DeepRob

Lecture 14
Imitation Learning I
University of Minnesota

HOW TO TRAIN YOUR DRAGON THE HIDDEN WORLD Kit Harington Auditions with Toothless

Project 3 — Releases today

- Instructions available on the website
  - Here: https://rpm-lab.github.io/CSCI5980-F24-DeepRob/projects/project3/
  - Uses PROPS Detection dataset
- Implement CNN for classification and Faster R-CNN for detection
- Autograder will be available soon!
- **Due Monday, October 28th 11:59 PM CT**

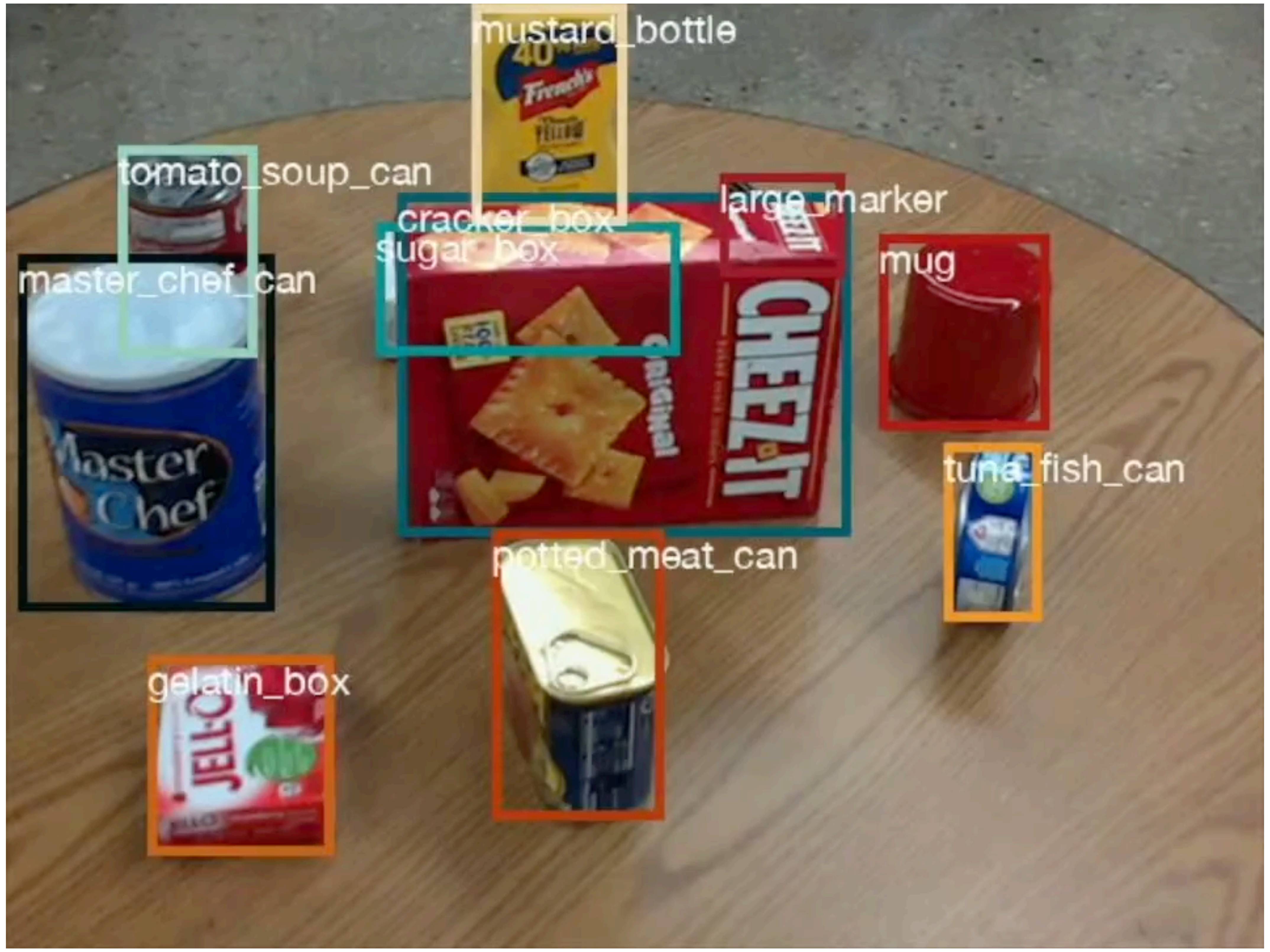
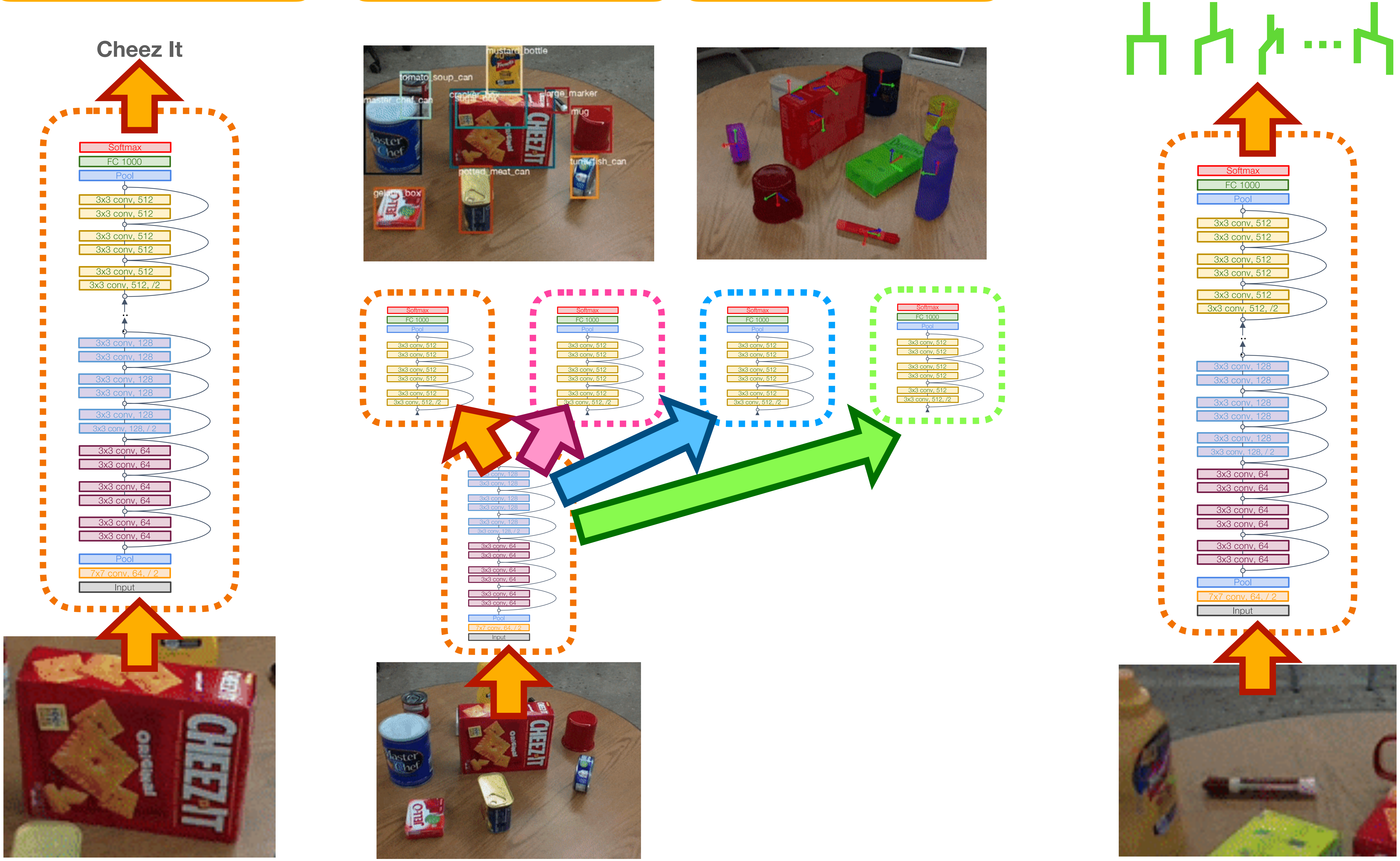


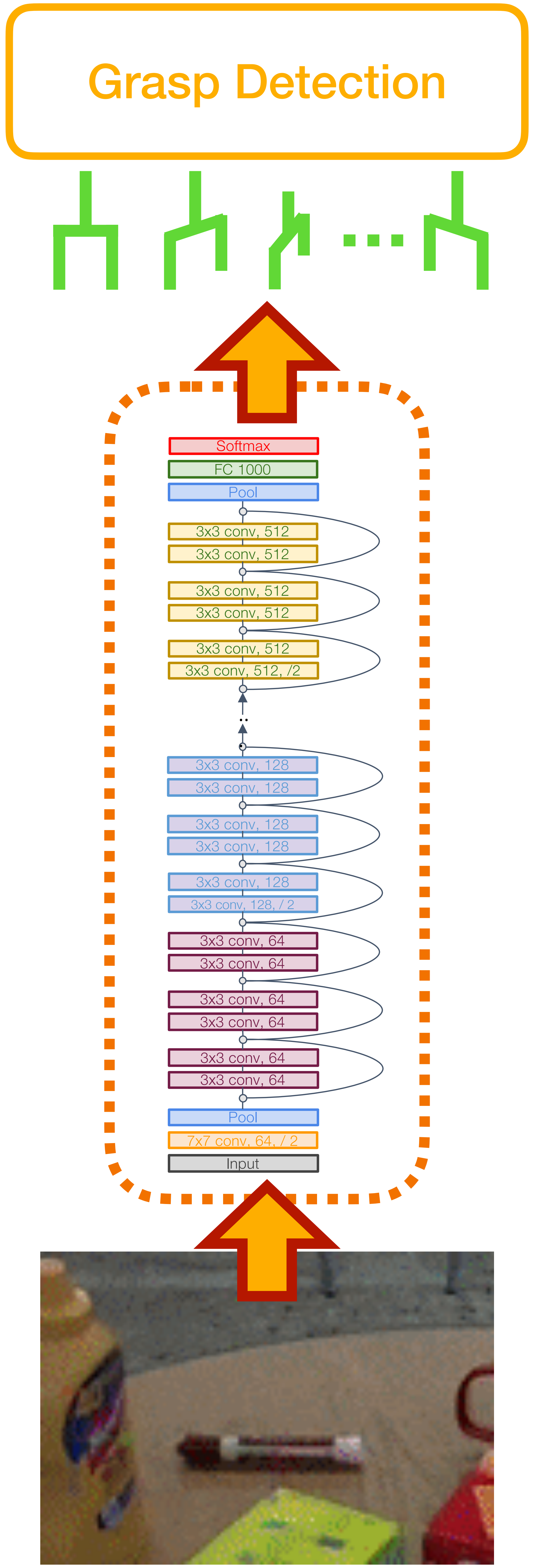
Image Classification

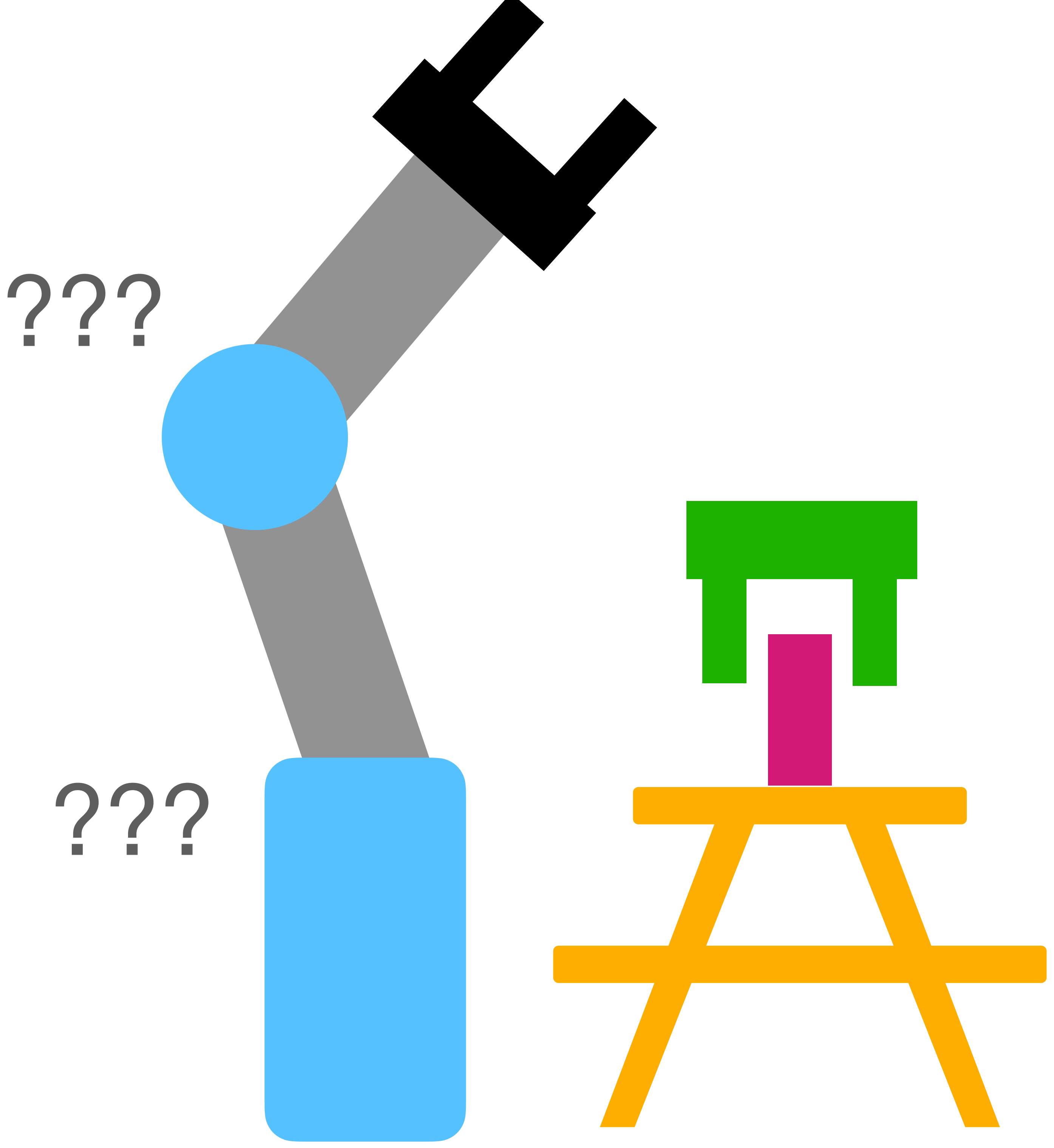
Object Detection

Segmentation & Pose Estimation

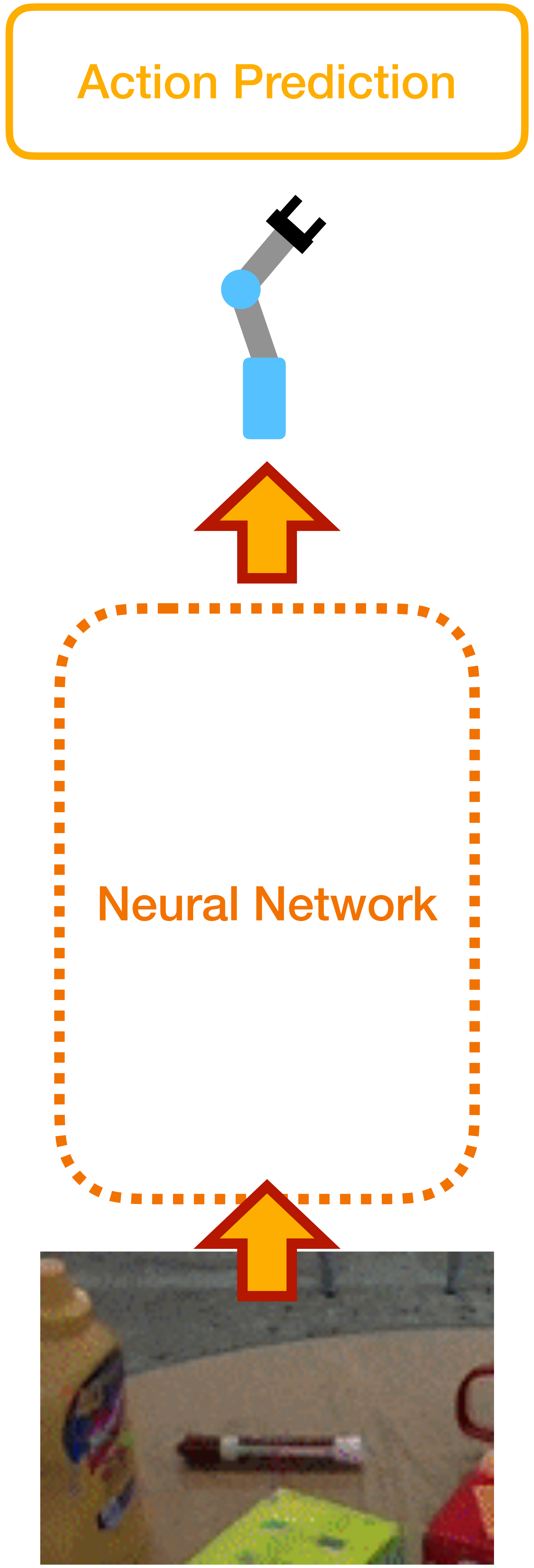
Grasp Detection

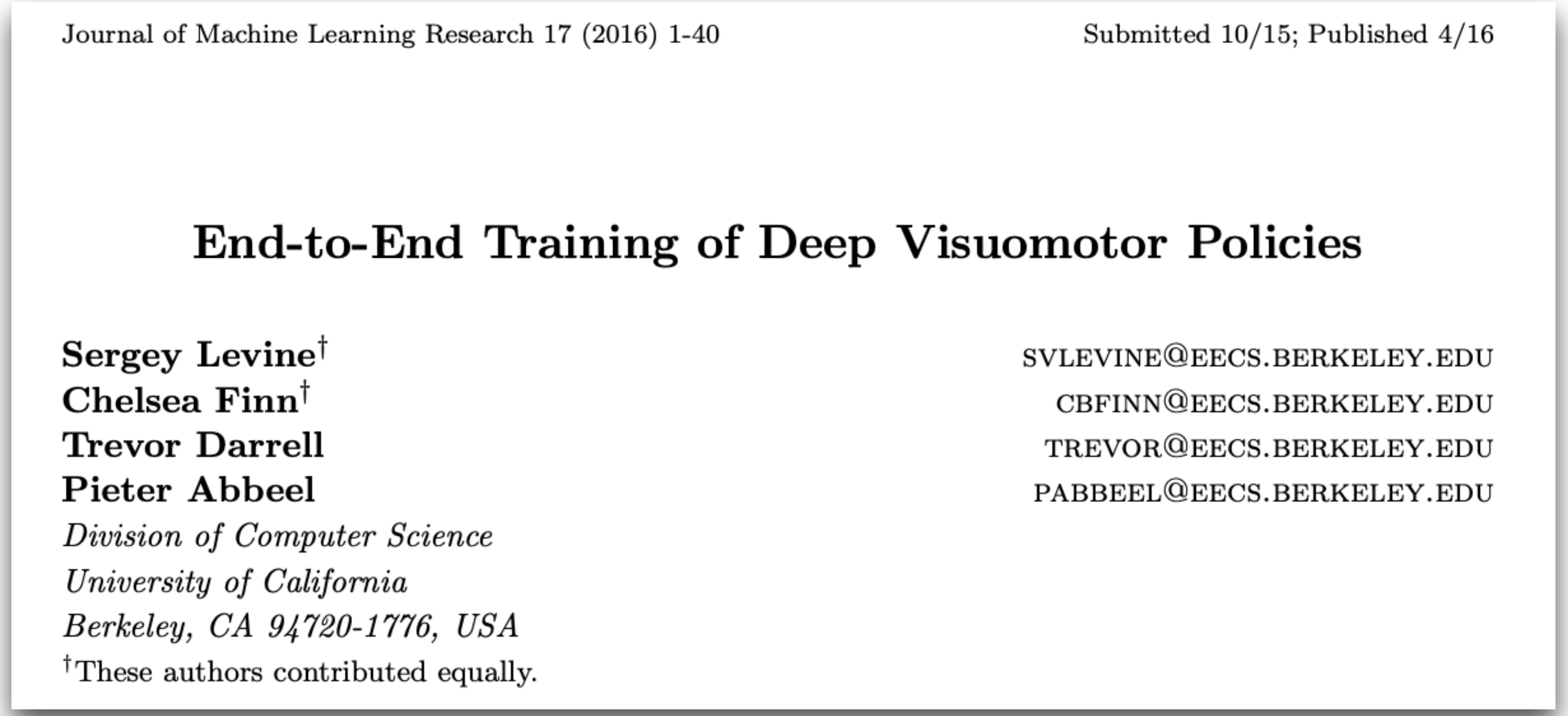






- After finding grasp poses, how to execute actions?
  - Remember! Inverse Kinematics (5551)
  - Remember! Motion Planning (5551)




Journal of Machine Learning Research 17 (2016) 1-40 Submitted 10/15; Published 4/16

End-to-End Training of Deep Visuomotor Policies

Sergey Levine[†] SVLEVINE@EECS.BERKELEY.EDU
Chelsea Finn[†] CBFINN@EECS.BERKELEY.EDU
Trevor Darrell TREVOR@EECS.BERKELEY.EDU
Pieter Abbeel PABBEEL@EECS.BERKELEY.EDU
Division of Computer Science
University of California
Berkeley, CA 94720-1776, USA
[†]These authors contributed equally.

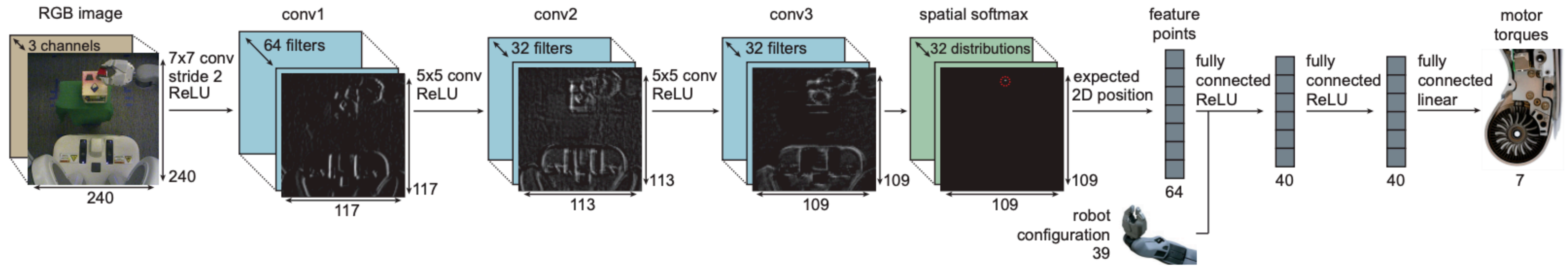
DR

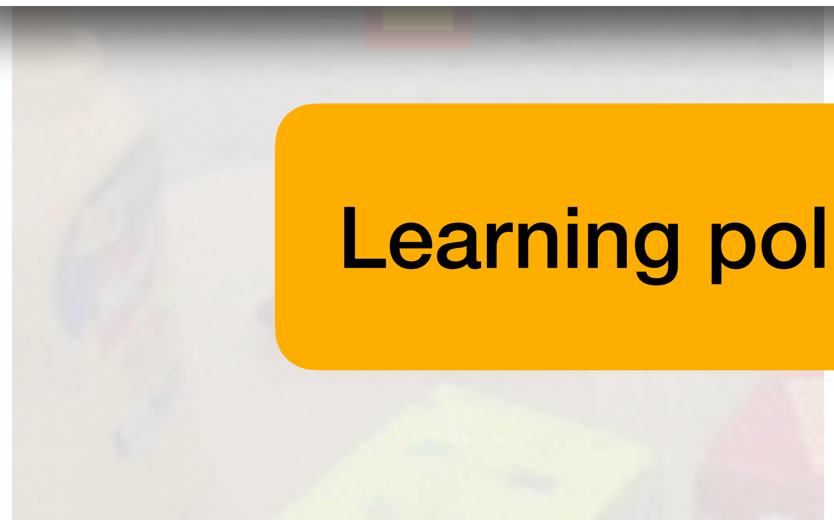
Action Prediction

Journal of Machine Learning Research 17 (2016) 1-40 Submitted 10/15; Published 4/16

End-to-End Training of Deep Visuomotor Policies

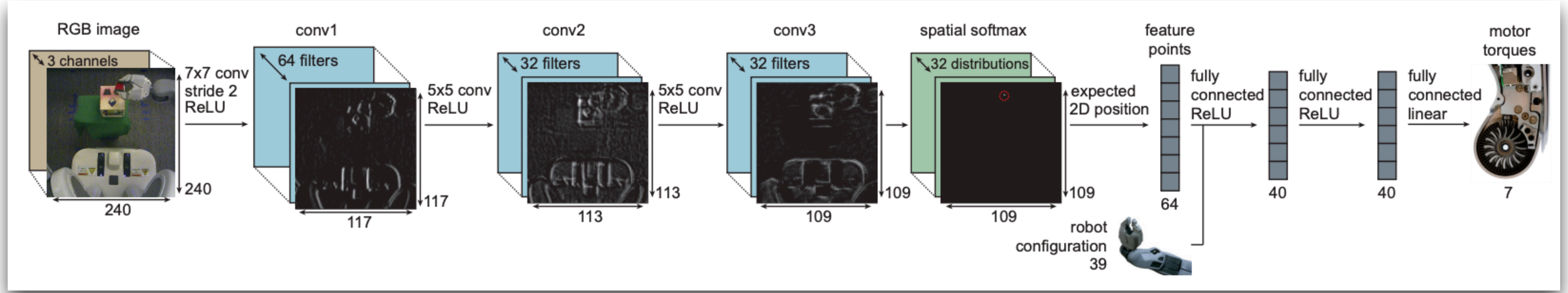
Sergey Levine† SVLEVINE@EECS.BERKELEY.EDU
Chelsea Finn† CBFINN@EECS.BERKELEY.EDU
Trevor Darrell TREVOR@EECS.BERKELEY.EDU
Pieter Abbeel PABBEEL@EECS.BERKELEY.EDU
Division of Computer Science
University of California
Berkeley, CA 94720-1776, USA
†These authors contributed equally.





Learning policies that map raw image observations directly to torques at the robot's motors

What does this entail?

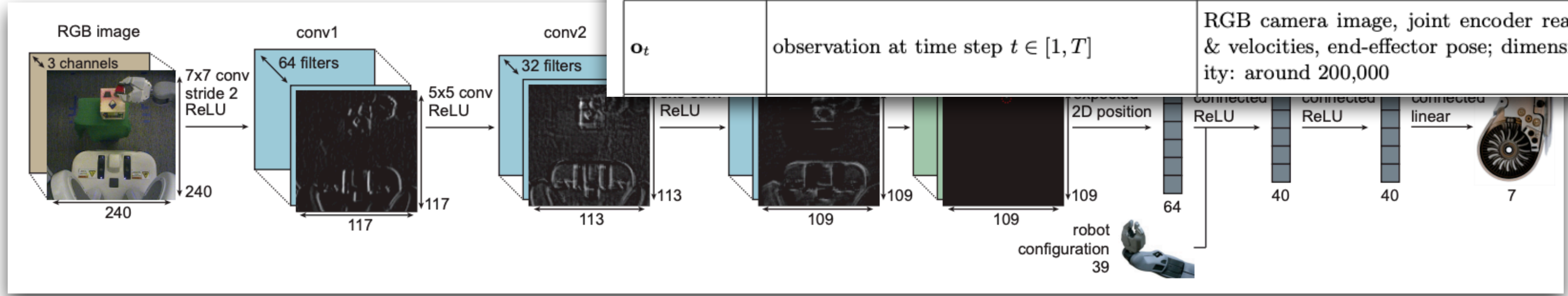


- Input: o_t
- Output: u_t
- Policy: $\pi_\theta(u_t \mid o_t)$

Levine, Sergey, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. "End-to-end training of deep visuomotor policies." *Journal of Machine Learning Research* 17, no. 39 (2016): 1-40.

What d

symbol	definition	example/details
$\mathbf{x}_t$	Markovian system state at time step $t \in [1, T]$	joint angles, end-effector pose, object positions, and their velocities; dimensionality: 14 to 32
$\mathbf{u}_t$	control or action at time step $t \in [1, T]$	joint motor torque commands; dimensionality: 7 (for the PR2 robot)
$\mathbf{o}_t$	observation at time step $t \in [1, T]$	RGB camera image, joint encoder readings & velocities, end-effector pose; dimensionality: around 200,000



- Input: o_t
- Output: u_t
- Policy: $\pi_\theta(u_t \,|\, o_t)$

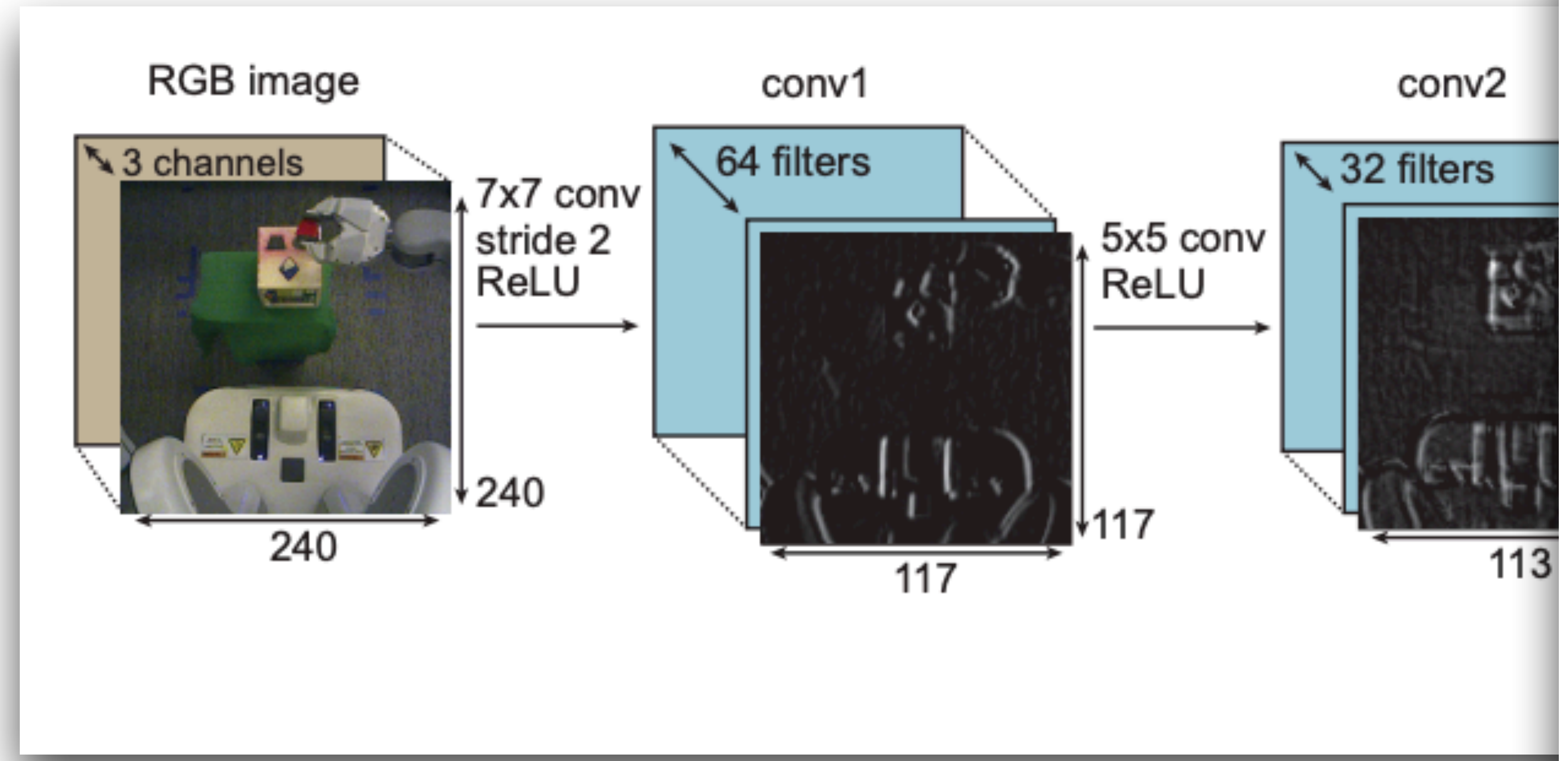
State x_t

Vs.

Observation o_t

Levine, Sergey, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. "End-to-end training of deep visuomotor policies." *Journal of Machine Learning Research* 17, no. 39 (2016): 1-40.

What c



symbol	definition	example/details
$\mathbf{x}_t$	Markovian system state at time step $t \in [1, T]$	joint angles, end-effector pose, object positions, and their velocities; dimensionality: 14 to 32
$\mathbf{u}_t$	control or action at time step $t \in [1, T]$	joint motor torque commands; dimensionality: 7 (for the PR2 robot)
$\mathbf{o}_t$	observation at time step $t \in [1, T]$	RGB camera image, joint encoder readings & velocities, end-effector pose; dimensionality: around 200,000
τ	trajectory: $\tau = \{\mathbf{x}_1, \mathbf{u}_1, \mathbf{x}_2, \mathbf{u}_2, \ldots, \mathbf{x}_T, \mathbf{u}_T\}$	notational shorthand for a sequence of states and actions
$\ell(\mathbf{x}_t, \mathbf{x}_t)$	cost function that defines the goal of the task	distance between an object in the gripper and the target
$p(\mathbf{x}_{t+1}\|\mathbf{x}_t, \mathbf{u}_t)$	unknown system dynamics	physics that govern the robot and any objects it interacts with
	ution	stochastic process that produces camera images from system state
	licy parameter-	convolutional neural network, such as the one in Figure 2
		notational shorthand for observation-based policy conditioned on state
	linear-Gaussian	time-varying linear-Gaussian controller has form $\mathcal{N}(\mathbf{K}_{ti}\mathbf{x}_t + \mathbf{k}_{ti}, \mathbf{C}_{ti})$
	for $\pi_\theta(\mathbf{u}_t\|\mathbf{x}_t)$: $\mathbf{x}_t, \mathbf{u}_t)$	notational shorthand for trajectory distribution induced by a policy

Table 1: Summary of the notation frequently used in this article.

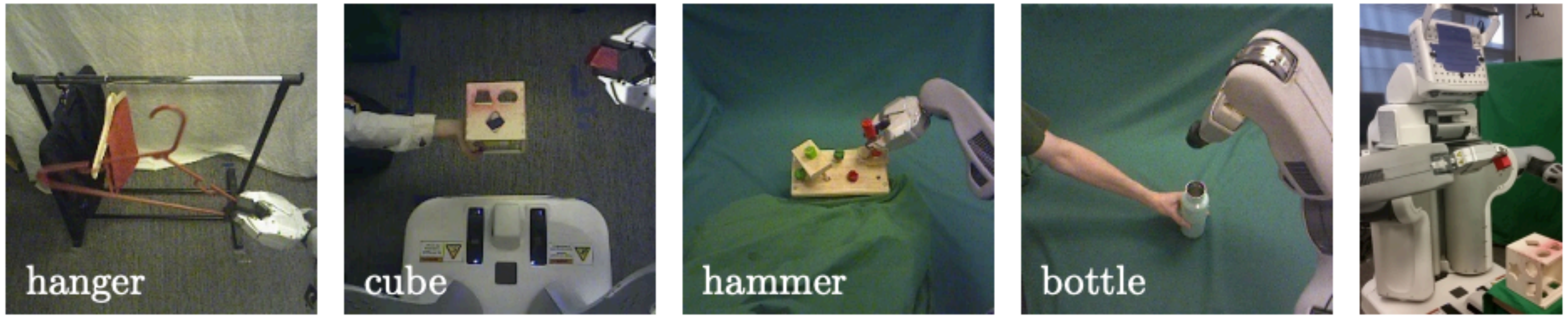
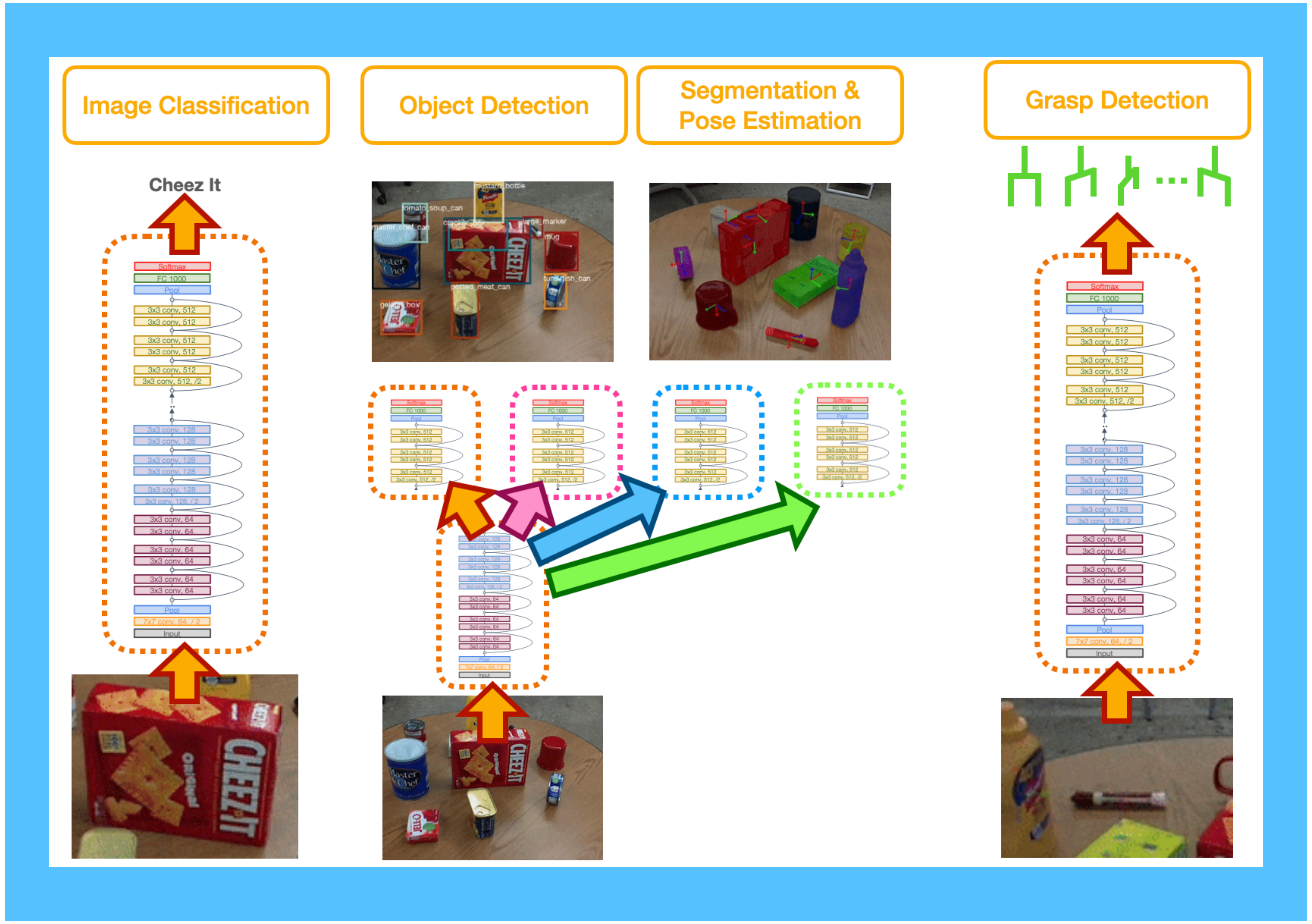


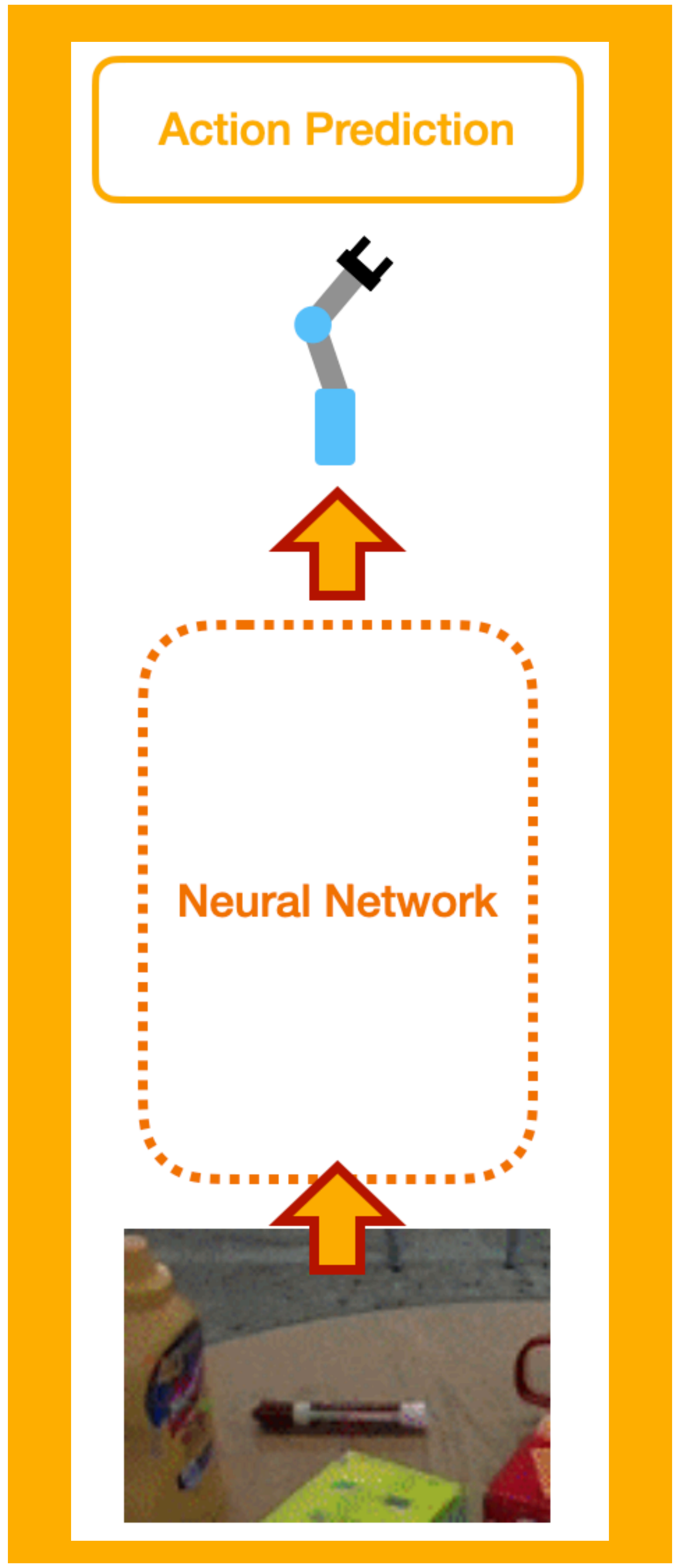
Figure 1: Our method learns visuomotor policies that directly use camera image observations (left) to set motor torques on a PR2 robot (right).

Levine, Sergey, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. "End-to-end training of deep visuomotor policies." *Journal of Machine Learning Research* 17, no. 39 (2016): 1-40.

Levine, Sergey, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. "End-to-end training of deep visuomotor policies." *Journal of Machine Learning Research* 17, no. 39 (2016): 1-40.
https://www.youtube.com/watch?v=Q4bMcUk6pcw

Challenges in going from Prediction to Control





Challenges in going from Prediction to Control

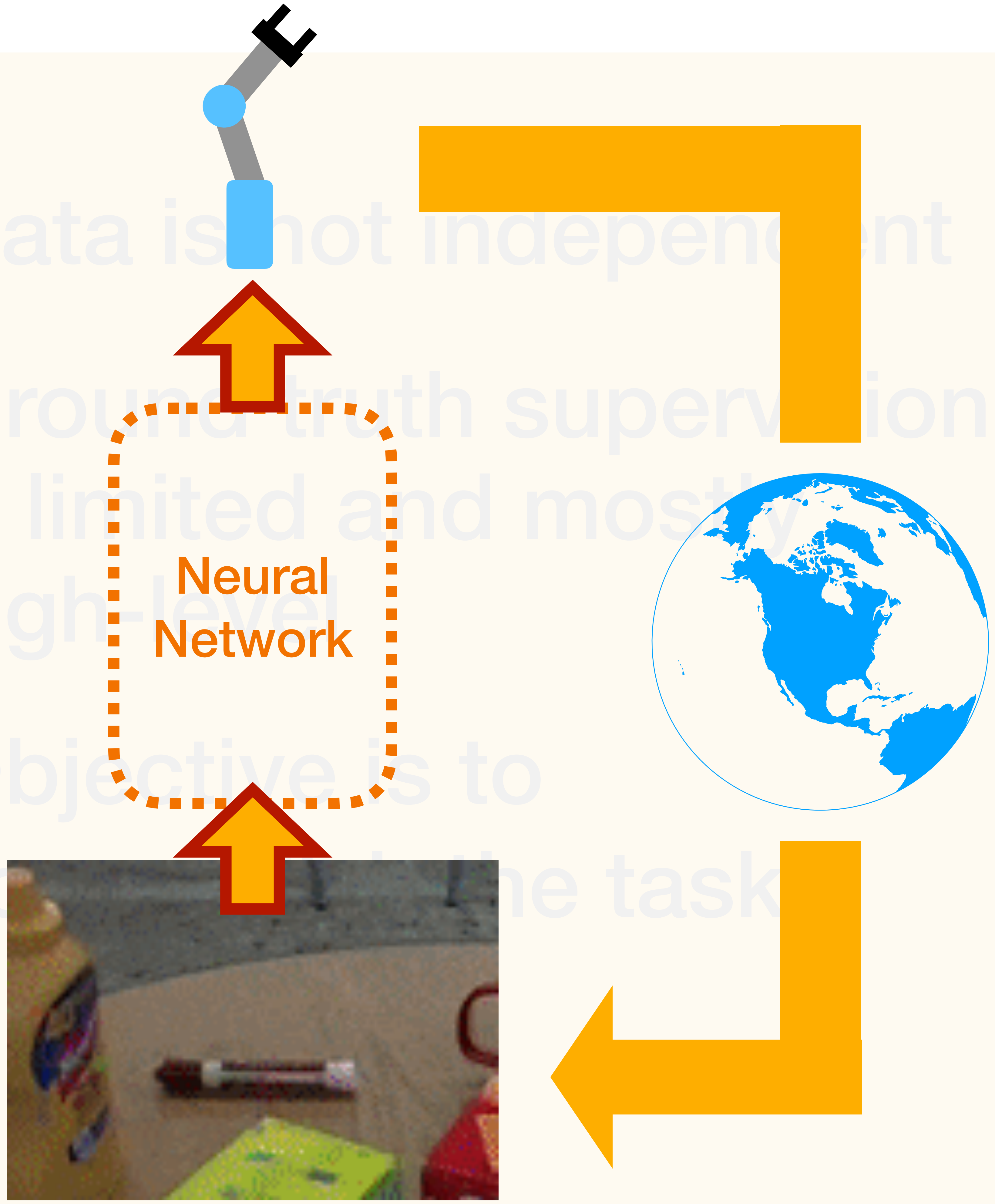
- Data is i.i.d distributed
- Ground truth supervision for the prediction is available
- Objective is to predict the right label or regress a value close to the ground truth

- Data is not independent
- Ground truth supervision is limited and mostly high-level
- Objective is to accomplish the task

Challenges in going from Prediction to Control

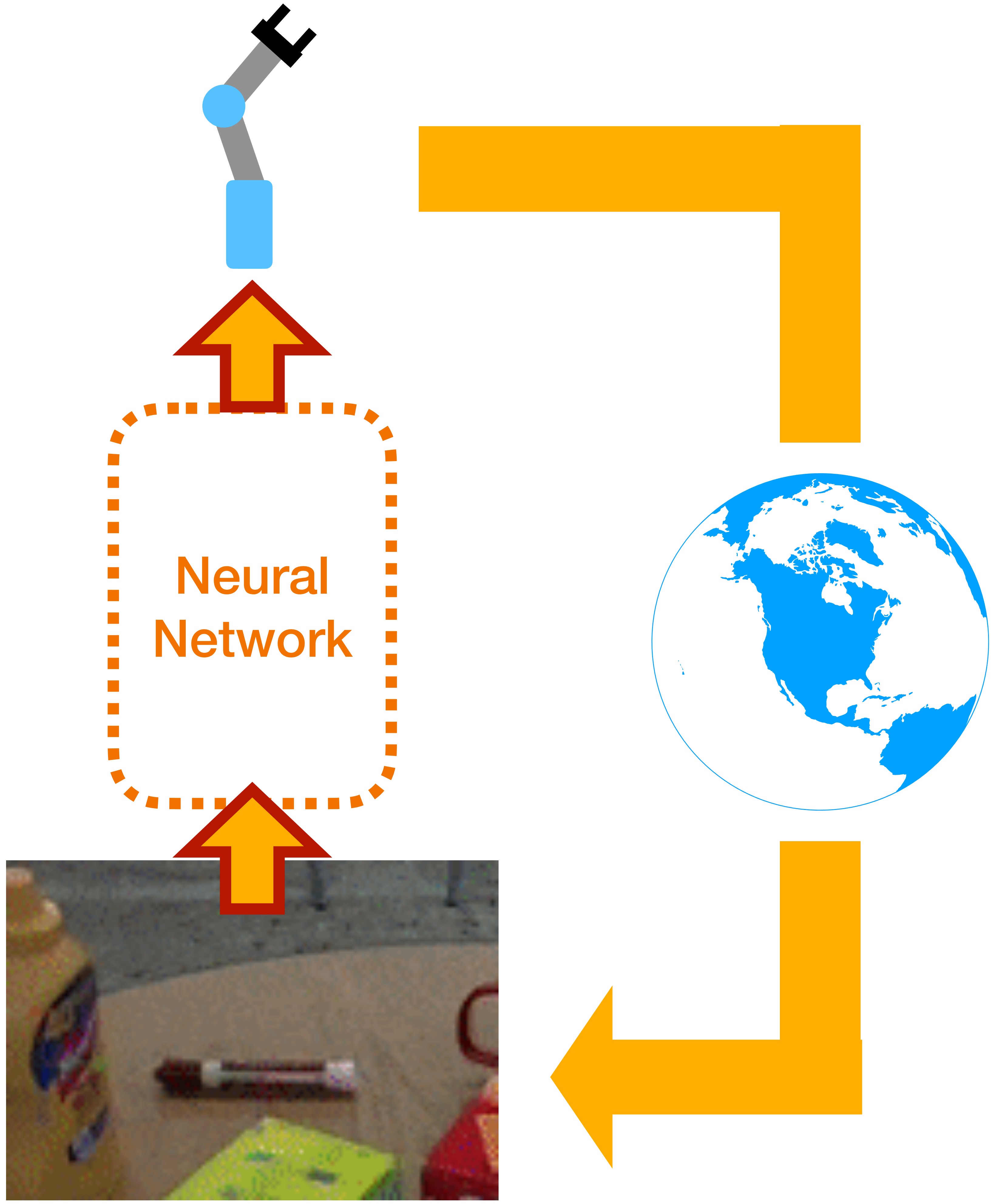
- Data is i.i.d distributed
- Ground truth supervision for the prediction is available
- Objective is to predict the right label or regress a value close to the ground truth

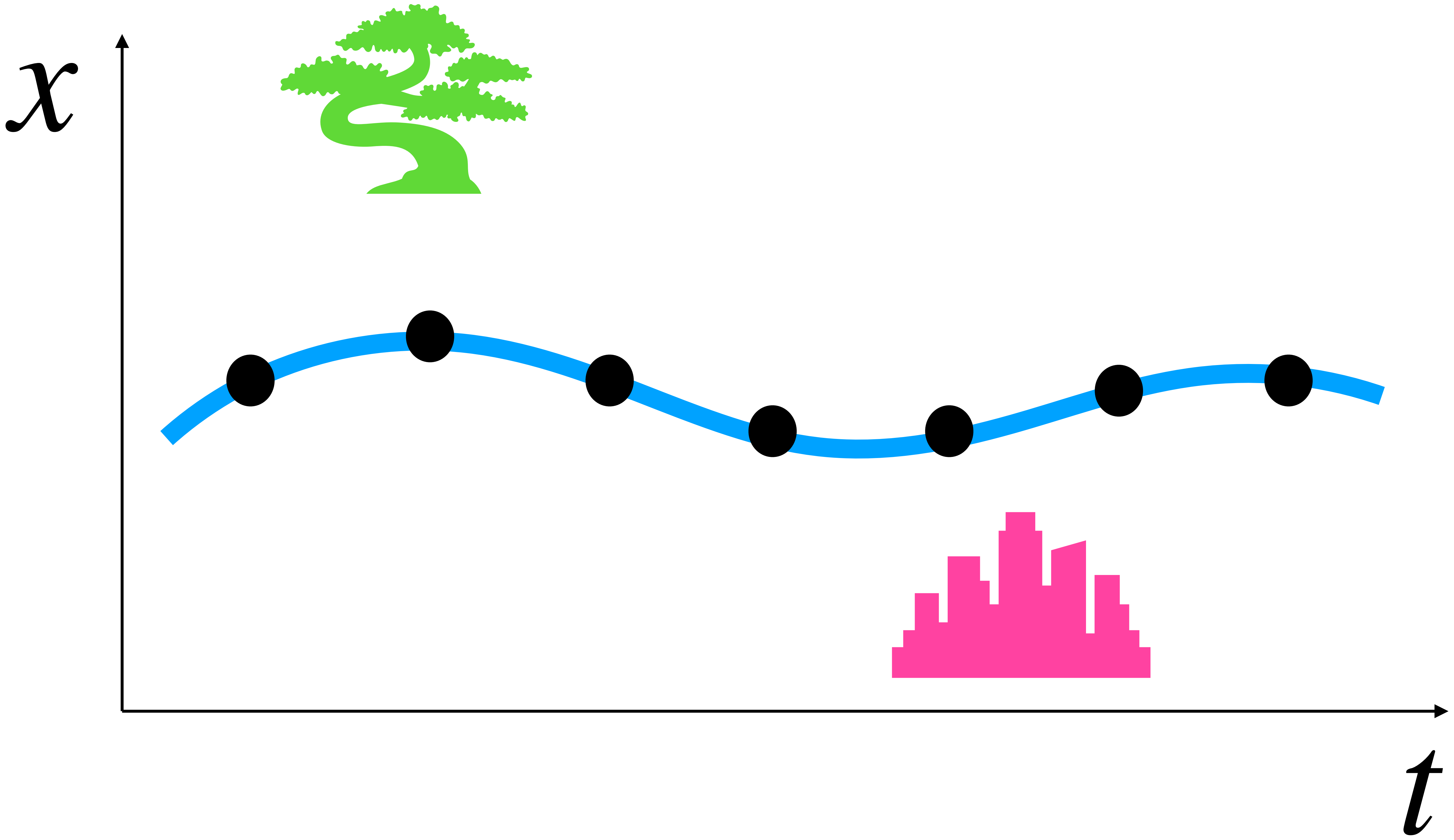
- Data is not independent
- Ground truth supervision is limited and mostly high-level
- Objective is to ac the task



There is feedback and associated issues!

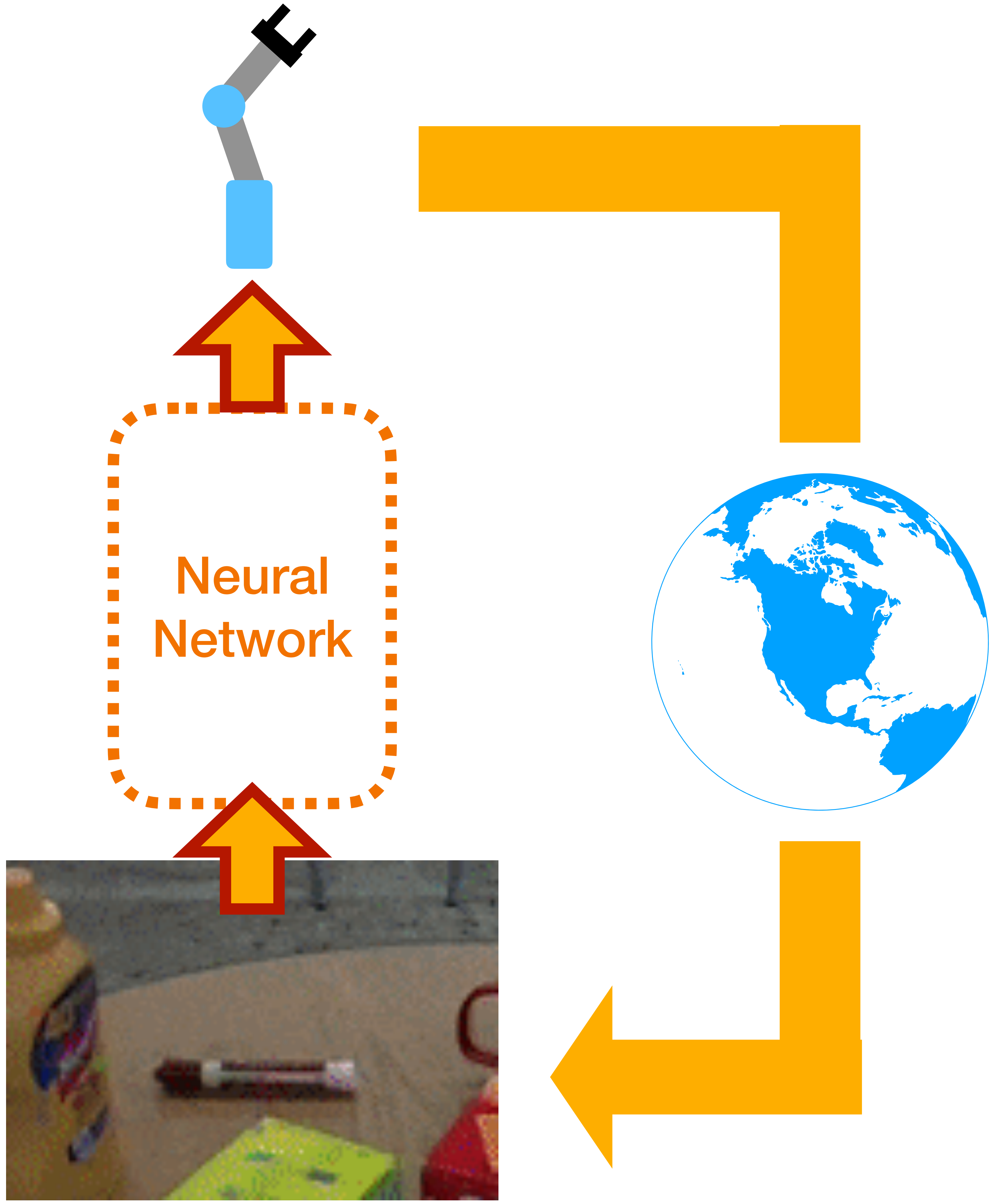
There is feedback and associated issues!

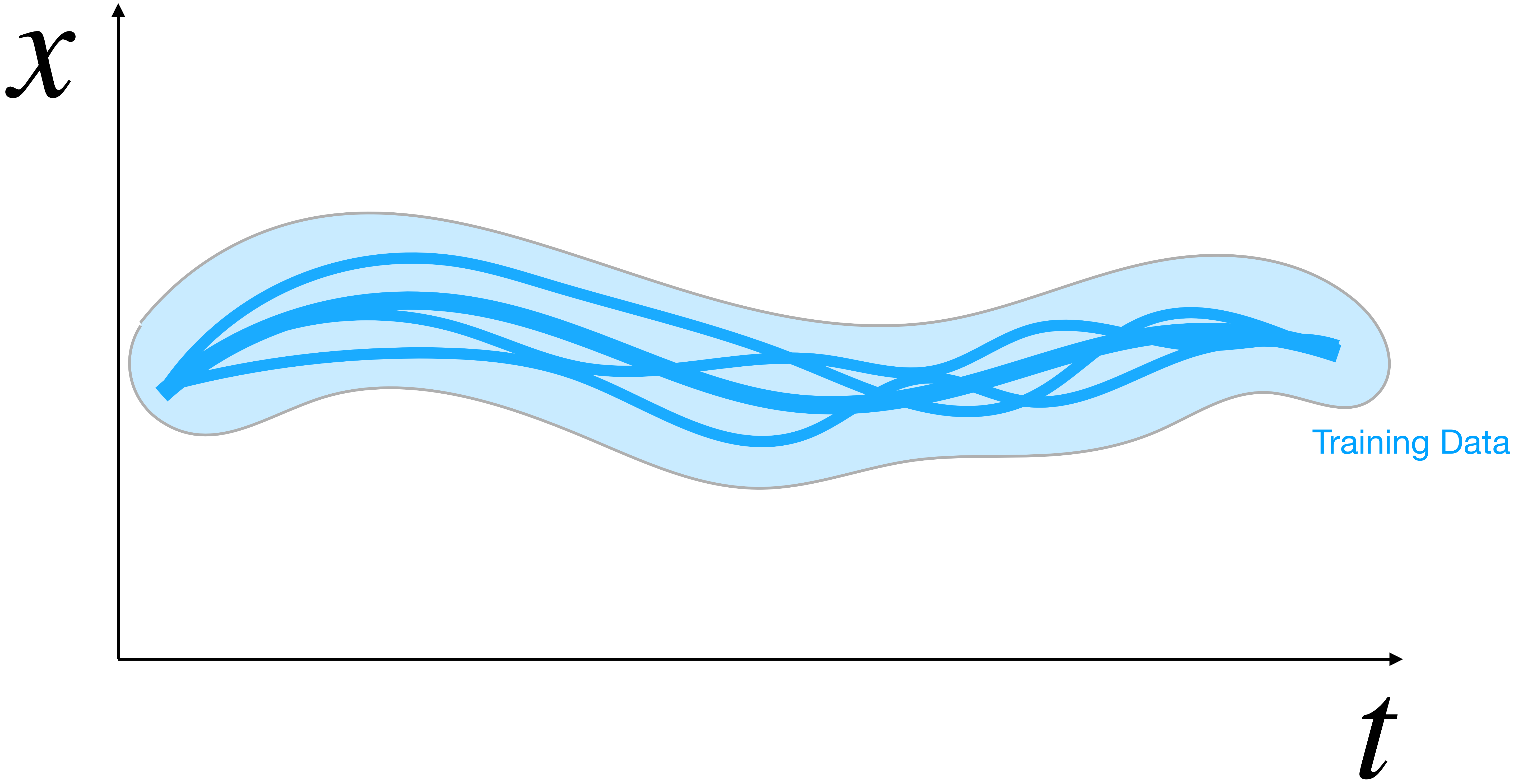




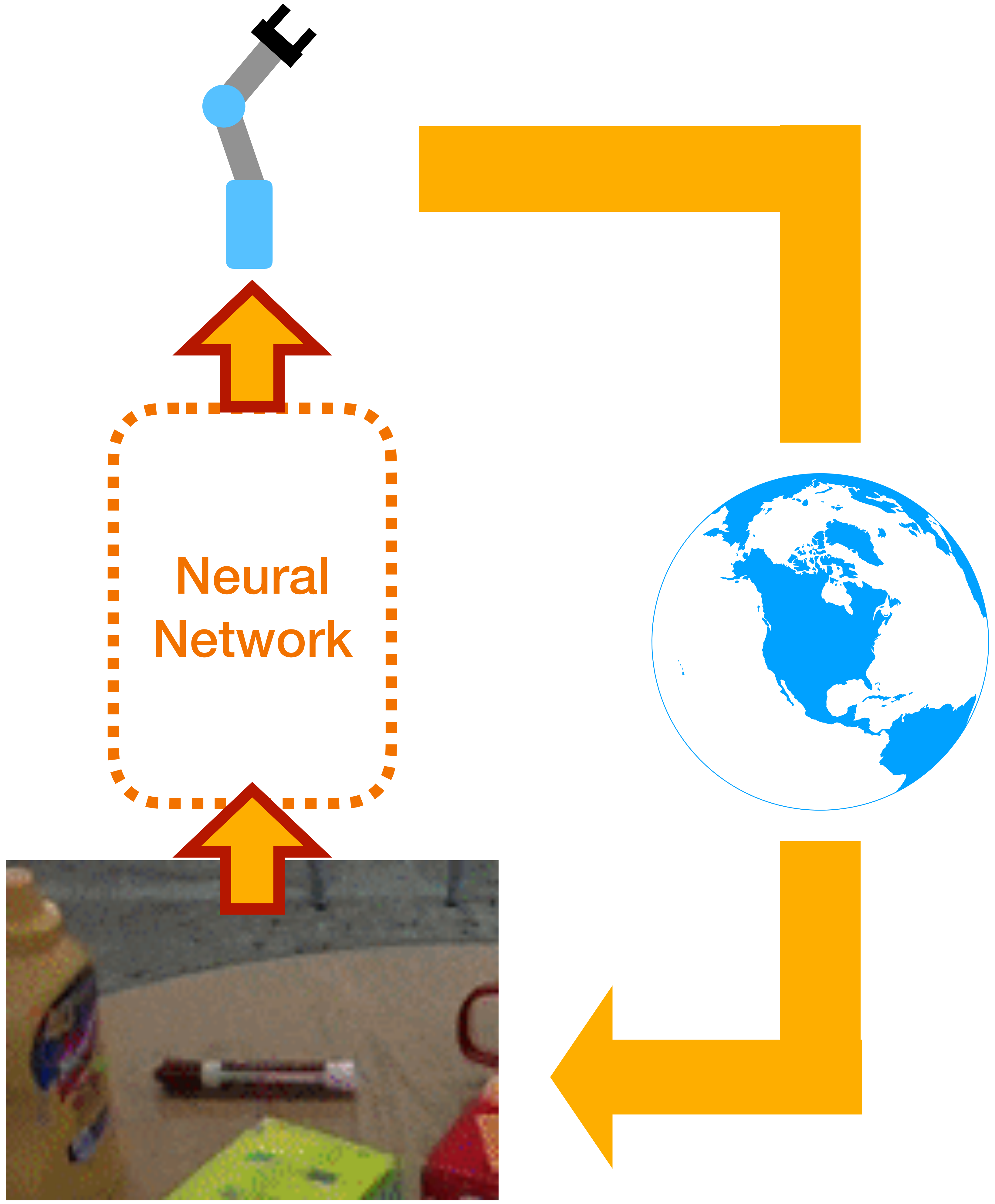
Data is dependent

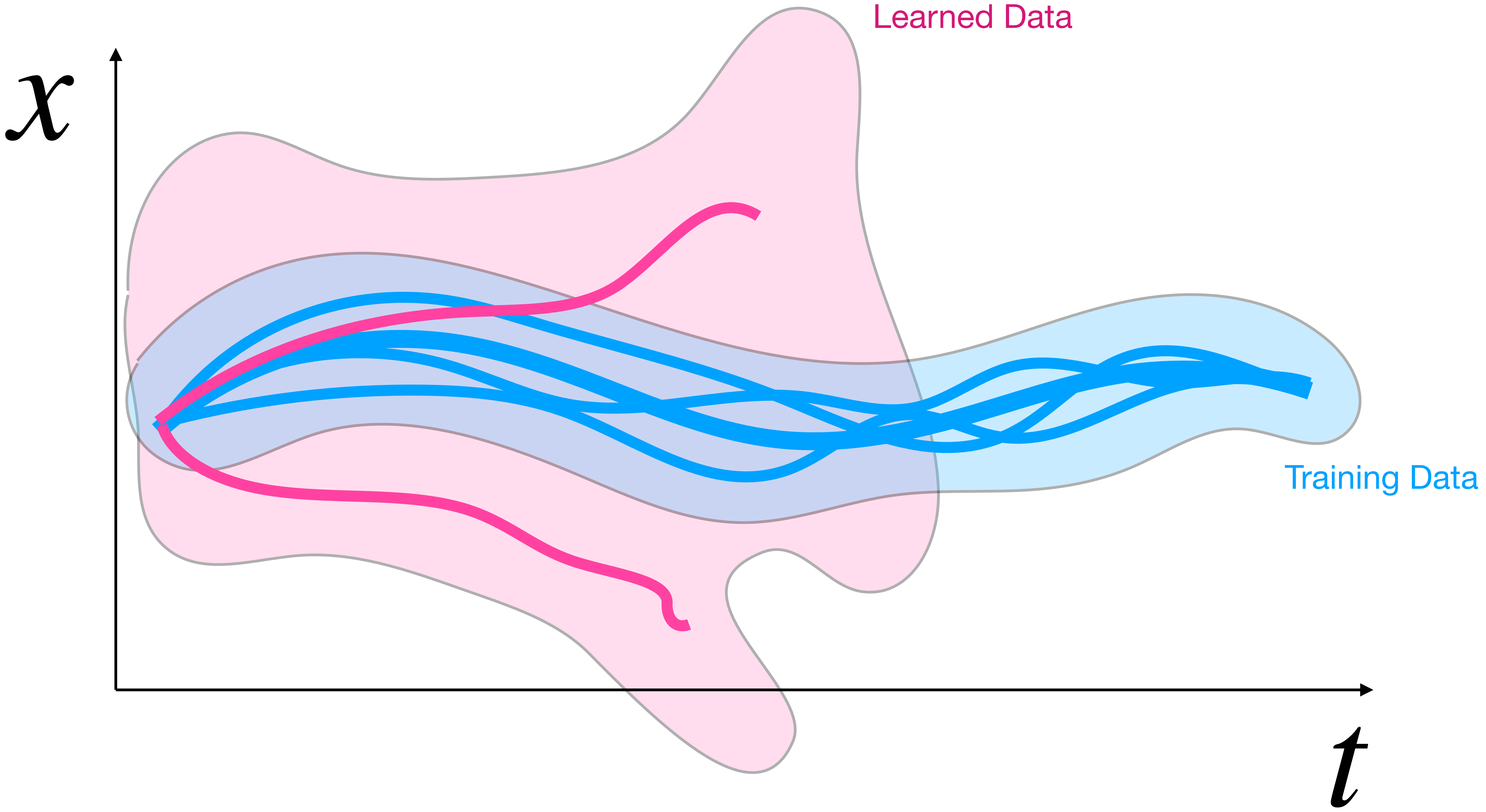
There is feedback and associated issues!





There is feedback and associated issues!





Covariance Shift due to Feedback

This is a commonly seen issue

2019 IEEE/CVF International Conference on Computer Vision (ICCV)

Exploring the Limitations of Behavior Cloning for Autonomous Driving

Felipe Codevilla *
Computer Vision Center (CVC)
Campus UAB, Barcelona, Spain
fcodevilla@cvc.uab.es

Eder Santana
Toyota Research Institute (TRI)
Los Altos, CA, USA.
edercsjr@gmail.com

Antonio M. López
Computer Vision Center (CVC)
Campus UAB, Barcelona, Spain
antonio@cvc.uab.es

Adrien Gaidon
Toyota Research Institute (TRI)
Los Altos, CA, USA.
adrien.gaidon@tri.global

Robotics: Science and Systems 2019
Freiburg im Breisgau, June 22-26, 2019

ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst

Mayank Bansal
Waymo Research
Mountain View, CA, USA
mayban@waymo.com

Alex Krizhevsky[†]
akrizhevsky@gmail.com

Abhijit Ogale
Waymo Research
Mountain View, CA, USA
ogale@waymo.com

2017 IEEE Intelligent Vehicles Symposium (IV)
June 11-14, 2017, Redondo Beach, CA, USA

Imitating Driver Behavior with Generative Adversarial Networks

Alex Kuefler[1], Jeremy Morton[2], Tim Wheeler[2], and Mykel Kochenderfer[2]

Causal Confusion in Imitation Learning

Pim de Haan[*1], **Dinesh Jayaraman**[†‡], **Sergey Levine**[†]
[*]Qualcomm AI Research, University of Amsterdam,
[†]Berkeley AI Research, [‡] Facebook AI Research

DAGGER

A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning

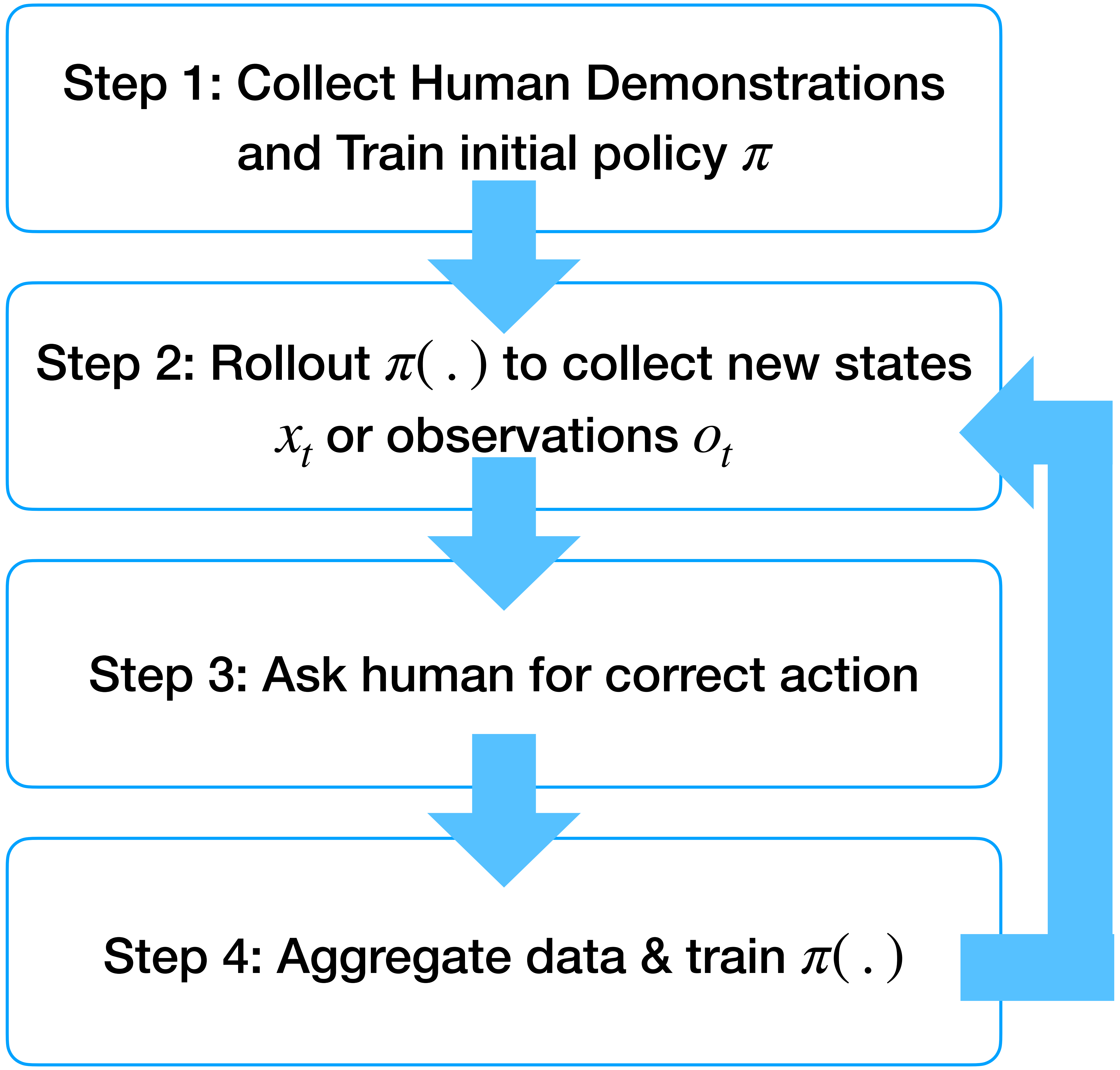
Stéphane Ross
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
stephaneross@cmu.edu

Geoffrey J. Gordon
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213, USA
ggordon@cs.cmu.edu

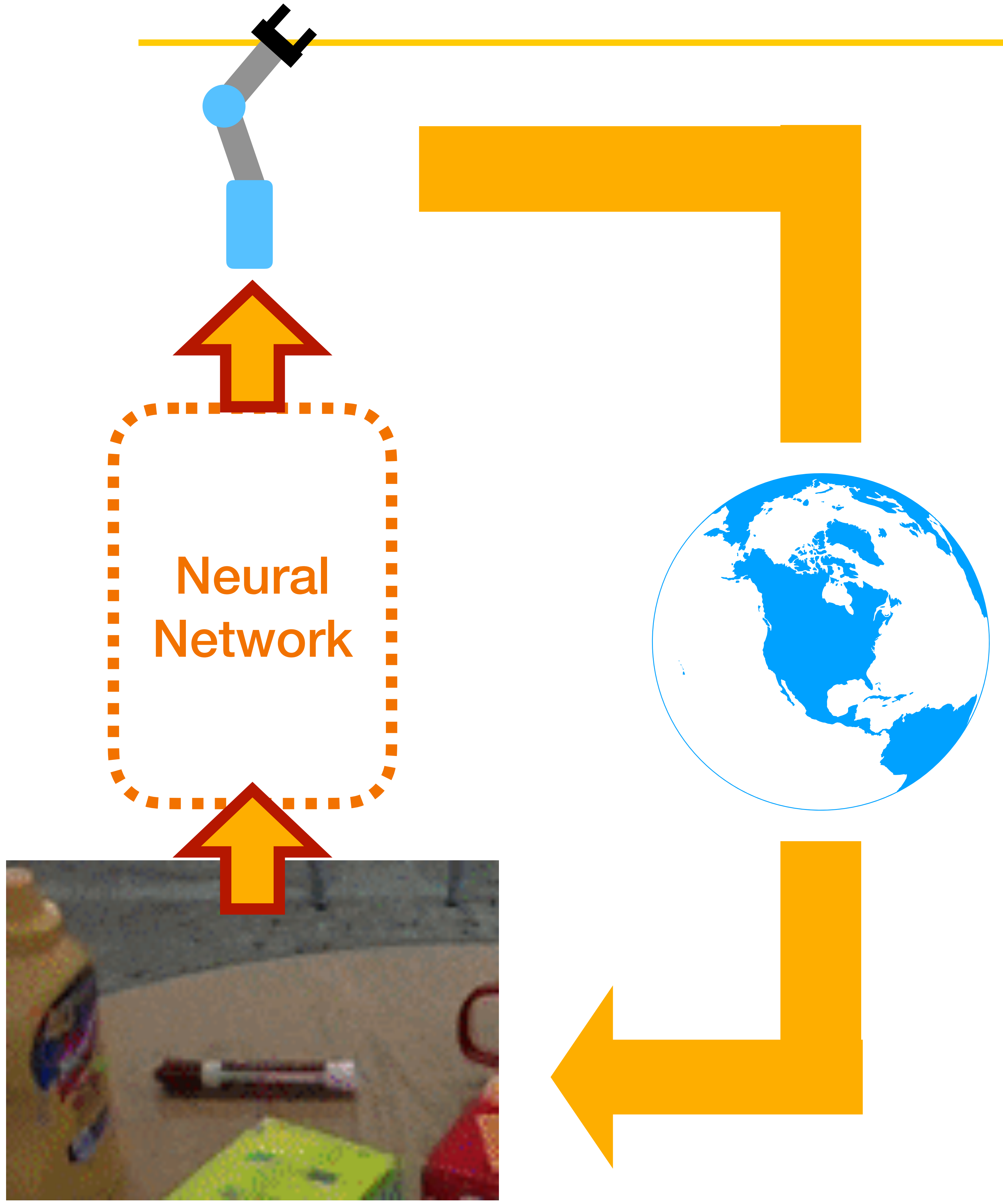
J. Andrew Bagnell
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
dbagnell@ri.cmu.edu

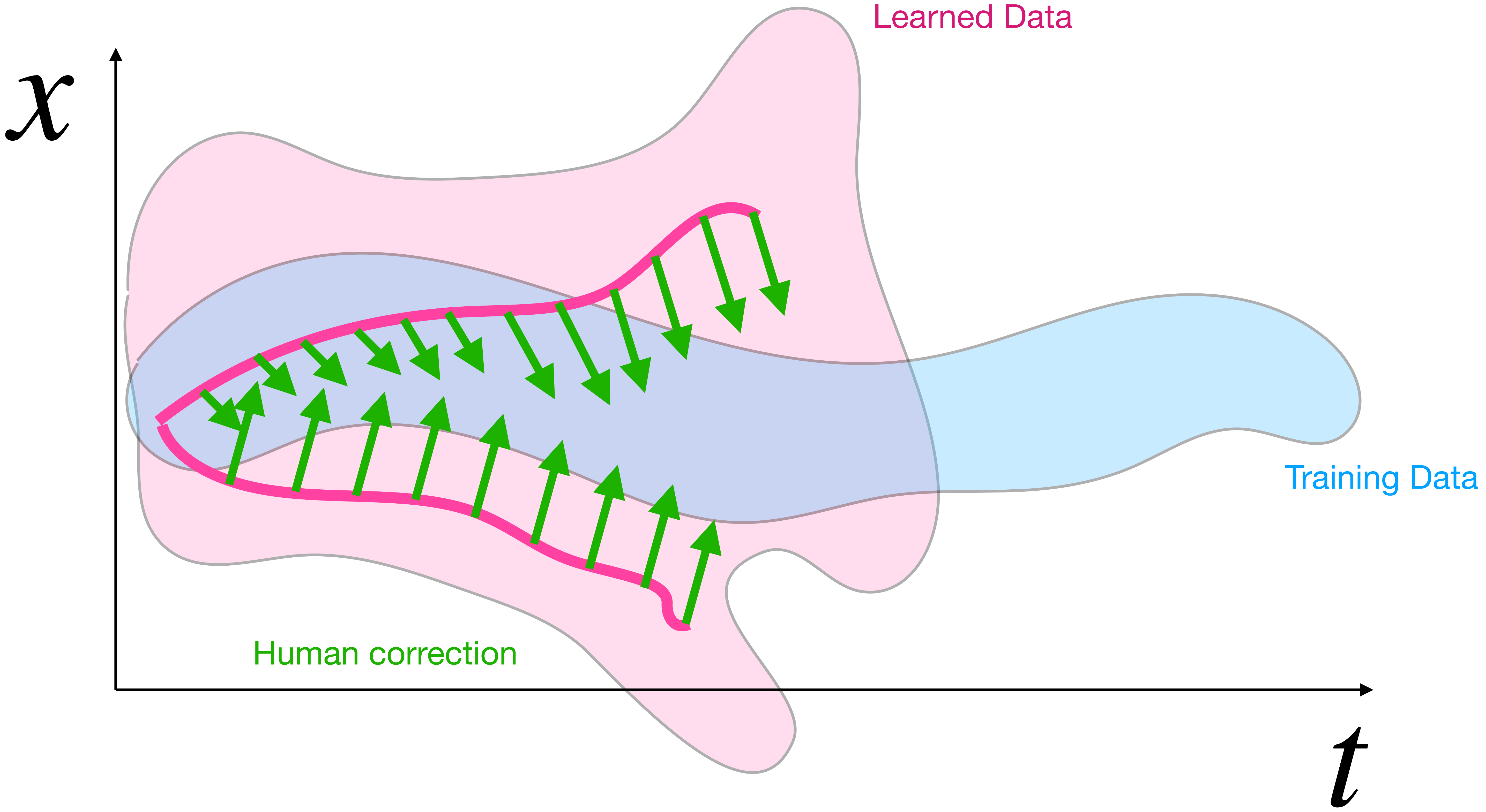
Initialize $\mathcal{D} \leftarrow \emptyset$.
Initialize $\hat{\pi}_1$ to any policy in Π.
for $i = 1$ **to** N **do**
  Let $\pi_i = \beta_i \pi^* + (1 - \beta_i)\hat{\pi}_i$.
  Sample T-step trajectories using π_i.
  Get dataset $\mathcal{D}_i = \{(s, \pi^*(s))\}$ of visited states by π_i and actions given by expert.
  Aggregate datasets: $\mathcal{D} \leftarrow \mathcal{D} \bigcup \mathcal{D}_i$.
  Train classifier $\hat{\pi}_{i+1}$ on $\mathcal{D}$.
end for
Return best $\hat{\pi}_i$ on validation.

Algorithm 3.1: DAGGER Algorithm.

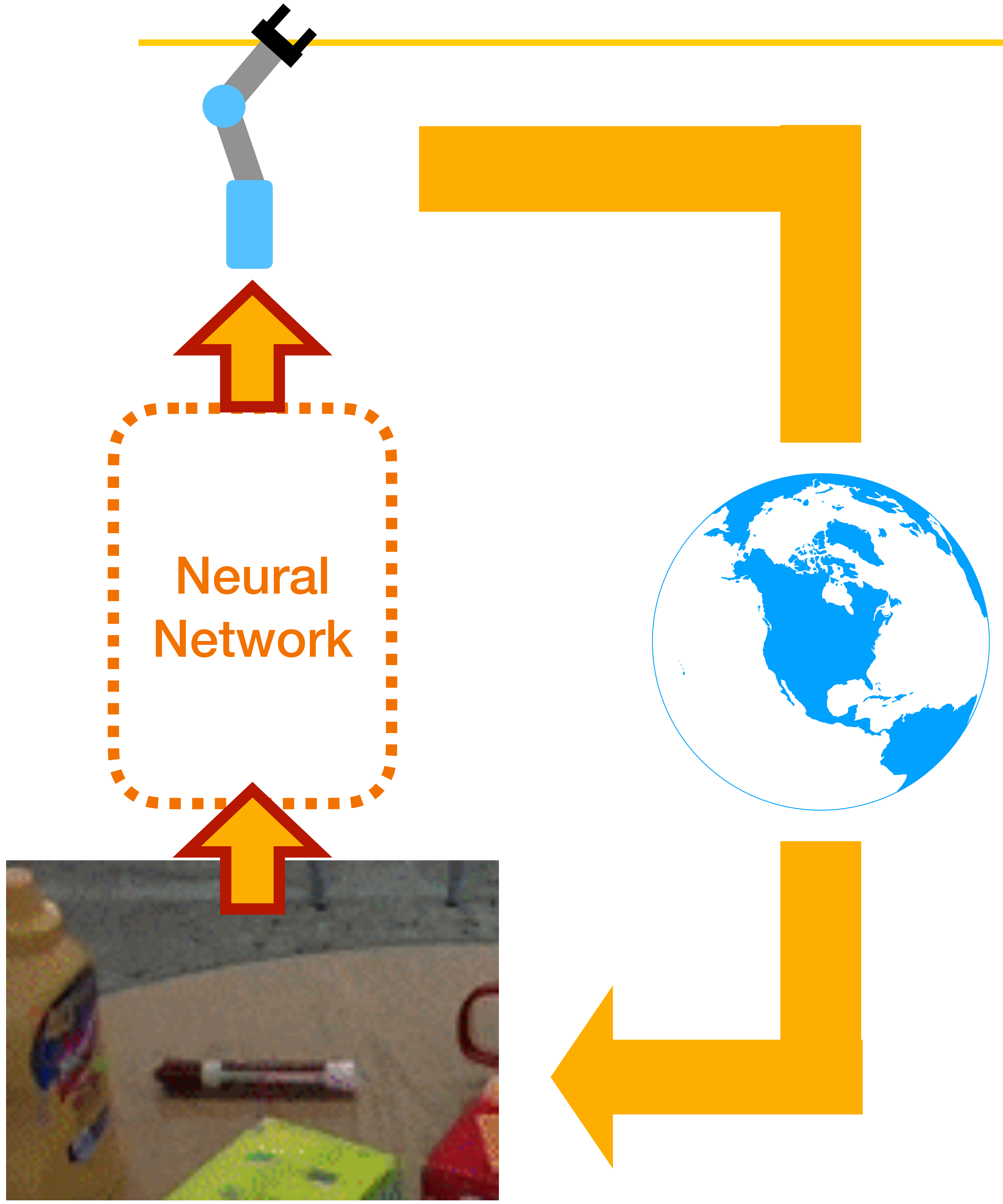


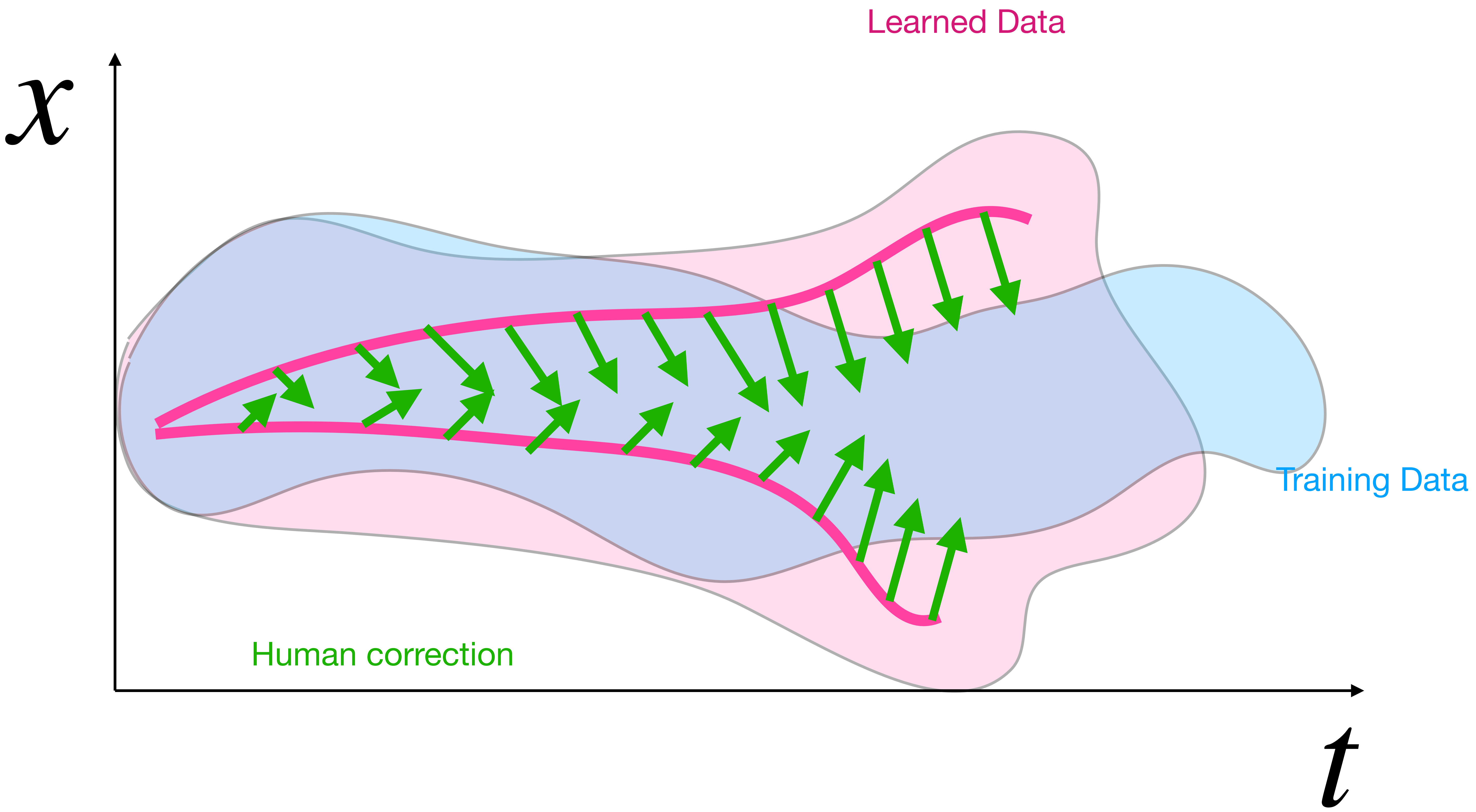
There is feedback and associated issues!



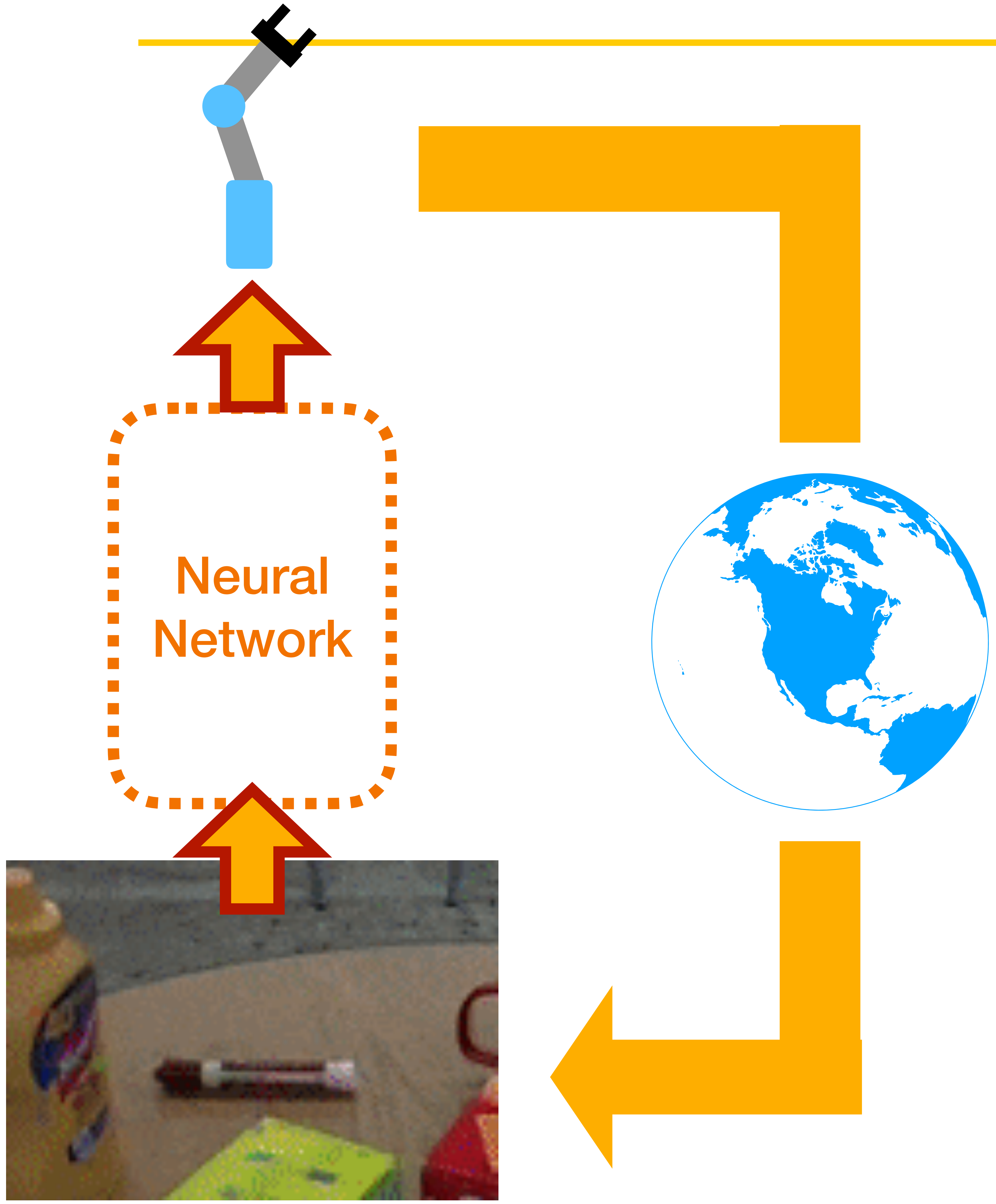


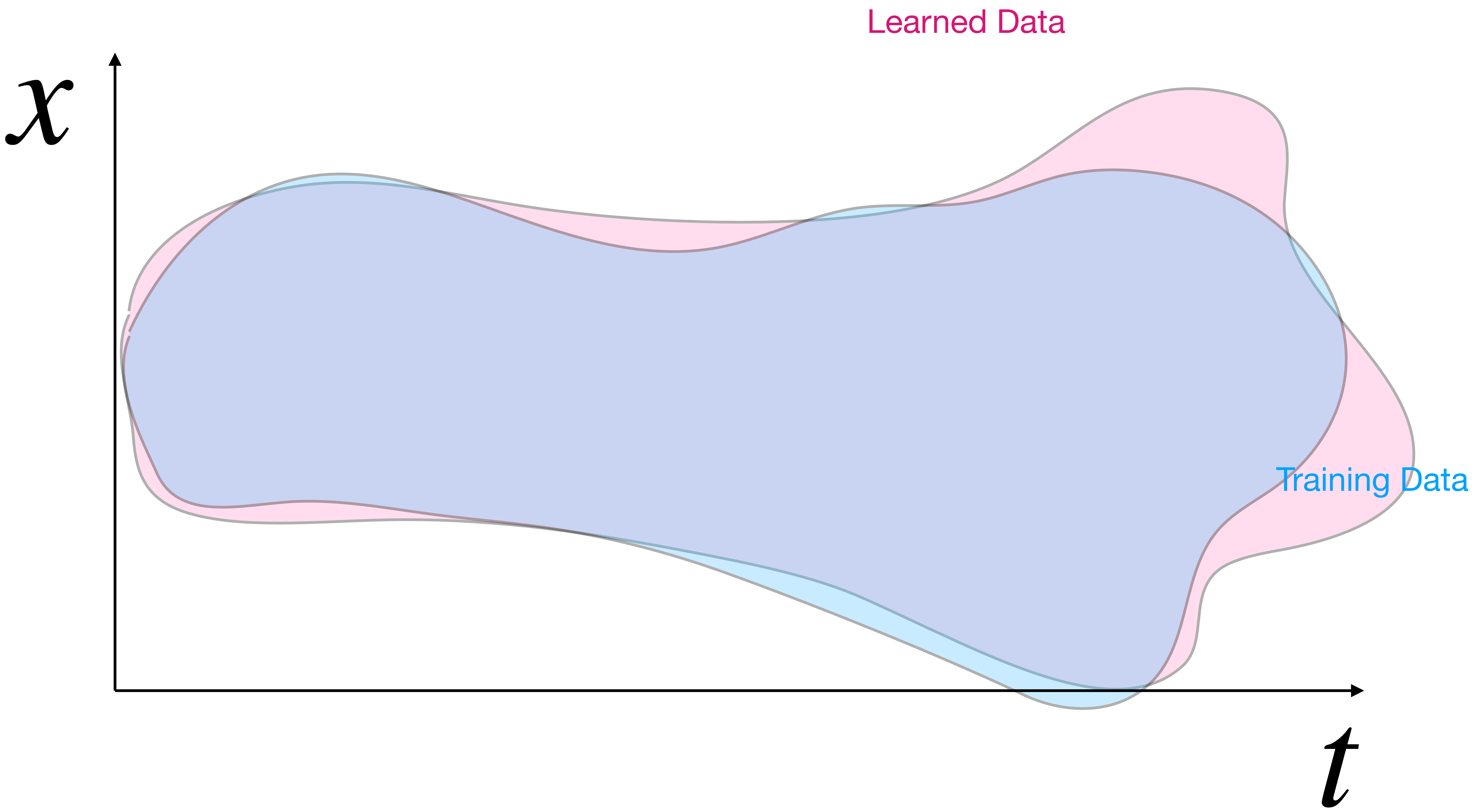
There is feedback and associated issues!





There is feedback and associated issues!





Next Lecture:
Imitation Learning II

DeepRob

Lecture 14
Imitation Learning I
University of Minnesota

HOW TO TRAIN YOUR DRAGON THE HIDDEN WORLD Kit Harington Auditions with Toothless