AUTOMATED BEHAVIORAL ANALYSIS OF ASLEEP FRUIT FLY

by
Ali Osman Berk Şapcı

Submitted to
the Faculty of Engineering and Natural Sciences
in partial fulfillment of the requirements for the degree of
Master of Science

Sabancı Üniversitesi
İstanbul, Türkiye
July 2022

AUTOMATED BEHAVIORAL ANALYSIS OF ASLEEP FRUIT FLY

Approved by

Asst. Prof. Öznur Taştan .....
(Thesis Advisor)

Prof. Sündüz Keleş .....
(Thesis Co-Advisor)

Asst. Prof. Onur Varol .....

Prof. Gözde Bozdağı Akar .....

Assoc. Prof. Ercüment Çiçek .....

Date of approval: July 27, 2022

ABSTRACT

AUTOMATED BEHAVIORAL ANALYSIS OF ASLEEP FRUIT FLY

ALI OSMAN BERK ŞAPCI

COMPUTER SCIENCE AND ENGINEERING MSC. THESIS, JULY 2022

Thesis Advisor: Asst. Prof. Öznur Taştan
Thesis Co-Advisor: Prof. Sündüz Keleş

Keywords: Sleep, Computational ethology, Behavioral analysis, Activity detection, Behavior mapping

Sleep is a highly conserved behavior program across the animal kingdom, hinting at its essential value. In order to decipher the functions of sleep, careful characterization of the underlying changes in behavior and physiology is needed in powerful genetic model systems such as Drosophila Melanogaster. Recent advances in machine learning have enabled tracking of body parts and robust pose estimation in videos; however, automated quantification of behaviors requires mapping from a pair of spatial coordinates to behavioral categories. Detection and successful mapping of behaviors exhibited during sleep come with unique challenges. Existing methods and pipelines are developed with behaviors defined by macro postural changes ignoring subtle movements during sleep. Our task of phenotyping sleep requires tackling behaviors defined by unobtrusive changes that sparsely occur during long sleep cycles. To this end, we develop `basty` (Automated Behavioral Analysis of Asleep Fruit Fly), a novel, end-to-end pipeline made public as a configurable, open source, and easy-to-use software package. Our pipeline enables several analyses, such as computing meaningful behavioral representations, detecting activities in long sleep experiments (14-16 hours), and nearest neighbors analysis in a low dimensional behavioral space. We evaluate our pipeline with a dataset of sleep deprivation and wild-type sleep experiments, focusing on five behavioral categories. Results show that our pipeline successfully maps behavioral categories, achieving an AUC score of 0.8, and it can detect unobserved behaviors and differences in behavioral repertoires. Furthermore, we present an analysis of the behavioral repertoire exhibited during sleep by examining

spatio-temporal characteristics of the behaviors and their temporal organization; in both wild-type sleep and sleep-deprived experiments.

ÖZET

UYKU HALINDEKI MEYVE SINEĞININ DAVRANIŞLARININ OTOMATIK ANALIZI

ALI OSMAN BERK ŞAPCI

BILGISAYAR BILIMI VE MÜHENDISLIĞI YÜKSEK LİSANS TEZİ, TEMMUZ 2022

Tez Danışmanı: Dr. Öğr. Üyesi Öznur Taştan
İkincil Tez Danışmanı: Prof. Dr. Sündüz Keleş

Anahtar Kelimeler: Uyku, Hesaplamalı etoloji, Davranış analizi, Aktivite tespiti, Davranış tasnifi

Uyku, hayvanlar aleminde evrimsel olarak korunmuş, elzem bir davranış biçimidir. Uykunun işlevlerini anlamak için, Drosophila Melanogaster gibi model organizmalarda uyku esnasında gözlemlenen davranışsal ve fizyolojik değişikliklerin başarılı bir şekilde karakterize edilmesi gerekmektedir. Makine öğrenimini alanındaki gelişmeler, vücut bölümlerinin otomatik olarak takip edilebilmesini ve yüksek başarımlı poz tahmini yapılmasını sağlamıştır. Ancak, davranışların tespit edilmesi ve sınıflandırılması, pozların ve pozlarda meydana gelen değişiklerin hangi davranış kategorilerine karşılık geldiğini hesaplamayı gerektirir. Uyku esnasında sergilenen davranışların tespiti ve başarılı bir şekilde tasnif edilmesi, kendine özgü zorluklara sahiptir. Mevcut yöntemler ve veri işleme yaklaşımları, uyku sırasındaki meydana gelen ve fark etmesi zor hareketlerden ziyade, makro ölçekte gerçekleşen postural değişikliklere odaklanarak geliştirilmiştir. Uykuda gerçekleşen davranışları analiz etme hedefimiz, uzun uyku döngüleri sırasında seyrek olarak meydana gelen küçük değişiklikleri başarılı bir şekilde saptamayı ve sınıflandırmayı gerektirir. Bu amaçla açık kaynak kodlu ve kullanımı kolay bir yazılım paketi olarak sunduğumuz bir veri işleme modeli olan basty'i geliştirdik. Modelimiz, anlamlı davranış temsillerini hesaplama, uzun uyku deneylerinde (14-16 saat) aktiviteleri tespit etme ve düşük boyutlu davranışsal gömme uzaylarında en yakın komşu çözümlemesi gibi çeşitli analizlere olanak tanır. Modelimizi, beş davranış kategorisine odaklanarak, uyku yoksunluğu ve vahşi tip uyku deneylerinin verileri ile değerlendirdik. Sonuçlar, geliştirdiğimiz yaklaşımın davranış kategorilerini başarılı bir şekilde tasnif ettiğini, ve 0,8 AUC puanı elde edebildiğini gösteriyor. Üstelik, yöntemimizin daha önce gözlemlenmeyen davranışları ve davranışsal

repertuvarlardaki farklılıkları da tespit edebildiğini de ortaya koyduk. Ayrıca, hem vahşi tip uyku hem de uyku yoksunluğu deneylerinde, davranışların uzamsal-zamansal niteliklerini ve zamana bağlı organizasyonlarını inceleyerek uyku sırasında sergilenen davranışsal repertuvarın bir analizini sunuyoruz.

ACKNOWLEDGMENT

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Dedicated to
x, y, and z.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Sleep is an essential behavioral program conserved across the animal kingdom, in diverse species ranging from jellyfish to humans, whose function remains unknown (Campbell and Tobler, 1984; Nath et al., 2017). In mammals, sleep consists of multiple stages marked by physiological changes, including reductions in muscle tone and distinct electrophysiological activity patterns in the brain (Corner, 1977; Sauer et al., 2003) In invertebrates, sleep has largely been studied as a unitary process and identified by bouts of consolidated immobility. Thus, careful characterization of underlying changes in behavior and physiology is needed for understanding the functional role of sleep and characterizing distinct sleep stages in powerful genetic model systems such as *Drosophila Melanogaster*.

Recent technical advances enabled automated quantification of behaviors, and this has opened up a new field of "computational ethology" (Anderson and Perona, 2014; Datta et al., 2019). Particularly, the recent progress in deep learning has led to the emergence of methods for tracking animal motion, which provided the opportunity of studying naturalistic behavior at an unprecedented resolution (Pereira, 2020). Developments of pose estimation and tracking tools such as DeepLabCut (Mathis et al., 2018) and SLEAP (Pereira et al., 2019, 2022) collecting spatio-temporal data about the animals possible. However, this data only consists of coordinate values in two or three-dimensional spatial domains, depending on the employed tracking tool. Quantifying the animals' rich and complex behavioral repertoires, considering their temporal structure and ambiguity, is inherently a difficult problem without clear ground truth, as discussed in Pereira (2020).

Because a static set of pose values are not sufficient to describe spatio-temporal complexities of behaviors, converting positional coordinates to meaningful spatio-temporal feature representations is an essential step for capturing the time-varying structure of

behaviors. The first step of many behavior mapping pipelines is generating hand-crafted features describing relative positions of body parts, distances between body parts, angles between body parts, and how these values change over time, e.g., velocities and angular velocities (Kabra et al., 2013; Hsu and Yttri, 2021; Marshall et al., 2021; Nilsson et al., 2020). Alternatively, there exist several studies which directly compute behavioral representations from videos by employing deep learning (Bohnslav et al., 2021), or advanced computer vision techniques (Berman et al., 2014; Wiltschko et al., 2015). For both approaches, resulting behavioral representations are high-dimensional time series that can also be used to generate a spectrogram representation as in Berman et al. (2014); Todd et al. (2017); Marshall et al. (2021).

A popular approach is projecting high-dimensional time series of behavioral representations into a low-dimensional behavioral embedding; using either autoencoders (Whiteway et al., 2021; Graving and Couzin, 2020) or manifold learning techniques, such as t-SNE (Maaten and Hinton, 2008), Isomap (Tenenbaum et al., 2000) and UMAP (McInnes et al., 2020) (Berman et al., 2014; Marshall et al., 2021; Hsu and Yttri, 2021; DeAngelis et al., 2019; Mearns et al., 2020). Having appropriate behavioral representations available, one can perform supervised learning, given user-provided examples of behavioral categories. Various algorithms are used for the task of learning behaviors, such as SVM Boser et al. (1992) by Hsu and Yttri (2021), LSTM (Hochreiter and Schmidhuber, 1997) by Wu et al. (2021), and random forest ensembles (Breiman, 2001) by Kabra et al. (2013) and Nilsson et al. (2020). Alternatively, unsupervised approaches are useful for discovering repeatedly and stereotypically exhibited behaviors, as "behavioral clusters" with clustering (Berman et al., 2014; Todd et al., 2017; Marques et al., 2018; Marshall et al., 2021) or "behavioral states" with state space models (Wiltschko et al., 2015). The advantage of benefiting from unsupervised learning is to avoid annotator bias, annotation cost, and various shortcomings of depending on human definitions of behavior.

Neuroscientists studying sleep have mainly focused on locomotor-type behaviors exhibited at night (Wiggin et al., 2020; Nath et al., 2017), and attempts to understand underlying sleep stages by measuring the overall displacement and quiescence. For example, the commercially available Drosophila Activity Monitor (DAMs, Inc., Waltham, Massachusetts) is used extensively to study circadian rhythms and sleep (Pfeiffenberger et al., 2010b,a). Relatively advanced systems, such as Ethoscopes (Geissmann et al., 2017), have also been developed in recent years. Featuring supervised machine learning, Ethoscopes learns to detect not only walking activity but also micro-activities (for example, in-place movements such as grooming and egg laying) but without a fine-grained categorization of micro-activities to more specific behaviors (Geissmann et al., 2019). Hinting at the im-

portance of analyzing rich behavioral repertoire exhibited during sleep, van Alphen et al. (2021), analyzed the functional role of proboscis pumping behavior, but without automatizing behavioral analysis, proboscis pumping behavior is quantified by manually scanning through the videos. Ad-hoc solutions, specifically dedicated to a single or a small subset of behaviors have also been introduced. For example, Itskov et al. (2014) presents a method that utilizes capacitive measurements for automated quantification of feeding behavior. Another such ad-hoc solution is developed by Qiao et al. (2018) and achieves automated analysis of long-term grooming behavior in *Drosophila Melanogaster*. Their method, including hardware and a platform for video recordings, maps fly activity onto a three-dimensional behavior space and utilizes a k-nearest neighbors classifier. However, detailed phenotyping of the behavioral repertoire of sleep, which we aim to achieve in this work, has not been addressed in the literature.

The inherent structure of sleep is mainly characterized by reduced muscle activity and long dormancy epochs. Thus, observable behaviors are exhibited much more scarcely and rarely, compared to wakefulness. Moreover, during sleep, the behavioral repertoire of fruit flies substantially consists of in-place behaviors rather than major postural and positional. For instance, proboscis pumping is an in-place behavior, which is not necessarily accompanied by a positional or postural change in the rest of the body. Another such example is the switch-like behavior of the haltere that is characterized by the haltere's positional change in the order of microns and eventuates rapidly within a second. Therefore, our task of phenotyping sleep requires tackling behaviors defined by unobtrusive changes that sparsely occur during long sleep cycles. Previously developed behavior mapping models do not cover these behaviors. Their main focus has been on behaviors that feature major postural and positional changes, such as walking and wing waggle.

In this work, we develop `basty` (Automated Behavioral Analysis of Asleep Fruit Fly), a novel, end-to-end pipeline made public as a configurable, open source, and easy-to-use software package[1]. `basty` directly operates on the output of the pose estimation tools, such as DeepLabCut (Mathis et al., 2018) and SLEAP Pereira et al. (2022). Unlike previous behavior mapping studies, we focus on behavioral repertoire exhibited during dormancy, i.e., in-place behaviors such as grooming, switch-like haltere movement, and postural adjustments. Similar to previous studies, our approach starts by computing meaningful behavioral representations. To this end, `basty` offers an easily configurable and flexible framework, which includes a novel preprocessing step, extensive spatio-temporal feature computation, and spectrogram generation. Enabling analysis of sleep experiments and biological inference, `basty` allows detecting activities in long sleep

[1] `basty`, implemented in Python, is publicly available at https://github.com/bo1929/basty

experiments (14-16 hours). Activity detection can be done both in a supervised manner and unsupervised manner by using the Gaussian mixture model and random forest ensembles, respectively. The proposed pipeline includes a novel supervised behavior mapping approach, which takes advantage of semi-supervised dimensionality reduction and nearest neighbors analysis in a low dimensional behavioral space.

We evaluate our pipeline with a dataset of sleep deprivation and wild-type sleep experiments, focusing on five behavioral categories: grooming, feeding, proboscis pumping, haltere switch, and postural adjustments. Notably, automated quantification of switch-like movement of haltere is achieved for the first time in this work. Results show that our pipeline successfully maps behavioral categories, achieving an AUC score of 0.8 with limited supervision. We also demonstrate that our approach enables detecting unobserved behaviors and differences in behavioral repertoires. Furthermore, we present an analysis of the behavioral repertoire exhibited during sleep. Our analysis reveals the spatio-temporal characteristics of the behaviors and their temporal organization; in both wild-type sleep and sleep-deprived experiments.

The thesis is organized as follows:

- Chapter 2 describes how the sleep experiments are conducted. We present the technical details for the collection of fly video recordings and pose estimation data.
- Feature extraction stage is described in the Chapter 3 in detail. This chapter includes preprocessing, computation of spatio-temporal features, wavelet transformation, and sliding window moving statistics for the generation of dynamic postural features.
- Chapter 4 explains our approach for quantifying activities and extracting bouts of micro-activities observed during sleep.
- We describe our approach for fine-grained categorization of the exhibited behaviors in Chapter 5. In particular, behavioral embedding generation is discussed in Section 5.2, and we explain nearest neighbor analysis in Section 5.3.
- Evaluation of the pipeline and configurations that we used in `basty` is given in the Chapter 6. We also analyze the behavioral repertoire of the asleep fruit fly in terms of spatio-temporal characteristics later in this chapter.
- We conclude our work and discuss future directions in Chapter 7.

2. EXPERIMENTS AND DATA COLLECTION

In this chapter, we describe how the sleep experiments are conducted and fly video recordings and pose estimation data were collected.

The experimental data were collected by Dr. Mehmet Fatih Keles at Wu Lab, Johns Hopkins University.

2.1 Sleep Experiments and Video Recordings

A custom imaging setup was used to perform high-resolution characterization of sleep-related behaviors in flies. This setup includes a custom 3D printed chamber (7.2X4.3X2.4 mm [WxHxL]) that is placed in front of an IR-sensitive (Flir) 30 FPS camera with a telecentric lens (Edmund Optics), an illustration of the experimental setup is given in the Figure 2.1.

Flies were recorded between ZT10-ZT2 (16 hours total) for wild-type sleep experiments, and between ZT10-ZT6 (20 hours total) for sleep-deprived experiments (see Table 2.1). Each chamber has a food port (1.5 mm diameter) that allows access to liquid food (2.5% yeast, 2.5% sugar). The recording setup is in a light-tight box and humidity control (60%) is achieved via a humidifier plugged into a humidity control switch. Experimental flies are loaded to individual chambers at ZT8-ZT9 via a mouth pipette or a small vacuum pump. Individual chambers are sealed with 7×7 mm acrylic windows. Windows are coated with SigmaCote to prevent flies from ventral or dorsal postural positions. 5-7 day old female

Experiment Name	Experiment Type	Fly Gender	# of Frames
FlyF1-03082020164520	wild type sleep	female	1,727,979
FlyF11-01182022175505	wild type sleep	female	1,727,979
FlyF14-08172021175459	wild type sleep	female	1,727,979
FlyF15-08182021173222	wild type sleep	female	1,727,979
FlyF8-08112021174107	wild type sleep	female	1,727,979
FlyM13-08172021175457	wild type sleep	male	1,727,979
FlyM4-03062020153616	wild type sleep	male	1,727,979
FlyF19SD-11052021164243	sleep-deprived	female	2,159,979
FlyF19SD-11152020170647	sleep-deprived	female	2,159,979
FlyF41SD-11192021170807	sleep-deprived	female	2,159,979
FlyF52SD-11242021161943	sleep-deprived	female	2,159,979

Table 2.1 Details of the collected experimental data.

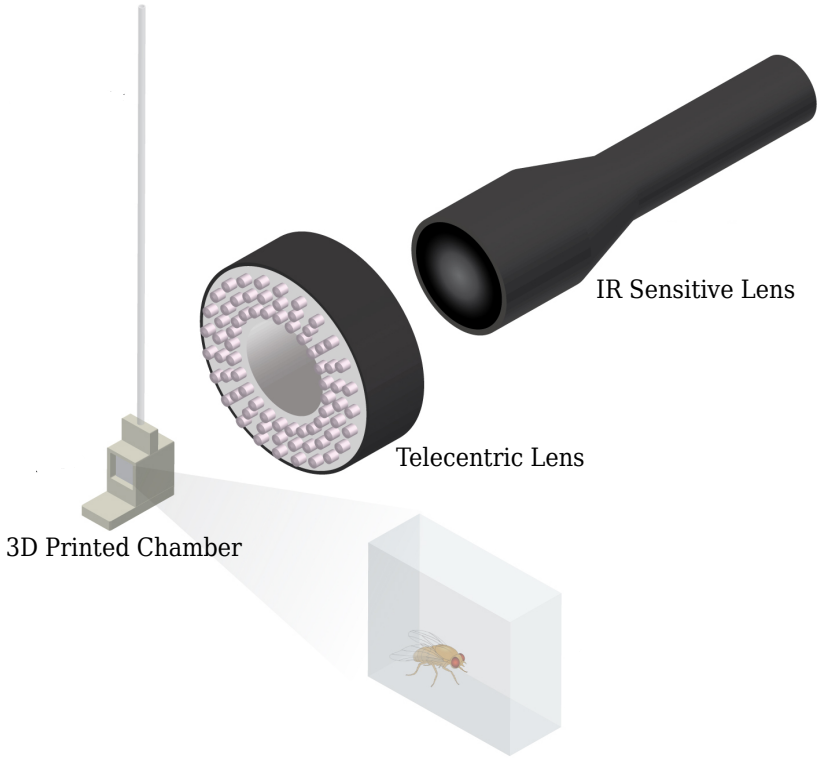


Figure 2.1 An illustration of the experimental setup used to perform high-resolution imaging of experiments.

and male flies are used in the experiments.

2.2 Pose Estimation and Tracking

A recently developed deep neural network based software, DeepLabCut (Mathis et al., 2018) was used to achieve markerless pose estimation. Over 20 body parts are first labeled in 1654 images from 28 animals (16 female, 12 male) to train the model with a 95/5 train/test split. An example labeled frame is given in Figure 2.2. We used a ResNet-50 (He et al., 2016) based neural network with a batch size of 4 and 200,000 training iterations. The rest of the settings were kept default. The resulting network has a test error of 3.67 pixels.

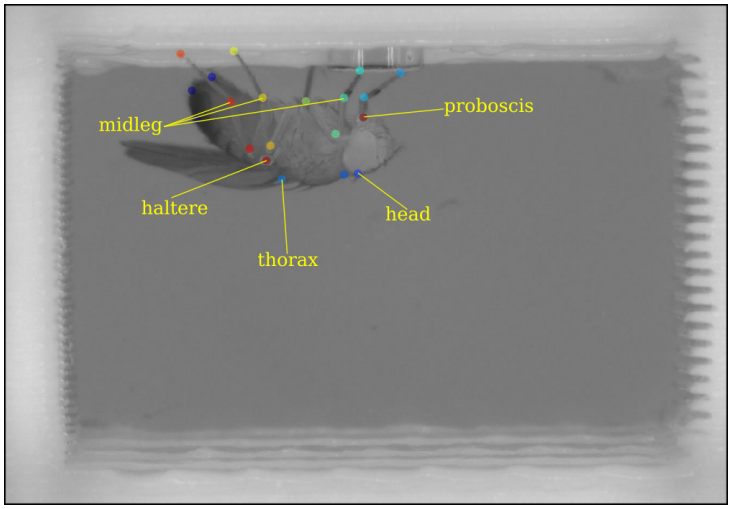
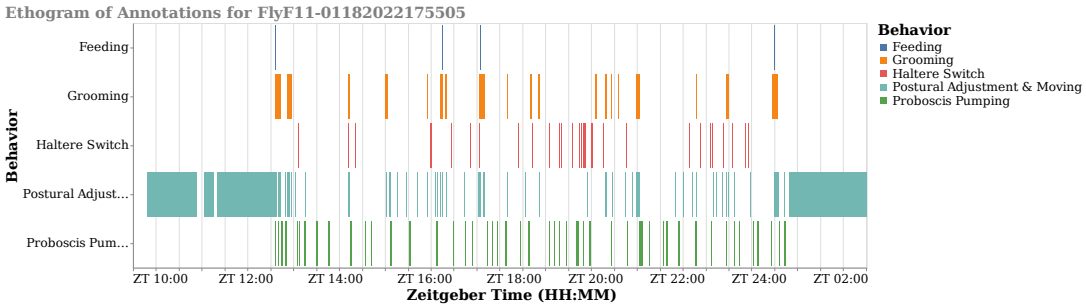


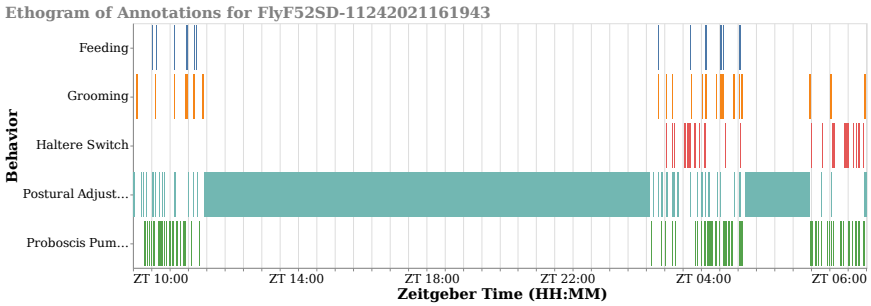
Figure 2.2 An example video recording frame of the fruit fly placed in a 3D printed chamber. Colorful markers indicate the tracked body parts.

2.3 Behavior Annotations

3 human annotators labeled 5 different behaviors (feeding, grooming, moving, haltere switch, proboscis pumping) across 11 videos (14 hours and 16 hours each, respectively for 7 wild-type sleep experiments and 4 sleep deprivation experiments). A single experiment is annotated by all 3 annotators to check rigor and overlap among annotators. We only used the annotations of a fly when each pair of annotators agreed at least on the 90% of the annotations. Two example ethograms generated from the annotations are given in Figure 2.3.



(a) Wild type.



(b) Sleep-deprived.

Figure 2.3 Two example ethograms of annotated behaviors observed during the sleep experiments. The dark period starts at ZT12 and ends at ZT0.

3. FEATURE EXTRACTION

3.1 Overview of Feature Extraction

For a single experiment data, i.e., a single fruit fly recorded between ZT10 and ZT2 (zeitgeber time 10 and 2), feature extraction consists of four consecutive steps, where the input in the present stage is the output of the previous one. In this study, the input of the first step is the raw output signal of the tracking and pose estimation model which is produced by DeepLabCut, a toolbox for markerless pose estimation. The feature extraction steps are as follows:

1.1 Constructing pose values and preprocessing; dealing with occluded body parts, alignment of different orientations, filtering, and imputation.

1.2 Computing spatio-temporal features, such as distances between body parts, velocity, and angular velocity from body part positions.

1.3 Computing dynamic postural features by extending spatio-temporal features to multiple timescales using wavelet transformation and sliding window statistics.

1.4 Computing normalized high-dimensional behavioral representations.

Matrices $\mathbf{X} \in \mathbb{R}^{T \times N}$ and $\mathbf{Y} \in \mathbb{R}^{T \times N}$ denote multivariate time series for x and y cartesian components of two-dimensional video recordings. These data are collected for N tracked body parts of a fly in T consecutive time stamps by a pose estimation model. This

multivariate time series matrices $\mathbf{X}$ and $\mathbf{Y}$ are the raw input data that goes into the first step of the feature extraction. Note that the number of body parts, N, must be the same among all experiments conducted with different fruit flies, but the number of time stamps, T, might differ. Each column of the $\mathbf{X} = \left[\mathbf{x}_1, \cdots, \mathbf{x}_N\right]^\top$ and $\mathbf{Y} = \left[\mathbf{y}_1, \cdots, \mathbf{y}_N\right]^\top$ can be written respectively as follows:

$$\begin{aligned} \mathbf{y}_i &= (y_{i,1}, y_{i,2}, \ldots, y_{i,t-1}, y_{i,t}, y_{i,t+1}, \ldots, y_{i,T}), \\ \mathbf{x}_i &= (x_{i,1}, x_{i,2}, \ldots, x_{i,t-1}, x_{i,t}, x_{i,t+1}, \ldots, x_{i,T}). \end{aligned} \tag{3.1}$$

Here i denotes the index of the body part, e.g., leg tip or proboscis.

In addition to $\mathbf{X}$ and $\mathbf{Y}$, a pose estimation model may report prediction scores for each tracked body part at each time step, which is the case for DeepLabCut as well. $L \in \mathbb{R}^{N \times T}$ denotes the time series of prediction scores, each column of the $L = [\mathbf{l_1}, \cdots, \mathbf{l_N}]^\top$ can be written as follows:

$$\mathbf{l}_i = (l_{i,1}, l_{i,2}, \ldots, l_{i,t-1}, l_{i,t}, l_{i,t+1}, \ldots, l_{i,T}). \tag{3.2}$$

The prediction scores tend to be very low when the body part is not visible. Thus, L provides valuable information about the occluded body parts. In the Section 3.2.1, how L is incorporated into construction of pose values is described in detail.

3.2 Preprocessing

This step involves preprocessing the signal by filtering and imputation of certain video frames. But in addition to this, there are a couple of optional procedures that can be beneficial for our task of learning stereotypical behaviors. These additional procedures deal with the occluded body parts of the fly, alignment of the fly orientations and defining new points of interest.

3.2.1 Occluded Body Parts

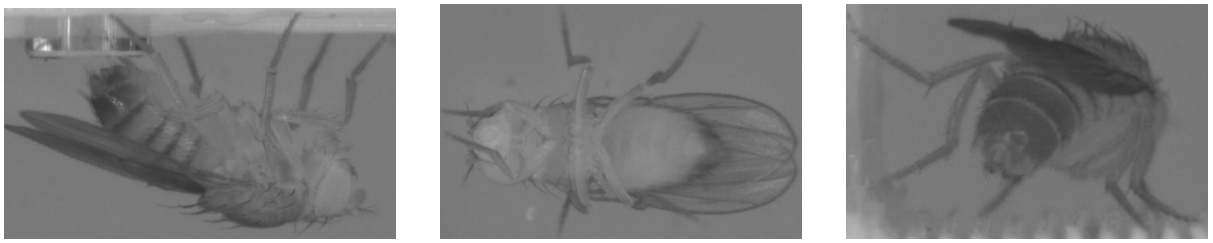
The two-dimensional nature of the video recordings introduces several significant challenges, one of which is the occluded body parts. Occlusion describes various cases where the tracked parts are eclipsed by other parts in the view area. As we are interested in behaviors that are characterized by particular body parts, occlusion introduces a great challenge. Obviously, it is impossible to capture a switch-like movement of haltere, when haltere is occluded.

There are many types of occlusions. One type is short occlusions that often results from postural changes. Imputation of the time series $\mathbf{X}$ and $\mathbf{Y}$ for such short occlusions is relatively easy since the number of consecutive missing data points is limited. However, this is not the case for long occlusions, which usually occur when the fly is in dormancy for an extended period of time. Especially for the body parts with left and right counterparts, fly's orientation can result in one of the counterparts being occluded for long dormancy periods. We use imputation and elimination of the corresponding data-points to deal with the occlusions. Before describing those approaches, we define a criterion for target occlusion.

3.2.1.1 Oriented Pose Values for Body Parts with Left & Right Counterparts

If the fly is oriented perpendicular to the camera perspective, as in Figure 3.1c, then one of the left and right body parts is often occluded. In other orientations (e.g., Figure 3.1a and Figure 3.1b), occlusion of both sides is possible as well as the case of no occlusion. However, in the conducted experiments, flies usually choose to stay dormant perpendicular to the camera perspective for long durations, as mentioned in Chapter 2. In such cases, one can concede using only one of the left or right counterparts to construct pose values. Therefore, this optional step is included in the behavioral mapping pipeline to reduce pose values of body parts with left and right counterparts to a single value.

We use prediction scores to determine which body part should be used to compute oriented pose values. Let i and j be a pair of body parts which are left and right counterparts of each other, e.g., left and right haltere. Then, one can use the $\mathbf{l}_i$ and $\mathbf{l}_j$ to predict if one of them is occluded at a particular time step t. Let orient($\mathbf{x}_i$,$\mathbf{x}_j$) (orient($\mathbf{y}_i$,$\mathbf{y}_j$)) be a new pose vector that will be computed based on $\mathbf{x}_i$ and $\mathbf{x}_j$ ($\mathbf{y}_i$ and $\mathbf{y}_j$), e.g., a vector of oriented haltere pose values, composed of left and right haltere pose values. The following conditional procedures are proposed to compute oriented pose values from the left and right pose values by deciding the orientation of the fly for a counterpart body pair. The



(a) Parallel. **(b)** Oblique. **(c)** Perpendicular.

Figure 3.1 Three examples demonstrating the different orientations viewed by the camera. Orientations similar to Figure 3.1b and Figure 3.1c results in occluded body parts and hence erroneous tracking.

procedures below can be used separately or successively.

- If $l_{i,t} - l_{j,t} \geq \epsilon$, then, without loss of generality, $\text{orient}(\mathbf{x}_i, \mathbf{x}_j)_t = x_{i,t}$ and $\text{orient}(\mathbf{y}_i, \mathbf{y}_j)_t = y_{i,t}$ for a threshold ϵ, typically $\epsilon > 0.5$.
- If $\left|\left\{ t' \,:\, l_{i,t'} > l_{j,t'}, t' \in [t-\tau \,..\, t+\tau] \right\}\right| > \tau$, then, without loss of generality, $\text{orient}(\mathbf{x}_i, \mathbf{x}_j)_t = x_{i,t}$ and $\text{orient}(\mathbf{y}_i, \mathbf{y}_j)_t = y_{i,t}$, for a window of size $2w+1$.
- If $l_{i,t} > l_{j,t}$ and if the nearest confident left orientation is closer than the nearest confident right orientation, i.e.,

$$\arg\min_{t'} \left\{ |t-t'| \,:\, l_{i,t'} - l_{j,t'} \geq \epsilon \right\} > \arg\min_{t'} \left\{ |t-t'| \,:\, l_{j,t'} - l_{i,t'} \geq \epsilon \right\},$$

  then, without loss of generality, $\text{orient}(\mathbf{x}_i, \mathbf{x}_j)_t = x_{i,t}$ and $\text{orient}(\mathbf{y}_i, \mathbf{y}_j)_t = y_{i,t}$.
- If simply $l_{i,t} > l_{j,t}$, then without loss of generality, $\text{orient}(\mathbf{x}_i, \mathbf{x}_j)_t = x_{i,t}$ and $\text{orient}(\mathbf{y}_i, \mathbf{y}_j)_t = y_{i,t}$.

Except for the direct comparisons based on prediction confidence scores as in the last procedure, some of the time points might be left with undecided orientations. If the number of such time points is manageably small, then direct comparisons of prediction scores for those time points are convenient and handy.

After applying the above procedures for a left and right counterpart pair i and j, we can define oriented multivariate time series as

$$\begin{aligned} \mathbf{X}^o &= \left(\left[(\mathbf{x}_k)_{k \notin \bigcup_{\{i,j\} \in \mathcal{O}} \{i,j\}} \right]^\top \,\middle|\, \left[\left(\text{orient}(\mathbf{x}_i, \mathbf{x}_j) \right)_{\{i,j\} \in \mathcal{O}} \right]^\top \right), \\ \mathbf{Y}^o &= \left(\left[(\mathbf{y}_k)_{k \notin \bigcup_{\{i,j\} \in \mathcal{O}} \{i,j\}} \right]^\top \,\middle|\, \left[\left(\text{orient}(\mathbf{y}_i, \mathbf{y}_j) \right)_{\{i,j\} \in \mathcal{O}} \right]^\top \right), \end{aligned} \tag{3.3}$$

where $\mathcal{O}$ is the set of index pairs of left and right counterparts and $\bigcup_{\{i,j\} \in \mathcal{O}} \{i,j\}$ is the

union of all indexes of such body part pairs. Applying the procedures described above for each left and right counterparts results in computing $\mathbf{X}^o$ and $\mathbf{Y}^o$. Such oriented versions of the original data matrices can be used instead of $\mathbf{X}$ and $\mathbf{Y}$ in the rest of the pipeline, if desired.

3.2.1.2 Finding Occlusions using Prediction Scores & Anomalous Pose Values

Major postural changes and overlapping body parts during movements result in some body parts being occluded for a short interval of time. These types of occlusions may span several frames to several seconds. Since the pose estimation model makes predictions for such data points, it is necessary to detect and process such data points. Our pipeline exploits two indicators for detection: prediction scores and anomalous changes in pose values. We first mark the body parts estimated to be occluded and the corresponding time intervals. In the ensuing step, the desired imputation method is used to fill those time intervals appropriately.

The following conditions are considered to detect occluded body parts, note that they can be used separately or in combination with each other.

- If the prediction confidence score at time step t is lower than a given threshold ϵ, then the t is marked to be imputed.
- If z-score of the prediction confidence score $l_{i,t}$ computed within a window with size τ, and centered at t, is lower than a given threshold ϵ_z, then the t is marked to be imputed.
- If the second-order gradient of the estimated pose value exceeds a given threshold δ, then the t is marked to be imputed.
- If the difference between the estimated pose value at time step t and the median of pose values within the window of size τ, centered at time step t, exceeds the threshold δ, then the t is marked to be imputed.

Here, the window sizes and threshold parameters are determined separately for each condition. The conditions described above are not only useful for the occluded body parts, but are also beneficial for tackling unnatural and abnormal predictions of the pose estimation model.

3.2.2 Aligning Different Orientations

It may be desirable to align different orientations and postures in some cases. For instance, it is not possible to interpret vertical and horizontal replacements of body parts separately without aligning them into a reference frame. Flies can position themselves in different orientations while exhibiting similar actions, as it can be seen from the Figure 3.1a.

To rotationally align each frame, we transform body part coordinate values into an egocentric reference frame centered in the middle of the fly's spine, e.g., a line along the thorax. Then, we performed a linear transformation on the pose values to center the spine at the origin, oriented along the y-axis.

Alignment of the frames is an optional step in the behavioral mapping pipeline. It is potentially beneficial, and reasonable, to perform alignment when the Cartesian pose values of the body parts are included as spatio-temporal features.

3.2.3 Filtering and Imputation

Estimated pose values are highly noisy because of the reasons explained in Chapter 1. Thus, filtering the pose values and imputation of the data points marked as occluded or abnormal is an essential step before proceeding with the rest of the pipeline, which contains stages that are sensitive to noise.

One of the most practical and effective approaches for time series imputation is applying interpolation. We also benefit from univariate interpolation to replace values at marked time points, where the exact algorithm is chosen to be one of the following; linear interpolation, spline interpolation, forward filling, or backward filling.

After imputation, we apply a median filter with appropriate window size and smooth each $\mathbf{x_i}$ and $\mathbf{y_i}$ separately. Finally, a boxcar filter (moving average) with a relatively small window size is used to filter out the rapidly changing signals by averaging. We observed that large window sizes may result in smoothing out some critical signals, which potentially define short-duration low-amplitude behaviors, e.g., switch-like haltere movement behavior.

A Rauch-Tung-Striebel Kalman smoother (Rauch et al., 1965) was employed in the development stage. However, no significant performance improvement is observed, and hence

later abandoned due to its computational cost and the requirement of setting the various state parameters to reasonable values.

The configuration and actual parameters of the imputation methods and the filters are provided in Chapter 6.1.

3.3 Computation of Spatio-temporal Features

After preprocessing of pose values, learning stereotypical behaviors becomes feasible. Although tracking of relevant body parts and processing corresponding pose values is an essential step for quantifying behavior, a set of coordinate values is not sufficient to represent and capture complex spatio-temporal dynamics of animal behavior. There are thousands of unique postures, and behaviors are not even exhibited by some static set of postures. Instead, they are defined by expressive and meaningful spatio-temporal features such as distances, velocities, angles, and angular velocities. Therefore, one needs to compute such features from the coordinate values of body parts in two-dimensional space.

The second stage of the feature extraction is the computation of spatio-temporal features from pose values. Two types of features are computed in this stage, as listed below.

2.1 **Snapshot features:** Spatio-temporal feature values computed at a snapshot of time, listed as follows:

- distances,
- angles,
- cartesian pose values (i.e., per body part features).

2.2 **Gradient features:** Spatio-temporal feature values computed based on how snapshot features change over time, listed as follows:

- change of distances,
- change of angles (i.e., angular velocities),
- change of cartesian pose values (i.e., body part velocities).

The gradients of snapshot features are computed using second-order accurate central differences in the interior points. The resulting gradient features have the same shape, i.e., the number of features and the number of time-stamps, as the snapshot features.

3.3.1 Distances Between Body Parts

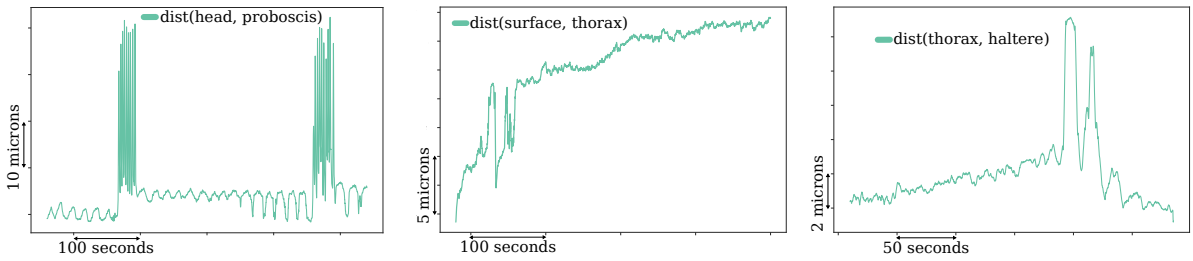
Given a body part pair (i, j), the distance between them at a time step t is calculated with the Euclidean distance, given below,

$$d_t^{i,j} = \sqrt{(x_{i,t} - x_{j,t})^2 + (y_{i,t} - y_{j,t})^2}. \tag{3.4}$$

The corresponding gradient feature, which is the change of distance between the body part i and j, is computed using the second-order gradient approximation,

$$\dot{d}_t^{i,j} = \begin{cases} \frac{|d_{t+1}^{i,j} - d_t^{i,j}|}{\Delta t} & \text{if t=0 or t=T,} \\ \frac{|d_{t+1}^{i,j} - d_{t-1}^{i,j}|}{2\Delta t} & \text{otherwise,} \end{cases} \tag{3.5}$$

where Δt is the sampling period, and it is equal to $^1/_{\text{FPS}}$ seconds.



(a) Pumping-like movement of the proboscis. **(b)** Postural adjustment and movement of the thorax. **(c)** Switch-like movement of the haltere.

Figure 3.2 Three spatio-temporal feature examples describing kinematics of different behaviors.

3.3.2 Joint Angles Between Body Parts

Given a triplet of body parts (i, j, k), the angle between i and k around j is calculated using the two-argument arctangent function as given below,

$$\omega_t^{i,j,k} = \text{atan2}\left(\det\begin{bmatrix} x_{i,t}-x_{j,t}, x_{k,t}-x_{j,t} \\ y_{i,t}-y_{j,t}, y_{k,t}-y_{j,t} \end{bmatrix}, \begin{bmatrix} x_{i,t}-x_{j,t} \\ y_{i,t}-y_{j,t} \end{bmatrix} \cdot \begin{bmatrix} x_{k,t}-x_{j,t} \\ y_{k,t}-y_{j,t} \end{bmatrix}\right) + \pi. \tag{3.6}$$

Then, similar to the change of distance features, the angular velocities are approximated by

$$\dot{\omega}_t^{i,j,k} = \begin{cases} \frac{|\omega_{t+1}^{i,j,k} - \omega_t^{i,j,k}|}{\Delta t} & \text{if t=0 or t=T,} \\ \frac{|\omega_{t+1}^{i,j,k} - \omega_{t-1}^{i,j,k}|}{2\Delta t} & \text{otherwise.} \end{cases} \tag{3.7}$$

3.3.3 Cartesian Pose Values of Body Parts

The cartesian pose values of a body part i are straightforwardly given by the x and y coordinate values as follows:

$$\begin{aligned} x_t^i &= x_{i,t}, \\ y_t^i &= y_{i,t}. \end{aligned} \tag{3.8}$$

Note that for such a single body part, two feature values are generated. Corresponding gradient features, namely the body part velocities along each cartesian component, are computed by

$$\begin{aligned} \dot{x}_t^i &= \begin{cases} \frac{x_{t+1}^i - x_t^i}{\Delta t} & \text{if t=0 or t=T,} \\ \frac{x_{t+1}^i - x_{t-1}^i}{2\Delta t} & \text{otherwise,} \end{cases} \\ \dot{y}_t^i &= \begin{cases} \frac{y_{t+1}^i - y_t^i}{\Delta t} & \text{if t=0 or t=T,} \\ \frac{y_{t+1}^i - y_{t-1}^i}{2\Delta t} & \text{otherwise.} \end{cases} \end{aligned} \tag{3.9}$$

In order to compute the overall two-dimensional velocity of a body part, one can always use the distance between the origin and corresponding body part.

3.3.4 Constructing Spatio-temporal Feature Matrices

Let $\mathcal{C}, \mathcal{D}$, and $\mathcal{A}$ denote the sets of body parts, body part pairs, and body part triplets; respectively defining cartesian pose values, distances, and angles, respectively. Similarly, let $\mathcal{C}', \mathcal{D}'$, and $\mathcal{A}'$ denote sets that define sets of respective gradient features. Then the snapshot feature matrix S constructed as follows:

$$\mathbf{S} = \Big(\Big[(\mathbf{x}^i)_{i \in \mathcal{C}} \Big] \Big| \Big[(\mathbf{y}^i)_{i \in \mathcal{C}} \Big] \Big| \Big[(\mathbf{d}^{i,j})_{\{i,j\} \in \mathcal{D}} \Big] \Big| \Big[(\boldsymbol{\omega}^{i,j,k})_{\{i,j,k\} \in \mathcal{A}} \Big] \Big), \tag{3.10}$$

where the vectors are defined as $\mathbf{x}^i = \big[x_1^i, \cdots, x_T^i\big]$, $\mathbf{y}^i = \big[y_1^i, \cdots, y_T^i\big]$, $\mathbf{d}^{i,j} = \big[d^{i,j}, \cdots, d_T^{i,j}\big]$, and $\boldsymbol{\omega}^{i,j,k} = \big[\omega_1^{i,j,k}, \cdots, \omega_T^{i,j,k}\big]$.

Similarly, for gradient features, the feature matrix is constructed by concatenating change of distances, angular velocities and body part velocities; given by

$$\mathbf{G} = \Big(\Big[(\dot{\mathbf{x}}^i)_{i \in \mathcal{C}'} \Big] \Big| \Big[(\dot{\mathbf{y}}^i)_{i \in \mathcal{C}'} \Big] \Big| \Big[\big(\dot{\mathbf{d}}^{i,j}\big)_{\{i,j\} \in \mathcal{D}'} \Big] \Big| \Big[(\dot{\boldsymbol{\omega}}^{i,j,k})_{\{i,j,k\} \in \mathcal{A}'} \Big] \Big), \tag{3.11}$$

where the vectors are defined as $\dot{\mathbf{x}}^i = \big[\dot{x}_1^i, \cdots, \dot{x}_T^i\big]$, $\dot{\mathbf{y}}^i = \big[\dot{y}_1^i, \cdots, \dot{y}_T^i\big]$, $\dot{\mathbf{d}}^{i,j} = \big[\dot{d}^{i,j}, \cdots, \dot{d}_T^{i,j}\big]$, and $\dot{\boldsymbol{\omega}}^{i,j,k} = \big[\dot{\omega}_1^{i,j,k}, \cdots, \dot{\omega}_T^{i,j,k}\big]$.

The resulting two feature matrices are $\mathbf{S} \in \mathbb{R}^{T \times (2|\mathcal{C}| + |\mathcal{D}| + |\mathcal{A}|)}$, namely snapshot feature matrix, and $\mathbf{G} \in \mathbb{R}^{T \times (2|\mathcal{C}'| + |\mathcal{D}'| + |\mathcal{A}'|)}$, namely gradient feature matrix. N_S denotes the number of snapshot features, which is given by $2|\mathcal{C}| + |\mathcal{D}| + |\mathcal{A}|$ and N_G denotes the number of gradient features, which is equal to $2|\mathcal{C}'| + |\mathcal{D}'| + |\mathcal{A}'|$.

3.4 Computation of Dynamic Postural Features

Instantaneous values of spatio-temporal features do not provide a sufficient description of complex postural dynamics of behaviors. Understanding the output of a complex biological system, in our case the behavior, can only be achieved by studying multiple timescales concurrently. Previous studies attempted to search behavioral motifs, e.g., repeated sub-sequences of actions with finite length, within the behavioral time series (Ye and Keogh, 2011; Brown et al., 2013). However, as Berman et al. (2014) states, this paradigm usually requires problems of temporal alignment and relative phasing between

different scales. Alternatively, extending spatio-temporal features to capture postural dynamics at different timescales eliminate the requirements of temporal alignment and motif-based analysis. In order to extend the spatio-temporal features to dynamic postural features, we applied wavelet transformation (similar to Berman et al. (2014)) and computed moving statistics at different timescales (similar to Kabra et al. (2013)), respectively for the snapshot feature set ($\mathbf{S}$) and the gradient feature set ($\mathbf{G}$).

3.4.1 Moving Statistics of Gradient Features

Gradient features only reflect the instantaneous values of velocities with respect to the sampling rate. In order to capture how these values change within a given interval, the moving statistics of gradient features such as the mean and the standard deviation are computed with a sliding window approach.

Let τ be the window size parameter, i.e., the timescale of interest, then the moving mean of the corresponding gradient feature $\mathbf{g}_i$ is given by the function μ_τ, as given below:

$$\mu_\tau(g_{i,t}) = {}^{1}\!/_{\min\{t+\tau,\,T\}-\max\{t-\tau,\,1\}+1} \sum_{t'=\max\{t-\tau,\,1\}}^{\min\{t+\tau,\,T\}} g_{i,t'}. \tag{3.12}$$

Similarly, the moving standard deviation of a gradient feature $g_{i,t}$ by is computed by σ_τ, as in the below equation.

$$\sigma_\tau(g_{i,t}) = \left({}^{1}\!/_{\min\{t+\tau,\,T\}-\max\{t-\tau,\,1\}+1} \sum_{t'=\max\{t-\tau,\,1\}}^{\min\{t+\tau,\,T\}} \left(\mu_\tau(g_{i,t}) - g_{i,t'}\right)^2 \right)^{1/2} \tag{3.13}$$

Moving statistics feature generation approach has been used to learn animal behavior by Kabra et al. (2013), and Marshall et al. (2021) has also included such features into the analysis.

3.4.2 Wavelet Transformation of Snapshot Features

The wavelet domain is a useful representation of postural dynamics due to the following reasons given by Berman et al. (2014), and the proposed spectrogram generation is used by others as well (Marshall et al., 2021; Todd et al., 2017).

- It describes dynamics over multiple timescales simultaneously by possessing a multi-resolution time-frequency trade-off.
- It eliminates the requirement of precise temporal alignment for capturing periodic behaviors by taking amplitudes of the continuous wavelet transform of each snapshot feature at different scales.

Given a function $s(t)$, the continuous wavelet transformation at a frequency $f > 0$ is expressed as the following integral:

$$W_{f,t'}\left[s\left(t\right)\right] = \frac{1}{\sqrt{a\left(f\right)}} \int_{-\infty}^{\infty} s\left(t\right) \Psi^{*}\left(\frac{t-t'}{a(f)}\right) \mathrm{d}t, \tag{3.14}$$

where Ψ is the wavelet function and a is a function for converting frequencies to wavelet scale factor. The Morlet wavelet is well suited for describing postural dynamics which is closely related to human aural and vision perception (Daugman, 1985), and it is used in the pipeline. The corresponding wavelet function is given by

$$\Psi(t) = \exp\left\{\frac{t^2}{2}\right\} \cos(w_0 t), \tag{3.15}$$

where w_0 is a non-dimensional parameter. The frequency-to-scale conversion function a for the Morlet wavelet is as follows:

$$a\left(f\right) = \frac{w_0 + \sqrt{2 + w_0^2}}{4\pi f}. \tag{3.16}$$

For the discrete sequence of snapshot feature $\mathbf{s}_i$ with sampling period Δt, $W_{f,t'}[s(t)]$ translates into

$$W_f\left(\mathbf{s}_i, t'\right) = \frac{1}{\sqrt{a(f)}} \sum_{t=1}^{T} \Delta t s_{i,t} \Psi^{*}\left(\frac{t-t'}{af}\right), \tag{3.17}$$

where $t', t \in \mathbb{Z}$ and $1 \leq t' \leq T$ (Torrence and Compo, 1998).

3.4.2.1 Normalization of Wavelet Power Spectrum

In order to ensure that wavelet transforms (Equation 3.17) at each frequency f, are directly comparable to each other and to the other transformed time series, the transformation W_f has to be normalized at each frequency f to have unit energy. This normalization for the Morlet wavelet at frequency f is as follows:

$$C(f) = \frac{\pi^{-\frac{1}{4}}}{\sqrt{2a(f)}} \exp\left\{ 1/4 \left(w_0 - \sqrt{w_0^2 + 2} \right)^2 \right\}. \tag{3.18}$$

The resulting normalized transformation, which is also used to generate the spectrogram in Berman et al. (2014), is given by

$$W_f^0(\mathbf{s}_i, t') = \frac{1}{C(f)} |W_f(\mathbf{s}_i, t')|. \tag{3.19}$$

In addition to the above conventionally used normalization, we alternatively adopted the normalization proposed by Liu et al. (2007). According to this alternative adjustment, the wavelet power spectrum should be equal to the transform coefficient squared divided by the scale it associates.

$$W_f^0(\mathbf{s}_i, t') = \frac{W_f(\mathbf{s}_i, t')^2}{a(f)} \tag{3.20}$$

We observed substantial improvements using this power spectrum.

3.4.2.2 Determining Spectrum Frequencies

We investigate two different approaches for computing a set of frequencies, and we include both of them in the behavior mapping pipeline. One set is dyadically spaced frequencies between $f_{\min}$ and $f_{\max}$ via

$$f_i = f_{\max} 2^{-\frac{i-1}{N_f - 1} \log \frac{f_{\max}}{f_{\min}}}, \tag{3.21}$$

where $f_{\max} = {}^{FPS}/_2$ Hz is the Nyquist frequency.

The other alternative set of frequencies is linearly spaced between $f_{\min}$ and $f_{\max}$ by

$$f_i = f_{\min} + \frac{f_{\max} - f_{\min}}{N_f - 1} i, \tag{3.22}$$

for $i = 1, 2, \ldots, N_f$, and their corresponding wavelet scales are computed by $a(f_i)$.

3.4.3 Constructing Dynamic Postural Feature Tensors

Let $\mathcal{T}_S = \{1/f_{\min}, \cdots 1/f_{\max}\}$ and $\mathcal{T}_G = \{\tau_{\min}, \cdots, \tau_{\max}\}$ denote the timescale sets respectively for wavelet transforms of snapshot features and moving statistics of gradient features. Then corresponding feature tensors are given as follows:

$$\begin{aligned} \mathbf{W} &= \left(W_f^0(\mathbf{s}_i, t)\right)_{1 \le t \le T,\, 1/f \in \mathcal{T}_S,\, 1 \le i \le N_S}, \\ \mathbf{M}^\mu &= \left(\mu_\tau\left(g_{i,t}\right)\right)_{1 \le t \le T,\, \tau \in \mathcal{T}_G,\, 1 \le i \le N_G}, \\ \mathbf{M}^\sigma &= \left(\sigma_\tau\left(g_{i,t}\right)\right)_{1 \le t \le T,\, \tau \in \mathcal{T}_G,\, 1 \le i \le N_G}. \end{aligned} \tag{3.23}$$

The resulting extended feature tensors of dynamic postural representations are $\mathbf{W} \in \mathbb{R}^{T \times |\mathcal{T}_s| \times N_S}$, $\mathbf{M}^\mu \in \mathbb{R}^{T \times |\mathcal{T}_v| \times N_G}$ and $\mathbf{M}^\sigma \in \mathbb{R}^{T \times |\mathcal{T}_v| \times N_G}$.

3.5 Computation of Behavioral Representations

After applying wavelet transformation or computing moving statistics to extend extracted spatio-temporal features to dynamic postural features, a couple of additional operations are required to continue in the behavior mapping pipeline.

3.5.1 Flattening Dynamic Postural Feature Tensors

As constructed in Section 3.4, dynamic postural feature tensors are $\mathbf{W} \in \mathbb{R}^{T \times |\mathcal{T}_S| \times N_S}$, $\mathbf{M}^\mu \in \mathbb{R}^{T \times |\mathcal{T}_G| \times N_G}$, and $\mathbf{M}^\sigma \in \mathbb{R}^{T \times |\mathcal{T}_G| \times N_G}$. In order to apply manifold learning-based dimensionality reduction algorithms or traditional machine learning algorithms such as decision trees, the last two dimensions of these feature tensors need to be contracted to obtain a matrix representation. As a result, feature matrices are $\tilde{\mathbf{W}} \in \mathbb{R}^{T \times (N_S |\mathcal{T}_S|)}$, $\tilde{\mathbf{M}}^\mu \in$

$\mathbb{R}^{T\times(N_G|\mathcal{T}_G|)}$ and $\tilde{\mathbf{M}}^{\sigma} \in \mathbb{R}^{T\times(N_G|\mathcal{T}_G|)}$ are obtained.

3.5.2 L_1 Normalization of Features

Dynamic postural feature distributions of similar behaviors may differ among flies due to different characteristics such as sex and sleep deprivation, Due to the two-dimensional nature of the video recordings, different orientations may cause observing different feature values for the same behavior. In order to have a homogeneous feature space among flies and throughout the temporal dimension, at each time step t, L_1 normalization is applied as follows:

$$\hat{\mathbf{w}}_i = \left(\frac{\tilde{w}_{t,i}}{\sum_{j=1}^{N_S|\mathcal{T}_s|} \tilde{w}_{t,j}} \right)_{1\leq t\leq T} \quad \text{and} \quad \hat{\mathbf{W}} = \left[(\hat{\mathbf{w}}_i)_{1\leq i\leq N_S|\mathcal{T}_S|} \right]. \tag{3.24}$$

Similarly, L_1 normalized versions of $\tilde{\mathbf{M}}^{\mu}$ and $\tilde{\mathbf{M}}^{\sigma}$, namely $\hat{\mathbf{M}}^{\mu}$ and $\hat{\mathbf{M}}^{\sigma}$ are obtained. Here $\hat{\mathbf{W}}$, $\hat{\mathbf{M}}^{\mu}$ and $\hat{\mathbf{M}}^{\sigma}$ are the final multivariate time series of normalized high dimensional behavioral representation of a single experiment data, i.e., single fruit fly recorded between ZT10 and ZT2. Notice that we may treat each time step, that is, the frame, as a discrete probability distribution after L_1 normalization.

4. ACTIVITY DETECTION

4.1 Overview of Activity Detection

Each experiment comprises sixteen hours of video recording spanning both awake and asleep epochs. Since we are only interested in the behavioral repertoire exhibited during sleep, time intervals where the animal is dormant, namely the dormancy epochs, should be detected before proceeding with the behavior mapping stage. We characterize dormancy epochs by lack of macro-activities, i.e., significant postural and locational changes, which can be detected by displacement of the animal. After detecting and excluding intervals of macro-activities, we end up with time points where the fly is dormant. An additional processing step is needed for the dormancy epochs, as we are not interested in the time points where the fly is totally quiescent. Our major focus is on the micro-activity bouts manifested during a dormancy epoch. In order to detect those bouts, we should distinguish micro-activities exhibited during dormant epochs from macro-activities by quantifying them with a closer look at various body parts. We use the term "bouts" and "epochs" respectively for micro-activities and macro-activities to reflect their difference in terms of duration. As it is shown in Section 6.2, behaviors categorized in micro-activities are tends to have shorter durations, compared to macro-activities.

At this stage, our ultimate goal is to extract bouts of micro-activities exhibited during the dormancy state. Extracted micro-activity bouts constitute the data points that are subject to behavior mapping. There are several benefits of reducing the data points to

this subset of dormancy and micro-activity, instead of using the entire experiment for behavior mapping. Considering the high frame rate and long length of video recordings, computational requirements are an important concern in our pipeline. Since at least the 90% of the frames are either totally quiescent or macro-activities, e.g., walking, this approach has the benefit of reducing the computational requirements significantly. Another critical point is that the quiescence frames contain only noise energy, and normalizing each frame amplifies me normalization amplifies this low-level noise energy, generating a uniform-like probability distribution for behavioral representation (Todd et al., 2017). Eliminating pure quiescence frames without any micro-activity avoids this. Also, as we are only interested in the behaviors exhibited during sleep, excluding macro-activity frames prevent the domination of large number of frames with walking and macro-activities in the behavioral embedding space.

In this chapter, we first describe the quantification of the macro-activities (Section 4.2.1) and micro-activities (Section 4.2.2) in the Section 4.2. After that, in the Section 4.3, we discuss two different approaches, unsupervised detection (Section 4.3.1) and supervised detection (Section 4.3.1), for detection of micro-activities and macro-activities, and constructing corresponding frame sets Dormancy, Macro-activity, and Micro-activity.

4.2 Quantifying Activities

4.2.1 Dormancy and Macro-activity Epochs

When a fly is awake, many behaviors are manifested by featuring major postural changes and displacement of the body in different ways. We categorize these types of behaviors under the umbrella term of macro-activity, and dormancy is defined as the lack of macro-activities and characterized by micro-activities. One can characterize macro-activities without considering their sub-categories by using the velocities, i.e., gradient features. In order to distinguish sub-categories of macro-activities, such as walking and rearing, more detailed and descriptive features are required. However, in our case, computing a single scalar value to capture the overall movement of a fly result is sufficient for detecting macro-activity epochs. We define this feature value by summing the gradient features for

all timescales, utilizing the dynamic postural feature tensor M^{μ}, as follows:

$$v_t = \sum_{\tau \in \mathcal{T}_G} \sum_{i=1}^{N_G} \mu_\tau (g_i, t), \tag{4.1}$$

where the resulting velocity-based feature vector is $\mathbf{v} = (v_1, \cdots, v_T)$. Essentially, high and low values of v_t indicate macro-activity and dormancy, respectively. Micro-activities can not be detected by using such a straightforward and general value, and therefore, can not be distinguished from dormancy by solely using only this quantity.

4.2.2 Micro-activity Bouts

Behaviors exhibited during dormancy epochs are not necessarily accompanied by postural changes and the displacement of the animal's body. For instance, the pumping-like movement of the proboscis or switch-like movement of haltere tends to happen independently from the remaining body parts. Thus, one needs to consider more than a single scalar value, as in the previous case of macro-activity. For each snapshot feature s_i, we sum all frequency channels f, i.e., timescales, utilizing the dynamic postural feature tensor W, as follows:

$$u_{t,i} = \sum_{1/f \in \mathcal{T}_G} W_f^0 (s_i, t) \quad \text{or} \quad u_{t,i} = \max_{1/f \in \mathcal{T}_G} W_f^0 (s_i, t). \tag{4.2}$$

The resulting feature vectors, $\mathbf{u}_i = \big(u_{i,1}, \cdots, u_{i,T}\big)$, are representative enough for capturing the micro-activities as an umbrella category. Such micro-activities potentially occur independently from the remaining body parts. For example, using the snapshot feature of the distance between haltere and posterior thorax, we are able to detect the switch-like movement of haltere in dormancy epochs, even if the rest of the body is at rest.

4.3 Detecting Activities

4.3.1 Unsupervised Approach

The straightforward approach for detecting macro activities would be determining a global threshold based on the distribution of the values of $\mathbf{v}$ and $\mathbf{u}$. Instead of determining such a threshold value, denoted by c, we use to treat the distribution of $\mathbf{v}$ and $\mathbf{u}$ as a mixture of Gaussian distributions to detect an appropriate threshold value, separately for each experiment. This approach avoids the hassle of dealing with the varying distributions of $\mathbf{v}$ and $\mathbf{u}$ among different experiments and constitutes a solid ground for the threshold value.

Consider a mixture of univariate Gaussian distributions with K components, $z = 1, \cdots, K$, sorted by their corresponding mean values $\mu_1, \cdots, \mu_K$, and with full covariance matrices. Then we can define $K-1$ many decision boundaries by:

$$\mathbb{P}\left(V = \delta_k \mid z = k\right) = \mathbb{P}\left(V = \delta_k \mid z = k+1\right), \tag{4.3}$$

where $k = 1, \cdots, K-1$, V denotes a random variable for values of v_t and δ_k is the threshold value for the kth decision boundary. After computing the decision boundaries for the mixture of Gaussian distributions, the macro-activity threshold is set to either one of the decision boundaries (δ_k for $k = 1, \cdots, K-1$) or one of the means (μ_k for $k = 1, \cdots, K$). For example, a reasonable choice would be $K = 2$ and threshold $c = \mu_2$, or $K = 3$ and threshold $c = \delta_2$. Classification is done by constructing the following frame sets:

$$\begin{aligned} \mathsf{Dormancy} &= \left\{ t : \ v_t \leq c, \ 1 \leq t \leq T \right\}, \\ \mathsf{Macro\text{-}activity} &= \left\{ t : \ v_t > c, \ 1 \leq t \leq T \right\}. \end{aligned} \tag{4.4}$$

We follow a similar approach for detecting micro-activities. Since we look for micro-activities in dormancy epochs, now we only consider the frames from the set $\mathsf{Dormancy}$. As micro-activity bouts are quantified by multiple feature vectors, we determine separate threshold values for each $\mathbf{u}_i$, using the same approach with macro-activity detection, utilizing a mixture of Gaussian distributions. After determining thresholds c_i for each $\mathbf{u}_i$, we construct the following frame sets:

$$\begin{aligned} \mathsf{Quiescence} &= \left\{ t : \ \bigwedge\nolimits_{i=1}^{N_G} u_{t,i} \leq c_i, \ t \in \mathsf{Dormancy} \right\}, \\ \mathsf{Micro\text{-}activity} &= \left\{ t : \ \bigvee\nolimits_{i=1}^{N_G} u_{t,i} > c_i, \ t \in \mathsf{Dormancy} \right\}. \end{aligned} \tag{4.5}$$

For a frame to be identified as micro-activity, it is sufficient if at least one feature is above

the corresponding threshold. This is due to the fact that micro-activities are manifested by the displacement of only a small subset of body parts' locations.

4.3.2 Supervised Approach

Annotations contain 5 behavioral categories, namely grooming, postural adjustments, proboscis pumping, haltere switch, and feeding. These behavioral categories correspond to the micro-activities that we are interested in. We formulate a binary classification problem by considering all behavioral categories as the positive class, and the rest as the negative class. After constructing the Dormancy set with the unsupervised approach described in Section 4.3.1, we train a random forest of decision trees (Breiman, 2001) with all the frames of annotated flies' Dormancy set. Similar to the unsupervised approach, we use $\mathbf{u}_i$ as the training features. For a single annotated fly, the resulting training feature matrix is $\mathbf{U} \in \mathbb{R}^{T \times N_G}$. In contrast to the original publication (Breiman, 2001), we use the Scikit-learn (Pedregosa et al., 2011) implementation which combines classifiers by averaging their probabilistic prediction, instead of letting each classifier vote for a single class. Micro-activity set is constructed with the frames predicted as the positive class, corresponding to the union of annotated behavioral categories. Similarly, the Quiescence set consists of the frames predicted as the negative class.

Configuration and the parameters of the random forest ensemble of decision trees are given in Section 6.1, and its performance is evaluated and compared with the unsupervised approach in the same section as well.

5. BEHAVIOR MAPPING

5.1 Overview of Behavior Mapping

Our ultimate goal is to have a fine-grained categorization of the behaviors observed during dormancy. Thus, it is not sufficient to detect the activities and construct the sets Micro-activity and Macro-activity. The behavior mapping stage is the most critical and novel part of the pipeline, and in this stage, we discover and predict stereotypical behaviors by mapping each frame in the sets Micro-activity.

The behavior mapping stage starts by generating low dimensional behavioral embeddings from the high dimensional behavioral representation matrix $\hat{\mathbf{W}}$, but only the rows corresponding to the Micro-activity are included in the mapping. Rest of the frames, namely the Quiescence set and Macro-activity set directly assigned to quiescent and moving categories. In Section 5.2, behavioral embedding generation is discussed in detail, including supervised, semi-supervised, and unsupervised approaches. We use semi-supervised pair UMAP embeddings, described in Section 5.2.3, to project behavioral representations into a low-dimensional behavior embedding.

The next step is a novel nearest neighbor analysis in the generated low-dimensional behavioral spaces, as we detail in Section 5.3. The nearest neighbor analysis consists of several parts. The first one is the computation of the behavioral weights (Section 5.3.1). Next, we combine behavioral weights provided by each "view", i.e., annotated experiment by forming a committee (Section 5.3.2). Finally, post-processing is described

in the Section 5.3.3.

5.2 Behavioral Embeddings

The dynamic postural features are able to capture many different timescales, however, their high-dimensional structure makes it challenging to directly exploit behavioral representations in analysis, learning, and visualization. For example, the behavioral representation matrix ($\hat{\mathbf{W}}$) computed from dynamic postural features ($\mathbf{W}$) with 20 spatio-temporal features and 25 timescales (i.e., frequency channels) has 20×25=500 columns. Since the correlation between different spatio-temporal features (e.g., the distance between head and proboscis and cartesian pose values of proboscis) and different timescales are often strong (see Figure 5.1), one may expect that the topological structure of the high dimensional behavioral representations ($\hat{\mathbf{W}}$) can be accurately represented in a lower dimensional space. Therefore, we would like to find a low-dimensional embedding that captures the important features of the dataset. The embedding we compute should minimize local distortions, since trajectories pause near a repeatable position whenever a particular stereotyped behavior is observed (Berman et al., 2014; DeAngelis et al., 2019; Ali et al., 2019).

Many dimensionality reduction algorithms seek to preserve the pairwise distance structure among all the data samples, the well-known examples of such algorithms are PCA (Hotelling, 1933), and MDS (Kruskal, 1964). Alternatively, some algorithms favor the preservation of local distances over global distance, such as UMAP (Uniform Manifold Approximation & Projection) (McInnes et al., 2020) t-SNE (Maaten and Hinton, 2008), Laplacian Eigenmaps (Belkin and Niyogi, 2003) and LargeVis (Tang et al., 2016). The latter category of algorithms aims to achieve to preserve important local structures and helps to improve classification performance when used in combination with learning algorithms where the function is only approximated locally, e.g., k-nearest neighbor classifier (McInnes et al., 2020). Similarly, it has been shown that manifold learning based dimensionality reduction algorithms can improve the clustering performance (Sainburg et al., 2021). Moreover, the trade-off of preserving local distances over global distances does not introduce any significant disadvantages in the latter stages of the behavioral pipeline.

We use UMAP for its superior performance in many aspects. McInnes et al. (2020) demonstrates that a k-nearest neighbor classifier trained on UMAP embeddings achieves

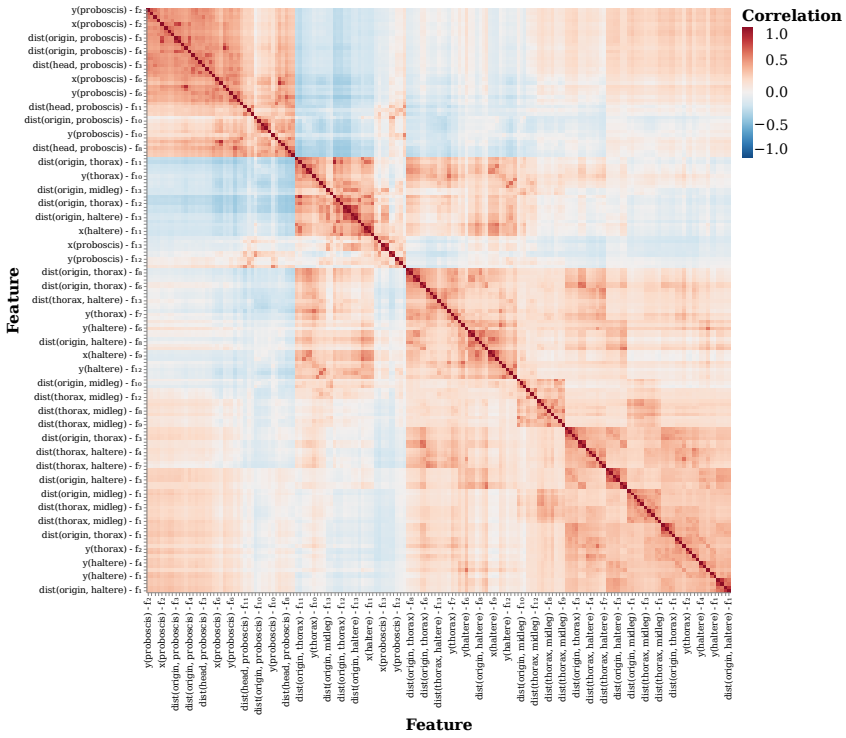


Figure 5.1 Spearman's correlation coefficient between each feature value of behavioral representations. Features are clustered and grouped based on the absolute value of the correlation values. Frequency channels $f_1, \cdots, f_20$ are dyadically spaced between 1 Hz and 20 Hz.

higher accuracy for large k values, compared to PCA, t-SNE, LargeVis, and Laplacian Eigenmaps since it captures non-local scales in a markedly effective way and local scales comparably or better. Thus, it can be argued that UMAP has captured more of the global and topological structure of the benchmark datasets than its counterparts, t-SNE and LargeVis. Another reason for choosing UMAP over other alternatives is that it tends to produce more stable embeddings. It is capable of producing sub-sample embeddings that are very close to the full embedding even for sub-samples of 5% of the dataset, outperforming the results of t-SNE and LargeVis. Finally, the computational performance of UMAP scales well with the number of samples, the embedding dimensionality, and the data dimensionality, compared to t-SNE LargeVis, and Eigenmaps, resulting in significantly lower run-times on various datasets such as COIL-20 (Nene et al., 1996), Fashion-MNIST (Xiao et al., 2017), and GoogleNews (Mikolov et al., 2013). Run-time performance of the dimensionality reduction algorithm is an important concern in the behavioral mapping pipeline, as the number of samples and the data dimensionality is often on the order of 10^6 and 10^2, respectively.

UMAP and other non-linear dimensionality reduction algorithms that attempt to use a

mathematical structure akin to a k-nearest neighbor graph to approximate a manifold, follow a similar basic structure as below (McInnes et al., 2020).

- Graph Construction
  - 3.1 Construct a weighted k-nearest neighbor graph.
  - 3.2 Apply some transformation on the edge weights to envelop local distances.
  - 3.3 Deal with the incompatibility of the local metrics, i.e., disagreeing weights of the edges.
- Graph Layout
  - 4.1 Define an objective function that preservers desired characteristics of the k-nearest neighbor graph.
  - 4.2 Compute a low dimensional representation by optimizing the defined objective function.

A distance metric is needed to construct a k-nearest neighbor (k-NN) graph. In our case, we normalized the feature matrices as described in Section 3.5.2, feature vector entries at each time step sum up to 1, and therefore the feature vectors can be considered as discrete probability distributions. We use the Hellinger distance (Hellinger, 1909) to quantify the similarity between "discrete probability distributions" of features, which is given by

$$\text{Hellinger}(P,Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{k} (\sqrt{p_i} - \sqrt{q_i})^2}. \tag{5.1}$$

Then, based on the Hellinger distances, UMAP constructs a weighted k-NN graph using nearest neighbor descent (Dong et al., 2011). Then, the k-NN graph is modified by making edges directed and defining the new weights using the local Riemannian metric of each data point. After this step, there exist two edges with disagreeing weights between the nearest neighbors, as the local metrics differ. UMAP combines those weights by using t-conorm and the resulting weight can be interpreted as the probability of at least one of the two directed edge to exist.

Finally, UMAP uses a force-directed graph layout algorithm in low dimensional space where attractive and repulsive forces are defined based on the gradients, optimizing the edge-wise cross-entropy between the original weighted graph and equivalent weighted graph induced by the embeddings. The algorithm proceeds by iteratively applying attractive and repulsive forces at each edge or vertex. Since the "true" graph captures the topology of the source data, the approximated graph also matches the overall topol-

ogy of the data, and thus produces a meaningful low-dimensional representation. A detailed mathematical and algorithmic description of the UMAP algorithm can be found in (McInnes et al., 2020), and it is skipped here as it goes beyond the scope of this work.

The algorithm described above is an unsupervised method, but it can be easily extended to work in a supervised or semi-supervised manner. Although we use a useful metric, e.g., Hellinger distance, defining the distance between a set of points, one can also define a simple metric for categorical values to extend UMAP further for supervised and semi-supervised cases. We can obtain a second view of the source data by using a metric where distances for points in the same and different categories as well as points without a category (for the semi-supervised case) are defined appropriately. A straightforward and simple example would be defining distances as follows: 1 if the points are in the same category, 0 if the points are in different categories, and 0.5 if either of the points is uncategorized. One can combine weighted graphs constructed using the two distance metrics (local metric and defined metric for categorical values), and arrive at a shared view of the data. In the behavioral mapping pipeline, we benefit from unsupervised UMAP, and its supervised and semi-supervised extensions for different purposes.

The utilized behavioral embeddings fall into three different categories, namely "disparate embeddings", "joint embeddings" and "pair embeddings", detailed descriptions and their applications are described respectively in Section 5.2.1, Section 5.2.2, and Section 5.2.3 respectively.

The resulting behavioral embedding

$$\mathbf{E} = \left[\mathbf{e}_1, \cdots \mathbf{e}_F\right] \in F \times N, \tag{5.2}$$

is a low-dimensional representation of

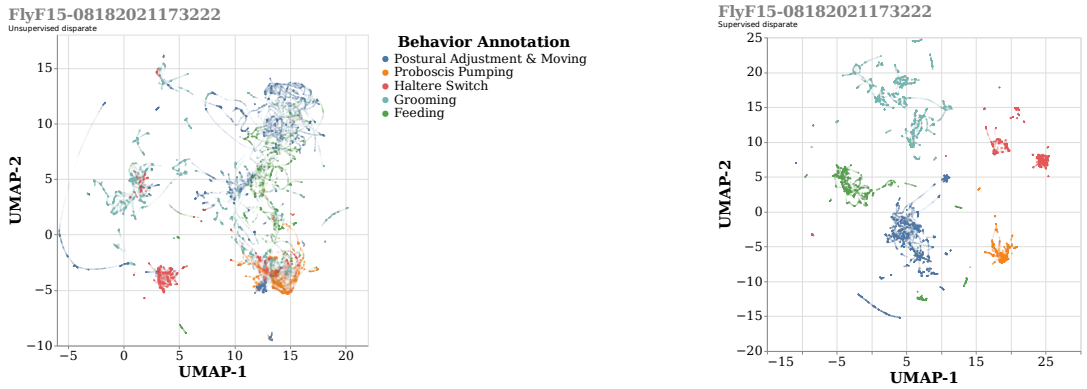
$$\text{row}_{t_f} \hat{\mathbf{W}} \in \mathbb{R}^{F \times (N_S |\mathcal{T}_S|)} \ \left(t_f \in \textsf{Micro-activity}\right), \tag{5.3}$$

where and $N << (N_S|\mathcal{T}_S|)$ and F is the numbers of frames estimated as dormant and active, being equal to $|\textsf{Micro-activity}|$. Each frame f, corresponds to a time point $t_f \in$ Micro-activity.

5.2.1 Disparate Embeddings

UMAP may be used to embed high-dimensional behavioral representations of each experiment separately to obtain disparate behavioral embeddings. Treating each experiment separately is useful for several purposes.

For example, using supervised UMAP for annotated experiments, we can explore behavioral sub-categories in annotations as annotations are very high-level, biased and general categorization of behaviors. For instance, one annotation category, e.g., "grooming", can be consisted of two different clusters in the behavioral embedding space, corresponding to "grooming of head" and "grooming of abdomen" (see Figure 5.2b). Defining very specific and low-level behavioral categories is not feasible and prone to error, a post-annotation analysis using disparate embeddings helps us to zoom in on annotated behaviors. Another scenario of benefiting disparate embeddings is using unsupervised UMAP for annotated and/or unannotated experiments separately. We can visualize how behavioral repertoire is represented in the low-dimensional embeddings space and analyze which features drive different regions and clusters of the behavioral embeddings (see Figure 5.2a).



(a) Unsupervised. **(b)** Supervised.

Figure 5.2 Two-dimensional supervised and unsupervised behavioral embeddings of FlyF15-08182021173222. Both embeddings reveal variations within annotated behavioral categories.

5.2.2 Joint Embeddings

Ideally, we would like to embed all experiments into a shared joint behavioral space, and using such an embedding would enable us to discover stereotypical and common behaviors among different flies. However, we observed that this is only possible to a certain extent and such joint embeddings tend to not mix well as the number of experiments increases. When we jointly compute a behavioral space using all available experiments, the resulting

embedding is usually not totally homogeneous in terms of flies and experiments. We observed that similar behaviors might end up embedded in different regions, and multiple clusters consisting of highly similar behaviors emerge.

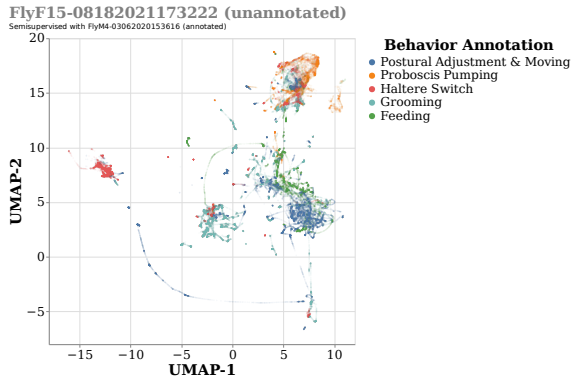
Especially for the joint treatment of annotated experiments, supervised UMAP fails to embed similar behaviors of different experiments closely and homogeneously. There are many factors contributing to this impairment of supervised joint embeddings, such as behavioral variations among flies and experiments, differences in orientation while exhibiting similar behaviors, and broad definitions of annotation categories. Unless the number of jointly embedded experiments does not exceed several, unsupervised UMAP performs relatively well, and enables a fully unsupervised and unbiased analysis of behaviors. We can exploit unsupervised joint behavioral embeddings for visualization purposes to discover how different feature combinations are exhibited, and density-based clustering to extract similar behavioral bouts. However, utilizing unsupervised joint embeddings becomes problematic as the number of experiments increases, and behavioral space gets "too crowded".

5.2.3 Pair Embeddings

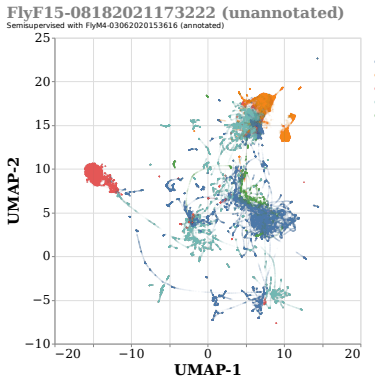
Using disparate embeddings does not allow one to embed an annotated and an unannotated experiment into a joint behavioral space, and joint embeddings poorly blend different experiments in a shared space. Instead, we propose a novel alternative approach to benefit from the semi-supervised dimensionality reduction capabilities of UMAP, while avoiding generating a hard-to-interpret embedding. In this approach, namely semi-supervised pair embeddings, we compute a joint behavioral space for each annotated and unannotated pair, using the available annotations. As a result, for R^- unannotated experiments and R^- annotated experiments, a semi-supervised pair embedding will be generated for each $R^- \times R^+$ pair.

For a single unannotated experiment, a semi-supervised behavioral embedding for each annotated experiment provides different "views". Especially when the behavioral repertoire of the annotated and the unannotated experiments are similar, the provided "view" turns out to be an accurate, easy-to-interpret low-dimensional representation of the exhibited behaviors of the unannotated experiment. When the behavioral repertoire and/or feature distribution are dissimilar, the resulting embedding may not provide useful information about the unannotated experiment, but an advantage of this approach is that

the other pair embeddings do not get distorted by poor matches. As described in Section 5.3, semi-supervised pair embeddings are utilized to predict behavioral categories of unannotated experiments by combining multiple view acquired from annotated ones.



(a) Only unannotated experiment.



(b) Unannotated and annotated experiments.

Figure 5.3 Semi-supervised pair embeddings with an annotated and an unannotated experiment. FlyM4-03062020 (annotated) provides a "view" on the behavioral repertoire of the FlyF15-08182021 (unannotated).

5.3 Nearest Neighbor Analysis and Classification

One of our ultimate goals is to annotate an experiment using already annotated ones. In previous sections and chapters, we described feature extraction, activity detection, and computation of behavioral embeddings. In this section, we describe a novel nearest neighbor based classification approach, which has to tackle the following challenges:

- sparsity of behavioral expressions,
- imbalanced distribution of behavioral categories,
- scarcity of annotated experiments, which are laborious to produce,
- variation between experiments.

Our approach consists of two steps, briefly stated as follows:

5.1 computing behavioral weights of an unannotated experiment, using its semi-supervised pair embeddings for each annotated experiment,

5.2 combining behavioral weights of the unannotated experiment with an annotated experiment committee by voting.

We compute the behavioral weights with our nearest neighbor based approach by considering the sparsity of behavioral expressions and imbalanced distribution of behavioral categories, as described in Section 5.3.1. In Section 5.3.2, we detail the committee-by-voting approach, which helps us to deal with variation between experiments by providing multiple "views" on observed behavioral expressions that we attempt to annotate. Finally, the post-processing of resulting predictions is described in Section 5.3.3.

5.3.1 Behavioral Weights

Consider two experiments, an annotated one expt$^+$, and an unannotated one expt$^-$, and their semi-supervised pair embedding, respectively $\mathbf{E}^+ = \left[\mathbf{e}^+_1, \cdots \mathbf{e}^+_{F^+}\right]$ and $\mathbf{E}^- = \left[\mathbf{e}^-_1, \cdots \mathbf{e}^-_{F^-}\right]$. Given the true annotations $\mathbf{y}^+$ of the frames in the Micro-activity$^+$ set of expt$^+$, and K behavioral categories, the goal is to compute $\widehat{\mathbf{b}}_f = \left[\hat{b}_{f,1}, \cdots \hat{b}_{f,K}\right]$, representing the weights (in other words, the similarity score) of each behavioral category for expt$^-$, using $\mathbf{E}^+, \mathbf{E}^-$ and $\mathbf{y}^+$.

The procedure start by querying k-nearest neighbors of expt$^-$'s each frame f in the joint behavioral embedding space consisting of $\mathbf{E}^+$ and $\mathbf{E}^-$, using the k-d trees for efficiency (Bentley, 1975). k-NN(f) denotes the set of indices of $\mathbf{e}^-_f$'s k nearest neighbors, and the k-NN weight $b_{f,i}$ for each query point (i.e., frame) f of expt$^-$, and behavioral category i, is computed by

$$b_{f,i} = \begin{cases} \displaystyle\sum_{f' \in k\text{-NN}_i(f)} \frac{1}{\mathrm{d}\left(\mathbf{e}^-_f, \mathbf{e}^+_{f'}\right)^p + \epsilon} & \text{if } |k\text{-NN}_i(f)| \neq 0, \\ 0 & \text{if otherwise,} \end{cases} \quad \text{where} \quad p \in \{0,1,2\}. \tag{5.4}$$

Here, $k\text{-NN}_i(f) = \left\{ f' \,:\, y^+_{f'} = i, \text{ and } f' \in k\text{-NN}(f) \right\}$, is the set of indices of data points of expt$^+$ whose annotation is the behavior category i and is one of the k nearest neighbors of $\mathbf{e}^-_f$. $\mathrm{d}\left(\mathbf{e}^-_f, \mathbf{e}^+_{f'}\right)$ is the euclidean distance between $\mathbf{e}^-_f$ and $\mathbf{e}^+_{f'}$, and p parameterizes the relation between distance and weight $b_{f,i}$. We add a small number ϵ (10^{-6}) to the denominator to avoid numerical errors. The resulting vector $\mathbf{b}_f = \left[b_{f,1}, \cdots, b_{f,K}\right]$ weights the similarity of the frame f to each behavioral category in the shared embedding space based on nearest neighbors.

Naturally, the number of occurrences or durations of the behavior bouts are different for each behavioral category, and therefore, $\mathbf{y}^+$ is highly imbalanced. As a result, the number of nearest neighbors and $\mathbf{b}_f$ are biased in favor of frequently occurring and long-bout behaviors. For instance, pumping-like movements of the proboscis occur more frequently in longer bouts than switch-like movements of the haltere. Especially when k is large, it becomes crucial to consider the imbalanced distribution of behavior occurrences, since the embedding space will be dominated by frequent behaviors. Thus, incorporating this imbalance into the formulation may help to improve the recall of rarely occurring short-bout behaviors and precision of frequently occurring long-bout behaviors. To achieve this, we normalize the scores previously computed, $b_{f,i}$, as a function of the number of occurrences of the behavioral category i as follows:

$$(5.5) \qquad ,b'_{f,i} = \frac{b_{f,i}}{(1+N_i^+)^p} \quad \text{or} \quad \frac{b_{f,i}}{\log_k(1+N_i^+)} \quad \text{where} \quad p \in \{0, 1/2, 1\}, k \in \{2, 10\},$$

where $N_i^+ = \left|\left\{ f : y_f^+ = i \right\}\right|$ is the number of frames annotated as behavioral category i. In the above equation, two different alternatives are given for this normalization step; a polynomial one and a logarithmic one, where p and k parameterize the relation between N_i^+ and $b'_{f,i}$. For instance, if one is mostly interested in achieving high recall for frequently occurring behaviors, low p values or using the logarithmic alternative might be more appropriate. It may be even desired to set $p = 0$, and not considering the number of occurrences in some cases, see Section 6.1 for more details.

The resulting vector $\mathbf{b}'_f \in \mathbb{R}^K$ is dependent on the annotated experiment expt$^+$, and the vectors computed based on different annotated experiments are not comparable with each other. Hence, we map the values of $b'_{f,i}$ to $[0,1]$ using either the softmax function or L_1 normalization as follows:

$$(5.6) \qquad \hat{b}_{f,i} = \frac{\exp\left\{b'_{f,i}\right\}}{\sum_{j=1}^K \exp\left\{b'_{f,j}\right\}} \quad \text{or} \quad \frac{b'_{f,i}}{\sum_{j=1}^K b'_{f,j}}.$$

The resulting behavioral weight vector $\hat{\mathbf{b}}_f \in [0,1]^K$ can be considered as a probability distribution. Here, the vector $\hat{\mathbf{b}}_f$ represents the behavioral characteristics of the frame f of expt$^-$ based on the behavioral repertoire of expt$^+$. The voting-like scheme, as described in Section 5.3.2, incorporates the behavioral weight vectors of all annotated experiments to finalize the classification for expt$^-$.

5.3.2 Experiment Committee by Voting

Consider all experiments: unannotated experiments $\text{expt}_1^-, \cdots, \text{expt}_{R^-}^-$, and annotated experiments $\text{expt}_1^+, \cdots, \text{expt}_{R^+}^+$, where R^- and R^+ are the number of experiments, respectively. The goal is to combine the behavioral weights of an unannotated experiment expt_k^-, calculated separately for each annotated experiment.

Let $\hat{\mathbf{b}}_f^{k,l}$ denote the behavioral weights of expt_k^- computed with expt_l^+. Each annotated experiment contributes to the overall behavioral score of expt_k^-; in other words, annotated experiments, forming a committee, vote according to $\hat{\mathbf{b}}_f^{k,l}$. Each annotated experiment provides a different view of the behavioral repertoire, as in the case of pair embeddings (see Section 5.2.3). Before describing hard voting and soft voting approaches, there is one more step to discuss.

There exists a significant variation among the exhibited behavioral repertoires in the experiments. An annotated experiment might lack some behavioral expressions or manifest some behaviors excessively. In such cases, if the behavioral weight vector $\hat{\mathbf{b}}_f^{k,l}$ is not confident, i.e., weights of behavioral categories are close to each other, one may want to decrease its contribution to the voting. To achieve this, we propose two optional approaches; penalizing the behavioral weights with the entropy of the "probability distribution" $\hat{\mathbf{b}}_f^{k,l}$, or with the uncertainty. The contribution of votes of expt_l^+ to the committee formed for expt_k^- is $\mathbf{vote}_f^{k,l} = \left[\mathsf{vote}_{f,1}^{k,l}, \cdots, \mathsf{vote}_{f,K}^{k,l}\right]$, and is given by

$$\mathsf{vote}_{f,i}^{k,l} = \left(\log_2(K) - \mathrm{H}\left(\hat{\mathbf{b}}_f^{k,l}\right)\right)\hat{b}_{f,i}^{k,l} \quad \text{or} \quad \left(1 - \max_{1\le j\le K} \hat{b}_{f,j}^{k,l}\right)\hat{b}_{f,i}^{k,l} \quad \text{or} \quad \hat{b}_{f,i}^{k,l}. \tag{5.7}$$

If $\max_i \hat{b}_{f,i}$ is close to $1/K$, which means that the computed vector weights the behaviors uniformly, then the factors $\left(\log_2(K) - \mathrm{H}\left(\hat{\mathbf{b}}_f^{k,l}\right)\right)$ or $\left(1 - \max_{1\le j\le K} \hat{b}_{f,j}^{k,l}\right)$ may be used to decrease the "importance" of the vote.

Now, after computing behavioral votes for each annotated and unannotated experiment pair, we can combine those votes for a single unannotated experiment expt_k^- by forming a committee of annotated experiments $\text{expt}_1^+, \cdots, \text{expt}_{R^+}^+$. The predicted behavioral category at frame k of expt_k^- is given by the combined votes of the committee. The predicted behavioral category at frame k of expt_k^-, denoted by $\hat{y}_f^k$, is given by combined votes of the committee. To obtain the overall aggregate view of the committee, we can follow two alternative voting schemes, namely hard voting (i.e., majority rule voting) or soft voting. At this point, we may want to have scores for each behavioral category rather than hard labels. Such score information might be desired to capture the complex and

hierarchical structure of the behaviors. Behaviors sometimes occur simultaneously, and sometimes observed behaviors might manifest similarities to more than one behavioral category (Berman et al., 2016). Hence, one may also use the combined vote scores directly, without computing the arg max .

5.3.2.1 Hard Voting

In the hard voting scheme, the prediction of a frame f is given by the majority behavioral category of the arg max of each annotated experiment's votes and computed by the following formula:

$$\widehat{y}_f^k = \underset{1\le i\le K}{\arg\max} \left\{ \underset{1\le j\le K}{\arg\max}\, \mathsf{vote}_{f,j}^{k,l} : \; j=i, 1\le l\le R^+ \right\}. \tag{5.8}$$

5.3.2.2 Soft Voting

In contrast to majority voting (hard voting), the soft voting scheme assigns the behavioral category as the arg max of the sum of each annotated experiment's vote vector, and $\widehat{y}_f^k$ is given by

$$\widehat{y}_f^k = \underset{1\le i\le K}{\arg\max} \sum_{l=1}^{R^+} \mathsf{vote}_{f,i}^{k,l}. \tag{5.9}$$

5.3.2.3 Behavioral Scores

Definitions of behavioral categories might be broad, and expressed behaviors sometimes are a combination of multiple behavioral categories. For example, a switch-like movement of the haltere can simultaneously occur with a postural adjustment. One may also want to examine and interpret top-k predictions, especially when there exist many narrow behavioral categories. Thus, in addition to assigning categories to frames, it is also useful

to compute and report scores for behavioral categories based on the votes contributed by each annotated experiment. Moreover, such scores can be utilized as confidence scores and might be helpful to analyze differences in the behavioral repertoire (see Figure 6.2 and Figure 6.5).

For a behavioral category i, such a score can be calculated by combining votes as follows:

$$\frac{\sum_{l=1}^{R^+} \mathsf{vote}_{f,i}^{k,l}}{\sum_{i=1}^{K}\sum_{l=1}^{R^+} \mathsf{vote}_{f,i}^{k,l}}. \tag{5.10}$$

5.3.3 Post-processing

Finally, we post-process predicted annotations to improve our nearest neighbor based algorithm by incorporating a couple of assumptions on the temporal organization of the behaviors and physical constraints. Before applying post-processing procedures, the frames in the sets Macro-activity and Quiescence should be recovered, as we only predicted the annotations of the frames in the set Micro-activity. Without having a temporally continuous vector of annotations, post-processing would lead to unintended consequences and erroneous results. Let $\textbf{pred-a}^k = (\hat{y}_1^k, \cdots, \hat{y}_T^k)$ be the expt_k's vector of predicted annotations for the entire experiment, defined as follows:

$$\mathsf{pred\text{-}a}_t^k = \begin{cases} 0 & \text{if } t \in \mathsf{Quiescence}_k, \\ \hat{y}_f^k & \text{if } t \in \mathsf{Micro\text{-}activity}_k \text{ and } t = t_f, \\ K+1 & \text{if } t \in \mathsf{Macro\text{-}activity}_k. \end{cases} \tag{5.11}$$

This vector of annotations spans an entire experiment of k and consists of detailed annotations of behavioral subcategories during dormancy, the general category of macro-activities, and quiescence.

Before discussing post-processing, we first need to define behavioral bouts. Informally, a behavioral bout is a segment of time, in which the same behavioral category is continu-

ously observed. We can formally define a behavioral bout for a given $\widehat{y}_t^k$ as follows:

$$\begin{aligned} \mathsf{bout}_t^0 &= \max\left\{\arg\max_{t'} \left(\widehat{y}_t^k \neq \widehat{y}_{t'}^k\right) \wedge (1 \leq t' \leq t) \vee (t' = 1)\right\}, \\ \mathsf{bout}_t^1 &= \min\left\{\arg\max_{t'} \left(\widehat{y}_t^k \neq \widehat{y}_{t'}^k\right) \wedge (t \leq t' \leq T) \vee (t' = T)\right\}, \end{aligned} \tag{5.12}$$

where bout_t^0 and bout_t^1 are respectively the beginning and the end of the behavioral bout, to which $\widehat{y}_t^k$ belongs.

We have sensible expectations about the duration of behavioral bouts of each behavioral category acquired by human annotators. For instance, a bout of proboscis's pumping-like movement should not be shorter than 200 millisecond, as bout duration distributions of annotations indicate. In addition to annotations, we may have reasons for physical and biological constraints. An example is that grooming behavior, bout duration of grooming can not exceed a couple of minutes, probability of observing a grooming behavior lasted longer than 2 minutes is extremely low (Qiao et al., 2018). Considering such constraints and temporal expectations, we apply the following post-processing procedures, and parameters should be determined based on the behavioral repertoire of interest. Post-processing procedures are especially helpful for avoiding misleading classification of the time points with erroneous tracking of body parts. Each procedure is optional but should be applied in the given order.

5.3.3.1 Pruning Interrupting Bouts

We prune short intervals interrupting possibly long and continuous behavioral bouts. Given a window size τ, the majority behavioral category in the surrounding window around a time point t is assigned that point by:

$$\widehat{\mathsf{pred\text{-}a}}_t^k = \arg\max_{\mathsf{a}} \left|\left\{ t' \,:\, \mathsf{pred\text{-}a}_{t'}^k = \mathsf{a},\ \max\{1, t-\tau\} \leq t' \leq \min\{t+\tau, T\} \right\}\right|. \tag{5.13}$$

The window size τ is typically less than a second.

5.3.3.2 Elimination of Overly Long and Overly Short Bouts

We eliminate behavioral bouts whose durations are extremely short or extremely long, and hence, contradict physical constraints and our temporal expectations. For each behavioral category i, we define two thresholds δ_i^{short} and δ_i^{long}, namely upper bound and lower bound. Then, the behavioral bouts whose duration is longer or shorter than the corresponding thresholds, are set to quiescence.

$$\widehat{\mathsf{pred\text{-}a}}_t^k = \begin{cases} 0 & \mathsf{bout}_t^0 - \mathsf{bout}_t^1 < \delta^{\text{short}}_{\mathsf{pred\text{-}a}_t^k}, \\ 0 & \mathsf{bout}_t^0 - \mathsf{bout}_t^1 > \delta^{\text{long}}_{\mathsf{pred\text{-}a}_t^k}, \\ \mathsf{pred\text{-}a}_t^k & \text{otherwise}, \end{cases} \tag{5.14}$$

For each behavioral category i, the thresholds δ_i^{short} (δ_i^{long}) should be greater (smaller) than the window size τ used in Section 5.3.3.1.

6. RESULTS

6.1 Employing the Pipeline and Evaluation

A total of 11 experiments (7 wild-type sleep and 4 sleep-deprived) and their annotations are split into 10 training experiments and 1 test experiment for all combinations. We report results and evaluate each split and demonstrate varying performance for different splits.

Our pipeline starts with the feature extraction stage, where raw tracking data generated by DeepLabCut is used to generate meaningful behavioral representations. Body parts used to extract features are proboscis, haltere, thorax, head (left and right), and three midlegs (left and right). We use oriented pose values for body parts with left and right counterparts, as described in Section 3.2.1.1. In order to detect occluded body parts, we set the low confidence score threshold to 0.075. Moving median window size τ is 15 frames, (0.5 seconds) and threshold τ set to 15 microns. If the position of a body part exceeds the median of the window centered around it by τ, it is marked as occluded, as described in Section 3.2.1.2.

Linear interpolation is used for the imputation of marked frames. After that, firstly, a median filter of size 6 frames, and then a boxcar filter of size 6 frames is applied to reduce the tracking noise and smooth the signal, without smoothing behaviors exhibited by rapid movements. Aligning different orientations and transforming the frames to be egocentric did not result in performance improvement in our case, hence, we did not

transform frames as egocentric.

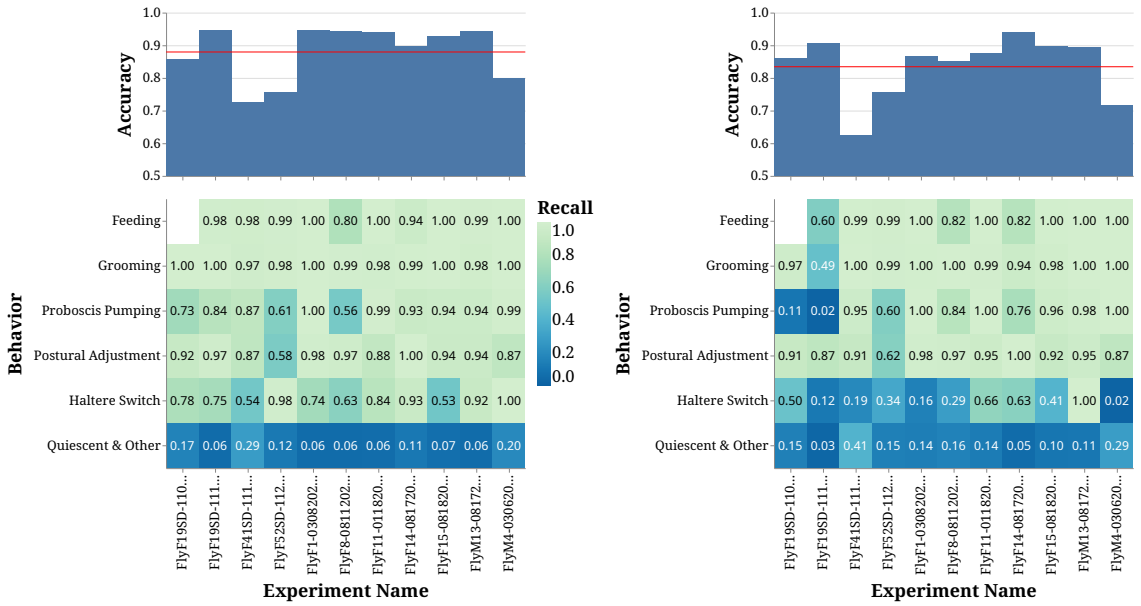
Then, we computed snapshot features and gradient features as described in Section 3.3, and given in Table 6.1. In order to generate postural dynamics, we applied wavelet transformation to snapshot features at 20 different frequency channels dyadically spaced between 1 Hz and 20 Hz (see Equation 3.21 for determining spectrum frequencies). Different timescales are normalized as given in Equation 3.20. After flattening and L_1 frame normalization, we ended up with a $13 \times 20 = 260$ dimensional behavioral representation matrix $\hat{\mathbf{W}}$. Similarly, for gradient features, moving mean values are computed for a single timescale which is 33 milliseconds, resulting in $9 \times 1 = 9$ dimensional behavioral representation matrix $\hat{\mathbf{M}}^{\mu}$, which is only used for activity detection and constructing Dormancy set.

	Snapshot Features	**Gradient Features**
Distance between	haltere and origin proboscis and origin thorax and origin head and proboscis haltere and thorax (average) midlegs and origin (average) midlegs and thorax	head and proboscis thorax and proboscis thorax and origin
Cartesian components of	haltere head thorax	head proboscis thorax

Table 6.1 Computed spatio-temporal features.

After constructing behavioral representation matrices, the activity detection stage of our pipeline takes place. Here reported recall scores are for the frame set Micro-activity, and it is desired to achieve high recall scores, except for the "quiescent and other" category. We evaluate both unsupervised and supervised approaches, recall, and accuracy values are shown in Figure 6.1.

For unsupervised activity detection, the threshold c is set to the decision boundary λ_1 of two Gaussian components for the construction of the Dormancy set. Similarly, thresholds c_i for each $\mathbf{u}_i$ value are the first decision boundaries of 3 Gaussian components, sorted by their means. The resulting detection performance is quite poor for the switch-like behavior of haltere since this behavior occurs very subtly, and it is similar to quiescent frames. 8 out of 11 splits have a recall score of less than 0.5. Also, recall scores for two of the sleep-deprived experiment splits (FlyF19SDs) are below 0.15, which makes it



(a) Supervised detection. **(b)** Unsupervised detection.

Figure 6.1 Performance summary of activity detection and micro-activity detection. The red line indicates the macro average of accuracy scores achieved for each split. High recall scores are desired for behavioral categories, as opposed to quiescent frames. Supervised and unsupervised detections are given respectively in the left subfigure and the right subfigure.

impossible to proceed with a successful mapping.

In the supervised approach, a random forest of decision trees (Breiman, 2001) is utilized with 10 estimators, where the maximum depth of each tree is 5 and the criterion is Gini impurity. For grooming, recall is always greater than 0.97, which implies that we do not lose an insignificant amount of annotated frames. Similarly, the recall score of feeding drops below 0.95 only once. For proboscis pumping, 6 out of 11 splits have a recall score greater than 0.9, and performance is relatively poor for two splits (0.61 and 0.56 recall). Again, haltere switch is the most challenging behavioral category, but recall is often greater than 0.75 as opposed to the unsupervised approach. Considering its superior performance over the unsupervised approach, we proceed to behavior mapping by employing supervised detection.

After activity detection, the next stage is behavior mapping, where the frames in the set Micro-activity are mapped to behavioral categories. The first step is the computation of behavioral embeddings. The semi-supervised pair embeddings described in the Section 5.2.3 computed for each annotated and unannotated experiment pair. We set the embedding space dimension to 2, which 260 dimensional behavioral representation matrix $\hat{\mathbf{W}}$ is reduced to. The number of neighbors parameter of UMAP is set to 75,

the minimum distance parameter is determined as 0, and the distance metric is Hellinger distance (Equation 5.1).

Then, with the generated behavioral embeddings, nearest neighbors analysis is performed to assign behavioral scores to unannotated frames. In each semi-supervised embedding, we computed 25 nearest annotated neighbors of the frames, and contributions of the neighbors are weighted disproportional to its distance, using the Euclidean distance, as formally given in the Equation 5.4. After that, weights are normalized with the $\log_2$ number of occurrences of behavioral category of contributing neighbors, and then L_1 normalized to make behavioral weights sum up to 1 (see Equation 5.5, 5.6). Finally, we summed each behavioral weight, obtained from a different "view" (i.e., annotated experiment), and L_1 normalized again to end up with the final behavioral score values.

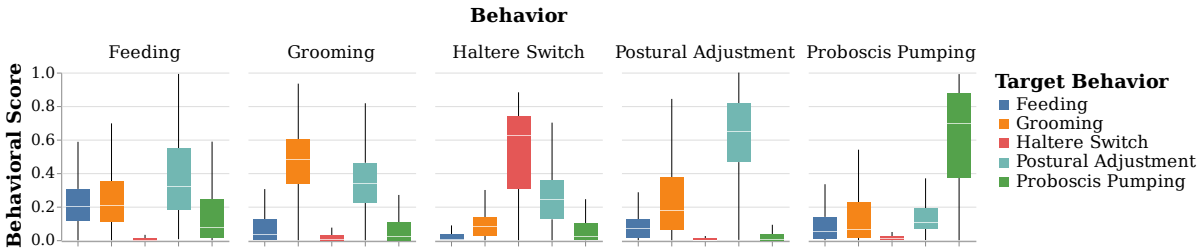
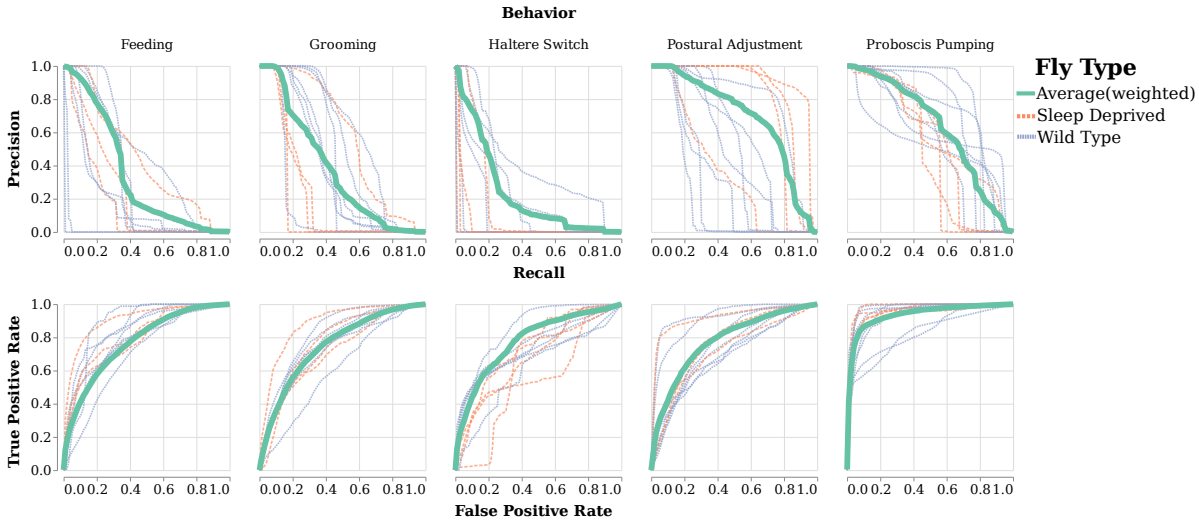
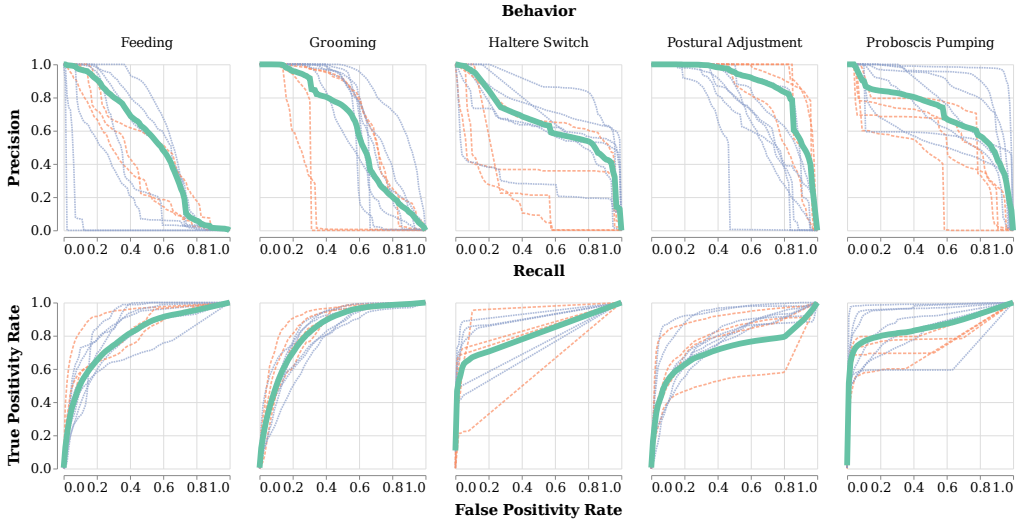


Figure 6.2 Distributions of behavioral score values of each behavioral category for all splits. Each box-plot column demonstrates the behavioral score distributions of target behaviors for the corresponding annotation.

The resulting score vectors are used for classification and also utilized as confidence scores. In Figure 6.2, behavioral score distribution for each category is given for all splits together. Except for the feeding category, the median score of the correct target behavioral category is greater than 0.5. Grooming and postural adjustment are the second most similar target behavior for each other, as expected. When the grooming behavior is short, it becomes more similar to short-duration moving and postural adjustment behavior. Behavioral scores are also helpful for revealing similarities and dissimilarities between behavioral categories and can be helpful for determining spatio-temporal feature sets. Although we normalized behavioral weights with the number of occurrences, which favors the relatively short duration and rare switch-like haltere behavior, we can see that its target behavior score does not exceed 0.1 for wrong categories. Compared to other behaviors, feeding has the poorest performance, with a median correct score of 0.2. The postural adjustment category wrongly has the highest median score for feeding. Frames with feeding behavior annotation are often misidentified as postural adjustment, grooming, and proboscis pumping. As proboscis movement is common both for feeding and proboscis pumping, confusion with the proboscis pumping is expected for feeding.



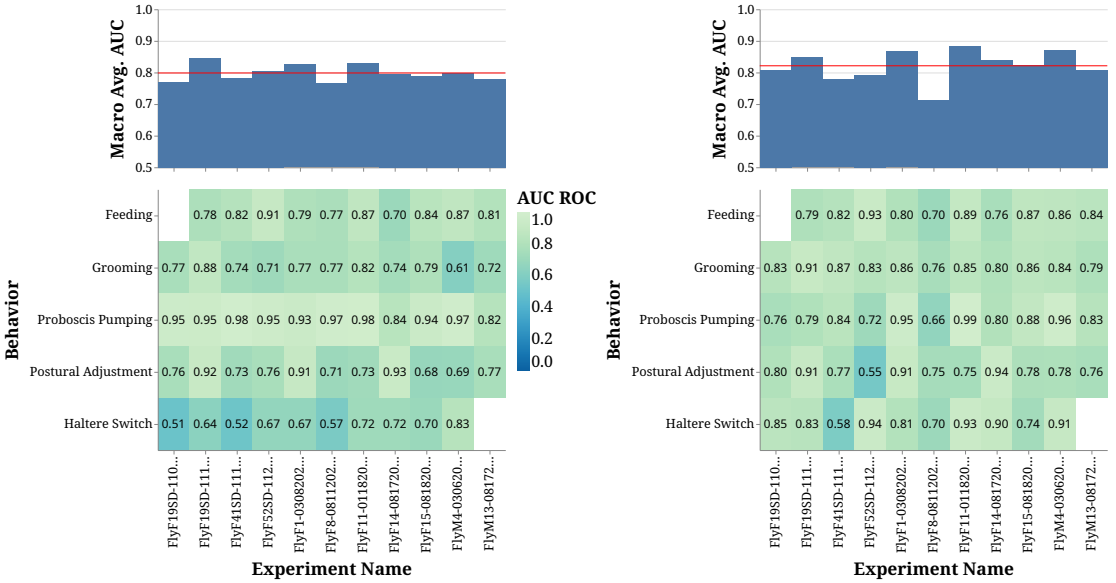
(a) ROC and precision-recall curves for frames detected as micro-activity.



(b) ROC and precision-recall curves for annotated frames.

Figure 6.3 Performance summary of behavior mapping demonstrated using receiver operating characteristic curve and precision-recall curve. The weighted averages of ROC and precision-recall curves are computed by interpolation. Curves for both frames estimated as micro-activity and annotated frames are given respectively in Figures 6.3a, 6.3b.

Using behavioral computed behavioral scores, we computed precision-recall and receiver operating characteristic curves for each split, and also computed interpreted weighted averages of curves, see Figure 6.3 and Figure 6.4. Evaluations are done by considering two subsets of frames, the first subset is the frames annotated as one of the behavioral categories and the other one is the Micro-activity set. The latter one contains false positive micro-activity frames, which is quantified as the recall score of quiescent and other category. False positive micro-activity frames affect haltere switch detection performance



(a) AUC scores for Micro-activity set.

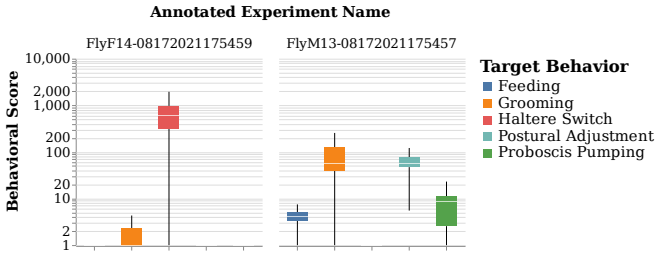
(b) AUC scores for annotated frames.

Figure 6.4 Performance summary of behavior mapping with the area under curve scores of ROC. The red line indicates the macro average of AUC scores for each split. Scores for both frames estimated as micro-activity and annotated frames are given respectively in Figures 6.4a, 6.4b.

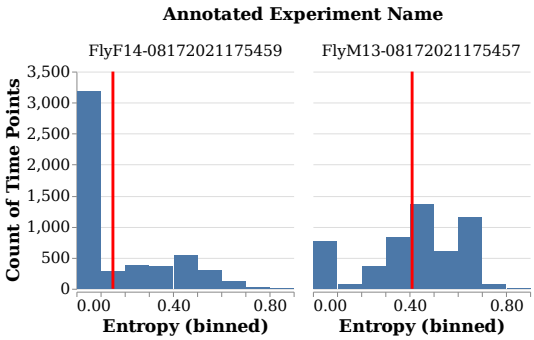
most, for instance, its weighted mean precision decreases by ∼0.5 for fixed recall at ∼0.65. This performance difference also shows the importance of achieving low recall for quiescent and other category in the activity detection stage. Since the evaluation is done with the Micro-activity set is more realistic, the following comments are made on considering it.

The most robust detection is achieved for proboscis pumping and postural adjustment behavioral categories, respectively, achieving F-1 greater than 0.8 and 0.85 scores for some splits. The maximum F-1 score achieved for grooming is 0.60 for FlyF1-03082020164520 split, whereas F-1 scores of haltere switch behavioral category do not exceed 0.4. AUC scores for the proboscis pumping often exceed 0.95, implying robust detection for almost all the splits. We observe that detection performance varies greatly among different experiments. This variation can be due to several reasons, but most likely reasons are related to tracking performance and annotation quality, as well as behavioral repertoires of flies. Considering all splits together, we achieve a macro average of 0.8 AUC score with a one-versus-rest approach.

As discussed in Chapter 1, the behavioral repertoire of the fruit fly is much richer than the behavioral categories that we considered. Moreover, there exists a great amount of variation among flies' behavioral repertoire and behavioral visits (see Section 6.2). Thus,



(a) Behavioral score values, before L_1 normalization.



(b) Entropy values. The red line is the mean.

Figure 6.5 Histogram of entropy values and box-plot of behavioral scores computed using one unannotated and two annotated experiments with varying behavioral repertoires. Here, the behavioral repertoire of FlyF1-03082020 is predicted separately with two different annotated experiments, namely FlyF14-08172021 and FlyM13-08172021. Behavioral scores and entropy values are computed for the haltere switch behavior. The latter one lacks the haltere switch behavior, and as a result, behavioral scores tend to have higher entropy. Results demonstrate the ability of the proposed pipeline to detect and discover unseen unannotated behavioral categories using behavioral scores.

we also investigate our pipeline's ability to detect unseen behavioral categories by utilizing behavioral scores. Such an ability is important for avoiding misleading the user of the pipeline. It may also be desired to reconsider defined and annotated behavioral categories, and may further lead to interesting biological observations. To this end, we consider the following experimental setting: behavior mapping of an unannotated experiment (FlyF1-03082020) is done separately for two annotated experiments, namely FlyF14-0817202 and FlyM13-08172021. Then, behavioral scores of the FlyF1-03082020's frames with haltere switch annotation are compared between FlyF14-0817202 guided mapping and FlyM13-08172021 guided mapping. As it can be seen from Figure 6.8c, FlyF1-03082020 and FlyF14-0817202 spent $\sim$4 minutes by exhibiting switch-like haltere movement behavior, whereas FlyM13-08172021 spent no more than 5 seconds. Therefore, compared to FlyF14-0817202, the haltere switch is an unseen and lacking behavioral category for FlyM13-08172021. We expect to identify this using behavioral scores. Indeed, behavioral scores reflect the difference in the behavioral repertoires of the annotated flies. As it can be seen in Figure 6.5a, behavioral scores obtained based on FlyF14-0817202's behavioral repertoire are much more confident about the correct behavioral category, whereas behavioral scores obtained based on FlyM13-08172021's behavioral repertoire are more uniform-like and total score is often shared among multiple behavioral categories. Using the entropy of the behavioral score, one can quantify the existence of unseen and new behavioral categories more systematically. In Figure 6.5b, we compare entropy distribu-

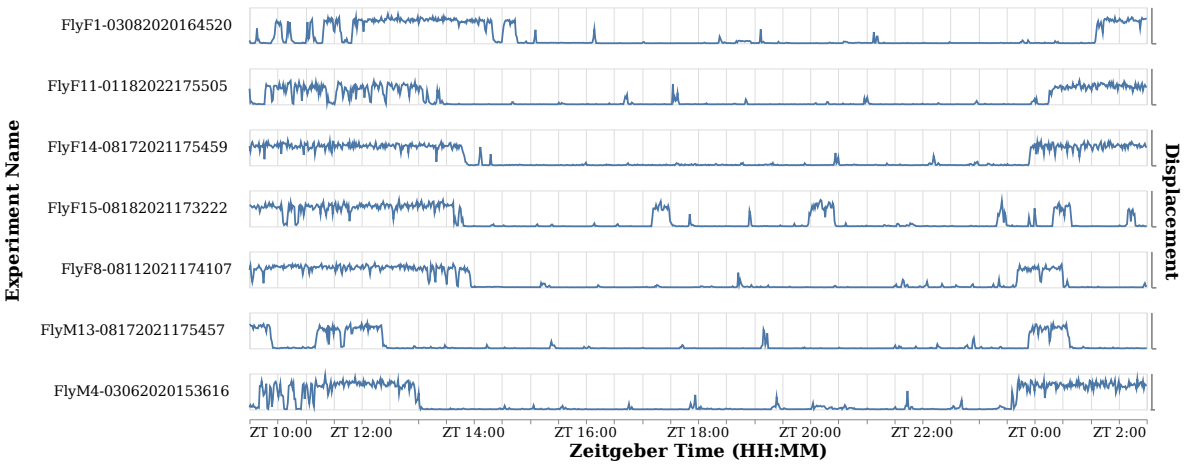
tions of the behavioral scores obtained using annotated experiments FlyM13-08172021 and FlyF14-0817202. We can immediately see that scores obtained based on the FlyM13-08172021's repertoire have higher entropy, with a mean of 0.41, whereas for the other one mean is 0.15. For instance, if the entropy of a frame's score is greater than 0.2, it may be recommended to visit them and investigate the reason. Accordingly, behavioral annotations can be extended.

6.2 Analyzing Behavioral Repertoires

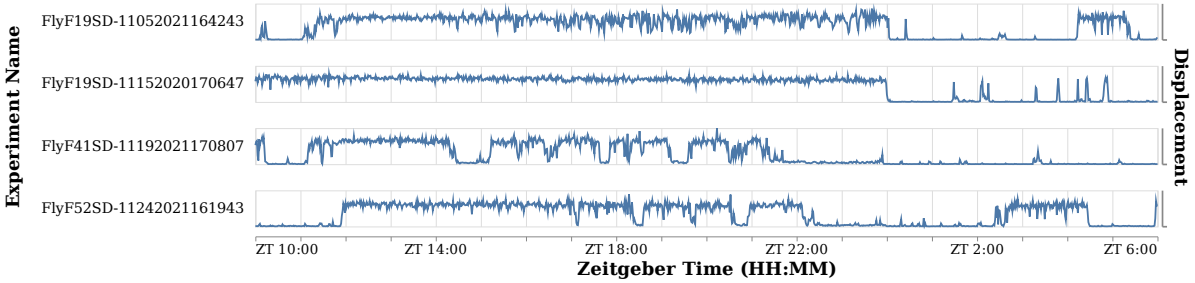
In this section, we analyze the collected data of sleep experiments to discover how behaviors that we are interested in are exhibited. Starting from a broad perspective, the briefest description of sleep experiments can be done by quantifying the displacement of the flies placed in the chamber. In Figure 6.6, we plot velocity-based feature vectors $\mathbf{u}$ between ZT10-ZT02 and ZT10-ZT06, respectively for each wild-type sleep and sleep-deprived experiment. We observe that the long dormancy epochs are interrupted by relatively short macro activity bouts during sleep. In addition to short macro activity bouts, we observe very small displacements all over the night, which indicates there exist micro-activities other than major positional and postural changes. Hence, a more fine-grained analysis and detailed examination are necessary to gain insight, as this work attempts to achieve.

Using annotated behavioral categories, we further examine more closely how activities manifested during sleep. In Figure 6.7, experiments are temporally divided into 5-minute bins, and activity is quantified as the ratio between the number of annotated frames to the total number of frames (900 frames). Such a quantification provides us a view of sleep which demonstrates that many activities occur without the displacement of the fly's body.

In addition to the temporal organization of the activities, we are also interested in the characteristics of each behavioral category such as total time spent, bout durations, kinematics, and feature value distributions. In order to visualize feature distributions of each behavioral category, we sum behavioral representation values of all frequency channels for each spatio-temporal feature. For a single frame, resulting values sum up to 1 as they are L_1 normalized. In Figure 6.8a, we see how behavioral categories are defined by spatial characteristics. As expected, proboscis-related features play a significant role in characterizing feeding and proboscis pumping. Similarly, features related to leg and



(a) Wild type.



(b) Sleep-deprived.

Figure 6.6 Overall displacement values over entire experiments. Displacement of the body reveals long dormancy and sleep epochs, and macro-activities. Displacement values are computed as described in Equation 4.1 and smoothed with a rolling mean of a 1-minute window.

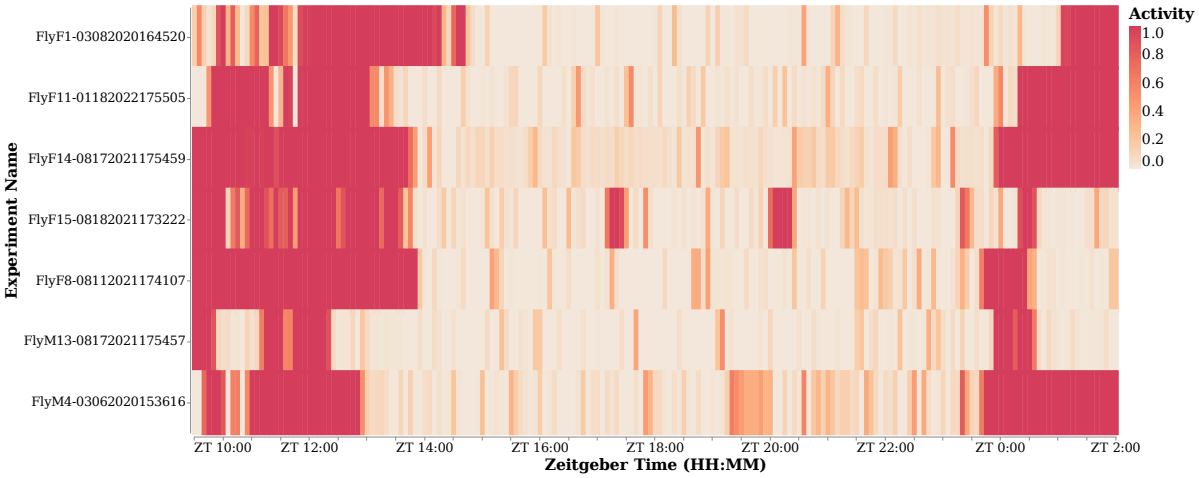
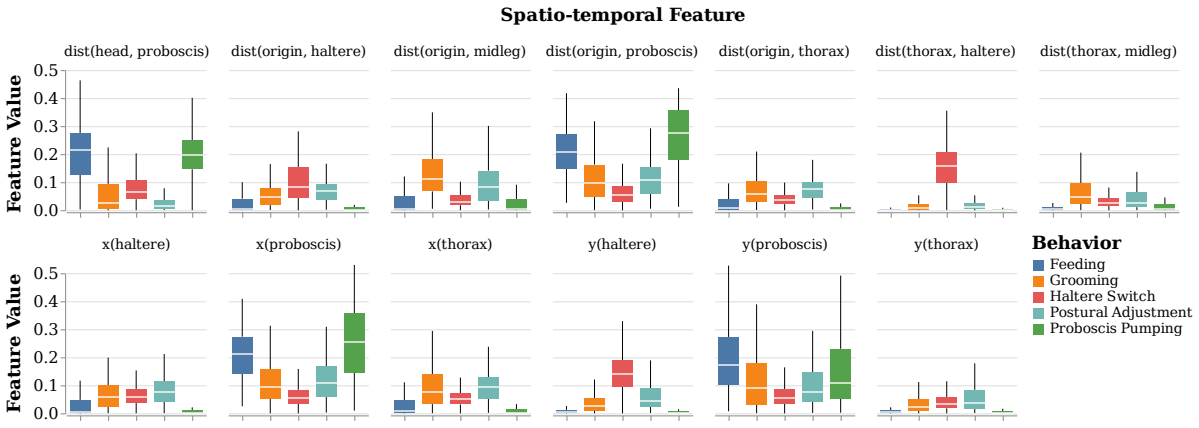
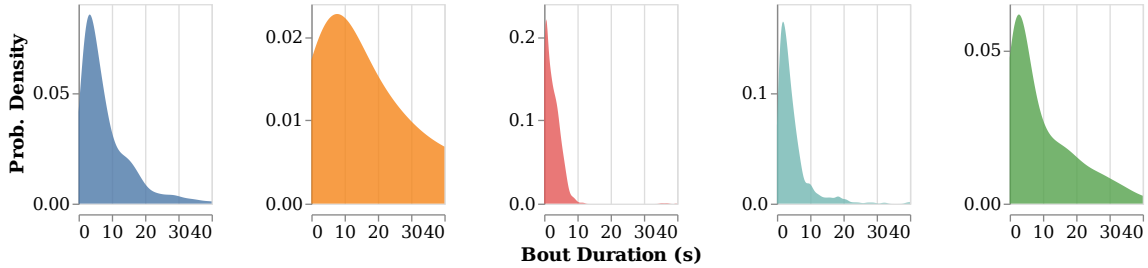


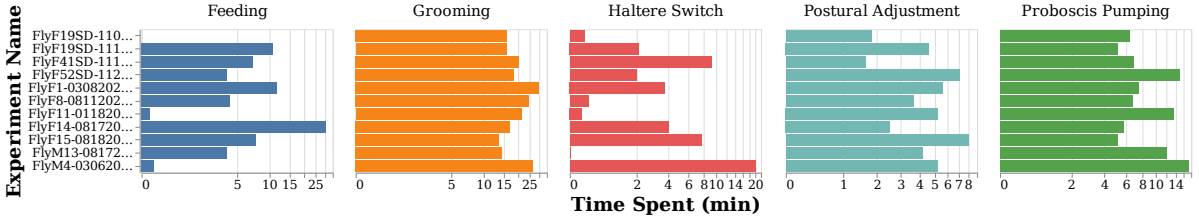
Figure 6.7 Binned temporal heatmap of activities. Each bin is 5 minutes, corresponding to 9000 frames. Activity value is computed as the ratio of the number of annotated frames and the total number of frames in that bin.



(a) Summation of all frequency channels of behavioral representation value for each spatio-temporal feature.



(b) Kernel density estimations of bout durations for each behavioral category. The variance of bout durations within and across behavioral categories demonstrates the rich behavioral repertoire.



(c) Time spent while exhibiting each behavioral category for all experiments.

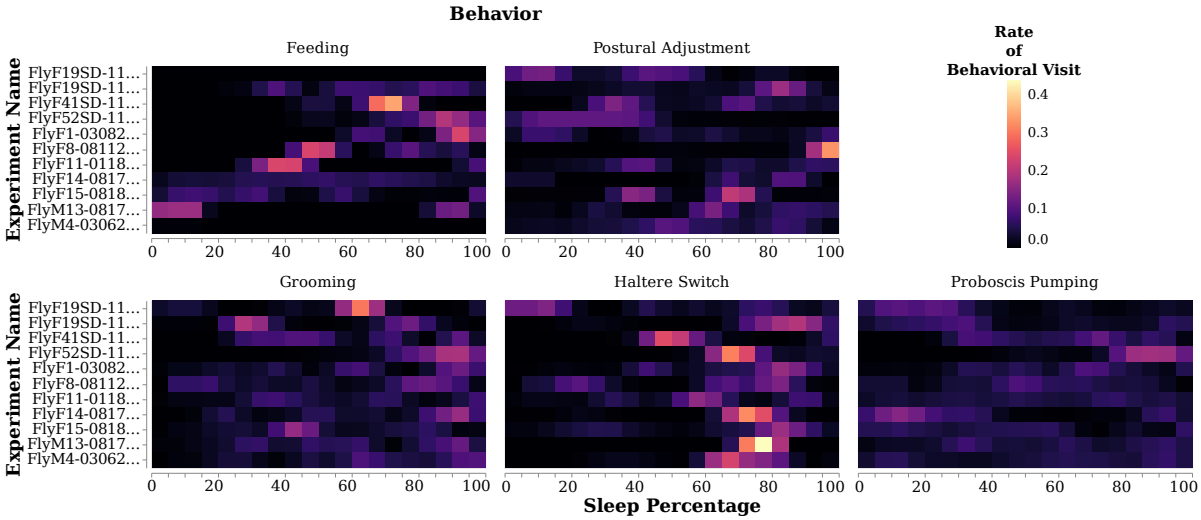
Figure 6.8 Summary of behavioral repertoires, demonstrated using both spatial and temporal characteristics.

thorax are more apparent for grooming and postural adjustment.

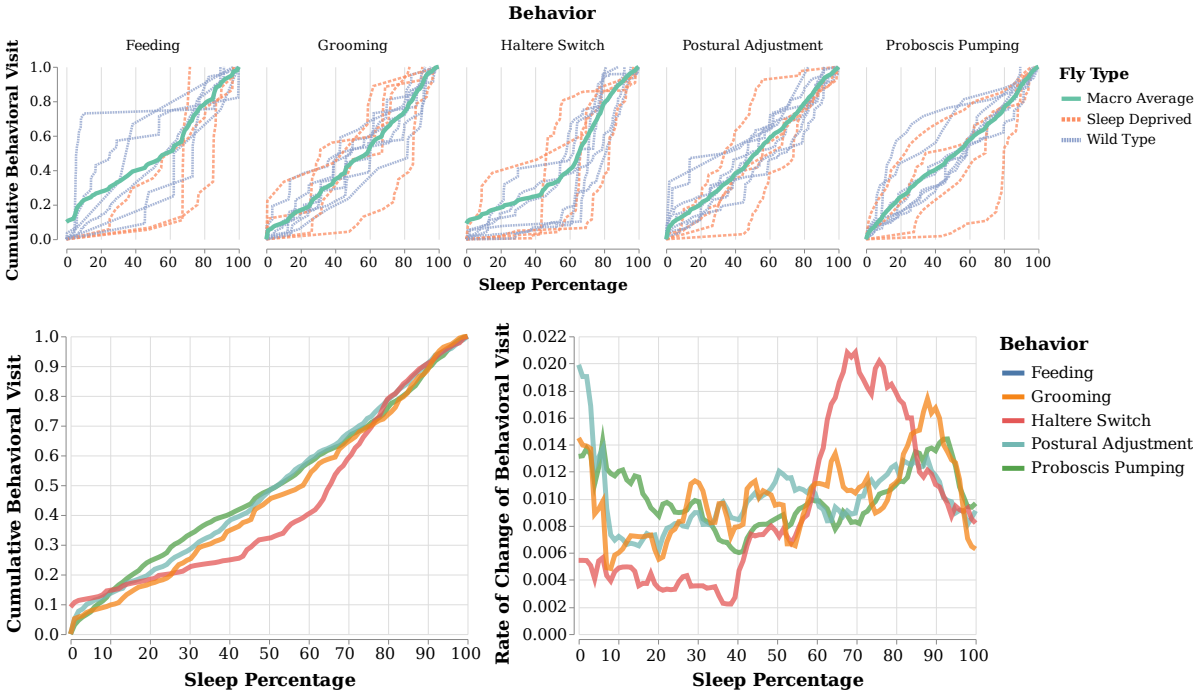
Behaviors differ in terms of bout durations as well. Kernel density estimations of bout duration distributions for each behavioral category are given in Figure 6.8b. Haltere switch rarely exceeds several seconds, and postural adjustments usually occur within ∼4 seconds, and both have relatively low variance 4.4, and 8.2 seconds, respectively. Feeding, proboscis pumping, and grooming tend to have longer bouts with greater variances. For example, the bout duration distribution of grooming has a variance of 15 seconds, and the mean bout duration is 6.2 seconds.

Another important aspect of fruit flies' behavioral repertoires is the total time spent while exhibiting each behavior (see Figure 6.8c). Grooming is the most exhibited behavior in terms of total time spent, with an average of 19.9 minutes, and a standard deviation of 5.2 minutes. We observe a great variance for time spent exhibiting haltere switch among experiments, the average, and the standard deviation are 4.5 minutes and 5.9 respectively.

Characterization of underlying changes in behaviors is essential for understanding sleep, and thus, we also examine behavioral visits that occurred during flies' sleep. To evaluate different experiments jointly, we align sleep bouts by dividing the longest dormancy epoch into bins. In Figure 6.9, behavioral visits are plotted with a joint x-axis, namely the sleep percentages. The most immediate observation about the temporal organization of the behavioral categories is of haltere switch. The occurrence of haltere switch is not uniform along the time, for 3 sleep-deprived and 3 wild-type sleep experiments 90% of the haltere switch occurrences are after 40 sleep percentage. For 8 out of 11 experiments, more than 40% of the switch-like haltere movements occur between 60 and 80 sleep percentages. Moreover, for 4 out of 11 experiments, 80% of haltere switches are observed between this interval. Another observation is that grooming behaviors are more likely to occur in the last 40 sleep percentage and very unlikely to occur within the first 20 sleep percentage.



(a) Rate of behavioral visits during the sleep within each 5 sleep percentage bin. Each row is normalized, to sum up to 1.



(b) Cumulative behavioral visits for each behavior, behaviors with less than 2 bouts are excluded.

Figure 6.9 Demonstration of behavioral visits during the sleep. Ethograms of flies are aligned by dividing the longest dormant epoch into 100 bins, corresponding to the sleep percentages. The total number of behavioral visits is normalized by the total number of bouts for each fly and behavioral category, separately.

7. CONCLUSION

Ethology, the scientific study of animal behavior, focuses on behavior under natural conditions, and is as a rapidly growing field. With its scientific roots in late 19th and 20th centuries, it became a well-recognized scientific discipline and yielded significant insight into general principles of the nervous system. Behavior, as arguably the most robust output of the brain, is needed to be deciphered to understand complex neurobiological phenomena, including sleep. Technical advances, ranging from imaging technologies to machine learning techniques, transformed ethology into a new computational discipline, and the field of computational ethology has emerged.

In the past decade, remarkable tools, which involve computer vision and machine learning, have been developed for automated quantification of animal behavior. Employing such tools provided unprecedented insight into complex behaviors, such as hunting and mating (Mearns et al., 2020; Janisch et al., 2021). Yet still, challenges remain open. While behaviors exhibited by awake animals tend to be characterized by major postural and positional changes, behavioral repertoire of asleep animals contains subtle movements and unobtrusive changes that sparsely occur during long sleep cycles. Thus,

BIBLIOGRAPHY

Ali, M., Jones, M. W., Xie, X., and Williams, M. (2019). TimeCluster: dimension reduction applied to temporal data for visual analytics. *The Visual Computer*, 35(6-8):1013–1026.

Anderson, D. J. and Perona, P. (2014). Toward a Science of Computational Ethology. *Neuron*, 84(1):18–31.

Belkin, M. and Niyogi, P. (2003). Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15(6):1373–1396. Conference Name: Neural Computation.

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517.

Berman, G. J., Bialek, W., and Shaevitz, J. W. (2016). Predictability and hierarchy in *Drosophila* behavior. *Proceedings of the National Academy of Sciences*, 113(42):11943–11948.

Berman, G. J., Choi, D. M., Bialek, W., and Shaevitz, J. W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99):20140672. Publisher: Royal Society.

Bohnslav, J. P., Wimalasena, N. K., Clausing, K. J., Dai, Y. Y., Yarmolinsky, D. A., Cruz, T., Kashlan, A. D., Chiappe, M. E., Orefice, L. L., Woolf, C. J., and Harvey, C. D. (2021). DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels. *eLife*, 10:e63377.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, COLT '92, pages 144–152, New York, NY, USA. Association for Computing Machinery.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Brown, A. E. X., Yemini, E. I., Grundy, L. J., Jucikas, T., and Schafer, W. R. (2013). A dictionary of behavioral motifs reveals clusters of genes affecting Caenorhabditis elegans locomotion. *Proceedings of the National Academy of Sciences*, 110(2):791–796. Publisher: Proceedings of the National Academy of Sciences.

Campbell, S. S. and Tobler, I. (1984). Animal sleep: a review of sleep duration across phylogeny. *Neuroscience and Biobehavioral Reviews*, 8(3):269–300.

Corner, M. A. (1977). Sleep and the beginnings of behavior in the Animal Kingdom—Studies of Ultradian motility cycles in early life. *Progress in Neurobiology*, 8:279–295.

Datta, S. R., Anderson, D. J., Branson, K., Perona, P., and Leifer, A. (2019). Computational Neuroethology: A Call to Action. *Neuron*, 104(1):11–24.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A*, 2(7):1160–1169. Publisher: Optica Publishing Group.

DeAngelis, B. D., Zavatone-Veth, J. A., and Clark, D. A. (2019). The manifold structure of limb coordination in walking Drosophila. *eLife*, 8:e46409.

Dong, W., Moses, C., and Li, K. (2011). Efficient k-nearest neighbor graph construction for generic similarity measures. In *Proceedings of the 20th international conference on World wide web - WWW '11*, page 577, Hyderabad, India. ACM Press.

Geissmann, Q., Beckwith, E. J., and Gilestro, G. F. (2019). Most sleep does not serve a vital function: Evidence from Drosophila melanogaster. *Science Advances*, 5(2):eaau9253. Publisher: American Association for the Advancement of Science.

Geissmann, Q., Rodriguez, L. G., Beckwith, E. J., French, A. S., Jamasb, A. R., and Gilestro, G. F. (2017). Ethoscopes: An open platform for high-throughput ethomics. *PLOS Biology*, 15(10):e2003026. Publisher: Public Library of Science.

Graving, J. M. and Couzin, I. D. (2020). VAE-SNE: a deep generative model for simultaneous dimensionality reduction and clustering. Pages: 2020.07.17.207993 Section: New Results.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Las Vegas, NV, USA. IEEE.

Hellinger, E. (1909). Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, 1909(136):210–271. Publisher: De Gruyter.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441. Publisher: Warwick & York.

Hsu, A. I. and Yttri, E. A. (2021). B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nature Communications*, 12(1):5188.

Itskov, P. M., Moreira, J.-M., Vinnik, E., Lopes, G., Safarik, S., Dickinson, M. H., and Ribeiro, C. (2014). Automated monitoring and quantitative analysis of feeding behaviour in Drosophila. *Nature Communications*, 5(1):4560. Number: 1 Publisher: Nature Publishing Group.

Janisch, J., Perinot, E., Fusani, L., and Quigley, C. (2021). Deciphering choreographies of elaborate courtship displays of golden-collared manakins using markerless motion capture. *Ethology*, 127(7):550–562. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/eth.13161.

Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S., and Branson, K. (2013). JAABA:

interactive machine learning for automatic annotation of animal behavior. *Nature Methods*, 10(1):64–67. Number: 1 Publisher: Nature Publishing Group.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27.

Liu, Y., San Liang, X., and Weisberg, R. H. (2007). Rectification of the Bias in the Wavelet Power Spectrum. *Journal of Atmospheric and Oceanic Technology*, 24(12):2093–2102.

Maaten, L. v. d. and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605.

Marques, J. C., Lackner, S., Félix, R., and Orger, M. B. (2018). Structure of the Zebrafish Locomotor Repertoire Revealed with Unsupervised Behavioral Clustering. *Current biology: CB*, 28(2):181–195.e5.

Marshall, J. D., Aldarondo, D. E., Dunn, T. W., Wang, W. L., Berman, G. J., and Ölveczky, B. P. (2021). Continuous Whole-Body 3D Kinematic Recordings across the Rodent Behavioral Repertoire - SI. *Neuron*, 109(3):420–437.e8.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9):1281–1289. Number: 9 Publisher: Nature Publishing Group.

McInnes, L., Healy, J., and Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]*. arXiv: 1802.03426.

Mearns, D. S., Donovan, J. C., Fernandes, A. M., Semmelhack, J. L., and Baier, H. (2020). Deconstructing Hunting Behavior Reveals a Tightly Coupled Stimulus-Response Loop. *Current biology: CB*, 30(1):54–69.e9.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, pages 3111–3119, Red Hook, NY, USA. Curran Associates Inc.

Nath, R. D., Bedbrook, C. N., Abrams, M. J., Basinger, T., Bois, J. S., Prober, D. A., Sternberg, P. W., Gradinaru, V., and Goentoro, L. (2017). The Jellyfish Cassiopea Exhibits a Sleep-like State. *Current Biology*, 27(19):2984–2990.e3.

Nene, S. A., Nayar, S. K., and Murase, H. (1996). Columbia Object Image Library (COIL-20). Technical Report CUCS-005-96, Columbia University.

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., and Golden, S. A. (2020). Simple Behavioral Analysis (SimBA) – an open source toolkit for computer classification of complex social behaviors in experimental animals. preprint, Animal Behavior and Cognition.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, . (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830.

Pereira, T. D. (2020). Quantifying behavior to understand the brain. *Nature Neuroscience*, 23:13.

Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M., and Shaevitz, J. W. (2019). Fast animal pose estimation using deep neural networks. *Nature Methods*, 16(1):117–125. Number: 1 Publisher: Nature Publishing Group.

Pereira, T. D., Tabris, N., Matsliah, A., Turner, D. M., Li, J., Ravindranath, S., Papadoyannis, E. S., Normand, E., Deutsch, D. S., Wang, Z. Y., McKenzie-Smith, G. C., Mitelut, C. C., Castro, M. D., D'Uva, J., Kislin, M., Sanes, D. H., Kocher, S. D., Wang, S. S.-H., Falkner, A. L., Shaevitz, J. W., and Murthy, M. (2022). SLEAP: A deep learning system for multi-animal pose tracking. *Nature Methods*, 19(4):486–495.

Pfeiffenberger, C., Lear, B. C., Keegan, K. P., and Allada, R. (2010a). Locomotor Activity Level Monitoring Using the Drosophila Activity Monitoring (DAM) System. *Cold Spring Harbor Protocols*, 2010(11):pdb.prot5518. Publisher: Cold Spring Harbor Laboratory Press.

Pfeiffenberger, C., Lear, B. C., Keegan, K. P., and Allada, R. (2010b). Processing sleep data created with the Drosophila Activity Monitoring (DAM) System. *Cold Spring Harbor Protocols*, 2010(11):pdb.prot5520.

Qiao, B., Li, C., Allen, V. W., Shirasu-Hiza, M., and Syed, S. (2018). Automated analysis of long-term grooming behavior in Drosophila using a k-nearest neighbors classifier. *eLife*, 7:e34497.

Rauch, H. E., Striebel, C. T., and Tung, F. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, 3(8):1445–1450.

Sainburg, T., McInnes, L., and Gentner, T. Q. (2021). Parametric UMAP Embeddings for Representation and Semisupervised Learning. *Neural Computation*, 33(11):2881–2907.

Sauer, S., Kinkelin, M., Herrmann, E., and Kaiser, W. (2003). The dynamics of sleep-like behaviour in honey bees. *Journal of Comparative Physiology A*, 189(8):599–607.

Tang, J., Liu, J., Zhang, M., and Mei, Q. (2016). Visualizing Large-scale and High-dimensional Data. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, pages 287–297, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.

Tenenbaum, J. B., Silva, V. d., and Langford, J. C. (2000). A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323. Publisher: American Association for the Advancement of Science.

Todd, J. G., Kain, J. S., and Bivort, B. L. d. (2017). Systematic exploration of unsuper-

vised methods for mapping behavior. *Physical Biology*, 14(1):015002. Publisher: IOP Publishing.

Torrence, C. and Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78.

van Alphen, B., Semenza, E. R., Yap, M., van Swinderen, B., and Allada, R. (2021). A deep sleep stage in Drosophila with a functional role in waste clearance. *Science Advances*, 7(4):eabc2999. Publisher: American Association for the Advancement of Science.

Whiteway, M. R., Biderman, D., Friedman, Y., Dipoppa, M., Buchanan, E. K., Wu, A., Zhou, J., Bonacchi, N., Miska, N. J., Noel, J.-P., Rodriguez, E., Schartner, M., Socha, K., Urai, A. E., Salzman, C. D., Laboratory, T. I. B., Cunningham, J. P., and Paninski, L. (2021). Partitioning variability in animal behavioral videos using semi-supervised variational autoencoders. *PLOS Computational Biology*, 17(9):e1009439. Publisher: Public Library of Science.

Wiggin, T. D., Goodwin, P. R., Donelson, N. C., Liu, C., Trinh, K., Sanyal, S., and Griffith, L. C. (2020). Covert sleep-related biological processes are revealed by probabilistic analysis in Drosophila. *Proceedings of the National Academy of Sciences*, 117(18):10024–10034. Publisher: Proceedings of the National Academy of Sciences.

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., Abraira, V. E., Adams, R. P., and Datta, S. R. (2015). Mapping Sub-Second Structure in Mouse Behavior. *Neuron*, 88(6):1121–1135.

Wu, Y. E., Dang, J., Kingsbury, L., Zhang, M., Sun, F., Hu, R. K., and Hong, W. (2021). Neural control of affiliative touch in prosocial interaction. *Nature*, 599(7884):262–267.

Xiao, H., Rasul, K., and Vollgraf, R. (2017). Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. Number: arXiv:1708.07747 arXiv:1708.07747 [cs, stat].

Ye, L. and Keogh, E. (2011). Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. *Data Mining and Knowledge Discovery*, 22(1):149–182.