

Current Deep RL Frontiers (+pointers to some representative recent work)

■ Off-policy Policy Gradients / Off-policy Actor Critic / Connect with Q-Learning

- DDPG [Lillicrap et al, 2015]; Q-prop [Gu et al, 2016]; Doubly Robust [Dudik et al, 2011]; Deep Energy Q [Haarnoja*, Tang* et al, 2016]
- PGQ [O'Donoghue et al, 2016]; ACER [Wang et al, 2016]; Q(λ) [Harutyunyan et al, 2016]; Retrace(λ) [Munos et al, 2016], Equivalence PG and Soft-Q [Schulman et al, 2017],...

■ Exploration

- VIME [Houthoofd et al, 2016]; Count-Based Exploration [Bellemare et al, 2016]; #Exploration [Tang et al, 2016]; Curiosity [Schmidhuber, 1991]; Parameter Space Noise for Exploration [Plappert et al, 2017]; Noisy Networks [Fortunato et al, 2017]

■ Auxiliary objectives

- Learning to Navigate [Mirowski et al, 2016]; RL with Unsupervised Auxiliary Tasks [Jaderberg et al, 2016], ...

■ Multi-task and transfer (incl. sim2real)

- DeepDriving [Chen et al, 2015]; Progressive Nets [Rusu et al, 2016]; Flight without a Real Image [Sadeghi & Levine, 2016]; Sim2Real Visuomotor [Tzeng et al, 2016]; Sim2Real Inverse Dynamics [Christiano et al, 2016]; Modular NNs [Devin*, Gupta*, et al 2016]; Domain Randomization [Tobin et al, 2017]

■ Language

- Learning to Communicate [Foerster et al, 2016]; Multitask RL w/Policy Sketches [Andreas et al, 2016]; Learning Language through Interaction [Wang et al, 2016]

Current Deep RL Frontiers (+pointers to some representative recent work)

■ Meta-RL / Learn-to-learn

- Learning to Learn by Gradient Descent by Gradient Descent [Andrychowicz et al 2016]; RL2: Fast RL through Slow RL [Duan et al., 2016]; Learning to Reinforcement Learn [Wang et al, 2016]; Learning to Experiment [Denil et al, 2016]; Learning to Learn for Black-Box Opt. [Chen et al, 2016], Model-Agnostic Meta-Learning (Finn et al, 2017) ...

■ 24/7 Data Collection

- Learning to Grasp from 50K Tries [Pinto&Gupta, 2015]; Learning Hand-Eye Coordination [Levine et al, 2016]; Learning to Poke by Poking [Agrawal et al, 2016]

■ Safety

- Survey: Garcia and Fernandez, JMLR 2015; Uncertainty-Aware RL for Collision Avoidance [Kahn et al, 2017]

■ Architectures

- Memory, Active Perception in Minecraft [Oh et al, 2016]; DRQN [Hausknecht&Stone, 2015]; Dueling Networks [Wang et al, 2016]; ...

■ Inverse RL

- Generative Adversarial Imitation Learning [Ho et al, 2016]; Guided Cost Learning [Finn et al, 2016]; MaxEnt Deep RL [Wulfmeier et al, 2016]; ...

■ Model-based RL

- Deep Visual Foresight [Finn & Levine, 2016]; Embed to Control [Watter et al., 2015]; Spatial Autoencoders Visuomotor Learning [Finn et al, 2015]; PILCO [Deisenroth et al, 2015]

■ Hierarchical RL

- Modulated Locomotor Controllers [Heess et al, 2016]; STRAW [Vezhnevets et al, 2016]; Option-Critic [Bacon et al, 2016]; h-DQN [Kulkarni et al, 2016]; Hierarchical Lifelong Learning in Minecraft [Tessler et al, 2016]; Feudal Networks [Vezhnevets et al, 2017]; Stochastic NNs [Florensa et al, 2017]