Introduction

Imagine a baby playing with an object she has never seen before. She feels the object with her fingertips. She shakes it and listens to the noise. She even puts it in her mouth to taste it. These actions might seem strange to adults, but babies learn by making use of all their senses. They intelligently explore new objects and situations without external guidance [1], allowing them to mature into adults. I believe we should create robots that can act and learn as babies do.

The artificial intelligence community has become increasingly interested in reinforcement learning (RL), and for good reason: RL allows systems to move beyond passive learning and interact with the world while learning from these interactions. In the standard RL paradigm, a researcher manually specifies a reward function (score), and an agent learns to maximize this reward. This setup works well in domains like Atari video games, or even complex strategy games like Go, where there are explicit reward signals and a simulator allows for the collection of millions of samples. However, in real-world applications like robotics, reward functions are hard to formulate and learning from real-world data requires sample efficiency, as it is cost-prohibitive to collect many samples from real, deployed physical systems.

I will leverage the strengths of computers by using self-supervision, a set of techniques allowing for automatic generation of training data without relying on the bottleneck of human operation or labeling. In particular, my key idea is that a robot has multiple sensing modalities, and being able to learn associations between them is a rich source of signal. Just as babies explore without external guidance, so too will an agent using this approach. I believe this will lead to significant progress when compared to the currently prevailing approach of *curiosity* [2], which uses visual prediction error as a source of supervision. Learning predictive models, especially with visual inputs, can be difficult to train, as I learned during my undergraduate years doing research on video prediction.

I aim to make real robot learning more tractable through efficient exploration with limited supervision. With this strategy, robots can adapt in unstructured situations, helping them assist humans with everyday tasks in homes and workplaces. My experiences tailoring algorithms to machine learning applications, including computer vision and autonomous driving, strongly position me to conduct this robot learning research.

Intellectual Merit

I propose an alternative to curiosity-based exploration: my **multimodal self-supervision** algorithm makes use of rich environment information beyond images without relying on prediction. Multimodal approaches are especially amenable to self-supervised methods, as we can use two complementary modalities as joint supervision, such as audio and video, vision and textual inputs, or depth and tactile sensor data. My initial experiments use vision and audio, as these are readily available in simulation. Again, imagine the baby, playing with a rattle. The baby shakes the rattle, which surprises our multimodal associative model. This surprise provides a high intrinsic reward for the shaking action and helps the baby to learn the relationships between visual changes and audio changes. This multimodal associative model is trained to output whether the modalities are aligned using a time-contrastive loss [3]. For example, this model could take in an image and an audio clip that is sometimes misaligned. The model is trained to output 1 in aligned cases and 0 otherwise. This associative model's error functions as the agent's reward, and the agent chooses actions to maximize this reward. This multimodal algorithm can behave like the baby: it acquires information across modalities to build a richer representation of the world.

My initial experiments in this direction are promising. Preliminary results show that my method using audio and visual modalities learns a video game twice as fast as the visual prediction (curiosity) baseline and reaches a four times higher peak performance. I plan to submit these results to the 2020 International Conference on Machine Learning.

After this initial demonstration of the approach using audio and video on simple video games, I plan to extend it to other modalities in simulation. I will test it using force sensing and visual inputs on simulated robotic tasks, including both manipulation and locomotion. After proving my method's success in simulation, I will implement it on real-world robots. I am particularly excited about my approach's potential for improving robotic manipulation using visual and tactile sensing. This could enable robots to learn more complex tasks than they can currently learn without supervision, including opening a door or using a hammer.

Through this research, I aim to answer a number of open questions: Is the act of prediction crucial for good exploration? What is the best way to combine modality representations for self-supervision? How can we make these methods sample-efficient enough to yield real-world success? Instead of hardcoding a robot to perform one specific task, we should build robots that adapt and learn to solve many tasks as babies do. My approach brings us closer to this goal and allows for the self-supervised learning of not only modality representations but also an agent policy that could be adapted for downstream use. This work builds on efficient exploration, which will be necessary for robots adapting quickly in complex or unknown environments.

Broader Impacts

The lasting impact of RL will not be from impressive video game AI players or simulated agents that do tricks–it will be from these algorithms working in the real world. I view research through a long-term lens which allows me to focus on designing algorithms with real-world transferability in mind. Both the multimodal and self-supervised aspects of my project move us towards tractable real robot learning. Robots that learn are a precursor to deployment in unstructured environments.

In the long term, I aim to improve people's quality of life through robotics. My goal is to enable robots to move beyond specialized roles in factories and other controlled environments and into homes and workplaces, interacting with and helping humans on a daily basis. I envision a future where these robots clean nuclear waste to keep humans out of dangerous jobs, assist with daily tasks like opening pill bottles to allow elderly people to live independently, and automate driving to save lives lost to human error.

References

[1] P. Rochat. *Object Manipulation and Exploration in 2- to 5-Month-Old Infants*. Developmental Psychology 1989.
[2] Y. Burda, H. Edwards, D. Pathak, et al. *Large-Scale Study of Curiosity-Driven Learning*. ICLR 2019.
[3] P. Sermanet, C. Lynch, et al. *Time-contrastive networks: Self-supervised learning from video*. ICRA 2018.