# Light My Cells : Bright Field to Fluorescence Imaging Challenge 2024

Dorian Kauffmann, Emmanuel Faure, Guillaume Gay, Edouard Bertrand, Thomas Walter, Christophe Zimmer

# Light My Cells : Bright Field to Fluorescence Imaging Challenge 2024: Structured description of the challenge design

## CHALLENGE ORGANIZATION

### Title

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

Light My Cells : Bright Field to Fluorescence Imaging Challenge 2024

### Challenge acronym

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

lightmycells

### 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 Light My Cells France-Bioimaging challenge aims to contribute to the development of new image-to-image 'deep-label' methods in the fields of biology and microscopy.
The main task is to predict the best focus image of multiple fluorescently labelled organelles from label-free transmitted-light images.
In order to make them usable, the aim of this challenge is to produce new open source methods which can handle a large acquisition variability: Z-focus, multiple channels,
acquisition sites, input-modalities (Bright Field, Phase Contrast & Differential Interference Contrast or DIC), instruments, magnifications, cells and markers.
The high variability of the database is possible thanks to the structuring role of the France-Bioimaging national infrastructure, which federates 23 imaging acquisition sites
distributed all over France.

Biomedical point of view :
In order to obtain fluorescence microscopy images, it is necessary to perform a manual biochemical labelling treatment - time-consuming and costly - over cells with specific
fluorescent probes and dyes. But, the cells studied may themselves be perturbed by the fluorescence microscopy process, both by exposure to excitation light (phototoxicity) and by the probes themselves.
As phototoxicity increases with light exposure, it impairs long term imaging. Similarly, fluorophore dimming through photobleaching limits the signal-to-noise ratio of the images.
Furthermore, adding markers is an invasive method. The fluorophore might hinder its target's molecular interactions and protein overexpression increases its concentration in the
cytoplasm, disrupting regulation processes. Worse, the fluorophores themselves can be cytotoxic.

As fluorescence microscopy induces temporal and functional perturbations, it is thus crucial for live microscopy to limit the number of fluorescent probes used in an experiment.
On the contrary, label-free transmitted light microscopy such as bright field, phase contrast and DIC is non-invasive, phototoxicity is sharply reduced, and the signal quality is conserved throughout the acquisition.
The biological aim of this challenge is to recover fluorescence images in silico from bright field images.

Technical point of view:
We want to give a boost for multi-output deep learning methods based on a single input, when the training database is made up of images that do not always include all the required channels and have a high degree of variability (e.g. magnification, depth of focus, numerical aperture). This leads participants to develop in particular new architectures and loss functions dedicated for sparse output.
The purpose is to offer a tool for biologists that can be robust on any acquisition protocol and effective for the whole community, irrespective of the size of the images, cell line, acquisition site, modality or instrument. In order to assess the generalisability of the methods developed, we will exclude one complete acquisition site from the training database and leave it for the final evaluation. For the "Light my cells" challenge, we want to evaluate the ability of the methods to predict the best Z-focus plane for any organelle even in bad acquisition conditions. To achieve this goal, participants will have the possibility to perform data augmentation provided by the acquisitions.
It consists in large Z-stacks images of transmitted light microscopy containing a majority of out focus planes.
We defined metrics for each of the 4 organelles and for each (5) deviations of the focus plane to measure the ability to perform the task. We will evaluate each participant on this 4x5 metrics matrix, and the winners will be the ones with the best average of all the metrics. Moreover, participants will get an additional bonus for : code quality and accessibility,
lightweight deep learning model, short time of training and prediction, and evaluation of the carbon footprint.
Among the current state-of-the-art approaches for image-to-image tasks in bio-imaging are "DeepHCS: Bright-field to fluorescence microscopy image conversion using multi-task learning with adversarial losses for label-free high-content screening" (2021) and "Label-free prediction of cell painting from bright field images" (2022), both of which focus their methodologies solely on the use of the bright-field imaging modality, while "In Silico Labelling: Predicting Fluorescent Labels in Unlabeled Images" (2018) uses the same three modalities as our approach.
While "DeepHCS" (2021) and "In Silico Labelling" (2018) use a wide range of metrics to assess image quality, "Label-free prediction of cell painting" (2022) uses a more restricted set of metrics.
However, these previous works present a very low diversity of applications and do not provide an easily accessible database.
In addition, "DeepHCS" (2021) faces limitations due to the fixed sizes and specific dyes of its database,"In Silico Labelling" (2018) uses fixed formats that are not typical of those used in microscopy and similarly, the authors of "Label-free prediction of cell painting" (2022) admit limitations in the size and diversity of their database.
Nevertheless, a more extensive and publicly accessible JUMP-CP database exists for cell painting, which can be used for pretraining the 'Light My Cells' challenge.
Yet to the best of our knowledge, the desired methods for the 'Light my cells' challenge have no open source equivalent, and aspires to be rooted with an open database and open algorithms.

Impact :
We want to contribute to open science by making available training and testing databases with high variability,

even in the case of different acquisitions. In order to make the models accessible for any biologist, we will integrate the best open source methods into open-science image processing and analysis software (e.g. BioImage Model Zoo, Napari).
The "Light My Cells" challenge will become a reference for 'deep-label' methods for fluo-free-labelling bio-imaging. This is the first France-Bioimaging challenge, which will be followed by several others each year.

### Challenge keywords

List the primary keywords that characterize the challenge.challenge_

microscopy, image-to-image, cell painting, wide field, transmitted light, bright-field, contrast phase, DIC, fluorescence, cells, nucleus, mitochondria, tubulin, actin, deep-labelling, label-free, mammalian, in silico labelling.

### 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: lightmycells

## 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 Light My Cells France-Bioimaging challenge aims to contribute to the development of new image-to-image 'deep-label' methods in the fields of biology and microscopy.
The main task is to predict the best focus image of multiple fluorescently labelled organelles from label-free transmitted-light images.
In order to make them usable, the aim of this challenge is to produce new open source methods which can handle a large acquisition variability: Z-focus, multiple channels,
acquisition sites, input-modalities (Bright Field, Phase Contrast & Differential Interference Contrast or DIC), instruments, magnifications, cells and markers.
The high variability of the database is possible thanks to the structuring role of the France-Bioimaging national infrastructure, which federates 23 imaging acquisition sites
distributed all over France.

Biomedical point of view :
In order to obtain fluorescence microscopy images, it is necessary to perform a manual biochemical labelling treatment - time-consuming and costly - over cells with specific
fluorescent probes and dyes. But, the cells studied may themselves be perturbed by the fluorescence microscopy process, both by exposure to excitation light (phototoxicity) and by the probes themselves.
As phototoxicity increases with light exposure, it impairs long term imaging. Similarly, fluorophore dimming through photobleaching limits the signal-to-noise ratio of the images.
Furthermore, adding markers is an invasive method. The fluorophore might hinder its target's molecular interactions and protein overexpression increases its concentration in the
cytoplasm, disrupting regulation processes. Worse, the fluorophores themselves can be cytotoxic.
As fluorescence microscopy induces temporal and functional perturbations, it is thus crucial for live microscopy to limit the number of fluorescent probes used in an experiment.
On the contrary, label-free transmitted light microscopy such as bright field, phase contrast and DIC is non-invasive, phototoxicity is sharply reduced, and the signal quality is conserved throughout the acquisition.
The biological aim of this challenge is to recover fluorescence images in silico from bright field images.

Technical point of view:
We want to give a boost for multi-output deep learning methods based on a single input, when the training database is made up of images that do not always include all the required channels and have a high degree of variability (e.g. magnification, depth of focus, numerical aperture). This leads participants to develop in particular new architectures and loss functions dedicated for sparse output.
The purpose is to offer a tool for biologists that can be robust on any acquisition protocol and effective for the whole community, irrespective of the size of the images, cell line, acquisition site, modality or instrument. In order

to assess the generalisability of the methods developed, we will exclude one complete acquisition site from the training database and leave it for the final evaluation. For the "Light my cells" challenge, we want to evaluate the ability of the methods to predict the best Z-focus plane for any organelle even in bad acquisition conditions. To achieve this goal, participants will have the possibility to perform data augmentation provided by the acquisitions. It consists in large Z-stacks images of transmitted light microscopy containing a majority of out focus planes.
We defined metrics for each of the 4 organelles and for each (5) deviations of the focus plane to measure the ability to perform the task. We will evaluate each participant on this 4x5 metrics matrix, and the winners will be the ones with the best average of all the metrics. Moreover, participants will get an additional bonus for : code quality and accessibility,
lightweight deep learning model, short time of training and prediction, and evaluation of the carbon footprint.
Among the current state-of-the-art approaches for image-to-image tasks in bio-imaging are "DeepHCS: Bright-field to fluorescence microscopy image conversion using multi-task learning with adversarial losses for label-free high-content screening" (2021) and "Label-free prediction of cell painting from bright field images" (2022), both of which focus their methodologies solely on the use of the bright-field imaging modality, while "In Silico Labelling: Predicting Fluorescent Labels in Unlabeled Images" (2018) uses the same three modalities as our approach.
While "DeepHCS" (2021) and "In Silico Labelling" (2018) use a wide range of metrics to assess image quality, "Label-free prediction of cell painting" (2022) uses a more restricted set of metrics.
However, these previous works present a very low diversity of applications and do not provide an easily accessible database.
In addition, "DeepHCS" (2021) faces limitations due to the fixed sizes and specific dyes of its database,"In Silico Labelling" (2018) uses fixed formats that are not typical of those used in microscopy and similarly, the authors of "Label-free prediction of cell painting" (2022) admit limitations in the size and diversity of their database.
Nevertheless, a more extensive and publicly accessible JUMP-CP database exists for cell painting, which can be used for pretraining the 'Light My Cells' challenge.
Yet to the best of our knowledge, the desired methods for the 'Light my cells' challenge have no open source equivalent, and aspires to be rooted with an open database and open algorithms.

Impact :
We want to contribute to open science by making available training and testing databases with high variability, even in the case of different acquisitions. In order to make the models accessible for any biologist, we will integrate the best open source methods into open-science image processing and analysis software (e.g. BioImage Model Zoo, Napari).
The "Light My Cells" challenge will become a reference for 'deep-label' methods for fluo-free-labelling bio-imaging. This is the first France-Bioimaging challenge, which will be followed by several others each year.

### Keywords

List the primary keywords that characterize the task.

microscopy, image-to-image, cell painting, wide field, transmitted light, bright-field, contrast phase, DIC, fluorescence, cells, nucleus, mitochondria, tubulin, actin, deep-labelling, label-free, mammalian, in silico labelling.

## ORGANIZATION

### Organizers

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

Dorian Kauffmann - Challenge Project engineer, at France BioImaging Infrastructure (FBI)
Emmanuel Faure - CNRS researcher (CNRS-UM) & FBI.data mission officer
Guillaume Gay - Research engineer for the FBI.data project
Edouard Bertrand - Research Director and Scientific Director of FBI
Thomas Walter - Professor at Mines Paris and Director of the Centre for Computational Biology (CBIO)
Christophe Zimmer - Research Director at Institut Pasteur

b) Provide information on the primary contact person.

dorian.kauffmann@france-bioimaging.org

## Life cycle type

Define the intended submission cycle of the challenge. Include information on whether/how the challenge will be continued after the challenge has taken place.Not every challenge closes after the submission deadline (one-time event). Sometimes it is possible to submit results after the deadline (open call) or the challenge is repeated with some modifications (repeated event).

Examples:

- One-time event with fixed conference submission deadline
- Open call (challenge opens for new submissions after conference deadline)
- Repeated event with annual fixed conference submission deadline

One-time event with fixed submission deadline

## Challenge venue and platform

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

challenge results will be announced at ISBI2024

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).

https://lightmycells.grand-challenge.org/ (coming soon)

## Participation policies

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

Participants will use fully automatic methods based on supervised learning to predict a multichannel output image from the input image (image to image) without manual
intervention. In the context of "Light My Cells", supervised learning involves training models with paired input-output data (images) to learn the dependencies between them.
The use of private data is not encouraged and will not win the prize. To ensure fairness, entrants must share links to external data, datasets or pre-trained models (all freely available)
on the challenge forum by 21.03.2024; after this date, no additional data may be used for fair comparison.

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.

The database is licenced as CC-BY, i.e. everyone (also non-participants of the challenge) are free to use the training data set in their respective work, given attribution in the publication. For participation in the challenge, all publicly available data or networks are allowed.

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

As we take part into national institutions, every member of CNRS, France BioImaging, Montpellier University and LIRMM may participate in the challenge and be eligible for prizes.
Therefore, members of the organizers team and everyone involved in the conception or acquisition of the training and test datasets may participate in the challenge, but are not
eligible for prizes or the final ranking in the Final Testing Phase - to avoid potential conflict of interest.

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

Award for top 3 winners:
- award certificate
- A challenge paper will be written with the organizing team's members for submission to journals
- invitation to publish their methods in the proceedings of the IEEE International Symposium on Biomedical Imaging 2024
- support and integration of open source code into open science image processing and analysis software (e.g. BioImage Model Zoo, Napari)

For the 1st:
- Invitation to 2024 France-Bioimaging annual meeting
- all paid hotel, food, plane (in economic cost and reserved at least 3 months in advance)
- graphic card (e.g. NVIDIA GeForce RTX 4070)
- android tablet

For the 2nd:
- graphic card (e.g. NVIDIA GeForce RTX 4070)
- android tablet

For the 3rd:
- android tablet

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.

Participating teams have the option to decide if their performance results will be disclosed publicly. However, only those teams that share open source code and algorithms will qualify for winning awards.
The top three teams, as well as potential other original methods developed, will be announced publicly at ISBI and

recognized as co-authors in the challenge document.

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).

Entities, including researchers and companies seeking to compare their AI models or products without entering into the competition, participating teams and individuals intending to use the challenge, are not permitted to publish until the challenge document is published (expected by the end of 2024).
Use of the database is free, but publication of an article on the use of the database is subject to prior publication of our data article.
After publication of our articles (both data and challenge), all parties are free to publish their results, but must cite the appropriate publication.
As the data will remain public, we plan to produce an article three years after the publication of the challenge paper with the new methods published.

### 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.

All participants will perform local training (and validation) before submitting to the test phases.
For both phases, participants will submit their algorithm in docker format to Grand-challenge.org, where metrics will automatically evaluate predictions on secret test datasets and rank participants.

When submitting, participants will have two choices:
- Submit open source code (Recommended)
- or only submit the inference part including the saved trained model (in h5 format) (in which case it will not be eligible for prizes)

In all cases, algorithms must predict and produce the best fluorescence Z-focus image of all four organelle channels with the same image size as the input image and in OME-TIFF format.

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.

We have defined the challenge as a single task with two phases:
- a preliminary test phase (on 30 images) to familiarize with the algorithm submission procedure, with the possibility to have five submissions (with a maximum of one by week)
- the final test phase (on 300 images) with only one submission accessible will not give the possibility to evaluate their algorithms before submitting.

## 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 challenge and website page opening will be officially launched on 08/01/2024, with training database releases scheduled for 15/01/2024.
Registration will be open from 08/01/2024 to 20/02/2024.
Submissions will follow two phases:
- Phase 1: From 15/02 to submission closure on 21/03.
- Final phase: From 15/04 to 30/04.
Results will be announced on 15/05 (allowing time for winner notifications), with associated workshop days potentially at ISBI on 27/05.
Paper publication is expected by the end of 2024.

## 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).

No ethics approval is necessary for the database used in this context.

## 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)

The database will be made publicly available as soon as it is accessible during the challenge.
However, there is an embargo on publishing any findings or results derived from these data until we publish the paper on this database.

After this, it will be necessary to cite the publication of the article on the database.
All will be released under the CC-BY-4.0 license.

### 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.

All metrics used for the evaluation will be open-source and accessible.
However, the evaluation software will run automatically on Grand-Challenge.org on the ground truth images for the test phases.
There are no links yet, but there will be soon.

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

All dockers will be stored and accessible on Grand-Challenges.org.
In addition, we will suggest that participants provide a link (which will be accessible on our challenge page on Grand-Challenges.org) to their source code in a public repository (e.g.
GitHub, Gitlab, etc.).
For non-free software, we cannot provide an accessible link to their code, but we will link to the corresponding company.

### 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.

There are no conflicts of interest to declare.
The challenge received specific funding from France BioImaging.
Access to test case labels during the preliminary phase is restricted to the organizing team. At the end of this phase, these labels will be accessible to all and included in the training dataset.
Before and during the final phase, access remains restricted to the organizing team until the end of the challenge, after which it will be open to all.

## 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

Biology, Microscopy

### Task category(ies)

State the task category(ies)

Examples:

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

Prediction

### 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 for the final biological and biomedical application comprises any type of cells observed through various microscopes among three transmitted light microscopy (bright field, phase contrast, and DIC) and various probes and dyes (GFP, Hoescht, Draq5, mCherry, MitoTracker Red and Green, Far Red, SIR, E83 cherry, MISP cherry and Spy555) in 2D (plated) cultures.

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

The challenge cohort comprises the following cell lines :
Cells:
- HeLa human cells
- HeLa stably expressing RAB6A fused with EGFP
- HeLa cells expressing stably histone H2B fused to EGFP
- Neurons from hippocampus of E18 rat embryos
- C4-2B human prostate cancer cells
- U2OS and HEP3B human liver carcinoma cells
- T24 Urinary bladder transitional human cell carcinoma
- MDCK dog cells expressing stably the EB3 protein fused to Cherry
- U2OS human cells stably expressing histone H2B fused to EGFP

### Imaging modality(ies)

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

Bright field, phase contrast and DIC and Epifluorescence microscopy.

### Context information

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

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

Image data are provided with metadata including sample preparation specifics, details of the imaging instrument and methodology used, channel content and information, NCBI
taxonomic classification and attribution to the sites/platform involved in image generation.

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

We provide a cell line reference.

### 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.

Not relevant for our challenge.

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.

Prediction of fluorescence images of 4 different organelles : Nucleus, Mitochondria, Tubulin, Actin

### 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)).

Predict the multichannel fluorescence images associated with the input-image, with the

## 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).

Microscopes:
- Inverted Zeiss Axio Observer 7
- Inverted Zeiss Axio Observer stand equipped with environmental control (37C and CO2) controlled by ZEN software
- Inverted Nikon Ti2 microscope
- Inverted Nikon Ti eclipse with NIS software and environmental control
- Inverted Leica DMi8 stand equipped with environmental control (37C and CO2) controlled by Inscoper software
- Leica HC PL APO CS2
- Inverted Leica DMI8 video-microscope
- Inverted Leica DMi8 stand equipped with environmental control (37C and CO2) controlled by Metamorph software
- Home-made setup equipped with an ORCA Flash 4 sCMOS (Hamamatsu) and a 63x Olympus phase contrast objective

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

Chosen model: Mammalian cells (human, rodent, etc.)
- Observation of different subcellular structures on a cellular scale
- Experimental conditions: LIVE cell culture (no fast phenomena)
- Wide field microscope
- Number of acquisition sites : 10
- Number of participants : 21
- Number of sample preparation by person : roughly 3
- Number of images / sample : roughly 45 (capturing several views from the same plate)
- Objective/Pixel size : free from 40x (60 nm/px) to 100x (300 nm/px)
- Type of imaging :
- 2D with few z slides with different focus (Delta z free)
- 1 stack with 2 (or more) channels (around 10 to 20 planes):

Channel :
1. Bright Field, Phase Contrast and/or DIC
2. Fluorescence imaging of Nucleus marker
3. (optional) Fluorescence imaging of structures (in order of preference):

- Mitochondria (ex: mitotracker or fluorescent protein)
- Tubulin (ex: SIR-tubulin or fluorescent protein)
- Actin (ex: SIR-actin or fluorescent protein)
- Multiple microscopes are encouraged (even for the same plate)
- Multiple objectives and magnifications are also welcome
- Warning : Minimal movement shift between fluorescent and phase contrast
- No time-lapse required
- No SPECIFIC wavelength (to be kept in the metadata)
- Format: OME-TIFF
- Metadata:
- Instrument metadata (objective magnification and NA, physical pixel size ...)
- Sample metadata (species, lineage, staining / contrast method, etc.)
- REMBI

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.

France BioImaging Platforms:
- Montpellier Ressources Imagerie (MRI) - Montpellier
- Centre de Recherches de Biochimie Macromoléculaire CRBM - Montpellier
- Centre de Biochimie Structurale(CBS) - Montpellier
- Institute of Human Genetics (IGH) - Montpellier
- L'Institut Génétique & Développement de Rennes (IGDR) - Rennes
- Institut de génétique et de biologie moléculaire et cellulaire (IGBMC) - Alsace
- Genethon - Île-de-France
- Light Imaging Toulouse CBI (TRI-LITC) Toulouse
- Bordeaux Imaging Center (BIC) - Bordeaux
- Institut Gustave Roussi - Île-de-France

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).

Platform engineer with expertise in microscopy and biological sample preparation.

### 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.

Both training and test cases represent a transmitted light microscopy image (with three possible modalities: bright field, phase contrast and DIC) combined with the equivalent
fluorescence microscopy image with the best focus for the specified organelles. For each image in both cases, acquisition metadata will also be supplied with the database.
However, whereas the training cases have the whole Z-stack available (as a natural augmentation of the data), the test cases consist only of single 2D images carefully selected to
represent variability. In order to predict the optimal Z-focus, we will not give the full stack during the test phases.

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

As our database is made up of roughly 10 000 new, previously unpublished 2D images.
95 % of the database will be dedicated to the training data set. The remaining 5% will be allocated as follows: roughly 10% for the preliminary test phase - and added to the training database at the end - and the remainder (90%) for the final test phase.

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

We've adopted a standard proportion approach for data spitting outlined in the paper "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil" (2021).

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 testing database will be carefully selected in order to represent all types of variabilities. To this end, it will contain few random partitions in the selected representative variability and unseen data: one acquisition site will be reserved for testing all the modalities during the final phase.

### Annotation characteristics

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

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

No human annotations: ground truth is performed by in vivo acquisition

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.

None (cf item 21-b as we have ground truth performed by in vivo acquisition )

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.

None (cf item 21-b as we have ground truth performed by in vivo acquisition)

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.

None (cf item 21-b as we have ground truth performed by in vivo acquisition)

### 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 pre-processing involved converting all data (images) into OME-TIFF files, a standardized microscopy format containing associated metadata. This conversion was accomplished using the python library bioformat, following the guidelines outlined in "Metadata matters: access to image data in the real world" (2010).
Furthermore, we implemented a method to select the optimal focusing plane for the fluorescent images, guided by the techniques detailed in the paper "Analysis of focus measure operators for shape-from-focus" (2013).

### 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.

Potential sources of error in image annotation may include small movements within Z-stacks or between transmitted light and fluorescence images, notably due to cell culture dynamics.
However, to the best of our knowledge, these errors appear to have minimal impact and do not significantly affect the "annotations". These problems may not show significant variations between training, validation or test cases.

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

Due to the unbalanced diversity of the database, participants can face model overfitting. In addition, due to sparse distribution of the image output (e.g. some images contain only nuclei and actin channels of the 4 required outputs), participants will need some high level knowledge to tune their loss functions.

## 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)

We defined metrics according to the 4 organelles and (0 to 5) defined deviations of the focus plane to measure the ability to perform the task. We will evaluate each participant on this 4x6 metrics matrix, and the winners will be the ones with the best average over all the metrics. We will use standard metrics that are used in the field to evaluate the submissions:
- Structural Similarity Index Measure (SSIM) of predicted and ground truth images.
- Pearson Correlation Coefficient (PCC) of predicted and ground truth images.
- Mean Absolute Error (MAE) of predicted and ground truth images.
- Euclidean and cosine distances between original and ground truth images (textures metrics).

For all deviation of the focus plane, we choose to evaluate the 4 organelles as follows:
- Nucleus and mitochondria: all metrics

- Actin and tubulin: only SSIM and PCC

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

We chose the MAE as it is less sensible to aberrations than the Mean Square Error (for which the square term enhances error from high intensity signal). PCC is frequently used in the
state-of-the art methods and is linked to pixel intensity, which makes sense here for fluorescence images with black background.

The Structural Similarity Index (SSIM) stands out as a metric closely aligned with human vision, factoring in luminance, contrast, and structure. Notably, it is more efficient (than FSIM e.g.) when applied to greyscale images compared to RGB images. SSIM makes sense too as both PCC and MAE, can also be good if the image is very blurred.

And PCC has, with SSIM, only an overall similarity: everything is equally important. Therefore, PCC does not take into account variations in structure or texture.
Then, we will evaluate texture features that are usually used for traditional cellular phenotyping. As there will be no manual annotation, we can only calculate them at the
image level, but they will still provide us with a useful additional metric in phenotypic space. Resulting from this, we choose two additional metrics: the Euclidean and the cosine distance between original and ground truth images.

We choose to use only SSIM and PCC for actin and tubulin evaluation because, as they are not visible to the naked eye on transmitted light images, we risk having a lot of "fuzzy" and therefore not being really sure of the meaning of texture metrics and MAE (even if it's better than MSE).

We have defined these different evaluations to best suit the organelles chosen. Even if we evaluate organelles individually according to their offset from the best focal plane, our
objective remains to identify the best algorithm in terms of generalization and adaptability to variability.

### Ranking method(s)

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

We will provide both an overall ranking by averaging these scores described and individual rankings.
Anyone can take part in the challenge. Participants will be displayed in order of scores, and winners will be determined by the highest overall average.

Scores will evaluate all participants, regardless of code type and model weight and availability, so as not to limit participation by companies even if they do not wish to distribute
their code for intellectual property or commercial reasons.

In addition, participants can earn bonus points for aspects such as code quality and accessibility, use of lightweight deep learning models, shorter training and prediction times,

and consideration of carbon footprint assessment.
Of course, the bonuses will have no impact on the final ranking, but the methods with the most bonuses will have their place and will be described in the article.

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

For each test phase, missing results will be taken into account (in the average score) with the lowest metric values. However, during the first preliminary test phase, we offer participants the possibility to submit 5 times their algorithms in order to avoid technical issues.

Of course, it's essential to encourage open communication between the organizers (ourselves) and the participants in order to resolve any problems contributing to the lack of results, if not irrelevant. To this end, a forum will be set up on the challenge web page.

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

We have chosen to independently measure the results of each deviation for the focus plane (see item 26). Thanks to this approach, we will have access to the relevance and detail
of every output of every method, even with missing outputs.
We have chosen to add a bonus at the end to promote smart work and GreenIT, as we want to be rooted in open science and sustainability.

### 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.

When we start to have results, we will carry out statistical analysis of the methods according to the following criteria:
- data variability
- Z-focusing, multiple channels, acquisition sites, input-modalities (Bright Field, Phase Contrast & DIC), instruments, magnifications, cells and markers
- computing times
- number of (light)weights
- bonus items:
- number of lightweights,
- consideration of carbon footprint
- computing time
- libraries (PyTorch, TensorFlow, bioformat, etc.) & programming language (Python, etc.) used
- ability to overcome the partial processing of the desired data, enabling the algorithms to learn despite this lack.
- gap of the deviation above or below cells

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

The statistical methods used were selected according to the current state of the art and the most important criterion, namely the variability of the data, which were chosen on the basis of real acquisition use cases.

### 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.

During the challenge, we plan to explore assembly techniques, assess inter-algorithm variability, evaluate biases in submitted methods and analyse ranking variability. The aim of
these analyses will be to deepen our understanding of the performance of the submitted methods and improve the overall evaluation.

## 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