2017 Research Highlights

PI: Prof. Sergey Levine

Overview

In 2017, we carried out fundamental research on deep reinforcement learning algorithms, deep models for sensorimotor control and touch sensing, autonomous learning of collision avoidance, and other robotics and reinforcement learning topics. Some of our results are summarized below.

2017 Research

Deep Object-Centric Representations for Generalizable Robot Learning

Coline Devin, Pieter Abbeel, Trevor Darrell, Sergey Levine

Robotic manipulation in complex open-world scenarios requires both reliable physical manipulation skills and effective and generalizable perception. In this paper, we propose a method where general purpose pretrained visual models serve as an object-centric prior for the perception system of a learned policy. We devise an object-level attentional mechanism that can be used to determine relevant objects from a few trajectories or demonstrations, and then immediately incorporate those objects into a learned policy. A task-independent meta-attention locates possible objects in the scene, and a task-specific attention identifies which objects are predictive of the trajectories. The scope of the task-specific attention is easily adjusted by showing demonstrations with distractor objects or with diverse relevant objects. Our results indicate that this approach exhibits good generalization across object instances using very few samples, and can be used to learn a variety of manipulation tasks using reinforcement learning.


Code: [https://github.com/cdevin/objectattention](https://github.com/cdevin/objectattention) (will be also linked in bdd repo)
Self-Supervised Deep Reinforcement Learning with Generalized Computation Graphs for Robot Navigation

Gregory Kahn, Adam Villaflor, Bosen Ding, Pieter Abbeel, Sergey Levine

Enabling robots to autonomously navigate complex environments is essential for real-world deployment. Prior methods approach this problem by having the robot maintain an internal map of the world, and then use a localization and planning method to navigate through the internal map. However, these approaches often include a variety of assumptions, are computationally intensive, and do not learn from failures. In contrast, learning-based methods improve as the robot acts in the environment, but are difficult to deploy in the real-world due to their high sample complexity. To address the need to learn complex policies with few samples, we propose a generalized computation graph that subsumes value-based model-free methods and model-based methods, with specific instantiations interpolating between model-free and model-based. We then instantiate this graph to form a navigation model that learns from raw images and is sample efficient. Our simulated car experiments explore the design decisions of our navigation model, and show our approach outperforms single-step and N-step double Q-learning. We also evaluate our approach on a real-world RC car and show it can learn to navigate through a complex indoor environment with a few hours of fully autonomous, self-supervised training.

Code: https://github.com/gkahn13/gcg (in BDD repo too)

EX2: Exploration with Exemplar Models for Deep Reinforcement Learning

Justin Fu*, John D. Co-Reyes*, Sergey Levine

Deep reinforcement learning algorithms have been shown to learn complex tasks using highly general policy classes. However, sparse reward problems remain a significant challenge. Exploration methods based on novelty detection have been particularly successful in such settings but typically require generative or predictive models of the observations, which can be difficult to train when the observations are very
high-dimensional and complex, as in the case of raw images. We propose a novelty detection algorithm for exploration that is based entirely on discriminatively trained exemplar models, where classifiers are trained to discriminate each visited state against all others. Intuitively, novel states are easier to distinguish against other states seen during training. We show that this kind of discriminative modeling corresponds to implicit density estimation, and that it can be combined with countbased exploration to produce competitive results on a range of popular benchmark tasks, including state-of-the-art results on challenging egocentric observations in the vizDoom benchmark.

Code: [https://repo.eecs.berkeley.edu/git/projects/bdd/ex2.git](https://repo.eecs.berkeley.edu/git/projects/bdd/ex2.git)

The Feeling of Success: Does Touch Sensing Help Predict Grasp Outcomes?

**Roberto Calandra**, Andrew Owens, Manu Upadhyaya, Wenzhen Yuan, **Justin Lin**, Edward H. Adelson, **Sergey Levine**

A successful grasp requires careful balancing of contact forces. Deducing whether a particular grasp will be successful from indirect measurements, such as vision, is therefore quite challenging, and direct sensing of contacts through touch sensing provides an appealing avenue toward more successful and consistent robotic grasping. However, in order to fully evaluate the value of touch sensing for grasp outcome prediction, we must understand how touch sensing can influence outcome prediction accuracy when combined with other modalities. Doing so using conventional model-based techniques is exceptionally difficult. In this work, we investigate the question of whether touch sensing aids in predicting grasp outcomes within a multimodal sensing framework that combines vision and touch. To that end, we collected more than 9,000 grasping trials using a two-finger gripper equipped with GelSight high-resolution tactile sensors on each finger, and evaluated visuo-tactile deep neural network models to directly predict grasp outcomes from either modality individually, and from both modalities together. Our experimental results indicate that incorporating tactile readings substantially improve grasping performance.

Dataset and Code: [https://sites.google.com/view/the-feeling-of-success/](https://sites.google.com/view/the-feeling-of-success/) (also BDD repo)

GPLAC: Generalizing Vision-Based Robotic Skills using Weakly Labeled Images
Avi Singh, Larry Yang, Sergey Levine
We tackle the problem of learning robotic sensorimotor control policies that can generalize to visually diverse and unseen environments. Achieving broad generalization typically requires large datasets, which are difficult to obtain for task-specific interactive processes such as reinforcement learning or learning from demonstration. However, much of the visual diversity in the world can be captured through passively collected datasets of images or videos. In our method, which we refer to as GPLAC (Generalized Policy Learning with Attentional Classifier), we use both interaction data and weakly labeled image data to augment the generalization capacity of sensorimotor policies. Our method combines multitask learning on action selection and an auxiliary binary classification objective, together with a convolutional neural network architecture that uses an attentional mechanism to avoid distractors. We show that pairing interaction data from just a single environment with a diverse dataset of weakly labeled data results in greatly improved generalization to unseen environments, and show that this generalization depends on both the auxiliary objective and the attentional architecture that we propose. We demonstrate our results in both simulation and on a real robotic manipulator, and demonstrate substantial improvement over standard convolutional architectures and domain adaptation methods.

CAD2RL: Real Single-Image Flight without a Single Real Image
Fereshteh Sadeghi, Sergey Levine
Deep reinforcement learning has emerged as a promising and powerful technique for automatically acquiring control policies that can process raw sensory inputs, such as images, and perform complex behaviors. However, extending deep RL to real-world robotic tasks has proven challenging, particularly in safety-critical domains such as autonomous flight, where a trial-and-error learning process is often impractical. In this paper, we explore the following question: can we train vision-based navigation policies entirely in simulation, and then transfer
Imitation from Observation: Learning to Imitate Behaviors from Raw Video via Context Translation

YuXuan Liu*, Abhishek Gupta*, Pieter Abbeel, Sergey Levine

Imitation learning is an effective approach for autonomous systems to acquire control policies when an explicit reward function is unavailable, using supervision provided as demonstrations from an expert, typically a human operator. However, standard imitation learning methods assume that the agent receives examples of observation-action tuples that could be provided, for instance, to a supervised learning algorithm. This stands in contrast to how humans and animals imitate: we observe another person performing some behavior and then figure out which actions will realize that behavior, compensating for changes in viewpoint, surroundings, and embodiment. We term this kind of imitation learning as imitation-from-observation and propose an imitation learning method based on video prediction with context translation and deep reinforcement learning. This lifts the assumption in imitation learning that the demonstration should consist of observations and actions in the same environment, and enables a variety of interesting applications, including learning robotic skills that involve tool use simply by observing videos of human tool use. Our experimental results show that our approach can perform imitation-from-observation for a variety of real-world robotic tasks modeled on common household chores, acquiring skills such as sweeping from videos of a human demonstrator.

Website: https://fsadeghi.github.io/CAD2RL/
Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations

Aravind Rajeswaran, Vikash Kumar, Abhishek Gupta, John Schulman, Emanuel Todorov, Sergey Levine

Dexterous multi-fingered hands are extremely versatile and provide a generic way to perform multiple tasks in human-centric environments. However, effectively controlling them remains challenging due to their high dimensionality and large number of potential contacts. Deep reinforcement learning (DRL) provides a model-agnostic approach to control complex dynamical systems, but has not been shown to scale to high-dimensional dexterous manipulation. Furthermore, deployment of DRL on physical systems remains challenging due to sample inefficiency. Thus, the success of DRL in robotics has thus far been limited to simpler manipulators and tasks. In this work, we show that model-free DRL with natural policy gradients can effectively scale up to complex manipulation tasks with a high-dimensional 24-DoF hand, and solve them from scratch in simulated experiments. Furthermore, with the use of a small number of human demonstrations, the sample complexity can be significantly reduced, and enable learning within the equivalent of a few hours of robot experience. We demonstrate successful policies for multiple complex tasks: object relocation, in-hand manipulation, tool use, and door opening.

Temporal Difference Models: Model-Free Deep RL for Model-Based Control

Vitchyr Pong*, Shixiang Gu*, Murtaza Dalal, Sergey Levine

Model-free reinforcement learning (RL) has been proven to be a powerful, general tool for learning complex behaviors. However, its sample efficiency is often impractically large for solving challenging real-world problems, even for off-policy algorithms such as Q-learning. A limiting factor in classic model-free RL is that the learning signal consists only of scalar rewards, ignoring much of the rich information contained in state transition tuples. Model-based RL uses this information, by training a predictive model, but often does not achieve the same asymptotic performance as model-free RL due to model bias. We introduce temporal difference models (TDMs), a family of goal-conditioned value functions that unify model-free learning and model-based learning. TDMs combine the benefits of model-free and model-based RL: they leverage the rich information in state transitions to learn very efficiently, while still attaining
asymptotic performance that exceeds that of direct model-based RL methods. Our experimental results show that, on a range of continuous control tasks, TDMs provide a substantial improvement in efficiency compared to state-of-the-art model-based and model-free methods. Paper: coming soon! Code: coming soon!

Learning Deep Composable Maximum-Entropy Policies for Real-World Robotic Manipulation

Tuomas Haarnoja, Vitchyr Pong, Aurick Zhou, Murtaza Dalal, Pieter Abbeel, Sergey Levine

Model-free deep reinforcement learning has been shown to exhibit good performance in domains ranging from video games to simulated robotic manipulation and locomotion. However, model-free methods are known to perform poorly when the interaction time with the environment is limited, as is the case for most real-world robotic tasks. In this paper, we study how maximum-entropy policies, trained using an algorithm called soft Q-learning, can be applied to real-world robotic manipulation. The application of this method to real-world manipulation is facilitated by two important features of soft Q-learning. First, soft Q-learning can learn multimodal exploration strategies by learning policies represented by expressive energy-based models. Second, we show that policies learned with soft Q-learning can be composed to create new policies, and that the reduction in performance from combining policies in this way, versus training new policies from scratch, can be bounded in terms of the divergences between the composed policies. This compositionality provides an especially valuable tool for real-world manipulation, where constructing new policies by composing existing skills can provide a large gain in efficiency over training from scratch. Our experimental evaluation demonstrates that soft Q-learning is substantially more sample efficient than prior model-free deep reinforcement learning methods, and that compositionality can be performed for both simulated and real-world tasks.

Paper: coming soon! Code: coming soon!
Divide and Conquer Reinforcement Learning

Dibya Ghosh, Avi Singh, Larry Yang, Aravind Rajeswaran, Vikash Kumar, Sergey Levine

Standard model-free deep reinforcement learning (RL) algorithms sample a new initial state for each trial, allowing them to optimize policies that can perform well even in highly stochastic environments. However, problems that exhibit considerable initial state variation typically produce high-variance gradient estimates for model-free RL, making direct policy or value function optimization challenging. In this paper, we develop a novel algorithm that instead optimizes an ensemble of policies, each on a different ‘slice’ of the state space, and gradually unifies them into a single policy that can succeed on the whole state space. This approach, which we term divide and conquer RL, is able to solve complex tasks where conventional deep reinforcement learning methods are ineffective. Our results show that divide and conquer RL greatly outperforms conventional policy gradient methods on challenging grasping, manipulation, and locomotion tasks, and exceeds the performance of a variety of prior methods.

Paper: coming soon!