Scoring rodent digging behavior with *Annolid*

Jiarui Fang, Chen Yang, Thomas A. Cleland

Computational Physiology Lab, Dept. of Psychology, Cornell University, Ithaca, NY USA

Computational Physiology Laboratory

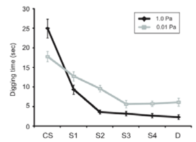
Overview

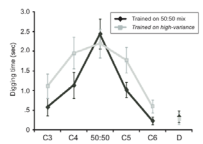
- Identifying complex behaviors in animal studies presents significant challenges, particularly for actions characterized by *sequences of movements* rather than static poses. A case in point is the behavior of '***digging***'.
- We developed an automated method to score rodent digging behavior using *Annolid*, a behavior analysis package grounded in deep learning and instance segmentation techniques. Our approach demonstrates that, with appropriate post-processing, *Annolid* achieves high accuracy in detecting digging behavior based on a training set of approximately 100 labeled frames.

Digging

Why might one want to score digging?

- First, digging is an innate, natural behavior in rodents; this simplifies the training process and is frequently utilized in various research contexts.
- Second, previous studies have indicated that the duration of digging correlates with the animal's certainty level, making it a valuable proxy for assessing cognitive processes such as generalization.
- Consequently, we sought to accurately and automatically quantify the entire period of digging based on video recordings of mice, lending meaningful insights into behavioral studies while minimizing user workload and potential scorer bias.



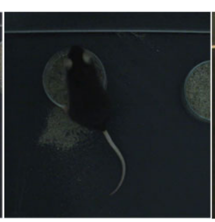


Digging time can be a measure of learning-dependent *generalization gradients* (Cleland et al. 2011)

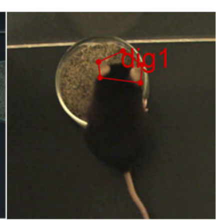
What is *digging*?

- We operationally defined '*digging*' as a continuous, investigative process characterized by two features. First, the **mouse must be orienting its head toward the substrate**. This definition is characteristic of both digging-with-paws and digging-with-nose behaviors. Second, that **substrate must be the digging medium** (play sand).
- To instantiate this, **we trained the model on a specifically tailored instance**, outlining the head in frames in which it was oriented down towards the substrate, while also **including a corona around the nose** so as to encode the visual texture of the substrate into the instance. Hence, we did not need to define the dishes as additional instances.
- Defining digging as a continuous, investigative process acknowledges the natural variability in the behavior. For example, interruptions in which the mouse raises its head are still considered part of the digging sequence if they are brief and occur between two instances of the head-lowering behavior. This was achieved in post-processing.

Head is not positioned over digging medium

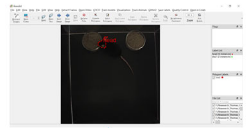


Not oriented (nose does not point down)



An instance labeled as *digging*

Annolid workflow to train a network to score digging



Draw polygon around the head, including some of the substrate surrounding the nose.

Label this instance as *digging* or *non-digging*.

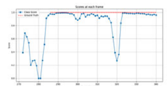
Export *Annolid*-generated metadata in COCO format.

COCO Common Objects in Context

Train *Annolid* instance segmentation model using Google Colab

colab

Apply model to testing video(s) for automatic scoring



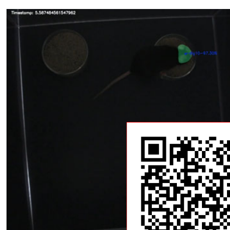
Compare results with human-scored ground truth to identify incorrectly classified frames

Import frames from training videos into *Annolid* GUI

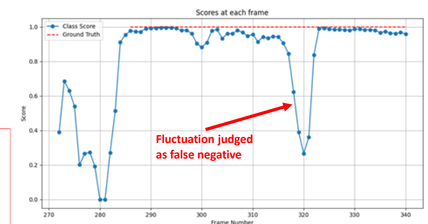
Outcome: a well-trained model applicable to the automatic scoring of any similar digging videos.

Annolid output (raw results)

- After applying a trained model to a video for automatic scoring, *Annolid* exports a CSV file containing frame-by-frame "*certainty scores*" [0..1] for digging behavior.
- Threshold-based classification into "*digging*" and "*non-digging*" frames can be used to **visually annotate original videos** for assessment and human-in-the-loop optimization.



Visually annotated video clip

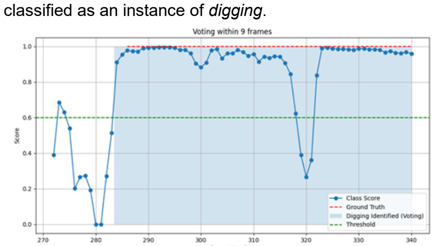


Digging scores across a section of this video. Ground truth is based on human classification.

- Importantly, the dynamic nature of rodent behavior (e.g., inconstant posture) generates fluctuations in these scores, which can disrupt the continuity of identified digging sequences.
- To achieve the continuity appropriate to our definition of digging behavior, we applied a post-processing filter on the instance segmentation results. **We evaluated two candidate filters.**

Post-processing: voting filter

- **Voting filter.** For each frame in turn, within a specified surrounding *range* of frames, if more than half of the frames in range are scored above the *threshold*, then the central frame is classified as an instance of *digging*.



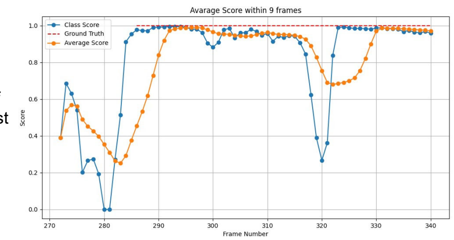
Voting filter with a range of 9 frames and a threshold of 0.6.

These parameters correctly distinguished *digging* from *non-digging* frames (compared to ground truth).

Post-processing: smoothing filter

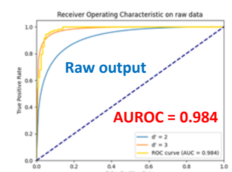
- **Smoothing filter.** Instead of a binary decision, this filter smooths fluctuations by replacing the score of each frame with the average score of its neighboring frames within a selected range.

A 9-frame range and a threshold of 0.6 separates most *digging* and *non-digging* frames



Result

- We compared these two filters, with ranges of 9 or 21 frames, using an Area Under the Receiver Operating Characteristic curve (AUROC) metric.
- All of these improved the match to ground truth, largely by selectively increasing the proportion of true positives identified.

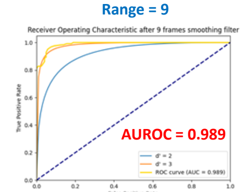


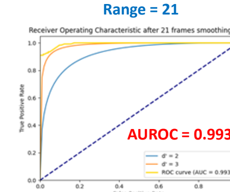
- The two filters are indistinguishable in their performance (while the voting filter here performs slightly better, in some other examples the smoothing filter yields better results).
- Importantly, filter improvements are not dependent on finely tuned parameters.

Range = 9

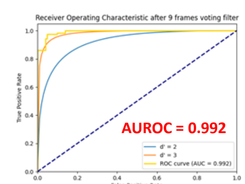
Range = 21

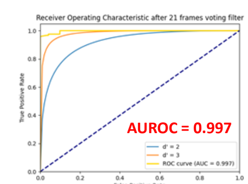
Smoothing filter





Voting filter





- While this model focuses on digging behavior, this method can be applied to all continuous behaviors that can be characterized by frame-identifiable features (such as "lowering head while over the correct substrate").

References & Acknowledgements

Cleland TA, Chen S-YT, Hozer KW, Ukatu HN, Wong KJ, Zheng F (2011). Sequential mechanisms underlying concentration invariance in biological olfaction. *Front Neuroeng.* 4:21. doi:10.3389/fneng.2011.00021.

Please refer to poster 512.02, "*Annolid: Annotate, Segment, and Track Anything You Need*" (next door) for more information about Annolid. Annolid is supported by R01 DC019124 and R01 DC014701.