

SLEAP into Behavioral Analysis: A Deep Learning Approach

M.S. Project

November 21, 2024

Madhumitha Sekamuri

Advisor: Dr. Rajiv Ramnath

Associate Professor: Dr. Laurence Coutellier, Behavioral Neuroscience

Department of Computer Science and Engineering

The Ohio State University



Introduction

1. Motivation
2. Problem Statement
3. Proposed Solution
4. Design Chart
5. Training
 - SLEAP – why?
 - Data Labelling
 - Model Parameters
 - Loss
6. Inference
 - Prediction & H5 File generation
7. Analysis
 - Results
8. Conclusion & Future work

Motivation

- Understanding how mice use sniffing behaviors to communicate can uncover fundamental principles of social interaction.
- Differentiating sniffing types provides insight into how age, sex, and individual traits influence social behaviors.
- Linking sniffing patterns to neural circuitry helps identify the neurological basis of social interaction.
- Analyzing temporal patterns of social behaviors offers a deeper understanding of how interactions evolve over time.
- This research can advance neuroscience by revealing mechanisms behind behavioral differences in animal models.
- Machine learning enables efficient and precise analysis of complex social behaviors, paving the way for innovative approaches to behavior research.

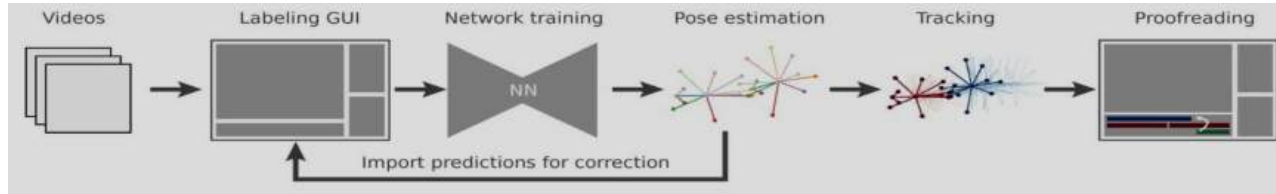
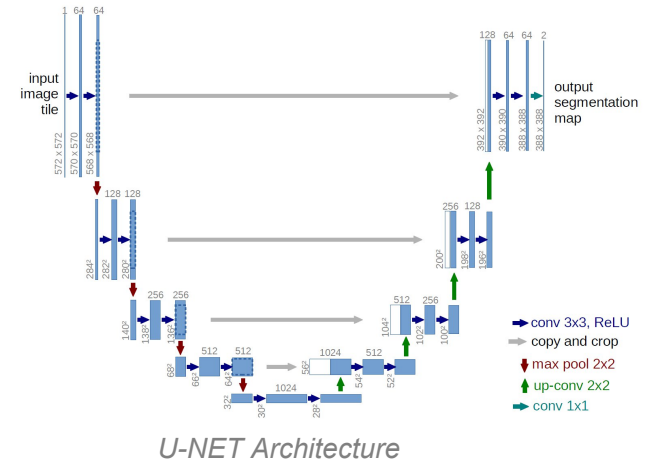
Problem Statement

- Develop methods to predict specific social behaviors (e.g., ano-genital sniffing, nose-to-nose, nose-to-neck, nose-to-left ear, nose-to-right ear) across individual videos and group categories, such as sex and age groups.
- Quantify the frequency of behavioral events, such as the number of occurrences of ano-genital sniffing.
- Measure the duration of each behavioral event to understand its temporal characteristics.
- Calculate the latency to the first occurrence of each behavior within observed interactions.
- Analyze behavior patterns observed during the initial three minutes of interaction, including corresponding latencies.
- Automate these analyses using machine learning to eliminate the need for manual tracking and annotation, improving efficiency and accuracy.



Proposed Solution

- Utilize transfer learning with pre-trained models such as U-Net, ResNet152, Hourglass, Pretrained Encoder and many more models to accurately capture and analyze complex social behaviors.
- Leverage open-source tools, including SLEAP and DeepLabCut, renowned for their effectiveness in animal pose estimation, to extract behavioral parameters.
- Implement SLEAP for automated behavioral tracking to streamline data collection and analysis.
- Employ U-Net as the backbone model to achieve precise identification and classification of specific behaviors.



Flowchart



SLEAP for APT

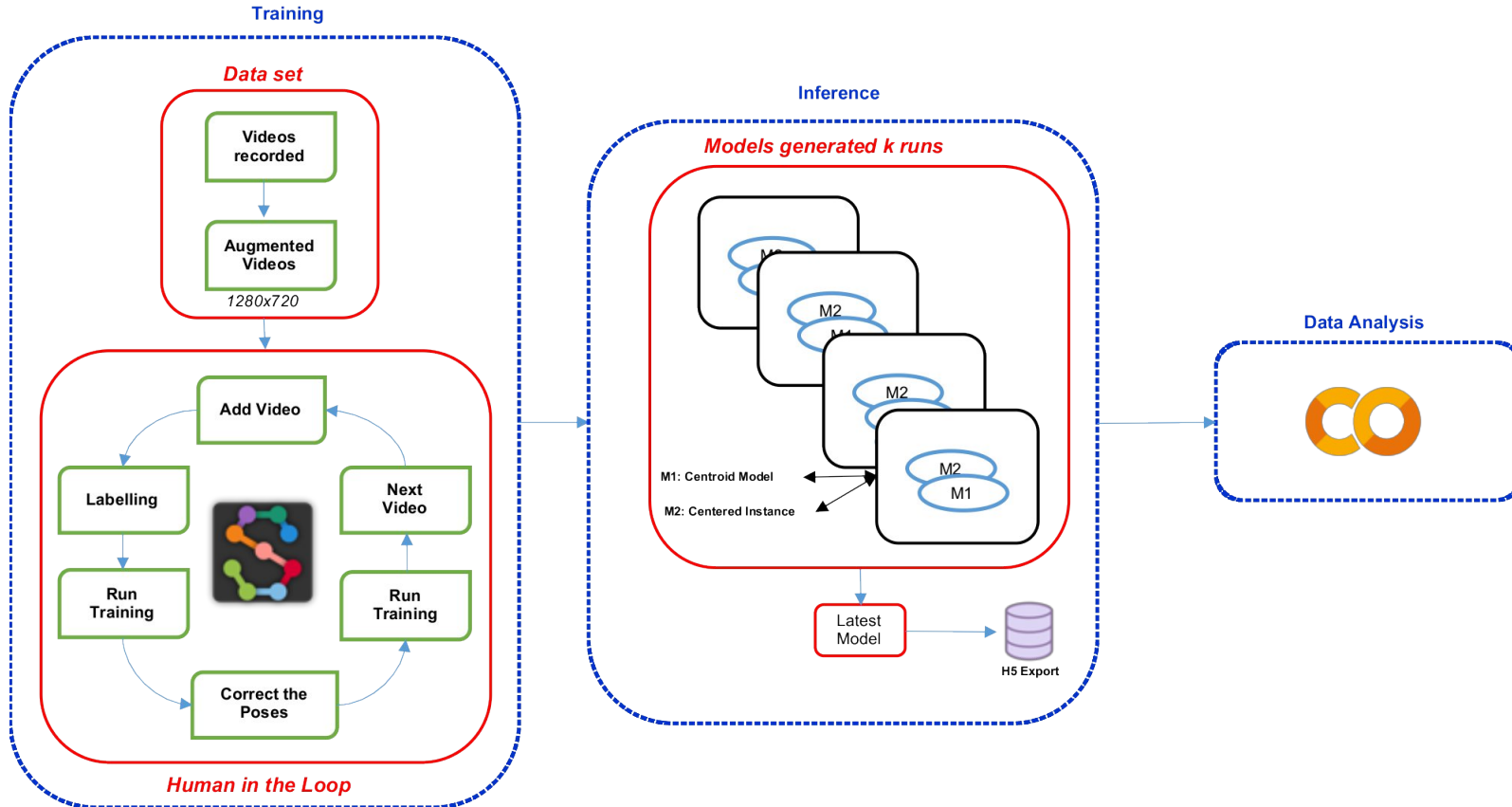
Approach



SLEAP – Why?

- Deep-learning-based framework for precise multi-animal pose tracking and behavioral analysis.
- Enables training and utilization of models to automatically track movements across diverse animal types.
- Provides a user-friendly interface for labeling, training, and proofreading pose estimation.
- Supports single- and multi-animal tracking with both top-down and bottom-up training strategies.
- Offers state-of-the-art pretrained and customizable neural network architectures, requiring minimal labeled data.
- Ensures fast training and inference, facilitating efficient analysis of recorded videos.
- Supports remote workflows and includes a flexible developer API for integration and customization.

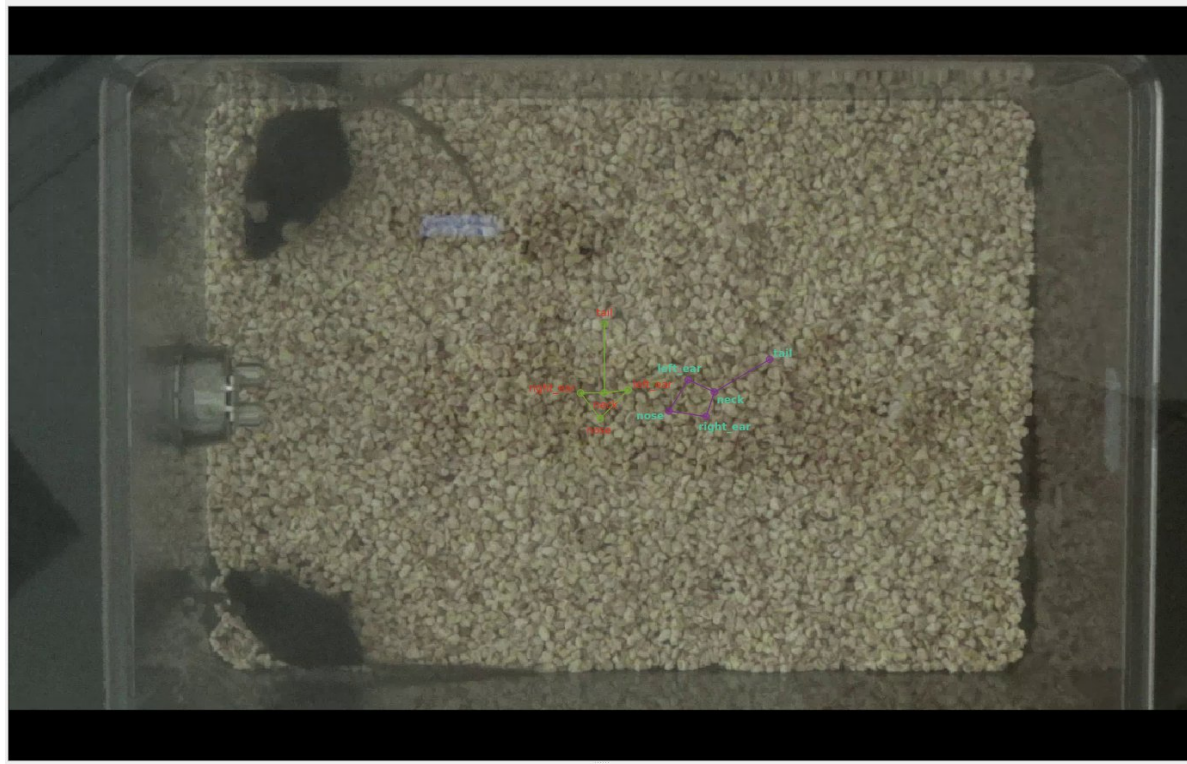
Design Chart



Training



Data Labelling - Before

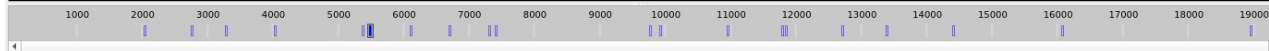


The screenshot displays a video labeling application. The main window shows a video frame of a mouse in a container, with a skeleton overlay identifying body parts: tail, left ear, right ear, nose, neck, and whiskers. The interface includes a timeline at the bottom with frame numbers from 1000 to 19000. Below the timeline, it shows 'Video 13/13', 'Frame: 5,491/19,380', 'Labeled Frames: 1 in video, 555 in project', and 'Current frame: 2 instances'. On the right, an 'Instances' panel contains a table with two rows of data.

| Instances | | | |
|-----------|--------|-------|------------|
| | Points | Track | Skeleton |
| 1 | 5/5 | | Skeleton-5 |
| 2 | 5/5 | | Skeleton-5 |

Below the table are buttons for 'New Instance' and 'Delete Instance'. At the bottom of the right panel are tabs for 'Videos', 'Skeleton', 'Instances', and 'Labeling Suggestions'.

Data Labelling – After Correction



Video 13/13 Frame: 5,491/19,380 Labeled Frames: 1 in video, 555 in project Current frame: 2 instances

Instances

| | Points | Track | Score | Skeleton |
|---|--------|-------|-------|------------|
| 1 | 5/5 | | | Skeleton-5 |
| 2 | 5/5 | | | Skeleton-5 |

New Instance Delete Instance

Videos Skeleton Instances Labeling Suggestions



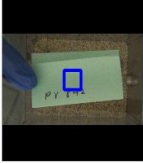
Model training Parameters

SLEAP v1.3.3

Training Pipeline Centroid Model Configuration Centered Instance Model Configuration

[Trained] 240930_225521.centroid.n=129(training_config.json)

☐ Use Trained Model
☐ Resume Training



Receptive Field for centroid:
76 pixels

Receptive field size is a function of the number of down blocks (4), the number of convolutions per block (2), and the convolution kernel size (3).

You can control the number of down blocks by setting the **Max Stride** (16).

The number of convolutions per block and the kernel size are currently fixed by your choice of backbone.

You can also control the receptive field size relative to the original image by adjusting the **Input Scaling** (0.5).

Data

Validation fraction: 0.10
Input Scaling: 0.50
Crop Size: 0 ☒ Auto

Augmentation

Rotate: ☒
Rotation Min Angle: -15.00
Rotation Max Angle: 15.00
Scale: ☐
Scale Min: 0.90
Scale Max: 1.10
Random flip: none
Uniform Noise: ☐
Uniform Noise Min Val: 0.00
Uniform Noise Max Val: 10.00
Gaussian Noise: ☐
Gaussian Noise Mean: 5.00
Gaussian Noise Stddev: 1.00
Contrast: ☐
Contrast Min Gamma: 0.50
Contrast Max Gamma: 2.00
Brightness: ☐
Brightness Min Val: 0.00
Brightness Max Val: 10.00

Model

Backbone: unet

Stem Stride: 0 ☒ None
Max Stride: 16
Filters: 16
Filters Rate: 2.00
Middle Block: ☒
Up Interpolate: ☒
Heads: centroid

Anchor Part: neck
Sigma: 2.50
Output Stride: 2

Copy to clipboard Save configuration files... Export training job package... Run Cancel


Centroid Model

SLEAP v1.3.3

Training Pipeline Centroid Model Configuration Centered Instance Model Configuration

[Trained] 240930_230309.centered_instance.n=129(training_config.json)

☐ Use Trained Model
☐ Resume Training



Receptive Field for centered_instance:
76 pixels

Receptive field size is a function of the number of down blocks (4), the number of convolutions per block (2), and the convolution kernel size (3).

You can control the number of down blocks by setting the **Max Stride** (16).

The number of convolutions per block and the kernel size are currently fixed by your choice of backbone.

You can also control the receptive field size relative to the original image by adjusting the **Input Scaling** (1.0).

Data

Validation fraction: 0.10
Input Scaling: 1.00
Crop Size: 336 ☐ Auto

Augmentation

Rotate: ☒
Rotation Min Angle: -15.00
Rotation Max Angle: 15.00
Scale: ☐
Scale Min: 0.90
Scale Max: 1.10
Random flip: none
Uniform Noise: ☐
Uniform Noise Min Val: 0.00
Uniform Noise Max Val: 10.00
Gaussian Noise: ☐
Gaussian Noise Mean: 5.00
Gaussian Noise Stddev: 1.00
Contrast: ☐
Contrast Min Gamma: 0.50
Contrast Max Gamma: 2.00
Brightness: ☐
Brightness Min Val: 0.00
Brightness Max Val: 10.00

Model

Backbone: unet

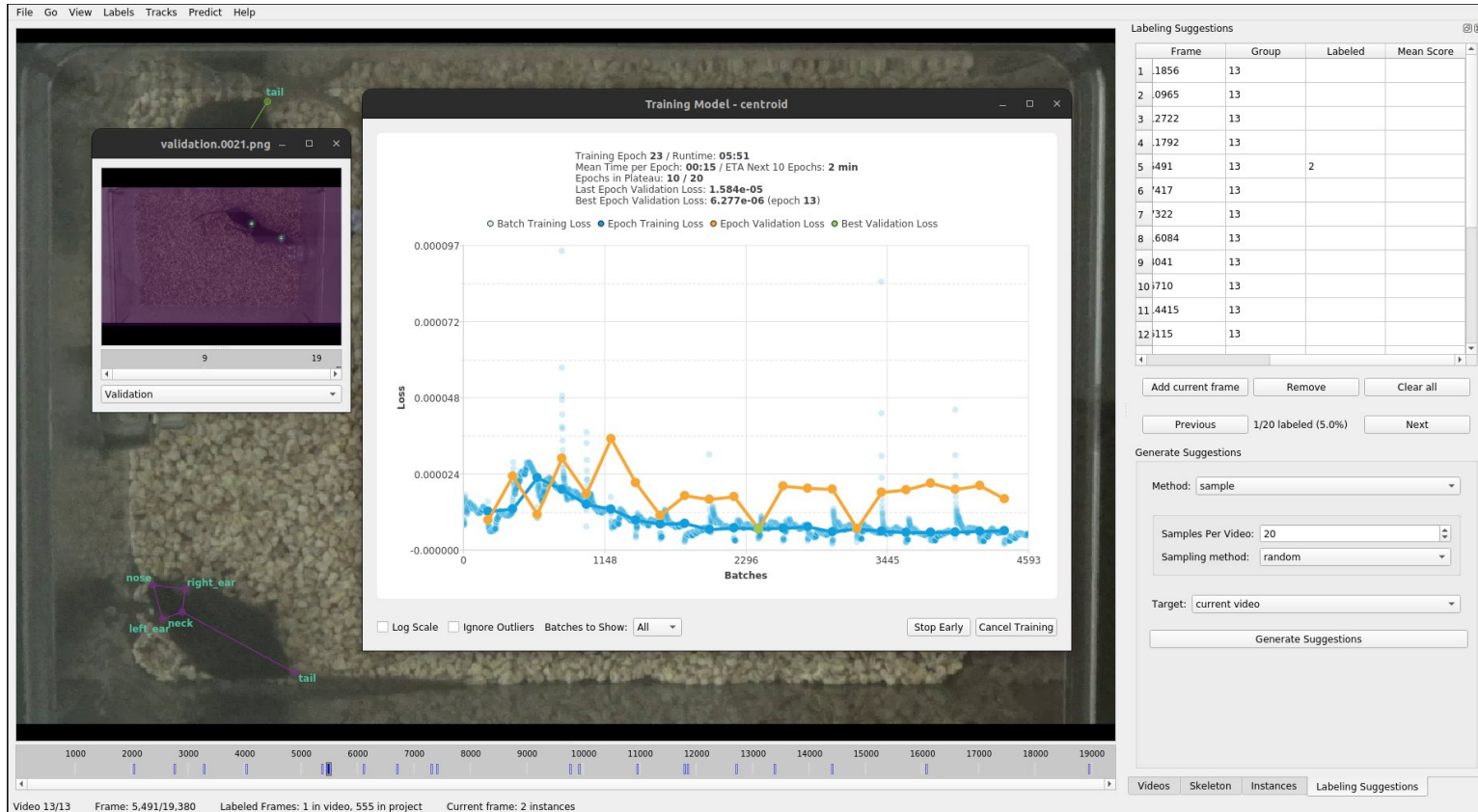
Stem Stride: 0 ☒ None
Max Stride: 16
Filters: 24
Filters Rate: 2.00
Middle Block: ☒
Up Interpolate: ☒
Heads: centered_instance

Anchor Part: neck
Sigma: 2.50
Output Stride: 4

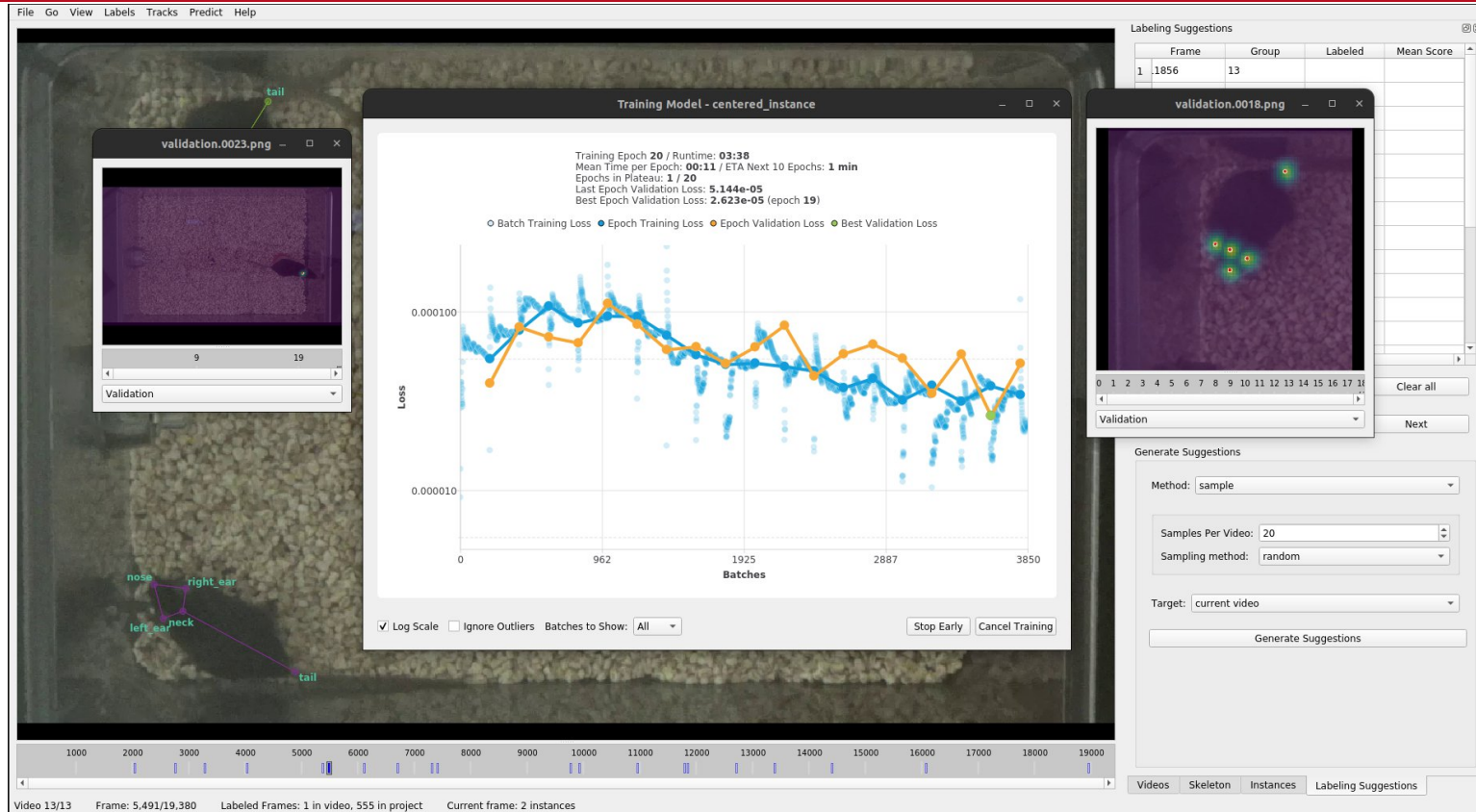
Copy to clipboard Save configuration files... Export training job package... Run Cancel

Centered Instance

Centroid Loss



Centered Instance Loss



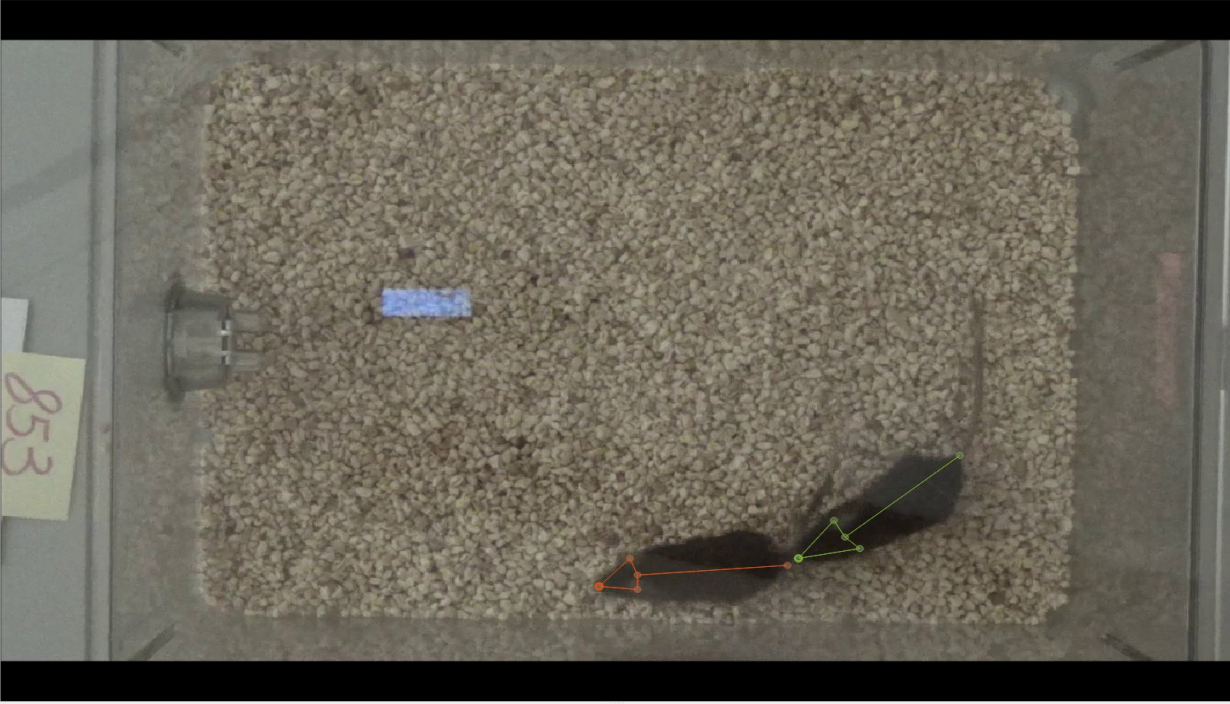
Inference



Prediction and Export H5 file

/home/sathya/Downloads/D1/842/D1.slp - SLEAP v1.3.4

File Go View Labels Tracks Predict Help



Instances

| | Points | Track | Score | Skeleton |
|---|--------|---------|-------|------------|
| 1 | 5/5 | track_0 | 0.75 | Skeleton-5 |
| 2 | 5/5 | track_1 | 0.96 | Skeleton-5 |

Video 2/12 Frame: 3,990/19,200 Labeled Frames: 10 in video, 554 in project Predicted Frames: 17,748 (92.44%) in video Current frame: 2 Instances

New Instance Delete Instance

Videos Skeleton Instances Labeling Suggestions



Analysis



Results

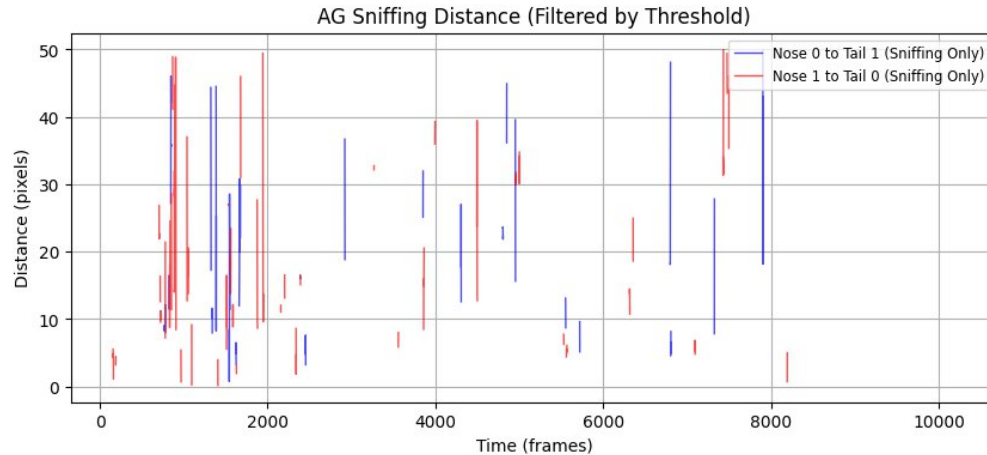
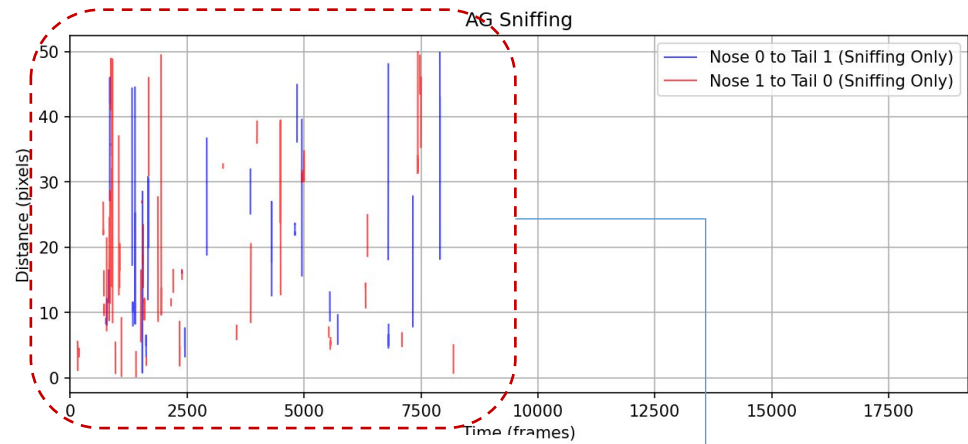
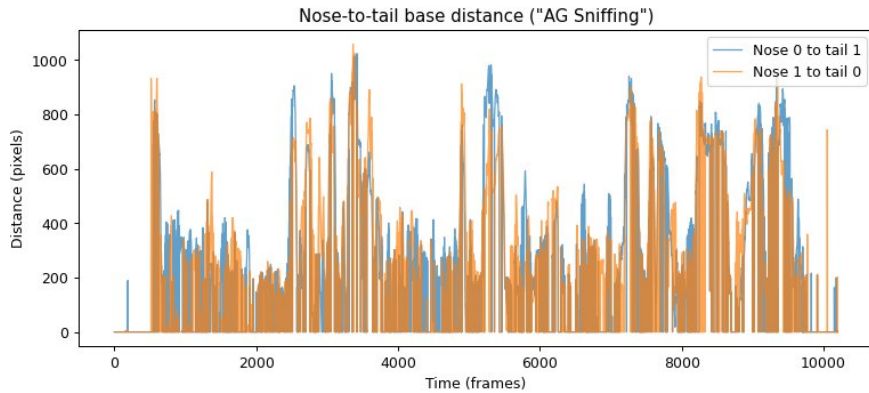
- Analyze the following behavioral parameters for both individual videos and groups of five videos:
 - Nose-to-tail sniffing (ano-genital sniffing).
 - Nose-to-nose sniffing.
 - Nose-to-right-ear sniffing.
 - Nose-to-left-ear sniffing.
 - Nose-to-neck sniffing.
- Determine the first occurrence of each behavior for one video
- Quantify the total occurrences of each behavior for one video
- Examine each behavior and its latency during the first three minutes of the video.

```
with h5py.File("/content/drive/MyDrive/Neuroscience/H5/Female\
adult C/853.000_853 interaction.analysis.h5", "r") as f:
    print("Available datasets:")
    def printname(name):
        print(name)
    f.visit(printname) # This will print all dataset names
```

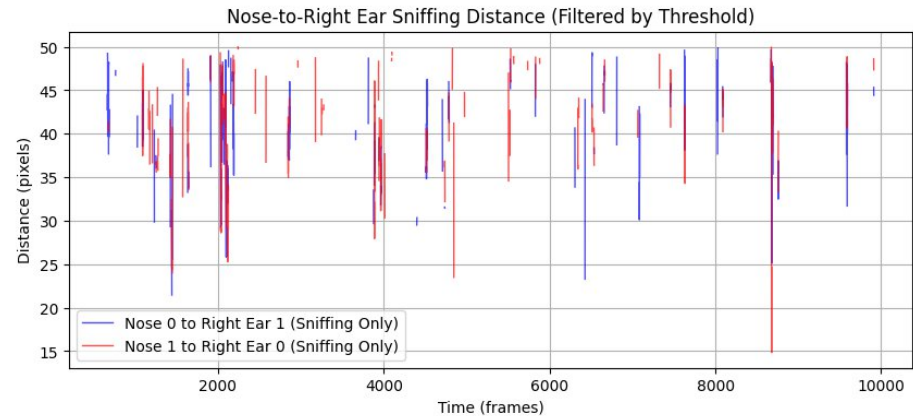
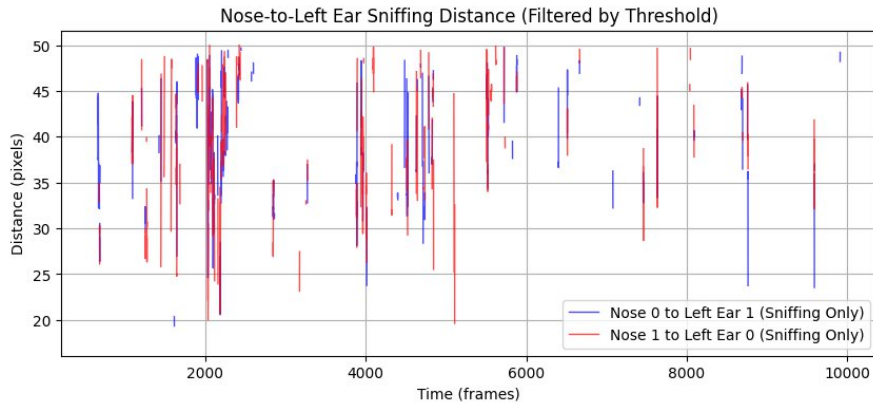
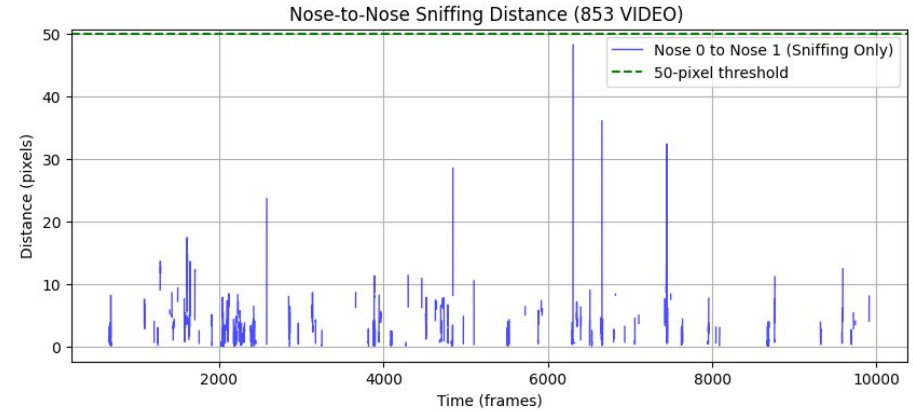
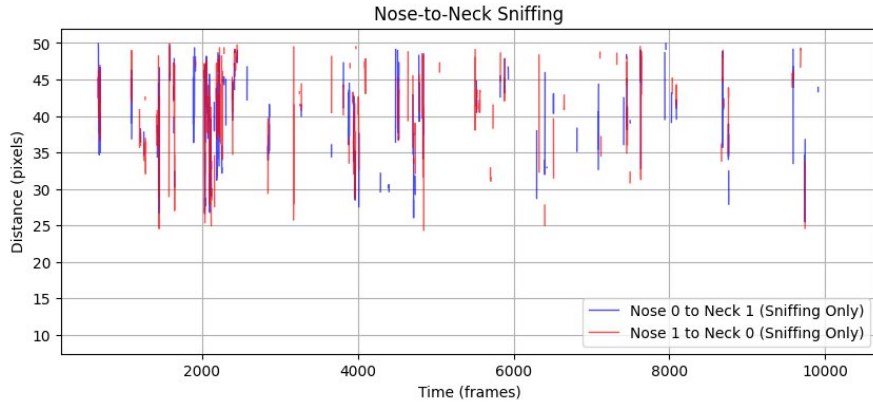
```
Available datasets:
edge_inds
edge_names
instance_scores
labels_path
node_names
point_scores
provenance
track_names
track_occupancy
tracking_scores
tracks
video_ind
video_path
```

Structure of H5 file

Results for video 853 from Female Adult C group



Results for video 853 from Female Adult C group



Results for video 853 from Female Adult C group

Number of AG sniffing events (Nose 0 to Tail 1): 146

Number of AG sniffing events (Nose 1 to Tail 0): 198

Total number of AG sniffing events: **344**

Nose-to-Neck sniffing events (Nose 0 to Neck 1): 392

Nose-to-Neck sniffing events (Nose 1 to Neck 0): 418

Total Nose-to-Neck sniffing events: **810**

Total Nose-to-Nose sniffing events: **406**

Nose-to-Left Ear sniffing events (Nose 0 to Left Ear 1): 352

Nose-to-Left Ear sniffing events (Nose 1 to Left Ear 0): 342

Total Nose-to-Left Ear sniffing events: **694**

Nose-to-Right Ear sniffing events (Nose 0 to Right Ear 1): 270

Nose-to-Right Ear sniffing events (Nose 1 to Right Ear 0): 304

Total Nose-to-Right Ear sniffing events: **574**

First occurrence of Nose 0 to Right Ear 1: Frame 663, Time 22.10 seconds

First occurrence of Nose 0 to Left Ear 1: Frame 663, Time 22.10 seconds

First occurrence of Nose to Nose: Frame 663, Time 22.10 seconds

First occurrence of Nose 0 to Neck 1: Frame 663, Time 22.10 seconds

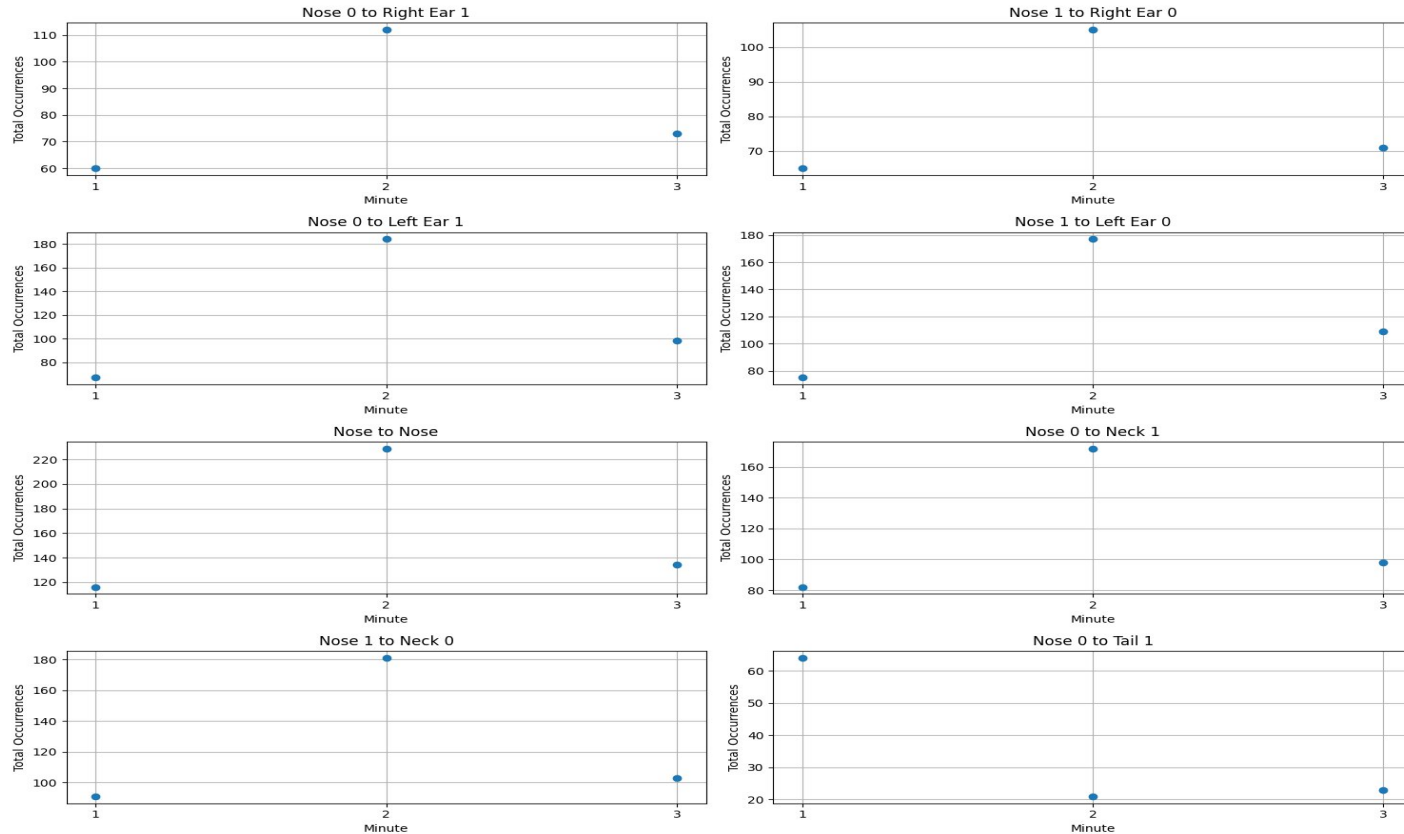
First occurrence of Nose 1 to Neck 0: Frame 665, Time 22.17 seconds

First occurrence of Nose 1 to Right Ear 0: Frame 679, Time 22.63 seconds

First occurrence of Nose 1 to Left Ear 0: Frame 679, Time 22.63 seconds

First occurrence of Nose 0 to Tail 1: Frame 690, Time 23.00 seconds

Results for video 853 from Female Adult C group



Results for video 853 from Female Adult C group

Observation of each behavior and its latency in the first 3 minutes of the video

Nose 0 to Right Ear 1:

- Minute 1: 60 occurrences, Total Time: 2.00 seconds
- Minute 2: 112 occurrences, Total Time: 3.73 seconds
- Minute 3: 73 occurrences, Total Time: 2.43 seconds

Nose 1 to Right Ear 0:

- Minute 1: 65 occurrences, Total Time: 2.17 seconds
- Minute 2: 105 occurrences, Total Time: 3.50 seconds
- Minute 3: 71 occurrences, Total Time: 2.37 seconds

Nose 0 to Left Ear 1:

- Minute 1: 67 occurrences, Total Time: 2.23 seconds
- Minute 2: 184 occurrences, Total Time: 6.13 seconds
- Minute 3: 98 occurrences, Total Time: 3.27 seconds

Nose 1 to Left Ear 0:

- Minute 1: 75 occurrences, Total Time: 2.50 seconds
- Minute 2: 177 occurrences, Total Time: 5.90 seconds
- Minute 3: 109 occurrences, Total Time: 3.63 seconds

Nose to Nose:

- Minute 1: 116 occurrences, Total Time: 3.87 seconds
- Minute 2: 229 occurrences, Total Time: 7.63 seconds
- Minute 3: 134 occurrences, Total Time: 4.47 seconds

Nose 0 to Neck 1:

- Minute 1: 82 occurrences, Total Time: 2.73 seconds
- Minute 2: 172 occurrences, Total Time: 5.73 seconds
- Minute 3: 98 occurrences, Total Time: 3.27 seconds

Nose 1 to Neck 0:

- Minute 1: 91 occurrences, Total Time: 3.03 seconds
- Minute 2: 181 occurrences, Total Time: 6.03 seconds
- Minute 3: 103 occurrences, Total Time: 3.43 seconds

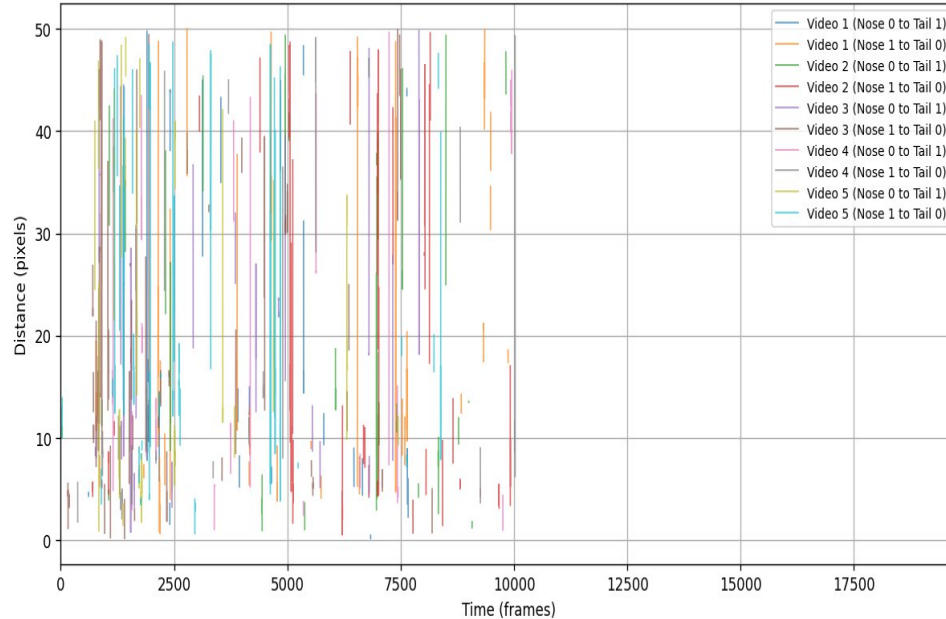
Nose 0 to Tail 1:

- Minute 1: 64 occurrences, Total Time: 2.13 seconds
- Minute 2: 21 occurrences, Total Time: 0.70 seconds
- Minute 3: 23 occurrences, Total Time: 0.77 seconds

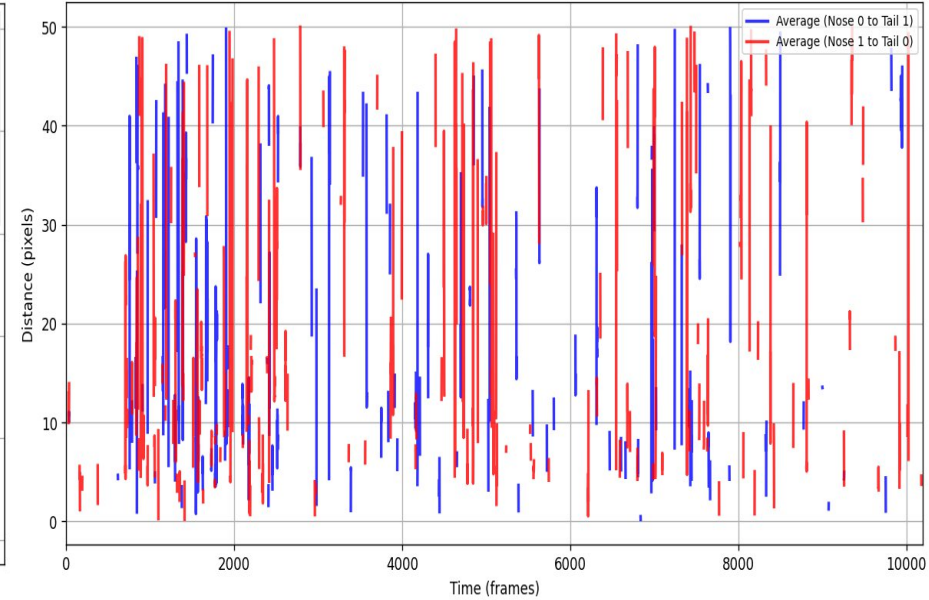


Results for all videos from Female Adult C group

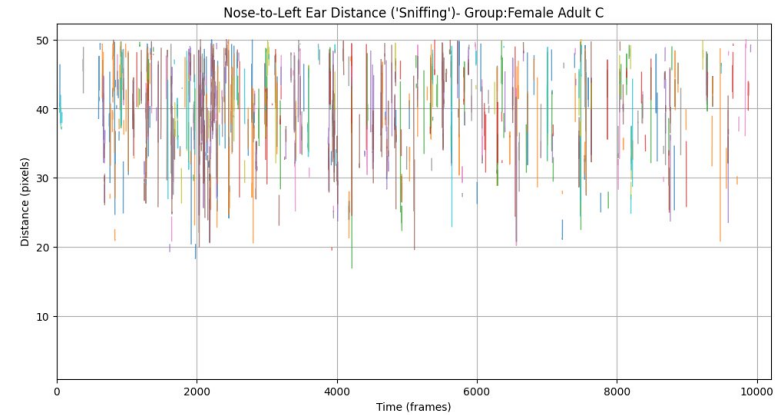
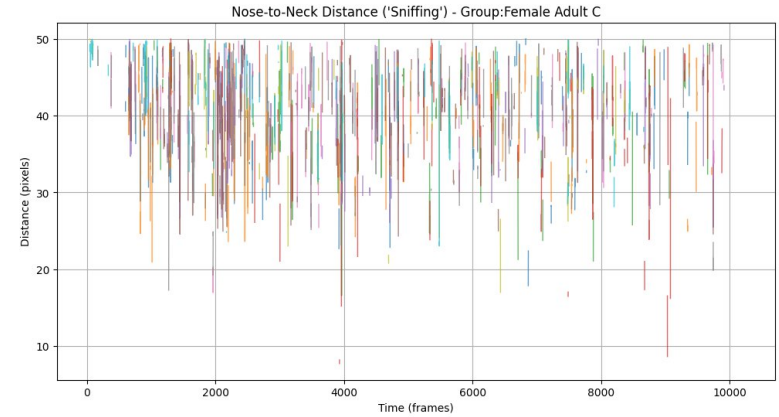
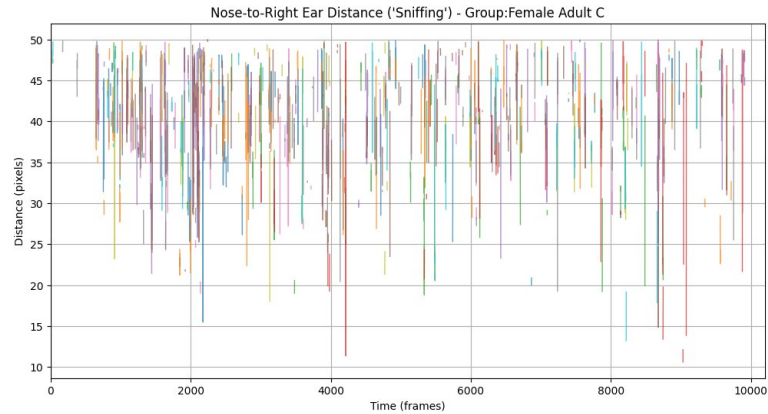
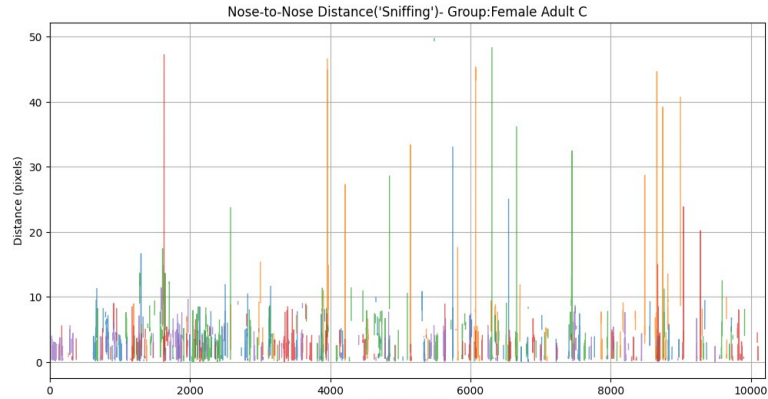
Nose-to-Tail Base Distance ('AG Sniffing') - Group:Female Adult C



Average AG Sniffing Distance - Combined 5 Videos (Female Adult C)



Results for all videos from Female Adult C group



Conclusion and Future Work

The study successfully utilized SLEAP for automated pose tracking to analyze mouse social behaviors, quantifying sniffing interactions with high accuracy. Key findings highlight patterns in behavior duration, frequency, and latency, providing insights into social and neural interactions. This methodology improves efficiency over manual tracking and lays a foundation for scalable behavioral neuroscience studies.

FUTURE WORK : MOSEQ (MOtion SEQuencing)

Learning Stereotyped Patterns: MoSeq identifies and learns repetitive, consistent movement patterns ("stereotyped behaviors") from keypoint tracking data, enabling automatic behavioral classification.

Temporal Analysis: MoSeq determines when these learned behaviors occur over time, providing insights into the sequence and timing of actions during social interactions.

Comprehensive Toolkit: MoSeq offers tools to fit its probabilistic model to keypoint tracking data (e.g., from SLEAP) and analyze the results for behavioral studies.

REFERENCES:

<https://sleap.ai/tutorials/tutorial.html>

<https://www.nature.com/articles/s41592-022-01426-1>

<https://keypoint-moseq.readthedocs.io/en/latest/>



Questions?

Thank You

