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-lapse sequence of frames
- **Field of interest** = a frame definition domain eroded in the lateral axes by a constant number of pixels/voxels to better tackle objects entering or leaving the field of view.

Each video available for either training or testing was annotated once at three different institutions by independent annotators to better understand inter-annotator variability. For each video, each annotator created two separate manual annotations, one for evaluating the SEG measure and one for evaluating the DET and TRA measures. The main reason why we separate annotations for individual measures is that each object must be labeled over all frames to correctly evaluate DET and TRA, but only a representative subset of all objects is sufficient to correctly evaluate SEG. At the same time, the segmentation accuracy is much more critical for evaluating SEG than for evaluating DET and TRA, especially near object contours. Therefore, we have developed different tools for manually creating detection, tracking, and segmentation annotations. Furthermore, to further facilitate the tuning of competing algorithms for some of the datasets, silver reference segmentation annotations are provided. 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 DIC-C2DH-HeLa, Fluo-C2DL-Huh7, Fluo-C2DL-MSC, Fluo-C3DH-H157, Fluo-N2DH-GOWT1, Fluo-N3DH-CE, Fluo-N3DH-CHO, and PhC-C2DH-U373,
- **E = 25** for BF-C2DL-HSC, BF-C2DL-MuSC, Fluo-C3DL-MDA231, Fluo-N2DL-HeLa, and PhC-C2DL-PSC,
- **E = 0** for the computer-generated datasets (i.e., Fluo-C3DH-A549-SIM, Fluo-N2DH-SIM+, and Fluo-N3DH-SIM+) and for Fluo-C3DH-A549, Fluo-N3DL-DRO, Fluo-N3DL-TRIC, and Fluo-N3DL-TRIF.

Gold reference detection and tracking annotation

The main requirement on the tool used for the preparation of the manual detection and tracking annotation was to speed up the manual object detection and tracking process as much as possible without any impact on the DET and TRA measures. This was achieved by manually placing object markers for each occurrence 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 be necessarily formed of a single connected component, allowing annotators to create quintessential volumetric markers.

- **Track** = the longest temporal sequence of markers of the same label without temporal gaps. Each track is represented by the label of its markers. We ensure that each single track has a unique label.
- **Track link** = a logical connection between two occurrences of the same object in successive frames
- **Parent link** = a logical connection between two tracks, corresponding to either mother-daughter relationships in case of division events or the object reappearance after a temporal gap.

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 used only to decide which 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 because of poor signal or being completely overlaid by other object) 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 its predecessor using a parent link.
- Keep the same label until the tracked object disappears or the last time-point is reached. If the object divides, track its daughters as new objects and link their initial markers to the terminal marker of their mother using parental links. If, due to an abnormal division, the daughters fuse into a single object afterwards, do not track them as new objects, but rather consider them as the continuation of the mother object without any division event.

Instructions specific for some datasets:

- In datasets with nuclear labeling, track only nuclei. Do not put markers for cells with invisible nuclei, 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 videos. Annotate all frames for the test videos.
- 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, Fluo-N3DL-TRIF: 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. Note that the image data was acquired using a light-sheet microscope and fused from multiple viewpoints. Some views in Fluo-N3DL-TRIF may not be registered perfectly (resulting in display of clearly false, outstanding nuclei) or the views altogether do not have to cover the full volume (resulting in an artificial localized dark patch that stretches along one image axis). Avoid, therefore, annotating the lineages that are or can be affected by these imperfections. In Fluo-N3DL-TRIC, the mentioned artifacts can also be found but to a substantially smaller extent.
- BF-C2DL-HSC, BF-C2DL-MuSC: Do not track cells that appear outside the central well.

Gold reference segmentation annotation

The main goal of manual segmentation is to obtain a representative gold-standard corpus (denoted as gold truth) for the evaluation of segmentation accuracy. The task for annotators was to mark pixels/voxels belonging to objects as accurately as possible. Each object was segmented as a set of pixels/voxels with the same unique label. The annotators could use two simple tools: (1) drawing/erasing closed contours with filling the surrounded area, and (2) drawing/erasing pixels/voxels 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 that 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 are considered a part of the cytoplasm.
- Fluo-N2DL-HeLa: Segment all non-interphase nuclei in the first five frames in the given random order. If, due to an abnormal mitosis, two daughter cells remain fused at all times, label the cluster of fused daughter cells as one single object.
- Fluo-N3DH-CHO: Nuclei should not have holes. The low-staining nucleoli are considered a part of the nucleus. However, possible background areas located inside the nucleus due to abnormal, amoeboid nuclear shapes are considered background.
- Fluo-N3DL-DRO, Fluo-N3DL-TRIC, Fluo-N3DL-TRIF: Segment only those nuclei that are highlighted by tracking markers. However, in contrast to other datasets, all highlighted nuclei need to be segmented irrespective of their number. Remember that the image data was acquired using a light-sheet microscope and fused from multiple viewpoints, therefore a visible blur is not necessarily linked to the axial direction.
- Fluo-C3DH-A549: Segment whole cell volumes, including filopodial protrusions, fully in 3D for the randomly selected frames, without the necessity of segmenting at least 100 objects.
- BF-C2DL-HSC, BF-C2DL-MuSC: The inner halves of dark regions around cells are considered a part of the cytoplasm.

Silver reference segmentation annotation

The main goal of a silver-standard corpus (denoted as silver truth) containing computer-generated segmentations is to obtain a denser set of cell segmentation masks than it is practicable to create manually. To this end, the algorithms of 39 challenge participants, who submitted their results before October 1st, 2020 and explicitly approved the usage of their results for this purpose, were analyzed for their performance. Up to 16 top-performing algorithms that simultaneously achieved the SEG and DET scores over the halves of the reported human performance for both the training and test datasets were selected per dataset. Their results were merged using a modified version of the label fusion approach to deal with clustered objects, producing quantitatively more reliable segmentation results than the SIMPLE and STAPLE algorithms.