DeepOF: a Python package for supervised and unsupervised pattern recognition in mice motion tracking data

Lucas Miranda [1], **Joeri Bordes** [2], **Benno Pütz** [1], **Mathias V Schmidt** [2], and **Bertram Müller-Myhsok** [1]¶

1 Research Group Statistical Genetics, Max Planck Institute of Psychiatry, Munich, Germany **2** Research Group Neurobiology of Stress Resilience, Max Planck Institute of Psychiatry, Munich, Germany ¶ Corresponding author

Summary

DeepOF (Deep Open Field) is a Python package that provides a suite of tools for analyzing behavior in freely-moving rodents. Specifically, it focuses on postprocessing time-series data extracted from videos using DeepLabCut (Mathis et al., 2018). The software encompasses a diverse set of capabilities, such as:

- Loading DeepLabCut data into custom objects and incorporating metadata related to experimental design.
- Processing data, including smoothing, imputation, and feature extraction.
- Annotating behavioral motifs in a supervised manner, such as recognizing huddling and climbing, and detecting fundamental social interactions between animals.
- Embedding motion tracking data in an unsupervised manner using neural network models, which also facilitate end-to-end deep clustering.
- Conducting post-hoc analysis of results and visualization to compare patterns across animals under different experimental conditions.

The package is designed to work with various types of DeepLabCut input (single and multi-animal projects), includes comprehensive documentation, and offers interactive tutorials. Although many of its primary functionalities (particularly the supervised annotation pipeline) were developed with top-down mice videos in mind, tagged with a specific set of labels, most essential functions operate without constraints. As demonstrated in the accompanying scientific application paper (Bordes et al., 2022), DeepOF has the potential to enable systematic and thorough behavioral assessments in a wide range of preclinical research settings.

Statement of need

The field of behavioral research has experienced significant advancements in recent years, particularly in the quantification and analysis of animal behavior. Historically, behavioral quantification relied heavily on tests that were designed with either one or a few readouts in mind. However, the advent of deep learning for computer vision and the development of packages such as DeepLabCut, which enable pose estimation without the need for physical markers, have rapidly expanded the possibilities for non-invasive animal tracking (Mathis et al., 2020).

By transforming raw video footage into time series data of tracked body parts, these approaches have paved the way for the development of software packages capable of automatically

annotating behavior following a plethora of different approaches, increasing the number of patterns that can be studied per experiment with little burden on the experimenters.

For example, several tools offer options to detect predefined behaviors using supervised machine learning. Along these lines, programs like SimBA (Nilsson et al., 2020), MARS (Segalin et al., 2021), or TREBA (Sun et al., 2021), allow users to label a set of behaviors and train classifiers to detect them in new videos. They employ different labelling schemes which require different amounts of user input, and offer high flexibility in terms of the number of behaviors that can be detected. On the other hand, packages such as B-SOiD (Hsu & Yttri, 2021), VAME (Luxem et al., 2022), and Keypoint-MoSeq (Weinreb et al., 2023), aim for a more exploratory approach that does not require user labelling, but instead relies on unsupervised learning to segment time series into different behaviors. These packages are particularly useful when the user is interested in detecting novel behaviors, or when the number of behaviors is too large to be annotated manually. Moreover, some approaches have been developed to combine the best of both worlds, such as the the A-SOiD active learning framework (Schweihoff et al., 2022), and the semi-supervised DAART (Whiteway et al., 2021). While a thorough discussion on the advantages and disadvantages of each package is beyond the scope of this paper, further information can be found in this recent review (Bordes et al., 2023).

In contrast to other available options, DeepOF offers both supervised and unsupervised annotation pipelines, that allow researchers to test hypotheses regarding experimental conditions such as stress, gene mutations, and sex, in a flexible way (Figure 1).
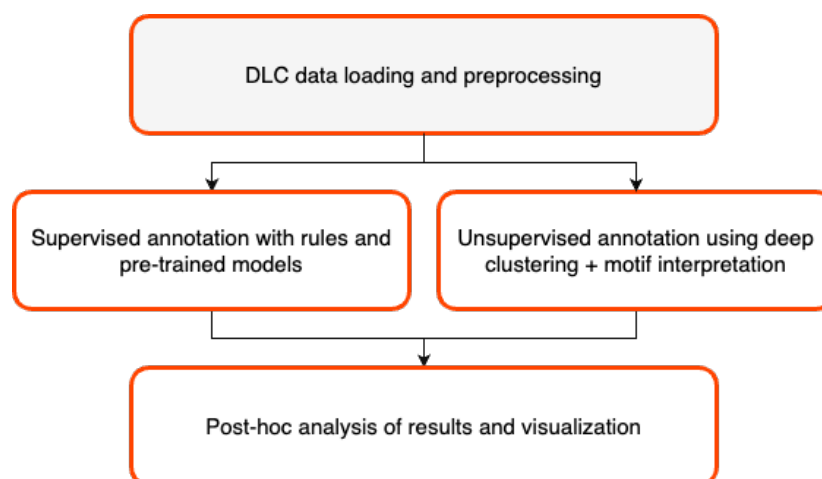
Figure 1: Scheme representing DeepOF workflow. Upon creating a project, DLC data can be loaded and preprocessed before annotating it with either a supervised pipeline (which uses a set of pre-trained models and rule-based annotators) or an unsupervised pipeline, which relies on custom deep clustering algorithms. Patterns retrieved with either pipeline can be passed to downstream post-hoc analysis tools and visualization functions.

The included supervised pipeline uses a series of rule-based annotators and pre-trained machine learning classifiers to detect when each animal is displaying a set of pre-defined behavioral motifs. The unsupervised workflow uses state-of-the-art deep clustering models to extract novel motifs without prior definition. DeepOF then provides an interpretability pipeline to explore what these retrieved clusters are in terms of behavior, which uses both Shapley Additive Explanations (SHAP) (Goodwin et al., 2022) and direct mappings from clusters to video. Moreover, regardless of whether the user chose the supervised annotation pipeline, the unsupervised one, or both, DeepOF provides an extensive set of post-hoc analysis and visualization tools.

When it comes to comparing it to other individual packages that use supervised and unsupervised

annotation, DeepOF stands out in several ways. First of all, it is the first package, to the best of our knowledge, to offer both options. Second, the supervised pipeline in DeepOF follows an opinionated philosophy, in the sense that it provides a set of pre-trained models that cannot be customized, but do not require user labels. This trades flexibility for ease of use, aiming at being a quick exploratory tool that can provide information on key individual and social behaviors with just a few commands. Furthermore, when it comes to the unsupervised pipeline, DeepOF provides three custom deep clustering algorithms capable of segmenting the behavioral time series, as well as the aforementioned built-in interpretability pipeline. If a user runs both pipelines, supervised annotations can be incorporated into this interpretability pipeline in quite a unique way, to detect associations between supervised and unsupervised patterns.

All in all, DeepOF is a comprehensive, end-to-end tool designed to transform DeepLabCut output into relatively quick, exploratory insights on behavioral shifts between experimental conditions, and pinpoint which behaviors are driving them.

Related literature

The DeepOF package has been used to characterize differences in behavior associated with Chronic Social Defeat Stress (CSDS) in mice, as presented in our preprint (currently in revision at the time of writing (Bordes et al., 2022)). There are several other ongoing projects involving the software, although none of them are published to this date.

Acknowledgements

Funding

References

Bordes, J., Miranda, L., Müller-Myhsok, B., & Schmidt, M. V. (2023). Advancing social behavioral neuroscience by integrating ethology and comparative psychology methods through machine learning. *Neuroscience & Biobehavioral Reviews*, *151*, 105243. https://doi.org/10.1016/J.NEUBIOREV.2023.105243

Bordes, J., Miranda, L., Reinhardt, M., Brix, L. M., Doeselaar, L. van, Engelhardt, C., Pütz, B., Agakov, F., Müller-Myhsok, B., & Schmidt, M. V. (2022). Automatically annotated motion tracking identifies a distinct social behavioral profile following chronic social defeat stress. *bioRxiv*. https://doi.org/10.1101/2022.06.23.497350

Goodwin, N. L., Nilsson, S. R. O., Choong, J. J., & Golden, S. A. (2022). Toward the explainability, transparency, and universality of machine learning for behavioral classification in neuroscience. *Current Opinion in Neurobiology*, *73*, 102544. https://doi.org/10.1016/j.conb.2022.102544

Hsu, A. I., & Yttri, E. A. (2021). B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-25420-x

Luxem, K., Mocellin, P., Fuhrmann, F., Kürsch, J., Miller, S. R., Palop, J. J., Remy, S., & Bauer, P. (2022). Identifying behavioral structure from deep variational embeddings of

animal motion. *Communications Biology 2022 5:1*, *5*(1), 1–15. https://doi.org/10.1038/s42003-022-04080-7

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience 2018 21:9*, *21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Mathis, A., Schneider, S., Lauer, J., & Mathis, M. W. (2020). A Primer on Motion Capture with Deep Learning: Principles, Pitfalls, and Perspectives. *Neuron*, *108*(1), 44–65. https://doi.org/10.1016/j.neuron.2020.09.017

Nilsson, S. R., Goodwin, N. L., Choong, J. J., Hwang, S., Wright, H. R., Norville, Z. C., Tong, X., Lin, D., Bentzley, B. S., Eshel, N., McLaughlin, R. J., & Golden, S. A. (2020). Simple Behavioral Analysis (SimBA) – an open source toolkit for computer classification of complex social behaviors in experimental animals. *bioRxiv*. https://doi.org/10.1101/2020.04.19.049452

Schweihoff, J. F., Hsu, A. I., Schwarz, M. K., & Yttri, E. A. (2022). A-SOiD, an active learning platform for expert-guided, data efficient discovery of behavior. *bioRxiv*. https://doi.org/10.1101/2022.11.04.515138

Segalin, C., Williams, J., Karigo, T., Hui, M., Zelikowsky, M., Sun, J. J., Perona, P., Anderson, D. J., & Kennedy, A. (2021). The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice. *eLife*, *10*, e63720. https://doi.org/10.7554/eLife.63720

Sun, J. J., Kennedy, A., Zhan, E., Anderson, D. J., Yue, Y., & Perona, P. (2021, June). Task Programming: Learning Data Efficient Behavior Representations. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr46437.2021.00290

Weinreb, C., Osman, M. A. M., Zhang, L., Lin, S., Pearl, J., Annapragada, S., Conlin, E., Gillis, W. F., Jay, M., Ye, S., Mathis, A., Mathis, M. W., Pereira, T., Linderman, S. W., & Datta, S. R. (2023). Keypoint-MoSeq: parsing behavior by linking point tracking to pose dynamics. *bioRxiv*. https://doi.org/10.1101/2023.03.16.532307

Whiteway, M. R., Schaffer, E. S., Wu, A., Buchanan, E. K., Onder, O. F., Mishra, N., & Paninski, L. (2021). Semi-supervised sequence modeling for improved behavioral segmentation. *bioRxiv*. https://doi.org/10.1101/2021.06.16.448685