Learning Visual Servoing with Deep Features and Fitted Q-Iteration

Alex X. Lee[1],

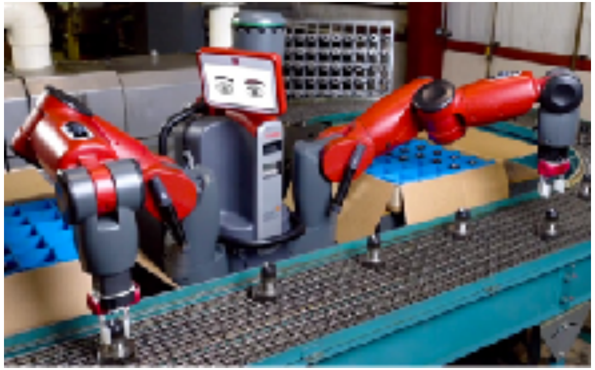
Sergey Levine[1], Pieter Abbeel[2,1,3]

[1]UC Berkeley, [2]OpenAI,

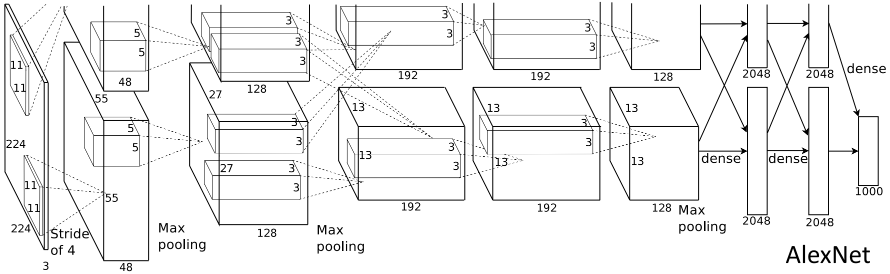
[3]International Computer Science Institute

BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Motivation



Deep Neural Networks in Computer Vision



object detection

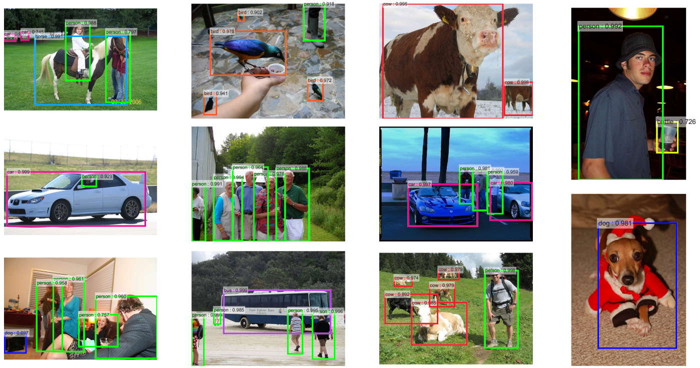
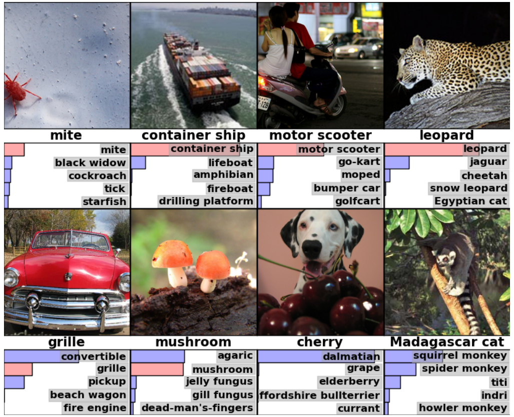
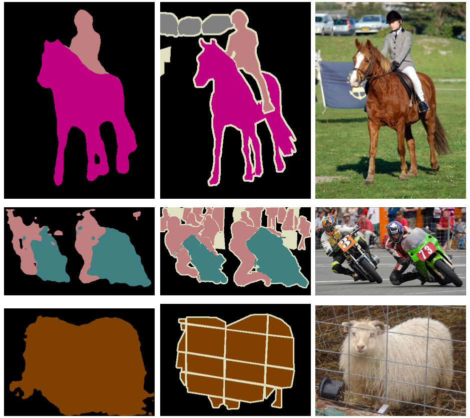


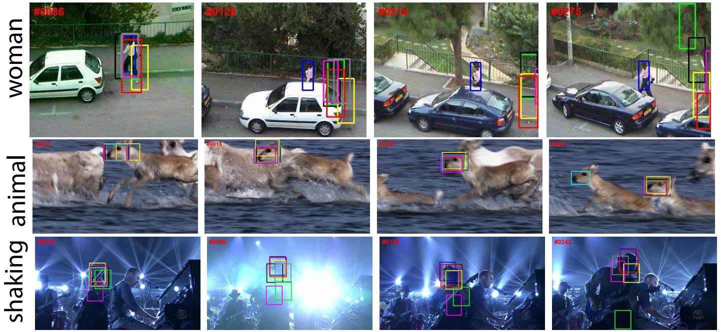
image classification



semantic segmentation



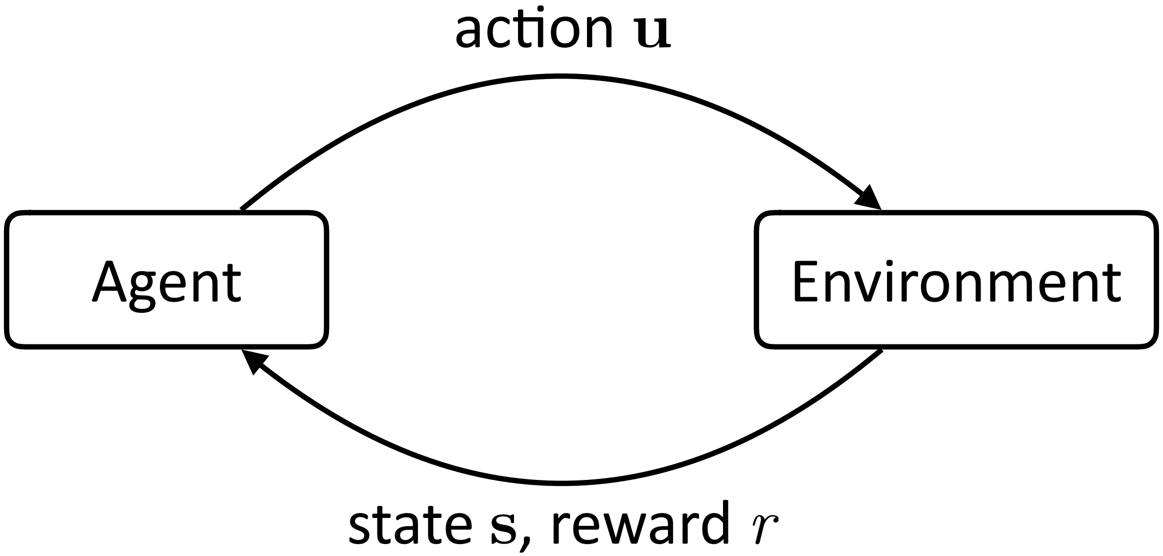
object tracking



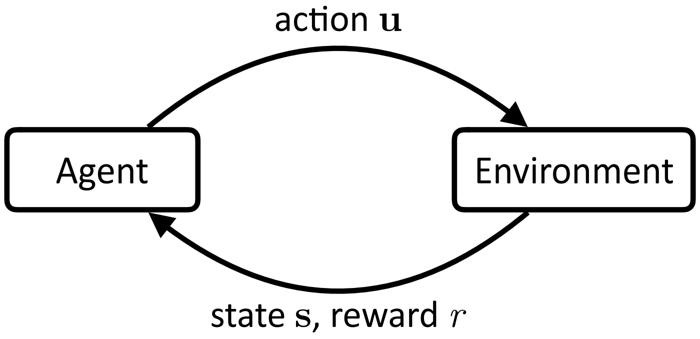
Outline

- 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

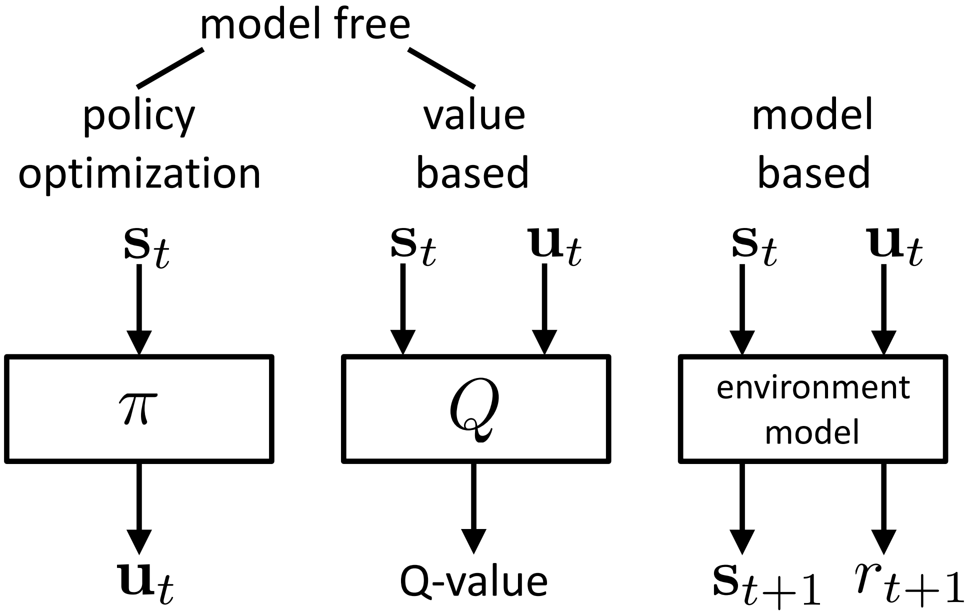
What is Reinforcement Learning?



Reinforcement Learning Approaches



$$\pi(\mathbf{u}_t|\mathbf{s}_t) = \arg\max_{\mathbf{u}} Q(\mathbf{s}_t, \mathbf{u})$$

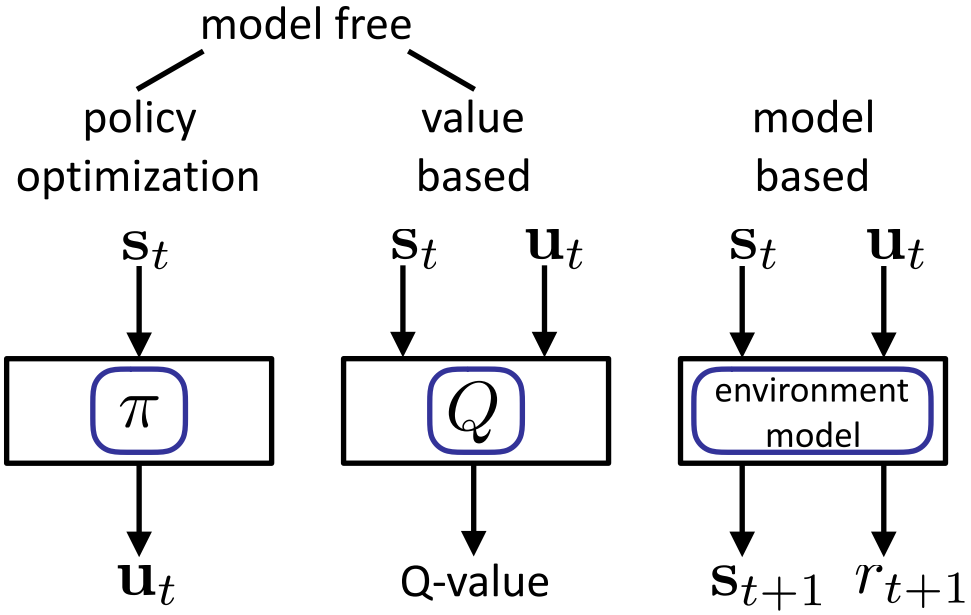


Policy optimization
- high sample complexity
- policy might be simpler than value or model

Value based
- medium sample complexity
- challenge for continuous and high-dimensional action spaces

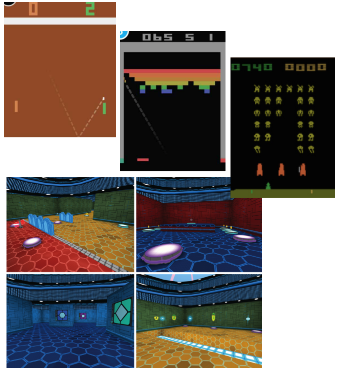
Model based
- low sample complexity
- relies on a good model

What is Deep Reinforcement Learning?

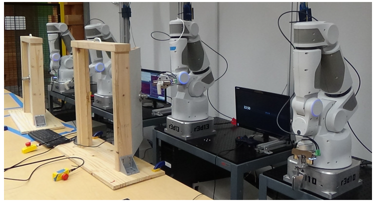


Examples of Deep Reinforcement Learning

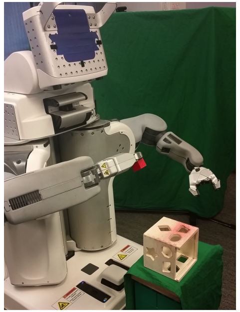
Silver et al, 2014
(DPG)
Lillicrap et al, 2015
(DDPG)



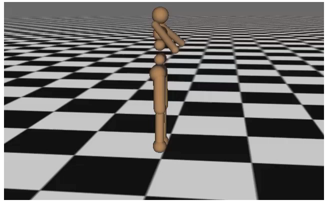
Mnih et al, 2015 (DQN)
Mnih et al, 2016 (A3C)



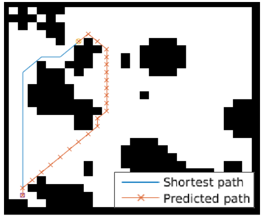
Gu*, Holly*, et al, 2016



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



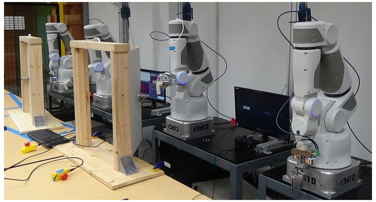
Schulman et al, 2016
(TRPO + GAE)



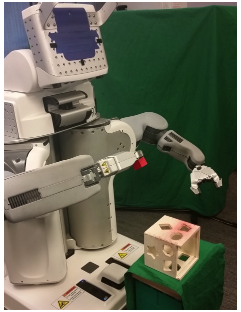
Tamar et al, 2016
(VIN)

Sadeghi et al, 2017 (CAD)^{2}RL

Deep Reinforcement Learning for Robotics



Gu*, Holly*, et al, 2016



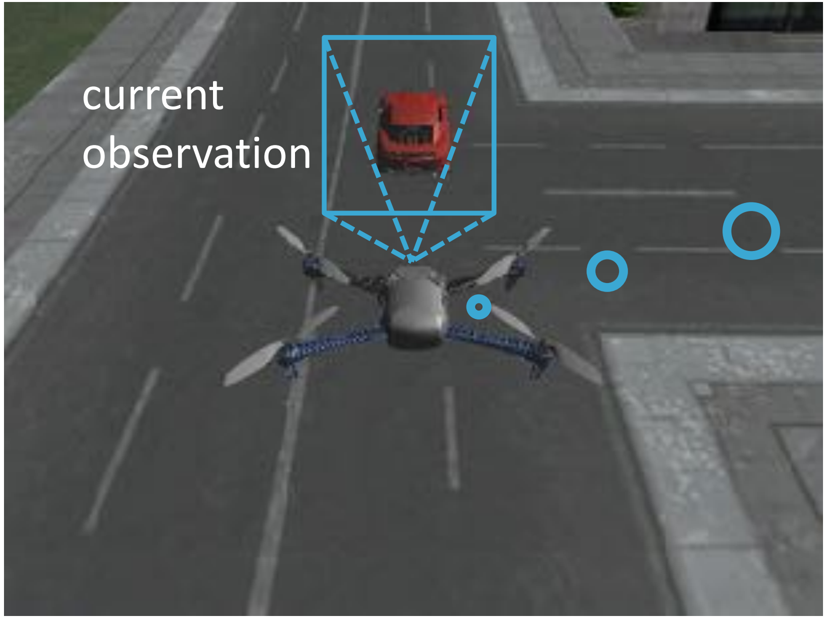
Levine*, Finn*, et al, 2016 (GPS)

Sadeghi et al, 2017 (CAD)^{2}RL

Outline

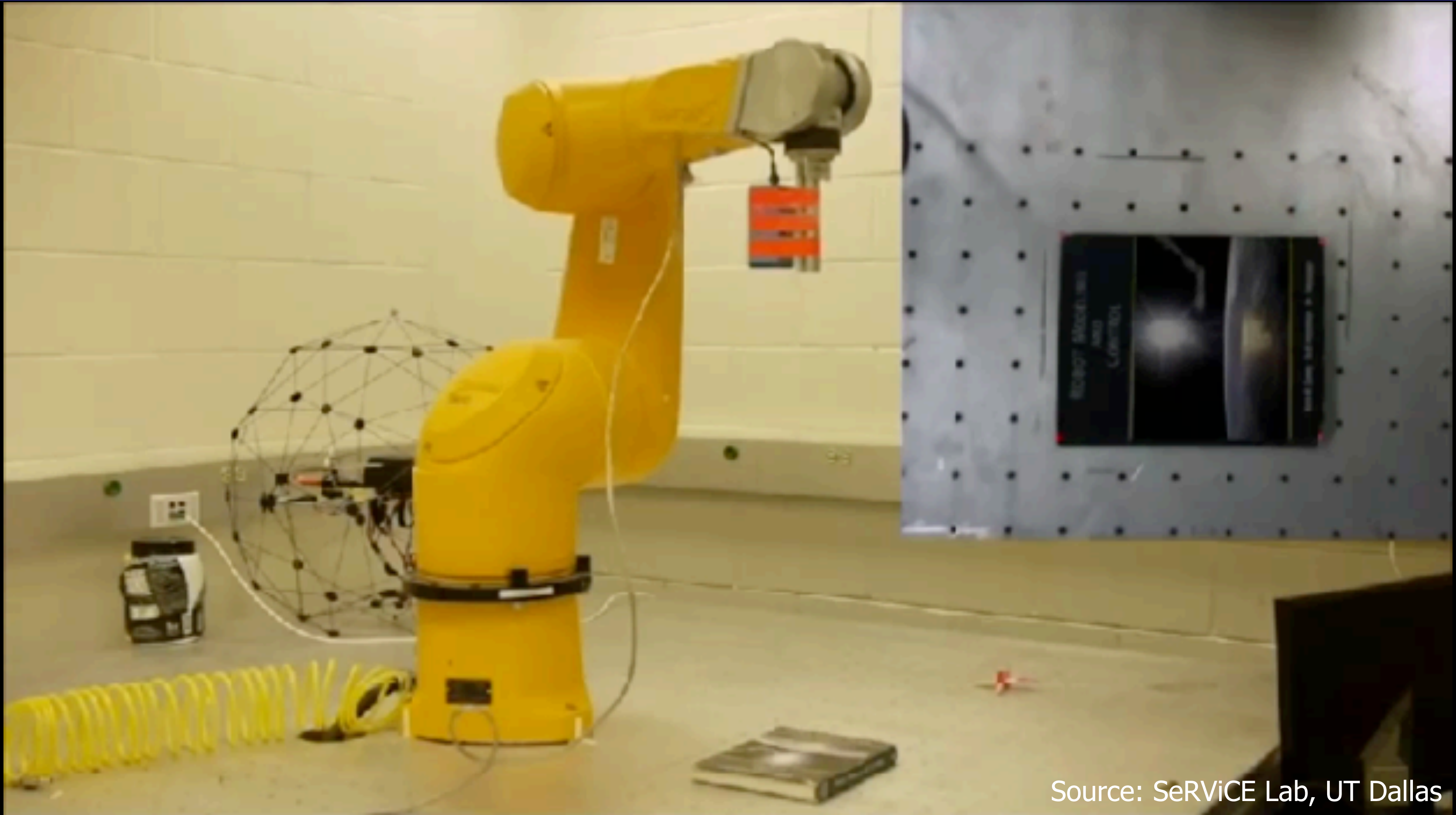
- Introduction
- Reinforcement learning and deep reinforcement learning
- **Visual servoing**
- Learn visual servoing with reinforcement learning
    - Policy optimization
    - Combine value and model based RL
        - Learn visual feature dynamics
        - Learn servoing policy with fitted Q-iteration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

Visual Servoing



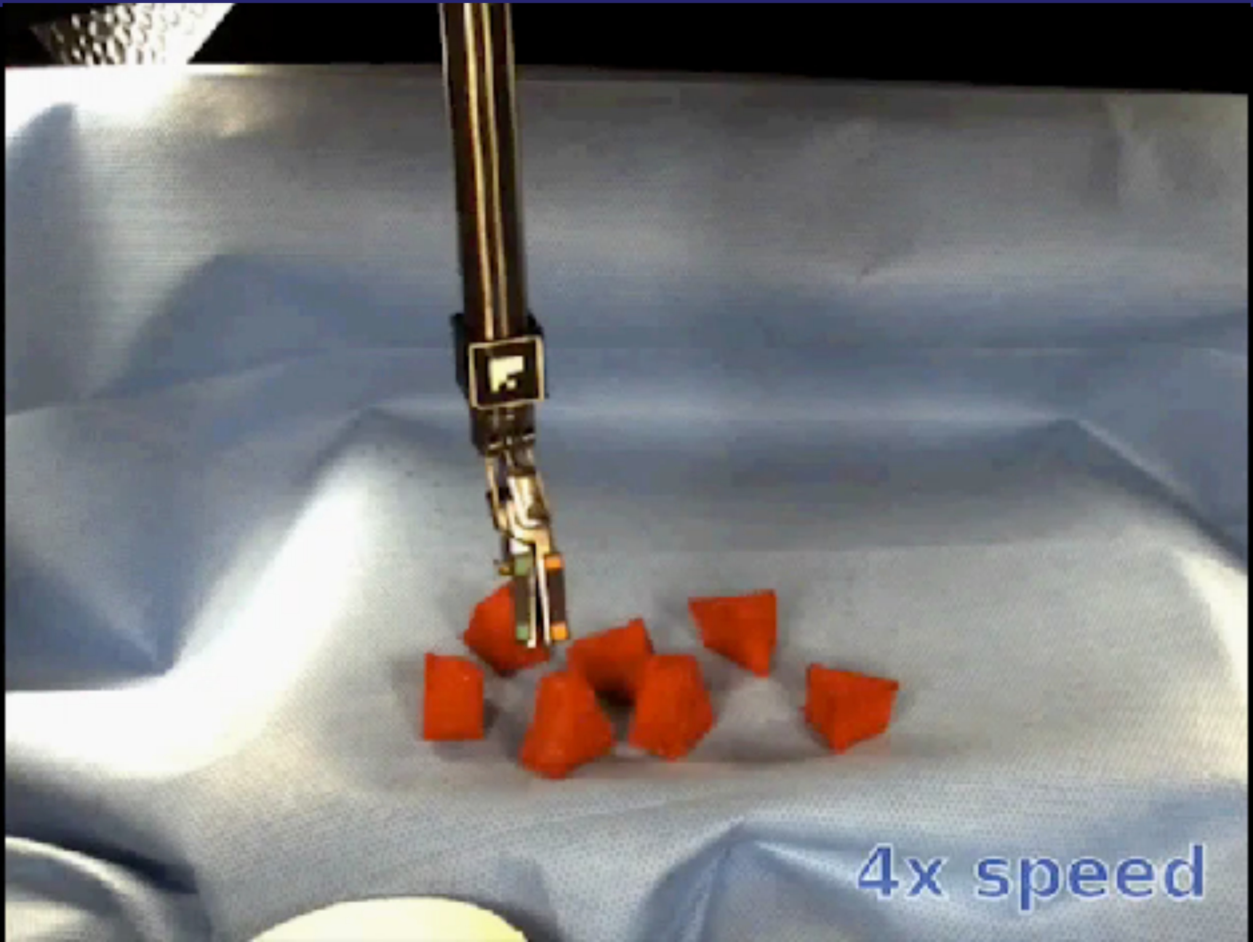
goal
observation

Examples of Visual Servoing: Manipulation



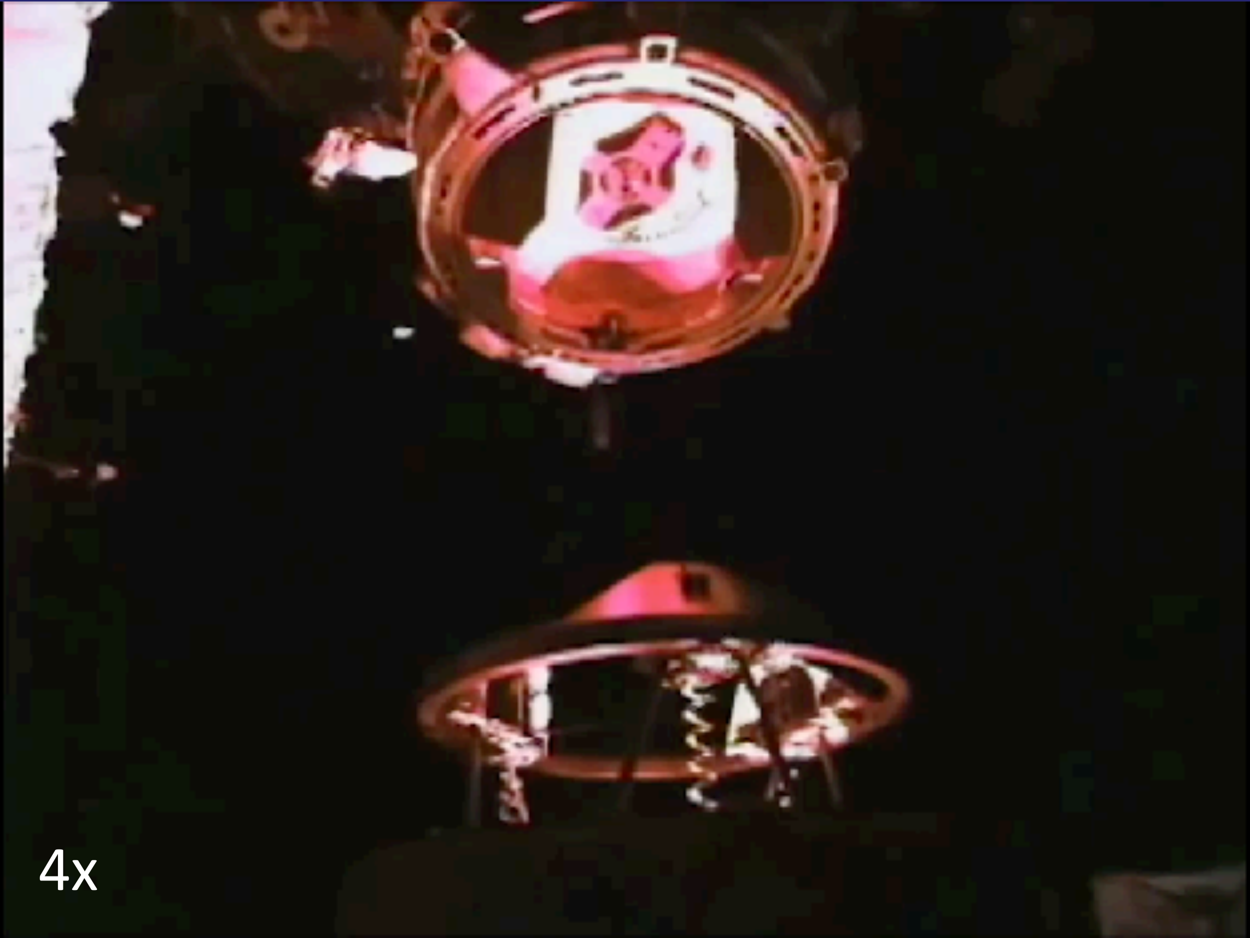
Source: SeRViCE Lab, UT Dallas

Examples of Visual Servoing: Surgical Tasks



Source:
Kehoe et al. 2016

Examples of Visual Servoing: Space Docking

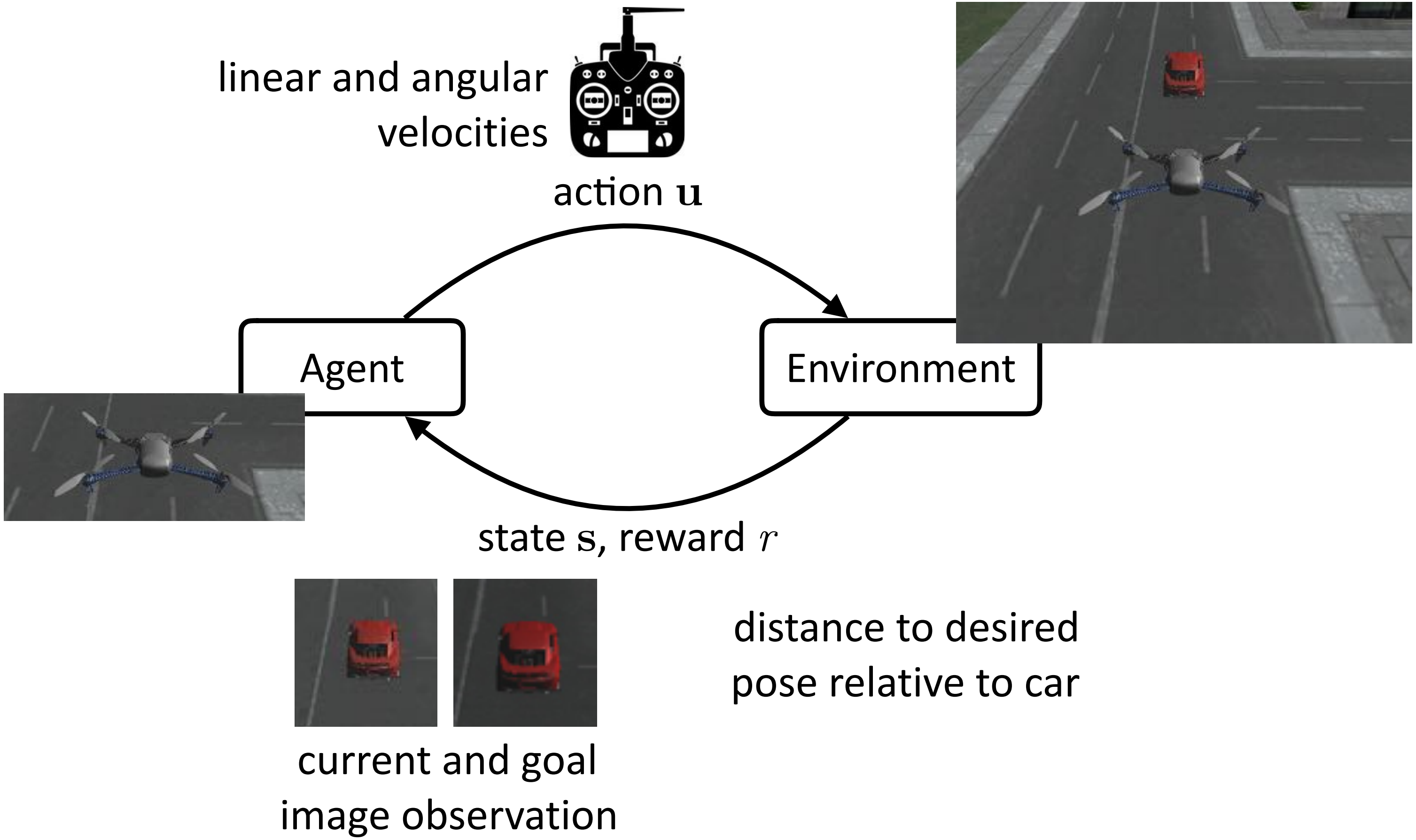


Source: NASA

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



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

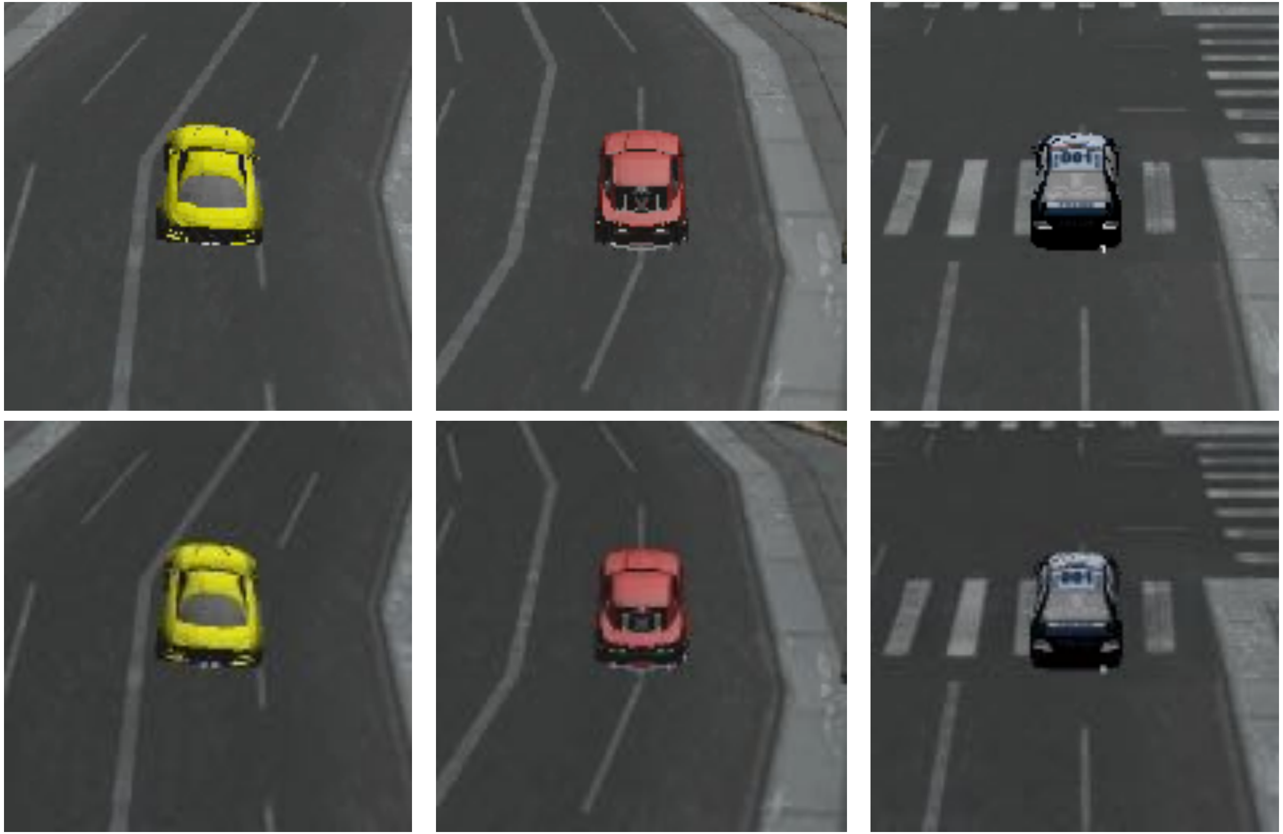
$$\mathbf{s}_t \rightarrow \boxed{\pi} \rightarrow \mathbf{u}_t$$

trained with more than 20000 trajectories!

example executions of trained policy

current observation

goal observation



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

Combining Value and Model Based Reinforcement Learning

State-action value based RL: $$\pi(\mathbf{s}_t) = \arg\max_{\mathbf{u}} Q(\mathbf{s}_t, \mathbf{u})$$

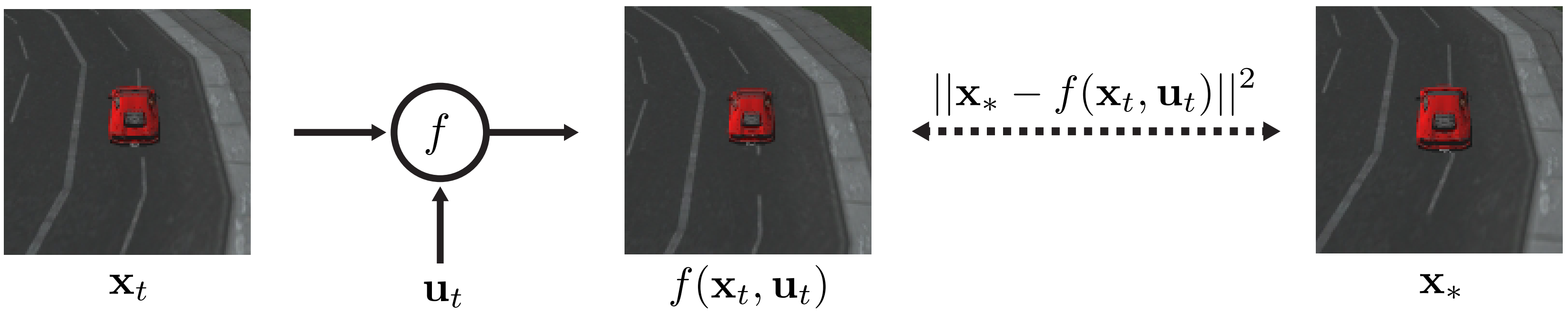
Combining Value and Model Based Reinforcement Learning

State-action value based RL: $$\pi(\mathbf{s}_t) = \arg\min_{\mathbf{u}} -Q(\mathbf{s}_t, \mathbf{u})$$

Visual servoing: $$\pi(\mathbf{s}_t) = \arg\min_{\mathbf{u}} \underbrace{||\mathbf{x}_* - f(\mathbf{x}_t, \mathbf{u}_t)||^2}_{-Q(\mathbf{s}_t, \mathbf{u})}$$

dynamics function

Servoing with Visual Dynamics Model

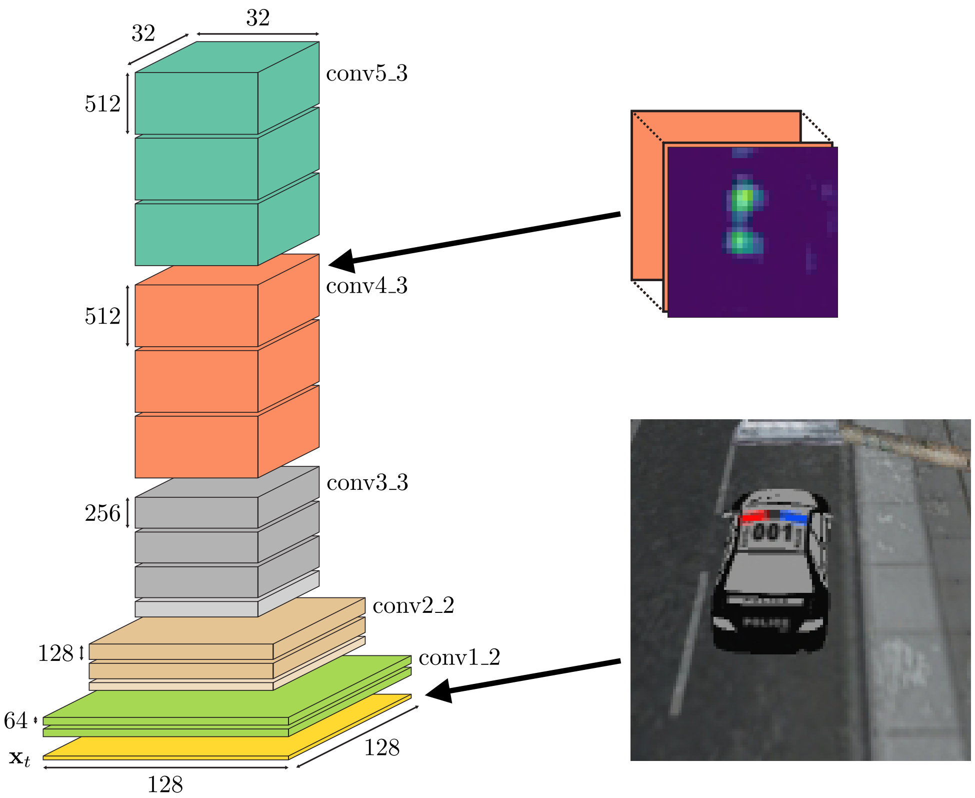


current observation

predicted observation

goal observation

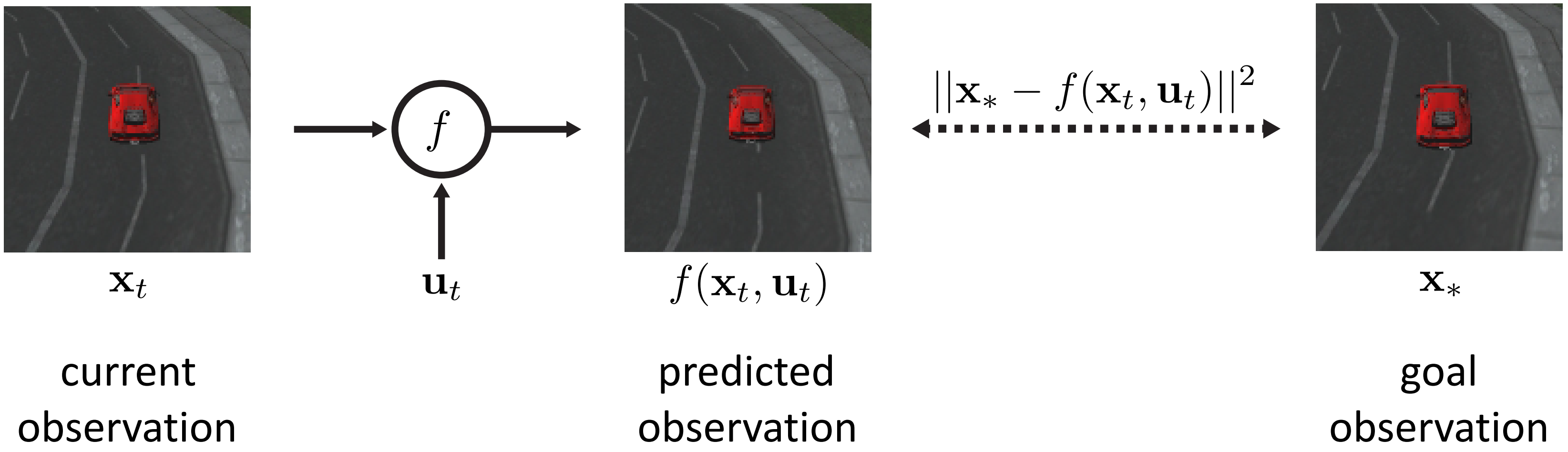
Features from Dilated VGG-16 Convolutional Neural Network



K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

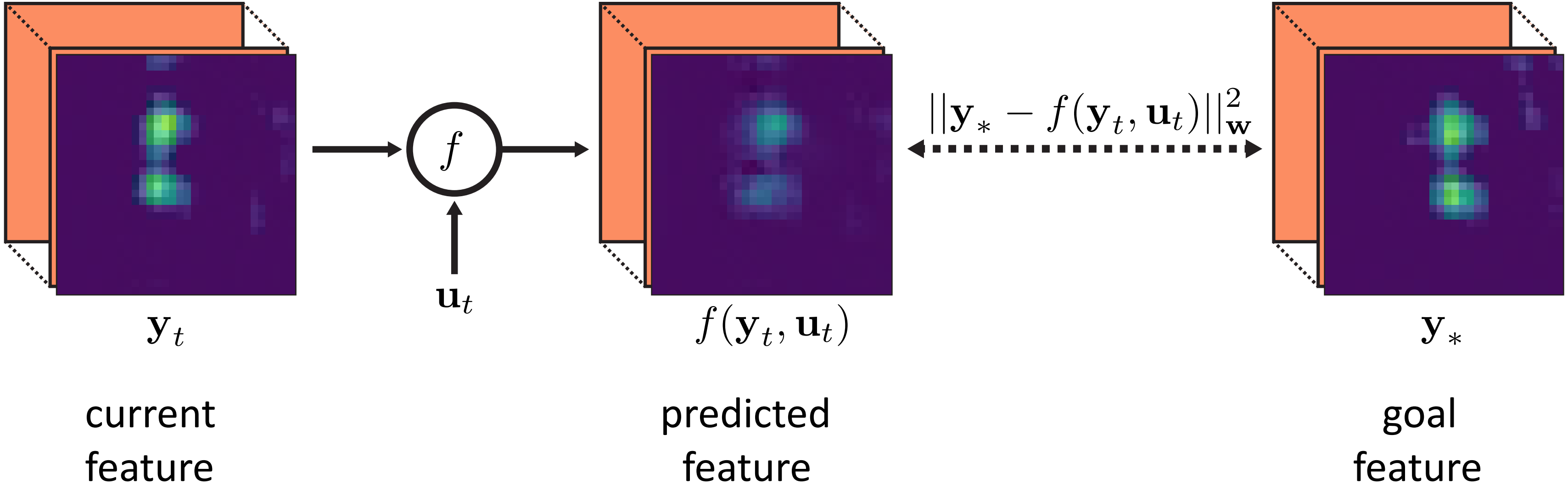
F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. In ICLR, 2016.

Servoing with Visual Dynamics Model



Servoing with Visual Dynamics Model

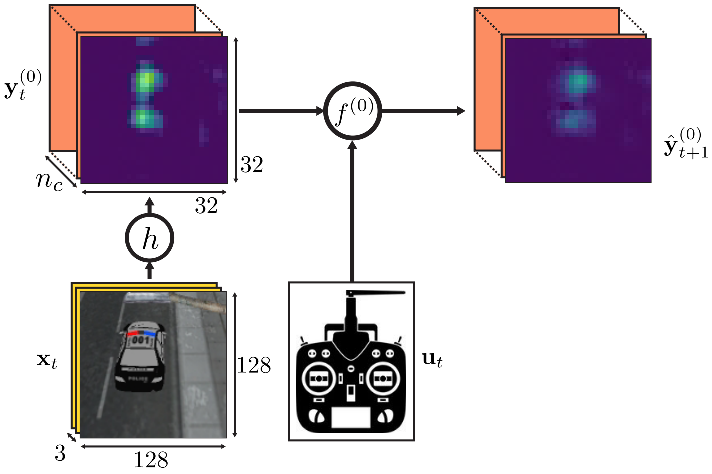
$$\pi(\mathbf{x}_t, \mathbf{x}_*) = \arg\min_{\mathbf{u}} \underbrace{\|\mathbf{y}_* - f(\mathbf{y}_t, \mathbf{u}_t)\|^2_{\mathbf{w}}}_{-Q_{\mathbf{w}}(\mathbf{s}_t, \mathbf{u})}$$



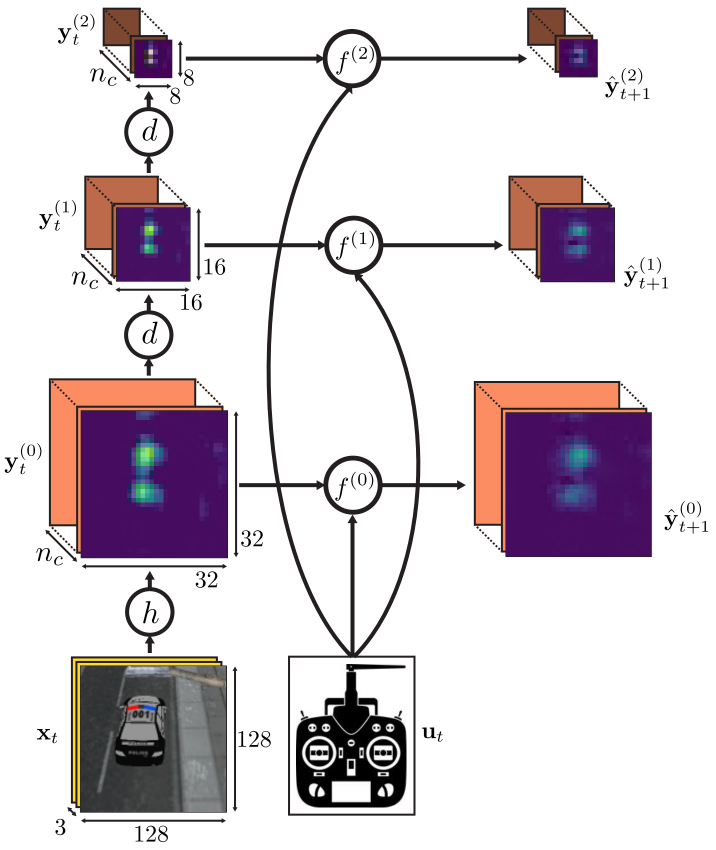
Outline

- 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

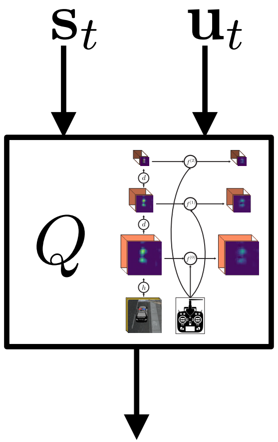
- Introduction
- Reinforcement learning and deep reinforcement learning
- Visual servoing
- Learn visual servoing with reinforcement learning
    - Policy optimization
    - Combine value and model based RL
        - Learn visual feature dynamics
        - **Learn servoing policy with fitted Q-iteration**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

Learning Model Based Policy with Fitted Q-Iteration

$$\pi(\mathbf{s}_t) = \arg\min_{\mathbf{u}} \underbrace{||\mathbf{y}_* - f(\mathbf{y}_t, \mathbf{u}_t)||^2_{\mathbf{w}}}_{-Q_{\mathbf{w}}(\mathbf{s}_t, \mathbf{u})}$$

Learning Visual Servoing with Deep Feature Dynamics and FQI

value based +
visual dynamics model



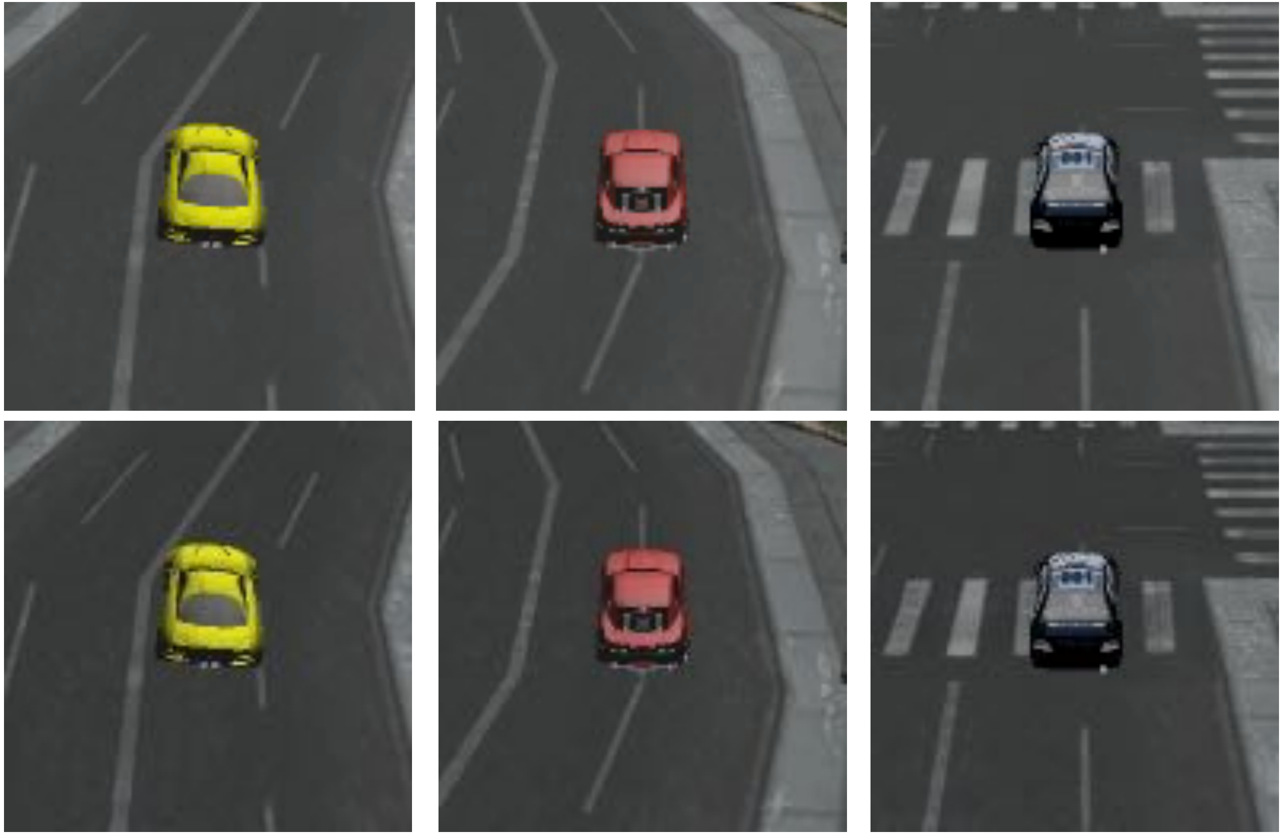
Q-value

trained with only
20 trajectories!

example executions of trained policy

current observation

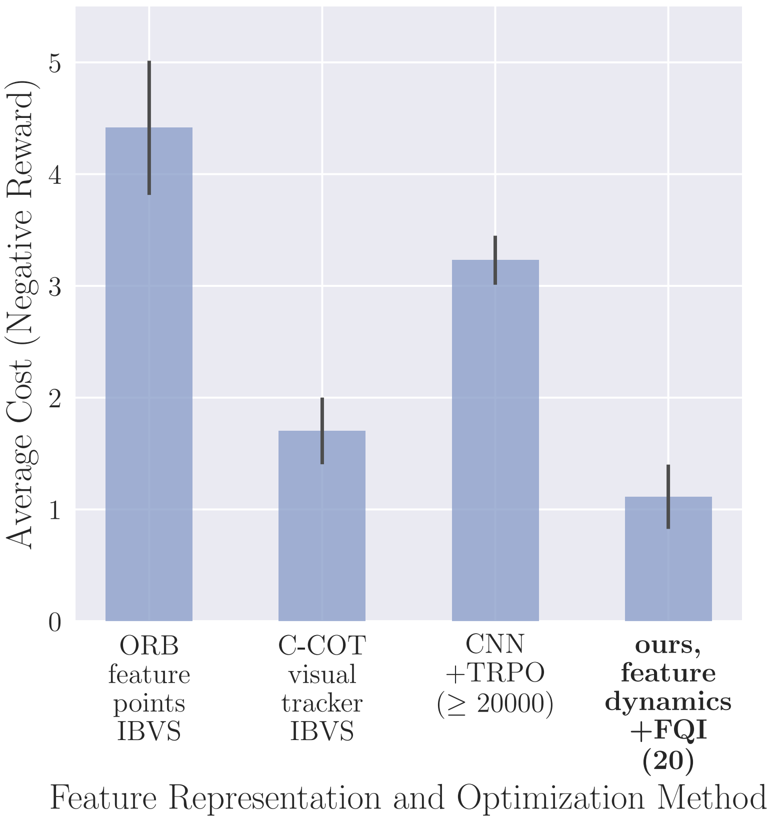
goal observation



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



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