

Titre de la thèse : De l'interaction Homme-Machine à l'entraînement de robots (InteracTraining)

Laboratoire d'accueil : Connaissance et Intelligence Artificielle Distribuées (CIAD) – <http://www.ciad-lab.fr>

Spécialité du doctorat préparé : Informatique

Mots-clefs : Interaction homme-machine, Apprentissage par renforcement, IA générative, Environnements simulés, Intégration de récompenses multi-sources, Mécanismes d'attention.

Contexte :

L'apprentissage par renforcement (AR) constitue une approche clé pour permettre aux robots d'apprendre de manière autonome à travers leurs interactions avec l'environnement. Cette approche repose sur l'optimisation d'une fonction de récompense pour guider les comportements dans des tâches complexes telles que la navigation et la manipulation. Elle s'est étendue récemment à d'autres domaines, notamment le traitement du langage naturel, où l'évaluation de l'humain (retour humain) est utilisée comme signal de récompense, comme dans le cas de ChatGPT-4 [1], enrichissant ainsi le processus d'apprentissage avec une dimension qualitative [2,3].

Parallèlement, les avancées en vision par ordinateur et en apprentissage profond permettent l'extraction avec précision des émotions humaines (via les expressions faciales) ou la reconnaissance des gestes [4,5]. Ces signaux peuvent constituer un retour qualitatif pertinent dans un contexte d'apprentissage robotique. Toutefois, les simulateurs actuels utilisés en AR, comme Habitat-Sim 3.0 [6], bien qu'intégrant des avatars humains, ne prennent pas encore en charge ce type de retour expressif.

Face à cette limitation, les modèles génératifs, tels que les GANs ou Stable Diffusion, offrent de nouvelles perspectives. Déjà utilisés dans les jeux vidéo ou les interfaces homme-machine, ces modèles permettent une génération conditionnelle réaliste d'expressions faciales et de gestes en fonction du contexte [7].

Objectifs :

Cette thèse vise à intégrer les modèles génératifs (GANs ou Stable Diffusion), dans des environnements simulés pour produire des retours expressifs humains (expressions faciales et gestes) réalistes et contextuellement adaptés, afin de les utiliser comme signaux (au sein des fonctions de récompense), dans le processus d'apprentissage par renforcement. L'objectif est de combiner ces signaux qualitatifs aux récompenses classiques (quantitatives) afin de guider plus finement les stratégies d'apprentissage de robots.

Concrètement, il s'agira de générer, via stable diffusion ou GAN, des retours expressifs humains cohérents avec l'état du robot et son environnement, puis d'intégrer ces retours dans la fonction de récompense du système AR. L'intégration de cette double récompense (quantitative et qualitative) devrait permettre un apprentissage plus fin, validé par l'humain et adapté à des tâches complexes de robotique interactive.

Le projet s'appuie sur la plateforme MOBILITECH-VAR, composée de robots et véhicules autonomes dotés de capteurs variés, pour effectuer la validation expérimentale et l'acquisition de données

réelles. L'ensemble du processus nécessite un support de calcul intensif, assuré par les serveurs graphiques du laboratoire CIAD à l'UTBM (site de Montbéliard), optimisés pour l'entraînement de modèles d'AR.

Défis :

Deux verrous scientifiques sont identifiés :

- Génération réaliste, cohérente et temps réel de retours expressifs (expressions faciales / gestuelles)
Il s'agit de générer des expressions faciales et des gestuelles réalistes, compatibles avec les contraintes temporelle et spatiale de l'environnement simulé. Il faudra garantir que ces retours synthétiques soient crédibles et exploitables par le robot en temps réel, tout en respectant la dynamique de la tâche de ce dernier.
- Fusion de récompenses qualitatives et quantitatives
Il s'agit de combiner efficacement la récompense qualitative issue de l'interprétation des signaux de retour de l'humain (expressions faciales et gestuelles) et les récompenses quantitatives issues de l'évaluation classique de la tâche du robot via des métriques. Cette combinaison hétérogène représente une difficulté majeure, largement reconnue dans la littérature sur l'AR [8], car une mauvaise pondération peut nuire aux performances globales de l'agent. Il s'agira donc de concevoir des stratégies d'intégration robustes et adaptatives.

Références :

1. Gill, Sukhpal Singh, and Rupinder Kaur. "ChatGPT: Vision and challenges." Internet of Things and Cyber-Physical Systems (2023).
2. Liu, Changchun, et al. "A mixed perception-based human-robot collaborative maintenance approach driven by augmented reality and online deep reinforcement learning." Robotics and Computer-Integrated Manufacturing 83 (2023): 102568.
3. Yu, Tian, and Qing Chang. "User-guided motion planning with reinforcement learning for human-robot collaboration in smart manufacturing." Expert Systems with Applications (2022): 118291.
4. Sun, Zhe, et al. "A discriminatively deep fusion approach with improved conditional GAN (im-cGAN) for facial expression recognition." Pattern Recognition 135 (2023): 109157.
5. Zhang, Ziyang, et al. "Enhanced discriminative global-local feature learning with priority for facial expression recognition." Information Sciences 630 (2023): 370-384.
6. Puig, Xavier, et al. "Habitat 3.0: A co-habitat for humans, avatars and robots." arXiv preprint arXiv:2310.13724 (2023).
7. Xia, Yifan, et al. "Local and global perception generative adversarial network for facial expression synthesis." IEEE Transactions on Circuits and Systems for Video Technology 32.3 (2021): 1443-1452.
8. Dann, Christoph, Yishay Mansour, and Mehryar Mohri. "Reinforcement Learning Can Be More Efficient with Multiple Rewards." Google Research. Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR202 (2023).

Profil demandé :

- Master en vision par ordinateur, apprentissage automatique, intelligence artificielle ou Robotique.
- Compétences solides en Python et PyTorch, avec une expérience dans les modèles génératifs profonds (GANs, modèles de diffusion comme Stable Diffusion).
- Bonne compréhension de l'apprentissage par renforcement, en particulier des fonctions de récompense et de leur optimisation.
- Connaissance des architectures d'attention et des mécanismes de fusion multi-sources.

- Familiarité avec les environnements simulés (type Habitat-Sim, iGibson ou Isaac Sim) et les environnements de robotisation ROS.
- Une bonne maîtrise de l'anglais (oral et écrit) est exigée.

Financement : Contrat doctoral financé par la région Bourgogne Franche-Comté (3 ans)

Dossier à envoyer pour le 25 août 2025

Date de début de la thèse : octobre/novembre 2025

Les candidatures doivent être envoyées par mail à Prof. Y. Ruichek (yassine.ruichek@utbm.fr) et Dr. M. Kas (mohamed.kas@utbm.fr).

Le dossier de candidature doit contenir : un CV détaillé, une copie du diplôme de Master ou tout document attestant du niveau de Master, une copie des bulletins de notes de Master, références et/ou une à deux lettres de recommandation

Direction / codirection de la thèse :

Yassine Ruichek (Directeur de thèse) et Mohamed Kas (Co-encadrant)

PhD title: From Human-Machine Interaction to Robot Training (InteracTraining)

Host laboratory: Connaissance et Intelligence Artificielle Distribuées (CIAD) –
<http://www.ciad-lab.fr>

Specialty of PhD: Computer Science

Keywords: Human-machine interaction, Reinforcement learning, Generative AI, Simulated environments, Multi-source reward integration, Attention mechanisms.

Context:

Reinforcement learning is an important approach that allows robots to learn on their own by interacting with their environment. It is based on optimizing a reward function to guide behavior in complex tasks such as navigation and object handling. Recently, this approach has also been used in other areas, such as natural language processing, where human feedback is used as a reward signal. This is the case with ChatGPT-4 [1], where the learning process is improved through human evaluation [2,3].

At the same time, progress in computer vision and deep learning has made it possible to recognize human emotions accurately through facial expressions and gestures [4,5]. These signals can be useful as feedback in the learning process for robots. However, current simulators used in reinforcement learning, such as Habitat Sim 3.0 [6], even though they include human avatars, do not yet support this type of expressive feedback.

To address this limitation, generative models like GANs and Stable Diffusion offer a good solution to make this feedback possible. These models are already used in video games and human-computer interfaces to create realistic and context-based facial and body expressions [7].

Objectives:

This research aims to include generative models such as GANs or Stable Diffusion in simulation environments to create realistic and meaningful human expressions (faces and gestures). These expressions will then be used as feedback signals in the reinforcement learning process. The goal is to combine these qualitative-based signals with traditional reward signals (metrics) to help guide the robot's learning more effectively.

More specifically, the project will create facial and gestural feedback using GAN or diffusion models, adapted to the robot's state and its environment. These will then be used in the reward system of the learning algorithm. By combining qualitative and quantitative feedback, the robot can learn better, with behavior that is more acceptable from a human point of view, especially in complex situations.

The project will benefit from the MOBILITECH VAR platform, which includes robots and self-driving vehicles equipped with different sensors, to test and validate the system in real settings. All of this will be supported by high-performance computing resources provided by the CIAD laboratory at UTBM in Montbéliard, which are specially set up for training learning models.

Challenges:

This research presents two main challenges. The first is the ability to generate realistic and expressive human feedback such as facial expressions and gestures in real time, while respecting the spatial and temporal constraints of the simulation environment. These synthetic expressions must be both believable and relevant to the robot's current context in order to be effectively used during the learning process. The second challenge lies in the integration of qualitative feedback derived from human signals with traditional quantitative rewards that measure task performance. Combining these heterogeneous types of feedback is complex and has been widely recognized in the literature on reinforcement learning as a critical issue, since improper weighting or fusion can negatively affect the learning efficiency and behavior of the agent [8]. The research will therefore focus on developing robust and adaptive strategies to merge both feedback types in a way that enhances learning and ensures that the robot develops behaviors that are both efficient and human aligned.

References:

1. Gill, Sukhpal Singh, and Rupinder Kaur. "ChatGPT: Vision and challenges." Internet of Things and Cyber-Physical Systems (2023).
2. Liu, Changchun, et al. "A mixed perception-based human-robot collaborative maintenance approach driven by augmented reality and online deep reinforcement learning." Robotics and Computer-Integrated Manufacturing 83 (2023): 102568.
3. Yu, Tian, and Qing Chang. "User-guided motion planning with reinforcement learning for human-robot collaboration in smart manufacturing." Expert Systems with Applications (2022): 118291.
4. Sun, Zhe, et al. "A discriminatively deep fusion approach with improved conditional GAN (im-cGAN) for facial expression recognition." Pattern Recognition 135 (2023): 109157.
5. Zhang, Ziyang, et al. "Enhanced discriminative global-local feature learning with priority for facial expression recognition." Information Sciences 630 (2023): 370-384.
6. Puig, Xavier, et al. "Habitat 3.0: A co-habitat for humans, avatars and robots." arXiv preprint arXiv:2310.13724 (2023).
7. Xia, Yifan, et al. "Local and global perception generative adversarial network for facial expression synthesis." IEEE Transactions on Circuits and Systems for Video Technology 32.3 (2021): 1443-1452.
8. Dann, Christoph, Yishay Mansour, and Mehryar Mohri. "Reinforcement Learning Can Be More Efficient with Multiple Rewards." Google Research. Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR202 (2023).

Candidate Profile:

- Master's degree in computer vision, machine learning, artificial intelligence, or robotics.
- Strong skills in Python and PyTorch, with experience in deep generative models (GANs, diffusion models such as Stable Diffusion).
- Good understanding of reinforcement learning, especially reward functions and their optimization.
- Knowledge of attention architectures and multi-source fusion mechanisms (qualitative and quantitative).
- Familiarity with simulated environments (such as Habitat-Sim, iGibson, or Isaac Sim) and robotic platforms (ROS).
- Advanced level in English writing and speaking is required.

Financing Institution: Doctoral contract funded by the région Bourgogne Franche-Comté (3 years)

Application deadline: August, 25th 2025

Expected start date: October/November 2025

Applications must be sent by email to Prof. Y. Ruichek (yassine.ruichek@utbm.fr) and Dr. M. Kas (mohamed.kas@utbm.fr).

The application must include: a detailed CV, a copy of the master's degree or any document attesting the master's level, a copy of the master's transcripts, references and/or one to two recommendation letters.

Supervisor(s):

Prof. Yassine Ruichek (supervisor), Dr. Mohamed Kas (co-supervisor)