View Reviews

Paper ID
2649

Paper Title
DArch: Dental Arch Prior-assisted 3D Tooth Instance Segmentation with Weak Annotations

Reviewer #1

Questions

1. Summary. In 5-7 sentences, describe the key ideas, experimental or theoretical results, and their significance.

This paper propose a 3D tooth instance segmentation on dental models with weak annotations (labelling all tooth centroids and only a few teeth). It takes a coarse-to-fine method to estimate dental arch, a dental arch-aware point sampling module to extract tooth centroid, and finally segment the tooth on the cropped patchs. The efficiency of annotation is an important topic in medical image segmentation. This work fuses the methods in point cloud, takes a weakly supervised method and reduces the annotation costs in tooth instance segmentation. However, the experiments are weak to evaluate their framework. And this is a highly specific work designed for teeth, and the proposed method lacks influence in the field.

2. Strengths. Consider the significance of key ideas, experimental or theoretical validation, writing quality, data contribution. Explain clearly why these aspects of the paper are valuable. Short bullet lists do NOT suffice.

1. The annotation-efficient work in tooth instance segmentation has great significance in model training.
2. This work embeds the regression of dental arch (Bezier curve regression) in the framework which is an effective domain knowledge. Although it was used in the existing clinical softwares, but it is interesting to integrate it into the deep learning pipeline.
3. The experimental partially shows the good effect of the proposed method.

3. Weaknesses. Consider the significance of key ideas, experimental or theoretical validation, writing quality, data contribution. Clearly explain why these are weak aspects of the paper, e.g. why a specific prior work has already demonstrated the key contributions, or why the experiments are insufficient to validate the claims, etc. Short bullet lists do NOT suffice.

1.The experiment is limited.
-The comparative experiment lacks extensiveness, and the author did not compare with some classic and latest instance segmentation methods, such as the works mentioned in the 'Related Work'. So the superiority of the proposed method is confusing.
-Ablation studies are lacking. Both the ablation studies for innovation and the analysis of hyperparameters are lacking. It is not clear how much the error in the first stage will affect the second stage in this two-stage model, and the impact of changes in some newly introduced hyperparameters on the overall performance of the model during the training process.
-Neither the data nor the code is public.

2. The author's misleading of some concepts. From Equ.4, this work not only needs tooth centroids, but also needs bounding box. So this is a typical object detection task which needs more annotation costs than the annotation of tooth centroid. This seems to be contrary to the annotation description of the data in the experimental part, which makes readers wonder how to calculate Equ.4 with only the tooth centroid and no bounding box.

4. Paper rating (pre-rebuttal).
Weak Reject

5. Justification of rating. What are the most important factors in your rating?

The avoidance of bounding boxes' description and the contradiction between Equ.4 and the description of the dataset's annotation made me question the rigor of this article. The private data sets and undisclosed code exacerbated my concerns. Although the idea of Bezier curve regression for dental arch made me interested.

6. Are there any serious ethical/privacy/transparency/fairness concerns? If yes, please also discuss below in Question 9.

No

7. Limitations and Societal Impact. Have the authors adequately addressed the limitations and potential negative societal impact of their work? Discuss any serious ethical/privacy/transparency/fairness concerns here. Also discuss if there are important limitations that are not apparent from the paper.

No.

8. Is the contribution of a new dataset a main claim for this paper? Have the authors indicated so in the submission form?

No dataset contribution claim

9. Additional comments to author(s). Include any comments that may be useful for revision but should not be considered in the paper decision.

See weakness.

Reviewer #2

Questions

1. Summary. In 5-7 sentences, describe the key ideas, experimental or theoretical results, and their significance.

The paper proposes a tooth instance segmentation framework which significantly reduces the annotation cost. The paper solves by detecting the centroid of each tooth and then cropping each tooth to perform segmentation. The authors finds smart and clever way to combine different state of the art methods to perform instance segmentation. The authors experiments with a fairly large dataset to experiment the proposed method. The authors have also performed ablation studies to explain the need of each components in the framework.

2. Strengths. Consider the significance of key ideas, experimental or theoretical validation, writing quality, data contribution. Explain clearly why these aspects of the paper are valuable. Short bullet lists do NOT suffice.

The paper is well written and easy to follow and the code can be reproduced.

3. Weaknesses. Consider the significance of key ideas, experimental or theoretical validation, writing quality, data contribution. Clearly explain why these are weak aspects of the paper, e.g. why a specific prior work has already demonstrated the key contributions, or why the experiments are insufficient to validate the claims, etc. Short bullet lists do NOT suffice.

1. Novelty: I feel the novelty is limited. The author has combined several state of the art methods to perform tooth instance segmentation. The authors have cited the reference papers in the related works. The contribution is the sampling mechanism specially designed for tooth instance segmentation.

2. Experimental setup: I could see that the dataset was split into train/test instead of train/val/test. I would like to know how the best model was saved ? Was it saved based on the test dataset or based on the training loss or anything different ?

3. Cross validation: In general 5-fold cross validation is a good way to overcome over-fitting. I could see that single fold cross validation was performed. It would be nice to 5-fold cross validation result for the same.

4. Standard deviations is not mentioned for any of the scores. I would not see any difference between 99.68 vs 99.41 (centroid detection) unless there is a significant difference in standard deviation.

5. Comparison with segmentation models: Since we have the GT instance segmentation mask it could also be fair to compare with instance segmentation models which does not use point cloud annotations. Such as [ToothNet: Automatic Tooth Instance Segmentation and Identification from Cone

Beam CT Images] cvpr 2019

6. Are tooth segmentation masks randomly removed from the original GT to perform weak annotation ? If so, does the model see all the different tooth while training ? What is the statistics (total) of different tooths does the model see in an epoch ? Will there be any imbalance of tooths shown to the model ? What will happen if a particular tooth GT is not at all shown to the model irrespective of its centroid shown.

4. Paper rating (pre-rebuttal).

Weak Reject

5. Justification of rating. What are the most important factors in your rating?

The resaon for Borderline rating is that the framework is not entirely novel but a careful combination of different components and the experimental setup is weak and needs naive segmentation models to be compared with. The dataset setup is also not clear and needs a further refinement.

6. Are there any serious ethical/privacy/transparency/fairness concerns? If yes, please also discuss below in Question 9.

No

7. Limitations and Societal Impact. Have the authors adequately addressed the limitations and potential negative societal impact of their work? Discuss any serious ethical/privacy/transparency/fairness concerns here. Also discuss if there are important limitations that are not apparent from the paper.

N/A

8. Is the contribution of a new dataset a main claim for this paper? Have the authors indicated so in the submission form?

No dataset contribution claim

Reviewer #3

Questions

1. Summary. In 5-7 sentences, describe the key ideas, experimental or theoretical results, and their significance.

They propose a novel detection-and-segmentation framework with weak annotation. The framework consists of tooth centroid detection and tooth instance segmentation. With limited annotation, they further focus on dental arch, design an APS module, and benefit the tooth detection.

2. Strengths. Consider the significance of key ideas, experimental or theoretical validation, writing quality, data contribution. Explain clearly why these aspects of the paper are valuable. Short bullet lists do NOT suffice.

Strength:

1.Expensive manual labeling is a significant problem but not well-studied for 3D dental models. The study of weak annotation on dental models can apply to real-world annotation and is practical.

2.The dental arch is often considered for dentists on patients' therapy. They propose a coarse-to-fine method to estimate dental arch and boost the task. Their novelty is worth noting.

The experiments are complete and demonstrate the effectiveness of their method. They conduct the experiments on 4,773 dental models, a relatively large data collection for dental models

3. Weaknesses. Consider the significance of key ideas, experimental or theoretical validation, writing quality, data contribution. Clearly explain why these are weak aspects of the paper, e.g. why a specific prior

work has already demonstrated the key contributions, or why the experiments are insufficient to validate the claims, etc. Short bullet lists do NOT suffice.

Weakness:

1.There are some typos and vague sentences in this paper.

2.The description of dental arch prediction method is not clear.

3.TSegNet[1] gives a good example of tooth detection and then segmentation. [1] shows its advantages on challenging dental models. I think that the author needs further elaborate on

a.the difference of APS and the distance-aware detection in [1];

b.the reason why they outperform [1] in weak annotation.

[1]. Cui Z, Li C, Chen N, et al. TSegNet: an efficient and accurate tooth segmentation network on 3D dental model[J]. Medical Image Analysis, 2021, 69: 101949.

4. Paper rating (pre-rebuttal).

Strong Accept

5. Justification of rating. What are the most important factors in your rating?

Justification: In general, the whole paper is easy to follow, and the idea is interesting and reasonable. I recommend acceptance of this paper.

6. Are there any serious ethical/privacy/transparency/fairness concerns? If yes, please also discuss below in Question 9.

No

7. Limitations and Societal Impact. Have the authors adequately addressed the limitations and potential negative societal impact of their work? Discuss any serious ethical/privacy/transparency/fairness concerns here. Also discuss if there are important limitations that are not apparent from the paper.

The application will have great impact in digital dentistry.

8. Is the contribution of a new dataset a main claim for this paper? Have the authors indicated so in the submission form?

No dataset contribution claim