Hackathon @ Harwell Oxfordshire (H2O)

3rd & 4th Oct 2023, CA20135 TEATIME

1. Introduction

1.1. Background

Hackathons typically bring together individuals with common interests to develop a software solution to a problem in a social, collaborative environment over the course of 24 - 48 hours. In recent years, machine learning (ML) hackathons have become popular events where participants, often grouped into teams, are provided with a well-curated benchmark training dataset and challenged to produce an ML solution using any techniques at their disposal. At the end of the challenge period, the “final” models are then evaluated against an unseen test dataset, and ranked according to accuracy (or similar).

1.2. Proposal

In support of CA20135 (TEATIME), we proposed an ML challenge hackathon focused on a common problem in home-cage monitoring: mouse behaviour classification. Several hours of video footage of single-housed mice were provided to participants alongside per-frame ground truth annotations of the behaviours being performed. Participants were invited to develop a supervised ML approach (i.e., using the known behaviour labels) using any methods at their disposal, including off-the-shelf techniques. At the end of the challenge, teams applied their models to a previously unseen testing dataset, and their results compared to the ground truth.

2. Challenge details

2.1. Dataset

A publicly available mouse behaviour dataset originally published by Jhuang et al. (2010) was provided to participants. This dataset comprises 4,200 short clips with each depicting a single, unambiguous behaviour of which eight are represented: *eat, drink, groom, walk, rear, hang, micromovement* and *rest.* Some of the behaviours are more visually discernible than others. The dataset itself was organised into twelve folders, with nine folders available to participants for model development (i.e., training) and the remaining three folders reserved for model evaluation (i.e., testing). Participants were requested not to use the latter dataset for any purpose other than final evaluation.

2.2. Code

Participants were provided with a “Jupyter Notebook” that contained starter code to minimise the effort required to load and pre-process the data, and instead focus on model training (https://lncn.ac/teatimeh2o). This also allowed participants to easily develop their solutions in Google Colaboratory (Colab) (https://colab.research.google.com) which provides free access to a Python programming environment as well as limited access to graphics processing units (GPUs). A brief tutorial of the starter code and Colab environment was provided to participants on the first day. In all but a few cases, participants opted to use this notebook as a starting point for developing their solutions.

3. Report on challenge

3.1. Participants

The hackathon was attended by 17 participants from 6 countries, with a gender balance of 29% Female to 71% Males. Most participants were young researchers, and thus at an early career stage (i.e. PhD or postdoctoral). The icebreaker activity indicated that most participants considered themselves at least moderately experienced in ML, with a few having published work in the field. Some participants had specific experience in the context of animal behaviour quantification (mice, macaques). Approximately half of participants opted to work in small teams, with the remaining half choosing to work individually. However, there was a strong collaborative atmosphere throughout the hackathon regardless of team participation. The first day of the hackathon was focused solely on model development.

3.2. Solutions

The second day of the hackathon was dedicated to final tweaks and model evaluation. A range of innovative solutions were then presented by participants that can be broadly categorised as follows:

- *Pixel-based*: raw video frames/clips used as inputs to a deep learning model
- *Landmark-based:* key points (2D coordinates) on the mouse's body were extracted
- *Segmentation-based*: video frames separated into parts (e.g., mouse, hopper, sawdust)

Pixel-based methods generally relied on training one or more convolutional neural networks (CNNs) to classify each clip into one the eight categories, with some methods opting to process the clips in some way prior to training (e.g., converting clips to single frames that reduce the data size while still preserving motion information). In general, these methods tended to perform best, achieving up to 70% accuracy on the unseen test set.

Several teams used a landmark-based approach whereby key points were extracted using DeepLabCut, and subsequently used to train a machine learning model (e.g., a random forest). Similarly, segmentation-based methods used techniques such as the Segment Anything Model (Kirillov et al. 2023) to extract regions of interest prior to downstream classification. In general, these methods did not perform as well as those methods that were trained directly on raw pixel data (50 - 60% accuracy on unseen data).

3.3. Awards

Two award categories were created by the organising committee:

- *Best solution*: the model with the highest evaluation accuracy was deemed the winner
- *Most innovative*: the most creative/ambitious approach, irrespective of how well it worked

The best solution awarded was given to Luis Perdigao whose pixel-based method was developed using the "fast.ai" library. The most innovative award was given to Juul Vossen and David Thonnard developing a landmark-based approach based on graph neural networks (GNNs).

4. Outcomes

Overall, the event was successful in delivering its aim of bringing together researchers from across the European community to work collaboratively on an open problem in home cage analysis. Feedback from participants was highly positive.

It became clear, during the planning and duration of the Hackathon, is what is missing from the field of home-cage monitoring are gold-standard reference sets of data. This, along with the lack of standardised data formats for the data and dispersion of data in inaccessible locations, prevents these data being used for AI purposes. Moreover, the datasets often lack critical metadata (e.g. environmental conditions such as light levels) making it almost impossible to merge and synthesise data across labs and limiting the reproducibility of the research. Together these became a driver for the establishment of the TEATIME working group to design a publicly available repository for home-cage monitoring data.

5. References

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y. and Dollár, P., 2023. Segment anything. *arXiv preprint arXiv:2304.02643*.

Jhuang, H., Garrote, E., Yu, X., Khilnani, V., Poggio, T., Steele, A.D. and Serre, T., 2010. Automated home-cage behavioural phenotyping of mice. Nature communications, 1(1), p.68.

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