Sensorimotor representation learning of manipulable objects

M2 internship proposal
2024-2025

Internship context and scientific objectives

There is "a fundamental misalignment between human and typical AI representations: while the former are grounded in rich sensorimotor experience, the latter are typically passive and limited to a few modalities such as vision and text" [1]. We propose in the MeSMRise (Multimodal deep SensoriMotor Representation learning) project to take inspiration from the way human babies learn to explore their environment through actions that shape their multimodal experience [2]. Especially, the sensorimotor contingencies (SMC) theory [3] combines coherent pieces of evidence from neuroscience, psychology, etc. of human perception and learning in a unified framework. The key claims are the learning of SMCs defined as "the structure of the rules governing the sensory changes produced by various motor actions" [3] and active perception as the "organism's exploration of the environment that is mediated by knowledge of SMCs" [4]. Some models implementing this theory are able to learn complex concepts such as containment [1] for instance.

Inspired by the SMC theory, the main objective of the project is to study how action can structure multimodal representations, learned with self-supervised learning methods. This will be applied to 3D objects, perceived by vision and point cloud, and manipulated in virtual environments. Specifically, we target the following properties:

- generalization to unknown environments and contexts
- robustness, e.g. to the orientation, background, shape ... of the object
- adaptability via the capacity of the model to autonomously find relevant information
- generality by using similar architectures and principles for all research questions

This proposed internship takes place in the second work package of the project (that also includes a PhD position opening in September 2025). This work package focuses on object representations as webs of anticipations for potential interaction [5] to address the following questions:

- how sensorimotor hierarchical representations can be integrated in network models (from detailed instances to abstract classes, from instant predictions to longer term integration)
- how higher level and path dependent features from graphical models can improve learning and perception (e.g., guiding actions for improved distinguishability of objects)

The generation of rich features in another project work package will rely on contrastive deep learning techniques, which will be exploited to represent objects. To this purpose, we may extend recent advances in deep neural networks such as Space State Models [6], but such models currently rely on unfolding of fixed input sequences to remain computationally tractable. Our key insights to improve object representation being based on action selection and bifurcations in interactions sequences, simulating the dynamics of interactions is required (in continuous or discrete time). Therefore, and to aim at realistic objectives for a 5-month internship project, object acquisition and manipulation may be simulated within a 2D discrete space and discrete time Tetris environment [7], with action selection being performed within the learned manifold [8].

Internship tasks

- Bibliographical study on distributed, sensorimotor and multimodal representations of objects;
- Within simulator implementation of simple sensorimotor interactions and learning of predictive models of objects;
- Model validation through the use of adequate metrics (graph theory, data mining, statistics);
- Participation in internship and project meetings (face-to-face and remotely), oral communication on the project and internship work;
- Writing of preparatory documents, interim and final reports.

Technical skills

- Strong algorithmic and programming skills (e.g., Python, R);
- AI / machine learning / deep learning techniques (e.g., using PyTorch/Tensorflow);
- Graphical models (e.g., graph theory, probabilistic);
- Parallelism and optimization (mainly software-based);
- Data analysis / statistics;
- Writing and oral skills in scientific English;
- Distributed version control systems (e.g., git) and Open Science practices.

Interpersonal skills

- Scientific and methodological rigor, including algorithms, programming and evaluation;
- Interdisciplinary teamwork and scientific interest in natural intelligence modelling;
- Ability to interact face-to-face and remotely with different members of the consortium;
- Autonomy and proactivity in research activities and reporting.

Practical internship conditions

The internship will start in February-March and last 5-6 months, funded by the MeSMRise ANR project, following national regulations in force at the time the internship agreement will be signed. The funding for a PhD fellowship has been secured following the topics of this internship (position to open from September 2025 for 3 years).

The intern will be located in Grenoble, supervised by Jean-Charles Quinton (LJK – UMR 5524 – Grenoble) in interaction with Mathieu Lefort and Frédéric Armetta (LIRIS – UMR 5205 – Lyon), and may regularly interact with other members and teams of the consortium. Some travels to Lyon and Clermont-Ferrand may therefore be organized. Access to hardware and an office will be guaranteed at LJK.

Contact

To apply or for any question on the internship, please contact Jean-Charles Quinton (quintonj@univ-grenoble-alpes.fr).

References

[1] Nicholas Hay, Michael Stark, Alexander Schlegel, Carter Wendelken, Dennis Park, Eric Purdy, Tom Silver, D Scott Phoenix, and Dileep George. Behavior is everything: Towards representing concepts with sensorimotor contingencies. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018.

[2] Linda Smith and Michael Gasser. The development of embodied cognition: Six lessons from babies. Artificial life, 11(1-2):13-29, 2005.

[3] J Kevin O'regan and Alva Noë. A sensorimotor account of vision and visual consciousness. Behavioral and brain sciences, 24(5):939-973, 2001.

[4] Erik Myin and J Kevin O'Regan. Perceptual consciousness, access to modality and skill theories. A way to naturalize phenomenology? Journal of consciousness studies, 9(1):27-46, 2002.

[5] Jedediah WP Allen, Bartuğ Çelik, Mark H Bickhard. Age 4 transitions: Reflection as a domain-general development for explicit reasoning. Cognitive Development, 2021.

[6] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396, 2021.

[7] Jean-Charles Quinton. Emergence of space from sensorimotor invariants: anticipatory network analysis in the context of the Tetris game. Cognitive Processing, 13, 48, 2012.

[8] Simon Forest, Jean-Charles Quinton, and Mathieu Lefort. Combining manifold learning and neural field dynamics for multimodal fusion. In 2022 International Joint Conference on Neural Networks (IJCNN), pages 1-8. IEEE, 2022.