TopCoW 2024 (2nd Edition): 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.

TopCoW 2024 (2nd Edition): 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).

TopCoW24

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 network of arteries connecting major circulations of the brain. Its vascular architecture is believed to affect the risk, severity, and clinical outcome of serious neuro-vascular diseases. However, characterizing the highly variable CoW anatomy is still a manual and time-consuming expert task. The CoW is usually imaged by two angiographic imaging modalities, magnetic resonance angiography (MRA) and computed tomography angiography (CTA), but there exist limited public datasets with annotations on CoW anatomy, especially for CTA. Therefore, we organized the TopCoW challenge [1] in 2023 with the release of an annotated CoW dataset. The TopCoW dataset was the first public dataset with voxel-level annotations for thirteen possible CoW vessel components, made possible by virtual-reality (VR) technology. It was also the first large dataset with paired MRA and CTA from the same patients. We invited submissions worldwide for the CoW segmentation task, which attracted over 140 registered participants from four continents and resulted in 18 high-quality algorithms.

TopCoW 2023 represented a first attempt at benchmarking the CoW anatomical segmentation task for MRA and CTA, both morphologically and topologically. It formalized the CoW characterization problem as a multiclass anatomical segmentation task with an emphasis on topological metrics. The top performing teams managed to segment many CoW components to Dice scores around 90%, but with lower scores for certain communicating arteries and rare variants. There were also topological mistakes for predictions with high Dice scores. Additional topological analysis revealed further areas for improvement in detecting certain CoW components and matching CoW variant topology accurately.

Based on the results in 2023, for the second edition of the TopCoW challenge in 2024, we propose to expand and improve the challenge in the following three aspects:

Firstly, in 2024, we will increase the dataset both in terms of size and site. In the new edition, we plan to increase the training set from 90 pairs of MRA-CTA dual modalities to 125 pairs, the test set size from 35 pairs to 70 pairs (in total 125+5+70=200 pairs). We also plan to gather an additional multi-institution internal test set to validate the generalizability and robustness of the top submissions. The additional multi-institutional test set will contain around 50 patients from multiple hospitals (around five centers) from several countries across Europe and Asia.

Secondly, we will also expand the quantitative evaluation for the anatomical segmentation task. Building on the experience from the first edition, we will include more direct evaluations on the topological properties of the segmentation, such as the topology matching of CoW variant graphs. Furthermore, overlap- and boundary-based evaluations will be for detected classes, and we will evaluate the detection performance of certain CoW components as a separate metric. The expanded segmentation metrics will hopefully guide the submissions to better preserve the topology of the CoW and tailor to closer clinical needs.

Thirdly, we will introduce two new tasks with added annotations. On top of the existing CoW segmentation task, we introduce a second segmentation task to segment the five efferent arteries supplied by the CoW, and a third (graph) classification task to classify representative CoW variants.

For Task 2, we will expand the CoW annotation to include the peripheral vascular trees distal to the CoW incoming arteries. The five efferent arterial trees of CoW are important anatomically and clinically, while their multiclass segmentation task has never been benchmarked. Anatomically, they represent most of the cerebral arterial network, and to be able to segment these vascular trees in images as multiclass labels has great anatomical relevance. Furthermore, many clinical diagnostic tasks involve locating and analyzing the distal vessels of CoW. For example, clinicians often need to identify and analyze the affected vessels and their branches in the five vascular trees in disease like ischemic stroke, aneurysm, and moyamoya disease. The localization information provided by the vascular trees can also be combined with other neurological analysis like brain parcellation and perfusion analysis like CT and MR perfusion to add cerebral arterial network information. Lastly, we remark that the multiclass CoW distal vascular tree segmentation task is new and has never been benchmarked. Our dataset will be one of the first to release such multiclass annotation labels to the community.

For Task 3, the CoW variant graph classification task has more immediate clinical impact, as the class labels of the CoW variant can be directly applied and used by clinicians in the clinical setting. Participants can transform the solutions to the multiclass segmentation task to solve for the graph classification task. Participants can also approach the classification task via other interesting paradigms such as naive image classification or graph learning. The newly proposed CoW variant graph classification task can help bridge vessel segmentation tasks with exciting graph problems in brain vasculature. Our graph classification labels share similarities with the labels from the partner-challenge CROWN last year, and we are interested to see how our version of more graph-based labels in the form of edge-list performs on our dataset as an extended benchmark.

As a first benchmark for CoW anatomical segmentation task, TopCoW 2023 gathered strong baseline results for further algorithm development and comparison. For the 2024 edition, TopCoW will continue to increase the paired CTA and MRA dataset size and with new and more enriched annotations. In the new edition, we will also introduce new metrics for more thorough evaluations and introduce new tasks to expand the challenge scope. For our new 2024 organization team, we have involved more clinical partners for data annotation and some of the participants from 2023 for pre-annotation. We look forward to organizing the second edition of the TopCoW challenge and welcoming submissions for an improved and more comprehensive CoW benchmark.

References:
[1] Yang, K., Musio, F., Ma, Y., Juchler, N., Paetzold, J. C., Al-Maskari, R., ... & Menze, B. (2023). TopCoW: Benchmarking Topology-Aware Anatomical Segmentation of the Circle of Willis (CoW) for CTA and MRA. arXiv preprint arXiv:2312.17670.

Challenge keywords

List the primary keywords that characterize the challenge.challenge_

Circle of Willis, Anatomical Segmentation, Brain CT Angiography, Brain MR Angiography, Vascular Topology, Graph Classification

Year

The challenge will take place in 2024

FURTHER INFORMATION FOR CONFERENCE ORGANIZERS

Workshop

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

No associated workshop.

Duration

How long does the challenge take?

Half day.

Expected number of participants

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

During the first iteration, TopCoW 2023 edition received 146 registered participants. We observed that many of the registrations occurred towards the end of the challenge. For example, over one-third (50+) of the participants registered in the last two months (August and September). Thus, we believe there is still strong momentum and room for more participation from the community for this year. For the 2024 edition, we estimate the registration to be at least 200 participants.
In 2023, there were 20 teams that submitted algorithms. Quite a few t eams from 2023 told us that they still had better algorithms not yet submitted due to time constraint. Thus, we anticipate many old teams to take part again. And estimating from the ratio of registration to submissions (around 10%), we expect at least 20 submitting teams from the 200 participants as a low estimate. As a high estimate we would foresee the participation of the same number of groups as in 2023, plus a reasonable increase resulting from higher visibility and better

preparation for a continuing two-year event, and one may reasonably anticipate 30-40 submissions

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 challenge results for both 2023 and 2024 in journal publications.

The TopCoW 2023 challenge results have been summarized in a 23-page pre-print on arXiv on December 29 2023 [1].

References:

[1] Yang, K., Musio, F., Ma, Y., Juchler, N., Paetzold, J. C., Al-Maskari, R., ... & Menze, B. (2023). TopCoW: Benchmarking Topology-Aware Anatomical Segmentation of the Circle of Willis (CoW) for CTA and MRA. arXiv preprint arXiv:2312.17670.

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 a camera/videoconferencing system 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 grand-challenge.org for the docker evaluation during the testing phase.

TASK 1: Multiclass Segmentation of CoW Anatomy

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.

Same as the challenge abstract, please see above.

Keywords

List the primary keywords that characterize the task.

Same as the challenge keywords, please see above.

ORGANIZATION

Organizers

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

[University of Zurich, Switzerland]
Kaiyuan Yang, Ibrahim Ethem Hamamci, Anjany Sekuboyina, Suprosanna Shit, Houjing Huang, Ezequiel De la Rosa, Chinmay Prabhakar, Bjoern Menze

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

[Imperial College London, UK]
Johannes C. Paetzold

[Harvard Medical School, USA]
Hongwei Bran Li

[University Hospital of Zurich, Switzerland]
Susanne Wegener, Laura Westphal, Elisa Colombo, Hakim Baazaoui

[Zhongnan Hospital of Wuhan University, China]
Yihui Ma

[Geneva University Hospitals, Switzerland]
Philippe Bijlenga, Julien Hämmerli, Catherine Wurster

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

[Technical University of Munich, Germany]
Diana Waldmannstetter, Florian Kofler, Fernando Navarro, Martin Menten, Ivan Ezhov, Daniel Rueckert, Bene Wiestler, Jan S. Kirschke

[University Hospital Berne and University of Berne, Switzerland]
Roland Wiest

[German Cancer Research Center (DKFZ) Heidelberg, Germany]
Maximilian R. Rokuss, Yannick Kirchhoff, Fabian Isensee, Klaus Maier-Hein

[Harbin Institute of Technology (Shenzhen), China]
Pengcheng Shi, Wei Liu, Ting Ma

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

[University of Toronto, Canada]
Jeroen Bisschop

[University of Chicago, USA]
Amrish Soundararajan

b) Provide information on the primary contact person.

Kaiyuan Yang (kaiyuan.yang@uzh.ch) and Fabio Musio (fabio.musio@uzh.ch)

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 for Task 1 (multiclass CoW anatomical segmentation task) from 2023 edition with new metrics and larger test set.
For MICCAI 2024, this challenge will have a fixed submission deadline to present the awards and milestones for the in-person event during the MICCAI 2024 conference.

Challenge venue and platform

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

MICCAI 2024

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).

TopCoW23 website: https://topcow23.grand-challenge.org/ TopCoW24 website: topcow24.grand-challenge.org (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 are allowed to use any other public datasets and private in-house data, or modify the supplied TopCoW training data, provided that they disclose any additional or modified training datasets in their description of the submitted algorithm.

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 (as baseline algorithms). However, they are not eligible for awards (e.g. Swiss cow toys).

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.

The collaborating winning teams from TopCoW23 will not be able to enter the challenge this year as usual participants, but instead will be tagged as organizers if they take part. And their methods will be publicly tagged as baseline or from the organizers.

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

Top three teams for each task in either MRA or CTA track (3 tasks and 2 tracks, or 6 leaderboards) will be publicly named as shown on the grand-challenge leaderboard webpage, and given a small Swiss wooden toy cow as a souvenir at the in-person challenge event. There will be no monetary awards given.

Given that there are two modality-tracks, each with three tasks, there will be six rankings to be announced and awarded. (TopCoW23 had four rankings announced, TopCoW24 will have six.)

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 TopCoW23 and TopCoW24 challenge results and benchmark will be summarized and published in a journal manuscript. All participants with a reasonable submission are invited to contribute to our challenge publication. Each submission can have maximum three co-authors for the challenge paper. Additional authors from the submissions can be included upon request with justification according to the ICMJE authorship guidelines.

Participating teams may publish their own 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 test datasets via submitted docker containers, i.e. type 2 submissions on Grand Challenge. Participants submit docker containers which are evaluated by organizers on grand-challenge.org on the test set (private test data and private annotations).

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. This summary is required for co-authorship in the final challenge journal paper.

The submission instructions for TopCoW23 are documented on this page:
https://topcow23.grand-challenge.org/participation-rules/

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 but without annotations. 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 validation set, the number of submissions for the validation set predictions is limited to one submission per team per day. The validation phase is not used for final evaluation, and participants are encouraged to use the validation phase to debug and validate their docker submission workflow.

The final test phases only allow for one submission per team for each phase. Each team is given only one opportunity to upload their containers for the hidden test set for a task in a track. 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 15th, 2024
- Release of New Annotations for TopCoW23 Data: June 01, 2024
- Release of TopCoW24 Training Data: July 01, 2024
- Submit to validation phases: July 01 until September 15, 2024
- Submit to final test phases: August 01 until September 15, 2024
- Contacting top performing teams and plan for the in-person session: Aug 01 until October 01 2024 (teams that require visas will be contacted earlier. We will coordinate with MICCAI to send invitation letters for visa clearance.)
- in-person challenge event: October 6th 2024

Note: the TopCoW23 training and validation data have been released since June 2023 under the CC BY-NC (Attribution-NonCommercial) license. The dataset access link is on our challenge website https://topcow23.grand-challenge.org/data/
Therefore, the Task 1 (multiclass CoW anatomical segmentation) which is continued from TopCoW23 can be considered to be in fact on-going now.

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 local ethical committee and that has been approved for public release. The data is anonymized (removal and anonymization of relevant DICOM patient information) in accordance with the IRB regulations. This includes 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 implementation of our evaluation metric code for TopCoW23 was open sourced at https://github.com/CoWBenchmark/TopCoW_ Eval_Metrics.

Our challenge evaluation will be transparent to all participants. We will use the same GitHub repo to update and synchronize any changes to the evaluation code for TopCoW24.

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 permissions from the participating teams. 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 will not give monetary awards.

Only the main organizers and their local annotation team will have access to all test and validation 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

Task-1 (CoW anatomical segmentation):
Segmentation

Additional points:
- 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.
- The evaluation of task-1 cow-segmentation results is limited to within the CoW region of interest (ROI). The CoW ROI is defined as the 3D bounding box containing the volume required for the diagnosis of the CoW variant [1].

The CoW ROI bounding boxes 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 data cohort was composed of patients admitted to the Stroke Center of the University Hospital Zurich (USZ) in 2018 and 2019. The inclusion criteria for the challenge data were: 1) both MRA and CTA scans were available for that patient; 2) at least the MRA or CTA allowed for an assessment of the CoW anatomy. The patients of the challenge cohort were admitted for or recovering from a stroke-related neurological disorder, including ischemic stroke, transient ischemic attack, stroke mimic, retinal infarct or amaurosis fugax, intracerebral hemorrhage, and cerebral sinus vein thrombosis. There is no evidence that the CoW appearance of this cohort would differ significantly from the general population.

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 due to the requirement for anonymization.

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

No clinical information about the patient will be made available due to the requirement for anonymization.

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 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.

Task-1 (CoW anatomical segmentation):
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 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)).

Task-1 (CoW anatomical segmentation):
The focus of task-1 is on topology-aware anatomical segmentations of CoW vessels. We assess the segmentation on both CTA and MRA modalities. Thus, the assessment of algorithms is by six metrics following our prior work [1,2,3,4], and for both modalities (CTA track and MRA track):
1. Class-average Dice similarity coefficient
2. Class-average centerline Dice (clDice)
3. Class-average Betti number in dimension 0 (zero-th Betti number) errors
4. Class-average Hausdorff Distance 95% Percentile (HD95)
5. Average F1 score (harmonic mean of the precision and recall) for detection of certain CoW components that are not always present (Group 2 components [1])
6. Variant-average topology match rate [1]

The evaluation of task-1 cow-segmentation results is limited to within the CoW region of interest (ROI). The CoW ROI is defined as the 3D bounding box containing the volume required for the diagnosis of the CoW variant [1]. The CoW ROI bounding boxes will be released for all training cases. We will not assess the segmentation performance on the peripheral and further downstream vessels outside the CoW ROI. Participants should focus on segmenting the CoW vessel components necessary to characterize the CoW angioarchitecture.

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. Standard clinical protocols as of 2019 were applied.

For CTA, the voxel size was around 0.45 mm in the X-Y dimension, and around 0.7 mm in the Z dimension.

For MRA, the voxel size was around 0.3 mm in the X-Y dimension, and around 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.

Training data and the test data uploaded to grand-challenge.org (130 pairs from TopCoW23 and another 70 pairs planned, or 200 pairs in total) are acquired from University Hospital Zurich, Switzerland.

The separate additional internal test set is pending approval from hospitals in Geneva, Singapore, and other centers in Switzerland and Germany.

d) Describe relevant characteristics (e.g. level of expertise) of the subjects (e.g. surgeon)/objects (e.g. robot) involved in the data acquisition process (if any).

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 image information necessary to diagnose the underlying CoW anatomical and geometric structure. All patients have both CTA and MRA modalities, one scan for each modality. A case for any task is the 3D angiographic imaging scan for that modality track.

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 section Annotation Characteristics). A bounding box of the CoW ROI is released for all training cases.

The task is to segment the CoW vessels and anatomical components in the CoW region in either MRA track or CTA track.

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

At least 200 patients with both CTA and MRA.
Both CTA and MRA of the same patient are provided, one scan for each modality. In total at least 200x2=400 angiographic imaging scans.

Training dataset: 125 patients (both image and annotations released)
Validation set: 5 patients (image released to public but without annotations)
Test set: 70 patients (not released to public but uploaded to grand-challenge.org)

One of the partnering hospitals in Geneva is open to releasing at least 30 patients to add to the test set that can be uploaded to grand-challenge.org, so the common test set can be 70+30=100 patients.

Potentially we plan to curate another in-house test set from partnering hospitals from other centers (pending approval), and we estimate around 50 patients.

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 200 patients should cover sufficient variability in the CoW angioarchitecture. We increase the test set used by grand-challenge.org by more than 100% (70-100 patients), and now the bigger test set will have more variability. We carefully inspect and verify all the annotations, and 200 patients are still within the capacity of our quality control.

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 for or recovering from a stroke-related neurological disorder, including ischemic stroke, transient ischemic attack, stroke mimic, retinal infarct or amaurosis fugax, intracerebral hemorrhage , and cerebral sinus vein thrombosis. For stroke patients in our dataset, their MRA imaging data is usually after treatment in case of any occlusion stroke but without visible interventional artifacts. In rare cases, the angiography scans can see the large vessel occlusions in one of the modalities.

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.

Task-1 (CoW anatomical segmentation):
For the TopCoW23 challenge, the initial 130 pairs of CoW annotations were manually labeled voxel-wise by two research staff who had gone through CoW anatomy education from the clinical experts. The cases that annotators were uncertain were flagged, and sent for further verification and approval by the clinical experts. The clinical expert team consisted of neuro-radiologists, neurologists, and neurosurgeons. All patients from the TopCoW challenge had both CTA and MRA modalities, with one scan for each modality. Annotators and clinical experts had both CTA and MRA modalities available to annotate and verify. The anatomy of the CoW was first inspected in both CTA and MRA to diagnose the anatomical components. Then the CoW vessels were annotated for each modality separately.

For the TopCoW24 challenge, we collaborate with top teams from the 2023 challenge and pre-annotate the new cases using the winning automatic methods. The model predictions are then sent for manual correction and further verification and approval by the same clinical experts as in the 2023 edition [1].

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.

Task-1 (CoW anatomical segmentation):
Annotators and clinical experts have both CTA and MRA modalities available to annotate and verify. The anatomy of the CoW is first inspected in both CTA and MRA to diagnose the anatomical components. Then the CoW vessels are annotated for each modality separately.

The manual annotation and verification are done in 3D using virtual-reality (VR) for efficient diagnosis and visualization.

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.

The annotation protocol on how to segment vessel components at bifurcation points such as ACA-ICA-MCA, ACAAcom, PCA-Pcom, Pcom-ICA, etc., were discussed and agreed upon by the clinical experts. The annotation protocol also covered CoW variants such as fetal PCA, triple ACA etc. The cases that annotators are uncertain are flagged, and sent for further verification and approval by the clinical experts.

All the vessel components of CoW necessary to diagnose the CoW angio-architecture are annotated voxel-wise. Multi-class labels contain a different voxel value for different CoW vessel segments:

0: Background, 1: BA, 2: R-PCA, 3: L-PCA, 4: R-ICA, 5: R-MCA, 6: L-ICA, 7: L-MCA, 8: R-Pcom, 9: L-Pcom, 10: Acom,11: R-ACA, 12: L-ACA, 15 : 3rd-A2

Task-1 (CoW anatomical segmentation) is continued from our 2023 challenge. For more information on the annotation protocol for task-1, please refer to [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.

The initial manual annotation and the following manual correction are verified and approved 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.

The automatic methods to pre-annotate new cases are from some of the top teams in 2023 challenge [1].

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.

N.A.

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 were anonymized (removal and anonymization of relevant DICOM patient information). Additional de-facing and cropping procedures were performed to ensure patient privacy in the image data after converting the DICOM to NIfTI format. The defaced image includes only the braincase region. Other than the defacing and the cropping to braincase region steps, no further preprocessing of the data is performed to keep the data as close to the clinical setting as possible. The image data are saved in NIfTI format, 16-bit signed, and in LPS+ orientation.

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.

Task-1 (CoW anatomical segmentation):
We investigated the inter-rater agreement on the CoW anatomical annotations in [1], and the agreement was 90% class-average Dice score, 99% clDice score, and near 0 class-average zero-th Betti number error.

We analyzed the interrater agreement in terms of voxel-wise annotations of individual CoW components after the CoW variant had been diagnosed and verified by the clinicians. Diagnosing the CoW variant can be an expert task that requires considerable highly-specialized clinical experience and knowledge. We will conduct further inter-rater agreement among senior clinicians for CoW component detection and CoW variant diagnosis.

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

CTA and MRA can 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 vessels to be over- or under-segmented, and might affect the variant diagnosis.

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)

Task-1 (CoW anatomical segmentation):
The focus of task-1 is on topology-aware anatomical segmentations of CoW vessels. We assess the segmentation on both CTA and MRA modalities. Thus, the assessment of algorithms is by six metrics following our prior work [1,2,3,4], and for both modalities (CTA track and MRA track):
1. Class-average Dice similarity coefficient
2. Class-average centerline Dice (clDice)
3. Class-average zero-th Betti number errors
4. Class-average Hausdorff Distance 95% Percentile (HD95)
5. Average F1 score (harmonic mean of the precision and recall) for detection of certain CoW components that are not always present (Group 2 components [1])
6. Variant-average topology match rate [1]

The evaluation of task-1 cow-segmentation results is limited to within the CoW region of interest (ROI). The CoW ROI is defined as the 3D bounding box containing the volume required for the diagnosis of the CoW variant [1]. The CoW ROI bounding boxes will be released for all training cases. We will not assess the segmentation performance on the peripheral and further downstream vessels outside the CoW ROI. Participants should focus on segmenting the CoW vessel components necessary to characterize the CoW angioarchitecture.

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

Task-1 (CoW anatomical segmentation):
Dice similarity coefficient and centerline-Dice [2] 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 [5]. clDice extends the traditional Dice by also measuring how much of the vessels are covered (coverage of vessel network).
We believe the classic Dice coefficient can still complement some aspects of the segmentation that clDice alone

may not capture. For example, while clDice is able to capture the connectivity aspect of the segmentation, the classic Dice is better suited to quantify accurate segmentation of the width and diameter which are related to over- and under-segmentation. For the CoW anatomy, we want to evaluate for both the volumetric overlap and the connectivity aspects. Thus we keep both Dice and clDice metrics for the assessment in order to evaluate the segmentation results holistically.

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

The Group 2 CoW components are smaller in diameter and volume, which are not best evaluated in overlap-based metrics like Dice and clDice scores as they are sensitive to small disagreement in voxels [6]. Thus, we introduce boundary-based metrics like Hausdorff distance to better evaluate CoW components with smaller structure size. 95% percentile Hausdorff distance (HD95) is less affected by extreme outliers.

Based on the results from 2023 edition, we felt a need to include more in-depth metrics on topological properties. The 2023 challenge using Dice, clDice and Betti number error metrics turned out to be inadequate to guide the submissions to preserve the topology of the CoW variants. Thus, for the new challenge we include direct evaluations on the topologies of the predicted segmentation, such as the connection conditions with relevant neighboring CoW components.

The above metrics are largely based on the evaluations performed in our prior work [1,2,3,4].

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.

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 [7].

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

Task-1 (CoW anatomical segmentation):

For missing results on test cases, i.e. a true negative, a false positive or a false negative predicted CoW anatomical component, the evaluations for Dice, clDice, Betti number error, and Hausdorff distance will be ignored. Therefore, those four above metrics (out of the six metrics) will only be evaluated when it is a true positive detection.

This is because based on our 2023 results, we found that detection seemed especially challenging for the Group 2 CoW components, i.e., the communicating arteries (Acom, R-Pcom and L-Pcom) and the rare 3rd-A2 segment. Therefore, we decide to use a separate, specific detection metric: a F1 score (harmonic mean of recall and precision), specifically for the Group 2 CoW components.

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

In TopCoW23, due to time constraint, the ranking for the awards of the MICCAI event was based on the leaderboards displayed on grand-challenge website, which rank the teams by the descending order of their per-case performance on each metric. The leaderboard used equal weights for each metric column and used the mean of the positions or ranks from metric columns, or rank-then-average.

In TopCoW24, we will explore the possibility of using Wilcoxon signed-rank test (with greater or lesser hypothesis as appropriate to the metric) on the test set to rank the submitted methods for each metric. Similarly, we will then use equal weights for each metric ranking, and use the mean of the ranks, or rank-then-average.

Bootstrapping (sampling with replacement) may also be performed on the test-set to analyze ranking robustness and stability post-challenge.
The bootstrapped 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.

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 [7].

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.

In the final publication, we will carry out additional topological 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. We will conduct ranking stability analysis post-challenge.

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 further analysis.

TASK 2: Multiclass Segmentation of CoW Distal Vascular Trees

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.

Same as the challenge abstract, please see above.

Keywords

List the primary keywords that characterize the task.

Same as the challenge keywords, please see above.

ORGANIZATION

Organizers

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

Same as task-1

b) Provide information on the primary contact person.

Same as task-1

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

Task-2 is a new task and will have a fixed submission deadline to present the awards and milestones for the in-person event during the MICCAI 2024 conference.

Challenge venue and platform

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

MICCAI 2024

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

Same as task-1

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

Same as task-1

Participation policies

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

Same as task-1

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.

Same as task-1

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.

Same as task-1

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

Same as task-1

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.

Same as task-1

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).

Same as task-1

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.

Same as task-1

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.

Same as task-1

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

Same as task-1

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).

Same as task-1

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)

Same as task-1

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.

Same as task-1

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

Same as task-1

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.

Same as task-1

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

Same as task-1

Task category(ies)

State the task category(ies)

Examples:

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

• Segmentation

• Tracking

Task-2 (CoW distal vascular tree segmentation)
Segmentation

Additional points:
- 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.
- The evaluation of task-2 vascular-tree-segmentation results is limited to outside the CoW region of interest (ROI). The CoW ROI is defined as the 3D bounding box containing the volume required for the diagnosis of the CoW variant [1]. The CoW ROI bounding boxes 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.

Same as task-1

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

Same as task-1

Imaging modality(ies)

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

Same as task-1

Context information

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

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

Same as task-1

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

Same as task-1

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.

Same as task-1

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.

Task-2 (CoW distal vascular tree segmentation):
To segment the distal vascular tree supplied by the CoW: Anterior vascular tree (ACA-tree), left middle cerebral artery (L-MCA-tree), right middle cerebral artery (R-MCA-tree), left posterior cerebral artery (L-PCA-tree), right posterior cerebral artery (R-PCA-tree)

There are two tracks for the 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)).

Task-2 (CoW distal vascular tree segmentation):
The focus of task-2 is on topology-aware segmentations of distal vascular trees supplied by the CoW. We assess the segmentation on both CTA and MRA modalities. Thus, the assessment of algorithms is by these metrics following our prior work [1,2,3,4], and for both modalities (CTA track and MRA track):
1. Class-average Dice similarity coefficient
2. Class-average centerline Dice (clDice)
3. Class-average Hausdorff Distance 95% Percentile (HD95)

The evaluation of task-2 vascular-tree-segmentation results is limited to outside the CoW region of interest (ROI). The CoW ROI is defined as the 3D bounding box containing the volume required for the diagnosis of the CoW variant [1]. The CoW ROI bounding boxes will be released for all training cases.

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).

Same as task-1

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

Same as task-1

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.

Same as task-1

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).

Same as task-1

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 image information necessary to diagnose the underlying CoW anatomical and geometric structure. All patients have both CTA and MRA modalities, one scan for each modality. A case for any task is the 3D angiographic imaging scan for that modality track.

The downstream distal vascular trees supplied by the CoW (five vascular trees) are annotated voxel-wise, and each vascular tree is labeled with a different voxel value (see section Annotation Characteristics). A bounding box of the CoW ROI is released for all training cases.

The task is to segment the distal vascular trees outside the CoW region in either MRA track or CTA track.

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

Same as task-1

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

Same as task-1

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.

Same as task-1

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.

Task-2 (CoW distal vascular tree segmentation):
For the TopCoW24 challenge, we collaborate with top teams from the 2023 challenge and adopt an active learning approach to annotate for task-2 based on the winning automatic methods. With a few manually annotated cases, the initial model is trained, and used for model prediction on the next batch of cases. Model predictions are then manually corrected and sent for further verification and approval by the clinical experts like in the 2023 edition [1]. The clinical expert team consists of neuro-radiologists, neurologists, and neurosurgeons. All patients from the TopCoW challenge have both CTA and MRA modalities, with one scan for each modality. Annotators and clinical experts have both CTA and MRA modalities available to annotate and verify. The anatomy of the distal vascular tree is first inspected in both CTA and MRA to diagnose the anatomical components. Then the vascular tree vessels are annotated for each modality separately.

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.

Task-2 (CoW distal vascular tree segmentation):
All patients from the TopCoW challenge have both CTA and MRA modalities, with one scan for each modality. Annotators and clinical experts have both CTA and MRA modalities available to annotate and verify. The anatomy of the distal vascular tree is first inspected in both CTA and MRA to diagnose the anatomical components. Then the vascular tree vessels are annotated for each modality separately.

The manual annotation and verification are done in 3D using virtual-reality (VR) for efficient diagnosis and visualization.

The five vascular trees distal to the CoW to be annotated are the anterior vascular tree (ACA-tree), left middle cerebral artery (L-MCA-tree), right middle cerebral artery (R-MCA-tree), left posterior cerebral artery (L-PCA-tree), right posterior cerebral artery (R-PCA-tree). The five vascular trees are annotated voxel-wise, and each vascular tree is labeled with a different voxel value:
0: Background, 2: R-PCA-tree, 3: L-PCA-tree, 5: R-MCA-tree, 7: L-MCA-tree, 13: ACA-tree

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.

Same as task-1

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.

N.A.

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.

Same as task-1

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.

Task-2 (CoW distal vascular tree segmentation):
The distal vascular tree downstream of CoW main vessels are easier to verify and diagnose than the complex CoW anatomy in task-1. Possible source of error are the peripheral regions in-between different vascular trees, such as the region supplied by both MCA and ACA or by both MCA and PCA. We have a hierarchical annotation process in which there are multiple levels of verification and approval based on
clinical seniority. Difficult cases 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.

Same as task-1

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)

Task-2 (CoW distal vascular tree segmentation):
The focus of task-2 is on topology-aware segmentations of distal vascular trees supplied by the CoW. We assess the segmentation on both CTA and MRA modalities. Thus the assessment of algorithms is by these metrics following our prior work [1,2,3,4], and for both modalities (CTA track and MRA track):
1. Class-average Dice similarity coefficient
2. Class-average centerline Dice (clDice)
3. Class-average Hausdorff Distance 95% Percentile (HD95)

The evaluation of task-2 vascular-tree-segmentation results is limited to outside the CoW region of interest (ROI). The CoW ROI is defined as the 3D bounding box containing the volume required for the diagnosis of the CoW variant [1]. The CoW ROI bounding boxes will be released for all training cases.

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

Task-2 (CoW distal vascular tree segmentation):
Dice similarity coefficient and centerline-Dice [2] 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 [5]. clDice extends the traditional Dice by also measuring how much of the vessels are covered (coverage of vessel network).
We believe the classic Dice coefficient can still complement some aspects of the segmentation that clDice alone may not capture. For example, while clDice is able to capture the connectivity aspect of the segmentation, the classic Dice is better suited to quantify accurate segmentation of the width and diameter which are related to over- and under-segmentation. For the CoW anatomy, we want to evaluate for both the volumetric overlap and the connectivity aspects. Thus we keep both Dice and clDice metrics for the assessment in order to evaluate the segmentation results holistically.

Task 2 involves the segmentation of distal vascular trees that often can have disconnected floating fragments especially at the peripherals. The connectivity of these peripheral vessels are not as emphasized by the clinicians as compared to the CoW components. Thus we discarded metrics like Betti number error for task 2.

Similar to Task 1, we also use boundary-based metrics like Hausdorff distance to evaluate vessel segmentations. 95% percentile Hausdorff distance (HD95) is less affected by extreme outliers.

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.

Same as task-1

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

Task-2 (CoW distal vascular tree segmentation):

For overlap-based metrics like Dice and clDice, 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, 0 for that class.

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

Same as task-1

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.

Same as task-1

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

Same as task-1

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.

Same as task-1

TASK 3: Classification of CoW Variant Graphs

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.

Same as the challenge abstract, please see above.

Keywords

List the primary keywords that characterize the task.

Same as the challenge keywords, please see above.

ORGANIZATION

Organizers

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

Same as task-1

b) Provide information on the primary contact person.

Same as task-1

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

Same as Task-2 which is a new task and will have a fixed submission deadline to present the awards and milestones for the in-person event during the MICCAI 2024 conference.

Challenge venue and platform

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

MICCAI 2024

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

Same as task-1

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

Same as task-1

Participation policies

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

Same as task-1

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.

Same as task-1

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.

Same as task-1

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

Same as task-1

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.

Same as task-1

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).

Same as task-1

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.

Same as task-1

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.

Same as task-1

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

Same as task-1

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).

Same as task-1

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)

Same as task-1

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.

Same as task-1

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

Same as task-1

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.

Same as task-1

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

Same as task-1

Task category(ies)

State the task category(ies)

Examples:

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

- Segmentation
- Tracking

Task-3 (CoW variant topology graph classification)
Classification

Additional points:
- 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.
- The CoW variants are classified according to whether the topological graphs are missing certain edges.
- Participants can skip the segmentation intermediate step and directly try image classification or image-to-graph learning.
- Segmentations for Task-1 can be post-processed to obtain predictions for Task-3.

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.

Same as task-1

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

Same as task-1

Imaging modality(ies)

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

Same as task-1

Context information

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

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

Same as task-1

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

Same as task-1

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.

Same as task-1

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.

Task-3 (CoW variant topology graph classification):
To classify the CoW variant topology graph in terms of whether an edge of the graph is missing or present. The CoW topology graph is represented by a graph with the following anterior and posterior edge lists. The anterior edge list is [L-A1, Acom, L-A2, 3rd-A2, R-A2, R-A1]. The posterior edge list is [L-P2, L-Pcom, L-P1, BA, R-P1, R-Pcom, R-P2]. The presence and absence of the edge of the edge list determines the topology graph class. The algorithm is expected to output the predicted anterior and posterior edge lists.

There are two tracks for the 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)).

Task-3 (CoW variant topology graph classification):

The focus of task-3 is to classify the CoW variant topology graph. The CoW variant graph is classified by the anterior and posterior edge-lists, and we assess the algorithm by two metrics:
1. Anterior class-average accuracy
2. Posterior class-average 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).

Same as task-1

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

Same as task-1

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.

Same as task-1

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).

Same as task-1

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 image information necessary to diagnose the underlying CoW anatomical and geometric structure. All patients have both CTA and MRA modalities, one scan for each modality. A case for any task is the 3D angiographic imaging scan for that modality track.

The annotation for the underlying CoW variant graph is an anterior edge list and a posterior edge list.

The task is to classify the anterior and posterior topology graphs in either MRA track or CTA track.

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

Same as task-1

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

Same as task-1

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.

Same as task-1

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.

Task-3 (CoW variant topology graph classification):
The annotations for task-3 graph edge list are obtained by simple neighborhood and connection analysis on the verified ground-truth anatomical segmentation masks in task-1.

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.

Task-3 (CoW variant topology graph classification):
The annotations for task-3 graph edge list are obtained by simple neighborhood and connection analysis on the verified ground-truth anatomical segmentation masks in task-1.

The CoW topology graph is represented by a graph with the following anterior and posterior edge lists.
- The anterior edge list is [L-A1, Acom, L-A2, 3rd-A2, R-A2, R-A1].
- The posterior edge list is [L-P2, L-Pcom, L-P1, BA, R-P1, R-Pcom, R-P2].
The presence and absence of the edge of the edge list determines the topology graph class. The annotations of the CoW variant graph for one case are the anterior and posterior edge lists.

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.

Same as task-1

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.

N.A.

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.

Same as task-1

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.

Task-3 (CoW variant topology graph classification):
The annotations for task-3 graph edge list are obtained by simple neighborhood and connection analysis on the verified ground-truth anatomical segmentation masks in task-1. The expected source of error is correlated with

the one in Task-1.

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

Same as task-1

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)

Task-3 (CoW variant topology graph classification):

The focus of task-3 is to classify the CoW variant topology graph. The CoW variant graph is classified by the anterior and posterior edge-lists, and we assess the algorithm by two metrics:
1. Anterior class-average accuracy
2. Posterior class-average accuracy

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

Task-3 (CoW variant topology graph classification):
One of the key applications of our challenge is to characterize representative CoW anterior and posterior variants. Therefore, the CoW components need to be correctly detected for their presence/absence; additionally, the extracted topology graph should also match that of the CoW variant.
Thus, we introduce the task-3 of graph classification to retain and match the CoW variant topology.
A class-average accuracy prevents the model from blindly predicting the most prevalent class and will give an overview of the variant classification performance, especially given that the CoW variants are highly variable and imbalanced.

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.

Same as task-1

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

Task-3 (CoW variant topology graph classification):

If the prediction is missing for the classification task, that case will be set as the most penalizing value, 0.

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

Same as task-1

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.

Same as task-1

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

Same as task-1

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.

Same as task-1

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:
[1] Yang, K., Musio, F., Ma, Y., Juchler, N., Paetzold, J. C., Al-Maskari, R., ... & Menze, B. (2023). TopCoW: Benchmarking Topology-Aware Anatomical Segmentation of the Circle of Willis (CoW) for CTA and MRA. arXiv preprint arXiv:2312.17670.
[2] Shit, S., Paetzold, J. C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., ... & Menze, B. H. (2021). clDice-a novel topology-preserving loss function for tubular structure segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16560-16569).
[3] Stucki, N., Paetzold, J. C., Shit, S., Menze, B., & Bauer, U. (2023, July). Topologically faithful image segmentation via induced matching of persistence barcodes. In International Conference on Machine Learning (pp. 32698-32727). PMLR.
[4] Menten, M. J., Paetzold, J. C., Zimmer, V. A., Shit, S., Ezhov, I., Holland, R., ... & Rueckert, D. (2023). A skeletonization algorithm for gradient-based optimization. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 21394-21403).
[5] Maier-Hein, Lena, et al. "Metrics reloaded: Pitfalls and recommendations for image analysis validation." ArXiv preprint arXiv:2206.01653 (2022).

[6] Reinke, A., Tizabi, M.D., Sudre, C.H., Eisenmann, M., Radsch, T., Baumgartner, M., Acion, L., Antonelli, M., Arbel, T., Bakas, S., et al., 2021. Common limitations of image processing metrics: A picture story. arXiv preprint arXiv:2104.05642 .
[7] 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 comments

Further comments from the organizers.

N/A