and pixel grayscale pixel values of 0–255. The 206 images were randomly divided into training, validation, and test sets, in the proportion 6:2:2, resulting in 123 3D images for the training set, 42 for the validation set, and 41 for the test set. The images were then separately split into 2D slices, resulting in 10,313 2D images for the training set, 3,502 images for the validation set and 3,351 images for the test set. The U-Net network was trained using the Adam optimization algorithm, with the learning rate initialized at 0.01, and other parameters set to recommended values. The discard rate from the dropout layer was 0.5, and 15 images were input for each small batch. The Dice loss values tended to become stable after 40 iterations.

In our study, a model obtained by training for 43 iterations was selected to segment all images in the test set, and the average Dice loss value obtained was 0.1056. The condyles and glenoid fossae in all 42 images of the test set were subjected to 3D reconstruction, and the Dice coefficients (DCs) of the condyles and glenoid fossae in the gold standard were counted, giving an average Dice index of 0.9179±0.0331 for the condyles, and 0.8985±0.0386 for the glenoid fossae.

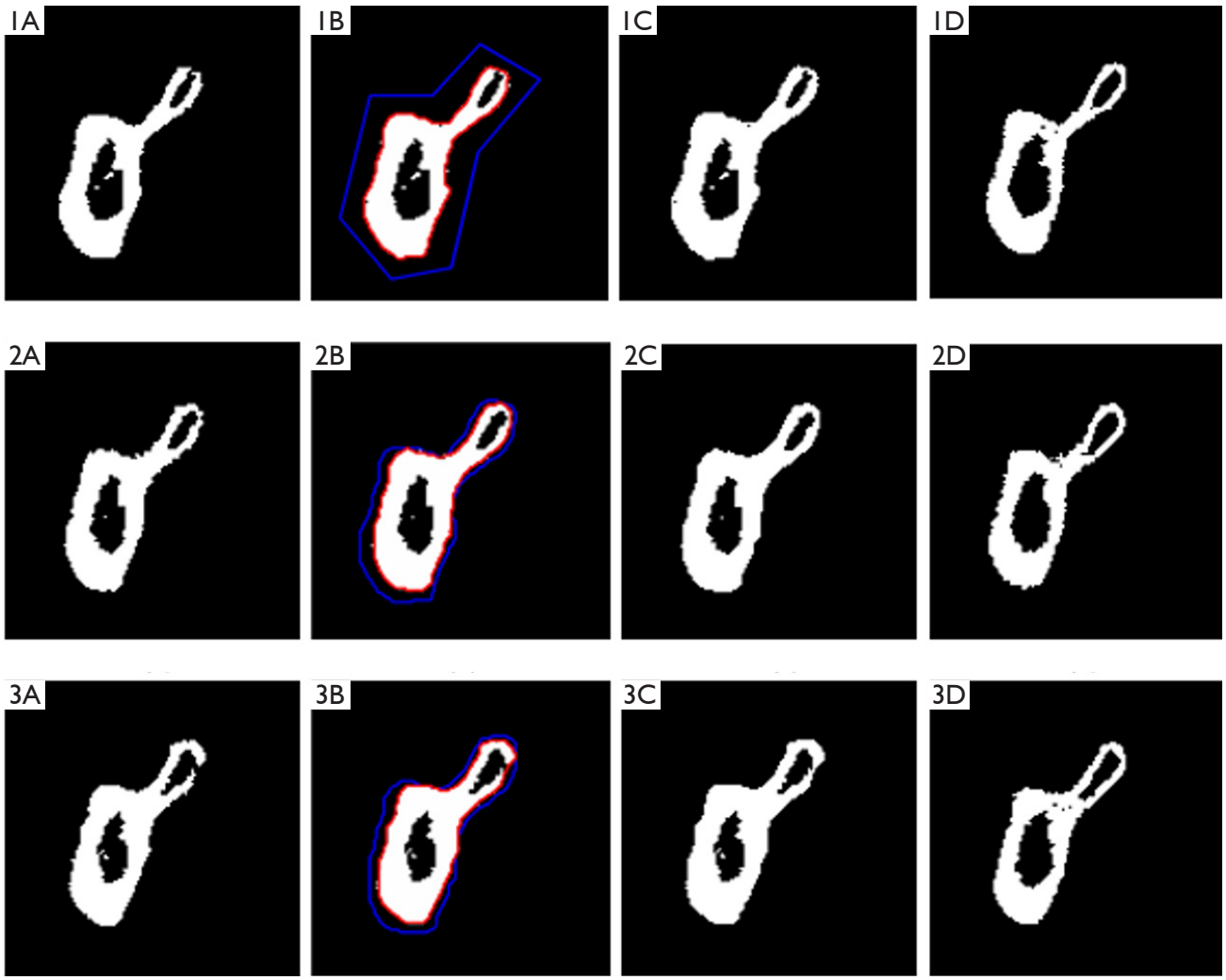

Figure 11 Condyle post-processing. The figure shows 3 consecutive images of a condylar structure. Each row shows the post-processing procedure for 1 image. (A) U-Net segmentation results; the blue curve in (B) is the initial boundary of the snake, wherein (1B) shows the delineated initial boundary, and the red curve is the convergence boundary; (C) shows the final results after replenishment of the boundary; (D) shows the ground truth.

Post-processing results

Figures 11 and *12* show the condyles and glenoid fossae image post-processing procedure using 3 images, with the initial boundary of the first image being delineated by the U-Net segmentation, as shown in *Figure 13*. Finally, the post-processed glenoid fossa and condyle structure images were subjected to 3D reconstruction and then smoothed with Gaussian low-pass filtering to obtain smooth, 3D surfaces.

Post-processing operations were performed on the preliminary results of these 42 images after U-Net segmentation. The surface distance from the gold standard for the experimental results before (U-Net segmentation results) and after post-processing is shown in *Table 1*, together with the mean surface distance (MSD) and maximum surface distances (Hausdorff distance, HD) (21,22). When the 2 sets of results are compared, it can be seen that the post-processing operation effectively reduced surface error between the experimental results and the gold standard, thereby improving boundary segmentation accuracy.

Table 2 shows the statistical results for experimental error calculations, including the Dice index, MSD, and HD for

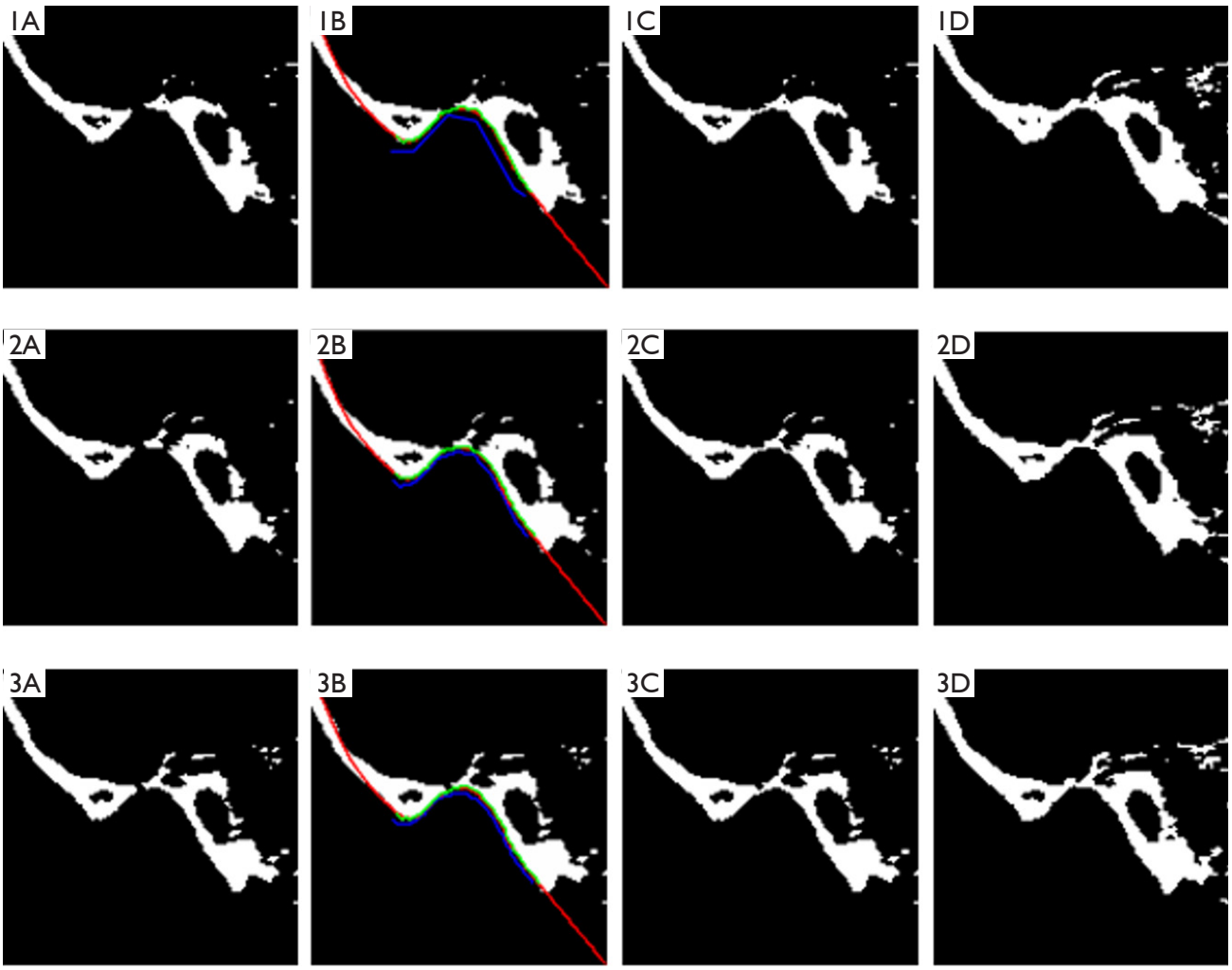

Figure 12 Glenoid fossa post-processing. The figure shows 3 consecutive images of a glenoid fossa structure, with each row showing the post-processing procedure for 1 image. (A) U-Net segmentation results; the blue curve in (B) is the initial boundary of the snake, wherein (1B) shows the delineated initial boundary; (C) shows the final results after replenishment of the boundary; (D) shows the ground truth.

condyles and glenoid fossae in relation to the gold standard in the segmentation results, as achieved for the 42 images.

Table 3 shows the U-Net and post-processing procedure computation times for a single CT slice.

Discussion

Glenoid fossa and condyle surface information is crucial in the context of TMJ disease diagnosis and treatment. A literature review indicated that most relevant studies targeted the complete mandible for segmentation. The condyles segmented in our study constituted just a small part of the mandible, which meant that existing mandible segmentation methods were not entirely suitable for condyle image segmentation.

Simultaneously, although low-dose CT reduces the harmful effects of radiation on the human body, it also introduces problems such as increased segmentation difficulty due to poor image quality, so that manual delineation is currently the main technique used for segmentation in clinical practice. To this end, our study involved the development of an automatic segmentation algorithm that combined U-Net with a tracking concept. In the new algorithm, a 3-class U-Net network, with Dice loss as the loss function, is used to segment the glenoid fossa, condyle, and background in an image, achieving highly

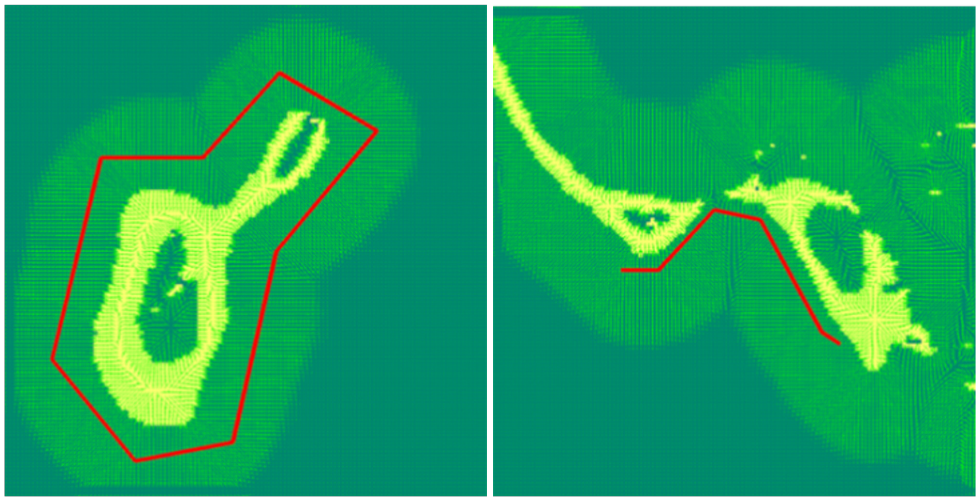

Figure 13 Delineation of (left) initial condyle boundaries, and (right) glenoid fossa.

Table 1 Comparison of surface error of experimental results before and after post-processing

Parameter	Before post-processing	After post-processing
Condyle (μ±σ)		
MSD (mm)	0.22±0.26	0.20±0.19
HD (mm)	5.77±5.30	3.90±4.35
Glenoid Fossa (μ±σ)		
MSD (mm)	0.21±0.08	0.19±0.08
HD (mm)	6.84±10.85	5.09±8.77

MSD, mean surface distance; HD, Hausdorff distance.

Table 2 Experimental statistical results

Parameter	Condyle (μ±σ)	Glenoid fossa (μ±σ)
FPR	0.003±0.003	0.006±0.007
FNR	0.083±0.071	0.127±0.092
Dice	0.92±0.03	0.90±0.04
MSD (mm)	0.20±0.19	0.19±0.08
HD (mm)	3.90±4.35	5.09±8.77

FPR, false positive rate; FNR, false negative rate; MSD, mean surface distance; HD, Hausdorff distance.

Table 3 Computation time for a single slice (UNIT = S)

Parameter	Condyle (μ±σ)	Glenoid fossa (μ±σ)
U-Net	0.023±0.003	0.023±0.003
Post-processing	0.618±0.051	0.563±0.047

accurate imaging of both glenoid fossae and condyles. In the algorithm, the GGVF snake model is used to extract the outer boundary of the structure, based on the U-Net segmentation results, and the internal force term constraint of the snake model is used to replenish the weak boundary of the fracture in the U-Net segmentation results, thus reducing the separation distance from the gold standard.

Although the algorithm can achieve relatively high accuracy and alleviate the low efficiency of manual delineation, it is only semi-automatic. Still, it requires some manual interaction—and so the next improvement goal is to avoid manual interaction and achieve fully automatic segmentation. At present, some simple ideas for achieving this include training the statistical shape models of glenoid fossae and condyles with a large number of existing gold standard datasets, and automatically obtaining the initial boundary of each image, in combination with an autoregressive model.

This algorithm also uses the 2D U-Net network, which is commonly used in medical image segmentation. For TMJ with 3D segmentation, 2D U-Net has less parameters and requires less graphics processing unit (GPU) memory, but it may ignore some inter-slice information. So, some 3D segmentation networks, such as 3D U-Net and V-Net (16), could be considered in future work. In addition, U-Net is relatively simple to use in feature extracting, making it more robust when applied to small datasets and easier to apply using general hardware. However, it may also limit network capacity at the same time. So, in our next phase, we could replace the core network used in this study with other, newer, and more complex networks, which may achieve more accurate segmentation.

Finally, since differences between the real image and that generated by the proposed model could be substantial in the medical field, the method should be considered as an initial rough segmentation procedure, with the doctor needing to simply modify and confirm the results, significantly decreasing the time required compared to fully manual contouring.

Acknowledgments

Funding: This work was supported by the National Key Research and Development Program of China (2020YFC2007104).

Footnote

Reporting Checklist: The authors have completed the MDAR reporting checklist. Available at http://dx.doi.org/10.21037/atm-21-319

Data Sharing Statement: Available at http://dx.doi.org/10.21037/atm-21-319

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-21-319). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All procedures performed in this study involving human participants were in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the ethics board of the First Affiliated Hospital of Sun Yat-sen University {NO. [2019]366} and informed consent was taken from all the patients.

References

1. Al-koshab M, Nambiar P, John J. Assessment of condyle and glenoid fossa morphology using CBCT in South-East Asians. PLoS One 2015;10:e0121682.
2. Enciso R, Memon A, Mah J. Three-dimensional visualization of the craniofacial patient: volume segmentation, data integration and animation. Orthod Craniofac Res 2003;6 Suppl 1:66-71; discussion 179-82.
3. Tognola G, Parazzini M, Pedretti G, et al, editors. Novel 3D reconstruction method for mandibular distraction planning. Proceedings of the 2006 IEEE International Workshop on Imagining Systems and Techniques (IST 2006); IEEE, 2006. doi: 10.1109/IST.2006.1650780.
4. Barandiaran I, Macía I, Berckmann E, et al. editors. An Automatic Segmentation and Reconstruction of Mandibular Structures from CT-Data. International

Conference on Intelligent Data Engineering and Automated Learning. Springer, Berlin, Heidelberg; 2009.
5. Xi T, Schreurs R, Heerink WJ, et al. A Novel Region-Growing Based Semi-Automatic Segmentation Protocol for Three-Dimensional Condylar Reconstruction Using Cone Beam Computed Tomography (CBCT). PLOS ONE 2014;9:e111126.
6. Wang L, Chen KC, Gao Y, et al. Automated bone segmentation from dental CBCT images using patch-based sparse representation and convex optimization. Medical Physics 2014;41:043503.
7. Fan Y, Beare R, Matthews H, et al. Marker-based watershed transform method for fully automatic mandibular segmentation from CBCT images. Dentomaxillofac Radiol 2019;48:20180261.
8. Chuang YJ, Doherty BM, Adluru N, et al. A Novel Registration-Based Semiautomatic Mandible Segmentation Pipeline Using Computed Tomography Images to Study Mandibular Development. J Comput Assist Tomogr 2018;42:306-16.
9. Wallner J, Hochegger K, Chen X, et al. Clinical evaluation of semi-automatic open-source algorithmic software segmentation of the mandibular bone: Practical feasibility and assessment of a new course of action. Plos One 2018;13:e0196378.
10. Gollmer ST, Buzug TM, editors. Fully automatic shape constrained mandible segmentation from cone-beam CT data. 2012 9th IEEE international symposium on biomedical imaging (ISBI); IEEE, 2012.
11. Kainmueller D, Lamecker H, Seim H, et al. editors. Automatic extraction of mandibular nerve and bone from cone-beam CT data. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Berlin, Heidelberg; 2009.
12. Lamecker H, Zachow S, Wittmers A, et al. Automatic segmentation of mandibles in low-dose CT-data. International Journal of Computer Assisted Radiology and Surgery 2006;1:393-5.
13. Wang L, Gao Y, Shi F, et al. Automated segmentation of dental CBCT image with prior-guided sequential random forests. Medical Physics 2016;43:336-46.
14. Ibragimov B, Xing L. Segmentation of organs at risks in head and neck CT images using convolutional neural networks. Medical Physics 2017;44:547-57.
15. Ronneberger O, Fischer P, Brox T, editors. U-Net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.
16. Milletari F, Navab N, Ahmadi SA, editors. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth International Conference on 3D Vision (3DV); IEEE, 2016.
17. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. Int J Comput Vis 1988;1:321-31.
18. Wang Y, Narayanaswamy A, Tsai CL, et al. A broadly applicable 3-D neuron tracing method based on open-curve snake. Neuroinformatics 2011;9:193-217.
19. Xu C, Prince JL. Generalized gradient vector flow external forces for active contours. Signal Processing 1998;71:131-9.
20. Xu C, Prince JL. Snakes, shapes, and gradient vector flow. IEEE Transactions on Image Processing 1998;7:359-69.
21. Aspert N, Santa-Cruz D, Ebrahimi T, editors. Mesh: Measuring errors between surfaces using the hausdorff distance. Proceedings. IEEE International Conference on Multimedia and Expo; IEEE, 2002.
22. Huttenlocher DP, Rucklidge WJ, Klanderman GA, editors. Comparing images using the Hausdorff distance under translation. Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; IEEE,1992.

(English Language Editor: J. Jones)

Cite this article as: Liu Y, Lu Y, Fan Y, Mao L. Tracking-based deep learning method for temporomandibular joint segmentation. Ann Transl Med 2021;9(6):467. doi: 10.21037/atm-21-319