Learning Visual Servoing with Deep Features and Fitted Q-Iteration

Alex X. Lee¹, Sergey Levine¹, Pieter Abbeel²,¹,³

¹UC Berkeley, ²OpenAI, ³International Computer Science Institute
Motivation
Deep Neural Networks in Computer Vision

image classification

semantic segmentation

object detection

object tracking

shaking animal

woman

AlexNet
Outline

- Introduction
- **Reinforcement learning and deep reinforcement learning**
  - Visual servoing
  - Learn visual servoing with reinforcement learning
    - Policy optimization
    - Combine value and model based RL
      - Learn visual feature dynamics
      - Learn servoing policy with fitted Q-iteration
  - Comparison to prior methods
- Conclusion
What is Reinforcement Learning?

- **Agent**
- **Environment**

- action $u$
- state $s$, reward $r$
Reinforcement Learning Approaches

\[ \pi(u_t|s_t) = \text{arg max}_u Q(s_t, u) \]

- **Agent**
- **Environment**
- **State** \( s_t \), **Reward** \( r \)

**Model-Free Approaches**
- **Policy Optimization**
  - \( \pi \) (policy)
  - \( Q(s_t, u) \) (Q-value)
  - \( u_t \) (action)
  - High sample complexity
  - Policy might be simpler than value or model

**Model-Based Approaches**
- **Value-Based**
  - \( Q(s_t, u_t) \) (Q-value)
  - \( s_t \rightarrow u_t \rightarrow s_{t+1}, r_{t+1} \)
  - Medium sample complexity
  - Challenge for continuous and high-dimensional action spaces

**Environment Model**
- \( \text{model free} \)
- \( \text{model based} \)
- \( \text{low sample complexity} \)
- Relies on a good model
What is Deep Reinforcement Learning?

- Model-free policy optimization
  - $s_t \rightarrow \pi \rightarrow u_t$

- Model-based value-based
  - $s_t \rightarrow Q \rightarrow Q$-value
  - $s_t \rightarrow r_{t+1}, s_{t+1}$

- Model-based
  - $s_t \rightarrow$ environment model
Examples of Deep Reinforcement Learning

Silver et al., 2014 (DPG)
Lillicrap et al., 2015 (DDPG)

Mnih et al., 2015 (DQN)
Mnih et al., 2016 (A3C)

Gu*, Holly*, et al., 2016
Levine*, Finn*, et al., 2016 (GPS)

Sadeghi et al., 2017 (CAD)$^2$RL

Schulman et al., 2016 (TRPO + GAE)

Tamar et al., 2016 (VIN)
Deep Reinforcement Learning for Robotics


Levine*, Finn*, et al, 2016 (GPS)

Sadeghi et al, 2017 (CAD)^2RL
Outline

- Introduction
- Reinforcement learning and deep reinforcement learning
- **Visual servoing**
  - Learn visual servoing with reinforcement learning
    - Policy optimization
    - Combine value and model based RL
      - Learn visual feature dynamics
      - Learn servoing policy with fitted Q-iteration
  - Comparison to prior methods
- Conclusion
Visual Servoing

current observation

goal observation
Examples of Visual Servoing: Manipulation

Source: SeRViCE Lab, UT Dallas
Examples of Visual Servoing: Surgical Tasks

Source: Kehoe et al. 2016
Examples of Visual Servoing: Space Docking

Source: NASA
Outline

- Introduction
- Reinforcement learning and deep reinforcement learning
- Visual servoing
- **Learn visual servoing with reinforcement learning**
  - Policy optimization
  - Combine value and model based RL
    - Learn visual feature dynamics
    - Learn servoing policy with fitted Q-iteration
- Comparison to prior methods
- Conclusion
Learning Visual Servoing with Reinforcement Learning

- **Agent**
  - Linear and angular velocities
  - Action $u$

- **Environment**
  - State $s$, reward $r$
  - Distance to desired pose relative to car
  - Current and goal pose relative to car
  - Current and goal image observation

The diagram illustrates the interaction between the agent and the environment, showing how the agent's actions are influenced by the state of the environment and how rewards are calculated based on the distance to the desired pose.
Outline

- Introduction
- Reinforcement learning and deep reinforcement learning
- Visual servoing
- Learn visual servoing with reinforcement learning
  - Policy optimization
  - Combine value and model based RL
    - Learn visual feature dynamics
    - Learn servoing policy with fitted Q-iteration
- Comparison to prior methods
- Conclusion
Learning Visual Servoing with Policy Optimization

Policy optimization

\[ S_t \rightarrow \pi \rightarrow u_t \]

trained with more than 20000 trajectories!

example executions of trained policy
Outline

- Introduction
- Reinforcement learning and deep reinforcement learning
- Visual servoing
- Learn visual servoing with reinforcement learning
  - Policy optimization
  - **Combine value and model based RL**
    - Learn visual feature dynamics
    - Learn servoing policy with fitted Q-iteration
- Comparison to prior methods
- Conclusion
State-action value based RL: $\pi(s_t) = \arg \max_u Q(s_t, u)$
Combining Value and Model Based Reinforcement Learning

State-action value based RL:  \( \pi(s_t) = \arg \min_u -Q(s_t, u) \)

Visual servoing:  \( \pi(s_t) = \arg \min_u ||x_* - f(x_t, u_t)||^2 - Q(s_t, u) \)

\( \pi(s_t) = \text{arg min}_{u} \) dynamics function
Servoing with Visual Dynamics Model

\[ x_t \quad \xrightarrow{f} \quad f(x_t, u_t) \quad \xrightarrow{||x_* - f(x_t, u_t)||^2} \quad x_* \]

- current observation
- predicted observation
- goal observation
Features from Dilated VGG-16 Convolutional Neural Network


Servoing with Visual Dynamics Model

\[ x_t \quad \xrightarrow{f} \quad f(x_t, u_t) \quad \xrightarrow{\text{predicted observation}} \quad ||x_* - f(x_t, u_t)||^2 \quad \xrightarrow{\text{goal observation}} \quad x_* \]

current observation

predicted observation

goal observation
Servoing with Visual Dynamics Model

\[
\pi(x_t, x_*) = \arg\min_u \left\| y_* - f(y_t, u_t) \right\|^2_w - Q_w(s_t, u)
\]
Outline

- Introduction
- Reinforcement learning and deep reinforcement learning
- Visual servoing
- Learn visual servoing with reinforcement learning
  - Policy optimization
  - Combine value and model based RL
    - Learn visual feature dynamics
    - Learn servoing policy with fitted Q-iteration
- Comparison to prior methods
- Conclusion
Feature Dynamics: Multiscale Bilinear Model
Feature Dynamics: Multiscale Bilinear Model
Outline

- Introduction
- Reinforcement learning and deep reinforcement learning
- Visual servoing
- Learn visual servoing with reinforcement learning
  - Policy optimization
  - Combine value and model based RL
    - Learn visual feature dynamics
    - **Learn servoing policy with fitted Q-iteration**
- Comparison to prior methods
- Conclusion
π(s_t) = \arg\min_u ||y_\ast - f(y_t, u_t)||^2_w - Q_w(s_t, u)
Learning Visual Servoing with Deep Feature Dynamics and FQI

value based + visual dynamics model

\[
Q(s_t, u_t) \rightarrow \text{Q-value}
\]

trained with only 20 trajectories!

example executions of trained policy
Outline

- Introduction
- Reinforcement learning and deep reinforcement learning
- Visual servoing
- Learn visual servoing with reinforcement learning
  - Policy optimization
  - Combine value and model based RL
    - Learn visual feature dynamics
    - Learn servoing policy with fitted Q-iteration
- Comparison to prior methods
- Conclusion
Comparison to Prior Methods

Feature Representation and Optimization Method
Conclusion

- Deep reinforcement learning allows us to learn complex robot policies that can process complex visual inputs
- Combine value based and model based for better sample complexity
- Visual servoing
  - Learn visual feature dynamics
  - Learn Q-values with fitted Q-iteration
Thank You

Resources

Paper: arxiv.org/abs/1703.11000
Code: github.com/alexlee-gk/visual_dynamics
Servoing benchmark code: github.com/alexlee-gk/citysim3d
More videos: rll.berkeley.edu/visual_servoing

Acknowledgements