Multi-sequence CMR based myocardial pathology segmentation challenge: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Multi-sequence CMR based myocardial pathology segmentation challenge

Challenge acronym

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

MyoPS 2020

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.

Assessment of myocardial viability is essential in the diagnosis and treatment management for patients suffering from myocardial infarction (MI). Cardiac magnetic resonance (CMR) is particularly used to provide imaging anatomical and functional information of heart, such as the late gadolinium enhancement (LGE) CMR sequence which visualizes MI, the T2-weighted CMR which images the acute injury and ischemic regions, and the balanced-Steady State Free Precession (bSSFP) cine sequence which captures cardiac motions and presents clear boundaries. Combining these multi-sequence CMR data can provide rich and reliable information as well as morphological information of the myocardium.

The target of this challenge is combining multi-sequence CMR data to segment each position of the myocardium into different pathologies, including normal myocardium, infarction and edema. This is referred to as the myocardium pathology classification/segmentation, which is crucial for the diagnosis and treatment management of patients. This is, however, still arduous, particularly due to the pathological myocardium. Since manual delineation is generally time-consuming, tedious and subjects to inter- and intra-observer variations, automating this segmentation is highly desired in clinical practice.

The proposed challenge will provide the three sequence CMR images from 45 patients [1, 2], for developing novel algorithms that can segment myocardial pathology combining the complementary information from these three-sequence CMR images. The challenge presents an open and fair platform for various research groups to test and validate their methods on these datasets acquired from the clinical environment. The aim of the challenge will not only be to benchmark various myocardial pathology segmentation algorithms, but also to cover the topic of general cardiac image segmentation, registration and modeling, and raise discussions for further technical development and clinical deployment.

Challenge keywords

List the primary keywords that characterize the challenge.

multi-sequence CMR; myocardial infarction; myocardial pathology segmentation

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.

The Statistical Atlases and Computational Modeling of the Heart (STACOM) workshop

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.

45

In the MS-CMR Seg challenge 2019 (before the ddl: 9 July, 2019) received 79 participants, namely 79 teams were registered, 23 results were submitted for evaluation and 14 works were selected for presentation in the workshop. The number of people attending the oral session was not counted, but the seminar room was fully occupied (for details, one can refer to the photos taken in the event: zmiclab.github.io/mscmrseg19/workshop.html). Therefore, we expect this year around 45 teams will participate in this event according to previous years.

Publication and future plans

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

We will consider to submit a benchmark paper to a top tier journal, e.g. Medical Image Analysis or IEEE Transactions on Medical Imaging.

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.

1 projector, 1 computer (or 1 laptop), 1 monitor, 2 loud speakers, 3 microphones. Note that the challenge is joined with the STACOM workshop, so these are shared with the workshop.

TASK: Myocardial Pathology Segmentation

SUMMARY

Keywords

List the primary keywords that characterize the task.

multi-sequence CMR; LGE CMR; myocardial infarction (MI); edema

ORGANIZATION

Organizers

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

-Xiahai Zhuang, School of Data Science, Fudan University, Shanghai China.
-Lei Li, School of Biomedical engineering, Shanghai Jiao Tong University, China.

b) Provide information on the primary contact person.

zxh@fudan.edu.cn
lilei.sky@sjtu.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.

Examples:

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

Open call.

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.

https://cmt3.research.microsoft.com/MyoPS2020

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

http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/MyoPS20/

Participation policies

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

Fully automatic., Semi automatic.

Additional points: The challenge will provide the manual segmentation of myocardium. The participants could optionally employ the manually labeled myocardium as an initialization, namely semi-automatic myocardial

pathology segmentation algorithm. Alternatively, the participants could also propose a fully-automatic myocardial pathology segmentation algorithm.

Two different designs of the framework, semi-automatic and fully-automatic, will be seperately evaluated.

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 not participate.

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

The best three works will be awarded with prizes, i.e., one best paper and two runner-ups. The selected papers will be published in Lecture Notes in Computer Science (LNCS).
Note: Paper, presentation and test results will all be taken into account when a piece of work is evaluated.

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.

Each participating team can choose whether their 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).

1. The participant who proposes a novel algorithm and obtains promising results will be selected as an author.
2. The participants need to cite the required papers when you use the data for publications.
3. There is no embargo time

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.

Participants send the results to the organizers via email or uploading to a specified 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.

Maximum number of submissions: 2 (Only the first two submissions are evaluated for a team.)

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

Training data release: April 1st, 2020
Test data release: April 1st, 2020
Submission deadline: July 15 th, 2020
Notification of acceptance: July 23 rd, 2020
Workshop Camera ready: July 31 st, 2020
Workshop Date: (TBA) Oct 4 th, 2020, Lima, Peru.

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).

No.
The study had been approved by the institutional review board and the data had been anonymized before Sept 2018, Shanghai, China.

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 SA.

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 evaluation code may be made public after the challenge.

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

The accessibility of the participating teams' code will be decided by the participants.

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 sponsors are not decided yet. The sponsor candidates are Circle and Sensetime. The logos and names of sponsors will be presented on the challenge website, and the sponsorship fee will be used to pay the awards of the winners.

All the participants who submit the signed data agreement can access the test data labels after 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

Treatment planning, Diagnosis, Decision support, 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 who underwent cardiomyopathy and got CMR scans including T2-weighted 3D acquisitions, balanced-Steady State Free Precession (bSSFP) cine sequence, and late gadolinium enhancement (LGE) cardiac magnetic resonance (CMR) sequence.

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

The challenge cohort is same as the target cohort.

Imaging modality(ies)

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

The imaging technique applied in this challenge is MRI. Specifically, the bSSFP CMR is a balanced steady-state, free precession cine sequence. The LGE CMR is a T1-weighted, inversion-recovery, gradient-echo sequence. The T2 CMR is a T2-weighted, black blood Spectral Presaturation Attenuated Inversion-Recovery (SPAIR) sequence.

Context information

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

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

Myocardial pathology

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

Myocardial infarction patient

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.

Heart shown in multi-sequence CMR, including the late gadolinium enhancement (LGE) CMR, the T2-weighted CMR and the balanced-Steady State Free Precession (bSSFP) cine sequence.

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.

Infarcted and edema regions in the myocardium.

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, Accuracy.

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).

Not available

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

The bSSFP CMR is a balanced steady-state, free precession cine sequence. Since both the LGE and T2 CMR were scanned at the end-diastolic phase, the same cardiac phase of the bSSFP cine data was selected for this study. The bSSFP images generally consist of 8 to 12 contiguous slices, covering the full ventricles from the apex to the basal plane of the mitral valve, with some cases having several slices beyond the ventricles. The typical parameters are as follows, TR/TE: 2.7/1.4 ms; slice thickness: 8-13 mm; in-plane resolution: reconstructed into 1.25×1.25 mm.

The LGE CMR is a T1-weighted, inversion-recovery, gradient-echo sequence, consisting of 10 to 18 slices and covering the main body of the ventricles. The typical parameters are as follows, TR/TE: 3.6/1.8 ms; slice thickness: 5 mm; in-plane resolution: reconstructed into 0.75×0.75 mm.

The T2 CMR is a T2-weighted, black blood Spectral Presaturation Attenuated Inversion-Recovery (SPAIR) sequence, generally consisting of a small number of slices. For example, among the 35 cases, 13 have only three slices, and the others have five (13 subjects), six (8 subjects) or seven (one subject) slices. The typical parameters are as follows, TR/TE: 2000/90 ms; slice thickness: 12-20 mm; in-plane resolution: reconstructed into 1.35×1.35 mm.

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 institute in which the data was acquired will not be provided for anonymization reasons.

The data providing platform/source is same as our last challenge, i.e., Multi-sequence Cardiac MR Segmentation Challenge 2019.

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).

Experienced physicists specialized in cardiac MRI.

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.

A case refers to a patient with three sequence CMR images, which includes LGE, T2 and bSSFP CMR.

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

Training + validation cases: 30
Test cases: 15

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

Firstly, the total number of cases is limited, because few patients underwent scanning of all three CMR sequence including LGE, T2 and bSSFP CMR. More importantly, labeling three sequences manually is laborious and time-consuming.
Secondly, the task of this challenge is arduous, so sufficient training data is required to ensure reasonable performance. Therefore, the proportion of training and test, i.e., 2:1, is chosen.

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.

None

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.

Each of the CMR images had been manually delineated three times by three independent, well-trained observers who were not aware of the methodology of this work.

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 annotators were instructed to segment the ventricles and myocardium slice-by-slice, using the brush tool in the ITK-SNAP.

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.

Experienced physicists specialized in cardiac MRI or Phd students who major in biomedical image processing.

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.

The final gold standard segmentation was achieved by averaging the three manual delineations using the shape-based average approach.

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.

The slice will be pre-aligned together by affine transformation to avoid slice shift between different sequences. The participants can choose whether to use pre-processed data or original data. The final evaluation of the submitted algorithms will be evaluated separately.

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.

Inter- and intra-annotator variability will be offered.

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

None

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)
Average Contour Distance (ACD)
Hausdorff Distance (HD)
Statistical measures include Accuracy, Sensitivity and Specificity.

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

Zhuang proposed a multivariate mixture model for simultaneous registration and segmentation of multi-source

images, where DSC, ACD and HD were used as evaluated metrics [1]. Besides, the task of the challenge is to segment the normal myocardium, infarction and edema, which are discrete and small targets. Therefore, the Dice score of the results tends to be low, which sometimes result in the misleading in evaluating the performance. Therefore, other statistical measurements are also included as indicators of quantification performance. This is inspired by the work of LearnGC which achieved the atrial scar segmentation (it is also a small target) [3].

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.

For each participant we will compute the average metric value for the DSC and HD. Build the rank for each participant by sorting the accumulated metric values. In addition, the participants using different combinations of training data will be evaluated separately. In case of tied positions, computational runtime is compared to perform the ranking.

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

Only complete submissions are evaluated.

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

Refer to previous experience on this task.

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.

We will provide the DSC, ACD and HD as evaluation metrics for each case submitted to the challenge by a particular method. From this detail of results, the mean and standard deviation, as well as all percentile if needed, will be calculated for each group of results.

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

Test the sensitivity of a ranking by leaving out different amounts of test data, which embodies the data variability. A t-test is a type of inferential statistical method used to determine if there is a significant difference between the data of two groups, which would be helpful for the final ranking.

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.

None

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] Xiahai Zhuang. "Multivariate mixture model for myocardial segmentation combining multi-source images." IEEE transactions on pattern analysis and machine intelligence 41.12 (2018): 2933-2946.
[2] Xiahai Zhuang. "Multivariate mixture model for cardiac segmentation from multi-sequence MRI." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2016.
[3] Lei Li, et al. "Atrial scar quantification via multi-scale CNN in the graph-cuts framework." Medical Image Analysis 60 (2020): 101595.

Further comments

Further comments from the organizers.

None