Automated Segmentation Of Coronary Arteries: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Automated Segmentation Of Coronary Arteries

Challenge acronym

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

ASOCA

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.

Cardiovascular disease is a major cause of death. Medical imaging such as Computed Coronary Tomography Angiography (CCTA) is often used to evaluate the severity of coronary artery disease. Due to the non-invasive nature of CCTA, it is also commonly used for evaluating and reconstructing heart and coronary vessel structures. Coronary artery tree models have a wide array of applications in anatomy, physiology and pathophysiology for educational, training and research purposes; Including study of anatomy of coronary vessels and effects of disease on anatomy, developing machine learning models for disease risk prediction, education and training of medical professionals, In-silico testing of medical devices and 3D printing models for testing and education.

Due to the small size of coronary arteries, possible disease and image artefacts, segmentation of coronary arteries has been focused on semi-automatic methods where a human expert guides the algorithm and corrects errors. This severely limits large scale processing of medical images and possibility of integration with clinical systems. Previous challenges have been focused on specific tasks such as centreline extraction, stenosis quantification and segmentation of specific artery segments. However, to our knowledge this is the first challenge focused on developing fully automatic segmentation methods of the full coronary artery tree. This challenge aims to establish a large standardized annotated dataset of healthy and diseased coronary vessels and utilize this dataset to develop fully automated segmentation algorithms. Automated segmentation would allow processing the large number of CCTAs available to create larger datasets for use cases mentioned above.

Challenge keywords

List the primary keywords that characterize the challenge.

CT, Cardiovascular disease, image segmentation, coronary artery disease

Year

The challenge will take place in ...

2020

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.

Based on similar challenges in the past (Coronary Artery Stenoses Detection and Quantification Evaluation Framework), we would expect 50 initial participants. The challenge will remain open after the conference so we would expect other participants.

Publication and future plans

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

A manuscript will be written detailing the top 10 ranking submissions with up to two authors from each top 10 submission invited as co-authors.

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.

We will be using the Grand Challenge platform (grand-challenge.org/) for hosting the challenge, distributing data and evaluating results. No computing resources will be provided to the participants. Required hardware will likely include a computer, potentially with GPUs should participant wish to use GPU accelerated algorithms.

TASK: Automated Segmentation Of Coronary Arteries

SUMMARY

Keywords

List the primary keywords that characterize the task.

Cardiovascular disease, image segmentation, coronary artery disease

ORGANIZATION

Organizers

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

Ramtin Gharleghi, School of Mechanical and Manufacturing, University of New South Wales
Dr. Gihan Samarasinghe, School of Computer Science and Engineering, University of New South Wales
Professor Arcot Sowmya, School of Computer Science and Engineering, University of New South Wales
Dr. Susann Beier, School of Mechanical and Manufacturing, University of New South Wales

b) Provide information on the primary contact person.

Ramtin Gharleghi <r.gharleghi@student.unsw.edu.au>

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.

Examples:

- One-time event with fixed submission deadline
- Open call
- Repeated event with annual fixed submission deadline

Challenge will be a one time event with fixed submission deadline, in order to have all results ready by MICCAI 2020 conference. The challenge will continue accepting submissions after the deadline in order to serve as a standardized reference, however these submissions will not be included in the manuscript submitted.

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.

The challenge will be run on the Grand Challenge platform. We will use this platform for participant registration, distributing challenge data and evaluating results.

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

https://asoca.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.

Data used must be publicly available. Use of pre-trained neural networks (or similar) is permitted provided that the training data and architecture is publicly available.

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.

A manuscript will be written detailing the top 10 ranking submissions with up to two authors from each top 10 submission invited as co-authors.

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.

A leader board with results of all participating will be posted publicly after the conclusion of the challenge. Teams can request to not be included on the leader board if they wish to do so.

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).

Participating teams may publish their own results separately from the challenge, however this must take place after the submission of the challenge paper described above.

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.

We will be using the valuation framework available on the Grand Challenge platform. Participants will submit final segmentation labels for images provided on https://asoca.grand-challenge.org/. Submission instructions will be available on the same platform.

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.

Up to three submissions will be allowed, with the results of the last submission used to compute challenge rankings.

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

May 1, 2019 --- Release of training dataset

August 15, 2019 --- Release of test dataset

September 15, 2019 --- Submission deadline

September 30, 2019 --- Announcement of results.

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 data was collected under ethics approvals by:
- University of New South Wales Human Research Ethics Committees, Ref. No. HC190145, Apr/2019 - Apr/2024
- University of Auckland Human Participants Ethics Committee, Ref. No. 022961, May/2019 - May/2022

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)

Participating teams are not allowed to:
- Share the data with others outside of their research team.
- Use provided data for any purpose other than evaluating coronary vessel image segmentation methods as part of this challenge.
- Use the data for development of a commercial system.

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 used to produce rankings will be provided to the participants for use with the training samples provided. The same evaluation code will be used to produce final rankings.

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

Participating teams must make their code available to the organizer for the purpose of assessing that meets required criteria (fully automatic algorithm, no use of private data, and any other requirements specified in the challenge rules). Participating teams maintain full ownership of developed algorithms and other intellectual property; The organisers will not use this code for any purpose other than assessment of conformance with challenge rules.

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.

The organizers declare that they have no conflicts of interests. The annotated test cases will be provided to the organizers along with the rest of the training data, however it will only be used to evaluate submissions at the end of the challenge.

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

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.

Patients requiring treatment or monitoring for Cardiovascular disease. The data-set could be used for a several purposes, including improving screening, and treatment methods, training medical professionals, testing medical devices (computational simulation or in 3D printed models), hence the inclusion criteria is very broad.

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

The subjects were adults presenting to Mercy Angiography (Auckland, NZ) with suspected cardiovascular disease. After evaluation by a cardiologist, 30 patients with disease and 30 showing no signs of disease were randomly selected for this challenge.
Details of patient demographics:
Age: 55.8±7.5
Weight: 78± 14.5
Height: 1.69±0.1
BMI: 27.4±4.36
Sex: 67% Female, 33% Male
Ethnicity: Mostly European (87%)

Imaging modality(ies)

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

Computed Coronary Tomography Angiography (CCTA).

Context information

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

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

Images correspond to standard CT Angiography images, showing the region around the heart and coronary vessels. A contrast agent (60-80ml Omnipaque 350) was used. Patient heart rate was lowered to below 60bpm by administration of beta blockers. Data was collected using retrospective ECG-gated acquisition (GE LightSpeed 64 slice CT Scanner,USA. The timepoint used for the challenge is late diastole (75% cardiac cycle).

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

Patients included in the challenge had been undergoing evaluation or treatment for cardiovascular disease at a hospital. Other than excluding under-age patients, no explicit inclusion/exclusion criteria was used. The patient cohort is mainly European (87%), female (67%) patients. The healthy cases had zero calcium score and no visible stenoses, calcification or other vascular disease. Diseased vessels showed significant calcification or stenosis reported by the cardiologist.

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 from CCTA images, which include the region around the heart. While some algorithms may applicable to other imaging modalities, this challenge will only focus on CCTA imaging with a contrast medium.

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 structure is the coronary vessels supplying blood to the heart. This includes the left and right coronary arteries, and their branching segments (Left Anterior Descending, Left Circumflex, Septal, Diagonal, Obtuse Marginal, and Ramus Intermedius if present).

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)).

Feasibility, Interaction, Robustness, Specificity, Sensitivity.

Additional points: Algorithms should be automated and have minimal requirements for user interaction and expert input; while providing acceptable segmentation quality and precise delineation of the artery wall. This would allow reproducible processing of the large amount of medical data available.

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).

Data was collected using a GE LightSpeed 64 slice CT Scanner (USA).

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

A contrast agent (60-80ml Omnipaque 350) was used. Patient heart rate was lowered to below 60bpm by administration of beta blockers. Data was collected using retrospective ECG-gated acquisition (GE LightSpeed 64 slice CT Scanner, USA. The timepoint used for the challenge is late diastole (75% cardiac cycle).

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.

The data was provided by Mercy Angiography (now known as Intracare), located in Epsom, Auckland, New Zealand.

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).

Established 1989, Mercy Angiography is a world leading provider of imaging and cardiology services.

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 and test cases refer to CCTA images of one patient. In both cases the coronary artery lumen is annotated by medical experts. Annotations for the test cases will not be available to participants to avoid over-fitting results to the test cases.

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

20 diseased and 20 healthy cases along with annotations will be provided to participants for training, with 10 separate diseased and 10 healthy cases used for testing and computing participant rankings.

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

30 diseased and 30 healthy cases are included in this challenge keeping in mind the manual time required to generate annotations. Using 33% of the data for testing allows providing more training data than similar challenges

while still leaving out enough test data to ensure robust and reliable evaluation of submitted algorithms.

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.

We have chosen equal class distribution between diseased and healthy vessels, in order to allow development of algorithms which are robust to presence of disease. The real world distribution of these cases would include significantly more healthy than diseased cases.

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 images were manually annotated by three annotators.

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 3D Slicer software. The annotators were instructed to segment the vessel lumen, not including calcified regions or other disease blocking the lumen.

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.

The cases were manually annotated by two cardiology registrars, with previous experience and training working with CCTA images, as well as a PhD candidate working on the project and experienced with image segmentation.

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.

Segments of the vessel that are included in at least two out of the three annotations will be added to the vessel mask.

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 pre-processing is required (other anonymization to remove personally identifiable information). Participants will be provided with DICOM files containing the volumetric CCTA data and annotation masks (only in training cases).

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.

The diameter of coronary arteries decreases the further away they are form the aortic arch. Due to resolution limits of CCTAs, some vessels will be only be 1-2 voxels, which would likely result in inconsistent annotation. Majority of the coronary artery tree is large enough to allow accurate annotation.

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

Deviation from the image acquisition protocol can harm the quality of the images collected. Given the experience level of Intracare staff, we do not expect this to be a significant problem.

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)

Dice Similarity Coefficient (DSC) and 95% Hausdorff Distance will be used to compute final rankings.

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

Dice Similarity Coefficient should provide a balance between the specificity and sensitivity assessment aims. It has been shown to provide a good measure of segmentation quality [1], and is widely used in image segmentation applications. For small structures and where accurate delineation of the boundaries is important, it is recommended to also use a distance based metric [2]. Since maximum Hausdorff distance is very sensitive to noise, we will be using the 95th percentile Hausdorff distance which tolerates small outliers [2].

[1] K. H. Zou et al., "Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports," Academic radiology, vol. 11, no. 2, pp. 178–189, 2004.
[2] A. A. Taha and A. Hanbury, 'Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool', BMC medical imaging, vol. 15, no. 1, p. 29, 2015.

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.

Dice similarity coefficient and 95% Hausdorff distance will be generated for each case based participant submission and expert annotations. Participants will be ranked separately based on the mean dice coefficient and mean Hausdorff distance. The average of these rankings will be used to generate the overall ranking of the participants.

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

Submissions with missing results will not be accepted. Accepting submissions with missing results may allow participants to not attempt difficult cases, especially diseased arteries.

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

Due to the use of two metrics with very different ranges, the ranking scheme used is a feasible method of combining rankings generated from each metric. Case-level aggregation follows the recommendation that the arithmetic mean rather than median be used to aggregate results [1].

[1] L. Maier-Hein et al., “Why rankings of biomedical image analysis competitions should be interpreted with care,” Nature communications, vol. 9, no. 1, p. 5217, 2018.

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 data provided to participants will not have missing cases or labels. Missing data in the algorithm output (e.g. not providing segmentations for some cases) will not be accepted.
The metrics used in the study (Dice score and Hausdorff distance) will be calculated, ranked and aggregated using python and statistical analysis performed using python and the statsmodels package.

Variability of ranking and segmentation output is not considered for the leaderboard, however it would calculated and reported in the manuscript of the challenge results. For example we would be comparing the algorithm error to the inter annotator variability and comparing variability of algorithm performance between diseased and healthy subgroups. Interclass correlation coefficent will be used as the measure of variability calculated through variance component analysis [1].

[1] T. A. Snijders and R. J. Bosker, Multilevel analysis: An introduction to basic and advanced multilevel modeling. Sage, 2011.

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

The aggregation methodology used has been recommended for combining multiple metrics in a challenge. Ranking variability is calculated based on the recommendations of Snijders and Bosker.

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 are no further analysis applicable that are not discussed above. Participants are free to use ensembling in their algorithms, however we will not be combining algorithms from multiple participants through ensembling.

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.

Collection of the medical images is described in the publications below:

[1] P. Medrano-Gracia et al., “Construction of a Coronary Artery Atlas from CT Angiography,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014, vol. 8674, P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, Eds. Cham: Springer International Publishing, 2014, pp. 513–520.
[2] P. Medrano-Gracia et al., “A computational atlas of normal coronary artery anatomy,” EuroIntervention, vol. 12, no. 7, pp. 845–854, Sep. 2016.