Annotation procedure

Basic terminology:

- **Object** = a cell or cell nucleus (depending on the dataset)
- **Frame** = a 2D or 3D image of objects at one time-point
- **Video** = a time sequence of frames
- **Field of interest** = frame definition domain eroded in the lateral axes by a constant number of pixels to better tackle objects entering or leaving the field of view.

Each video available for either training or competition was annotated once at three different places by independent experts to better understand inter-expert variability. For each video at each place, an expert created two separate ground truths (GTs), one for evaluating SEG measure and one for evaluating DET and TRA measures. The main reason why we separate ground truths for DET, SEG, and TRA is that each object must be labeled over all time-points to correctly evaluate DET and TRA measures but only a representative subset of all objects is sufficient to correctly evaluate SEG measure. At the same time, the segmentation accuracy is much more critical for evaluating SEG measure than DET and TRA measures, especially near object contours. Therefore we have developed different tools for creating detection, tracking and segmentation GTs. The annotation strategies are described below.

Field of interest specification

In order to better tackle objects entering the field of view, a frame definition domain was virtually eroded in the lateral axes (x and y) by a constant number of pixels (voxels) E depending on a dataset. The border of the eroded domain was available to annotators as an overlay over the frame definition domain. The considered values of E were following:

- **E = 50** for the datasets DIC-C2DH-HeLa, Fluo-C2DL-MSC, Fluo-C3DH-H157, Fluo-N2DH-GOWT1, Fluo-N3DH-CE, Fluo-N3DH-CHO, and PhC-C2DH-U373,
- **E = 25** for the datasets Fluo-C3DL-MDA231, Fluo-N2DL-HeLa, and PhC-C2DL-PSC,
- **E = 0** for the simulated datasets (i.e., Fluo-C3DH-A549-SIM, Fluo-N2DH-SIM+ and Fluo-N3DH-SIM+) and for the datasets Fluo-C3DH-A549, Fluo-N3DL-DRO, and Fluo-N3DL-TRIC.

Ground truth for DET and TRA measures

The main requirement on the tool used for the preparation of the detection and tracking GT was to speed up the manual object detection and tracking as much as possible without any impact on the DET and TRA measures. This was achieved by object markers, which were manually placed for each appearance of an object in the field of interest. We use the following terminology:

- **Marker** = a set of pixels (voxels) of the same label in a frame related to a particular object. We ensure that there is at most one marker of the same label in each 2D frame. In 3D image data, the marker does not have to necessarily be formed of a single connected component.
- **Track** = the longest time sequence of markers of the same label without time gaps. Each track is represented by the label of its markers. We ensure that each single track has a unique label.

The annotators followed these instructions:

- Draw a marker for each object that is fully or partly visible within the field of interest. It is allowed to draw the marker (or only some of its pixels/voxels) outside of the field of interest. The field of interest is just used to decide what objects to mark.
- If an object disappears (either leaves the field of interest or is not visible in the frame at all for a couple of time-points due to poor signal) do not put a marker for the time-points where the object was invisible. If it is clear that the same object reappears in the field of interest, link its track to the correct parent marker.
- Keep the same label until the tracked object disappears or the last time-point is reached. If the object divides select any of its daughters and continue with tracking without changing the label. Note that with this strategy the labeling is temporarily inconsistent with challenge requirements (because the parent object has the same label as a daughter), but this causes no problems because a track is automatically split into two when any of its non-end object markers is selected as a parent.
- If appropriate, link a track (its first marker) to the correct parent marker.
- In datasets with nuclear labeling, track only nuclei. Do not put a marker for cells with invisible nucleus but use the information to correctly link the tracks.
- Fluo-N2DL-HeLa: If, due to an abnormal mitosis, the daughter cells remain fused at all times, the cluster of fused daughter cells will be labeled and tracked as one single object.
- Fluo-N3DH-CE: Annotate as much frames as you can for the training data. Annotate all frames for the challenge data.
- PhC-C2DL-PSC: Annotate only the frames with temporal indices ranging from 150 to 250 for the training data. Annotate all frames for the challenge data.
- Fluo-N3DL-DRO: Annotate only the lineages of cells that form the neural tube during an early embryogenesis of the fruit fly (*Drosophila melanogaster*).
- Fluo-N3DL-TRIC: Annotate only the lineages of cells in the blastoderm of the beetle embryo (*Tribolium castaneum*), which are at the border of the embryonic and extra embryonic tissues. These lineages can either end up in the embryo or form the serosa.

Ground truth for SEG measure

The main goal of manual segmentation is to obtain representative ground truth for the evaluation of segmentation accuracy. The task for annotators was to mark pixels belonging to objects as accurately as possible. Each object was segmented as a set of pixels with the same unique label. In opposite to TRA measure, it was allowed to enter non-continuous sets. The annotators could use two simple tools: (1) drawing/erasing closed contours with filling the surrounded area, and (2) drawing/erasing pixels using a pen-like tool with a possibility to change the stroke size.

For each video we have randomly permutated all its frames. For each 3D frame, we also randomly selected at least one of its 2D z-slices which contained some objects (empty slices were excluded).

The annotators followed these instructions:

- Segment all objects within each frame in the given random order until at least 100 objects are segmented and at least 2 frames are fully segmented.
- After reaching the given limit inspect the rest of images in the given random order and additionally segment all difficult, somehow interesting or previously unseen object phenotypes, which can be expected to cause segmentation problems. Try to segment at least 20 instances of each interesting phenotype if possible.
- For each segmented object draw the segmentation mask in the whole field of view (image domain), not just in the field of interest. Do not segment objects visible only outside the field of interest.
- In case of overlapping objects, segment the visible parts of each object.
- Try to keep the contours of the segmented objects as close as possible to real visible object contours.

Instructions specific for some datasets:

- In datasets with nuclear labeling, segment only nuclei.
- DIC-C2DH-HeLa: Do not follow spikes that sometimes extend from the cytoplasm.
- Fluo-C2DL-MSC: Filopodial areas with low staining will be considered part of the cytoplasm.
- Fluo-N2DL-HeLa: Segment all non-interphase nuclei in the first 5 frames. If, due to an abnormal mitosis, two daughter cells remain fused at all times, the cluster of fused daughter cells will be labeled as one single object.
- Fluo-N3DH-CHO: Nuclei should not have holes. This means the low-staining nucleoli will be considered as part of the nucleus. However, possible background areas located inside the nucleus due to abnormal, amoeboid nuclear shapes will be considered background.
- Fluo-N3DL-DRO: Segment only those nuclei that are highlighted by tracking markers. However, in contrast to the other datasets, all the highlighted nuclei need to be segmented irrespective of their number.
- Fluo-N3DL-TRIC: Segment only those nuclei that are highlighted by tracking markers. However, in contrast to the other datasets, all the highlighted nuclei need to be segmented irrespective of their number.
- Fluo-C3DH-A549: Segment whole cell volumes, including filopodial protrusions, fully in 3D for the randomly selected frames.