3DTeethLand: 3D Teeth Landmarks Detection Challenge: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

3DTeethLand: 3D Teeth Landmarks Detection Challenge

Challenge acronym

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

3DTeethLand

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.

Two years ago, we successfully introduced the '3DTeethSeg' challenge dealing with teeth segmentation and labeling tasks from intraoral 3D scans.

Continuing from our previous challenge and striving for in depth perception of intraoral scans, we intend to address within this version of the challenge a more complex task, teeth landmark detection.

This task holds significant importance in modern clinical orthodontics. These crucial landmarks, including features such as cusps and mesial distal locations, play a fundamental role in advancing orthodontic treatment planning and assessment in clinical dentistry [1]. However, several significant challenges could be present given the intricate geometry of individual teeth and substantial variations between individuals [2]. To address these complexities, the development of advanced techniques, particularly through the application of deep learning, is required for the precise detection of 3D tooth landmarks.

This challenge introduces the first publicly available dataset for 3D teeth landmarks detection, encouraging community involvement in a topic with important clinical implications. It plays a key role in advancing automation and leveraging AI for optimizing orthodontic treatments.

[1] B. Woodsend, E. Koufoudaki, P. A. Mossey, and P. Lin, "Automatic recognition of landmarks on digital dental models", Computers in Biology and Medicine, vol. 137, p. 104819, 2021.
[2] S. Triarjo, R. Sarno, S. C. Hidayati, and G. Sihaj, "Automatic 3D Digital Dental Landmark Based on Point Transformation Weight," in 5th International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2023, 2023, pp. 336 341.

Challenge keywords

List the primary keywords that characterize the challenge.challenge_

digital orthodontics, landmark detection, 3D intraoral scan, 3D point cloud

Year

The challenge will take place in 2024

FURTHER INFORMATION FOR CONFERENCE ORGANIZERS

Workshop

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

This challenge will be part of a half-day Thematic Challenge Event focused on dentistry, jointly organized in collaboration with two other challenges.

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.

In the previous version of the challenge, we had a minimum of 400 individual registration, and a final number of 10 participating teams attended the final evaluation phase.

Publication and future plans

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

We plan to coordinate a common publication with team members participating to the final evaluation phase.

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.

Following the previous setup, participating teams will submit their Docker containers to the challenge platform, requiring GPU-equipped servers for inference. Additionally, for the satellite event organization at the MICCAI conference, a projector will be needed for the presentation of the competing solutions.

TASK 1: 3D Teeth Landmarks detection

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.

Two years ago, we successfully introduced the '3DTeethSeg' challenge dealing with teeth segmentation and labeling tasks from intraoral 3D scans.

Continuing from our previous challenge and striving for in-depth perception of intraoral scans, we intend to address within this version of the challenge a more complex task, teeth landmark detection.

This task holds significant importance in modern clinical orthodontics. These crucial landmarks, including features such as cusps and mesial-distal locations, play a fundamental role in advancing orthodontic treatment planning and assessment in clinical dentistry [1]. However, several significant challenges could be present given the intricate geometry of individual teeth and substantial variations between individuals [2]. To address these complexities, the development of advanced techniques, particularly through the application of deep learning, is required for the precise detection of 3D tooth landmarks.

This challenge introduces the first publicly available dataset for 3D teeth landmarks detection, encouraging community involvement in a topic with important clinical implications. It plays a key role in advancing automation and leveraging AI for optimizing orthodontic treatments.

[1] B. Woodsend, E. Koufoudaki, P. A. Mossey, and P. Lin, "Automatic recognition of landmarks on digital dental models," Computers in Biology and Medicine, vol. 137, p. 104819, 2021.
[2] S. Triarjo, R. Sarno, S. C. Hidayati, and G. Sihaj, "Automatic 3D Digital Dental Landmark Based on Point Transformation Weight," in 5th International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2023, 2023, pp. 336-341.

Keywords

List the primary keywords that characterize the task.

digital orthodontics, landmark Detection, 3D intraoral scan, 3D point cloud

ORGANIZATION

Organizers

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

Organizing team:
Achraf Ben-Hamadou, Digital Research Center of Sfax, Tunisia
Sergi Pujades, Inria, Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, France

Ahmed Rekik, Digital Research Center of Sfax, Tunisia
Oussama Smaoui, Udini, Aix-en-Provence, France
Nour Neifar, Digital Research Center of Sfax, 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:
Udini, Aix-en-Provence, France
CRNS Digital Research Center of Sfax, Tunisia

b) Provide information on the primary contact person.

Achraf Ben-Hamadou, Digital Research Center of Sfax, Tunisia

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

This edition is a continuation of the 3DTeethSeg challenge. Repeated event with annual 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

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

None at this moment. Webpage will be similar to the 3DTeethSeg Challenge (https://3dteethseg.grand-challenge.org/)

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.

Not limited to data provided by the challenge. However, participants are asked to provide a description of all datasets used for training to ensure a fair comparison.

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.

Members of the organizers team 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 the evaluation platform grand-challenge.com.

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 5-th March, and will remain open until 31th August 2024, deadline of the final docker submission.
10th April 2024: Expected release of the training data.
1st July 2024: Opening of the submission for the Preliminary Algorithm Submission Phase
30th July 2024: Preliminary Algorithm Submission closes. Final Docker submission portal opens.
31th August 2024. Final 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 (Tunisia) 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 with the contribution of the CRNS. Only the organizers 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

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

Task category(ies)

State the task category(ies)

Examples:

- Classification

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

Detection: for each tooth detect dental landmarks: cusps points, mesial points, distal points, facial axis points, inner points , outer points.

Note that although teeth identification or segmentation is not explicitly requested, the number of landmarks on a tooth depends on the tooth category. For example, a premolar can have two cusps points while a molar could have five cusps.

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)).

Find a dental landmarks detection algorithm with high sensitivity and specificity for 3D intraoral scans.

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 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, Belgium and Tunisia. 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 (72%), public testing set (4%), and a hidden testing set used for participant ranking (24%).
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 900 patients leading to 1800 individual scans
Public testing set: 3D scans acquired for 5 patients leading to 10 individual scans.
Hidden testing set: 3D scans acquired for 300 patients leading to 600 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 around 300 scans [1], which means potentially under 200 patient use-cases. None of these are made publicly available. We believe that providing more than 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 (900 patient use-cases) for training allowing enough variability between scans and subjects. We provide a set of 5 test use-cases to let the participants evaluate their algorithms in the preliminary test phase. Finally, the ranking score is computed based on the hidden testing set consisting of

300 patient use-cases. These will be accessible only to the organizers and the reviewers.

[1] T. Kubík and M. Spanel, "Robust Teeth Detection in 3D Dental Scans by Automated Multi-view Landmarking:," in Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies, 2022, pp. 24-34.

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 patients requiring either orthodontic (50%) or prosthetic treatment (50%). The patient age and gender distributions are: 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 both training and hidden 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. 7 annotators are involved in the project. Each single 3D scan is annotated by only one annotator and the annotation systematically controlled and validated 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 carried out using a custom annotation tool provided by our clinical partner, Udini. Annotators were instructed to identify and recognize every tooth landmark class on any tooth in an intraoral 3D scan. The provided tool enables precise selection of landmark positions on the teeth surfaces and also allows for assigning the corresponding labels; each landmark is attributed one single label.
Udini conducted online training sessions for annotators to familiarize them with the annotation tool and the definition of each landmark category. During these sessions, difficult cases identified by our clinical partners were presented, and annotators were guided on how to address them. Additionally, all annotations are systematically reviewed by our dentist partners for quality assessment.

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. We decided to systematically validate all the annotations with the three clinical validators in order to ensure accurate landmark annotations. From the corrected cases, we learnt the acceptance threshold on the landmark position accuracy.

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)

mAP: mean average precision: is computed by aggregating the average precision(calculated using the area under the curve (AUC) of the Precision x Recall curve for each distance threshold) of each landmark category. This means we have as many mAP values as landmark categories.

mAR: mean average recall: is calculated by aggregating the average recall (the area under the curve of the recall x distance threshold) of each landmark category. This means we have as many mAR values as landmark categories.

Notes:
- Landmark categories: We consider 4 landmark categories within this challenge.

- Localization criterion and thresholds: the localization criterion is based on the Euclidean distance between predicted and reference landmarks. A prediction is considered as a hit if the distance is smaller than a given threshold. As landmarks detection could be valuable for different clinical applications with different localization requirements, the localization performance will be assessed considering different predefined thresholds.

- Assignment strategy: a greedy strategy will be used for the matching between predictions and reference landmarks where all predicted landmarks are ranked by their predicted class scores and iteratively assigned (starting with the highest score) to the closest reference landmarks. NB: A reference landmark can only be assigned to one predicted landmark.

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

Within the existing literature, various metrics such as root-mean-square error [1], average distance error [2] and mean absolute error[3], have been traditionally used for the assessment of landmarks localization, where a fixed number of landmark points is assumed. However, our challenge context differs from this convention as we are addressing a detection task involving a variable number of landmarks per tooth. In this particular context and following [4], we contend that precision (AP) and recall (AR) metrics offer superior performance, aligning more effectively with this task.

[1] S. Triarjo, R. Sarno, S. C. Hidayati, and G. Sihaj, "Automatic 3D Digital Dental Landmark Based on Point Transformation Weight," in 5th International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2023, 2023, pp. 336-341.

[2] G. Wei et al., "Dense representative tooth landmark/axis detection network on 3D model," Computer Aided Geometric Design, vol. 94, p. 102077, 2022.

[3] T.-H. Wu et al., "Two-Stage Mesh Deep Learning for Automated Tooth Segmentation and Landmark Localization on 3D Intraoral Scans," IEEE Transactions on Medical Imaging, vol. 41, pp. 3158-3166, 2022.

[4] Y. You et al., "KeypointNet: A Large-Scale 3D Keypoint Dataset Aggregated From Numerous Human Annotations," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, pp. 13644-13653.

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.

This ranking method was used in several challenges such as BraTS [1], Medical Segmentation Decathlon [2], VerSe'20 [3] and Dentex [4]. Participants have provided positive feedback, primarily for the stability of this method in outlying performances.

[1] B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," IEEE Trans Med Imaging, vol. 34, no. 10, pp. 1993-2024, 2015.
[2] M. Antonelli et al., "The Medical Segmentation Decathlon," Nature Communications, vol. 13, no. 1, p. 4128, 2022.
[3] A. Sekuboyina et al., "VerSe: A Vertebrae labelling and segmentation benchmark for multi-detector CT images," Medical Image Analysis, vol. 73, p. 102166, 2021.
[4] I. E. Hamamci et al., "DENTEX: An Abnormal Tooth Detection with Dental Enumeration and Diagnosis Benchmark for Panoramic X-rays." arXiv, 2023.

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

Missing data in the algorithm output is allowed but it implicitly penalizes mAP and mAR scores.

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

We will adopt a point-based ranking method coupled with the bootstrapping technique to enhance the ranking robustness.
The procedure involves:
1. Computation of metrics (in section 26.a)
2. For each metric, teams are pairwise compared using the Wilcoxon Signed Rank Test. The team deemed 'statistically better' (p-value < 0.001) in each comparison receives one point. This process yields a 'total point count' for each team, reflecting the number of times it outperformed its counterparts.
3. Bootstrapping: Apply bootstrapping as 10% of the data is resampled and step 2 is repeated for the remaining data. This generates a 'total point count' per team for each dropped data.
4.Aggregation of 'total point counts' to determine the final ranking.

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 based on Wilcoxon Signed Rank Test. 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.

Same answer as in 27.c

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.

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.

N/A

Further comments

Further comments from the organizers.

3DTeethLand is a continuation of the previous 3DTeethSeg challenge; however, this version will primarily focus on more complex attributes (landmarks) visible on teeth in intraoral 3D scans.