Using Deep Learning Pose Estimation Tool to Analyse Naturalistic Behaviours in Rodents

Candidate Number: MFLG5

BSc Neuroscience

University College London

Abstract

A cognitive map is an internal representation of the world, supported at the neural level by spatially tuned cells in the hippocampal formation. However, how this system is constructed — whether it is innate or shaped by experience — remains unclear. Addressing this question requires a deeper understanding of the sensory and motor experiences of animals during early development. Yet, most rodent studies have relied on isolated testing conditions or simplified behavioural measures that may not reflect natural development. In this study, we used DeepLabCut (hereafter referred to as DLC), a deep learning-based pose estimation tool, to track rat dam and pups in a naturalistic home-cage environment from postnatal day 8 (P8) to P23. We trained a DLC model and evaluated its performance in pose estimation, animal assembly, and tracking. The model achieved high accuracy in pose estimation and generalised well to unseen data. However, the automated animal assembly and tracking pipeline in DLC did not perform well, requiring manual tuning to obtain smooth and continuous tracking data. Following technical evaluation, we used the trained model to investigate the emergence of exploratory behaviour and movement in rat pups. Exploratory behaviour was measured as the time spent outside the huddle area, while movement was defined as periods in which the pup's velocity exceeded a predefined threshold for a minimum duration. We found a marked increase in the time spent outside the huddle between P17 and P18, followed by a sharp rise in moving time between P19 and P21. In summary, our technical evaluation supports the feasibility of using DLC to study naturalistic behaviours in developing rat pups within a home-cage environment. Moreover, the results suggest distinct developmental trajectories for exploration and movement, with exploration emerging earlier than sustained movement.

Declaration of Contribution

I assisted Marco Abrate and Asra Albu-Swailim with the collection of the video recording data. I performed the training and validation of the DLC model independently, and conducted all subsequent data analysis. The custom Python scripts used for analysis were written by myself. Prof. Tom Wills provided scientific input and guidance throughout the project.

Contents

Introduction

In mammals, the ability to navigate in space is fundamental to survival, as it allows animals to locate resources, avoid danger, and move efficiently through the environment. To navigate flexibly in an environment, a cognitive map — the internal representations of the world — is necessary for learning and remembering important locations. This concept was first proposed by Edward Tolman, who suggested that somewhere in the brain there must be a representation of the environment. Decades later, John O'Keefe provided direct neural evidence for this idea with the discovery of place cells in the hippocampus, showing that the firing of these hippocampal pyramidal cells is tuned to specific locations in the environment. This led him to propose that the hippocampus provides the neural basis of cognitive mapping, which represents the allocentric (world-centred) view of space and allows for flexible navigation (O'Keefe & Nadel, 1978). In the following years, the discovery of additional classes of spatially tuned neurons, such as grid cells, head direction cells, and boundary vector cells, further demonstrated how the role of hippocampal formation in spatial cognition is supported at the neural level.

The Role of Experience in Cognitive Map Development

While the idea of the hippocampus as a cognitive map is well established, how such a complex cognitive system is constructed and what is the role of sensory experience and endogenous factors remain open questions. In John O'Keefe and Lynn Nandel's original proposal, they argued that the hippocampal neural representation of space is a Kantian *synthetic a priori* system, meaning that it does not require empirical experience for its construction (O'Keefe & Nadel, 1978). In this view, the development of a hippocampal cognitive map is innate and should be independent of experience with space. On the other hand, the network models of these spatially tuned neurons, such as the continuous attractor network of head direction cells, suggested that such networks are created by experience-dependent learning, with inputs from visual landmarks and the vestibular system (Hahnloser, 2003).

To explore this idea, research has been conducted on the emergence and maturation of spatially tuned cells in laboratory rats, which remain in the nest for the first two weeks of life, and begin to explore the space during the third weeks of life (Gerrish & Alberts, 1996).

These studies have shown that different classes of spatially tuned neurons emerge at different ages, mature at different rates, and are all present by P21 (Figure 1). Comparing their emergence and maturation timeline with the onset of active exploration offers some insights into the role of experience in the development of a cognitive map. For example, head direction (HD) cells, which encode the animal's current allocentric head direction, are the first to emerge and mature. Adult-like HD responses can be recorded from the dorsal presubiculum as early as P14 (Tan et al., 2010). Since head direction responses emerge before active exploration, their construction might not depend on experience with space. In contrast, grid cells, which fire in multiple locations arranged in a hexagonally symmetrical grid covering the environment, do not emerge until P20 but mature rapidly to adult levels (Wills et al., 2010). The late emergence of grid cells implies that their normal development could rely on experience with the environment. On the other hand, place cells emerge as early as P16 but initially exhibit high variability in their place responses and undergo gradual maturation over time (Wills et al., 2010). While a subset of place cells displays adult-like spatial stability and specificity at P16, most remain immature, requiring over two weeks to reach maturity (Scott et al., 2010). Thus, while their early emergence suggests that place cell development could be innate, the full maturation of the place cell network may depend on spatial experience.

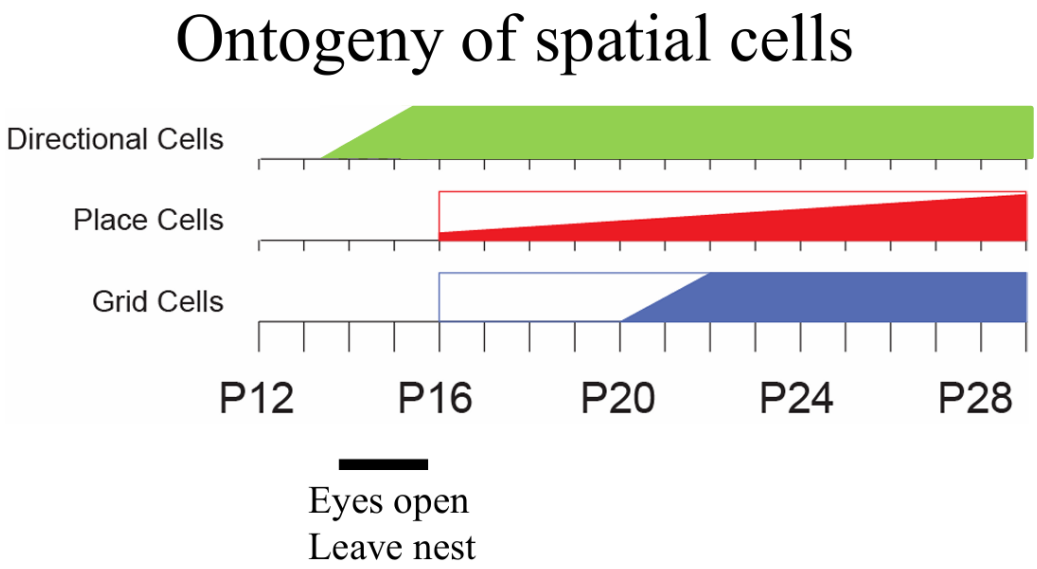

Figure 1. The emergence and development of head-direction, place, and grid cells. (O'Keefe et al., 2014)

A recent study provided direct evidence for the role of experience in the development of grid cells. Ulsaker-Janke et al. (2023) demonstrated that in rats deprived of geometric experience during development—raised in opaque spherical environments and darkness—grid-like firing patterns were initially absent when they explored an open square environment. However, after seven days of training, these patterns gradually emerged and developed into a clear grid-like structure. This suggests that the formation of regular grid-like firing requires exploration of space with light and well-defined geometric boundaries (Ulsaker-Janke et al., 2023).

However, whether sensory experience is necessary for the development of head direction cells and place cells remains unclear. While their early emergence suggests an innate basis, as predicted by the cognitive map theory, testing this idea is challenging due to the difficulty of completely ruling out the influence of experience. For instance, Wills et al. (2010) found no correlation between experience in the testing environment and the spatial firing of place cells and head direction cells, supporting the idea of an innate cognitive map. However, this does not exclude the possibility that experience in the home cage contributes to their development. Similarly, Bassett et al. (2018) showed that the continuous attractor network underlying head direction cells is already established before eye opening in rat pups, indicating that visual input is not required. Yet, it remains difficult to rule out the role of other sensorimotor experiences, such as vestibular input from head-turning movements and other sensory inputs such as whiskers.

Thus, experience can take many forms and influence the development of the hippocampal cognitive map in multiple ways. Beyond active exploration outside the nest, early sensorimotor activities such as crawling, turning, and diving motions may be crucial for establishing vestibular and motor inputs into the head direction system (Taube, 2007). Additionally, interactions with the rat dam and littermates may also shape hippocampal development and, in turn, impact the formation of the cognitive map in rat pups. For example, Liu et al. (2000) reported that the variations in maternal care in rat led to difference in spatial learning, memory, and hippocampal synaptogenesis. Specifically, offspring of mothers that showed higher levels of licking, grooming, and nursing performed better in the Morris water maze test and showed enhanced expression of NMDA receptor subunits, brain-derived neurotrophic factor (BDNF) mRNA, and increased cholinergic innervation in the hippocampus (Liu et al. 2000).

Therefore, assessing the "innate cognitive map" hypothesis requires more than studying the development of spatially tuned neurons in isolation. Instead, it is essential to consider their development within the broader context of sensory and spatial experiences to achieve a more integrated understanding of the emergence of hippocampal function and spatial cognition (Wills et al., 2013). However, our current understanding of these experiences remains limited, as most studies have been conducted in non-naturalistic settings that may not accurately reflect normal development. This report reviews existing research on the development of sensory-motor systems and spatial behaviours in rats, and discusses the limitations in those studies.

The Development of Sensory-Motor Systems in Rats

Rats are altricial animals, born with limited sensory perception and motor abilities, similar to humans. The development of their sensory systems follows the general mammalian blueprint of sensory ontogenesis, with vestibular and olfactory functions emerging first, followed by tactile, auditory, and visual functions (Alberts, 1984). At birth, rats exhibit a rudimentary righting reflex, reflecting the emergence of the vestibular system (Altman & Sudarshan, 1975). Similarly, rat pups show a preference for their mother's odour by P2, signalling the emergence of olfaction (Polan & Hofer, 1998). Tactile sensation develops around P4, when passive whisker movement can elicit activity in rat pups, and whisker clipping at P4–P5 disrupts suckling and huddling behaviours (Sullivan et al., 2003). The auditory system begins functioning around P8, as indicated by observable hair cell stimulation (Uziel et al., 1981). Vision, a crucial sense for navigation, is the last to develop with eyes opening happening between P13 and P15 (Altman & Sudarshan, 1975).

The development of motor skills in isolated rat pups was comprehensively described by Altman & Sudarshan (1975), who assessed pups individually in an open field every day from P1 to P21. Their findings showed that pup movement progressed through distinct stages. Between P3 and P7, the primary form of movement was pivoting, where pups turned on the spot using their forelimbs while their hindlimbs remained immobile. From P7 onward, they transitioned to crawling, a translational movement with still limited hindlimb involvement. Starting from P14, pups became capable of quadrupedal walking, with coordinated use of hindlimbs. By P21, they were capable of performing complex motor skills, marking the maturation of their locomotor abilities. However, it is important to note that motor skills

assessed in isolation may differ from natural behaviours observed in the home cage, where dam and littermates are present (Wills et al., 2013). This highlights the need to study motor development in a more naturalistic setting.

The Development of Spatial Behaviour in Rats

Beyond sensory-motor development, understanding the natural emergence of spatial behaviour is essential. Studies in adult rats have shown that animals with an intact hippocampus engage in systematic exploratory behaviours when exposed to a novel environment and exhibit habituation upon repeated exposure (Berlyne, 1996). In contrast, hippocampal-lesioned rats display hyperactivity, reduced systematic exploration, and impaired habituation (O'Keefe & Nadel, 1978). Therefore, pinpointing the emergence of spontaneous exploration may provide key insights into the development of hippocampal functions.

The study by Bronstein et al. (1974) examined activity levels and behaviours in isolated rat pups of different ages by placing them in an open field for prolonged trials. They found that P15 pups exhibited significantly higher activity levels and less habituation than P21 pups, a behavioural pattern resembling that of hippocampal-lesioned adult rats. However, this difference is likely driven by the anxiety response to social isolation rather than a reflection of spontaneous exploration. Supporting this, Randall & Campbell (1976) found that the heightened activity levels of P15 pups diminished when anesthetized littermates were present. Further evidence came from studies on ultrasonic distress vocalizations, which decrease significantly from P15 to P20 (Hofer & Shair, 1978). These findings suggest that testing pup behaviour in isolation may introduce stress-related confounds, highlighting the importance of using more naturalistic settings. Another study by Nadel et al. (1992) investigated the development of motor activity and active exploration by placing single rat pups in an open field containing several objects, including a running wheel. A key finding was that the onset of active exploration occurred abruptly, in an all-or-none fashion. However, because the pups were tested in isolation, further research is needed in more naturalistic settings to validate these results. Gerrish and Alberts (1996) examined exploratory behaviour in a more naturalistic setting. In this study, a rat nest was connected to an open field accessible to both the dam and pups. The researchers found that pups spent almost no time in the open field at P14. Between P15 and P17, they spent less than 10% of their time there,

followed by a sudden increase to approximately 30% at P18. However, their measurement of exploratory behaviour was restricted to the amount of time spent in the open field, which provided limited insight into other kinds of pups' behaviours.

Collectively, although these studies have provided valuable insights, our current understanding remains limited, as most studies have relied on isolated testing conditions or simplified measures that may not fully capture the sensory and spatial experience of rat pups in normal development. To overcome these limitations, we explored the use of DeepLabCut, a deep learning-based pose estimation tool, to quantify the natural emergence of pup movement and spatial behaviour in the home cage.

Studying Natural Emergence of Movement and Exploratory Behaviour with DeepLabCut Toolbox

DeepLabCut (DLC) is a deep learning pose estimation tool designed to track animal movement and posture (Mathis et al., 2018). Using transfer learning, DLC fine-tunes pretrained deep neural networks to recognize self-defined body parts with minimal training data. This enables high-precision pose estimation that generalises across species and experimental conditions (Mathis et al., 2018). Extended from this framework, multi-animal DLC further incorporates identity-preserving tracking, allowing the tool to distinguish and follow individual animals in dynamic environments (Lauer et al., 2022).

In recent years, the DLC have been widely adopted for studying laboratory rodents as an alternative to manual annotation, which is time-consuming and labour-intensive. DLC has been used in various behavioural studies, including research on fear expression (Chanthongdee et al., 2024), alcohol intoxication (Zahran et al., 2024), and ketamine-induced behaviours (Popik et al., 2024). Notably, DLC has also been used to analyse naturalistic behaviours. For instance, Lapp et al. (2023) developed a pipeline using DLC to track key points on rat dam and pups from P1 to P10, assisting the analysis of mum-pup interactions in a naturalistic setting. This allowed them to study how early social experiences influence neural development. Given its potential, DLC holds great promise for facilitating a more detailed examination of the development of pup movement and exploratory behaviour in naturalistic contexts. However, several challenges remain, including the

dramatic morphological changes pups undergo during development, as well as issues related to occlusion and individual re-identification.

Therefore, the overall aim of this study is to train a DeepLabCut model to track rat pups in a naturalistic home cage, evaluate the performance of the model, and use the tracking data generated by the model to study the emergence of pup movement and exploratory behaviour. The following sections detail the model training, evaluation, and downstream analyses.

Material and Methods

Animals

All animal protocols were approved by the Home Office, and all scientific procedures of living animals were under the ANIMALS (SCIENTIFIC PROCEDURES) ACT 1986. The Lister Hooded rats used for the experiment were bred in-house, and rat dams were individually housed throughout pregnancy. The date of birth of the litter was considered as P0. We controlled for the litter size by reducing the pup number to 8 pups. Animals were housed in a conventional cage with size 45cm in width, 60cm in length, and 20cm in height. The temperature in the BSU room was held between 19 and 23 degrees, and the humidity between 55 RH +/- 10. The light in the room was kept on a 12:12h cycle, with dawn to dusk setting, so at 10:00 am the light gradually turned off, and at 9:30 pm the light gradually turned on. Dam was enriched with a plastic shelter and a wooden chewing bar. Chip wood pieces and tissue paper were provided as bedding material. All enrichments were glued to the to make sure they could not be moved by the animals and to ease background recognition by the DeepLabCut model (Figure 2). Animals had access to food and water *ad libitum*. Weaning took place at P21, where the rat dam was taken away and the litter mates remained housed in the home cage.

Figure 2. The home-cage environment with the enrichments.

Video Recording

The video recording setup consisted of a Raspberry Pi microcomputer and an infrared camera mounted on a custom-built overhead recording rig, positioned directly above the home cage. This allowed us to capture top-down video of the animals in the cage. Moreover, infrared lights were used for illumination during dark conditions. Specifically, two pairs of IR lights were mounted symmetrically on the longer side of the recording rig, and one IR light on each shorter side, providing uniform coverage of the home cage (Figure 3). With this setting, we were able to record rat behaviours in a naturalistic cage environment in both light and dark conditions with minimal intervention.

The video data was recorded at 14 frames per second (fps) at a resolution of 1920 x 1080 pixels. Video recordings started on P8 and ended on P23. 10 minutes videos were recorded at 10 minutes past every hour for 16 days, with a fixed interval of 1 hour between recordings' starts. The recording scheduled to begin at 10:10 am was skipped because the laboratory lights were gradually turning off during that time. Thus, 23 videos were recorded per day, for a total of 368 videos. Videos were uploaded to the group folder on the UCL cloud storage and converted from H264 to Mp4 format.

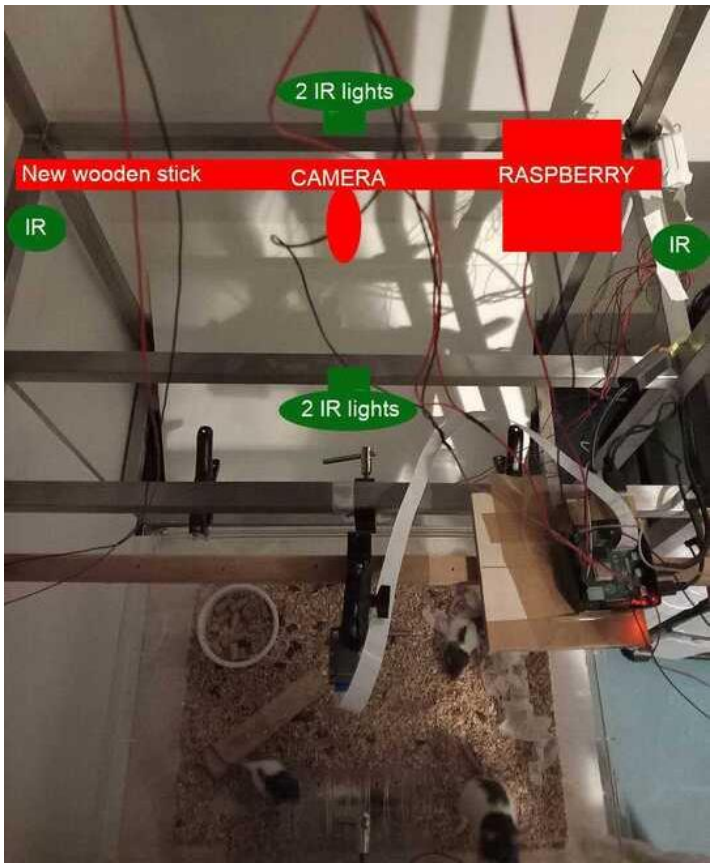

Figure 3. The video recording set up. Infrared lights were mounted above the home cage to provide uniform top-down illumination for the camera with minimal disturbance, as rats cannot see infrared light (Muntz, 1967).

Multi-Animal Pose Estimation Model

Overview of Multi-Animal DeepLabCut and Workflow

Multi-Animal DeepLabCut is a deep learning-based pose estimation tool designed for tracking multiple animals in video recordings. It employs a deep neural network to localise user-defined body parts in each frame, then groups these key points into individual animals (assembly) and tracks them across frames (tracking). The standard DLC workflow consists of the following steps: project creation, frame labelling, model training, pose estimation, animal assembly, and tracking. The output of DLC is tracking data containing the coordinates of all key points across frames, enabling downstream analysis.

Project Creation

Pose estimation data for the rat dam and eight pups in the home cage was generated using Multi-Animal DeepLabCut, version 2.3.9 (Lauer et al., 2022). The installation and setup followed the user guide available on the official website (https://github.com/DeepLabCut). Only videos recorded during the dark condition brightened by IR lights were used to train the model, because rats tend to be more active during this period, making it more suitable for capturing the emergence of movement and social interactions.

Frame Labelling

A total of 800 frames were extracted from 41 dark condition videos (around 20 frames per video) recorded from P8 to P23. This was done automatically by DLC using k-means algorithm, which clustered videos based on visual appearance in order to capture a broad range of scenarios. Next, the extracted frames were labelled using the interactive graphic user interface (GUI) of DLC.

For the rat dam, seven body parts were labelled (snout, left ear, right ear, shoulder, spine1, spine2, and tail base) to provide sufficient coverage and reduce the impact of occlusion. These were defined as "unique body parts" in the configuration file because a single dam was present at all times. For the rat pups, five body parts were labelled (snout, left ear, right ear, shoulder, tail base). We have used these five body parts systematically even across the dramatical morphological stages that pups go through. These body parts were defined as "multi-animal body parts" in the configuration file. To ensure consistency and accuracy during manual labelling, a set of landmarking criteria was followed for each body part (Figure 4):

Snout: Marked at the most anterior visible point of the head.

Left and right ear: Labelled at the base of each ear, and the side was determined based on the animal's orientation in the frame.

Shoulder: Identified as the junction between the head and the main body, typically visible as a transition in fur colour.

Spine1 and Spine2 (dam only): Placed along the midline of the back, equally spaced between shoulder and tail base.

Tail base: Marked at the junction between the tail and the body, aligned with the spinal midline.

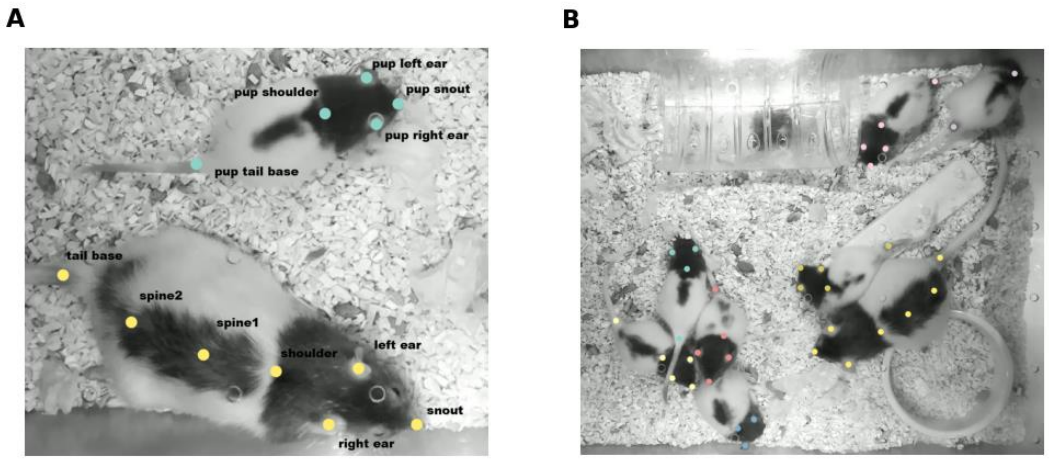

Figure 4. (A) Example of labelled body parts for a rat dam and a pup. **(B)** Example of a labelled frame. Only visible animals and body parts were labelled.

It is important to note that body parts were labelled only when they were clearly visible in the frame, as this improves accuracy. However, we also noticed that this might give the model less confidence in predicting joints when they are partially visible. Moreover, only pups that were visible in a given frame were labelled, i.e. pups playing in the plastic tube were skipped. Finally, since the eight pups were considered to be visually indistinguishable, the identify label (e.g. pup1, pup2…) wasn't kept consistent between frames, and the parameter "identity" was set to False in the configuration file. All the labelled frames were used as ground truth for the training and validation of the maDLC model.

Model Training

The TensorFlow engine was used for DLC model training. The model architecture was based on ResNet-50 with ImageNet pre-trained weights. The training fraction was set to 0.90, meaning that 90% of the labelled frames were used to train the model and the remaining 10% were used for evaluation. The model was trained for 100,000 iterations using a batch size of 8, saving a snapshot every 10,000 iterations. Model training was performed on a Lenovo Legion Y9000P with an Intel Core i9 14900HX, 2.2GHz processor, Win 11, and an NVIDIA GeForce RTX 4060 Laptop GPU.

Pose Estimation

Through training, the model learned to perform pose estimation, i.e., predicting the locations of key points from video frames. The performance of the model in pose estimation was evaluated through two metrics — loss and root-mean-square error (hereafter referred to as error). Loss was defined as the sum of the cross-entropy loss between target score maps and predicted score maps, the local refinement loss, and the limb length loss (Mathis et al., 2018). It was calculated every 1,000 iterations during training. In general, loss quantifies the error between the model predictions and the ground truth values. It typically decreases over time and converges as the model learns to make predictions more accurately. The loss values were found in the "train" folder of the project. The second metric, error, measures the Euclidean distance (in cm) between a manual label (ground truth) and the corresponding model prediction. The errors for all key points were computed using the "*evaluate_network*" function in DLC. Here, the median error was calculated at each snapshot (every 10,000 iterations), separately for the train and the test frames. The evaluation results are presented in the Results section.

Animal Assembly and Tracking

With estimated key points, animal assembly and tracking were performed to generate trajectories for each animal across frames. In general, DLC provides two approaches for animal assembly and tracking. The first approach uses the *auto_track = True* option in *analyze_video* function, which automates the process using default parameters optimised for general cases. The second approach (*auto_track = False*) performs animal assembly and tracking in a stepwise manner, requiring users to run each step of the pipeline separately while allowing greater manual control (users can iteratively change the parameters). Specifically, two processing steps were involved: (1) converting detections into tracklets (short trajectory segments), and (2) stitching these tracklets into continuous trajectories. These steps were performed using the DLC functions *convert_detections2tracklets* and *stitch_tracklets*. A full description of the parameters used at each step is provided in the appendices.

To compare the performance of the two tracking approaches of DLC, we applied both methods to the same set of videos and evaluated their performance in generating reliable trajectories. Unlike pose estimation, there was no ground truth tracking data, as it was not feasible to manually annotate every frame in a video (~8400 frames in a 10-minute

recording). Therefore, the trajectories were evaluated through visual inspection and by quantifying the occurrence of sudden jumps in positions. The evaluation results are presented in the Results section.

Once the trajectories were confirmed to be reliable based on the above evaluations, the final step of the DLC tracking pipeline was performed to generate the output .csv files. Each .csv file contained the x and y coordinates and the associated confidence scores (ranging from 0 to 1) for all body parts across all frames in a video. These files served as the input for the subsequent analyses of pup movement and exploratory behaviour.

Post-Tracking Analysis

The tracking data of rat pups generated by DLC were used for two analyses: (1) the emergence of exploratory behaviour, measured by the time spent outside the huddle, and (2) the emergence of movement, quantified by the time spent moving at a velocity exceeding predefined speed and duration thresholds. Animal trajectories were smoothed using a rolling average filter with a window of 5 frames. Velocity was computed as the distance between consecutive frames divided by the frame duration (1/14 s). To avoid artefacts from tracking errors, frames where velocity exceeded 100cm/s were excluded. All analyses were performed using custom Python scripts, which are provided in the appendices.

Statistics

Statistical analyses were performed using Python (version 3.12), primarily with the SciPy and NumPy packages for statistical tests, and Matplotlib and Seaborn for data visualization. The normality of the data was assessed using the Shapiro–Wilk test, and the equality of variances among multiple groups was evaluated using Levene's test. If the data met the assumptions of normality and homogeneity of variances, parametric tests were applied, including one-way ANOVA for group comparisons, followed by pairwise t-tests where appropriate. Otherwise, the Kruskal–Wallis test was used for multiple group comparisons, followed by pairwise Mann–Whitney U tests when appropriate. Statistical significance was defined as $p < 0.05$.

Results

The Results section is divided into two parts. In the first part, we focus on the technical evaluation of the DLC model's performance in tracking rat dam and pups in a naturalistic home-cage environment. In the second part, we apply the model to study the naturalistic behaviours of rat pups. Specifically, we investigate the emergence of exploratory behaviour and movement across postnatal days.

Pose Estimation Performance

As mentioned in the Method section, we trained a DLC model using 800 manually labelled frames extracted from 41 dark condition videos from P8 to P23. The model was trained for 100,000 iterations, and its performance in pose estimation was evaluated through loss and errors (i.e. the Euclidean distance between each manual label and model prediction). Loss was calculated every 1,000 iterations. As the number of training iterations increased, the performance of the model improved and eventually plateaued, as reflected by the decrease and convergence of loss (Figure 5A). In addition, the median error was calculated every 10,000 iterations separately for the train and test frames (Figure 5B). Starting from 30,000 iterations, the median training error converged at around 0.15cm, and the median test error converged at around 0.23 cm. This showed that the model not only fitted the training data well but also generalised effectively to unseen data. Taken together, these results demonstrated that our model achieved strong performance in pose estimation for rat dam and pups in home cage.

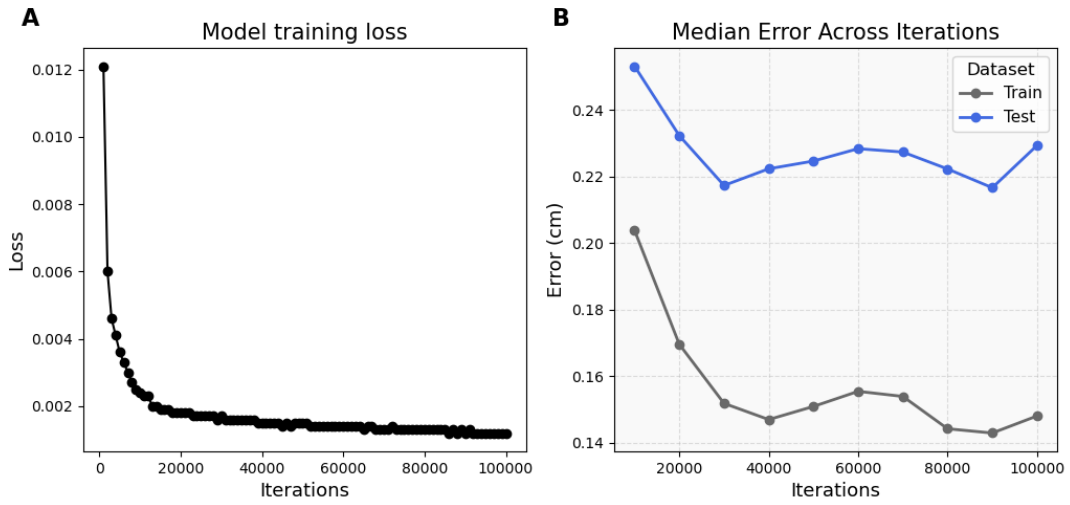

Figure 5. (A) Model training loss decreased with each training iteration. **(B)** The median error (in cm) converged for both the training and test sets.

Effect of Training Set Size on Pose Estimation Performance

We showed that a model trained using 800 labelled frames (720 for training and 80 for evaluation) performed well in pose estimation. However, frame labelling is time-consuming and labour-intensive — it took over 50 hours to label 800 frames. According to the DeepLabCut User Guide, the number of labelled frames required to train a high-performance model depends on many factors, including the complexity of the behaviour, the experimental context, and the quality of the video. Indeed, previous studies using DLC have reported a wide range of labelled frames, from 200 (Popik et al., 2024) to 1800 (Lapp et al. 2023). Therefore, it is important to determine how many labelled frames are sufficient for training a high-performance model for our study.

To address this, we trained and compared three models using different numbers of labelled frames: 200, 400, and 800. All frames were randomly selected subsets from the original pool of 800 frames, using a custom Python script that randomly sampled frame indices. The number of training frames used for each model was 120, 320, and 720, respectively (i.e., 80 frames were always held out for evaluation). All models were trained for 50,000 iterations, as prior analysis showed that both the loss and the median error had converged by this point. The errors on test frames for each model are shown as boxplots in Figure 6. Kruskal-Wallis test revealed a significant difference across the groups ($H = 53.44$, $p < 0.001$). To further explore these differences, pairwise Mann–Whitney U tests were conducted for each pair of models. The results showed significant differences between the 200-frames model and both the 400-frames and 800-frames models (200 vs 400: $p < 0.001$; 200 vs 800: $p < 0.001$), while no significant difference was found between the 400-frame and 800-frame models ($p = 0.0961$). For visual clarity, extreme outliers were excluded in Figure 6 using an IQR-based filter.

These results suggested that increasing the number of labelled frames from 200 to 400 significantly improved the performance of the model in pose estimation. However, further increasing the number of frames to 800 made no difference. Therefore, using 400 labelled frames appears sufficient for training a high-performance model for this study.

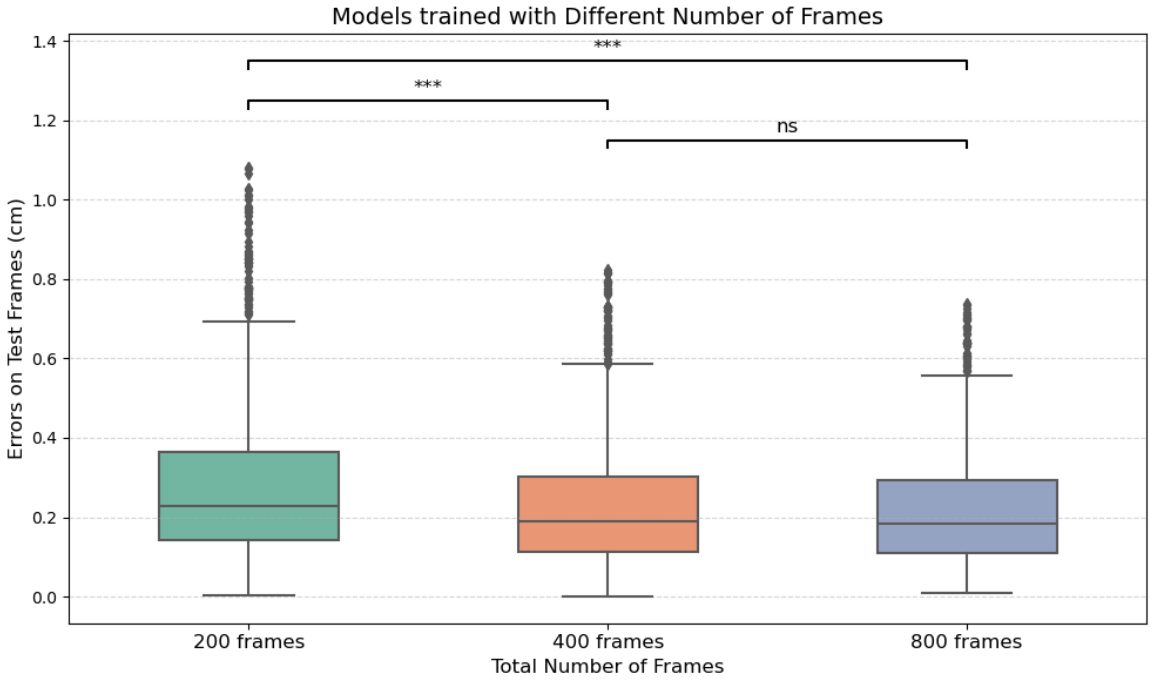

Figure 6. Comparison of test errors across models using 200, 400, and 800 labelled frames. To enhance visual clarity, outliers beyond Q3 + 1.5×IQR or below Q1 – 1.5×IQR were excluded. Statistical comparisons were conducted using the Kruskal–Wallis test followed by pairwise Mann–Whitney U tests. Asterisks indicate significant differences (*** = $p < 0.001$; ns = not significant).

Pose Estimation Error Across Body Parts

Next, we further examined the model performance in pose estimation across different key points. From our observations, some body parts, such as the snout and tail base, exhibited greater visual variability and were more frequently occluded. Therefore, we expected that the model's pose estimation performance might differ across body parts. To explore this, we evaluated the 800 frames model and plotted the errors for each body part (left ear, right ear, shoulder, snout, and tail base) on both the train and test frames (Figure 7). The median training errors were similar across body parts, ranging from 0.13 cm to 0.17 cm, suggesting that the model fit the training data equally well. However, the test errors showed greater variability, particularly for the tail base. While the left ear, right ear, shoulder, and snout all

had median test errors below 0.23 cm, the tail base exhibited a notably higher median test error of 0.46 cm. This indicated that the model had greater difficulty generalising to unseen data for the tail base. The higher error observed for the tail base is likely due to its greater movement variability and increased likelihood of occlusion during natural behaviours. In addition, the less consistent visual appearance of the tail base may also affect the model's pose estimation performance.

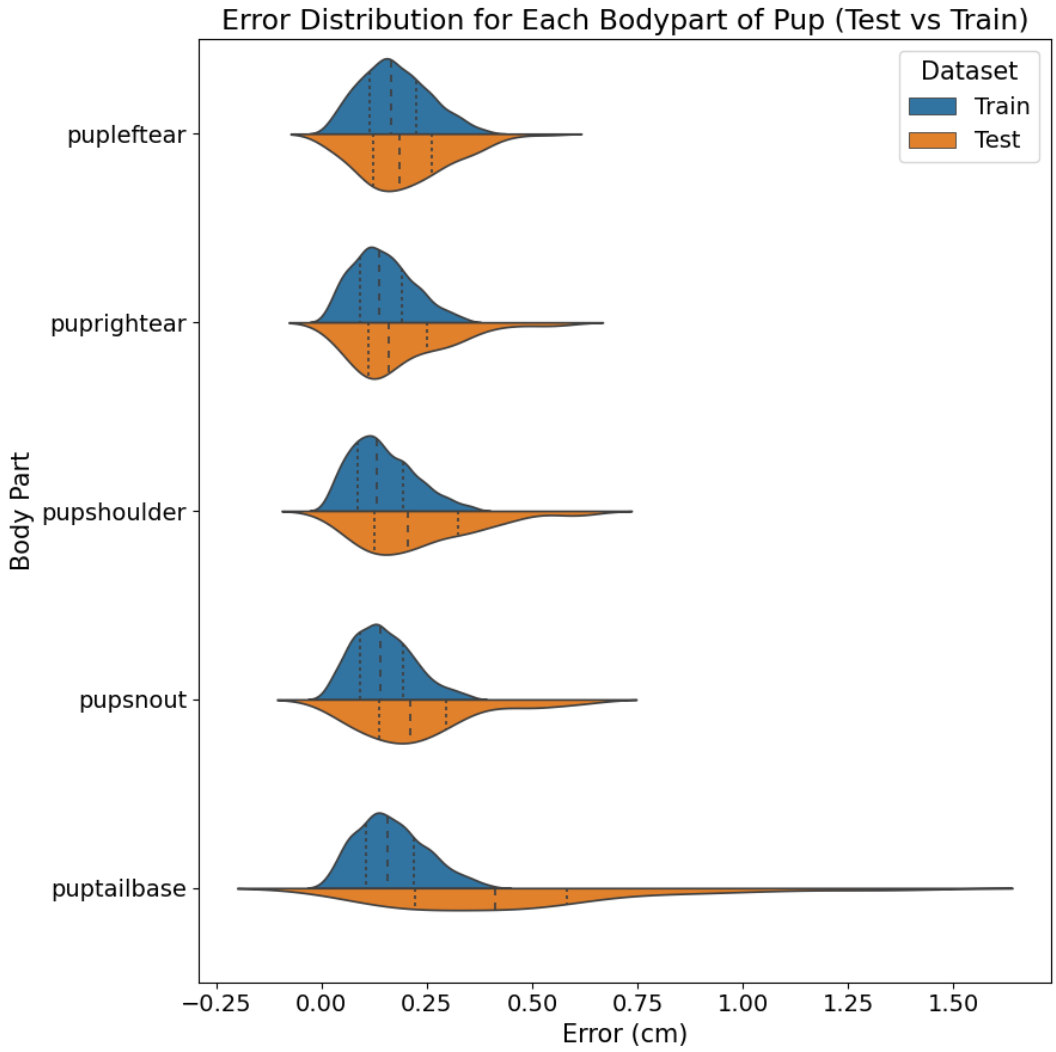

Figure 7. Violin plots showing the distribution of pose estimation errors (in cm) for each key point on the train (top) and test (bottom) frames. Vertical dotted lines indicate the first quartile, median, and third quartile of the error distribution.

Comparison of Tracking Performance between Automatic and Manual Approaches

Next, we evaluated the model performance in animal assembly and tracking. In this step, the estimated key points were assembled into animals, and trajectories were generated for each animal across frames. As detailed in the Methods section, DLC provides two approaches for animal assembly and tracking: an automatic approach and a manual approach. The automatic approach substantially reduces processing time, while the manual approach allows more control over tracking parameters. To compare their performance, we applied both approaches to the same set of videos (four randomly selected videos) and evaluated the generated trajectories through visual inspection and by measuring the occurrence of sudden jumps in position. One example of the trajectories generated by the two approaches is shown in Figure 8. We found that the automatic approach did not perform animal assembly and tracking well: the resulting trajectories were often discontinuous, and the plots of x and y positions exhibited numerous artefactual "spikes", corresponding to sudden jumps in position (Figure 8B). In contrast, the manual approach generated substantially higher-quality trajectories, with very few such "jumps" (Figure 8D).

This difference was likely caused by differences in key tracking parameters. For example, in the step that converts body point detections into short segments of trajectories, a key parameter is *max_age* — maximum duration of a lost trajectories before it's considered a "new animal". For the automatic approach, *max_age* was by default 0, meaning that even a one-frame disappearance would result in the reappearance being assigned to a different animal. This contributed to the sudden jumps observed in the trajectories. For the manual approach, *max_age* was set to 100, which allowed for temporary occlusion without splitting the trajectory.

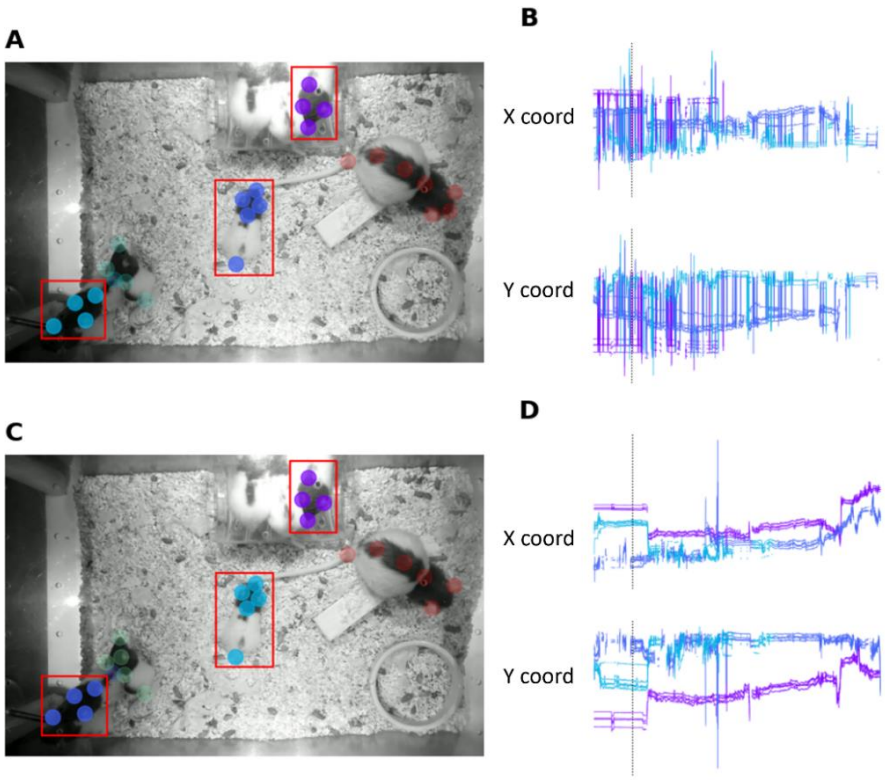

Figure 8. Example trajectories of three rat pups in the same video generated by the automatic approach (top) and the manual approach (bottom). **(A)** An example frame from the video. The rat pups are circled in red, and each coloured dot represents one body part of a pup. **(B)** Trajectories of the body parts generated by the automatic approach. The upper and lower plots show the x- and y-coordinates of the body parts, respectively. Each body part has one trajectory in the x-coordinate plot and one in the y-coordinate plot, spanning 8400 frames (corresponding to a 10-minute video). The vertical dotted line marks the frame shown in (A) in the video timeline. **(C)** The same frame, with the same rat pups circled. **(D)** Trajectories generated by the manual approach with key parameters tuned.

To further compare the trajectories generated by the two approaches, we quantified the occurrence of large discontinuities ("jumps") in the trajectories. A "jump" was defined as a sudden movement of a key point exceeding 30 cm between two consecutive frames. For each key point, we computed the distance between neighbouring frames and calculated the proportion of "jumps" in the trajectories. The result is shown in Figure 9. The Mann–Whitney U test revealed a significant difference in the proportion of "jumps" between the trajectories generated by the two approaches ($p < 0.001$). These results suggest that the manual approach provided better tracking performance with fewer artefactual "jumps".

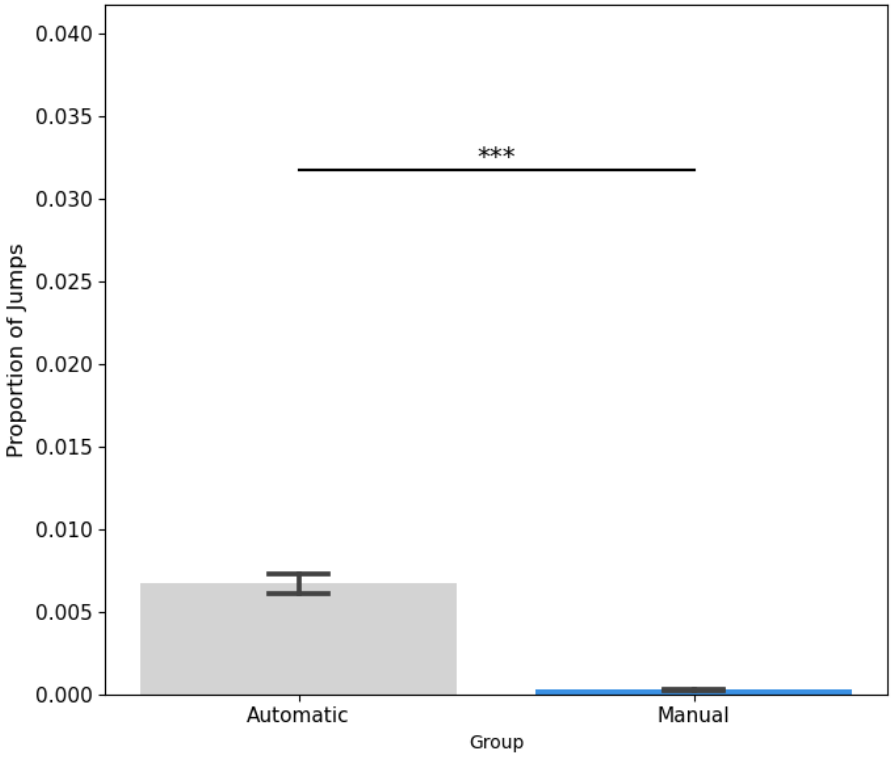

Figure 9. Comparison of the proportions of "jumps" in trajectories generated by the automatic and manual approaches. Each bar represents the proportion of "jumps" in key point trajectories under the two conditions. The significance bar indicates the result of the Mann–Whitney U test.

The Development of Exploratory Behaviour in Rat Pups

Following technical evaluation of DLC, we used the trained model to assess the emergence of exploratory behaviour in rat pups. Specifically, we examined the average time rat pups spent outside the huddle (the cluster where pups aggregated together), as this reflected the pups' tendency to leave the home area and engage in spatial exploration. Videos from P9 to P19 were used for analysis, as these were the days when a clearly identifiable huddle was present. Videos from later days were excluded because the huddle was no longer visually distinguishable. Tracking data generated by DLC was processed by computing each animal's trajectory based on the average position of the left and right ears. These two key points were used because they are more stable and less likely to be occluded during movement, providing a reliable estimate of the animal's trajectory.

Using a custom Python script, we manually defined the huddle area for each video and calculated the average time rat pups spent outside the huddle on each postnatal day (Figure 10). We found a marked increase in the time spent outside the huddle between P17 and P18. A 1-way ANOVA revealed a significant difference in exploration time across postnatal days ($F = 7.59$, $p < 0.001$). To further examine these differences, pairwise t-tests were performed between adjacent days. The results showed a significant difference between P17 and P18 ($p = 0.006$), but not for other days. Taken together, these results showed a clear developmental shift in exploratory behaviour. Although exploration outside the huddle started as early as P9 in some pups, the average outside-huddle time remained low and stable from P9 to P17. However, between P17 and P18, there was a sudden increase in exploration time. This result is consistent with the findings from Gerrish & Alberts (1996) and Nadel et al. (1992), where the abrupt onset of exploration was also found at P18. Such a behavioural transition may signal specific neural changes in the underlying hippocampal circuitry, aligning with its emerging role in spatial navigation and exploration.

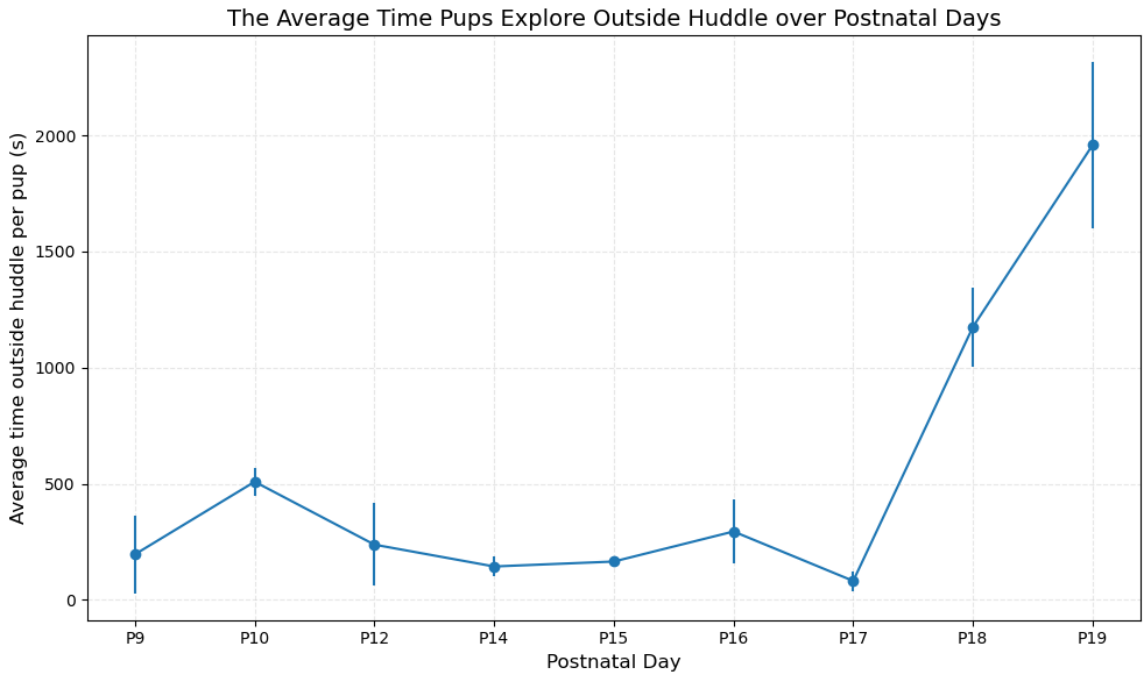

Figure 10. Average time pups spent outside the huddle across postnatal days (P9–P19). Error bars indicate the standard error of the mean (SEM). Both the Shapiro-Wilk test and Levene's test yielded p-values greater than 0.05.

The Development of Movement in Rat Pups

Next, we applied the trained model to investigate the emergence of movement in rat pups from P10 to P23. Using a custom Python script, we analysed the tracking data by computing each animal's trajectory and velocity across all frames. Movement was defined as periods where the animal's velocity exceeded a certain threshold for at least a minimum duration. For each animal, we calculated the proportion of time spent moving relative to the total video duration.

Since the definition of movement involved both a velocity threshold and a duration threshold, we tested the effect of each parameter separately. First, we held the velocity threshold constant at 3 cm/s and varied the duration threshold from 6 frames (0.43 s) to 14 frames (1 s) (Figure 11). Then, we held the duration threshold constant at 10 frames (0.71 s) and varied the velocity threshold from 2 cm/s to 4 cm/s (Figure 12). Across both analyses, we found that while the absolute amount of movement varied depending on the threshold, the overall developmental pattern remained consistent. In all cases, there was a significant increase in movement between P19 and P21. This suggests that the onset of sustained movement occurs around the time of weaning at P21. These findings are consistent with previous studies on motor skill development in rats, where P21 represents a key developmental milestone characterized by mature motor capabilities (Altman & Sudarshan, 1975).

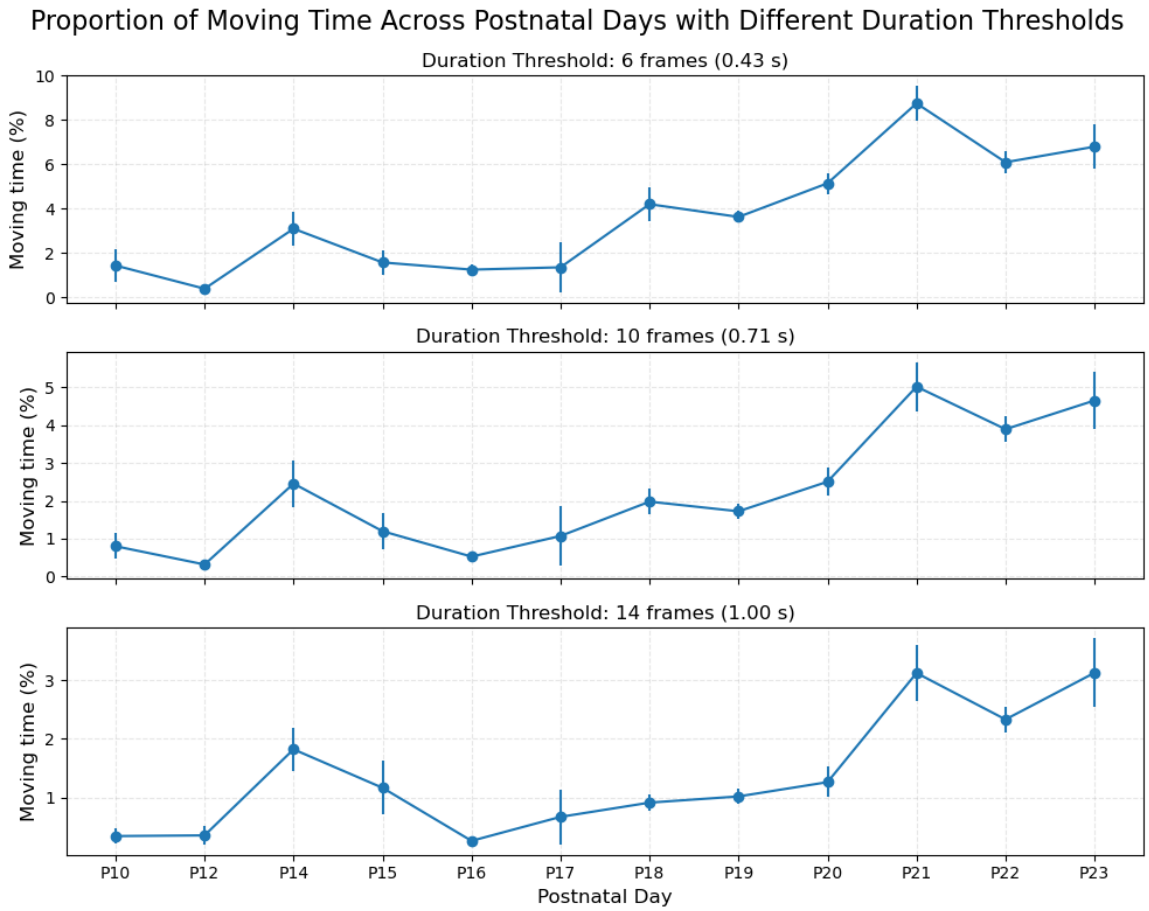

Figure 11. Proportion of moving time across postnatal days (P10–P23) under different duration thresholds (0.43–1.00 s). Velocity threshold is held constant at 3cm/s. Error bars represent the standard error of the mean (SEM).

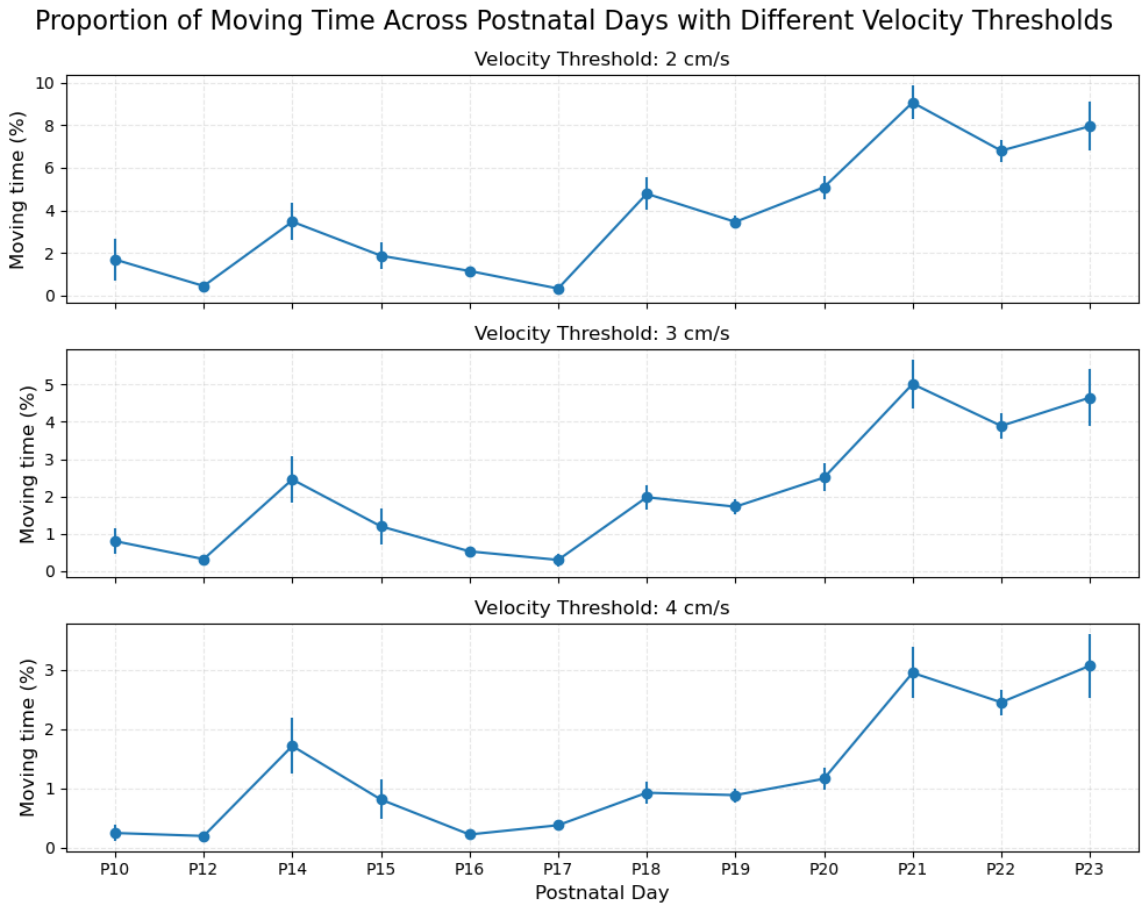

Figure 12. Proportion of moving time across postnatal days (P10–P23) under different velocity thresholds (2-4 cm/s). Duration threshold is held constant at 10 frames (0.71s). Error bars represent the standard error of the mean (SEM).

Discussion

Overall, this study demonstrates that multi-animal DeepLabCut, a deep learning pose estimation tool, is a feasible approach for tracking rat dam and pups in a naturalistic home-cage environment. The tracking data generated by DLC can be used for downstream analysis of pup movement and behaviour, which may contribute to a better understanding of the development of hippocampal cognitive map.

Despite the potential of DLC, there are several key limitations associated with its use in this context. First, during the animal assembly and tracking stage of the pipeline, the automated

function in DLC often fails to generate reliable trajectories, particularly in complex environments. As a result, manual tuning is required, which can be time-consuming and labour-intensive. Second, re-identification — the process of consistently assigning the same identity to an individual animal across all frames in the video — remains a major challenge. In our application, we found that when multiple animals were in close proximity or engaged in physical interaction, the model frequently confused their identities. These identity swaps persisted even after manual refinement of the trajectories. This limits the possibility of performing within-litter comparisons. One potential solution is to equip each animal with an identification chip (e.g., RFID) and combine this with DLC tracking, enabling consistent identity assignment alongside pose estimation (Fong et al., 2023). Third, motion blur caused by the limited frame rate impaired tracking accuracy. While rat pups are capable of moving at high velocities, the videos were recorded at only 14 frames per second. As a result, many frames involving rapid movement were blurred, making it difficult for the model to reliably detect key points. This often led to fragmented or missing trajectories. Although post hoc interpolation can partially restore these trajectories, it may not accurately reflect the true movement dynamics. To overcome this limitation, future recordings may benefit from using high-speed cameras with higher frame rates.

By analysing the tracking data of developing rat pups, we identified two distinct developmental patterns: an abrupt increase in exploratory behaviour at P18, followed by a later increase in sustained movement at P21. These findings align with previous studies of exploratory onset at around P18 (Gerrish & Alberts, 1996; Nadel et al., 1992) and motor maturation around P21 (Altman & Sudarshan, 1975). By evaluating both behaviours within the same naturalistic context, our study shows that exploration and movement follow separate developmental trajectories, with exploration emerging earlier than sustained movement. One possible explanation is that rat pups begin to leave the nest and explore their surroundings at P18, but lack the motor capacity to move over longer periods. As their motor abilities mature over the following days, more continuous movement becomes possible around P21.

Interestingly, we also observed occasional exploratory behaviour as early as P9, despite the pups' limited motor capabilities. Although rare, these early outside-huddle movements indicate that exploratory behaviour may emerge earlier than the time (P15) suggested by Gerrish & Alberts (1996) using open field assays. Therefore, these early movement and exploration may play an important role in the development of hippocampal circuitry related

to the formation of a cognitive map. Finally, with DLC's ability to track multiple body parts, there is great potential to automate the detection of more complex behaviours, such as rearing on the hind legs—a key marker of novelty-induced exploration (Lever et al., 2006).

To summarise, despite certain limitations, DeepLabCut provides a feasible approach for tracking and analysing rodent behaviour in naturalistic home-cage environment — enabling investigation into how early-life experience contributes to cognitive map development.

Acknowledgements

I would like to thank Prof. Tom Wills for his guidance and support throughout the project. I am also grateful to the members of the Wills Lab, especially Marco Abrate and Asra Albu-Swailim, for their helpful discussions and feedback during data analysis.

References

Altman, J., & Sudarshan, K. (1975). Postnatal development of locomotion in the laboratory rat. *Animal Behaviour*, *23*, 896–920.

Bassett, J. P., Wills, T. J., & Cacucci, F. (2018). Self-Organized Attractor Dynamics in the developing Head direction circuit. *Current Biology*, *28*(4), 609-615.e3. https://doi.org/10.1016/j.cub.2018.01.010

Berlyne, D. E. (1966). Curiosity and exploration. *Science*, *153*(3731), 25–33. https://doi.org/10.1126/science.153.3731.25

Bronstein, P. M., Neiman, H., Wolkoff, F. D., & Levine, M. J. (1974). The development of habituation in the rat. *Animal Learning & Behavior*, *2*(2), 92–96. https://doi.org/10.3758/bf03199129

Chanthongdee, K., Fuentealba, Y., Wahlestedt, T., Foulhac, L., Kardash, T., Coppola, A., Heilig, M., & Barbier, E. (2024). Comprehensive ethological analysis of fear expression in rats using DeepLabCut and SimBA machine learning model. *Frontiers in Behavioral Neuroscience*, *18*. https://doi.org/10.3389/fnbeh.2024.1440601

Chen, Y., & Baram, T. Z. (2015). Toward understanding how Early-Life stress reprograms cognitive and emotional brain networks. *Neuropsychopharmacology*, *41*(1), 197–206. https://doi.org/10.1038/npp.2015.181

Environmental temperature modulates onset of independent feeding: warmer is sooner. (1996). *PubMed*. https://doi.org/10.1002/(SICI)1098-2302(199609)29:6

Fong, T., Hu, H., Gupta, P., Jury, B., & Murphy, T. H. (2023). PyMouseTracks: flexible computer vision and RFID-Based system for multiple mouse tracking and behavioral assessment. *eNeuro*, *10*(5), ENEURO.0127-22.2023. https://doi.org/10.1523/eneuro.0127-22.2023

Hahnloser, R. (2003). Emergence of neural integration in the head-direction system by visual supervision. *Neuroscience*, *120*(3), 877–891. https://doi.org/10.1016/s0306-4522(03)00201-x

Hofer, M. A., & Shair, H. (1978). Ultrasonic vocalization during social interaction and isolation in 2-week-old rats. *Developmental Psychobiology*, *11*(5), 495–504. https://doi.org/10.1002/dev.420110513

Lapp, H. E., Salazar, M. G., & Champagne, F. A. (2023). Automated maternal behavior during early life in rodents (AMBER) pipeline. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-45495-4

Lauer, J., Zhou, M., Ye, S., Menegas, W., Schneider, S., Nath, T., Rahman, M. M., Di Santo, V., Soberanes, D., Feng, G., Murthy, V. N., Lauder, G., Dulac, C., Mathis, M. W., & Mathis, A. (2022). Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nature Methods*, *19*(4), 496–504. https://doi.org/10.1038/s41592-022-01443-0

Liu, D., Diorio, J., Day, J. C., Francis, D. D., & Meaney, M. J. (2000). Maternal care, hippocampal synaptogenesis and cognitive development in rats. *Nature Neuroscience*, *3*(8), 799–806. https://doi.org/10.1038/77702

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

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a cognitive Map*. https://repository.arizona.edu/handle/10150/620894

O'Keefe, J., University College London, Tolman, E. C., Morris, R., Garrud, P., Rawlins, J., & O'Keefe, J. (2014). Spatial cells in the hippocampal formation. In Nobel Prize Lecture, *Spatial Cells in the Hippocampal Formation*. https://www.nobelprize.org/uploads/2018/06/okeefe-lecture-slides.pdf

Olfactory preference for mother over home nest shavings by newborn rats. (1998, July 1). PubMed. https://pubmed.ncbi.nlm.nih.gov/9664168/

Popik, P., Cyrano, E., Piotrowska, D., Holuj, M., Golebiowska, J., Malikowska-Racia, N., Potasiewicz, A., & Nikiforuk, A. (2024). Effects of ketamine on rat social behavior as analyzed by DeepLabCut and SimBA deep learning algorithms. *Frontiers in Pharmacology*, *14*. https://doi.org/10.3389/fphar.2023.1329424

Randall, P. K., & Campbell, B. A. (1976). Ontogeny of behavioral arousal in rats: Effect of maternal and sibling presence. *Journal of Comparative and Physiological Psychology*, *90*(5), 453–459. https://doi.org/10.1037/h0077211

Renner, M. J., & Pierre, P. J. (1998). Development of Exploration and Investigation in the Norway Rat (Rattus norvegicus). *The Journal of General Psychology*, *125*(3), 270–291. https://doi.org/10.1080/00221309809595550

Scott, R. C., Richard, G. R., Holmes, G. L., & Lenck-Santini, P. (2010). Maturational dynamics of hippocampal place cells in immature rats. *Hippocampus*, *21*(4), 347–353. https://doi.org/10.1002/hipo.20789

Sullivan, R. M., Landers, M. S., Flemming, J., Vaught, C., Young, T. A., & Polan, H. J. (2003). Characterizing the functional significance of the neonatal rat vibrissae prior to the onset of whisking. *Somatosensory & Motor Research*, *20*(2), 157–162. https://doi.org/10.1080/0899022031000105190

Tan, H. M., Bassett, J. P., O'Keefe, J., Cacucci, F., & Wills, T. J. (2015). The Development of the Head Direction System before Eye Opening in the Rat. *Current Biology*, *25*(4), 479–483. https://doi.org/10.1016/j.cub.2014.12.030

Taube, J. S. (2007). The Head Direction Signal: Origins and Sensory-Motor Integration. *Annual Review of Neuroscience*, *30*(1), 181–207. https://doi.org/10.1146/annurev.neuro.29.051605.112854

Ulsaker-Janke, I., Waaga, T., Waaga, T., Moser, E. I., & Moser, M. (2023). Grid cells in rats deprived of geometric experience during development. *Proceedings of the National Academy of Sciences*, *120*(41). https://doi.org/10.1073/pnas.2310820120

Uziel, A., Romand, R., & Marot, M. (1981). Development of cochlear potentials in rats. *International Journal of Audiology*, *20*(2), 89–100. https://doi.org/10.3109/00206098109072687

Wills, T. J., Barry, C., & Cacucci, F. (2012). The abrupt development of adult-like grid cell firing in the medial entorhinal cortex. *Frontiers in Neural Circuits*, *6*. https://doi.org/10.3389/fncir.2012.00021

Wills, T. J., Cacucci, F., Burgess, N., & O'Keefe, J. (2010). Development of the hippocampal cognitive map in preweanling rats. *Science*, *328*(5985), 1573–1576. https://doi.org/10.1126/science.1188224

Wills, T. J., Muessig, L., & Cacucci, F. (2013). The development of spatial behaviour and the hippocampal neural representation of space. *Philosophical Transactions of the Royal Society B Biological Sciences*, *369*(1635), 20130409. https://doi.org/10.1098/rstb.2013.0409

Zahran, M. A., Manas-Ojeda, A., Navarro-Sánchez, M., Castillo-Gómez, E., & Olucha-Bordonau, F. E. (2024). Deep Learning-Based Scoring Method of the Three-Chamber social behaviour Test in a mouse model of alcohol intoxication. A comparative analysis of DeepLabCUT, Commercial automatic Tracking and manual scoring. *Heliyon*, *10*(17), e36352. https://doi.org/10.1016/j.heliyon.2024.e36352

Appendices

All Python scripts used for the analysis are available at:

https://github.com/BrantYHX/Brant-Yin-Dissertation.git