

A Pre-trained Transformer Model for Animal Behavior

September 2024

Project description

The scientific study of animal behaviour relies heavily on human observation of animals to define the units of their behaviour for subsequent analysis. However, whether and how this human-defined perspective affects our ability to understand how animals interpret and respond to each others' behaviour is unknown. For social behaviour like courtship, this is crucial. In addition, manual annotation of videos is typically used to quantify behaviours, and scientists are limited to analysing behavioural counts and/or durations. Machine Learning (ML) offers a solution to the subjectivity in human-created definitions of behaviour, and to reduce human labour involved in manual annotations of behaviour, but progress has been slow.

Motivation Large pre-trained transformer models [1] are a cutting-edge machine learning approach that uses unsupervised learning of very large data-sets (e.g. Large Language Models like GPT and T5) combined with subsequent, less computationally intensive fine-tuning to solve specific domain-relevant tasks (e.g. building a chat-bot like ChatGPT). This approach has been successfully applied to the problem of Human Action Recognition, e.g. accelerometer data from wearable sensors can be used to estimate movement behaviours (walking, running, lying down, etc.) [2].

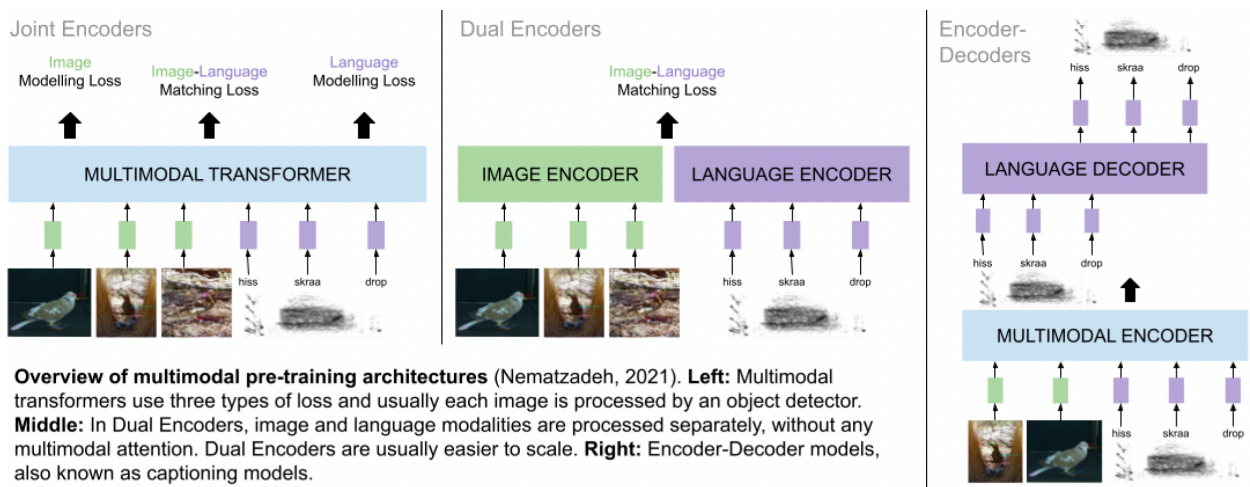


Figure 1: Overview of multimodal pre-training architectures.

Goal The aim of this project is to apply lessons learned from ongoing ML work into self-supervised learning and Human Action Recognition to the domain of animal behaviour, to unlock the revolutionary potential of ML frameworks for animal behavioural science.

You will use two large sets of accelerometer data already collected by behavioral biologists from free-ranging grey-lag geese (University of Vienna) and captive common quail (KLIVV, University of Veterinary Medicine, Vienna). As the first step, the human foundation model in [2] can be used to investigate the plausibility of using fine-tuning of Action Recognition models for these two bird species, which reduces the amount of data needed for training. We will also check the feasibility of developing our own dedicated foundation model, which is much more computationally intensive and data-hungry. Finally, we can also use the two data-sets to investigate generalization across species.

The topic is broad, but we will proceed step by step with the task.

Additional information

- **What will you learn?** The student will work on the SOTA transformer architectures for representation learning, and gain hands-on experience on foundation models when evaluating the performance of the developed methodology.
- **Requirements:** Deep learning knowledge (experience with self-supervised learning is a big plus), good Python and PyTorch skills, version control (git).
- **Supervisors:** Dr. Yun Cheng (yun.cheng@sdsc.ethz.ch), Prof. Dr. Olga Saukh (saukh@tugraz.at), Dr. Cliodhna Quigley (cliodhna.quigley@univie.ac.at)

References

- [1] Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., et al.: On the opportunities and risks of foundation models (2022), 2108.07258
- [2] Yuan, H., Chan, S., Creagh, A.P., Tong, C., Clifton, D.A., Doherty, A.: Self-supervised learning for human activity recognition using 700,000 person-days of wearable data (2023), 2206.02909