Unsupervised Perceptual Rewards for Imitation Learning

Pierre Sermanet, Kelvin Xu, Sergey Levine (2017)
Presented by Kevin Chen <kxc2103@columbia.edu>
Background
Barriers to reinforcement learning

- Needs many iterations and demonstrations
- Hand-designed reward functions
  - Manually reward intermediate steps
- Really sucks compared to human performance
Problem statement: learn a perceptual reward function for robotic learning from a few video demonstration
Entire training set for pouring!
Convolutional Neural Networks
Convolution
Convolutional neural network
Hubel and Wiesel
Convolutional neural network
Architecture
Architecture

- Use CNN to extract features
- Cluster features to find important steps in the task
- Robot rewarded for matching feature vector of each step
- Run RL as usual
CNN pre-trained on ImageNet

**Synset:** means
**Definition:** an instrumentality for accomplishing some end.

**Synset:** plant, works, industrial plant
**Definition:** buildings for carrying on industrial labor; "they built a large plant to manufacture automobiles".

**Synset:** food, nutrient
**Definition:** any substance that can be metabolized by an animal to give energy and build tissue.

**Synset:** building, edifice
**Definition:** a structure that has a roof and walls and stands more or less permanently in one place; "there was a three-story building on the corner"; "it was an imposing edifice".
CNN pre-trained on ImageNet

(Many layers omitted)
Map video into CNN feature space
Clustering algorithm finds steps

Step 1

Step 2

Step 3
Fit Gaussian to each step’s vector

Step 1

Step 2

Step 3
Reward function is value of Gaussian

\[ R(a) = \sum_{i=2}^{n} R_i(a) \times 2^{(i-1)} \]
Do RL as usual

\[ R(a) = \sum_{i=2}^{n} R_i(a) \times 2^{(i-1)} \]

\[ R(a) = 0.72 \]
Summary

- Generate multi-step reward function from images
- CNN enables generalization
  - Tiny training sets
  - Human hand vs. robot grasper
My Opinions

• Cool way to design reward functions, a huge RL barrier
• Great application of pre-trained neural networks
• Vague writing
• Shortcomings hidden in appendix
Questions?