Course 71254 - Introduction to Image Processing (2025)

Final Project

Using Deep Neural Network for Automated Mouse Tracking and Pose Estimation with DeepLabCut

Shani Orhof

16/02/2025

Project's Git repository

Project's Google Drive folder

Introduction

The question “What is personality?” is complex and widely debated. One approach to answering this is trait theory, which defines personality in terms of enduring characteristics, or traits. Traits such as consideration and honesty are specific aspects of the self that influence an individual’s behavior, emotions, and cognition. These traits remain relatively stable across time and contexts for a given individual but vary between people. Personality can be understood as a collection of an individual’s traits, each existing on a continuum. Different people exhibit varying levels of different traits, and measuring these levels allows for the assessment of personality. This idea underpins the most widely used personality framework—the Big Five personality traits, also known as the Five-Factor Model. The Big Five model groups numerous smaller traits into five broad dimensions (1) . For example, a person who is highly responsible and stubborn but low in spontaneity would score high in conscientiousness. Self-report questionnaires provide a way to quantify these traits by assessing how they manifest in thoughts, feelings, and behaviors.

One does not need to be a person to have personality. However, the scientific study of these differences as "personality" only emerged in the 20th century (2). Research on animal personality is heavily influenced by human personality studies, leading to the development of animal-specific adaptations of the Big Five model. Unlike humans, animals cannot self-report, and their internal states cannot be directly observed. Therefore, researchers must infer personality traits from behavioral observations. To do this, various tests have been designed to assess animal personality by measuring individual behavioral responses under controlled conditions. For example, the Novel Environment Test evaluates exploratory tendencies by analyzing how an animal reacts to being placed in an unfamiliar setting(3). This approach is widely used because it is both practical and allows for comparability and repeatability across research groups and species. As a result, it remains the most common method for studying animal personality (4).

However, this method also has significant limitations. While it reduces the influence of individual researchers' subjective interpretations, its foundation still relies on human assumptions about the motives behind animal behavior. The Novel Environment Test has been criticized for not accurately measuring exploration but rather reflecting an anxiety response (5) or neophobia (6). Another major criticism of behavioral testing is the nature of the test environments themselves. These tests are typically conducted in a setting separate from the animal’s natural living environment. The unfamiliar surroundings and handling process can cause significant stress, potentially skewing results. Additionally, behavioral tests examine only a limited range of actions. this narrow scope may fail to capture the full spectrum of meaningful behaviors, leading to an incomplete or misleading assessment of personality.

To address these criticisms while maintaining the trait theory approach to personality research, my lab employs a home cage monitoring system that continuously records and categorizes behavior. Groups of four mice are housed in semi-naturalistic enriched arenas where their activity is filmed. To visually distinguish between individuals, their fur is painted in one of four bright colors—red, green, blue, or yellow. Based on the "Social Box" method described by Forkosh et al. (2019) (7), we take an alternative approach to trait classification. Instead of categorizing behaviors according to human-defined traits, we use linear discriminant analysis (LDA). LDA processes behavioral data to identify dimension spaces that align with the definition of a personality trait—maximizing variability between individuals while minimizing variability within individuals. This results in uncorrelated axes, which we refer to as Identity Domains (ID), spanning a distinct behavioral subspace. Each mouse receives a score on each ID based on its behavior, providing an objective, data-driven measurement of personality traits.

This process is not without its limitations—we are constrained by the number of behaviors the system can automatically recognize. Currently, the system identifies mice as clusters of pixels that contrast with a predefined static black background. Each mouse's identity is determined by its assigned color, and its center of mass is defined as the cluster's midpoint. By tracking the movement of this center across the arena, we can recognize more than 60 unique behaviors. However, the system is unable to detect behaviors that do not involve movement through space. One of my primary goals for my master's research is to identify and incorporate these stationary behaviors into our Identity Domain (ID) analysis. Achieving this requires tracking the movement of different body parts from video data, rather than relying solely on center-of-mass movement.

The goal of this project is to automatically identify mice by their color and track the positions of their body parts using our existing monitoring videos. To achieve this, I have implemented DeepLabCut (DLC), an open-source Python package for animal pose estimation. As described by Mathis et al. (2018) (8), DLC is a deep neural network built on feature detection tools from the human pose estimation algorithm DeeperCut (9). Being open-source, DLC offers significant flexibility for customizing the code. This, along with its multi-animal tracking mode (10) , makes it the ideal tool for my project.

However, running DLC requires GPU, and an extensive environment installation process, meaning that currently the project can only be run on our lab computer. To facilitate replication, the project's google drive folder includes all necessary files to run the project on any system with DLC installed. I would be happy to invite you to the lab for a live demonstration of the code. Additionally, the folder holds example pairs of analyzed videos and their analysis result for performance assessment with the code in the project's git repository. The DLC filles could not have been added to the git due to their considerable size.

Database

The database for this project consists of video recordings collected from our experimental setup. Each mouse arena is continuously recorded by a suspended camera (i-Pro WV-S1136 and i-Pro WV-S1136A), generating consecutive 4-hour MP4 videos at 640×480 resolution, with an average file size of ~2GB per video. These recordings are later stitched together into 12-hour videos (~8GB each) to represent full day and night phases.

To ensure the network's robustness to natural variations in input, videos used throughout the project were sourced from various experiments conducted over the past two years. Due to long processing times, 4-hour video segments were primarily used, while a 10-minute excerpt from a full video was created for quicker testing during the performance assessment phase.

The labeled dataset consists of 50 randomly selected frames (saved as JPG files) from 16 different 12-hour videos, resulting in a dataset of 800 images. Due to large file sizes, only the labeled images—not the full videos—have been uploaded to the project's Git repository.

A former lab member, Leah Niv, previously attempted to apply DeepLabCut to our lab's videos last year but faced challenges due to incompatible package versions. Some of the training data was initially labeled by Leah and later reviewed by me, while the remainder was labeled by me and reviewed by other lab members. I have personally reviewed or labeled every image in the dataset.

Results

As of now, the project has yet to achieve the desired accuracy.

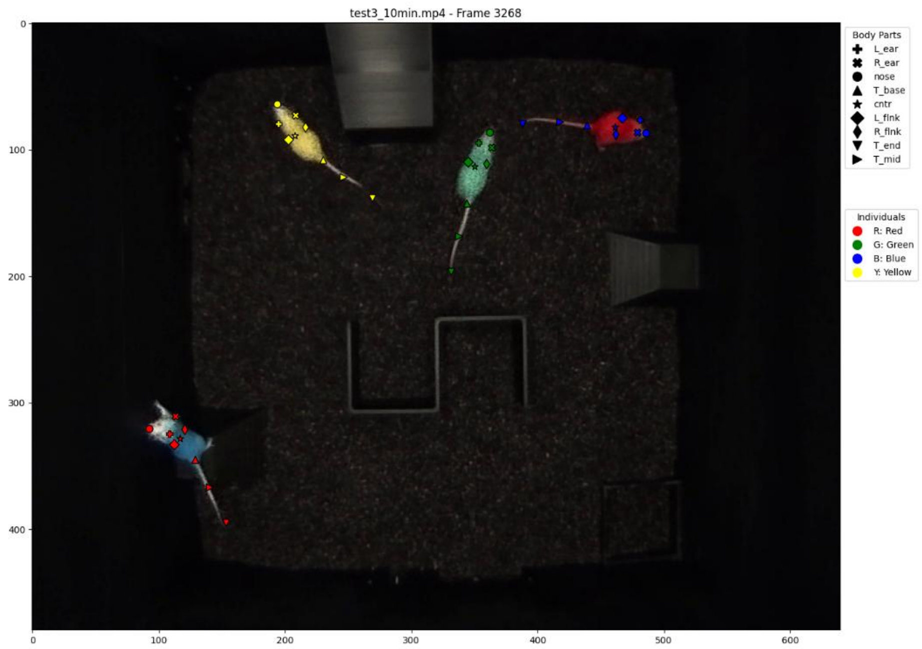
The video analysis outputs a CSV file specifying each mouse's identity, body part coordinates (X, Y), and prediction confidence. Overall, body part identification is fairly accurate. No mislabeling of body parts as other parts or labeling on a non-mouse object have been observed. Missing labels occur but are uncommon and typically coincide with other significant errors. Correct labels are at time slightly misplaced and appear adjacent to their true location, but this issue appear to very minor and could be negligible.

The primary challenge lies in accurately identifying individual mice. While the latest training version has significantly improved recognition—correctly identifying mice based on color most of the time—errors still occur. These mistakes can happen sporadically, even when a mouse remains within the camera's view. No clear pattern has emerged, and errors do not appear to be associated with specific mouse colors. Importantly, each label appears only once per frame, meaning no duplicate identities are assigned in a single image.

Types of Labeling Errors

1. Label Switching

One of the most common early issues—greatly reduced through training improvements but still prevalent —was label switching, where one mouse is entirely labeled as another. When both mice are visible, the error appears symmetrical, meaning they are effectively swapping identities.

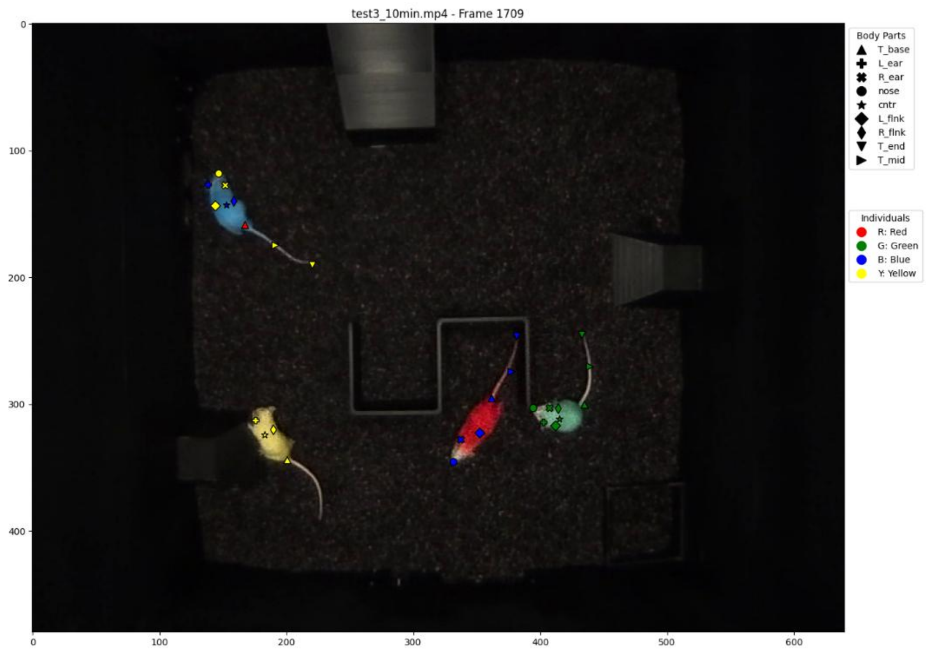


2. Label Drifting

Another issue is label drifting, which occurs when two mice are physically touching. In these cases, the labels of one mouse incorrectly mark the other, typically in the correct anatomical positions, most common for contacting body parts.

3. Label Jumping

The most concerning error is label jumping, where a few specific body parts of one mouse are labeled with another mouse's identity. Interestingly, this issue did not appear in earlier training versions and may be an unintended consequence of recent adjustments.



Discussion

From the initial training attempts, it was evident that while pose estimation was highly accurate, identity labeling remained problematic. In early versions, label switching was frequent, with mouse identities appearing almost random.

To address this, I hypothesized that preprocessing adjustments could improve identity recognition. Specifically, I modified the augmentation pipeline to reduce artificial color alterations, limiting them to mild saturation adjustments. Additionally, I minimized transformations unrelated to expected variation—such as cropping, rotation, and brightness changes. These modifications significantly reduced label switching, leading to more consistent identity tracking.

However, label drifting remained an issue, suggesting that physical proximity between mice continues to challenge identity tracking in a way unrelated to the training preprocessing. It is possible that expending the training data set to include more close contact examples could aid in improving the module. Additionally, label jumping emerged as a new problem in later versions. This error—where individual body parts of one mouse are misassigned to another—directly contradicts DLC's underlying skeleton structure. Understanding why this occurs is crucial, as DLC's built-in constraints should theoretically prevent it. I am currently collaborating with other researchers using DLC and plan to engage with the online community to explore possible causes, such as potential issues with our implementation of the multi-animal tracking module.

Despite these ongoing challenges, the model shows clear improvements. While the results are not yet suitable for research applications, continued refinements in training strategies, dataset augmentation, and model constraints will likely yield further progress.

Conclusion

DeepLabCut is an incredibly powerful tool, but its implementation is a complex process. If nothing else this project has taught me the importance of good documentation. Despite initial setbacks, substantial improvements have been made, particularly in reducing label switching. However, new challenges, such as label jumping, highlight the complexity of multi-animal tracking. Moving forward, further refinements and collaboration with the DLC community will be essential to achieving a fully reliable tracking system, which I intend to do. The potential of this program to help us expand the scope of behaviors we can track is enormous, greatly improving the research performed in our lab.

Bibliography

1. *The five-factor model of personality and its relevance to personality disorders.* **Costa, Paul T. and McCrae, Robert R.** 1992, Journal of Personality Disorders, pp. 343-359.

2. *From mice to men: what can we learn about personality from animal research?* **SD, Gosling.** 2001, Psychol Bull, pp. 45-86.

3. *Revisiting the Open-Field Test: What Does It Really Tell Us About Animal Personality?* **Perals, D., Griffin, A. S., Bartomeus, I., & Sol, D.** 2017, Animal Behaviour, pp. 69-79.

4. *Animal Personality: What Are Behavioural Ecologists Measuring?.* **Carter AJ, Feeney WE, Marshall HH, Cowlishaw G, Heinsohn R.** 2013, Biological ReviewS, pp. 465-475.

5. *Evaluating animal personalities: Do observer assessments and experimental tests measure the same thing?* **Carter, A. J., Marshall, H. H., Heinsohn, R., & Cowlishaw, G.** 2012, Behavioral Ecology and Sociobiology, pp. 77-84.

6. *Neophobia Is Not Only Avoidance: Improving Neophobia Tests by Combining Cognition and Ecology.* **Greggor, Alison L., Alex Thornton, & Nicola S. Clayton.** 2015, Current Opinion in Behavioral Sciences, pp. 82-89.

7. *Identity domains capture individual differences from across the behavioral repertoire.* **Forkosh, O., Karamihalev, S., Roeh, S. et al.** 2019, Nature Neuroscience, pp. 2023-2028.

8. *DeepLabCut: markerless pose estimation of user-defined body parts with deep learning.* **Alexander Mathis, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis , Matthias Bethge.** 2018, Nature Neuroscience, pp. 1281-1289.

9. *DeeperCut: a deeper, stronger, and faster multi-person pose estimation model.* **Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M. , Schiele, B.** 2016, European Conference on Computer Vision, pp. 34-50.

10. *Multi-animal pose estimation, identification and tracking with DeepLabCut.* **Jessy Lauer, Mu Zhou, Shaokai Ye, William Menegas, Steffen Schneider, Tanmay Nath, Mohammed Mostafizur Rahman, Valentina Di Santo, Daniel Soberanes, Guoping Feng, Venkatesh N. Murthy, George Lauder, Catherine Dulac, Mackenzie Weygandt Mathis & Alexander M.** 2022, Nature Methods, pp. 496-504.