PhD proposal on Diffusion Model-Based Robot Visuomotor Policy Learning for General-Purpose and Multi-Task Robot Manipulation and Assembly Conditioned by Instructions

@Imagine Team of LIRIS lab, Ecole Centrale de Lyon

The advancements in robotics and artificial intelligence have paved the way for sophisticated robotic systems capable of performing complex manipulation and assembly tasks. However, achieving efficient and robust visuomotor policies for robots remains a significant challenge, especially in dynamic and unstructured environments. This proposal aims to explore the potential of diffusion models in learning visuomotor policies, leveraging their ability to handle multimodal action distributions, high-dimensional action spaces, and general-purpose multi-task conditioning by natural language instructions.

Background and Motivation

Visuomotor policy learning involves mapping visual observations to motor actions, enabling robots to interact with their environment effectively. Traditional approaches often struggle with the complexities of multimodal action distributions and the need for high precision and temporal consistency. The recent introduction of diffusion policies offers a promising solution by employing a conditional denoising diffusion process to generate robot actions. These models have demonstrated superior performance in generating complex behaviors, making them an ideal candidate for robot manipulation and assembly tasks. Moreover, integrating natural language processing (NLP) allows for versatile task conditioning, enabling robots to perform a variety of tasks based on human instructions.

Goals

The primary objectives of this research are:

- To develop a diffusion model-based framework for visuomotor policy learning in robotic manipulation and assembly.
- To extend the framework for general-purpose multi-task learning conditioned by natural language instructions.
- To evaluate the performance of the proposed framework across various benchmarks and real-world tasks.
- To identify and address the challenges associated with integrating diffusion models and NLP into robotic control systems.

Skills

To successfully undertake the proposed PhD project, a succesful candidate should possess a combination of skills and expertise across several domains. Here are the key skills best suited for this PhD proposal:

- Machine Learning and AI:
  - Strong understanding of machine learning principles and algorithms.
  - Proficiency in deep learning frameworks (e.g., TensorFlow, PyTorch).
  - Experience with reinforcement learning and imitation learning techniques.
- Robotics:
  - Knowledge of robotic systems, kinematics, and dynamics.
  - Experience with robot manipulation and control algorithms.
  - Familiarity with robotic simulation environments (e.g., Gazebo, PyBullet, MuJoCo).
- Computer Vision:
  - Expertise in computer vision techniques for object detection, tracking, and scene understanding.
  - Experience with visual perception frameworks and libraries (e.g., OpenCV, ROS perception packages).
- Natural Language Processing (NLP):
  - Understanding of NLP techniques for language understanding and generation.
  - Experience with NLP frameworks (e.g., NLTK, spaCy, Hugging Face Transformers).
- Programming Languages:
  - Proficiency in programming languages such as Python and C++.
  - Strong software engineering skills, including code versioning (e.g., Git) and documentation.
- be fluent in french and/or english

Environment

The successful candidate will work in direct collaboration with researchers having an established expertise in computer vision, machine learning and robotics in partnership with high profile international partners (Darmstadt University of Technology in Germany, Nagoya University and Osaka University in Japan). Ecole Centrale de Lyon is part of the top ten engineering schools in France (Grande Ecoles), part of the elite of "Grande Ecoles" offering access to excellent quality graduate and undergraduate students.

Application

Prospective applicants should prepare a comprehensive application package that includes a detailed resume or CV highlighting relevant technical, analytical, and research skills, as well as any prior experience in robotics, machine learning, computer vision, and natural language processing.

Applicants should also submit a statement of purpose that outlines their research interests, motivation for applying, and how their background aligns with the proposed project.

Additionally, candidates are required to provide academic transcripts, two letters of recommendation from faculty or professionals familiar with their work, and any publications or project reports that demonstrate their capabilities.

Applications and letters should be sent via electronic mail to:

- Pr Liming Chen (liming.chen@ec-lyon.fr)

References

- Ho, J., Jain, A., & Abbeel, P. (2020). "Denoising Diffusion Probabilistic Models." In this seminal paper, the authors introduce a new class of generative models called Denoising Diffusion Probabilistic Models (DDPMs). Link to the paper
- Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2021). "Score-Based Generative Modeling through Stochastic Differential Equations." Link to the paper
- Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel, Shuran Song. "Diffusion Policy: Visuomotor Policy Learning via Action Diffusion." Link to the paper.