Topology-Aware Anatomical Segmentation of the Circle of Willis for CTA and MRA: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Topology-Aware Anatomical Segmentation of the Circle of Willis for CTA and MRA

Challenge acronym

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

TopCoW'23

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.

The Circle of Willis (CoW) is an important anastomotic network of arteries connecting the anterior and posterior circulations of the brain, as well as the left and right cerebral hemispheres [1]. The CoW is located at the skull base and is the joining area of major arteries to the brain. Due to its centrality, the CoW is commonly involved in pathologies like aneurysms and stroke. Clinically, the vascular architecture of the CoW is believed to impact the occurrence and severity of stroke [2, 3]. An accurate characterization of the CoW is therefore of great clinical relevance.

However, clinicians have articulated an unmet demand for efficient software tools to analyze and compare the angio-architecture of the CoW. Today, assessing the anatomy and vascular components of the CoW from medical angiography images is still an expert task and time-consuming. Furthermore, the CoW naturally has many variants of which certain principal artery components are hypoplastic or absent. It is estimated that only around half of our population has a complete CoW [4, 5]. These anatomical CoW variants should not be misinterpreted as a vascular disease, and it is not an exception to see the CoW vary markedly from person to person. An automated and personalized CoW vascular characterization will be of significant interest to both the clinical and the research communities.

Here, we propose the first public challenge on CoW angio-architecture extraction and brain vessel segment annotation on two common angiographic imaging modalities, namely magnetic resonance angiography (MRA) and computed tomography angiography (CTA). Although there have been two publicly available datasets on MRA modality, the CASILab and IXI datasets [6,7], their MRA scans are acquired from fairly old machines (from the year 2004 to 2006). More importantly, very limited annotations were provided if any. Annotated dataset on the other important modality, CTA, is even rarer. Thus, we identify a gap in good quality anatomical annotations, newer imaging datasets, and datasets of dual modalities. By releasing a public dataset covering both CTA and MRA with

careful annotations on the vasculature, our challenge can foster research and better benchmark comparisons on the CoW vascular segmentation results.

The aim of the challenge is threefold; first, to extract the CoW angioarchitecture from 3D angiographic imaging; second, to automatically annotate the vessel components; and third, to characterize the CoW topologically. We release a new dataset of joint-modalities, CTA and MRA of the same patient cohort, both with annotations of the underlying anatomy of CoW. There is a temporal relationship between the two modalities from the same patients in the form of follow-up scans. The inclusion criteria of the patient cohort is that at least one of the modalities allow for a diagnosis of the underlying CoW anatomical and geometric characterization. Joint-modalities from the same patient will serve as additional reference and provide supplementary anatomical information on the CoW vessels. We believe our temporal joint-modality annotations on MRA and CTA can ensure good quality anatomical annotations. Such a joint-modality dataset also opens doors to other medical imaging research questions such as registration and unpaired modality network design. Our challenge has two tracks for the same segmentation task with one track for each modality. We leverage both modalities during our annotation. And participants can leverage whichever modality they want, both CTA and MRA, and choose to tackle the task for either modality.

A technical emphasis of this challenge is on topology-aware segmentation. The extracted vessels should retain the topology of the underlying anatomy. We especially hope to raise awareness on the importance of evaluating performance beyond pixel-based or volumetric metrics. In particular, we will evaluate the segmentation performance on topology-based, junction-based, and graph-based metrics. The objects of interest will include centerlines, bifurcation points, and geometric shapes. Segmentation gives a complete description of the geometry of the vasculature. Extracting features such as centerline, radii, bifurcation points from segmentation has been frequently done. The challenge aims for vascular characterizations that capture the underlying topology and geometric variability of the CoW.

We believe this is a timely and new challenge that can be of benefit to both clinicians and medical imaging researchers. An automated CoW characterization on vessel anatomical labelings and topological properties can lead to many interesting downstream tasks and applications. Apart from enabling neurovascular disease management and personalized treatment planning, this challenge can also impact research and discovery. There are gems encoded in the angio-architecture of CoW such as hemodynamic implications and flow analysis. Profiling CoW variants can also reveal demographic geometric risk factors for vascular pathologies. An accurate CoW anatomical segmentation helps in development of more complex models, more efficient quantification, and reduction of cognitive workload and more consistent labeling procedure. Based on the clinical impact, data value, and research potential, we look forward to organizing this challenge (and its repeats) and welcoming submissions.

References:

[1] AG, Osborn. "Osborn's brain: imaging, pathology, and anatomy." Salt Lake City, Utah: Amirsys (2013): 932-940.
[2] Chuang, Yu-Ming, et al. "Configuration of the circle of Willis is associated with less symptomatic intracerebral hemorrhage in ischemic stroke patients treated with intravenous thrombolysis." Journal of critical care 28.2 (2013): 166-172.
[3] van Seeters, Tom, et al. "Completeness of the circle of Willis and risk of ischemic stroke in patients without cerebrovascular disease." Neuroradiology 57.12 (2015): 1247-1251.
[4] Krabbe-Hartkamp, Monique J., et al. "Circle of Willis: morphologic variation on three-dimensional time-of-flight MR angiograms." Radiology 207.1 (1998): 103-111.
[5] Iqbal, S. "A comprehensive study of the anatomical variations of the circle of willis in adult human brains."

Journal of clinical and diagnostic research: JCDR 7.11 (2013): 2423.
[6] Bullitt, Elizabeth, et al. "Vessel tortuosity and brain tumor malignancy: a blinded study1." Academic radiology 12.10 (2005): 1232-1240.
[7] IXI - Information eXtraction from Images. https://brain-development.org/ixi-dataset/

Challenge keywords

List the primary keywords that characterize the challenge.

Circle of Willis, Topology Aware Segmentation, Anatomical Segmentation, CT Angiography, MR Angiography, Vessel Profiling, Vessel Labeling

Year

The challenge will take place in ...

2023

FURTHER INFORMATION FOR MICCAI ORGANIZERS

Workshop

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

We plan to partner with Iris Vos and Hugo Kuijf from University Medical Center Utrecht, who are hosting a similar CoW challenge called CROWN at MICCAI 2023, to host a half-day sister-event in the same physical room for the in-person event.

No associated workshop (only a joint sister-event with CROWN challenge).

Duration

How long does the challenge take?

Half day.

Expected number of participants

Please explain the basis of your estimate (e.g. numbers from previous challenges) and/or provide a list of potential participants and indicate if they have already confirmed their willingness to contribute.

Based on recent publications and research interest on vessel and CoW segmentation, and an upcoming potential similar challenge also on CoW (by Utrecht, see above), we estimate at least 10 participating teams.

We intend to conduct a series of both online and in-person seminars to publicize and broadcast our challenge to attract more participants, as well as to clarify any questions from the participants (actively monitor the forum on grand-challenge website).

Publication and future plans

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

Yes. We plan to summarize and publish the TopCoW'23 challenge results and benchmark in a journal manuscript.

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.

For the in-person event, we need projectors, microphones, loudspeakers, and also cameras for hybrid participation.

The challenge will be off-line, but hosted on grand-challenge.org. Participants use their own computing resources for the algorithm training and development. The organizers will use the grand-challenge.org for the docker evaluation during the testing phase.

TASK: Topology Aware Anatomical Segmentation of the CoW for CTA and MRA

SUMMARY

Keywords

List the primary keywords that characterize the task.

Circle of Willis, Topology Aware Segmentation, Anatomical Segmentation, CT Angiography, MR Angiography, Vessel Profiling, Vessel Labeling

ORGANIZATION

Organizers

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

[University of Zurich, Zurich, Switzerland]
Kaiyuan Yang, Hongwei Bran Li, Anjany Sekuboyina, Bjoern Menze

[University Hospital of Zurich, Zurich, Switzerland]
Susanne Wegener, Yihui Ma, Laura Westphal

[Helmholtz Munich, Bavaria, Germany]
Rami Al-Maskari, Luciano Höher

[Zurich University of Applied Sciences, Zurich, Switzerland]
Fabio Musio, Norman Juchler, Sven Hirsch

[Technical University of Munich, Bavaria, Germany]
Johannes C. Paetzold, Suprosanna Shit, Diana Waldmannstetter, Florian Kofler, Ivan Ezhov

[University of Zurich, Switzerland]
Ibrahim Hamamci

[National University of Singapore, Singapore]
Andrew Makmur, James Hallinan

[Geneva University Hospitals, Switzerland]
Philippe Bijlenga

[University of Toronto, Canada]
Jeroen Bisschop

[Technical University of Munich, Germany]
Daniel Rueckert, Bene Wiestler

b) Provide information on the primary contact person.

Kaiyuan Yang (kaiyuan.yang@uzh.ch) and Prof. Bjoern Menze (bjoern.menze@uzh.ch) from University of Zurich

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

For MICCAI 2023, this challenge will have a fixed submission deadline to present the awards and milestones for the in-person event during the MICCAI 2023 conference.

Subject to the response to this challenge, we intend to repeat and scale up the challenge for MICCAI 2024 and beyond. Possible extensions to the challenge include diversifying and extending the dataset to multiple centers in Europe and Asia. We also plan to annotate deeper and more peripheral regions of the cerebral vasculature.

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).

To appear

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.

Participants may use any other public datasets and private in-house data, or modify the supplied TopCoW 2023 training data, provided that they disclose and mention any additional or modified training datasets in their description of the submitted algorithm. They also have to report and discuss the differences in their results from before and after using only the TopCoW 2023 data.

The participants also need to disclose the total amount of compute and the type of resources used (e.g., type and number of GPUs, internal cluster, or cloud provider, and high-performance computing cluster) used for their training.

We encourage the participants to be mindful of their use of computational resources and energy consumption.

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.

Members of the organizers' research groups can participate and their results can be included in the publications and the leaderboard. However, they are not eligible for awards.

People not from the organizers' research groups, i.e. from other labs/departments, may participate and are eligible for the awards and to be listed in the leaderboard.

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

The top three teams for segmentation performance from each of the two tracks (CTA track and MRA track) will be publicly named and given a certificate along with a box of Swiss chocolate. We will not give monetary awards but we plan to compensate the CO2 emissions of all participants with a suitable sponsor. We intend to talk with NVIDIA or AWS to sponsor winners a hardware award in the form of GPUs or cloud-credits.

Given that there are two modality-tracks, each with two sub-tasks, there will be four rankings to be announced and awarded.

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.

Top performing submissions are announced at the in-person challenge event. However, the participating team can choose whether their results will be made public any time before the day of announcement. The top 3-5 teams will be invited to prepare a 5-10 minute presentation for the challenge session to present and discuss their methods.

After the public announcement, a detailed analysis of the submitted results will be available upon request.

If a participant wishes to retract their submission after results are made public, their submitted performance will either be reported in an anonymized fashion both online and in the publication, or removed from the leaderboard and publication.

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).

The TopCoW'23 challenge results and benchmark will be summarized and published in a journal manuscript. At least two authors from the top ten submissions will be listed in the challenge paper and included in the authors list. Additional authors from the top submissions will be included upon request with justification according to ICMJE authorship guidelines.

Participating teams may submit their results separately without any publication embargo.

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 for evaluation will be done on the private test datasets via submitted docker containers.

Validation set: (public test data and private annotations): Participants submit predictions on our validation set images. This validation set is for participants to adjust their methods to unseen new data and to gauge their performance relative to others. But to prevent participants from submitting manually labeled clinical annotations on the validation images, we also introduce a hidden test set. The validation set is not used for final evaluation.

Test set: (private test data private and private annotations): Participants submit docker containers which are evaluated by organizers on grand-challenge.org.

Along with the docker containers, each participating team is encouraged to submit a 1-page summary describing their methods and approaches one week after the docker submission deadline.

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.

Participants will have access to the validation set images. Predictions on the validation data can be submitted to grand-challenge.org to get them automatically evaluated. However, in order to prevent overfitting and re-training on the test set, the number of submissions for the validation set predictions is limited to one submission per team per day.

The validation data will be used for the public leaderboard to get a preliminary unofficial ranking. Note that the final ranking will be based on the combined performance on the test data only and on more thorough evaluation metrics.

Each team is given only one opportunity to upload their containers for the hidden testset. In case of technical issues, we allow the participants to try again their docker submissions.

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

Preliminary Schedule:
- Challenge Website Online: April 14th 2023

- Release of Training data: May 1st, 2023
- Release of public validation data: July 15th, 2023
- Opening of submission system and leaderboard for validation data: July 15th, 2023
- Closing of submission system: Aug 1st, 2023
- Contacting top performing teams and plan for the in-person session: Aug 14th 2023 (teams that require visas will be contacted earlier. We will coordinate with MICCAI to send invitation letters for visa clearance.)
- in-person challenge event: Oct 8th 2023

Additional point: we intend to conduct a series of both online and in-person seminars to publicize and broadcast our challenge and to clarify any questions from the participants.

Ethics approval

Indicate whether ethics approval is necessary for the data. If yes, provide details on the ethics approval, preferably institutional review board, location, date and number of the ethics approval (if applicable). Add the URL or a reference to the document of the ethics approval (if available).

The data used in this challenge is a subset of research data from a study that has been approved by the relevant ethical committee. No additional ethics approval is required. The data is anonymized (removal and anonymization of relevant DICOM patient information). We perform additional de-facing and cropping procedures to ensure patient privacy in the image data.

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.

Code availability

a) Provide information on the accessibility of the organizers' evaluation software (e.g. code to produce rankings). Preferably, provide a link to the code and add information on the supported platforms.

The evaluation and ranking code will be made publicly available on Github and the challenge website.

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

The submitted docker containers will be made publicly available with the participating teams' permission. We encourage the participants to make their code public.

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.

We may expect one or two sponsors to provide hardware or cloud-service awards, but we will not give monetary awards.

Only the main organizers and their local annotation team will have access to all test labels and the private test datasets.

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, Diagnosis, Education, Screening, Training.

Task category(ies)

State the task category(ies).

Examples:

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

Segmentation (and depending on the approaches to tackle the task, intermediate steps could possibly involve registration, but it is not required)

Additional points:

1. Participants can choose from two tracks, one track for CTA modality and one track for MRA modality. They are encouraged to take part in both tracks.
2. The segmentation task is divided into two sub-tasks: binary segmentation of CoW vessels, and multi-class CoW vessel anatomical segmentation.
3. Participants can choose to only tackle the multi-class segmentation sub-task, as the multi-class segmentation can be converted into a binary segmentation evaluation. If participants only submit for the multi-class segmentation, the results will be automatically evaluated for binary segmentation by us as well, thus considered taking part in both sub-tasks.
4. The evaluation of segmentation results is limited to the CoW region of interest. The CoW region of interest (ROI) is defined as the 3D bounding box capturing the anatomical variant of the CoW. The annotated CoW ROI will be released for all training cases.

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 is the general population, as an accurate CoW vasculature characterization can be beneficial in many clinical applications, ranging from screening of patients being at a higher risk of stroke, to improving the clinical management of patients with cerebrovascular diseases

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

The challenge cohort is composed of patients admitted to the Zurich stroke registry (Department of Neurology and Stroke Center, University Hospital Zurich) in the year 2019. The inclusion criteria for the challenge data are: 1) Both MRA and CTA scans are available for that patient; 2) at least the MRA or CTA allows for an assessment of the CoW anatomy and geometry. The patients of the challenge cohort were recovering from a stroke-related neurological disorder, like large vessel occlusion (LVO) stroke, transient ischemic attack (TIA), stroke mimic, retinal infarct or cerebral vein thrombosis.

Imaging modality(ies)

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

Computed tomography Angiography (CTA) and Time of Flight Magnetic Resonance Angiography (TOF-MRA)

Context information

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

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

No contextual clinical information will be made available.

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

No clinical information about the patient will be made available.

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.

The dataset consists of brain angiographic CT and MR scans (CTAs and MRAs). The original CTA scans usually cover neck and head regions, and sometimes just the head region. The original MRA scans typically cover the patients' brain region. The MRA is usually a follow-up scan after a previous CTA of the same patient, but the time stamp information will not be made available.

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.

There are two objectives for the participating algorithms: Firstly, they should be able to segment the vessels of the CoW region in 3D angiograms. Secondly, they should be able to segment the different anatomical components of the CoW: Left and right internal carotid artery (ICA), left and right anterior cerebral artery (ACA), left and right middle cerebral artery (MCA), anterior communicating artery (Acom), left and right posterior communicating artery (Pcom), left and right posterior cerebral artery (PCA), and basilar artery (BA).

There are two tracks for the vessel segmentation task, one track for CTA modality and one track for MRA modality.

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)).

The focus of this challenge is on topology-aware and anatomical segmentations of CoW vessels. We assess the segmentation on both CTA and MRA modalities. Thus the assessment of algorithms are divided into three categories following our prior work [Shit et al. 2021], and for both modalities (CTA track and MRA track):

1. Volumetric metrics: Dice coefficient and centerline-Dice [Shit et al. 2021] of the CoW vessels
2. Topology-based metrics: errors of Betti numbers such as connected components and circular holes
3. Graph-based metrics: Junction/landmark-based F1 score [Citraro et al. 2020]

The segmentation results will be evaluated only for the CoW region. We will not assess the segmentation performance on the peripheral and further downstream vessels outside the CoW region. Participants should focus on segmenting the CoW vessel components necessary to characterize the CoW angio-architecture.

Further clarification on the CoW region for evaluation, the CoW region of interest (ROI) is defined as the 3D bounding box capturing the anatomical variant of the CoW. The annotated CoW ROI will be released for all

training cases.

References:
Shit, Suprosanna, et al. "clDice-a novel topology-preserving loss function for tubular structure segmentation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
Citraro, Leonardo, Mateusz Koziski, and Pascal Fua. "Towards reliable evaluation of algorithms for road network reconstruction from aerial images." European Conference on Computer Vision. Springer, Cham, 2020.

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).

MRA scans were typically acquired by SIEMENS Skyra model or Avanto Fit model, with magnetic field strength of 3 Tesla or 1.5 Tesla, and with TOF-3D mode or TOF-3D multi slab mode.

CTA is typically acquired by SIEMENS SOMATOM Definition Flash using Dual Energy (DE).

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

The clinical data was acquired during routine examinations from patients admitted to the University Hospital Zurich during the year 2019. Standard clinical protocols as of 2019 were applied.

For CTA, the voxel size ranges from around 0.34 to 0.53 mm in the X-Y dimension, and ranges from around 0.62 to 0.75 mm in the Z dimension.

For MRA, the voxel size ranges from around 0.29 to 0.35 mm in the X-Y dimension, and ranges from around 0.5 to 0.6 mm in the Z dimension.

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.

University Hospital Zurich, Switzerland

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).

The data was acquired at the University Hospital of Zurich during routine examinations following standard procedures for MRA and CTA.

Training and test case characteristics

a) State what is meant by one case in this challenge. A case encompasses all data that is processed to produce one result that is compared to the corresponding reference result (i.e. the desired algorithm output).

Examples:

- Training and test cases both represent a CT image of a human brain. Training cases have a weak annotation (tumor present or not and tumor volume (if any)) while the test cases are annotated with the tumor contour (if any).
- A case refers to all information that is available for one particular patient in a specific study. This information always includes the image information as specified in data source(s) (see above) and may include context information (see above). Both training and test cases are annotated with survival (binary) 5 years after (first) image was taken.

A case in this challenge is a 3D angiographic imaging scan of a human brain. Both CTA and MRA modalities are provided for the same human patient. At least one of the modalities contain the pixel-level information necessary to diagnose the underlying CoW anatomical and geometric structure. All patients have both CTA and MRA modalities, one scan for each modality. The joint-modality aspect is that for each MRA, there is a corresponding CTA pair from the same patient. Dual modalities should provide supplementary sources of information for annotation, and potentially for training.

All the vessel components of CoW necessary to diagnose the CoW angio-architecture are annotated voxel-wise, and different vessel segments are labeled with a different voxel value (see Item 23 Annotation Characteristics).

Training and test cases all have the MRA and CTA joint-modality pair. The task is to segment the CoW vessels and anatomical components on the CoW region in either MRA track or CTA track.

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

Total 130 patients with both CTA and MRA
Both CTA and MRA of the same patient are provided, one scan for each modality. So in total 130x2=260 angiographic imaging scans.

Training dataset: 90 patients (both image and annotations released)
Validation set: 10 patients (image released to public but without annotations)
Test set: 30 patients (not released to public)

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

Availability and quality control. These 130 patients' data with temporal joint-modalities originate from a larger research project. These 130 patients should also allow sufficiently diverse variability in the CoW angioarchitecture.

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 angiography scans are obtained from patients admitted to the Department of Neurology and can be of different indications, such as LVO, TIA, stroke mimic, retinal infarct, and cerebral sinus vein thrombosis. For stroke patients in our dataset, their MRA imaging data is usually after vessel clearance in case of any occlusion stroke but without visible interventional artifacts.

Both training and test set will include CTA and MRA joint-modality pairs of the same patients.

Annotation characteristics

a) Describe the method for determining the reference annotation, i.e. the desired algorithm output. Provide the information separately for the training, validation and test cases if necessary. Possible methods include manual image annotation, in silico ground truth generation and annotation by automatic methods.

If human annotation was involved, state the number of annotators.

Each case will be annotated by an annotator-approver workflow. The segmentation ground truth is first bootstrapped with initial manual annotations by a team of research staff experienced with cerebrovascular anatomy and CTA and MRA research. Then the initial annotations are verified and fine-tuned by one clinical expert on neurovascular disease. The clinical expert team consists of neurologists, neuro-radiologists, and a neurosurgeon. The clinical leaders in our organizing team will then go through all the annotations in the final round and approve the annotation.

All MRA scans are checked with the paired CTA scan and vice versa to make use of the supplementary source of information from both modalities during annotation.

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

Annotators have both CTA and MRA modalities available to annotate.

Initial bootstrap annotations are manually labeled in 3D using the SyGlass software. The initial annotations are done by research staff who have gone through anatomy training from the clinical experts specifically on the CoW structure. Then the annotations are verified, fine tuned and approved by our team of clinicians either in 3D using SyGlass or in 2D using ITK-Snap or Slicer.

The vessel components of the CoW to be annotated are left and right internal carotid artery (ICA), left and right anterior cerebral artery (ACA), left and right middle cerebral artery (MCA), anterior communicating artery (Acom), left and right posterior communicating artery (Pcom), left and right posterior cerebral artery (PCA), and basilar artery (BA). Note that only vessel components and regions necessary to diagnose the CoW angio-architecture and variants will be annotated.

A common annotation protocol on how to segment vessel components at bifurcation points such as ACA-ICAMCA, ACA-Acom, PCA-Pcom, Pcom-ICA etc are provided to all annotators and verifiers. When annotators are unsure how to assign vessel labels to the bifurcations, those cases are flagged and then annotated by consensus agreement.

Protocols are also given for special CoW variants such as fetal PCA, triple ACA etc.

Below is the annotation mask in pixel value for each vessel objects of interest:
Background: 0
Right-ACA: 1
Left-ACA: 2
Acom: 3
Right-MCA: 4

Left-MCA: 5
Right-ICA: 6
Left-ICA: 7
Right-Pcom: 8
Left-Pcom: 9
Right-PCA: 10
Left-PCA: 11
BA: 12

For the binary segmentation sub-task, the vessel components from above are combined into two classes without separate annotations:
Background: 0
CoW Vessels: 1

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.

After an initial manual annotation, all scans are verified and fine tuned by a team of clinical experts on neurovascular disease. Our clinical team is made up of neurologists, neuro-radiologists, and neuro-surgeon from the university hospital and co-organizing clinical institutions who are specializing in neuro-vascular diseases.

d) Describe the method(s) used to merge multiple annotations for one case (if any). Provide the information separately for the training, validation and test cases if necessary.

In cases where discussions among the clinicians are required, the annotation is merged mutually. All annotations will be checked and approved by the clinical leader in the final round and in case of disagreement, the clinical leader will flag the cases for another round of discussion.

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 3D angiographic scans in both MRA and CTA are first cropped and de-faced to only the head region, if necessary, to anonymize the patients. Then both the raw image and the vessel segmentation masks of this head region will be exported as image stacks with relevant headers.

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.

We have a hierarchical annotation process in which there are multiple levels of verification and approval based on clinical seniority. Each case may have multiple annotators and verifiers involved and all cases are approved by our clinical team leaders, thus we expect the variations to be small.

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

CTA and MRA might have different resolutions. MRA might amplify the effect of stenosis in vessels. Both

modalities might contain common artifacts such as flow-dependent signal cancellation artifacts, noise artifacts, ringing artifacts, pulsation artifacts, and beam hardening artifacts. These artifacts might cause the vessel segmentation to be over- or under-segmented.

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)

Three evaluations metrics for sub-task 1: Binary (CoW Vessel vs Background) Segmentation
1. cl-DICE [Shit et al. 2021]
2. Betti number errors (similar to [Shit et al. 2021])
3. Junction/landmark-based F1 score [Citraro et al. 2020] (similar to [Shit et al. 2021])

One evaluations metric for sub-task 2: Multi-class (CoW anatomical vessels) Segmentation
1. Weighted average of cl-DICE [Shit et al. 2021] and DICE similarity coefficient

References:
Shit, Suprosanna, et al. "clDice-a novel topology-preserving loss function for tubular structure segmentation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
Citraro, Leonardo, Mateusz Koziski, and Pascal Fua. "Towards reliable evaluation of algorithms for road network reconstruction from aerial images." European Conference on Computer Vision. Springer, Cham, 2020.

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

Dice similarity coefficient and centerline-Dice [Shit et al. 2021] measure the voxel overlap between the ground truth and the segmentation prediction. In particular, we highlight the centerline-Dice metric (clDice), which is suitable for evaluating voxel-wise overlap for tubular and curvilinear structures such as CoW vessels [Maier-Hein et al. 2022]. clDice extends the traditional Dice by also measuring how much of the vessels are covered (coverage of vessel network). Thus clDice is used for both the binary and the multi-class segmentation subtasks.

Betti numbers measure the topological properties such as connected components and circular holes.

Junction/landmark-based F1 scores is a graph-based metric that also measures the topological properties between two graphs [Citraro et al. 2020]. It is sensitive to end-point nodes and missing connections. It was originally developed for evaluating the quality of road networks.

The above metrics are largely based on the evaluations performed in our prior work clDice [Shit et al. 2021].

References:
Maier-Hein, Lena, et al. "Metrics reloaded: Pitfalls and recommendations for image analysis validation." arXiv preprint arXiv:2206.01653 (2022).
Shit, Suprosanna, et al. "clDice-a novel topology-preserving loss function for tubular structure segmentation."

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
Citraro, Leonardo, Mateusz Koziski, and Pascal Fua. "Towards reliable evaluation of algorithms for road network reconstruction from aerial images." European Conference on Computer Vision. Springer, Cham, 2020.

Ranking method(s)

a) Describe the method used to compute a performance rank for all submitted algorithms based on the generated metric results on the test cases. Typically the text will describe how results obtained per case and metric are aggregated to arrive at a final score/ranking.

For the binary segmentation sub-task, we sum the ranks across the three evaluation metrics and use this averaged rank for the final leaderboard.

For multi-class segmentation sub-task, the ranking will be based on the one metric.

Bootstrapping (sampling with replacement) will be performed on the test-set to get robust rankings.

The ranking for each metric is obtained via Wilcoxon signed-rank test (with 'greater' or 'lesser' hypothesis as appropriate to the metric) on the test set.

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

For volumetric metrics, if the submitted method fails to produce a result on a test case, the metric for that test case will be set to its most penalizing value.

For topology-based metrics, missing results on Betti numbers will be calculated as absolute Betti errors.

For graph-based metrics, missing results for junction/landmark-based F1 score will be set to its most penalizing value.

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

Wilcoxon signed-rank test (with 'greater' or 'lesser' hypothesis as appropriate to the metric) indicates if there is any statistical significance on the test data between two teams being compared. Similar ranking schemes were used in other recent medical challenges like BraTS challenge (http://braintumorsegmentation.org/) and VerSe challenge (https://verse2020.grand-challenge.org/), and also as recommended by the literature [Maier-Hein et al. 2018].

References:
Maier-Hein, Lena, et al. "Why rankings of biomedical image analysis competitions should be interpreted with care." Nature communications 9.1 (2018): 1-13.

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.

Each team will be compared with the other teams using Wilcoxon signed-rank test to determine if there is a statistically significant difference between the two compared teams.

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

Similar statistical methods were used in other challenges such as BraTS challenge (http://braintumorsegmentation.org/) and VerSe challenge (https://verse2020.grand-challenge.org/) with positive feedback from the participants, and as recommended by the literature [Maier-Hein et al. 2018].

References:
Maier-Hein, Lena, et al. "Why rankings of biomedical image analysis competitions should be interpreted with care." Nature communications 9.1 (2018): 1-13.

Further analyses

Present further analyses to be performed (if applicable), e.g. related to

- combining algorithms via ensembling,
- inter-algorithm variability,
- common problems/biases of the submitted methods, or
- ranking variability.

There are three categories of metrics assessed in this challenge, ranging from volumetric metrics to graph and topology based metrics. In the final publication, we will carry out a combined analysis on methods that can perform well for volumetric metrics while preserving topological and geometric properties. The resulting summary publication may conduct a post-challenge ranking that considers more properties to re-rank the submissions.

Segmentation gives a complete description of the geometry of the vasculature. We plan to extract features such as centerline, radii/diameters, bifurcation points from segmentation in the further analysis.

The anatomical segmentation allows us to further analyze the detection performance as well, i.e. how well the vascular connections in the brain can be detected.

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.

References are inserted in-place for the relevant text-fields.