Sifting through the haystack - efficiently finding rare behaviors in large-scale datasets

Shir Bar[1,2], Or Hirschorn[3], Roi Holzman[1,2], Shai Avidan[3]

[1] School of Zoology, Faculty of Life Sciences, Tel Aviv University, [2] The Interuniversity Institute for Marine Sciences in Eilat, [3] School of Electrical Engineering, Faculty of Engineering, Tel Aviv University

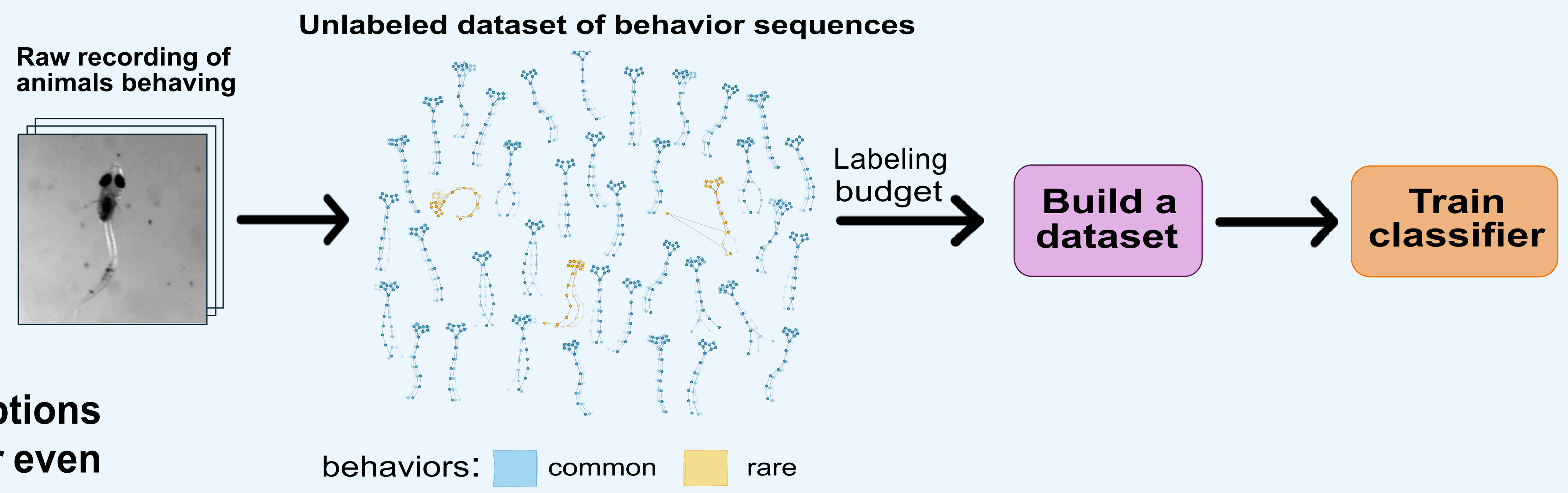
Problem setup

Continuous recording is crucial to find rare behaviors, but also creates an analysis bottleneck.

We ask:

1. How do we efficiently find rare behaviors on a budget?

2. Can we make no assumptions regarding the type, frequency or even existence of such behaviors?

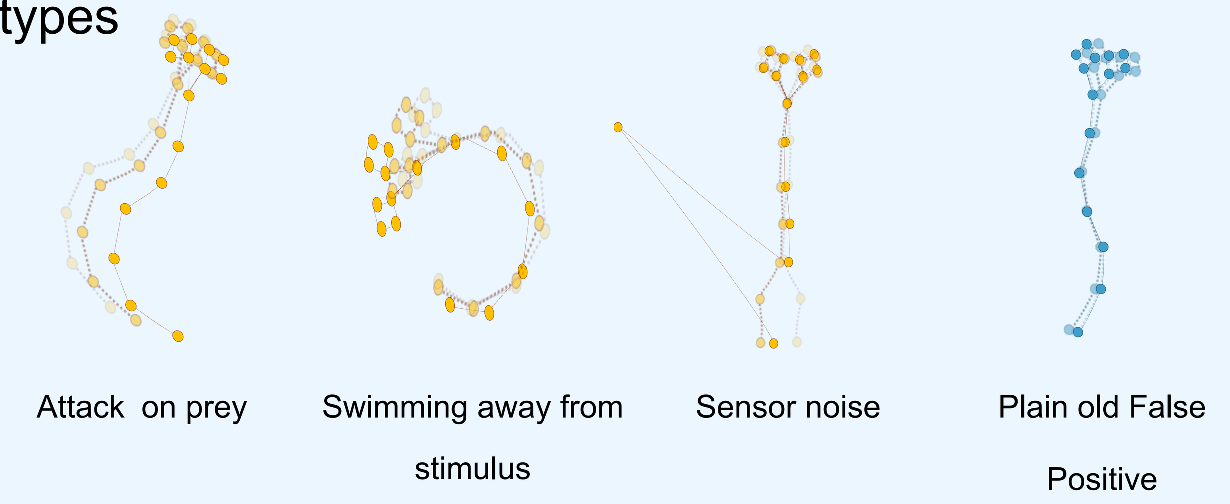


Anomaly detection alone is not the answer, there're many different types of anomalies!

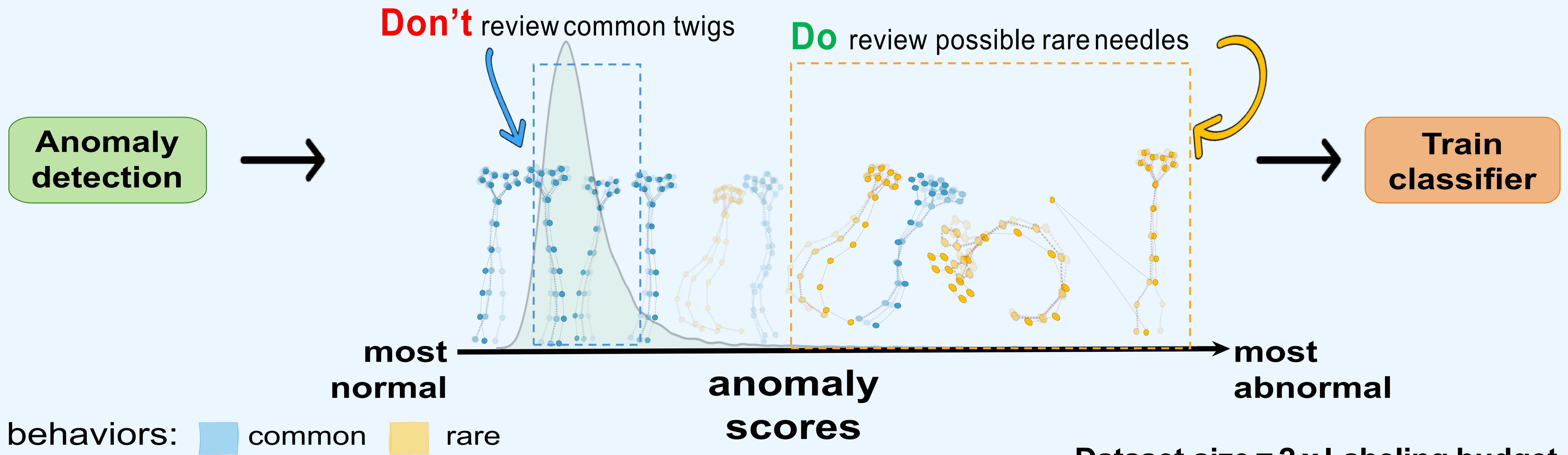
Here we use pose data, and focus on motion anomalies, but we'd like to note that:

A) Motion anomalies are not semantic anomalies, but are a good proxy for us

B) You can choose a different representation for your system and use the same concepts!



Approach



Dataset size = 2 x Labeling budget

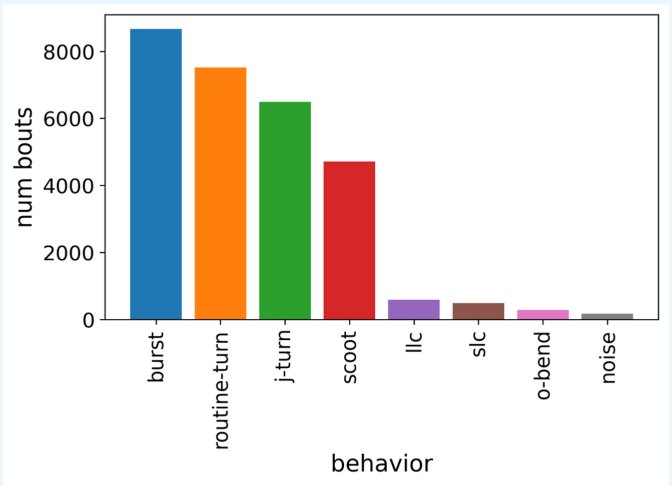
Our proposed method:

1. Rank the samples using an anomaly detector
2. Sample and review top scoring clips to look for anomalies of interest using the entire labeling budget
3. Get "freebies" from the center of the distribution and assume they are normal without review
4. Train a rare behavior binary classifier to get interesting samples from the dataset

Example dataset

PoseR - a fish behavior dataset [1]

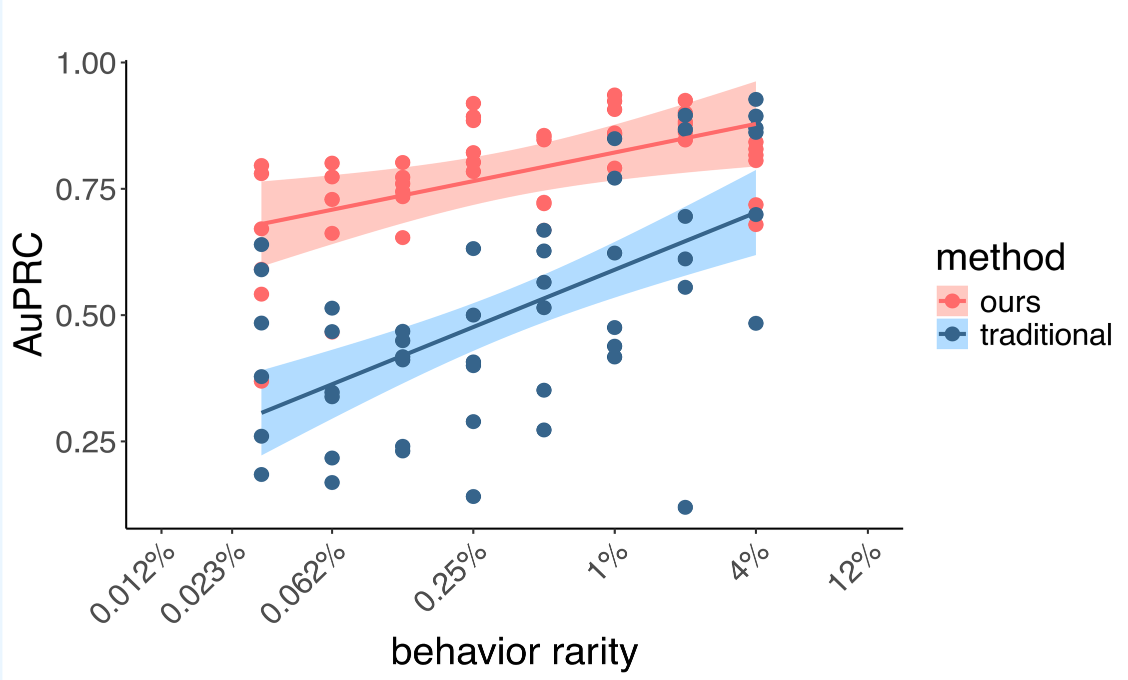
- 28K clips from neurobiology labs
- Rare behaviors make up ~5% of the data



https://doi.org/10.1101/2023.04.07.535991

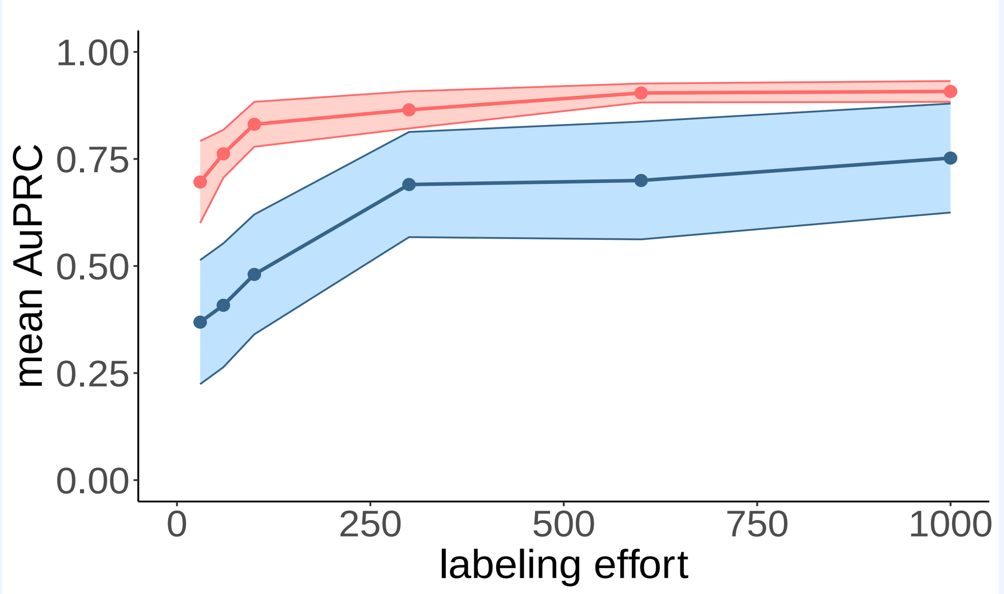
[1] Mullen, P.N., Bowlby, B., Armstrong, H.C. and Zwart, M.F., 2023. **PoseR-A deep learning toolbox for decoding animal behavior.** BioRxiv, pp.2023-04.

Pipeline performance

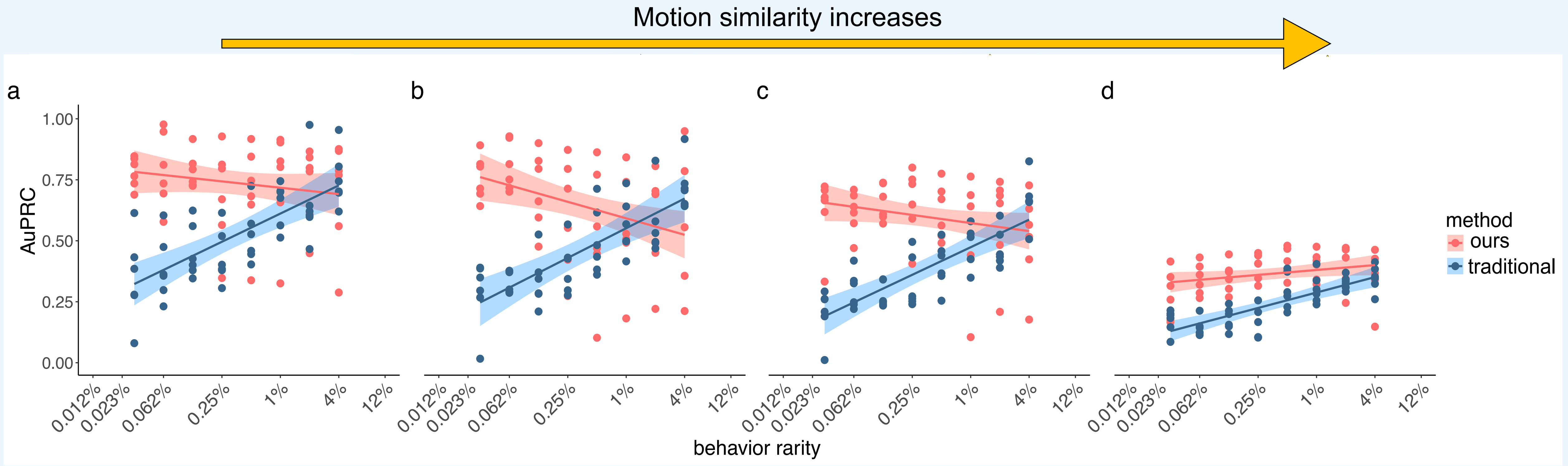


Performance as a function of rarity. Statistical modeling of classifier performance (in AuPRC) as a function of rarity (x-axis), labeling effort, and sampling method. We plot the predicted performance for a labeling budget of 200 samples.

Effect of labeling effort on performance. Classifier performance (AuPRC) as a function of labeling effort averaged across all tested rarities, the ribbon represents the 95% confidence intervals.



Synthetic experiments



Performance under increasing motion similarity. Using a **synthetic** dataset, we modeled classifier performance (y-axis) as a function of motion similarity (panels), rarity (x-axis), labeling effort and sampling method. Results presented here are for a labeling budget of 200 samples. Our method is most beneficial when behavior frequency is <1%.

Check out our paper and code for more: https://shir3bar.github.io/sifting-the-haystack-page/