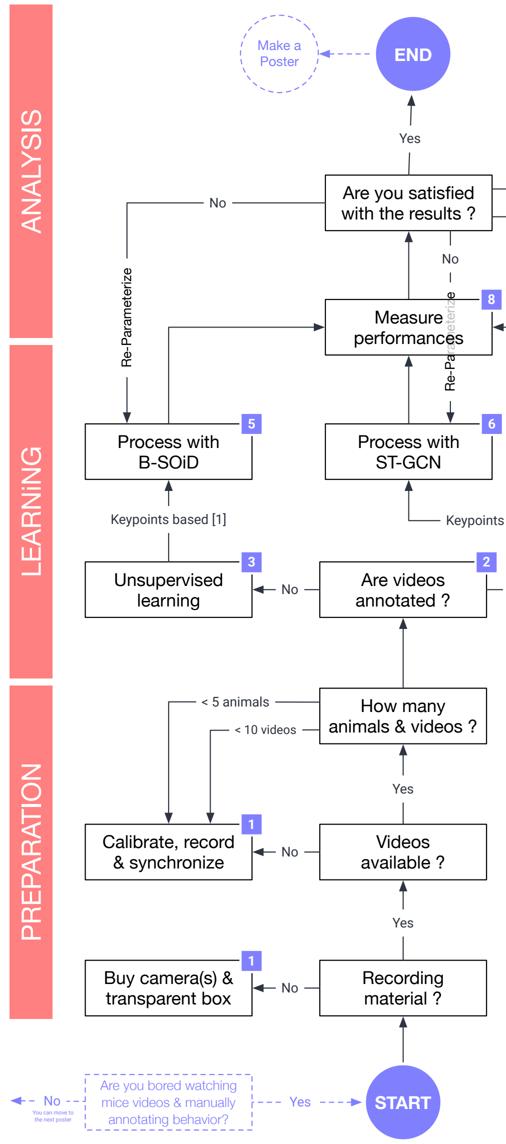


A Practical Guide to the

Automatic quantification of normal & pathological behavior in mice

For Hire
vleroy005@etsc.fr

VICTOR LEROY
LISE GUILHEMSANG
NICOLAS ROUGIER
NICOLAS MALLET

INSTITUTE OF NEURODEGENERATIVE DISEASES
BORDEAUX UNIVERSITÉ - CNRS - INRIA
NICOLAS.MALLET@U-BORDEAUX.FR

TL;DR: It works (somehow)

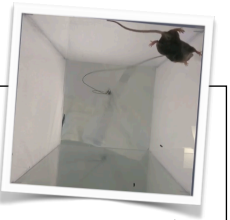
#162186

BORDEAUX IMN neurocampus

START

Abstract. The intake of the precursor levodopa in Parkinson's disease patients triggers abnormal and involuntary movements known as **levodopa-induced dyskinesias (LIDs).** To date, the main technique for assessing and quantifying LIDs is based on human experts that score them based on observations of patients (live or video). This method presents several drawbacks: it is both time consuming and tedious, but it is also prone to errors and inter-examiner variability. In this work, we explored a machine learning / computer vision approach in order to detect and quantify LIDs on video recordings of Parksinian mice treated with levodopa.

To do so, we tested three methods: (i) **an unsupervised method,** based on mouse pose estimation [1] and clustering of pertinent behavioural features [5] (ii) **a supervised method** which learns the segmentation from the videos [2] using convolutional neural networks and (iii) **a second supervised learning model** [3] based on the keypoints and originally designed to recognize human actions. We proceeded by investigating **how reliable are these three approaches for LIDs quantification from a practical point of view** (learning sets, parameters tuning, analysis of results, estimation of ground truth).



The unsupervised method successfully recognizes global dyskinetic patterns, (namely, contralateral rotations and axial dyskinesia) but can hardly detect subtler dyskinesias such as forelimbs tremors. Only in a few specific cases the method was able to properly characterize them. For the other cases, the supervised methods, that are especially tuned to recognize difficult behaviors, can provide the missing information.

However, this method requires the manual labelling of videos in order to train the model and suffers consequently from the same drawbacks as for the traditional methods. Such drawbacks can be largely mitigated using a standard training set (that is yet to be build). From our early results, this automated approach seems to be a viable alternative even though more tests are needed to assess the quality of the quantification when compared to human expertise.

1 Material & Methods

Transparent box allows to record from below. Be careful, the shape of the box induces different behaviors.

One GoPro or Smartphone cameras are good but depth camera or event camera might be better. If you have two synchronized cameras, it is even better.

Calibrate your cameras by setting the position, framerate, aperture, zoom level, etc. We used a GoPro with 120 fps (recommended for LIDs), 1/240s shutter speed, 1080p resolution, 3200k white balance.

Synchronize camera and any experimental setup you might use such as optogenetics (e.g. an always visible on/off light in the box).

Split your dataset in training (60%), validation (20%) and test (20%).

Check your dataset to make sure to not have an over-represented behavior or else, learning will overfit. Check also each beahvior is present in both training, validation and test.

Parameterize your model hyperparameters (e.g. using HyperOpt or Optuna)

2 Annotations

The estimated annotation time varies according to the level of (ethological) expertise, especially when labeling LIDs. Count around 45min for 10 000 frames, between 2h and 3h for a 10min video at 60fps.

Choosing the correct list of behaviors is critical for the quality of your analysis and will directly determine the results you get.

For normal behavior we use this list : body scratch, hind paw bite, supported/unsupported rear, turn left/right, walk, grooming, micromovement

For LIDs, after discussion with experts, we recommend to use the conventional 4 types categorization.

Labeling behavior is difficult because ambiguous sequences occur, especially during LIDs.

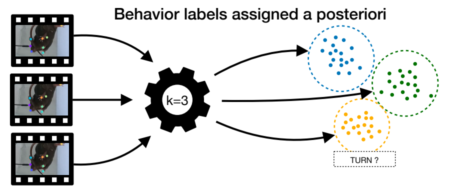
Bootstraping annotations is possible by using temporary results for a first approximation of behavior.

Annotating Tools: DeepEthogram, SimBA

Acknowldegments to the annotating experts: Gregory Porras, Lorena Delgado-Zabalza & Cristina Miguelez Palomo.

3 Unsupervised learning

Behavior labels assigned a posteriori



4 Supervised learning

Behavior labels assigned a priori

5 B-SOiD

B-SOiD - an open-source, unsupervised algorithm that identifies behavior without user bias. By training a machine classifier on pose pattern statistics clustered using new methods, B-SOiD approach achieves greatly improved processing speed and the ability to generalize across subjects or labs.

6 DeepEthogram

DeepEthogram uses supervised machine learning to convert raw video pixels into an ethogram, the behaviors of interest present in each video frame. DeepEthogram is designed to be general-purpose and applicable across species, behaviors, and video-recording hardware.

7 ST-GCN

Spatial-Temporal Graph Convolutional Networks, which moves beyond the limitations of previous methods by automatically learning both the spatial and temporal patterns from data. This formulation not only leads to greater expressive power but also stronger generalization capability.

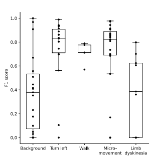
8 Results

B-SOiD **(qualitative evaluation)**

For unsupervised methods there is no ground truth because the behaviors are not known a priori. The first and easiest way to assess the performances is thus to directly look at the videos. Example of behavioral segmentation with B-SOID trained on 19 videos of 6 dyskinetic mice:

DeepEthogram **(quantitative evaluation)**

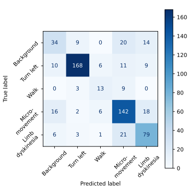
Performances of deepethogram (frame-wise F1 score) on the test set for 5 behaviors. Each dot represents the F1 score on one video (only if the behavior is occurring during the recording time). The model was trained on 19 videos of 8 dyskinetic mice.



ST-GCN **(quantitative evaluation)**

Performances of ST-GCN: clip-wise confusion matrix and F1 score on the test set for five behaviors. Same dataset as deepethogram.

F1 Scores:
Background: 0.48, Turn left: 0.86,
Walk: 0.51, Micro-movement: 0.73,
Limb Dyskinesia. 0.69



References

[1] Mathis, A., Mamidanna, P., Cury, K.M. et al. **DeepLabCut: markerless pose estimation of user-defined body parts with deep learning.** Nat Neurosci 21, 1281–1289 (2018)

[2] Bohnslav J.M., Wimalasena, N.K., Clausing, K. J. et al. **DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels.** eLife 10:e63377 (2021)

[3] Yan, S., Xiong, Y. and Lin, D. **Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition.** AAAI (2018).

[4] Wiltschko A.B., Johnson M.J., Iurilli G. et al. **Mapping Sub-Second Structure in Mouse Behavior.** Neuron, 88(6),1121-1135 (2015).

[5] Hsu, A.I., Yttri, E.A. **B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors.** Nat Commun 12, 5188 (2021).