Pulmonary Artery Segmentation Challenge 2022: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Pulmonary Artery Segmentation Challenge 2022

Challenge acronym

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

Parse2022

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.

It is of significant clinical interest to study pulmonary artery structures in the field of medical image analysis. The pulmonary arteries used in this challenge are located in the chest cavity of the human body, including the pulmonary trunk coming out of the pulmonary valve of the right ventricle, the left and right pulmonary arteries, and their main branches in the lungs. The primary function is to transport the human blood that's low in oxygen and high in carbon dioxide to the pulmonary capillaries of the lungs for the exchange of oxygen and carbon dioxide, which is extremely important. One prerequisite step is to segment pulmonary artery structures from CT with high accuracy and low time-consumption. The segmentation of pulmonary artery structures benefits the quantification of its morphological changes for diagnosis of pulmonary hypertension and thoracic surgery. However, due to the complexity of pulmonary artery topology, automated segmentation of pulmonary artery topology is a challenging task.

Besides, the open accessible large-scale CT data with well labeled pulmonary artery are scarce (The large variations of the topological structures from different patients make the annotation an extremely challenging process). The lack of well labeled pulmonary artery hinders the development of automatic pulmonary artery segmentation algorithm. Hence, we try to host the first Pulmonary ARtery SEgmentation challenge in MICCAI 2022 (Named Parse2022) to start a new research topic and make a solid benchmark for pulmonary artery segmentation task.

We have collected 200 3D volumes with refined pulmonary artery labeling from 10 clinicians, 100 for the training dataset, 70 for the closed testing dataset and 30 for the opened validated dataset. Multi-level Dice, Multi-level HD95, Maximum used memory, and time-cost are adopted as evaluation metrics. This challenge will also promote the pulmonary disease treatment, interactions between researchers and interdisciplinary communication.

Challenge keywords

List the primary keywords that characterize the challenge.

Pulmonary Artery, Segmentation, Quantification

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

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.

This challenge can not only be used to develop new algorithms, but also has clinical significance. To now, more 20 team have applied for registration. Besides, we will reward the top teams with more 1000 dollars to improve the enthusiasm of the contestants. According to our survey of the challenges in the past five years, we expect more than 70 participants. In the past five years, an average of 70 teams participated in medical imaging challenges such as FLARE2021, BraTS, and so on, which is most similar to ours. Therefore, in the Parse2022 challenge, we expect more than 70 participants.

Publication and future plans

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

We plan to publish at least one paper in the top journal of medical image analysis to summarize the challenge results and discuss the future research direction. At most two members in the top-10 teams will be co-authors. In the paper, we will publish a substantial summary of the challenge results.

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.

To achieve the fair comparison and computing environment, we utilize the following measures:
a) Huawei Cloud (https://activity.huaweicloud.com/) will be used as the primary and unified on-site platform. All algorithms are run on the same computing platform based on Docker to guarantee the authenticity of the results and the fairness of the comparison.
b) The website (https://parse2022.grand-challenge.org/) will be used to introduce the challenge and to publish the codes for baseline, evaluation and ranking methods. We will host the challenge, upload data, evaluate predictions and announce the results on the Huawei Cloud website.

TASK: Pulmonary Artery Segmentation

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.

It is of significant clinical interest to study pulmonary artery structures in the field of medical image analysis. The pulmonary arteries used in this challenge are located in the chest cavity of the human body, including the pulmonary trunk coming out of the pulmonary valve of the right ventricle, the left and right pulmonary arteries, and their main branches in the lungs. The primary function is to transport the human blood that's low in oxygen and high in carbon dioxide to the pulmonary capillaries of the lungs for the exchange of oxygen and carbon dioxide, which is extremely important. One prerequisite step is to segment pulmonary artery structures from CT with high accuracy and low time-consumption. The segmentation of pulmonary artery structures benefits the quantification of its morphological changes for diagnosis of pulmonary hypertension and thoracic surgery. However, due to the complexity of pulmonary artery topology, automated segmentation of pulmonary artery topology is a challenging task.

Besides, the open accessible large-scale CT data with well labeled pulmonary artery are scarce (The large variations of the topological structures from different patients make the annotation an extremely challenging process). The lack of well labeled pulmonary artery hinders the development of automatic pulmonary artery segmentation algorithm. Hence, we try to host the first Pulmonary ARtery SEgmentation challenge in MICCAI 2022 (Named Parse2022) to start a new research topic and make a solid benchmark for pulmonary artery segmentation task. And we will reward the top teams to improve the enthusiasm of the contestants.

We have collected 200 3D volumes with refined pulmonary artery labeling from 10 clinicians, 100 for the training dataset, 70 for the closed testing dataset and 30 for the opened validated dataset. Multi-level Dice, Multi-level HD95, Maximum used memory, and time-cost are adopted as evaluation metrics. This challenge will also promote the pulmonary disease treatment, interactions between researchers and interdisciplinary communication.

Keywords

List the primary keywords that characterize the task.

Pulmonary Artery, Segmentation, Automated

ORGANIZATION

Organizers

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

Kuanquan Wang (The head of Perceptual Computing Research Center, The School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, E-mail: wangkq@hit.edu.cn .)
Zhaowen Qiu (The head of Heilongjiang Tuomeng Technology Co., Ltd. of Image Science and Technology, Northeast Forestry University, Harbin 150040, China, E-mail: qiuzw@nefu.edu.cn .)
Wei Wang (The School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001,

China, E-mail: wangwei2019@hit.edu.cn)
Tao Song (Harbin Medical University, Harbin, China, E-mail: 30296557@qq.com)
Shaodong Cao (the Department of Radiology, The Fourth Hospital of Harbin Medical University, Harbin, China, E-mail: shaodong_cao@163.com)
Yi Zhao (Harbin Medical University, Harbin, China)
Jun Liu (The PHD Candidate of Perceptual Computing Research Center, The School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China. E-mail: liujun665@hotmail.com)
Yingte He (The PHD Candidate of Perceptual Computing Research Center, The School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China. E-mail:hyt971206@sina.com)
Shaowei Gan (Northeast Forestry University, Harbin 150040, China, E-mail:gshaowei1996@163.com)
Xinjie Liang (The PHD Candidate of Perceptual Computing Research Center, The School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China. E-mail: xinjiel.hit@gmail.com)
Mingwang Xu (The PHD Candidate of Perceptual Computing Research Center, The School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China. E-mail:1140990160@qq.com)
Ziyu Guo(Northeast Forestry University, Harbin 150040, China, E-mail: 786133379@qq.com)

b) Provide information on the primary contact person.

Kuanquan Wang (wangkq@hit.edu.cn)

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

Open call challenge.

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.

1. Huawei Cloud (https://activity.huaweicloud.com/) will be used as the primary and unified on-site platform.
2. The website (https://parse2022.grand-challenge.org/) will be used to introduce the challenge and to publish the codes for baseline, evaluation and ranking methods.

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

https://parse2022.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.

No additional data allowed.

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 and not listed in leaderboard.

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

1) Successful participation awards, which are electronic certificates, will be awarded to all teams that obtain valid test scores in the challenge leaderboard and complete technical paper submissions reviewed by the organizing committee.
2) The top-1 team will receive 500 dollars or electronic products with similar prices. The exquisite certificates will be awarded to all members of the Top-1 team.
3) The team that wins the second place will receive 300 dollars or electronic products with similar prices. The exquisite certificates will be awarded to all members of the Top-2 team.
4) The team that wins the third place will receive 200 dollars or electronic products with similar prices. The exquisite certificates will be awarded to all members of the Top-3 team.
5) The team achieving the first place in the single index (such as Dice, HD95, or time-consumption) will be awarded to all members with electronic certificates.
6) We are looking for sponsors of other companies (Such as Sensetime and UNITED IMAGING) for various individual awards for teams that make a special contribution in some ways.

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.

1) All the results will be shown publicly on the leaderboard.
2) The top 10 teams will be invited to make a 5-10 minute presentation for the MICCAI2022 challenge session

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).

1) The challenge organizers will publish at least one journal paper with high impact. At most 2 members of the top-10 teams will be co-authors.
2) All participating teams are encouraged to publish their results separately after the challenge, but they should cite the assigned paper.
3) No embargo time is defined.

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.

Docker container on the Huawei Cloud platform. Link to submission instructions: https://parse2022.grand-challenge.org/

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.

At most 1 submission is allowed for opening testing leaderboard. Wrong submissions will not be counted and computing challenge results. After the submission, the scores and leaderboard will be updated after a while.

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: March 28 (11:59PM GMT), 2022
Dataset release: April 6 (11:59PM GMT), 2022
Opened validation leaderboard submission: July 15 (11:59PM GMT), 2022
Challenge leaderboard submission: July 20 (11:59PM GMT), 2022
Submission deadline: July 30 (11:59PM GMT), 2022
Winner and invitation speakers: September 18 (11:59PM GMT), 2022

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).

The ethics approval was obtained by the Ethics Committee of Harbin Medical University

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.

Additional comments: Teams should sign a data usage agreement before downloading data.

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.

The codes to produce the baseline results, evaluations and rankings are available on our Github repository (The website will be open once our proposal is accepted.). Link to the code and documentation will be added to the online platform. The evaluation will be released to contestant before opened validation leaderboard submission. In this way, all the participants will verify the evaluation code by validation dataset and ensure the fairness of the challenge.

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

Teams will be encouraged to share their code, but they have the choice to do so or not.

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.

1) The National Natural Science Foundation of China provides the funding for open-accessible research topic.
2) Only the members of the organization have access to the labels of test dataset.

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, Treatment planning, Assistance, Surgery, Intervention planning, Training.

Task category(ies)

State the task category(ies).

Examples:

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

Segmentation.

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 is composed of patients who require thoracic surgery or treatment, and clinicians who want to make a surgical planning for patients.

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

The challenge cohort is composed of 200 3D CT volume data from the Harbin Medical University. The data were well labelled by 10 clinicians with more than 5 years clinical experience.

Imaging modality(ies)

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

Contrast Enhanced CT Pulmonary Angiography (CTPA)

Context information

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

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

The image sizes are between 512*512*228and 512*512*376. Pixel sizes of these images are between 0.50mm/pixel and 0.95mm/pixel, and their slice thicknesses are 1mm/pixel. The images will be stored in .nii.gz files. Voxel-level segmentation annotations are:
0 - Background
1 - Pulmonary artery

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

Not limited on patients with certain characteristic.

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.

Contrast Enhanced CT Pulmonary Angiography (CTPA) data from the dual-source 64-slice CT scanner in Harbin Medical University will be acquired in the final biomedical application.

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 algorithm achieves segmentation of pulmonary artery on Contrast Enhanced CT Pulmonary Angiography (CTPA) data

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, Runtime, Hardware requirements, Robustness.

Additional points: a) Multi-level Dice Similarity Coefficient (Dice)
b) Multi-level 95% Hausdorff distance (HD95)
In this challenge we will use two levels according to medical standards: the first level includes the aorta trunk and the left and right pulmonary arteries; the second level contains the other major pulmonary artery branches in the lungs.

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).

Contrast Enhanced CT Pulmonary Angiography (CTPA) data from the dual-source 64-slice CT scanner.

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

The image sizes are between 512*512*228 and 512*512*376. Pixel sizes of these images are between 0.50mm/pixel and 0.95mm/pixel, and their slice thicknesses are 1mm/pixel.

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.

Harbin Medical University, Harbin, China

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).

The data were well labelled by 10 clinicians with more than 5 years clinical experience.

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.

Training, validation, and test cases all represent a 3D pulmonary artery CT image and have well labelled groundtruth corresponding to pulmonary artery.

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

Training cases: 100
Opened validation cases: 30
Closed test cases: 70

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

Training cases: 100 (The relatively large number of data were used for training a robust model).
Opened validation cases: 30 (The relatively small number of data were used for validation of algorithm from different participants to verify the evaluation code by validation dataset and ensure the fairness of the challenge. At the same time, the relatively small number of data can avoid the disclosure of test set data distribution).
Closed test cases: 70 (The relatively large number of data were used for a fair final leaderboard).

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.

The data distribution in pulmonary artery segmentation tasks chosen according to real-world distribution

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.

10 clinical experts participated in the labeling work , we divide the 10 clinical experts into two groups, each group includes 5 experts , each case was labelled by 5 experts.

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.

1) The annotation is performed using MIMICS software. Adjust window width and window level so that structures show clearly.
2) The annotation of the pulmonary artery is semi-automatic. It is based on the method of region growing. The seed point is selected iteratively and manually.
3) Finally, all clinical experts fine-tune the annotation results of the pulmonary artery structure and check the annotation mutually.

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.

In each group, 10 experts with more than 5-year clinical experience annotated the cases and checked the labeling results by each other.

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.

After mutual inspection, the consistency of label is very high. Finally, the method of voting is used to merge 5 annotations for one case.

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.

For all cases, the preprocessing methods involve:
a) Remove the private information and convert DICOM files into nii.gz files.
b) Manually review each image for quality

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.

A small portion of tiny vessels will be only 1-2 voxels, which would possibly result in the inconsistent annotation.

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

Because the pulmonary artery and pulmonary vein are intertwined, it is challenging to distinguish the terminal vessels some times.

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)

a) Multi-level Dice Similarity Coefficient (Dice)
b) Multi-level 95% Hausdorff distance (HD95)
c) Running time
d) Maximum used memory

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

We graded the pulmonary artery blood vessels into multiple levels according to the radius of blood vessels, and evaluated the accuracy of trunk and branch respectively. The segmentation performance is evaluated in four aspects following:
1) Multi-level Dice Similarity Coefficient (Dice): Dice is used to evaluate the area-based overlap index.
2) Multi-level 95% Hausdorff distance (HD95): HD95 is used to evaluate the coincidence of the surface for stability and is sensitive to outliers.
3) Running time: Low time-consumption is preferred for good algorithm.
4) Maximum used memory: Low memory-consuming indicates the good application potential in more widely used computing platform.

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 ranking scheme includes the following steps:
1) Calculate the Dice, HD95, Maximum used memory, and Running time for all cases.
2) Rank the Dice, HD95, Maximum used memory, and Running time separately.
3) Average these rankings.
4) Tie if the rankings are equal.

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

Report errors and teams need to help participant to complete the missing results in the first time.

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

This ranking scheme is developed to take four important metrics into account. It provides a balanced scheme to judge whether the method can achieve accurate segmentation with lower runtime and memory simultaneously.

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.

When missing data, report errors and teams need to help participants to complete the missing results in the first time.
For each method, more than 4 rankings will be made and each ranking will be decided by the t-test. Then the final ranking will be calculated as the average score of the more than 4 rankings. Besides, we will check if the rankings are equal.

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

1. This ranking scheme avoids that just several testing cases perform better than other methods. The scheme tends to choose more stable and robust models which are useful for most cases. Therefore, the average operation is adopted.
2. This ranking scheme avoids that just one kind of metric performs better than other methods. The scheme tends to choose more accurate and less sensitive to outliers models which meet medical requirements and great clinical significance. Therefore, multiple evaluation metrics including the distance-based and area-based metrics are adopted.

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.

The combining algorithms via ensembling, inter-algorithm variability, common problems/biases of the submitted methods, or ranking variability are all will be analyzed in the future to publish in top journal in the field of medical image analysis.

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. Maier-Hein et al. (2018). Why rankings of biomedical image analysis competitions should be interpreted with care. Nature Communications, 9(1), 5217. doi: https://doi.org/10.1038/s41467-018-07619-7
2. Reinke et al. (2018). How to exploit weaknesses in biomedical challenge design and organization. Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 (pp. 388-395). Springer International Publishing. doi: https://doi.org/10.1007/978-3-030-00937-3_45
3. Maier-Hein et al. (2020) BIAS: Transparent reporting of biomedical image analysis challenges. Medical Image Analysis, 101796. doi: https://doi.org/10.1016/j.media.2020.101796