Cell Tracking Challenge 2024: Structured description of the challenge design

CHALLENGE ORGANIZATION

Title

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

Cell Tracking Challenge 2024

Challenge acronym

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

CTC

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 Cell Tracking Challenge (CTC) was launched in 2012, with the aim of fostering the development of novel, robust cell segmentation and tracking algorithms, and helping the
developers with the evaluation of their new algorithmic developments. Over its more than a decade long existence, six fixed-deadline ISBI challenge editions have been organized, and
since February 2017, the challenge is open for online submissions that are monthly evaluated, ranked, and posted on the challenge website. So far, two benchmarks have been offered,
namely segmentation-and-tracking benchmark (evaluating segmentation and tracking performance) and segmentation-only benchmark (evaluating purely segmentation
performance, no tracking part is required). A detailed description of the focus and history of the CTC can be found at http://celltrackingchallenge.net/ and in the new open-access
Nature Methods summary of the 10 years of its existence. The CTC is in constant evolution, and - as we did in the previous six editions attached to ISBI 2013-2015 and ISBI 2019-2021 - we plan to introduce some novelties in this new ISBI-sponsored challenge event.
Specifically, in this new 7th edition, the participants will be encouraged to submit further solutions to the recently opened generalizability tasks - either in the frame of the
segmentation-and-tracking benchmark (Task 1) or the segmentation-only benchmark (Task 2). The generalizability tasks focus on the development of methods that exhibit better
generalizability and work across most, if not all, of the existing datasets, instead of being optimized for one or a few datasets only. These tasks were established for the ISBI 2021 edition, and their first results were reported in the abovementioned paper, but no further results have been received since then.
Furthermore, a new tracking-only - more precisely linking-only benchmark (Cell Linking Benchmark) will be introduced to complement the segmentation-only benchmark for those who want to evaluate purely the object linking methods without having to supply segmentation results. Such a benchmark has been missing in the CTC portfolio and it is demanded by the CTC participants and the scientific community at large. Participants will be

encouraged to supply ideally generalizable solutions (Task 3) working across 13 preselected datasets but will also be able to submit dataset-specific solutions (Task 4) for datasets of their choice.

Challenge keywords

List the primary keywords that characterize the challenge.challenge_

cell detection, cell segmentation, cell linking, cell tracking, 2D time-lapse, 3D time-lapse

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.

N/A

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.

/

Publication and future plans

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

/

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.

N/A

TASK 1: CTC

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.

The Cell Tracking Challenge (CTC) was launched in 2012, with the aim of fostering the development of novel, robust cell segmentation and tracking algorithms, and helping the
developers with the evaluation of their new algorithmic developments. Over its more than a decade long existence, six fixed-deadline ISBI challenge editions have been organized, and
since February 2017, the challenge is open for online submissions that are monthly evaluated, ranked, and posted on the challenge website. So far, two benchmarks have been offered,
namely segmentation-and-tracking benchmark (evaluating segmentation and tracking performance) and segmentation-only benchmark (evaluating purely segmentation
performance, no tracking part is required). A detailed description of the focus and history of the CTC can be found at http://celltrackingchallenge.net/ and in the new open-access
Nature Methods summary of the 10 years of its existence. The CTC is in constant evolution, and - as we did in the previous six editions attached to ISBI 2013-2015 and ISBI 2019-2021 - we plan to introduce some novelties in this new ISBI-sponsored challenge event.
Specifically, in this new 7th edition, the participants will be encouraged to submit further solutions to the recently opened generalizability tasks - either in the frame of the
segmentation-and-tracking benchmark (Task 1) or the segmentation-only benchmark (Task 2). The generalizability tasks focus on the development of methods that exhibit better
generalizability and work across most, if not all, of the existing datasets, instead of being optimized for one or a few datasets only. These tasks were established for the ISBI 2021 edition, and their first results were reported in the abovementioned paper, but no further results have been received since then.
Furthermore, a new tracking-only - more precisely linking-only benchmark (Cell Linking Benchmark) will be introduced to complement the segmentation-only benchmark for those who want to evaluate purely the object linking methods without having to supply segmentation results. Such a benchmark has been missing in the CTC portfolio and it is demanded by the CTC participants and the scientific community at large. Participants will be encouraged to supply ideally generalizable solutions (Task 3) working across 13 preselected datasets but will also be able to submit dataset-specific solutions (Task 4) for datasets of their choice.

Keywords

List the primary keywords that characterize the task.

cell detection, cell segmentation, cell linking, cell tracking, 2D time-lapse, 3D time-lapse

ORGANIZATION

Organizers

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

The Steering Committee (that is responsible for this proposal) of the Cell Tracking Challenge currently consists of the following researchers:
Michal Kozubek, Main Coordinator, Masaryk University, Brno (Czech Republic)
Alexandre Cunha, California Institute of Technology, Pasadena, CA (USA)
Martin Maska, Masaryk University, Brno (Czech Republic)
Erik Meijering, University of New South Wales, Sydney (Australia)
Arrate Muñoz-Barrutia, Universidad Carlos III de Madrid, Madrid (Spain)
Carlos Ortiz de Solórzano, Center for Applied Medical Research, Pamplona (Spain)
Tammy Riklin Raviv, Ben-Gurion University of the Negev, Beer-Sheva (Israel)
Johannes Stegmaier, RWTH Aachen University, Aachen (Germany)
Virginie Uhlmann, BioVisionCenter, University of Zurich (Switzerland) and European Bioinformatics Institute (EMBL-EBI), Hinxton (United Kingdom)
For a full list including collaborators, see http://celltrackingchallenge.net/organizers/

b) Provide information on the primary contact person.

Michal Kozubek, Masaryk University, Brno (Czech Republic)
email: kozubek@fi.muni.cz

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

Open call

Challenge venue and platform

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

ISBI 2024

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

our own accessed via https://celltrackingchallenge.net

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

https://celltrackingchallenge.net

Participation policies

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

only automatic methods are allowed

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.

Tasks 1 and 2: restricted to the data provided by the challenge (to be able to compare behaviour over silver truth, gold truth and gold+silver truth).
Tasks 3 and 4: no restrictions (to allow, e.g., pre-training on particle tracking datasets) but participants must declare which training data they used.

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

May participate but are not eligible for awards. Moreover, method developers must not take part in the evaluation of the submissions, must not participate in the generation of the reference annotations for the test data, and must not have premature access to the data.

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

The three best-performing award-eligible competitors for each task will be announced and awarded a diploma and possibly some small prize. If we manage to find a sponsor, the
prize would consist of a travel and/or registration grant to (partly) cover the expenses of the competitors with active participation in the challenge workshop.

e) Define the policy for result announcement.

Examples:

- Top 3 performing methods will be announced publicly.
- Participating teams can choose whether the performance results will be made public.

The performance of all methods (submitted according to the rules) for all tasks will be publicly announced on the challenge website.

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 authors of the top-3 methods for any of the four tasks will be invited to participate in a challenge-summarizing paper. This yields 12 teams with possible repetitions, we assume about 10 unique teams. This restriction is motivated by the fact that methods not included in the leaderboard are not visible on the website and are seldom used by others; moreover, authors of low-performing methods do not contribute to the advancement of the field and do not deserve co-authorship.
The number of co-authors per each top-3 team will be limited to three (four if the team is among top-3 in multiple tasks, i.e. has won several times). This restriction is motivated by the fact that method development is typically done by one or two people possibly supervised by a third one; this policy prevents the inclusion of a whole research group among co-authors.

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.

http://celltrackingchallenge.net/submission-of-results/ (both program/code and results need to be submitted)
For the new linking-only benchmark (Tasks 3 and 4), the same rules will apply. However, participants must submit only their program (code or a docker container) that works on
unseen data in the specified format. This requirement applies to both regular and generalizable solutions, and they must also provide training data results to ensure reproducibility by the organizers. Results for test data will not be required, as participants will not be able to supply them because they will not have access to the test data annotations needed for these two tasks. The new benchmark will evaluate linking methods on common segmentation outputs, including access to original gray-scale images. The common initial segmentation will be the silver-standard corpus for real datasets and the gold-standard corpus for computer-generated datasets, with anonymized cellular identities over time, which is provided for training but remains undisclosed for testing. The objective is to link cells (synchronize IDs, trace cell division lineages) and potentially correct initial segmentation errors (that result in temporal gaps between tracklets) to enhance lineage tree accuracy.
Participants may simplify segmentation masks to centroids and apply linking methods to detection inputs rather than segmentation.

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.

No pre-evaluation is allowed. All participants will have to submit by the deadline, and only the latest submission will be evaluated.

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

The website and registration (required to submit results) are already running. The datasets, including annotations (except for annotations for test data), are free to download (no new datasets will be added). The deadline for submitting results will be March 25, 2024. The received submissions will undergo a technical check by April 1, 2024, and participants will be given a chance to correct any potential problems with their submissions by April 6, 2024. The results will be announced during the ISBI 2024 challenge workshop.

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).

Not required (responsibility of the labs that provided us with the 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)

See Conditions of use at http://celltrackingchallenge.net/datasets/ that can be briefly summarized as follows:
1) It is required to acknowledge the source of data (CTC website + latest CTC paper).
2) CTC-related use does not require any permission.
3) Public non-CTC-related, scientific use requires explicit permission from the challenge organizers.
4) Commercial use requires explicit permission from the challenge organizers and the data providers.
5) Cloning of datasets or their parts, including reference annotations, is strictly forbidden.

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.

Command-line software packages and Fiji plugins that implement the evaluation measures are made publicly available, along with the instructions required to run the software.
The command-line packages are used for the official evaluation of the algorithms by the challenge organizers and can be used by the participants to evaluate and tune their
algorithms, too. For more details, please refer to the following section:
http://celltrackingchallenge.net/evaluation-methodology/

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

The participants have always been required to upload their code or executable for the organizers' verification purposes. Since August 1, 2023, we require that the solution
becomes publicly available in a reusable form before we publish the results on the CTC website. The instructions for making the solution reusable are available in this document:
public.celltrackingchallenge.net/documents/Guidelines%20for%20submission%20reusability.pdf

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.

A sponsor will be negotiated later on. The test data is freely available to everybody.
The test data labels (gold as well as silver reference) are accessible only to the challenge organizers.

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

Task category(ies)

State the task category(ies)

Examples:

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

- Tracking

Detection, Segmentation, Tracking

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.

Model organisms, human volunteers, or just standardized cell lines

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

Mostly just standardized cell lines

Imaging modality(ies)

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

Optical microscopy: fluorescence (Fluo), brightfield (BF), phase contrast (PhC) and differential interference contrast (DIC) modes - see the prefix of each dataset

Context information

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

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

NA

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

NA

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.

Manually selected fields of view with cells

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.

Individual cells

Assessment aim(s)

Identify the property(ies) of the algorithms to be optimized to perform well in the challenge. If multiple properties are assessed, prioritize them (if appropriate). The properties should then be reflected in the metrics applied (see below, parameter metric(s)), and the priorities should be reflected in the ranking when combining multiple metrics that assess different properties.

- Example 1: Find highly accurate liver segmentation algorithm for CT images.
- Example 2: Find lung tumor detection algorithm with high sensitivity and specificity for mammography images.

Corresponding metrics are listed below (parameter metric(s)).

Robustness, Recall, Precision, 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).

Various optical microscopes, see http://celltrackingchallenge.net/datasets/ for more details.

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

For pixel size and time step, see http://celltrackingchallenge.net/2d-datasets/ and http://celltrackingchallenge.net/3d-datasets/.

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.

- Baxter Laboratory for Stem Cell Biology, Stanford University, USA
- Erasmus Medical Center, Rotterdam, The Netherlands
- Centre for Integrative Infectious Disease Research (CIID), University Hospital Heidelberg, Germany
- Center for Applied Medical Research (CIMA), Pamplona, Spain
- Institute of Biophysics, Academy of Sciences, Brno, Czech Republic
- Mitocheck Consortium (led by EMBL, Heidelberg, Germany)
- Department of Bioengineering, University of California at Berkeley, Berkeley, USA
- Fraunhofer Institution for Marine Biotechnology, Lübeck, Germany
- Centre for Biomedical Image Analysis (CBIA), Masaryk University, Brno, Czech Republic
- Dept. of Biological Engineering, Massachusetts Institute of Technology, Cambridge, USA
- Waterston Lab, University of Washington, Seattle, USA
- Howard Hughes Medical Institute, Janelia Farms Research Campus, Ashburn, USA
- Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

See http://celltrackingchallenge.net/2d-datasets/ and http://celltrackingchallenge.net/3d-datasets/ for more details.

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).

Human experts with biological background

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 2D or 3D time-lapse image series (2D+t or 3D+t datasets) acquired using microscopy

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

4 for each of the 20 included datasets (out of which 13 have silver truth available that is required for all four tasks)

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

The videos are quite long and contain many cells (including many annotated cells), so the performance of the methods is averaged over many cells (hundreds). Two videos are provided for training and two for test purposes for each dataset to ensure the representativeness of both training and test data.

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.

Both training and test datasets have similar properties and contain cells of similar shapes, textures, or behavior.

Annotation characteristics

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

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

All annotations were created manually (with the exception of silver segmentation corpus). For both tracking and segmentation, three annotators from three different institutions
created the annotations (complete for tracking and > 100 cells in selected frames for segmentation) that were merged afterward using a majority voting scheme. The silver
segmentation corpus was created by merging the results of the best available automatic methods. Details were published in the Nature Methods 2023 paper: https://www.nature.com/articles/s41592-023-01879-y

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.

See Annotation procedure PDF document at http://celltrackingchallenge.net/annotations/

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.

Both training and test cases were annotated by people with a background in biology and microscopy experience after initial training. Parts of the annotations underwent quality
control by challenge organizers.

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.

For the gold corpus, computed from 3 human annotations, we used the majority voting scheme for both segmentation and tracking. For the silver corpus, we used a modified
majority voting scheme described in the Nature Methods 2023 paper:
https://www.nature.com/articles/s41592-023-01879-y

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.

None

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.

As for the tracking, the most frequently occurring annotation error is moving to a
neighboring cell between time points, i.e., tracking suddenly continues on another cell. For the segmentation, different annotators adopt different boundary recognition policies leading to less agreement on the particular shape of the cell boundary.
Inter-annotator variability was measured for each dataset. The inter-annotator variability can be found on the challenge website under the top-3 leaderboards:
http://celltrackingchallenge.net/latest-ctb-results/
http://celltrackingchallenge.net/latest-csb-results/

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

None

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)

Tasks 1 and 2: The same as in the previous editions, i.e., SEG (stands for segmentation), DET (stands for detection), and TRA (stands for tracking), see
http://celltrackingchallenge.net/evaluationmethodology/
Tasks 3 and 4: The technical metric will be a new measure LNK (stands for linking) defined analogously to DET: while DET is the detection part of TRA, LNK will be the linking part of TRA. All these measures (TRA, DET, LNK) are based on the Acyclic Oriented Graph Matching (AOGM) measure (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144959) normalized to the interval from 0 (worst) to 1 (best).
In addition, we will report four biological measures introduced earlier in the CTC (in the Nature Methods 2017 paper: https://www.nature.com/articles/nmeth.4473) for all tasks where relevant (Tasks 1, 3, 4) as additional information:
Complete tracks (CT) measures the fraction of reference cell tracks that a given method can reconstruct entirely from the frame in which they appear to the frame in which they
disappear. CT is especially relevant when a perfect reconstruction of the cell lineages is required.
Track fractions (TF) averages, for all detected tracks, the fraction of the longest continuously matching algorithm-generated tracklet with respect to the reference track. Intuitively, this can be interpreted as the fraction of an average cell's trajectory that an algorithm reconstructs correctly once the cell has been detected.
Branching correctness (BC(i)) measures a method's efficiency at detecting division events with a tolerance of i frames.
Cell cycle accuracy (CCA) measures how accurate an algorithm is at correctly reconstructing the length of cell cycles (that is, the time between two consecutive divisions). Both BC(i) and CCA are informative about the ability of the algorithm to detect cell population growth (that is, proliferation).
All of the biologically inspired measures take values in the [0,1] interval, with higher values corresponding to better performance. These measures are especially relevant for the new linking benchmark, but we decided to make them available also for the standard segmentation-and-tracking benchmark (for previously submitted methods, the values can be found in XLS files downloadable from the leaderboard page).
In order to keep a balance between technical and biologically inspired performance, we will first average all biologically inspired measures applicable to the given dataset (yielding summarizing BIO measure) and then average LNK and BIO measures for the given dataset (yielding OP_CLB measure that stands for the overall performance of the Cell Linking Benchmark).

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

All measures are chosen so that they are backward compatible with the currently available CTC results.

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.

The ranking scheme was chosen as backward-compatible (and old-benchmarks-compatible for the new linking benchmark).

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

Missing results (on the level of datasets for Tasks 1-3 or on the level of videos for all tasks) are not allowed (such submissions are not accepted). Missing cells (false negatives) are
allowed but highly penalized in all measures (directly for detection/segmentation or indirectly by spoiling the lineage trees for linking/tracking).

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

Generalizability tasks (Tasks 1, 2, 3): The overall performance of a particular method for a particular measure is obtained by averaging its performance over all datasets (and also
over six training data configurations for Tasks 1 and 2). The overall performance then yields the ranking for the given measure.
Dataset-specific task (Task 4): No global winner will be published on the CTC website. Global winners for the ISBI 2024 competition will be determined by a weighted sum of the method's occurrences in the top-3 leaderboard of the new linking benchmark (first place - weight 3, second place - weight 2, third place - weight 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.

Thanks to the high number of cells, it is possible to perform standard tests of statistical significance. Such tests will be performed similarly to those published in our recent Nature
Methods 2023 paper (https://www.nature.com/articles/s41592-023-01879-y).

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

Backward-compatible with the previous study published in Nature Methods

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.

Common problems/biases of the submitted methods will be studied

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.

N/A

Further comments

Further comments from the organizers.

N/A