3D Teeth Scan Segmentation and Labelling Challenge: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

Use the title to convey the essential information on the challenge mission.

3D Teeth Scan Segmentation and Labelling Challenge

Challenge acronym

Preferable, provide a short acronym of the challenge (if any).

3DTeethSeg22

Challenge abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Computer-aided design (CAD) tools have become increasingly popular in modern dentistry for highly accurate treatment planning. In particular, in orthodontic CAD systems, advanced intraoral scanners (IOSs) are now widely used as they provide precise digital surface models of the dentition. Such models can dramatically help dentists simulate teeth extraction, move, deletion, and rearrangement and ease therefore the prediction of treatment outcomes. Hence, digital teeth models have the potential to release dentists from otherwise tedious and time consuming tasks.
Although IOSs are becoming widespread in clinical dental practice, there are only few contributions on teeth segmentation/labelling available in the literature [1,2,3] and no publicly available database. A fundamental issue that appears with IOS data is the ability to reliably segment and identify teeth in scanned observations.
Teeth segmentation and labelling is difficult as a result of the inherent similarities between teeth shapes as well as their ambiguous positions on jaws.

In addition, it faces several challenges:
1- The teeth position and shape variation across subjects.
2- The presence of abnormalities in dentition. For example, teeth crowding which results in teeth misalignment and thus non-explicit boundaries between neighboring teeth. Moreover, lacking teeth and holes are commonly seen among people.
3- Damaged teeth.
4- The presence of braces, and other dental equipment.
The challenge we propose will particularly focus on point 1, i.e. the teeth position and shape variation across subjects. With the extension of available data in the mid and long term, the other points will also be addressed in further editions of the challenge.

[1] Lian, Chunfeng, et al. "MeshSNet: Deep multi-scale mesh feature learning for end-to-end tooth labeling on 3D

dental surfaces." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019.
[2] Xu, Xiaojie, Chang Liu, and Youyi Zheng. "3D tooth segmentation and labeling using deep convolutional neural networks." IEEE transactions on visualization and computer graphics 25.7 (2018): 2336-2348.
[3] Sun, Diya, et al. "Automatic Tooth Segmentation and Dense Correspondence of 3D Dental Model." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020.

Challenge keywords

List the primary keywords that characterize the challenge.

3D Teeth segmentation, 3D segmentation, 3D object detection, 3D intraoral scans, dentistry

Year

The challenge will take place in ...

2022

FURTHER INFORMATION FOR MICCAI ORGANIZERS

Workshop

If the challenge is part of a workshop, please indicate the workshop.

None for the moment

Duration

How long does the challenge take?

Half day.

Expected number of participants

Please explain the basis of your estimate (e.g. numbers from previous challenges) and/or provide a list of potential participants and indicate if they have already confirmed their willingness to contribute.

We expect 40 participants

Publication and future plans

Please indicate if you plan to coordinate a publication of the challenge results.

We intend to coordinate with all participants a journal article summarizing the main results and conclusions drawn from the challenge. All participants can be authors of the article.
Participating teams can anyway publish their own results independently after an embargo time of 6 months.

Space and hardware requirements

Organizers of on-site challenges must provide a fair computing environment for all participants. For instance, algorithms should run on the same computing platform provided to all.

the challenge will be organized on grand-challenge.org for publicity and own online submission site for final model submission.

TASK: 3DTeethSeg22 3D Teeth Scan Segmentation and Labelling Challenge

SUMMARY

Abstract

Provide a summary of the challenge purpose. This should include a general introduction in the topic from both a biomedical as well as from a technical point of view and clearly state the envisioned technical and/or biomedical impact of the challenge.

Computer-aided design (CAD) tools have become increasingly popular in modern dentistry for highly accurate treatment planning. In particular, in orthodontic CAD systems, advanced intraoral scanners (IOSs) are now widely used as they provide precise digital surface models of the dentition. Such models can dramatically help dentists simulate teeth extraction, move, deletion, and rearrangement and ease therefore the prediction of treatment outcomes. Hence, digital teeth models have the potential to release dentists from otherwise tedious and time consuming tasks.
Although IOSs are becoming widespread in clinical dental practice, there are only few contributions on teeth segmentation/labelling available in the literature [1,2,3] and no publicly available database. A fundamental issue that appears with IOS data is the ability to reliably segment and identify teeth in scanned observations.
Teeth segmentation and labelling is difficult as a result of the inherent similarities between teeth shapes as well as their ambiguous positions on jaws.

In addition, it faces several challenges:
1- The teeth position and shape variation across subjects.
2- The presence of abnormalities in dentition. For example, teeth crowding which results in teeth misalignment and thus non-explicit boundaries between neighboring teeth. Moreover, lacking teeth and holes are commonly seen among people.
3- Damaged teeth.
4- The presence of braces, and other dental equipment.
The challenge we propose will particularly focus on point 1, i.e. the teeth position and shape variation across subjects. With the extension of available data in the mid and long term, the other points will also be addressed in further editions of the challenge.

[1] Lian, Chunfeng, et al. "MeshSNet: Deep multi-scale mesh feature learning for end-to-end tooth labeling on 3D dental surfaces." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019.
[2] Xu, Xiaojie, Chang Liu, and Youyi Zheng. "3D tooth segmentation and labeling using deep convolutional neural networks." IEEE transactions on visualization and computer graphics 25.7 (2018): 2336-2348.
[3] Sun, Diya, et al. "Automatic Tooth Segmentation and Dense Correspondence of 3D Dental Model." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020.

Keywords

List the primary keywords that characterize the task.

3D Teeth segmentation, 3D segmentation, 3D object detection, 3D intraoral scans, dentistry

ORGANIZATION

Organizers

a) Provide information on the organizing team (names and affiliations).

Organizing team:
Achraf Ben-Hamadou, Digital Research Center of Sfax, Sfax University, Tunisia
Houda Chaabouni, Digital Research Center of Sfax, Sfax University, Tunisia
Sergi Pujades INRIA , Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, France
Edmond Boyer INRIA Grenoble Rhône Alpes, France
Ahmed Rekik, Digital Research Center of Sfax, Sfax University, Tunisia

Clinical Evaluators and Annotation Approvers:
Dr Julien Strippoli, Orthodontic Clinic, Lyon, France
Dr Hugo Setbon, Dental Clinic, Brussels, Belgium
Dr Aurélein Thollot, Dental Clinic, Lyon, France

Data Contributor:
Edouard Ladroit, Udini, Aix-en-Provence, France

b) Provide information on the primary contact person.

Achraf Ben-Hamadou, Digital Research Center of Sfax, Tunisia
achraf.benhamadou@crns.rnrt.tn

Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One time event with fixed submission deadline.

Challenge venue and platform

a) Report the event (e.g. conference) that is associated with the challenge (if any).

MICCAI.

b) Report the platform (e.g. grand-challenge.org) used to run the challenge.

grand-challenge.org for publicity and own online submission site for final model submission.

c) Provide the URL for the challenge website (if any).

None at this moment.

Participation policies

a) Define the allowed user interaction of the algorithms assessed (e.g. only (semi-) automatic methods allowed).

Fully automatic.

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

No policy defined.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but not eligible for awards.

d) Define the award policy. In particular, provide details with respect to challenge prizes.

Prospective sponsorship from NVIDIA

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

All performance results will be made public.

f) Define the publication policy. In particular, provide details on ...

- ... who of the participating teams/the participating teams' members qualifies as author
- ... whether the participating teams may publish their own results separately, and (if so)
- ... whether an embargo time is defined (so that challenge organizers can publish a challenge paper first).

We intend to coordinate with all participants a journal article summarizing the main results and conclusions drawn from the challenge. All participants can be authors of the article.
Participating teams can anyway publish their own results independently after an embargo time of 6 months.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

Submission instructions will be on the webpage. We will let the participants submit their Docker submissions to our evaluation server.

b) Provide information on the possibility for participating teams to evaluate their algorithms before submitting final results. For example, many challenges allow submission of multiple results, and only the last run is officially counted to compute challenge results.

Multiple submissions are allowed. However, only the last submission will be considered for the challenge results. Also, the number of Docker submissions will be limited to one per day.

Challenge schedule

Provide a timetable for the challenge. Preferably, this should include

- the release date(s) of the training cases (if any)
- the registration date/period
- the release date(s) of the test cases and validation cases (if any)
- the submission date(s)
- associated workshop days (if any)
- the release date(s) of the results

Registration is open from 1st March to 15th March 2022.

15th March 2022: Expected release of the first batch of training data.

15th April 2022: Expected release of the second batch of training data.

1st July 2022: Expected release of the public test set + opening of the submission portal for the public test set.

15h July 2022: Submission of testset closes. Docker submission portal opens.

30th July 2022. Docker submission deadline.
Announcement of results at Workshop/MICCAI.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

All data is anonymized.
Ethics approval by the local ethics committee of the Digital Research Center, Sfax University granted 28.10.2020, No. PV-CS-28/10/2020
Study is validated by “DELSOL Avocats” (Paris, France) law firm spécilized in GDPR regulation.

Data usage agreement

Clarify how the data can be used and distributed by the teams that participate in the challenge and by others during and after the challenge. This should include the explicit listing of the license applied.

Examples:

- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-ND (Attribution-NoDerivs)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivs)

CC BY NC ND.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

Evaluation code will be made public before the system is open for submission.

b) In an analogous manner, provide information on the accessibility of the participating teams' code.

Not required, however we encourage the participants to publish their code on github or the challenge platform.

Conflicts of interest

Provide information related to conflicts of interest. In particular provide information related to sponsoring/funding of the challenge. Also, state explicitly who had/will have access to the test case labels and when.

There are no conflicts of interest.
Udini is the principal sponsor of the challenge by collecting and providing clinical data.
Only the organisers and members of their immediate team have access to test case labels.

MISSION OF THE CHALLENGE

Field(s) of application

State the main field(s) of application that the participating algorithms target.

Examples:

- Diagnosis
- Education
- Intervention assistance
- Intervention follow-up
- Intervention planning
- Prognosis
- Research
- Screening
- Training
- Cross-phase

Assistance, Training, Diagnosis, Prevention, CAD, Research, Screening, Surgery, Education.

Task category(ies)

State the task category(ies).

Examples:

- Classification
- Detection
- Localization
- Modeling
- Prediction
- Reconstruction
- Registration
- Retrieval
- Segmentation
- Tracking

Localization: for each visible tooth, give its center as a 3D point.
Identification: for each visible tooth, give the label of its center.
Segmentation: for each point in the point-cloud scan, give its label.

Cohorts

We distinguish between the target cohort and the challenge cohort. For example, a challenge could be designed around the task of medical instrument tracking in robotic kidney surgery. While the challenge could be based on ex vivo data obtained from a laparoscopic training environment with porcine organs (challenge cohort), the final biomedical application (i.e. robotic kidney surgery) would be targeted on real patients with certain characteristics defined by inclusion criteria such as restrictions regarding sex or age (target cohort).

a) Describe the target cohort, i.e. the subjects/objects from whom/which the data would be acquired in the final biomedical application.

The target cohort are patients requiring orthodontic and/or prosthetic treatment.

b) Describe the challenge cohort, i.e. the subject(s)/object(s) from whom/which the challenge data was acquired.

Intraoral 3D scans for patient dentition requiring orthodontic and/or prosthetic treatment.

Imaging modality(ies)

Specify the imaging technique(s) applied in the challenge.

Intraoral 3D scans.

Context information

Provide additional information given along with the images. The information may correspond ...

a) ... directly to the image data (e.g. tumor volume).

No additional context information will be given.

b) ... to the patient in general (e.g. sex, medical history).

No additional context information will be given.

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

Data is acquired for the intraoral cavity. For a given patient, two 3D scans will be acquired covering the upper and lower jaws.

b) Describe the algorithm target, i.e. the structure(s)/subject(s)/object(s)/component(s) that the participating algorithms have been designed to focus on (e.g. tumor in the brain, tip of a medical instrument, nurse in an operating theater, catheter in a fluoroscopy scan). If necessary, differentiate between target and challenge cohort.

The target of the participating algorithms is all visible teeth in a given intraoral 3D scan.

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Accuracy, Precision.

Additional points: - Detect in 3D any visible tooth in a given 3D scan with a high localization accuracy [please see TLA metric in assessment methods section].
- Correctly identify localized teeth [please see TIR metric in assessment methods section].
- Precisely segment and label all 3D points belonging to all visible teeth [please see TSA metric in assessment methods section].

All corresponding metrics are listed below in section assessment methods (parameter 26).

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

All data are acquired with state of the art intraoral 3D scanners. In practice the IOS scanners used are: the Primescan from Dentsply, the Trios3 from 3Shape, and the iTero Element 2 Plus. These scanners are representative and generate 3D scans with an accuracy between 10 and 90 micrometers and a point resolution between 30 and 80 pts/mm2. No additional equipment other than the IOS itself was used to during the acquisitions.

b) Describe relevant details on the imaging process/data acquisition for each acquisition device (e.g. image acquisition protocol(s)).

In general, no particular protocol is defined for the 3D intraoral scans except that we require the scans to cover the whole upper and lower jaws.

c) Specify the center(s)/institute(s) in which the data was acquired and/or the data providing platform/source (e.g. previous challenge). If this information is not provided (e.g. for anonymization reasons), specify why.

3D scans are originally acquired in partener dental clinics located mainly in France and Belgium. All acquired clinical data are anonymized.

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

Data are acquired by orthodontists/dental surgeons with more than 5 years of professional experience. Also, the clinical evaluators of the challenge have more than 10 years of expertise in orthodontistry, dental surgery and endodontistry.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

All data refer to intraoral 3D scans of one patient's jaw.
We made a data split to get: public training set (60%), public testing set (10%), and a hidden testing set used for participant ranking (30%).
Any visible tooth in all data is annotated by medical experts.
Participant teams will have access only to public sets, i.e., training and testing.

b) State the total number of training, validation and test cases.

- Public training set: 3D scans acquired for 600 patients leading to 2x600 individual scans
- Public testing set: 3D scans acquired for 100 patients leading to 2x100 individual scans.
- Hidden testing set: 3D scans acquired for 300 patients leading to 2x300 individual scans.

c) Explain why a total number of cases and the specific proportion of training, validation and test cases was chosen.

Most state-of-the art research works in this domain use a very limited number of scans, usually under 100 scans, which means potentially under 50 patient use-cases. None of these are made publicly available. We believe that providing 1000 patient use-cases considerably increases the number of publicly available data and can therefore trigger a boost in the research in this field.

The dataset is divided in a way to have enough data (600 patient use-cases) for training allowing enough variability between scans and subjects. We provide a set of 100 test use-cases to let the participants evaluate their algorithms.

Finally, the ranking score is computed based on the hidden testing set consisting of 300 use-cases. These will be accessible only to the organizers and the reviewers.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

Data are collected for patient requiring either orthodontic (50%) or prosthetic treatment (50%)
We will try to tend the patient age and gender distribution to the real-world patient age distribution: 50% male 50% female, about 70% under 16 years-old, about 27% between 16-59 years-old, about 3% over 60 years old).

These proportions are respected in the three datasets.

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

The annotation was performed iteratively by professional annotators including a validation iteration by dentists. 10 annotators are involved in the project. Each single 3D scan is annotated by only one annotator and the annotation systematically controlled by the three clinical evaluators of the challenge

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The annotations were performed using a self made annotation tool kindly provided by our clinical partner Udini. Annotators were instructed to individually outline each tooth border. This step leads to localize precisely separate teeth. While doing the visual check of annotated borders, dentists are also in charge of correcting the attributed tooth labels because of some ambiguous uses-cases.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

Professional annotation company Infolks https://infolks.info/ is in charge of performing the annotation. The annotator team members are highly qualified for such projects with large experience in medical imaging segmentation. In addition, the visual check step, requiring medical expertise, is carried out by our partners dentists.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

No Aggregation.

Data pre-processing method(s)

Describe the method(s) used for pre-processing the raw training data before it is provided to the participating teams. Provide the information separately for the training, validation and test cases if necessary.

No preprocessing steps are required.

Sources of error

a) Describe the most relevant possible error sources related to the image annotation. If possible, estimate the magnitude (range) of these errors, using inter-and intra-annotator variability, for example. Provide the information separately for the training, validation and test cases, if necessary.

Each 3D scan is attributed to only one annotator. In previous experiments manual annotation appeared to be the same in 99.5% of the total teeth vertices, the 0.5% variation arising at teeth boundaries. We decided to systematically validate all the annotations with the three clinical validators in order to ensure accurate boundary annotations.

b) In an analogous manner, describe and quantify other relevant sources of error.

N/A

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

-Teeth localization accuracy (TLA): mean of normalized Euclidean distance between ground truth (GT) teeth centroids and the closest localized teeth centroid. Each computed Euclidean distance is normalized by the size of the corresponding GT tooth. In case of no centroid (e.g. algorithm crashes or missing output for a given scan) a nominal penalty of 5 per GT tooth will be given. This corresponds to a distance 5 times the actual GT tooth size. As the number of teeth per patient may be variable, here the mean is computed over all gathered GT Teeth in the two testing sets.

- Teeth identification rate (TIR): is computed as the percentage of true identification cases relatively to all GT teeth in the two testing sets. A true identification is considered when for a given GT Tooth, the closest detected tooth centroid :
is localized at a distance under half of the GT tooth size, and
is attributed the same label as the GT tooth

- Teeth segmentation accuracy (TSA): is computed as the average F1-score over all instances of teeth point clouds. The F1-score of each tooth instance is measured as: F1=2*(precisionrecall)/(precision+recall)

b) Justify why the metric(s) was/were chosen, preferably with reference to the biomedical application.

- TLA and TIR together capture the identification and localisation capabilities of the algorithm.
- TSA is a fine grained evaluation of 3D points labelling and directly reflects the quality of the teeth segmentation.

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

The global ranking will be based on an averaged sum of the teeth localization accuracy(TLA), teeth identification rate (TIR) and teeth segmentation accuracy (TSA).

Score = [Exp(-TLA) + TIR + TSA]/3.

In addition, we will provide intermediate rankings based on each metric in order to highlight the performances of the competing methods in each of the three subtasks, i.e., localization, identification, and segmentation. Winners will be the top global ranking methods. Although the challenge tasks are not expected to run in realtime, we will add statistics on runtime performances for each participant algorithm as estimated on the challenge platform.

b) Describe the method(s) used to manage submissions with missing results on test cases.

Missing data in the algorithm output (e.g. not providing segmentations for some cases) are allowed but it implicitly penalizes TSA and TRI scores. TLA is handled as described in 26.a.

c) Justify why the described ranking scheme(s) was/were used.

We do not make a difference between the task categories (i.e.,localization, identification and segmentation). Therefore, considering a simple weightenning scheme would be easy to interpret and will lead to a fair ranking. In addition, we will provide intermediate rankings based on each metric in order to highlight the performances of the competing methods in each of the three subtasks.

Statistical analyses

a) Provide details for the statistical methods used in the scope of the challenge analysis. This may include

- description of the missing data handling,
- details about the assessment of variability of rankings,
- description of any method used to assess whether the data met the assumptions, required for the particular statistical approach, or
- indication of any software product that was used for all data analysis methods.

The proposed statistical analysis to rank participants is a weighted mean (27.a). Ranking variability will be characterized using the bootstrap method. Missing data is handled as described in 26.a.

b) Justify why the described statistical method(s) was/were used.

N/A

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

There is no further analysis applicable that is not discussed above.

The ensuing journal article about the challenge will have a detailed analysis on inter-algorithm variability, algorithm-human variability, and an evaluation of the ensemble of algorithms.

ADDITIONAL POINTS

References

Please include any reference important for the challenge design, for example publications on the data, the annotation process or the chosen metrics as well as DOIs referring to data or code.

[1] Lian, Chunfeng, et al. "MeshSNet: Deep multi-scale mesh feature learning for end-to-end tooth labeling on 3D dental surfaces." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019.

[2] Xu, Xiaojie, Chang Liu, and Youyi Zheng. "3D tooth segmentation and labeling using deep convolutional neural networks." IEEE transactions on visualization and computer graphics 25.7 (2018): 2336-2348.

[3] Sun, Diya, et al. "Automatic Tooth Segmentation and Dense Correspondence of 3D Dental Model." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2020.