Cross-Modality Domain Adaptation for Medical Image Segmentation: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Cross-Modality Domain Adaptation for Medical Image Segmentation

Challenge acronym

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

crossMoDA

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.

Domain Adaptation (DA) has recently raised strong interests in the medical imaging community. By encouraging algorithms to be robust to unseen situations or different input data domains, Domain Adaptation improves the applicability of machine learning approaches to various clinical settings. While a large variety of DA techniques has been proposed for image segmentation, most of these techniques have been validated either on private datasets [4,5] or on small publicly available datasets [6,7,8,9]. Moreover, these datasets mostly address single-class problems.

To tackle these limitations, the crossMoDA challenge introduces the first large and multi-class dataset for unsupervised cross-modality Domain Adaptation. The goal of the challenge is to segment two key brain structures involved in the follow-up and treatment planning of vestibular schwannoma (VS): the tumour and the cochlea. Specifically, the segmentation of the tumour and the surrounding organs at risk, such as the cochlea, is required for radiosurgery, a common VS treatment. Moreover, tumour volume measurement has also been shown to be the most accurate measurements for the evaluation of VS growth [3]. While contrast-enhanced T1 (ceT1) Magnetic Resonance Imaging (MRI) scans are commonly used for VS segmentation, recent work [1,2] has demonstrated that high-resolution T2 (hrT2) imaging could be a reliable, safer, and lower-cost alternative to ceT1. For these reasons, we propose an unsupervised cross-modality challenge (from ceT1 to hrT2) that aims to automatically perform VS and cochlea segmentation on hrT2 scans. The training source and target sets are respectively unpaired annotated ceT1 and non-annotated hrT2 scans. This challenge will be the first medical segmentation benchmark of unsupervised DA techniques and promote the development of new unsupervised domain adaptation solutions for medical image segmentation. It will also contribute to the development of new algorithms for the follow-up and treatment planning of VS using hrT2 scans only.

Challenge keywords

List the primary keywords that characterize the challenge.

Domain Adaptation, Segmentation, Cross-Modality

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.

In discussion with Domain Adaptation and Representation Transfer (DART).

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.

We are hoping for 20 participants.

We identified 4 medical imaging groups involved in cross-modality domain adaptation and we expect their participation to the challenge (Department of Computer Science and Engineering, The Chinese University of Hong Kong; BioMedIA Group, Department of Computing Imperial College London; Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara; Cooperative Medianet Innovation Center, Shanghai Jiao Tong University). We will directly contact these groups to present them the challenge.

Moreover, domain adaptation has recently raised great interests in the MICCAI community. For example, we found that 36 papers on domain adaptation were presented last year at MICCAI (main conference, DART and MLMI workshops). To promote the challenge, we plan: 1/ to list the main DA papers for medical image segmentation and directly contact their authors; 2/ use traditional tools such as social networks and medical imaging community mailing lists

Publication and future plans

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

A joint manuscript summarizing the results of the challenge will be submitted to a high-impact journal in the field (e.g., Medical Image Analysis, IEEE Transactions on Medical Imaging).

Space and hardware requirements

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

CrossMoDA is an off-site challenge and we will make use of the grand-challenge.org website to host the challenge, evaluation engine and leaderboards. Nvidia GPU support would be greatly appreciated for the testing phase.
On the day of the challenge, a projector, a computer and two microphones will be needed for the presentations of the winning methods.

TASK: Unsupervised Cross-Modality Domain Adaptation for Vestibular Schwannoma and Cochlea Segmentation

SUMMARY

Keywords

List the primary keywords that characterize the task.

Domain Adaptation, Cross-Modality, Segmentation, Vestibular Schwannoma, MRI

ORGANIZATION

Organizers

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

Reuben Dorent, King's College London, School of Biomedical Engineering & Imaging Sciences, London, UK

Aaron Kujawa, King's College London, School of Biomedical Engineering & Imaging Sciences, London, UK

Jonathan Shapey, King's College London, School of Biomedical Engineering & Imaging Sciences, London, UK

Samuel Joutard, King's College London, School of Biomedical Engineering & Imaging Sciences, London, UK

Jorge Cardoso, King's College London, School of Biomedical Engineering & Imaging Sciences, London, UK

Marc Modat, King's College London, School of Biomedical Engineering & Imaging Sciences, London, UK

Nicola Rieke, NVIDIA

Ben Glocker, Imperial College London, Department of Computing, BioMedIA Group, London, UK

Spyridon Bakas, University of Pennsylvania, Center for Biomedical Image Computing and Analytics, Philadelphia, USA

Tom Vercauteren, King's College London, School of Biomedical Engineering & Imaging Sciences, London, UK

b) Provide information on the primary contact person.

Reuben Dorent, reuben.dorent@kcl.ac.uk

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

Repeated 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).

crossmoda-challenge.ml (In construction)

Participation policies

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

Fully automatic.

b) Define the policy on the usage of training data. The data used to train algorithms may, for example, be restricted to the data provided by the challenge or to publicly available data including (open) pre-trained nets.

No additional data allowed.

c) Define the participation policy for members of the organizers' institutes. For example, members of the organizers' institutes may participate in the challenge but are not eligible for awards.

May participate but not eligible for awards and not listed in leaderboard.

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

In discussion with an industry partner.

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 ranking of winning teams will be publicly available on the challenge website.
- Participating teams can ask to be removed from the leaderboard.
- The top 3 ranked teams will be invited to present their method during the DART workshop. (TBC)
- The challenge results will be summarized in a joint manuscript.

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).

Up to two team members are qualified as author for the joint publication.
Participants are free to publish their own results separately, but can only mention overall results once the challenge paper is published (submission to arXiv is considered as a sufficient waiting period).

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.

The validation and test submission processes are different:
- Validation leaderboard: the participating teams are required to submit their predictions as a single compressed zip file via the grand-challenge.org interface. The tree directory structure will be described once the challenge is online.
- Test leaderboard: the participating teams are required to submit a Docker container on the Synapse Platform. Instructions will be given based on the BraTS example: https://github.com/BraTS/Instructions.

Based on the BraTS challenge, this choice is a trade-off between reducing the participant workload and limiting the risk of cheating for the final ranking.

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 are allowed to evaluate their algorithms on the validation set. While we expect the participants to try different approaches (e.g., network architecture, domain adaptation techniques), we don't want them to select their model hyper-parameters in a supervised manner, i.e. by computing the prediction accuracy. For this reason, the validation submissions will be limited to 2 per week.

Only one submission on the testing set will be allowed. To avoid any pitfalls (e.g., docker container issues), participants will get the opportunity to submit two submissions evaluated on a small subset of the test data.

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: February 2021
Release date of the training set: February 2021
Release date of the validation set: April 2021
Release of the test set: mid July 2021
Contributed challenge papers: end July 2021 (short description of the proposed technique, results on the validation set, technique for choosing the hyper-parameters)

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 contributions to this study were based on approval by the NHS Health Research Authority and Research Ethics Committee (18/LO/0532) and were conducted in accordance with the 1964 Declaration of Helsinki.

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 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.

Evaluation code will be made available with the training data.

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

To have access to the test leaderboard, all participating teams must provide an open-access to their training and testing code.

The participating teams are required to submit a docker container for performing the inference on the test data.

The top 3 ranked teams are required to submit their training and testing codes in a docker container for verification after the challenge submission deadline to ensure that the challenge rules have been respected.

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.

This challenge is supported by Wellcome Trust (203145Z/16/Z, 203148/Z/16/Z) and EPSRC (NS/A000050/1, NS/A000049/1) funding. All the organizers will have access to the test set if needed.

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

Surgery, Research, Intervention planning, Treatment planning.

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 consists of subjects with vestibular schwannoma from whom hrT2 scans have been acquired.

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

The challenge cohort consists of subjects with vestibular schwannoma from whom either ceT1 or hrT2 scans have been acquired.

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).

The image and segmentation data will be released in as compressed Nifti1Image objects stored in .nii.gz files. The image orientation and the voxel size can be found in these Nifti1Image files.

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

No additional patient information will be released.

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 MRI.

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.

Vestibular Schwannoma, Cochlea

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)).

Specificity, Sensitivity, Robustness, Accuracy.

DATA SETS

Data source(s)

a) Specify the device(s) used to acquire the challenge data. This includes details on the device(s) used to acquire the imaging data (e.g. manufacturer) as well as information on additional devices used for performance assessment (e.g. tracking system used in a surgical setting).

All images were obtained on a 32-channel Siemens Avanto 1.5T scanner using a Siemens single-channel head coil.

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

Contrast-enhanced T1-weighted imaging was performed with an MPRAGE sequence with in-plane resolution of 0.4×0.4 mm, in-plane matrix of 512×512, and slice thickness of 1.0 to 1.5 mm (TR=1900 ms, TE=2.97 ms, TI=1100 ms).

High-resolution T2-weighted imaging was performed with a 3D CISS or FIESTA sequence in-plane resolution of 0.5x0.5 mm, in-plane matrix of 384x384 or 448x448, and slice thickness of of 1.0 to 1.5 mm (TR=9.4 ms, TE=4.23ms).

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.

All images were acquired at the Queen Square Radiosurgery Centre (Gamma Knife).

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).

All structures were manually segmented in consensus by the treating neurosurgeon and physicist using both the ceT1 and hrT2 images. All segmentations were performed using the GammaPlan software that employs an in-plane semi-automated segmentation method. Using this software, delineation was performed on sequential 2D axial slices to produce 3D models for each structure.

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.

The training, validation and test cases represent MR images of a human brain. Specifically:
- A source training case represents a contrast-enhanced T1 (ceT1) scan with a manual segmentation of the tumour (vestibular schwannoma) and the cochlea.
- A target training case represent an unlabeled hrT2 scan.
- A target validation case represents a hrT2 scan with a manual segmentation of the tumour (vestibular schwannoma) and the cochlea. Note that participating teams won't have access to the target validation segmentations.
- A target test case represents a hrT2 scan with a manual segmentation of the tumour (vestibular schwannoma) and the cochlea. Note that participating teams won't have access to the target test cases.

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

Source training cases: 105 ceT1 scans

Target training cases: 105 hrT2 scans
Target validation cases: 32 hrT2 scans
Target test cases: 80 hrT2 scans

Importantly, each scan comes from a different patient (N=322). This especially mean there is no overlap between the source and target training cases.

Note that we already have access to the training and validation sets and have submitted an ethic amendment to include the 80 target test hrT2 scans. We are confident that this will be approved soon.

To further increase the data we intend to make available during this challenge, we are currently working with neuroradiologists at the University of Pennsylvania to identify additional multi-parametric MRI scans and manual annotations.

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

We already have access to the training and validation sets (N=242). We requested 80 other annotated scans to have a 75%-25% train-test split.

We chose the above split to ensure that:
- each data comes from a different patient (unsupervision)
- the training source and target sets are balanced (105 cases per set)
- the training, validation and test data respectively represent 65%, 10% and 25% of the overall dataset, according to common practices in machine learning.

d) Mention further important characteristics of the training, validation and test cases (e.g. class distribution in classification tasks chosen according to real-world distribution vs. equal class distribution) and justify the choice.

The training and validation data are being made publicly available on TCIA. This especially means that the segmentations of the training and validation T2 scans could be found somewhere else. To minimise the chance that the participant teams cheat by using these labels:
1/ the participant teams will be required to release their training and testing code and explain how they fine-tuned their hyper-parameters.
2/ the top 3 ranked teams will be required to submit their training and testing codes in a docker container for verification after the challenge submission deadline in order to ensure that the challenge rules have been respected.

Note that the test data won't be available on TCIA and will remain private.

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.

All segmentations were performed using the GammaPlan software that employs an in-plane semi-automated segmentation method. Using this software, delineation was performed on sequential 2D axial slices to produce 3D

models for each structure.

b) Provide the instructions given to the annotators (if any) prior to the annotation. This may include description of a training phase with the software. Provide the information separately for the training, validation and test cases if necessary. Preferably, provide a link to the annotation protocol.

The segmentation of the vestibular schwannoma and the cochlea were performed within the standard stereotactic radiosurgery worfkflow. We didn't give specific instructions to the annotators.

c) Provide details on the subject(s)/algorithm(s) that annotated the cases (e.g. information on level of expertise such as number of years of professional experience, medically-trained or not). Provide the information separately for the training, validation and test cases if necessary.

The tumour volume (vestibular schwannoma) and the main organ at risk (cochlea) were delineated in consensus by the patient's Consultant Neurosurgeon and treating Physicist with expert input from a Consultant Neuroradiologist.

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.

Imaging data: data were fully de-identified by removing all health information identifiers and de-faced [11]. Since the data were acquired in a consistent way (similar voxel spacing, same scanner), no further image pro-processing is required. The data will be released in the NIfTI format.

Annotations: the segmentation files will be released in the NIfTI format after converting the contours into label maps using plastimatch (https://plastimatch.org/plastimatch.html).

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.

Forty-six images were also manually segmented using both ceT1 and hrT2 scans by another neurosurgeon, who was blinded to the original manual segmentation of the tumour volume in order to test inter-observer variability. Inter-observer variability testing between clinical annotators recorded a Dice score of 93.82% (SD 3.08%) and an ASSD score of 0.269 (SD 0.095) mm.

Inter-observer variability for cochlea segmentation hasn't been evaluated yet.

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

The contours were originally drawn using: 1/ ceT1 scans as reference for VS segmentation; 2/ hrT2 scans as reference for cochlea segmentation. Consequently, the VS segmentation of the hrT2 scans and the cochlea segmentation of the ceT1 scans had been obtained after resampling, leading to some potential minor errors.

ASSESSMENT METHODS

Metric(s)

a) Define the metric(s) to assess a property of an algorithm. These metrics should reflect the desired algorithm properties described in assessment aim(s) (see above). State which metric(s) were used to compute the ranking(s) (if any).

- Example 1: Dice Similarity Coefficient (DSC)
- Example 2: Area under curve (AUC)

- Dice Similarity Coefficient (DSC)
- Average symmetric surface distance (ASSD)

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

The metrics (DSC, ASSD) were chosen because of their simplicity, their popularity, their rank stability [10], and their ability to assess the accuracy of the predictions.

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.

We will apply the ranking methods described in the BraTS challenge. Participating teams are ranked for each testing subjects, for each evaluated region (i.e., VS and cochlea), and for each measure (i.e., DSC and ASSD). The final ranking score for each team is then calculated by firstly averaging across all these individual rankings for each patient (i.e., Cumulative Rank), and then averaging these cumulative ranks across all patients for each participating team.

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

All the predictions for the test set will be locally performed using the teams' docker containers. No missing results is thus expected on the test set.

Concerning the validation set, all the predictions must be present in the submission zip file. Incomplete submissions won't be evaluated and counted as a validation submission.

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

This ranking scheme is standard and has been adopted in other challenge with satisfactory results, such as BraTS [12] and ISLES (http://www.isles-challenge.org/).

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.

Submission with missing data won't be accepted.

Ranking variability will be characterized using the Wilcoxon signed-rank test and bootstrapping.
This analysis will be performed in Python 3.7.

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

The two techniques are both simple non-parametric statistical methods.
Bootstrapping allows for simulating the performance of methods for a new set drawn from the distribution, to quantify ranking stability. The Wilcoxon signed-rank test allows for testing significance as we will have paired comparisons between the different techniques.

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.

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] Shapey, J., et al: An artificial intelligence framework for automatic segmentation and volumetry of Vestibular Schwannomas from contrast-enhanced t1-weighted and high-resolution t2-weighted MRI. In: Journal of Neurosurgery JNS. (2019)

[2] Wang, G., et al: Automatic segmentation of Vestibular Schwannoma from T2-weighted MRI by deep spatial attention with hardness-weighted loss. In: MICCAI 2019. (2019)

[3] Van de Langenberg, R., et al: Follow-up assessment of vestibular schwannomas: volume quantification versus two-dimensional measurements. In: Neuroradiology 51, 517 (2009).

[4] Kamnitsas, K., et al: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: IPMI (2017)

[5] Yang, J., et al: Unsupervised Domain Adaptation via Disentangled Representations: Application to Cross-Modality Liver Segmentation. In: MICCAI 2019. (2019)

[6] Chen, C., el al: Synergistic Image and Feature Adaptation: Towards Cross-Modality Domain Adaptation for Medical Image Segmentation. In: AAAI-19. (2019)

[7] Dou, Q., et al: Pnp-adanet: Plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation. ArXiv. (2018)

[8] Orbes-Arteaga, et al: Multi-domain adaptation in brain MRI through paired consistency and adversarial learning. In: DART - Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels

and Imperfect Data. (2019)

[9] Xue, Y., el al: Dual-task Self-supervision for Cross-Modality Domain Adaptation. In: MICCAI 2020. (2020)

[10] Maier-Hein, L., Eisenmann, M., et al: “Is the winner really the best? A critical analysis of common research practice in biomedical image analysis competitions”. ArXiv:1806.02051, (2018)

[11] Milchenko, M., Marcus, D. Obscuring Surface Anatomy in Volumetric Imaging Data. In: Neuroinform 11, 65–75 (2013).

[12] Bakas, S., et al: "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge". ArXiv:1811.02629 (2018).