Annolid: Annotate, Segment, and Track Anything You Need

512.02

Chen Yang, Matthew Einhorn, Jeremy Forest, and Thomas A. Cleland
Computational Physiology Lab, Dept. of Psychology, Cornell University, Ithaca, NY USA

Computational Physiology Laboratory

The problem and its existing solutions

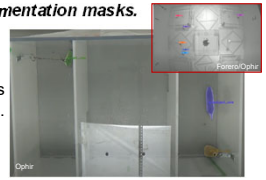
- Quantitative studies of animal behavior can be greatly facilitated by computational tools that partially or fully automate **behavior analysis from video recordings**. Such tools can greatly improve the efficiency of analysis, and reduce the potential for bias in human coders.
- However, natural behavior is rich and complex. Scientific questions of interest often require the development of new, customized behavioral tasks. **Analysis methods need to be able to embrace this complexity and adapt to the needs of diverse tasks, rather than constraining the behavioral task to the capabilities and limitations of the tool.**
- Deep learning-based strategies are particularly promising. The most prominent deep learning-based behavior analysis packages to date [7,8] are designed for the study of motor control, and use a ***pose estimation*** strategy based on keypoint tracking. This is a powerful approach, but does not always generalize to the many off-label use cases to which it has been applied.

Annolid is based on *instance segmentation*

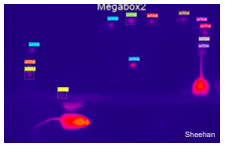
- We present **Annolid**,* an open-source software package for animal behavior analysis, based on a deep learning strategy of ***instance segmentation***.
- Instance segmentation is tremendously flexible, enabling a wide variety of automated behaviour assessments:

The fundamental object is an ***instance*** – a subject or object of interest in a video that is defined by the end user.

The most common approach is to define individual animals as instances. Here, two voles are outlined with ***segmentation masks.***



By leveraging ***overtraining***, multiple animals can be reliably and separately tracked.

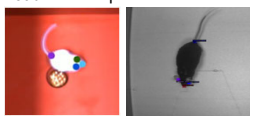


In contrast, broader training yields a more ***generalized*** instance (e.g., "mouse", or "urine spot"), enabling tracking of animals or elements not included in the training set.

Deformations of the segmentation mask can be used to define identified behaviors, such as *rearing*, *grooming*, *investigating*, or more specialized behaviors such as *huddling* in pups.

See poster 512.01 / XX32 - Scoring rodent digging behavior with *Annolid*

Instances also can be defined more narrowly – e.g., as *portions* of an animal. Here, focusing on the shape of the *head* mask improves the identification of digging behavior.



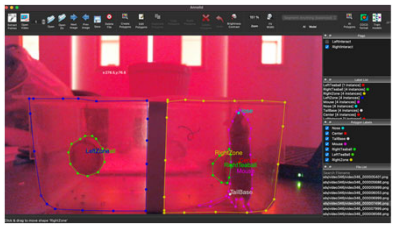
Still more narrowly, *keypoint tracking* (and pose estimation) can be performed as a special case of instance segmentation.

*Annolid is a hilarious portmanteau of "annotate" and "annelid", referring to segmentation.

Annolid workflow: segmentation and labeling

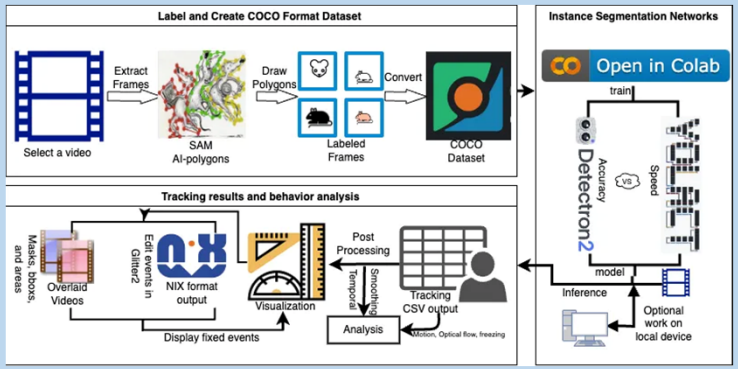
- From an experimental video, users employ Annolid's LabelMe-based GUI to identify instances and tag behaviors of interest, within a limited number of selected frames of video.

  Instances can be defined using polygons, instance objects, object parts, semantic segmentations, bounding boxes, keypoints and event timestamps.



- Meta's **Segment Anything Model** (SAM) also is now directly supported in Annolid. This tool automatically segments visual objects with polygons, greatly reducing end-user workload.

- A human-in-the-loop strategy for iterative updating of masks and labels via direct editing is also supported.
- The result is a COCO-formatted data file (*Common Objects in COntext*, an annotation standard)



Annolid workflow: training segmentation models

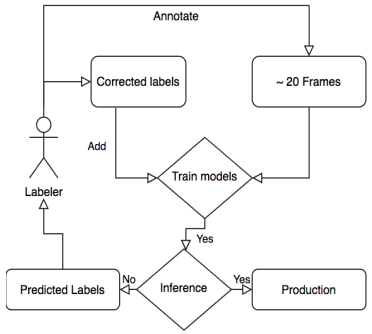
- Annolid model training and inference can be performed either locally or in the cloud (Google Colab), using either a YOLACT/YOLACT++ network for greater speed, or a Mask R-CNN network using Meta AI's Detectron2 platform for greater accuracy.

colab
Detectron2

- Different behavior analysis problems can require very different numbers of labeled frames to achieve the desired accuracy. Annolid's human-in-the-loop strategy can minimize researchers' labeling effort.

  First, label ~20 frames and train a model. Then, use the trained model for *inference* on the full video, visually identifying outcomes such as segmentation masks and behaviors.

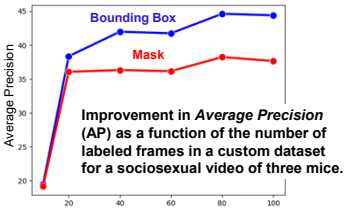
  Then, identify key errors and add or correct the labels on those error frames. Add the corrected predictions to the training set and train another model. Repeat until the model reaches desired performance.



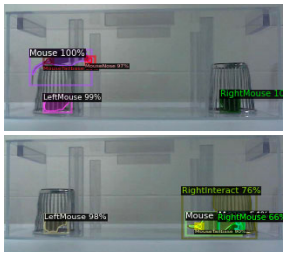
Tracking and Evaluation Results

Tracking metrics for a video featuring two voles interacting:

Total number of Frames in the video: 10575
Total labeled frames for training: 368
ID switches: 52 (52/10575=0.49%)
False negatives: 38 (38/10575=0.36%)
False positives: 11 (11/10575=0.1%)



Improvement in *Average Precision* (AP) as a function of the number of labeled frames in a custom dataset for a sociosexual video of three mice.

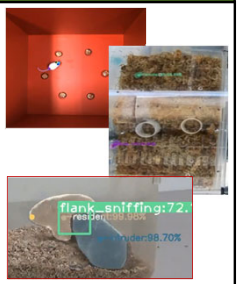


More information

More information:
https://cplab.science/annolid
https://cplab.science/annolid-intro

Open-source software:
https://cplab.science/annolid-git

Sample videos:
https://cplab.science/annolidtube



References & Acknowledgements

Supported by NIH/NIDCD grants R01 DC 019124 and R01 DC 014701.

Thank you to all beta testers and early adopters of Annolid, some of whose work is illustrated above:
Michael Sheehan, Rikki Laser, Santiago Forero, Alexander Ophir, Jessica Nowicki, Lauren O'Connell, and members of the CPLab.

References:

1. Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. & Girshick, R. Detectron2. https://github.com/facebookresearch/detectron2 (2019).
2. Kirillov, A. et al. Segment anything. arXiv:2304.02643 (2023).
3. Bolya, D., Zhou, C., Xiao, F. & Lee, Y. J. Yolact: Real-time instance segmentation. In ICCV (2019).
4. Bolya, D., Zhou, C., Xiao, F. & Lee, Y. J. Yolact++: Better real-time instance segmentation (2019). 1912.06218.
5. Wada, K. LabelMe: Image polygonal annotation with Python. https://github.com/wkentaro/labelme (2016).
6. He, K., Gkioxari,.G., Dollár, P. & Girshick, R. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, 2961–2969 (2017).
7. Mathis, Alexander, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning." Nature neuroscience 21, no. 9 (2018): 1281-1289.
8. Pereira, Talmo D., Nathaniel Tabris, Arie Matsliah, David M. Turner, Junyu Li, Shruthi Ravindranath, Eleni S. Papadoyannis et al. "SLEAP: A deep learning system for multi-animal pose tracking." Nature methods 19, no. 4 (2022): 486-495.