Brain MRI reconstruction challenge with realistic noise: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Brain MRI reconstruction challenge with realistic noise

Challenge acronym

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

RealNoiseMRI

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, there is a growing focus on the application of fast magnetic resonance imaging (MRI) based on prior knowledge. In the 1980s and 2000s the community used either purely mathematical models such as the partial Fourier transform or solutions derived through advanced engineering such as parallel imaging to speed up MRI acquisition. Since the mid-2000's, compressed sensing and artificial intelligence have been employed to speed up MRI acquisition. These newer methods rely on under sampling the data acquired in Fourier (aka k-) space and then interpolating or augmenting k-space data based on training data content.
One of the underlying problems for the development of fast imaging techniques, that just as in e.g. [1], it is common to use a fully sampled image as ground truth and then under sample it in k-space in order to simulate under sampled data. The problem with this approach is that in cases were the under sampled data is corrupted, through e.g. motion, this under sampling is unrealistic. This could easily happen in a clinical setting. Hence, the robustness of the new fast MRI reconstruction methods is often not evaluated in a realistic setting.
We propose a challenge that aims at checking the robustness of fast MRI reconstruction methods by providing image and k-space datasets that consist of ground truth as well as motion degraded scans. We aim to closely mimic the ongoing challenge organized by Facebook at NeurIPS: https://fastmri.org/, but provide motion degraded data to realistically test robustness. While the fastMRI challenge provides participants with a large dataset, the way the under-sampled k-space data is constructed is artificial and not realistic, since patient motion corrupts the k-space differently than removing certain lines of k-space.
Our setup would be the following: For 25 healthy volunteers, we are collecting brain imaging data with standard cerebral protocols including axial, 2D-encoded T1- and T2-weighted sequences (Short-TI-Inversion Recovery and Turbo Spin Echo). We have already scanned 14 of the 25 volunteers and the remaining 11 scans are scheduled within the next 2-3 months. In our study, we are assessing the utility of motion correction sequences and hence, we acquire data where the participants lie still and where they move in a pre-described fashion. We can therefore provide challenge participants with the ground truth (still) as well as degraded (due to motion) T1- and T2-weighted images and the corresponding k-space data. The participants are expected to train their models on the

challenge data provided by the fastMRI challenge. We will release a set of 5 k-space data and corresponding ground truth images for optimization and validation of the pre-trained models using fastMRI data. For the remaining 20 volunteers, we will withhold the ground truth images and only distribute the degraded k-space data as a test set.
The goal of our challenge for participants would then be to send in estimates of the ground truth images. We can evaluate the contributions based on the available ground truth data using the structural similarity measure (SSIM) as quantitative performance metric together with a visual evaluation through radiologists.
[1] Duan J, Schlemper J, Qin C, Ouyang C, Bai W, Biffi C, Bello G, Statton B, O'Regan DP, Rueckert D. VS-Net: Variable splitting network for accelerated parallel MRI reconstruction. In International Conference on Medical Image Computing and Computer-Assisted Intervention 2019 Oct 13 (pp. 713-722). Springer, Cham.

Challenge keywords

List the primary keywords that characterize the challenge.

MRI, reconstruction, noise, motion artefacts, compressed sensing, deep learning

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.

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.

25 to 50. At the public leaderboard of the fastMRI challenge for brain data, 22 successful models are listed. In the Machine Learning for Medical Image Reconstruction (MLMIR) workshop proceedings 2019 25 papers and for 2020 15 papers were published, which shows a high interest of the community in the this topic.

Publication and future plans

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

The aim is to summarize the challenge results in an article geared at a high impact journal, such as NeuroImage or Magnetic Resonance in Medicine.

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.

Data would be stored locally on the challenge organizers' servers and made available for download from there. The actual computations would need to be carried out in participants' local computational environment. The metrics for the reconstructed test images will be calculated by the organizers. The challenge would then on-site just reveal the outcome of the submissions and invite the winning teams to present their algorithms.

TASK: Reconstruction of motion corrupted T1 weighted MRI data

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.

see general abstract description, details will be very similar to the challenge presented in https://fastmri.org/

Keywords

List the primary keywords that characterize the task.

MRI, T1-weighted, motion corrupted

ORGANIZATION

Organizers

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

The MoCo study team at Rigshospitalet, Copenhagen, Denmark
https://sites.google.com/view/melanieganz/research-projects/imaging-children-without-anesthesia

b) Provide information on the primary contact person.

Melanie Ganz, mganz@nru.dk, Neurobiology Research Unit, Rigshospitalet, Copenhagen, Denmark

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.

grand-challenge.org

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

not available yet

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.

Publicly available data is 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 first placed team will receive an award. We are in negotiations with Siemens Healthineers about a free spot in a Siemens neuro imaging training course (appr. value of 600€).

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 Top 5 performing methods will be announced publicly. Participating teams can choose whether the 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).

There will be an embargo time of 12 months after the challenge submission date for the challenge organizers to publish a challenge paper first. Up to 3 members of the 10 best teams will qualify as co-author on the challenge paper, given that they fulfil the Vancouver Group Recommendations individually, which are:
1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
2. Drafting the work or revising it critically for important intellectual content; AND
3. Final approval of the version to be published; AND
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
Afterwards the participating teams may publish their own results separately.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

Submission of reconstructed images of the test data via web platform hosted on NRU server, no link to submission instructions available yet, but will be heavily like this challenge: https://fastmri.org/submission_guidelines

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.

The participating teams will have access to training data and we will allow submission of multiple results, but even if the participants submit multiple times, the test results will be hidden until after the submission deadline.

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 period: 1. April – 31. July 2021
Submission deadline: 31. Juli 2021

1. April: Challenge website with instructions opens. Applicants can apply for fastMRI training data.
1. May 2021: The challenge dataset is released for optimization / validation of the pre-trained models.
27. July 2021: The degraded k-space data sets are released.
27. July – 31. July 2021: The submission of the reconstructed test images is open.
August 2021: Calculation of metrics and visual image quality scoring of the submitted test images.
1. September 2021: The teams of the top-ranked methods are contacted and invited to present at the MICCAI challenge workshop.
27. September – 01. October 2021: MICCAI challenge takes place.
The competition organizers reserve the right to update the timelines.

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 protocol is approved by the Danish national committee on health research ethics and ratified by the Danish Data protection agency.

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.

For assessing image quality of the degraded images (reconstructed by participants), the still "clean" images will be segmented and cortical boundary delineations will be derived via FreeSurfer by the organizers. This then enables the use of FreeSurfer's registration tools for transferring the still image into the space of the degraded image (submitted by participants). A brain mask of the still scan will be extracted by Freesurfer and then applied to both scans. The two aligned images can then be evaluated on Structural Similarity Measure (SSIM), a popular measure developed in 2001 by Zhou Wang and Al Bovik to evaluate quality of broadcasted images. SSIM attempts to quantify changes in structural information of an image. The metric is calculated using the ssim function in the challenge organizers github repository. The SSIM scores of the submitted entries will be made available on the challenge website after the challenge submission period ends.

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

We will encourage the participating teams to make their code available on Github, but we will not require it.

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.

By entering, except where prohibited by law, each entrant grants NRU, Rigshospitalet, Denmark the right and permission to use, edit, modify, copy, reproduce, and distribute entrants' submitted images and image descriptions (abstracts) in connection with the challenge.
Data ownership (aka access to data) is protected by Danish data protection law and lies with the Capital region of Denmark. Only the people listed on the data collection agreement which consists of the people running this study will have access to the test data until after the MICCAI challenge has taken place. Afterwards the anonymized test data will be publicly released on the challenge website.
Therefore, Employees of NRU, Rigshospitalet are not eligible to participate in the challenge.
Furthermore, due to the award being offered by Siemens Healthineers, employees of Siemens Healthineers are not eligible to participate in 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.

Additional points: Magnetic resonance imaging, reconstruction

Task category(ies)

State the task category(ies).

Examples:

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

Reconstruction.

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.

Clinical patients that undergo MRI examinations

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

Healthy control subjects that undergo MRI examinations and that are instructed first to lie still and then to degrade the scan quality by following predescribed motion patterns

Imaging modality(ies)

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

Magnetic resonance imaging, axial T1-weighted Short-TI Inversion Recovery, no contrast. Flip Angle 150°. TR/TE/TI 3500/8.6/1387 ms. Voxel size 0.9x0.9x5.0 mm. GRAPPA (accel. Factor PE: 2). Partial Fourier Off.

Context information

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

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

image raw data (k-space) as well as reconstructed images (in the case of validation dataset) will be provided

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

no additional information will be given

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

brain shown in magnetic resonance imaging (MRI) data

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.

brain shown in magnetic resonance imaging (MRI) 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)).

Submitted reconstructions are evaluated on Structural Similarity Measure (SSIM). In order to reassure clinical relevancy, the top five submissions from each task with the highest median rank profile on the challenge dataset will be evaluated by a panel of radiologists (single blind assessment) who will ultimately select one winner in each task. This step of qualitative evaluation will be based on the following criteria: contrast to noise, artifacts, sharpness, and diagnostic confidence.

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).

3T Siemens Magnetom Prisma MRI scanner

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

The subject will undergo a standard 2D-encoded, T1-weighted imaging sequence (axial).

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.

Rigshospital, Copenhagen, Denmark

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).

Healthy adult controls

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.

Validation and test cases both represent a MRI image of a human brain. Validation cases consist of a "clean" image as well as a degraded image with the accompanying k-space data, while the test cases only consist of the defaced k-space data (degraded due to motion).

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

No training subjects will be provided, but fastMRI challenge provides 409 axial T1-weighted scans at 3T (https://fastmri.med.nyu.edu).
5 validation subjects for optimization and validation of the pre-trained models and 20 test subjects (without ground truth image) for final evaluation will be provided.
2: No training subjects will be provided, but fastMRI challenge provides 2524 axial T2-weighted scans at 3T (https://fastmri.med.nyu.edu).
5 validation subjects for optimization and validation of the pre-trained models and 20 test subjects (without ground truth image) for final evaluation will be provided.

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

The aim of this challenge is to evaluate how robust fast MRI reconstruction methods are to realistic noise / artefacts caused by real motion with a unique dataset from motion experiments, where the participants are instructed to lie still for one scan and to perform a predefined motion for the second. Therefore, we only provide a validation and test data set and expect the models to be pre-trained on the fastMRI challenge data set. Our study is set up to acquire 25 healthy adult controls, so we will provide 5 of those for validation and keep 20 subjects for final evaluation with a significant sample size.

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.

All subjects are healthy adult controls and we will aim for an age and gender balance between the training and test data.

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.

No reference annotation required.

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.

not applicable.

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.

not applicable.

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.

not applicable.

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 data preprocessing aka anonymization procedure will follow exactly the same steps as the fastMRI challenge. In order to provide anonymized k-space data it will be provided in the vendor neutral HDF5 format. Furthermore, in order to remove facial features, parts of k-space will need to be removed. We only use axial 2D-encoded images and the slices smaller/equal 5mm below the orbital rim will be replaced with zeroes. All processed k-spaces will then be reconstructed to images in DICOM format, loaded into a picture archival communication system (PACS) and visually checked by certified MR technologists to confirm exclusion of identifying facial features.
The ground truth images will be provided in BIDS MR format. This is a completely standard procedure for any public MRI dataset [2] (e.g. all MRI datasets on OpenNeuro have been treated like this.

[2] Gorgolewski KJ et. al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data. 2016;3: 160044. PMID: 27326542.

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.

not applicable.

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

Since the data will come directly from the scanner, there are of course noise sources related to the MRI acquisition such as e.g. scanner drift. But those are always present in real clinical data and to a little extent in the proposed short structural scans.
For each subject (train and test) we are acquiring a still as well as motion degraded scan. The validation still scans will be distributed in order to enable the assessment of validation / optimization data. Assessment of the test set will happen by running FreeSurfer on the still scans and then using a FreeSurfers registration tools to register them to the motion degraded scans. A brain mask of the still scan will be extracted by Freesurfer and then applied to both scans. This way, SSIM can be evaluated. While there are noise sources in this registration process, Freesurfer offers well established tools for multimodal registration and we will register all submissions based on exactly the same still image cortical delineation by FreeSurfer.

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)

Structural Similarity Measure (SSIM), compare eq. 9 in Zbontar, Jure et al. (2018). “fastMRI: An Open Dataset and Benchmarks for Accelerated MRI”. In: pp. 1–35. arXiv: 1811.08839.

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

SSIM is commonly used for the evaluation of reconstruction data.

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.

To identify the best performing algorithm across the test set, one proceeds in the following way: (a) For test subject k, obtain SSIM score for all algorithms. (b) For test subject k, rank the algorithms from optimal (SSIM=1) with highest rank indicating the best performance. (c) Obtain algorithm ranking for all 10 test subjects, then take the median rank of each algorithm, across all test subjects. This produces the median-rank profile.
Based on the median rank profile the top five submissions from each task will be evaluated by a panel of

radiologists who will ultimately select one winner in each task. The single blind qualitative evaluation will be based on the following criteria: contrast to noise, artifacts, sharpness, and diagnostic confidence. Based on these criteria submissions will be ranked and scores averaged to determine the winners.

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

Submission with missing results on test cases will be punished by setting the SSIM value for these test cases to zero.

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

In the assessment of different algorithms we are proposing to use the median rank. This is a common way of ranking e.g. preprocessing pipelines in the fMRI and PET literature.
Churchill, Nathan W., et al. "Optimizing preprocessing and analysis pipelines for single subject fMRI. I. Standard temporal motion and physiological noise correction methods." Human brain mapping 33.3 (2012): 609-627.
Nørgaard, Martin, et al. "Optimization of preprocessing strategies in Positron Emission Tomography (PET) neuroimaging: A [11C] DASB PET study." NeuroImage 199 (2019): 466-479.

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.

Since we will only have 20 test subjects, as already mentioned in the case of missing data the missing subject will enter the ranking with a SSIM of zero.
Regarding the comparison of different rankings we will identify the algorithm that maximizes the median rank across all test subjects. We will also report the variance of the median rank for each algorithm. It will be assessed if there is a significant difference of the top performing algorithm compared to the median ranking of the other algorithms via a Friedman test.
All statistical analysis will be carried out in Python and made available in the GitHub repository.

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

As stated above in the assessment of different algorithm or processing pipeline performance it is common in neuroimaging to use the median rank profile for assessment. Furthermore, there are established statistical tests such as the Friedmann test that then can be used to test which algorithms can be considered to be significantly different from the top performing algorithm. This will be used to assess if there is any significant difference between the top algorithm and the rest or not.

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.

We will quantitatively evaluate common problems/biases of the submitted methods. We will compare the SSIM values used for ranking with values for comparable image metrics like Peak Signal-to-Noise ration or Tenengrad, as well as the visual assessments by the radiologists.

TASK: Reconstruction of motion corrupted T2 weighted MRI data

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.

see general abstract description, details will be very similar to the challenge presented in https://fastmri.org/

Keywords

List the primary keywords that characterize the task.

MRI, T2-weighted, motion corrupted

ORGANIZATION

Organizers

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

The MoCo study team at Rigshospitalet, Copenhagen, Denmark
https://sites.google.com/view/melanieganz/research-projects/imaging-children-without-anesthesia

b) Provide information on the primary contact person.

Melanie Ganz, mganz@nru.dk, Neurobiology Research Unit, Rigshospitalet, Copenhagen, Denmark

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.

grand-challenge.org

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

not available yet

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.

Publicly available data is 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 first placed team will receive an award. We are in negotiations with Siemens Healthineers about a free spot in a Siemens neuro imaging training course (appr. value of 600€).

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 Top 5 performing methods will be announced publicly. Participating teams can choose whether the 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).

There will be an embargo time of 12 months after the challenge submission date for the challenge organizers to publish a challenge paper first. Up to 3 members of the 10 best teams will qualify as co-author on the challenge paper, given that they fulfil the Vancouver Group Recommendations individually, which are:
1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
2. Drafting the work or revising it critically for important intellectual content; AND
3. Final approval of the version to be published; AND
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
Afterwards the participating teams may publish their own results separately.

Submission method

a) Describe the method used for result submission. Preferably, provide a link to the submission instructions.

Examples:

- Docker container on the Synapse platform. Link to submission instructions: <URL>
- Algorithm output was sent to organizers via e-mail. Submission instructions were sent by e-mail.

Submission of reconstructed images of the test data via web platform hosted on NRU server, no link to submission instructions available yet, but will be heavily like this challenge: https://fastmri.org/submission_guidelines

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.

The participating teams will have access to training data and we will allow submission of multiple results, but even if the participants submit multiple times, the test results will be hidden until after the submission deadline.

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 period: 1. April – 31. July 2021
Submission deadline: 31. Juli 2021

1. April: Challenge website with instructions opens. Applicants can apply for fastMRI training data.
1. May 2021: The challenge dataset is released for optimization / validation of the pre-trained models.
27. July 2021: The degraded k-space data sets are released.
27. July – 31. July 2021: The submission of the reconstructed test images is open.
August 2021: Calculation of metrics and visual image quality scoring of the submitted test images.
1. September 2021: The teams of the top-ranked methods are contacted and invited to present at the MICCAI challenge workshop.
27. September – 01. October 2021: MICCAI challenge takes place.
The competition organizers reserve the right to update the timelines.

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 protocol is approved by the Danish national committee on health research ethics and ratified by the Danish Data protection agency.

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.

For assessing image quality of the degraded images (reconstructed by participants), the still "clean" images will be segmented and cortical boundary delineations will be derived via FreeSurfer by the organizers. This then enables the use of FreeSurfer's registration tools for transferring the still image into the space of the degraded image (submitted by participants). A brain mask of the still scan will be extracted by Freesurfer and then applied to both scans. The two aligned images can then be evaluated on Structural Similarity Measure (SSIM), a popular measure developed in 2001 by Zhou Wang and Al Bovik to evaluate quality of broadcasted images. SSIM attempts to quantify changes in structural information of an image. The metric is calculated using the ssim function in the challenge organizers github repository. The SSIM scores of the submitted entries will be made available on the challenge website after the challenge submission period ends.

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

We will encourage the participating teams to make their code available on Github, but we will not require it.

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.

By entering, except where prohibited by law, each entrant grants NRU, Rigshospitalet, Denmark the right and permission to use, edit, modify, copy, reproduce, and distribute entrants' submitted images and image descriptions (abstracts) in connection with the challenge.
Data ownership (aka access to data) is protected by Danish data protection law and lies with the Capital region of Denmark. Only the people listed on the data collection agreement which consists of the people running this study will have access to the test data until after the MICCAI challenge has taken place. Afterwards the anonymized test data will be publicly released on the challenge website.
Therefore, Employees of NRU, Rigshospitalet are not eligible to participate in the challenge.
Furthermore, due to the award being offered by Siemens Healthineers, employees of Siemens Healthineers are not eligible to participate in 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.

Additional points: Magnetic resonance imaging, reconstruction

Task category(ies)

State the task category(ies).

Examples:

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

Reconstruction.

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.

Clinical patients that undergo MRI examinations

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

Healthy control subjects that undergo MRI examinations and that are instructed first to lie still and then to degrade the scan quality by following predescribed motion patterns

Imaging modality(ies)

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

Magnetic resonance imaging, axial T2-weighted Turbo Spin Echo, no contrast. Flip Angle 150°. TR/TE 4400/117 ms. Voxel size 0.4x0.4x5.0 mm. GRAPPA (accel. Factor PE: 2). Partial Fourier Off.

Context information

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

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

image raw data (k-space) as well as reconstructed images (in the case of validation dataset) will be provided

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

no additional information will be given

Target entity(ies)

a) Describe the data origin, i.e. the region(s)/part(s) of subject(s)/object(s) from whom/which the image data would be acquired in the final biomedical application (e.g. brain shown in computed tomography (CT) data, abdomen shown in laparoscopic video data, operating room shown in video data, thorax shown in fluoroscopy video). If necessary, differentiate between target and challenge cohort.

brain shown in magnetic resonance imaging (MRI) data

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.

brain shown in magnetic resonance imaging (MRI) 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)).

Submitted reconstructions are evaluated on Structural Similarity Measure (SSIM). In order to reassure clinical relevancy, the top five submissions from each task with the highest median rank profile on the challenge dataset will be evaluated by a panel of radiologists (single blind assessment) who will ultimately select one winner in each task. This step of qualitative evaluation will be based on the following criteria: contrast to noise, artifacts, sharpness, and diagnostic confidence.

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).

3T Siemens Magnetom Prisma MRI scanner

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

The subject will undergo a standard 2D-encoded, T2-weighted imaging sequence (axial).

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.

Rigshospital, Copenhagen, Denmark

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).

Healthy adult controls

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.

Validation and test cases both represent a MRI image of a human brain. Validation cases consist of a "clean" image as well as a degraded image with the accompanying k-space data, while the test cases only consist of the defaced k-space data (degraded due to motion).

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

No training subjects will be provided, but fastMRI challenge provides 409 axial T1-weighted scans at 3T (https://fastmri.med.nyu.edu).
5 validation subjects for optimization and validation of the pre-trained models and 20 test subjects (without ground truth image) for final evaluation will be provided.
2: No training subjects will be provided, but fastMRI challenge provides 2524 axial T2-weighted scans at 3T (https://fastmri.med.nyu.edu).
5 validation subjects for optimization and validation of the pre-trained models and 20 test subjects (without ground truth image) for final evaluation will be provided.

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

The aim of this challenge is to evaluate how robust fast MRI reconstruction methods are to realistic noise / artefacts caused by real motion with a unique dataset from motion experiments, where the participants are instructed to lie still for one scan and to perform a predefined motion for the second. Therefore, we only provide a validation and test data set and expect the models to be pre-trained on the fastMRI challenge data set. Our study is set up to acquire 25 healthy adult controls, so we will provide 5 of those for validation and keep 20 subjects for final evaluation with a significant sample size.

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.

All subjects are healthy adult controls and we will aim for an age and gender balance between the training and test data.

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.

No reference annotation required.

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.

not applicable.

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.

not applicable.

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.

not applicable.

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 data preprocessing aka anonymization procedure will follow exactly the same steps as the fastMRI challenge. In order to provide anonymized k-space data it will be provided in the vendor neutral HDF5 format. Furthermore, in order to remove facial features, parts of k-space will need to be removed. We only use axial 2D-encoded images and the slices smaller/equal 5mm below the orbital rim will be replaced with zeroes. All processed k-spaces will then be reconstructed to images in DICOM format, loaded into a picture archival communication system (PACS) and visually checked by certified MR technologists to confirm exclusion of identifying facial features.
The ground truth images will be provided in BIDS MR format. This is a completely standard procedure for any public MRI dataset [2] (e.g. all MRI datasets on OpenNeuro have been treated like this.

[2] Gorgolewski KJ et. al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data. 2016;3: 160044. PMID: 27326542.

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.

not applicable.

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

Since the data will come directly from the scanner, there are of course noise sources related to the MRI acquisition such as e.g. scanner drift. But those are always present in real clinical data and to a little extent in the proposed short structural scans.
For each subject (train and test) we are acquiring a still as well as motion degraded scan. The validation still scans will be distributed in order to enable the assessment of validation / optimization data. Assessment of the test set will happen by running FreeSurfer on the still scans and then using a FreeSurfers registration tools to register them to the motion degraded scans. A brain mask of the still scan will be extracted by Freesurfer and then applied to both scans. This way, SSIM can be evaluated. While there are noise sources in this registration process, Freesurfer offers well established tools for multimodal registration and we will register all submissions based on exactly the same still image cortical delineation by FreeSurfer.

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)

Structural Similarity Measure (SSIM), compare eq. 9 in Zbontar, Jure et al. (2018). “fastMRI: An Open Dataset and Benchmarks for Accelerated MRI”. In: pp. 1–35. arXiv: 1811.08839.

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

SSIM is commonly used for the evaluation of reconstruction data.

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.

To identify the best performing algorithm across the test set, one proceeds in the following way: (a) For test subject k, obtain SSIM score for all algorithms. (b) For test subject k, rank the algorithms from optimal (SSIM=1) with highest rank indicating the best performance. (c) Obtain algorithm ranking for all 10 test subjects, then take the median rank of each algorithm, across all test subjects. This produces the median-rank profile.
Based on the median rank profile the top five submissions from each task will be evaluated by a panel of

radiologists who will ultimately select one winner in each task. The single blind qualitative evaluation will be based on the following criteria: contrast to noise, artifacts, sharpness, and diagnostic confidence. Based on these criteria submissions will be ranked and scores averaged to determine the winners.

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

Submission with missing results on test cases will be punished by setting the SSIM value for these test cases to zero.

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

In the assessment of different algorithms we are proposing to use the median rank. This is a common way of ranking e.g. preprocessing pipelines in the fMRI and PET literature.
Churchill, Nathan W., et al. "Optimizing preprocessing and analysis pipelines for single subject fMRI. I. Standard temporal motion and physiological noise correction methods." Human brain mapping 33.3 (2012): 609-627.
Nørgaard, Martin, et al. "Optimization of preprocessing strategies in Positron Emission Tomography (PET) neuroimaging: A [11C] DASB PET study." NeuroImage 199 (2019): 466-479.

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.

Since we will only have 20 test subjects, as already mentioned in the case of missing data the missing subject will enter the ranking with a SSIM of zero.
Regarding the comparison of different rankings we will identify the algorithm that maximizes the median rank across all test subjects. We will also report the variance of the median rank for each algorithm. It will be assessed if there is a significant difference of the top performing algorithm compared to the median ranking of the other algorithms via a Friedman test.
All statistical analysis will be carried out in Python and made available in the GitHub repository.

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

As stated above in the assessment of different algorithm or processing pipeline performance it is common in neuroimaging to use the median rank profile for assessment. Furthermore, there are established statistical tests such as the Friedmann test that then can be used to test which algorithms can be considered to be significantly different from the top performing algorithm. This will be used to assess if there is any significant difference between the top algorithm and the rest or not.

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.

We will quantitatively evaluate common problems/biases of the submitted methods. We will compare the SSIM values used for ranking with values for comparable image metrics like Peak Signal-to-Noise ration or Tenengrad, as well as the visual assessments by the radiologists.

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.

none

Further comments

Further comments from the organizers.

We would like to closely mimic the ongoing challenge organized by Facebook: https://fastmri.org/, but motion degrade the data to realistically test robustness. This is also why we want to make our assessment criteria comparable to the fastMRI challenge (use of SSIM). Additionally, compared to the fastMRI challenge, we have detailed a common and statistically sound way of aggregating different performance measures (median rank profile) and kept the clinical relevance of this whole problem in mind (adding the radiologist expert panel).
Furthermore, we encourage the participants to use the publicly available training data from the fastMRI challenge to pre-train their models (https://fastmri.med.nyu.edu). We are aware of the fact that the participants need to apply for access, but we have scheduled time for the application, since our website with challenge instructions opens at 1st of April and the validation/optimization training set is first published 1st of May (compare the challenge time schedule below).
The organizers might also be concerned about the legal requirements for sharing data. The anonymization will follow exactly the same steps as the fastMRI challenge.
In order to provide anonymized k-space data it will be provided in the vendor neutral HDF5 format. Furthermore, in order to remove facial features, parts of k-space will need to be removed. All processed k-spaces will then be reconstructed to images in DICOM format, loaded into a picture archival communication system (PACS) and visually checked by certified MR technologists to confirm exclusion of identifying facial features.
The ground truth images will be provided in BIDS MR format. This is a completely standard procedure for any public MRI dataset [2] (e.g. all MRI datasets on OpenNeuro have been treated like this).
[2] Gorgolewski KJ et. al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data. 2016;3: 160044. PMID: 27326542.