Multi-Disease, Multi-View & Multi-Center Right Ventricular Segmentation in Cardiac MRI (M&Ms-2): Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Multi-Disease, Multi-View & Multi-Center Right Ventricular Segmentation in Cardiac MRI (M&Ms-2)

Challenge acronym

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

M&Ms-2

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.

In recent years, many machine/deep learning models have been proposed to accurately quantify and diagnose cardiac pathologies [1-4]. These models rely on accurate segmentation of the ventricles to allow the extraction of robust clinical features in practice. In particular, the segmentation of the right ventricle (RV) is a challenging task due to the highly complex and variable shape of the RV and its ill-defined borders in cardiac MR images [5]. Yet, quantitative features and clinical indices of right ventricular function are sensible to subtle changes in shape and texture [6]. Furthermore, when machine/deep learning models are tested on unseen datasets acquired from distinct disease groups or clinical centers, the segmentation accuracy can be greatly reduced [7]. Despite recent advances in deep learning, robust segmentation of the RV continues to pose challenges in practice and there is a need for new methods to handle the inherent geometrical and textural complexities, in particular in the presence of RV related pathologies (e.g. Dilated Right Ventricle, Tricuspid Insufficiency, Arrhythmogenesis, Tetralogy of Fallot and Interatrial Comunication).

The last MICCAI challenge to focus on right ventricular segmentation took place in 2012, well before the deep learning revolution, and was based on 48 cases from a single clinical center. In this challenge, we invite participants to implement and evaluate advanced approaches based on machine/deep learning for right ventricular segmentation in a multi-disease, multi-view and multi-center setting. Specifically, we extend last year's M&Ms challenge and dataset, which included mostly pathologies of the left ventricle, by presenting a large dataset of 450 cardiac MRI datasets which will includes pathologies that are associated with various right ventricular abnormalities and remodelling, including Dilated Right Ventricle, Tricuspid Insufficiency, Arrhythmogenesis and Tetralogy of Fallot and Interatrial Comunication. A novel aspect of this challenge is the inclusion of long-axis images to help the automatic definition of the basal plane of the RV, which can be confused with the right atrium. The challenge will be supported by the H2020 euCanSHare project (www.eucanshare.eu), which is building a multi-center big data platform for cardiovascular personalised medicine research.

Challenge keywords

List the primary keywords that characterize the challenge.

Cardiac image segmentation – Right ventricle – Generalizable deep learning – Machine/deep learning - Multiview Segmentation

Year

The challenge will take place in ...

2021

FURTHER INFORMATION FOR MICCAI ORGANIZERS

Workshop

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

STACOM (Statistical Atlases and Computational Modelling of the Heart)

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.

Between 12 and 20 teams, based on the number of participants in the past edition of the M&Ms challenge.

Publication and future plans

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

A challenge journal paper will be prepared and submitted. We propose two co-authors per participating center (e.g. junior and senior author). Participants are free to publish their results separately.

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.

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

The platform used to run the competition will be CodaLab. CodaLab is an open-source platform that provides an ecosystem for conducting computational research in a more efficient, reproducible, and collaborative manner and it allows to implement competitions. In a typical CodaLab competition, participants compete to find the best

approach for a particular problem. The M&Ms challenge will be conducted in development and final phases. The appropriate data is made available to participants at each phase of the competition. During the development phase, participants have access to training data to develop and refine their algorithms. During the validation and testing phases, the participants will submit their codes in Docker containers to ensure a fully blinded competition. Results will be calculated at the end of each phase, at which point participants can see the competition results on the leaderboard.

All challenge information will be hosted on a website created for that purpose managed by Universitat de Barcelona in the next months.

TASK: Segmentation

SUMMARY

Keywords

List the primary keywords that characterize the task.

Cardiac image segmentation – Multiple MRI machines – Machine/deep learning

ORGANIZATION

Organizers

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

Carlos Martín-Isla (Universitat de Barcelona, Spain)
José F. Rodríguez Palomares (Vall d'Hebron Hospital, Barcelona, Spain)
Andrea Guala (Vall d'Hebron Hospital, Barcelona, Spain)
Sergio Escalera (Universitat de Barcelona, Spain, Computer Vision Center, Spain)
Karim Lekadir (Universitat de Barcelona, Spain)

b) Provide information on the primary contact person.

Contact: Carlos Martín-Isla (carlos.martinisla@ub.edu)

Life cycle type

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

Examples:

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

One time event with fixed submission deadline.

Challenge venue and platform

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

MICCAI.

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

codalab.org

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

https://www.ub.edu/mnms-2/

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 not participate.

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

200 euros for the best performing method and 200 euros for the best paper award.

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.

The best performing method and the most novel method (best paper award) will be announced publicly. The final rank for all the sumbissions will be also 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).

A challenge journal paper will be prepared and submitted. We propose two co-authors per participating centre (e.g. junior and senior author). Participants are free to publish their results separately. Every team is required to write and submit a paper about their method.

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.

Algorithm output will be submitted to codalab.org platform. Submission instructions will be defined with the dataset release by e-mail.

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.

Evaluation over a 40 cases validation set will be allowed up to 5 times per day. The final solution will be evaluated once over a 160 cases test set.

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

Release of training data: 1st May 2021
Registration period: 1st May - 31st May 2021
Last submission: 15th July 2021
Workshop day: 3rd October 2021
Results release: 3rd October 2021

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 ethical approvals will be obtained from the participating clinical centres. All datasets are fully anonymized.

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.

Additional comments: Additional comments: The data will be distributed under the license CC BY and the challenge participants will need to commit to not disseminate the data during the challenge duration.

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 code for computing the performance scores used by the organizers is based in the public repository for medpy library: https://github.com/loli/medpy/.

The code to evaluate the submissions will be released in the challenge webpage.

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

Participants will be highly encouraged to post their code openly in a public repository.

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.

No conflicts of interest are declared. Test labels will only be handled by the organizers and the clinical collaborators, and will be released publicly after the workshop.
The challenge will be supported by the H2020 euCanSHare project (www.eucanshare.eu), which is building a multi-centre big data platform for cardiovascular personalised medicine research.

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

Prognosis, Decision support, Diagnosis, CAD, Research.

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 will be patients scanned in different centres with different vendors and having different cardiomyopathies, focusing in the pathological right ventricles.

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 patients with five right ventricle pathologies (Dilated Right Ventricle, Tricuspidal Insufficiency, Arrhythmogenesis, Tetralogy of Fallot and Interatrial Comunication), 2 left ventricle pathologies (Dilated Left Ventricle, Hypertrophic Cardiomyopathy) and a control group that were scanned in four clinical centres using three different magnetic resonance vendors (Siemens, General Electric and Philips).

Imaging modality(ies)

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

Magnetic resonance imaging (MRI)

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 will be stratified by pathology using anonymized indices.

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

Gender and age will be provided for each 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.

Short-axis view of cardiac magnetic resonance imaging.
Horizontal long-axis view of cardiac magnetic resonance imaging (4 chambers).

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 target will be the cardiac right ventricle anatomic structure for both short-axis and long-axis views. The participants may combine both views or use each view indepentdently.

Additionally, the participants will be provided with the left ventricle and myocardium labels for the training set.

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)).

Robustness, Accuracy.

Additional points: Find RV segmentation algorithms for cardiac MR images that can generalize well across pathologies and centres.

Find RV segmentation algorithms for cardiac MR images that can use both long- and short-axis images to obtain more robust segmentation results.

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).

Four different scanners from three different vendors (Siemens, General Electric and Philips) were used during the acquisition. The field strength is 1.5T for all of them.

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

Standardized cardiovascular magnetic resonance (CMR) acquisition protocols are applied in order to evaluate biventricular function.

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.

Since the data will be publicly available and for anonymization reasons, the exact centre and scanner will not be matched. No information about the study centre will be provided.

250 studies were reused from from last year's M&Ms challenge:
- 91 studies were acquired at Hospital Vall d'Hebron (Spain).
- 70 studies were acquired at Clinica Sagrada Familia (Spain).
- 30 studies were acquired at Hospital Universitari Dexeus (Spain).
- 59 studies were acquired at Universitätsklinikum Hamburg-Eppendorf (Germany).

200 new studies with pathological right ventricle were acquired for the M&Ms-2 challenge:

- 116 studies were acquired at Hospital Vall d'Hebron (Spain).
- 45 studies were acquired at Clinica Sagrada Familia (Spain).
- 39 studies were acquired at Hospital Universitari Dexeus (Spain).

Overall, a total of 450 subjects with the following distribution per center:
- 45% of the studies were acquired at Hospital Vall d'Hebron (Spain).
- 25% of the studies were acquired at Clinica Sagrada Familia (Spain).
- 15% of the studies were acquired at Hospital Universitari Dexeus (Spain).
- 15% of the studies were acquired at Universitätsklinikum Hamburg-Eppendorf (Germany).

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).

Each centre will have its own expert clinician for their acquisitions, with experiences ranging from 2 to more than 10 years.

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 represents a patient study, that is, the set of two-dimensional short-axis images and four chambers long-axis that cover the heart from the apex to the base. All the cases have right ventricle, left ventricle and myocardial segmentations for both long-axis and short-axis views for the end diastolic and end sistolic frames extracted from both short-axis and long-axis cine CMR acquisitions..

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

Training: 250
Validation: 40
Testing: 160

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

The minimum amount of validation cases are selected keeping the training set large enough while valildaton can provide a good sense of generalization in the developement stage.
The test set is large enough to generalise and to provide a fair comparison between participants in the final stage.

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.

There will be cases from four different centres in the training , validation, and test sets with the distribution described before.
There will be cases of 3 rigth ventricle pathologies (Dilated Right Ventricle, Tricuspid Insufficiency and Tetralogy of Fallot, Dilated Left Ventricle), 2 left ventricle pathologies (Dilated Left Ventricle, Hypertrophic Cardiomyopathy) and a control group in the training set .

There will be cases of 5 rigth ventricle pathologies (Dilated Right Ventricle, Tricuspid Insufficiency, Interatrial Comunication, Arrhytmogenesis and Tetralogy of Fallot), 2 left ventricle pathologies (Dilated Left Ventricle, Hypertrophic Cardiomyopathy) and a control group in the testing and validation sets.

Keeping 2 right ventricle pathologies out of the training set will allow to analyze the generalisation of automatic right ventricle segmentation algorithms to unseen pathologies. The following pathology distribution will allow to study the performance of automatic CMR segmentation algorithms across diseases, focusing on the most conflictive region of the three main cardiac structures up to date, the right ventricle:

Dilated Right Ventricle
Total size: 40 - Training size: 20 - Validation size: 5 - Testing size: 15

Tetralogy of Fallot
Total size: 40 - Training size: 20 - Validation size: 5 - Testing size: 15

Interatrial Comunication
Total size: 40 - Training size: 0 - Validation size: 5 - Testing size: 35

Tricuspidal Regurgitation
Total size: 40 - Training size: 20 - Validation size: 5 - Testing size: 15

Congenital Arrhythmogenesis
Total size: 40 - Training size: 0 - Validation size: 5 - Testing size: 35

Dilated Left Ventricle
Total size: 90 - Training size: 70 - Validation size: 5 - Testing size: 15

Hypetrophic Cardiomiopathy
Total size: 70 - Training size: 50 - Validation size: 5 - Testing size: 15

Normal:
Total size: 90 - Training size: 70 - Validation size: 5 - Testing size: 15

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 case will be segmented by an experienced clinician from each institution. This manual annotation will be used as segmentation ground truth.

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.

Annotators were asked to annotate the images using the Circle cvi42 software.

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.

Each centre will have its own expert clinician for their annotations, with experiences ranging from 2 to more than 10 years.

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.

In order to reduce the inter-observer and inter-centre variability in the contours, in particular at the apical and basal regions, a detailed revision of the provided segmentations was performed by four researchers in pairs. They applied the same SOP across all CMR datasets to obtain the final ground truth.

To generate consistent annotations for the research community, we chose to apply a SOP as follows:

a) The LV and RV cavities must be completely covered, including the papillary muscles.

b) No interpolation of the MYO boundaries must be performed at the basal region, since the estimation of partial regions in the myocardium tend to have a minimum impact in clinical practice.

c) The RV must have a larger surface at the ED time-frame compared to ES.

d) The RV short-axis does not include the pulmonary artery nor the tricuspid valve. This is achieved by verifying the intersection of the SA contours on LA images, thus removing incorrect overlaps.

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.

All cases will be transformed from dicom to nifti format and no further pre-processing will be applied to images.

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 most relevant source of error is the delineation of the right ventricle blood pool in short-axis. This task often involves a secondary axial acquisition where the definition of the contours is of higher quality, and particularly when a right ventricle pathology is involved. Finally, the partial delineation of the slice where the tricuspid valve appears could bring disagreement between the annotators. The final annotations are performed with the consensus of simultaneous annotators to minimize these errors as expressed before.

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

Additional sources of error are not expected.

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)

In order to assess the quality of the automatically segmented masks P with respect to the ground truth G, two measures were proposed, namely:

Dice similarity coefficient (DSC):
DSC(P, G) = 2|P intersection G| / (|P| + |G|)

Hausdorff distance (HD):
HD(P, G) = max { sup(p c P) d(p, G), sup(g c G) d(g, P) }

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

DSC mesures the degree of overlapping between two volumes and it is widely accepted as a measure of similarity in biomedical segmentation applications.

HD measures the largest disagreement between the volumes and it is useful for identifying small outliers and penalizing contourings rather than volumes.

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.

HD measures will be normalizedin the range [0,1] for all the participants.
Mean and standard deviation for both DSC and HD metrics will be computed for all the test subjects in both end

diastolic and end sistolic pahses for both short-axis volumes and long-axis images.

Short-axis and long-axis metrics will be averaged with 0.75 and 0.25 weights, respectively in order to balance the contribution of overall contour lengths for both axis.

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

Missing results in the submission will get a zero for Dice and the equivalent worst value for Hausdorff distance (i.e. when normalized in the range [0,1], being 1 the worst and 0 the best value, missing results will get a 1).

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

The ranking is computed in the aforementioned way to motivate participants to force generalization of their models and accurate segmentation of minority vendors.

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.

Welch's t-tests will be performed to assess significance in the difference of mean segmentation accuracy between scanners and diseases.

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

Welch's t-test is chosen because it allows to test two populations with possibly different variances and unequal number of samples.

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 intra-class correlation coefficient will be computed for all the test set to identify those with larger disagreement when segmented with automatic methods.

The added value of long axis information when segmenting short-axis will be analyzed, focusing on the basal region.

Generalisation across clinical centers will be analyzed.
Generalization across pathologies will be analyzed, including unseen right ventricle diseases.

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] Bernard, Olivier, et al. "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?." IEEE transactions on medical imaging 37.11 (2018): 2514-2525.
[2] Tran, Phi Vu. "A fully convolutional neural network for cardiac segmentation in short-axis MRI." arXiv preprint arXiv:1604.00494 (2016).
[3] Isensee, Fabian, et al. "Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features." International workshop on statistical atlases and computational models of the heart. Springer, Cham, 2017.
[4] Zotti, Clément, et al. "GridNet with automatic shape prior registration for automatic MRI cardiac segmentation." International Workshop on Statistical Atlases and Computational Models of the Heart. Springer, Cham, 2017.
[5] Petitjean, Caroline, et al. "Right ventricle segmentation from cardiac MRI: a collation study." Medical image analysis 19.1 (2015): 187-202.
[6] Gillies, Robert J., Paul E. Kinahan, and Hedvig Hricak. "Radiomics: images are more than pictures, they are data." Radiology 278.2 (2016): 563-577.
[7] Zhuang, Xiahai et al. “Evaluation of algorithms for Multi-Modality Whole Heart Segmentation: An open-access grand challenge.” Medical Image Analysis (2019).