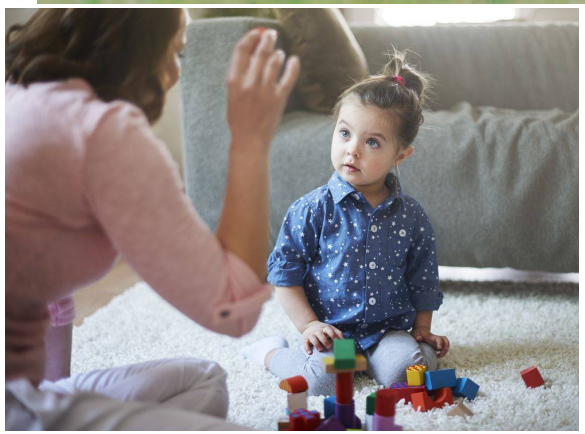
Imitation Learning

Alvaro Soto

Computer Science Department (DCC), PUC

Imitation Learning: An intrinsic human and animal skill



Imitation Learning:
An intrinsic human and animal skill

An innate ability

Imitation Learning: An intrinsic human and animal skill

A multimodal ability

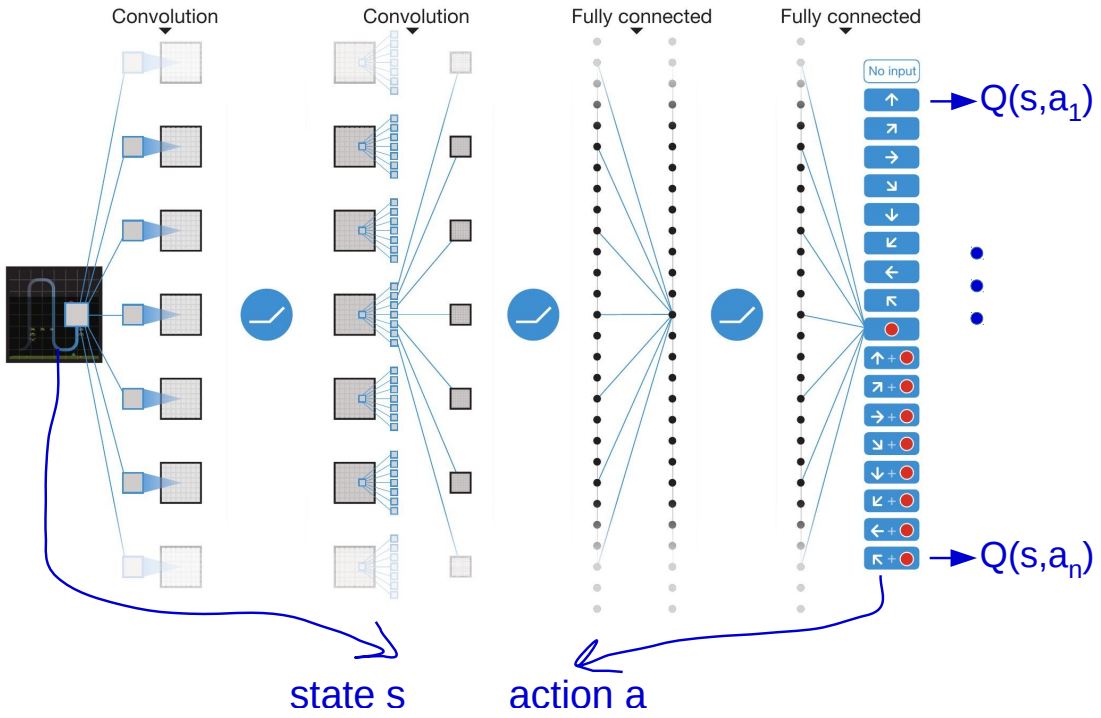
Photo Credit: Meltzoff & Moore, 1977, Science

If we want to create an intelligent machine,
we should imitate nature
and provide our machines with robust imitation learning abilities.

How can we provide imitation learning abilities to a machine ?

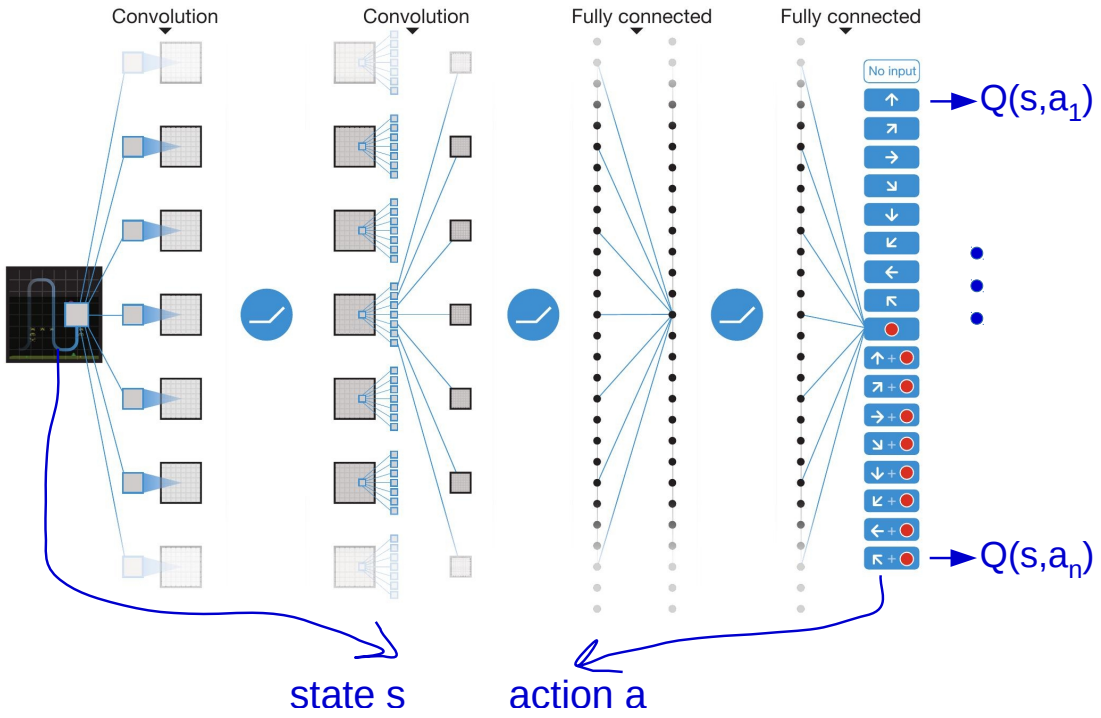
Recall from DQN

- We use a neural network to learn the Q-function.



Recall from DQN

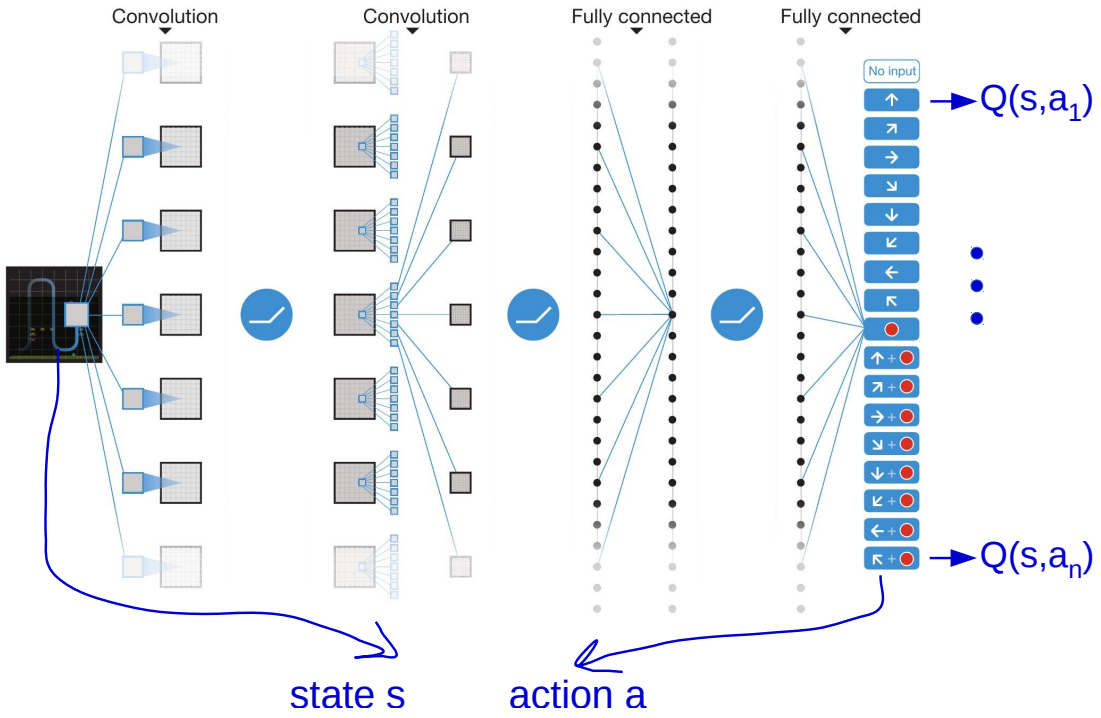
- We use a neural network to learn the Q-function.



- We did this without labels. How?

Recall from DQN

- We use a neural network to learn the Q-function.



- We did this without labels. How?
- We use a bootstrapping strategy: refining labels and building the Q-function as the agent explores the environment.

Recall from DQN

Main Trick: We use the current NW estimation of the Q-function to estimate a dynamic target value y_i:

$$y_i = \mathbb{E}_{s' \sim env}[r(s_i, a_i) + \gamma \arg\max_{a'} Q(s', a' | \theta_i^-)]$$

Recall from DQN

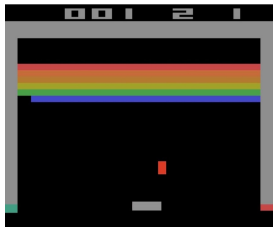
Main Trick: We use the current NW estimation of the Q-function to estimate a dynamic target value y_i:

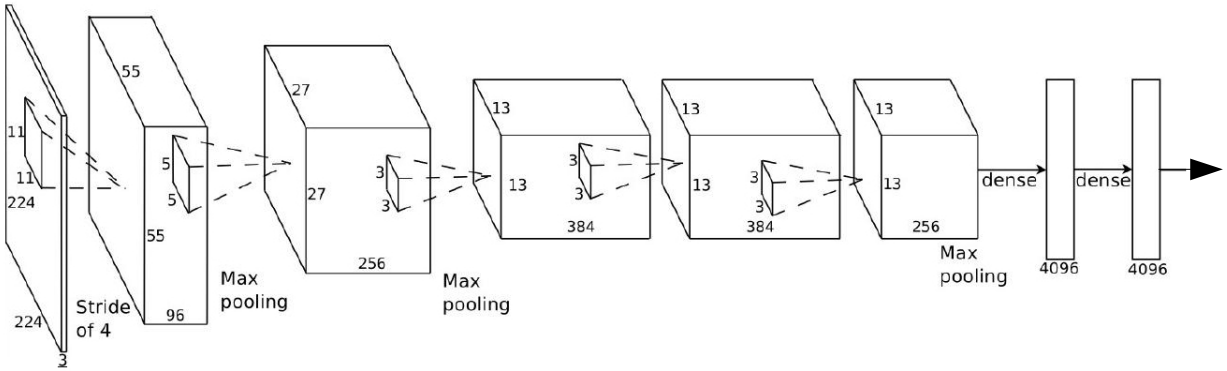
$$y_i = \mathbb{E}_{s' \sim env}[r(s_i, a_i) + \gamma \arg\max_{a'} Q(s', a' | \theta_i^-)]$$

Problem: inestability during learning is a big issue.

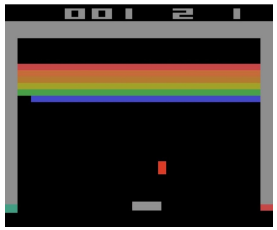
Any alternative?

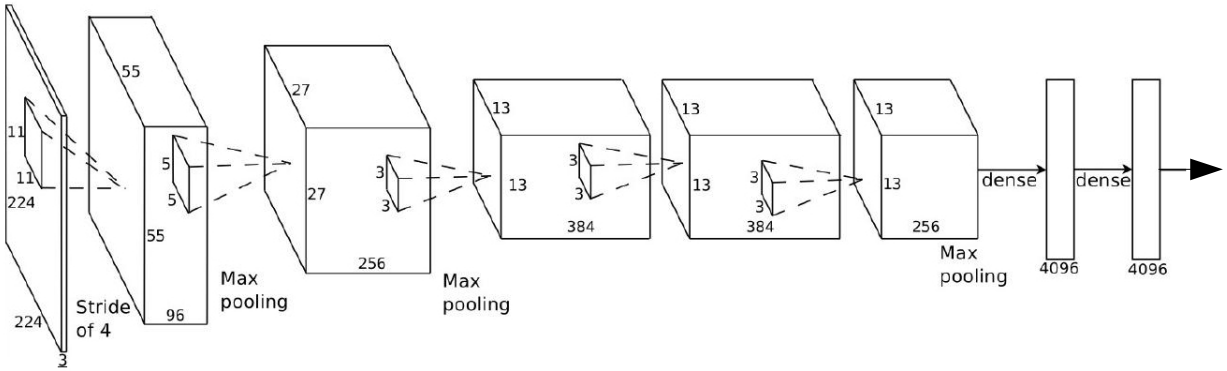
Idea: Use supervised learning





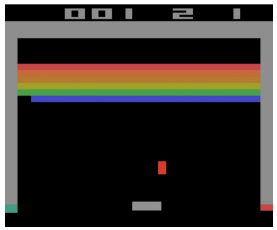
Idea: Use supervised learning

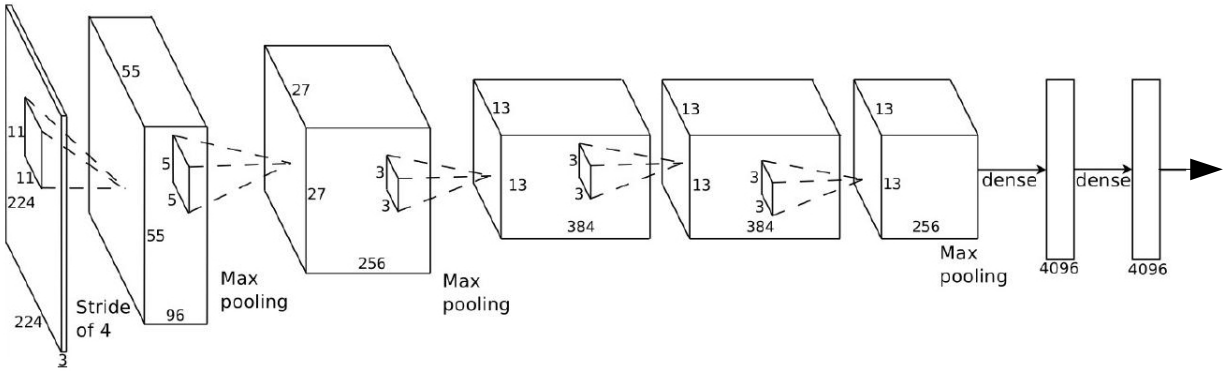




Q: Any potential problem?

Idea: Use supervised learning

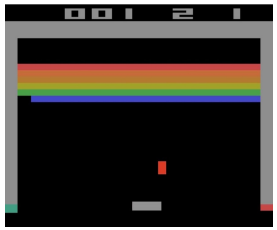


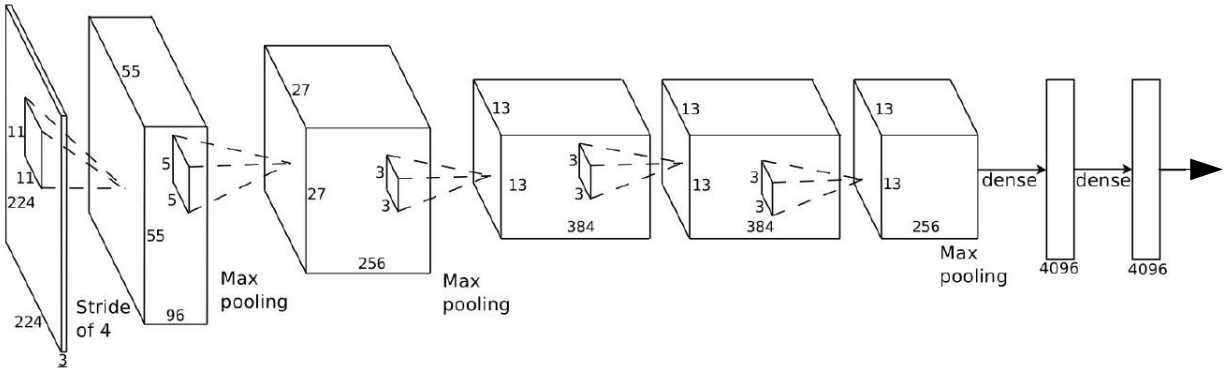


Q: Any potential problem?

Ans: Difficult/expensive to generate labels for a huge set of inputs

Idea: Use supervised learning



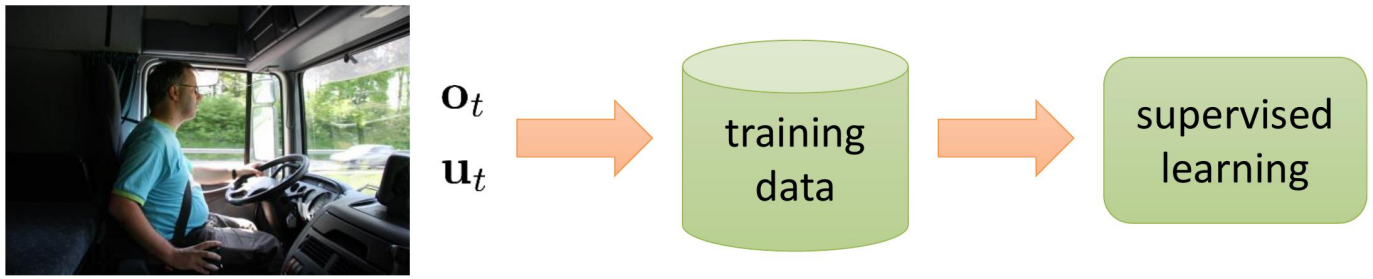


Q: Any potential problem?

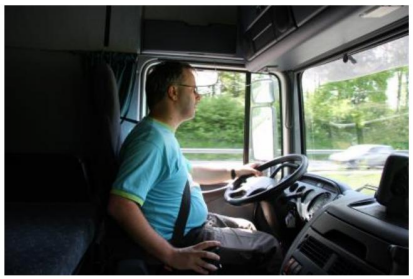
Ans: Difficult/expensive to generate labels for a huge set of inputs

What about using the dynamic of the game i.e. let's people play freely?

Let's consider the case of steering a car.
We can observe human drivers to easily generate training data to feed a supervised learning approach

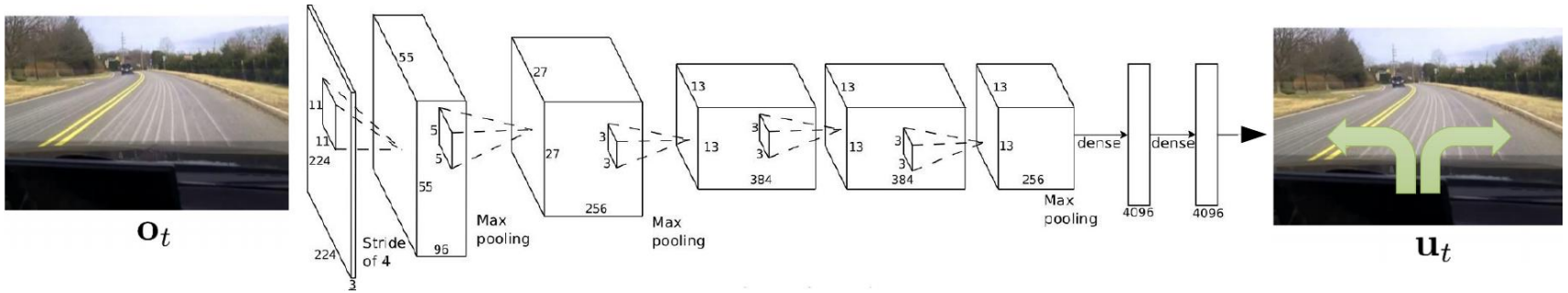


Let's consider the case of steering a car. We can observe human drivers to easily generate training data to feed a supervised learning approach

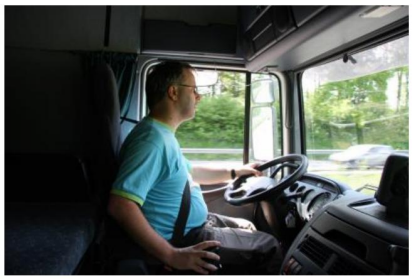


$\mathbf{o}_t$

$\mathbf{u}_t$

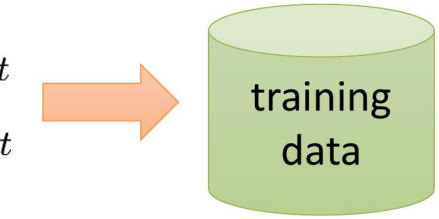


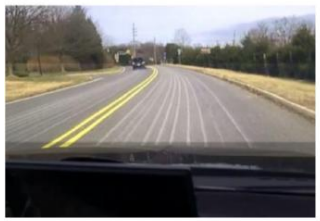
Let's consider the case of steering a car.
We can observe human drivers to easily generate training data to feed a supervised learning approach



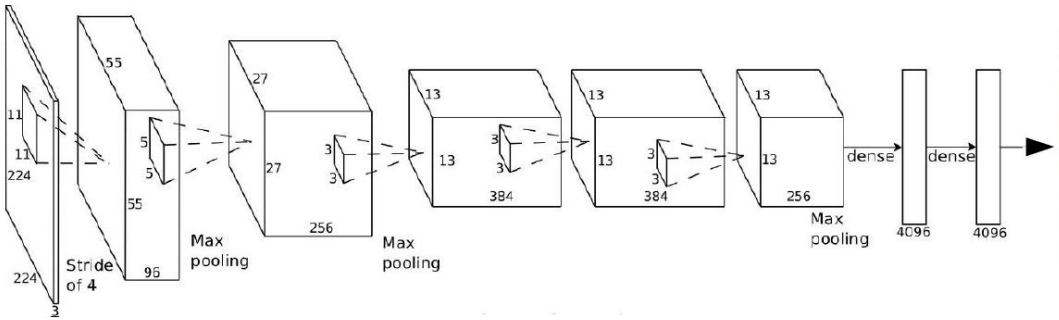
$\mathbf{o}_t$

$\mathbf{u}_t$





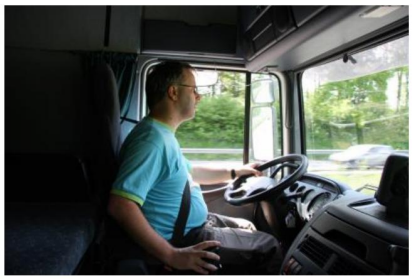
$\mathbf{o}_t$



$\mathbf{u}_t$

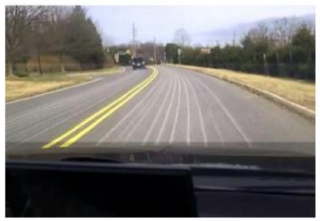
Do this work?

Let's consider the case of steering a car.
We can observe human drivers to easily generate training data to feed a supervised learning approach

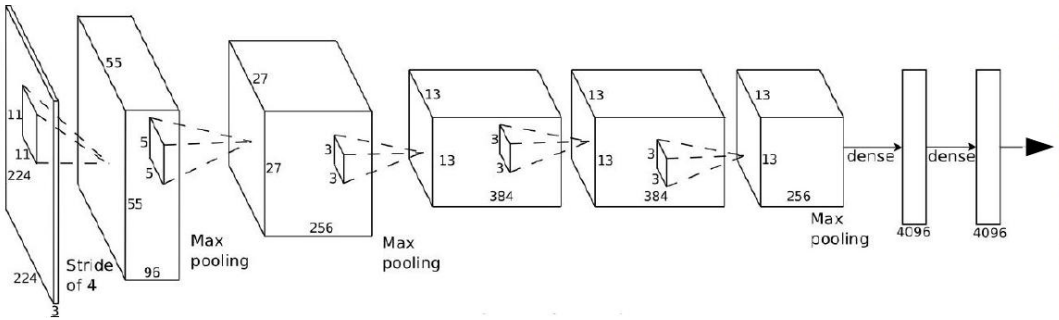


$\mathbf{o}_t$

$\mathbf{u}_t$



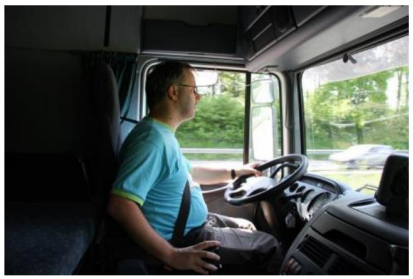
$\mathbf{o}_t$



$\mathbf{u}_t$

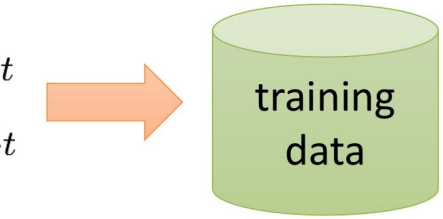
Do this work? No

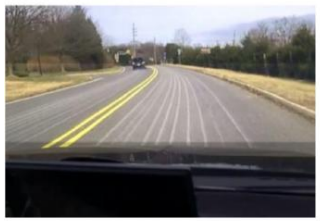
Let's consider the case of steering a car.
We can observe human drivers to easily generate training data to feed a supervised learning approach



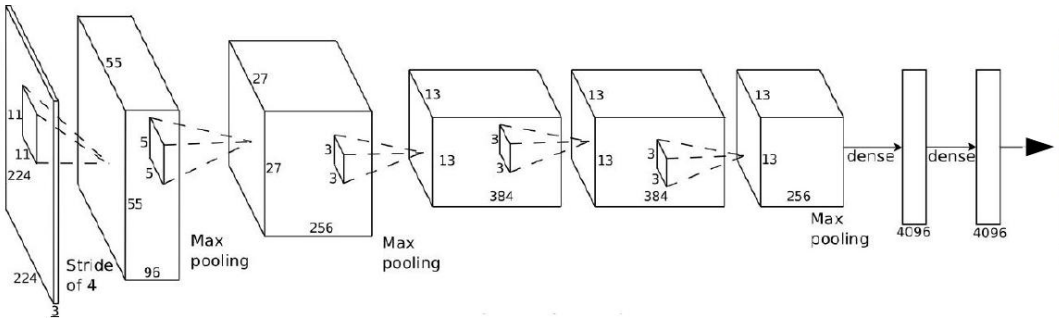
$\mathbf{o}_t$

$\mathbf{u}_t$





$\mathbf{o}_t$

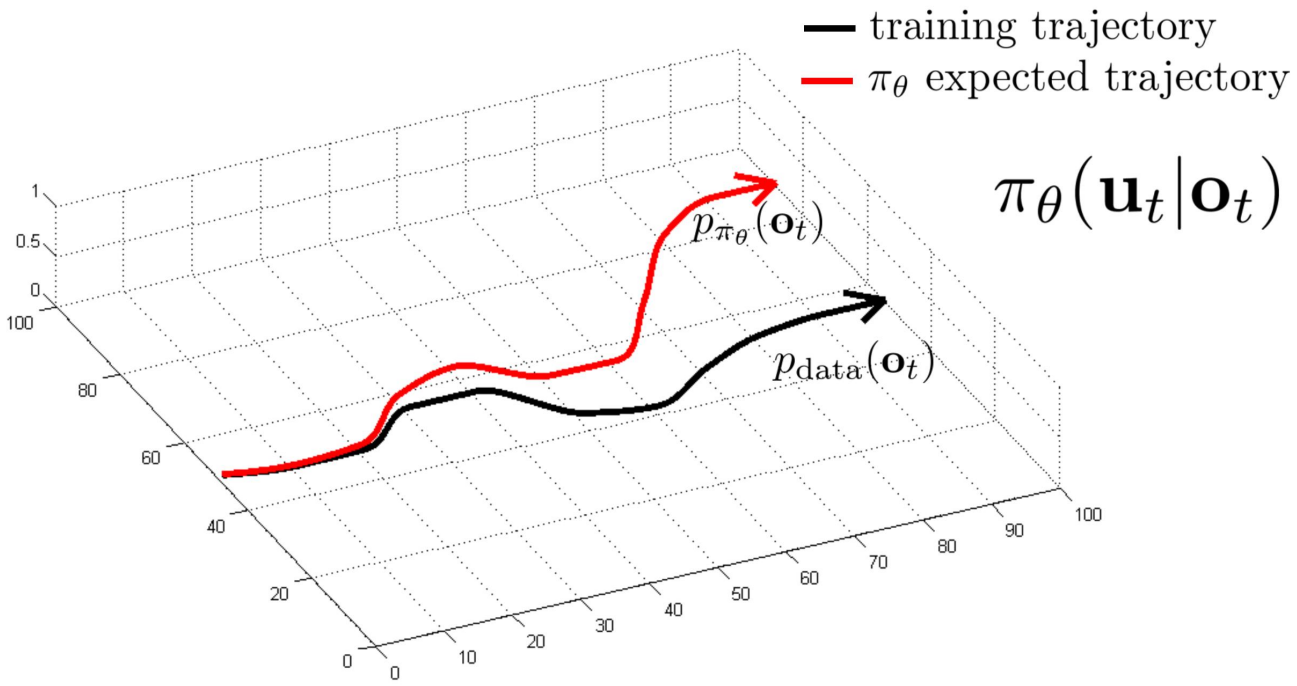


$\mathbf{u}_t$

Do this work? No

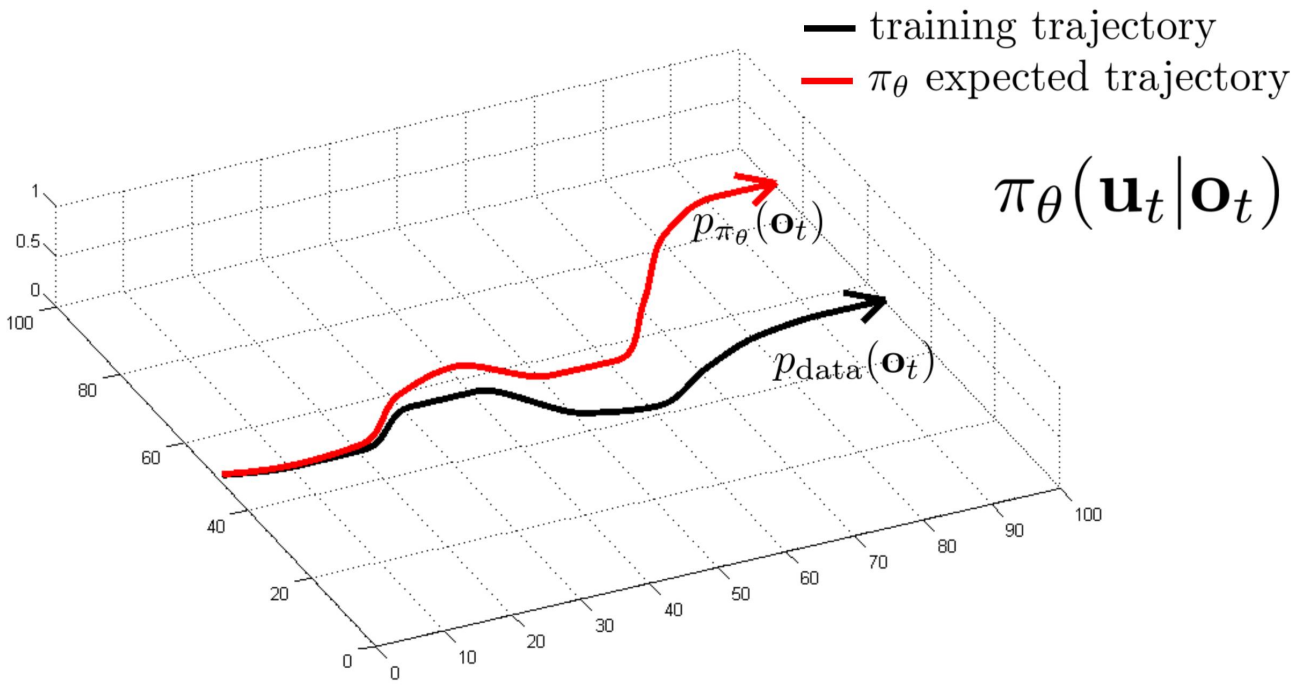
Why? It is very difficult to generate good training data (good:= diverse and expert behavior)

Trayectory matching problem



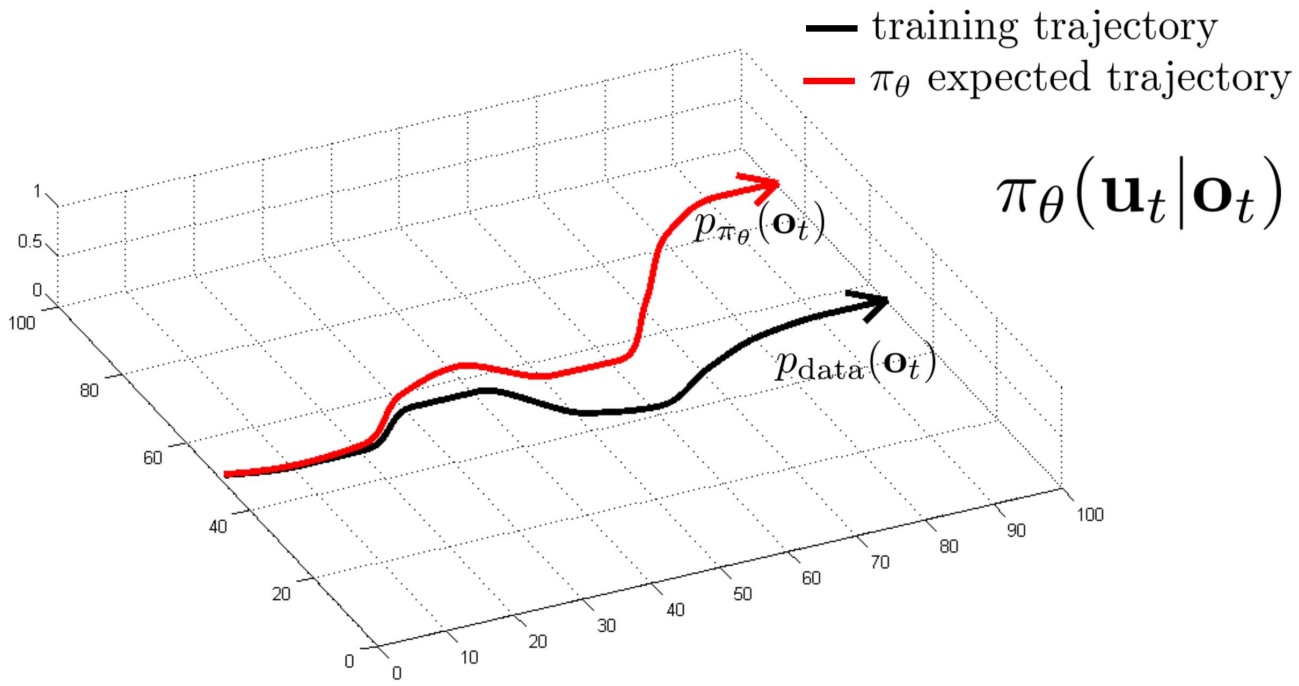
- At test time the dynamic of the game can diverge from similar trayectories in the training data.

Trayectory matching problem



- At test time the dynamic of the game can diverge from similar trayectories in the training data.
- In other words, there is a distribution mismatch between the training and test data.

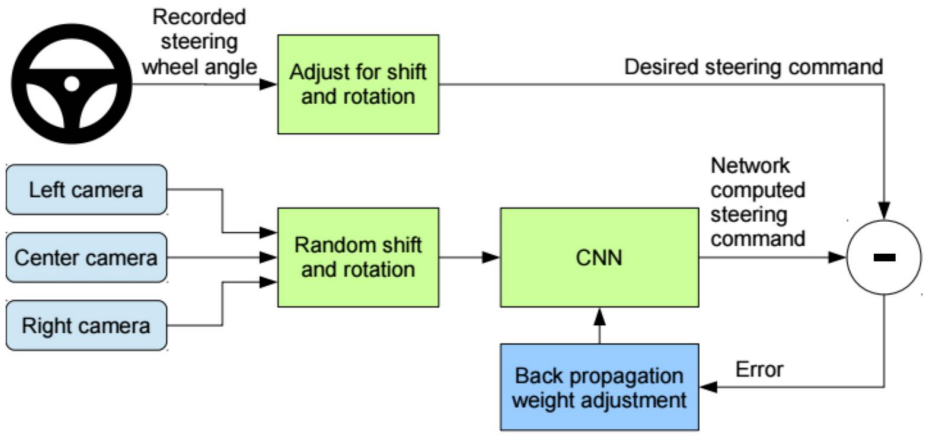
Trayectory matching problem

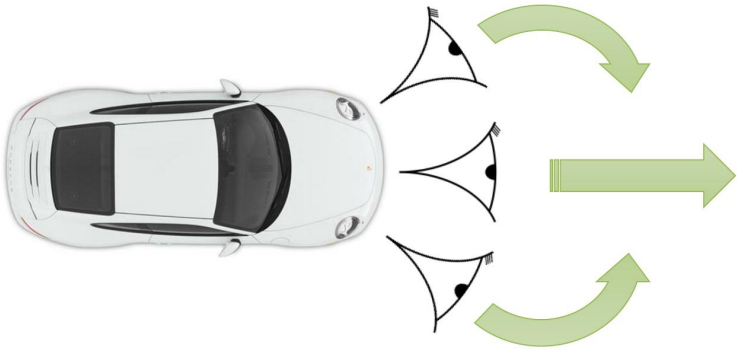


- At test time the dynamic of the game can diverge from similar trayectories in the training data.
- In other words, there is a distribution mismatch between the training and test data.
- Ex. human driving data won't generate much examples of critical situations.

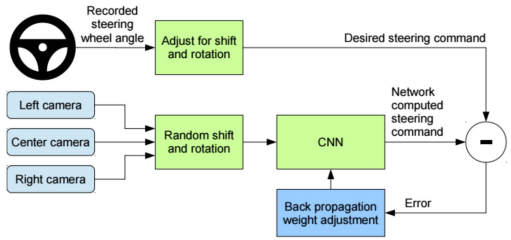
Solution 1: Being clever

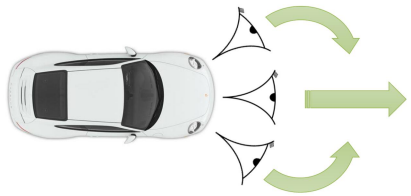
This do work !

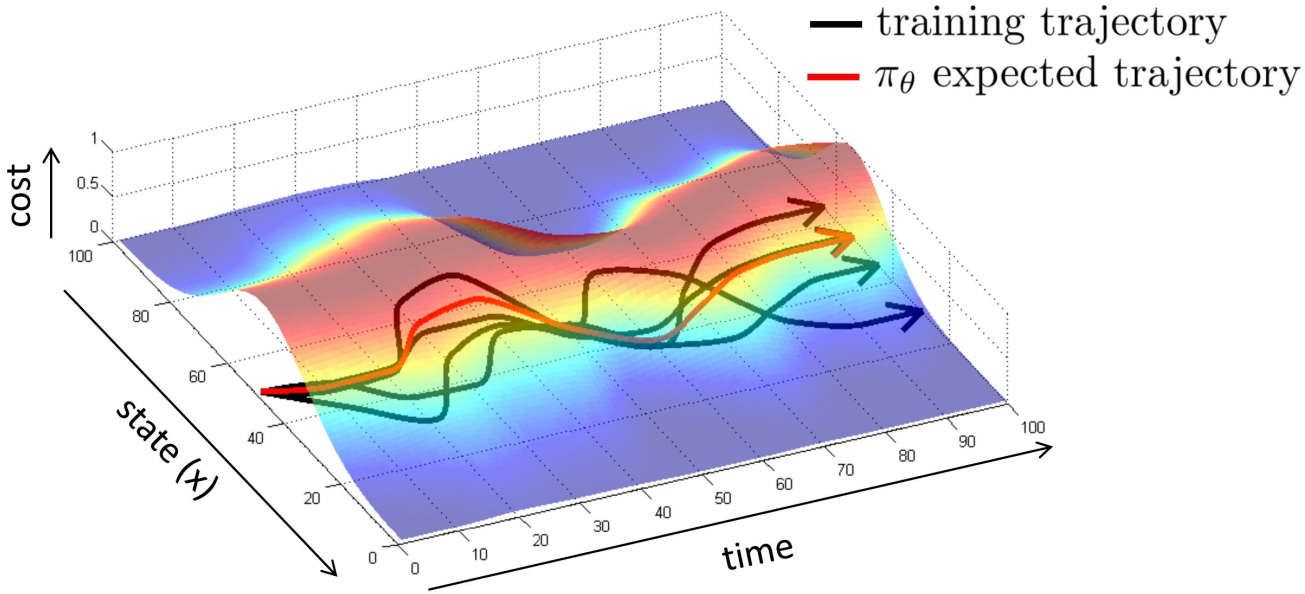




Solution 1: Being clever

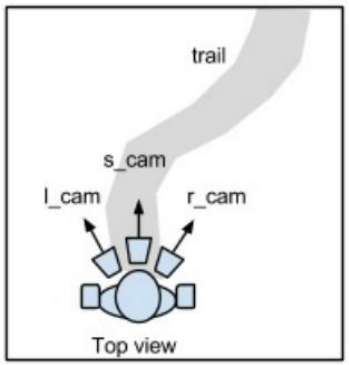


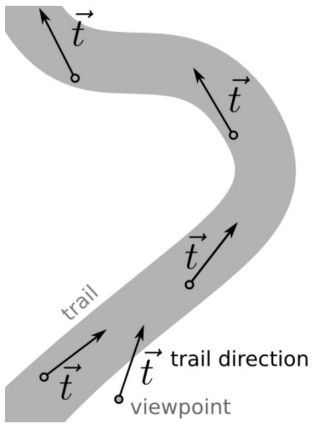


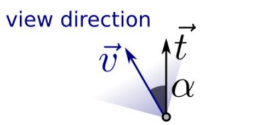


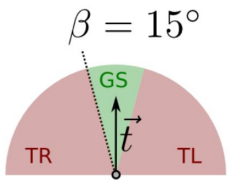
Being clever: Example 2

Dron Path Following in the Forest
Giusti et al., 2015



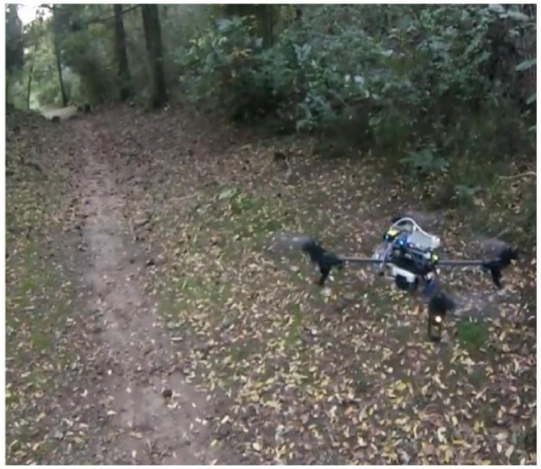


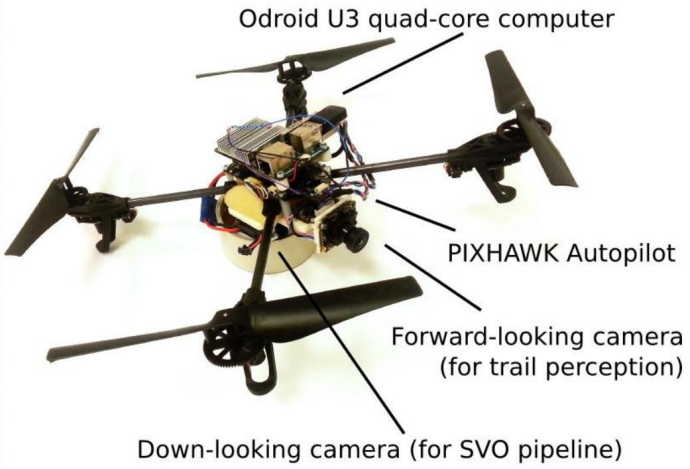




Being clever: Example 2

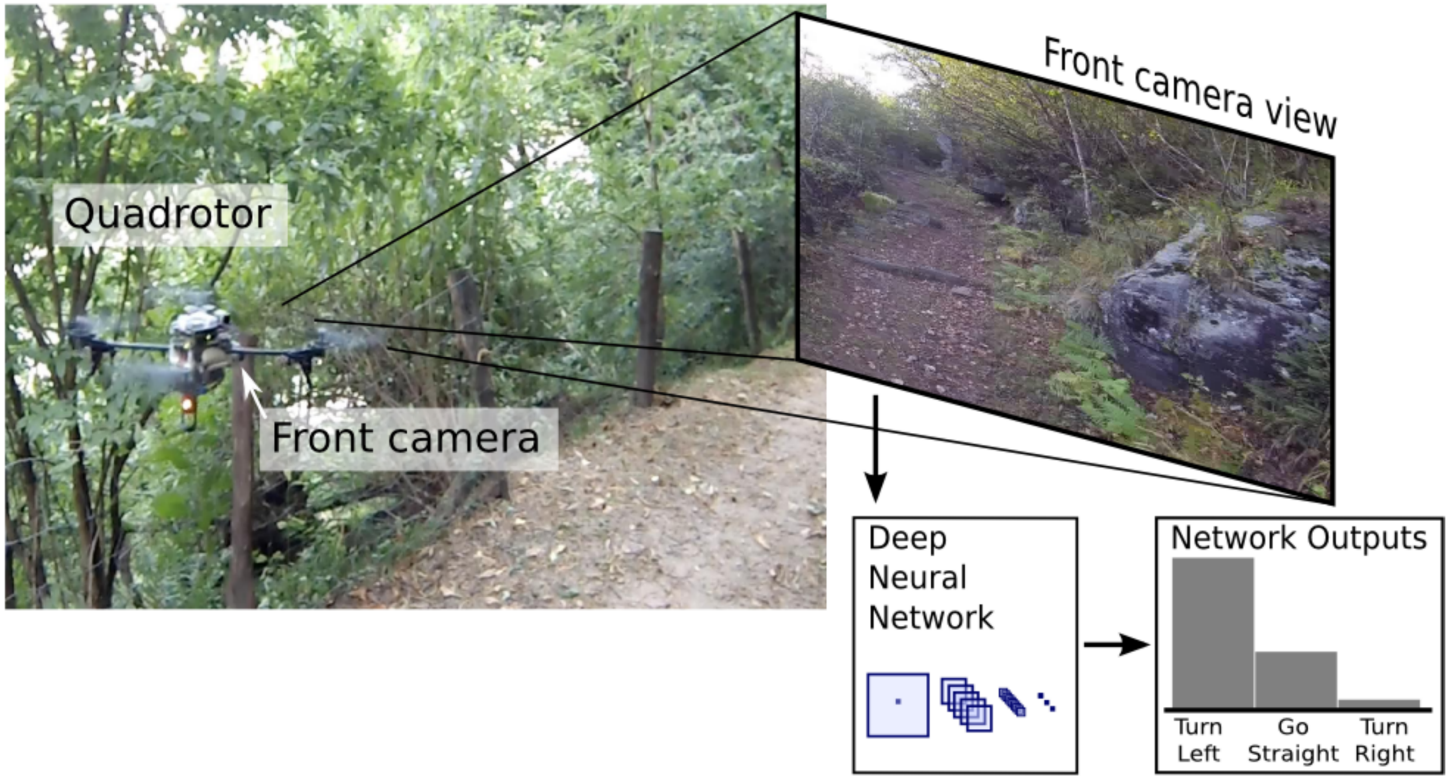
Dron Path Following in the Forest
Giusti et al., 2015





Being clever: Example 2

Dron Path Following in the Forest
Giusti et al., 2015



Being clever: Example 2

Dron Path Following in the Forest
Giusti et al., 2015

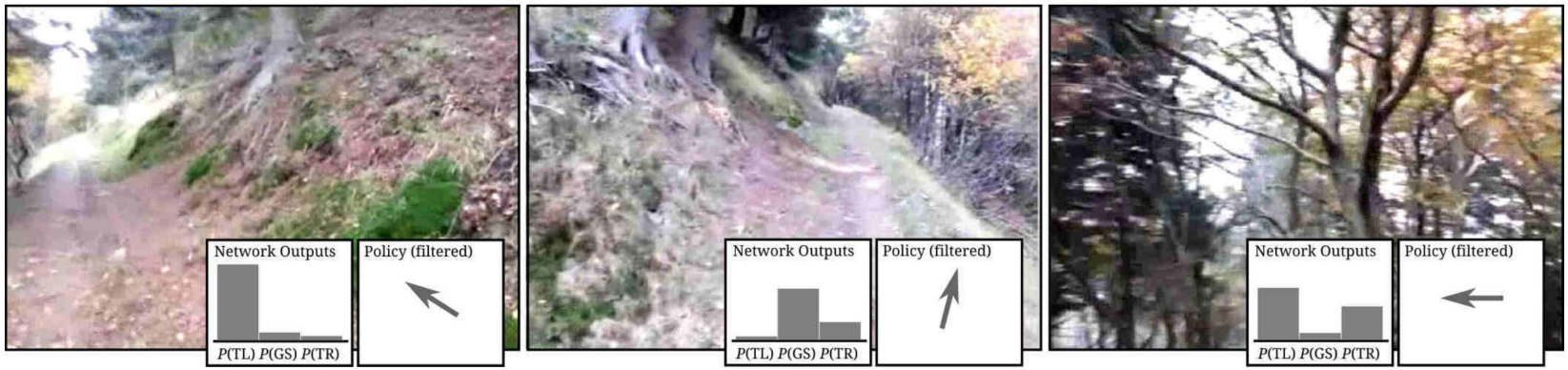


TABLE I: Results for the three-class problem.

	DNN	Saliency	[12]	Human1	Human2
Accuracy	85.2%	52.3%	36.5%	86.5%	82.0%

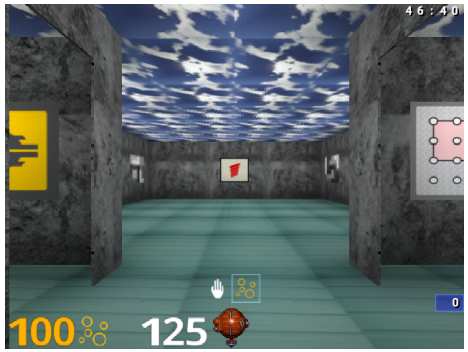
TABLE II: Results for the two-class problem.

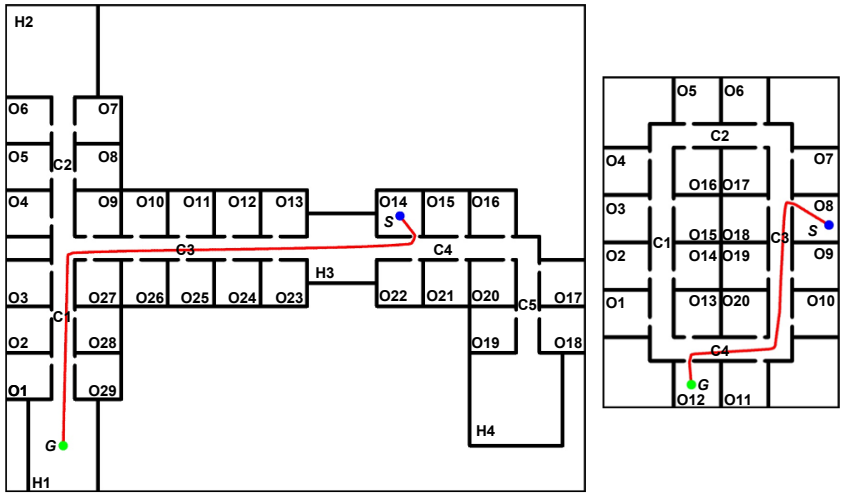
	DNN	Saliency	[12]	Human1	Human2
Accuracy	95.0%	73.6%	57.9%	91.0%	88.0%
Precision	95.3%	60.9%	39.8%	79.7%	84.0%
Recall	88.7%	46.6%	64.6%	95.1%	81.6%
AUC	98.7%	75.9%	–	–	–

Solution 2: Simulation

Imitation learning for indoor autonomous robot navigation

G. Sepúlveda et al., 2018

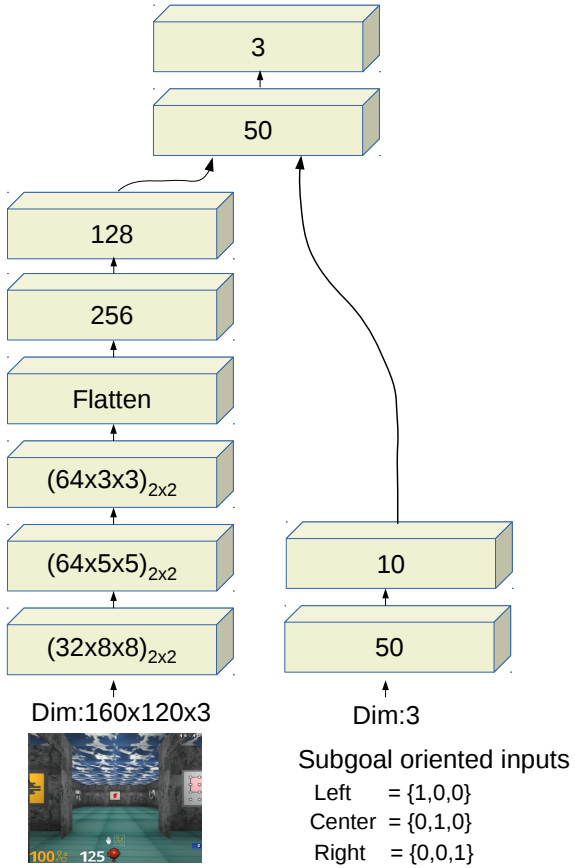




Solution 2: Simulation

Imitation learning for indoor autonomous robot navigation

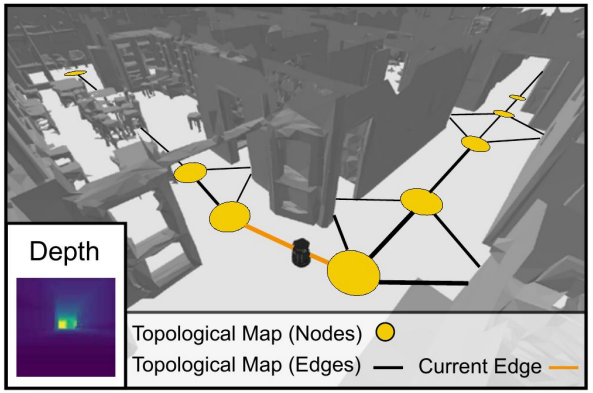
G. Sepúlveda et al., 2018



Solution 2: Simulation

Imitation learning for indoor autonomous robot navigation

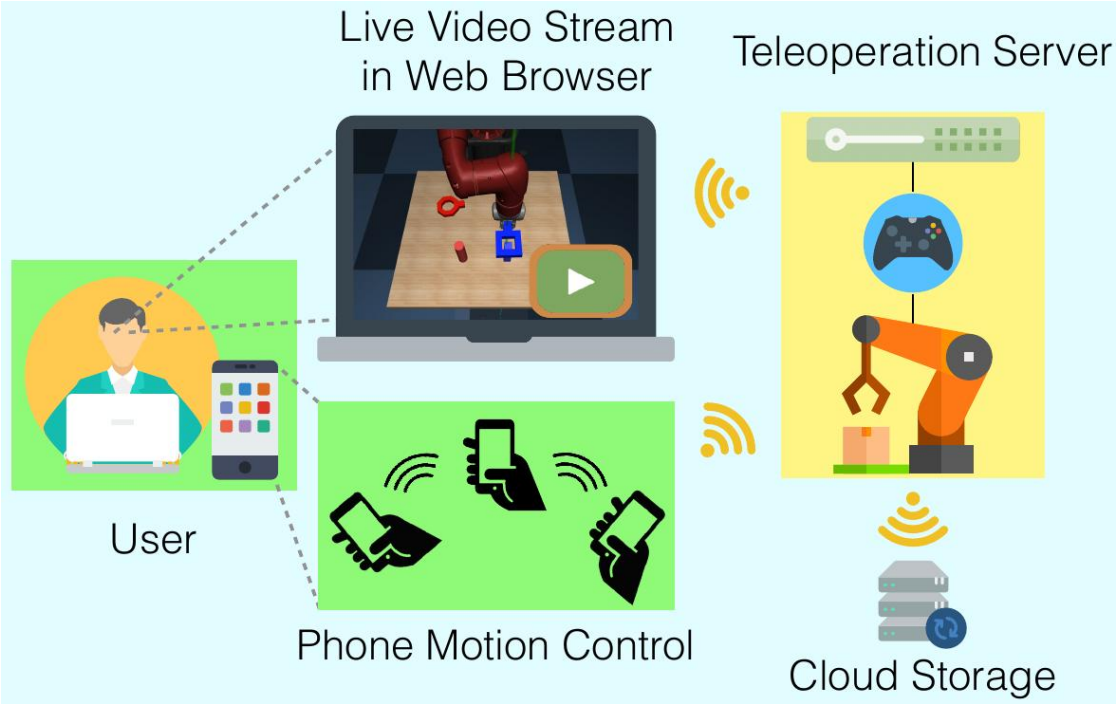
G. Sepúlveda et al., 2018



Solution 3: Creativity + Simulation

RoboTurk: A Crowdsourcing Platform for Robotic Skill Learning through Imitation

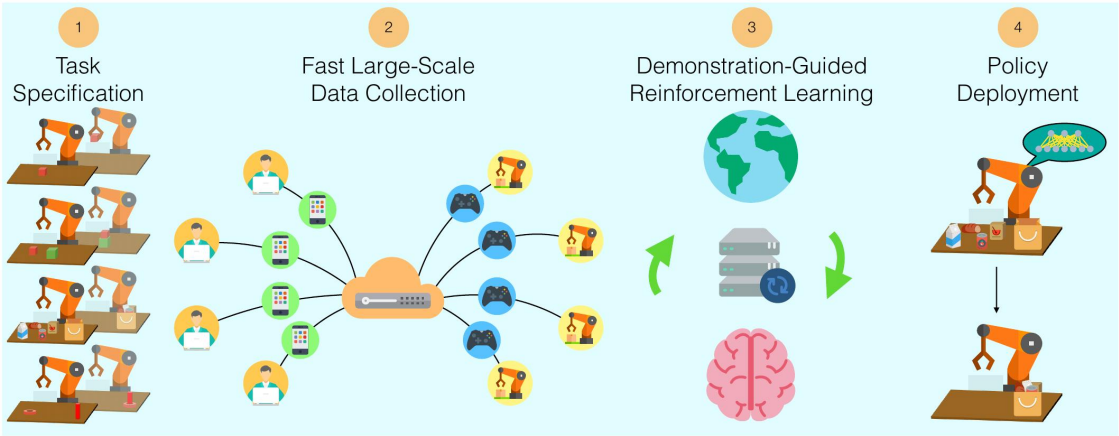
Ajay Mandlekar et al., 2018

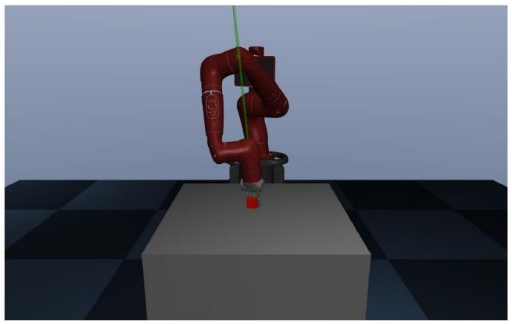


Solution 3: Creativity + Simulation

RoboTurk: A Crowdsourcing Platform for Robotic Skill Learning through Imitation

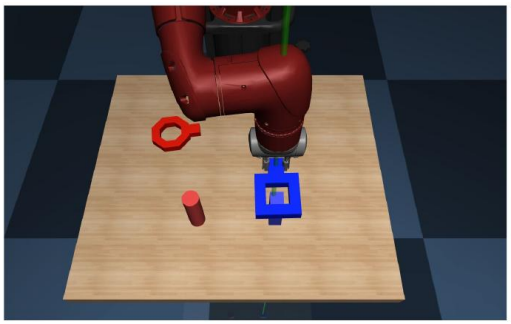
Ajay Mandlekar et al., 2018





Lifting

Picking



Assembly

Solution 3: Creativity + Simulation

- Approaches like RoboTurk are useful, however, they open relevant challenges:

Solution 3: Creativity + Simulation

- Approaches like RoboTurk are useful, however, they open relevant challenges:
  - How to guarantee diversity?

Solution 3: Creativity + Simulation

- Approaches like RoboTurk are useful, however, they open relevant challenges:
  - How to guarantee diversity?
  - How to guarantee expert behavior (optimality)?

Solution 3: Creativity + Simulation

- Approaches like RoboTurk are useful, however, they open relevant challenges:
  - How to guarantee diversity?
  - How to guarantee expert behavior (optimality)?
  - How to move from simulation to real situations?

Solution 3: Creativity + Simulation

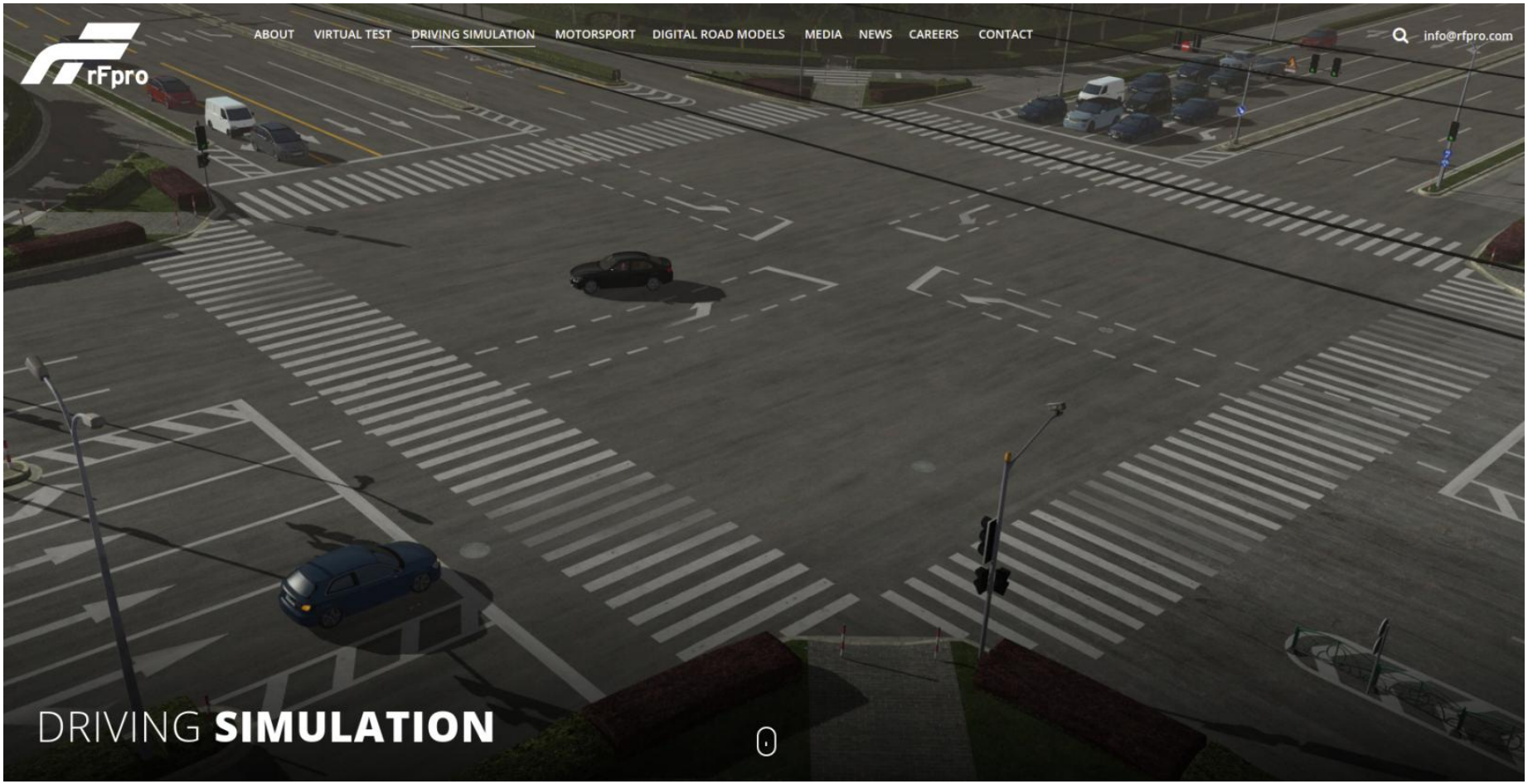
- Approaches like RoboTurk are useful, however, they open relevant challenges:
  - How to guarantee diversity?
  - How to guarantee expert behavior (optimality)?
  - How to move from simulation to real situations?
    - One possibility is the use of domain adaptation techniques.

Solution 3: Creativity + Simulation

- Approaches like RoboTurk are useful, however, they open relevant challenges:
  - How to guarantee diversity?
  - How to guarantee expert behavior (optimality)?
  - How to move from simulation to real situations?
    - One possibility is the use of domain adaptation techniques.
    - Another one is the use of good simulations.

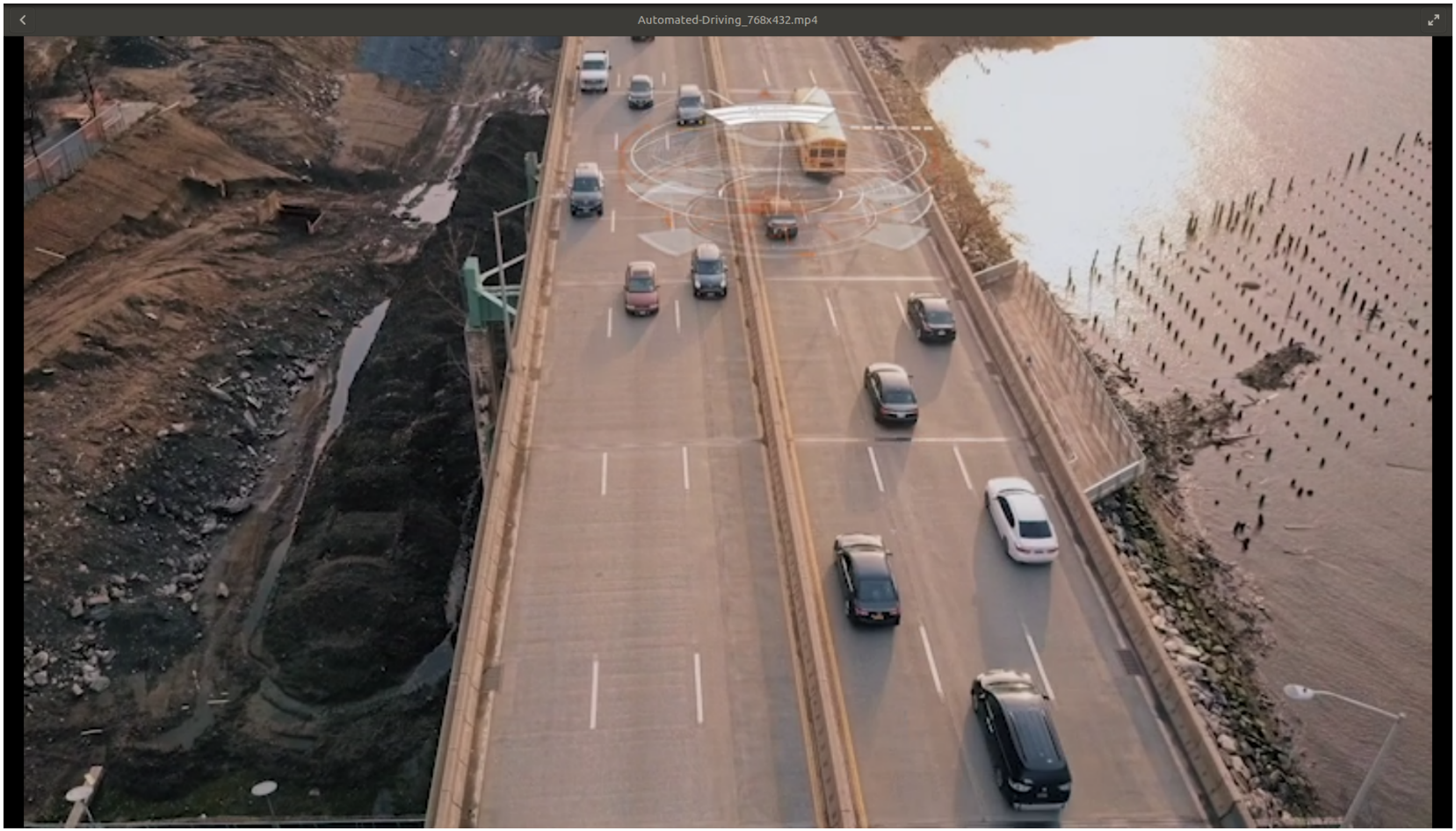
Simulation closing the loop: photo realistic scenarios

Simulations for autonomous driving



Simulation closing the loop: photo realistic scenarios

Simulations for autonomous driving

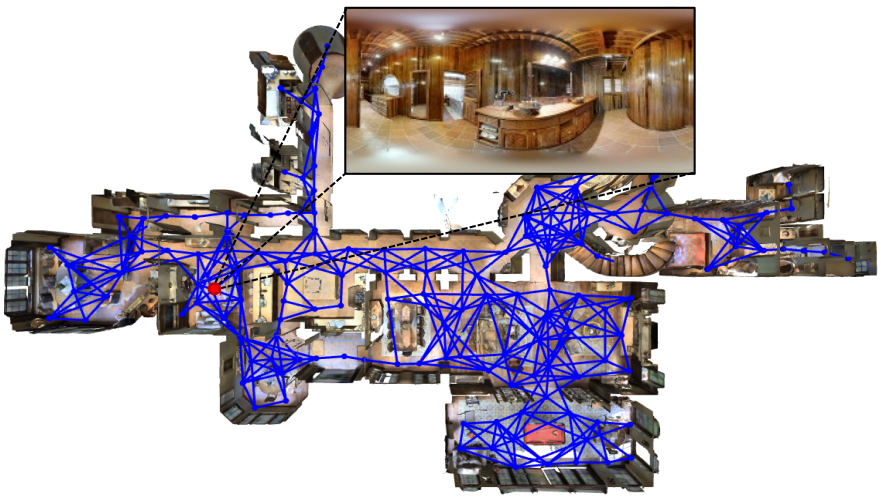


Simulation closing the loop: photo realistic scenarios

AI2-Thor: Simulations for indoor robot applications

Simulation closing the loop: photo realistic scenarios

MatterPort: Simulations for indoor robot applications



Solution 4: Mix of simulation and human feedback

- Training agent in simulation can help, however, it can suffer from problems:

Solution 4: Mix of simulation and human feedback

- Training agent in simulation can help, however, it can suffer from problems:

Solution 4: Mix of simulation and human feedback

- Training agent in simulation can help, however, it can suffer from problems:
  - Creating a good simulation is not trivial.

Solution 4: Mix of simulation and human feedback

- Training agent in simulation can help, however, it can suffer from problems:
  - Creating a good simulation is not trivial.
  - Transfering knowledge to real world is not an easy task.

Solution 4: Mix of simulation and human feedback

- Training agent in simulation can help, however, it can suffer from problems:
  - Creating a good simulation is not trivial.
  - Transfering knowledge to real world is not an easy task.
- Training agent using data from expert can work, however, it can suffer from problems:

Solution 4: Mix of simulation and human feedback

- Training agent in simulation can help, however, it can suffer from problems:
  - Creating a good simulation is not trivial.
  - Transfering knowledge to real world is not an easy task.
- Training agent using data from expert can work, however, it can suffer from problems:
  - Lack of diversity (distribution mismatch problem)

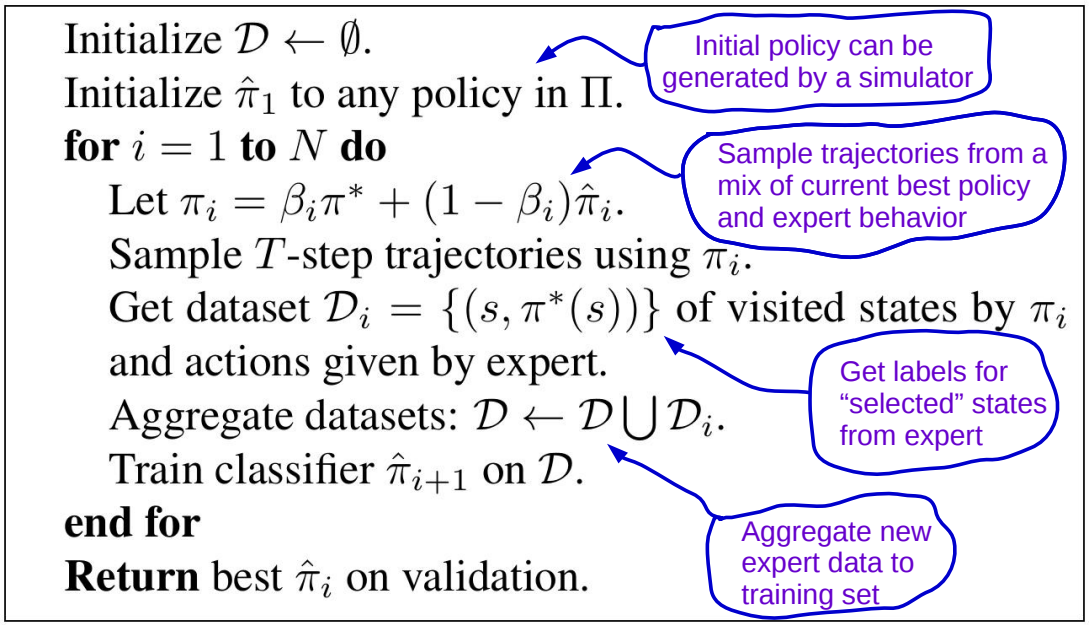
Solution 4: Mix of simulation and human feedback

- Training agent in simulation can help, however, it can suffer from problems:
  - Creating a good simulation is not trivial.
  - Transfering knowledge to real world is not an easy task.
- Training agent using data from expert can work, however, it can suffer from problems:
  - Lack of diversity (distribution mismatch problem)
  - Difficult to obtain enough expert data.

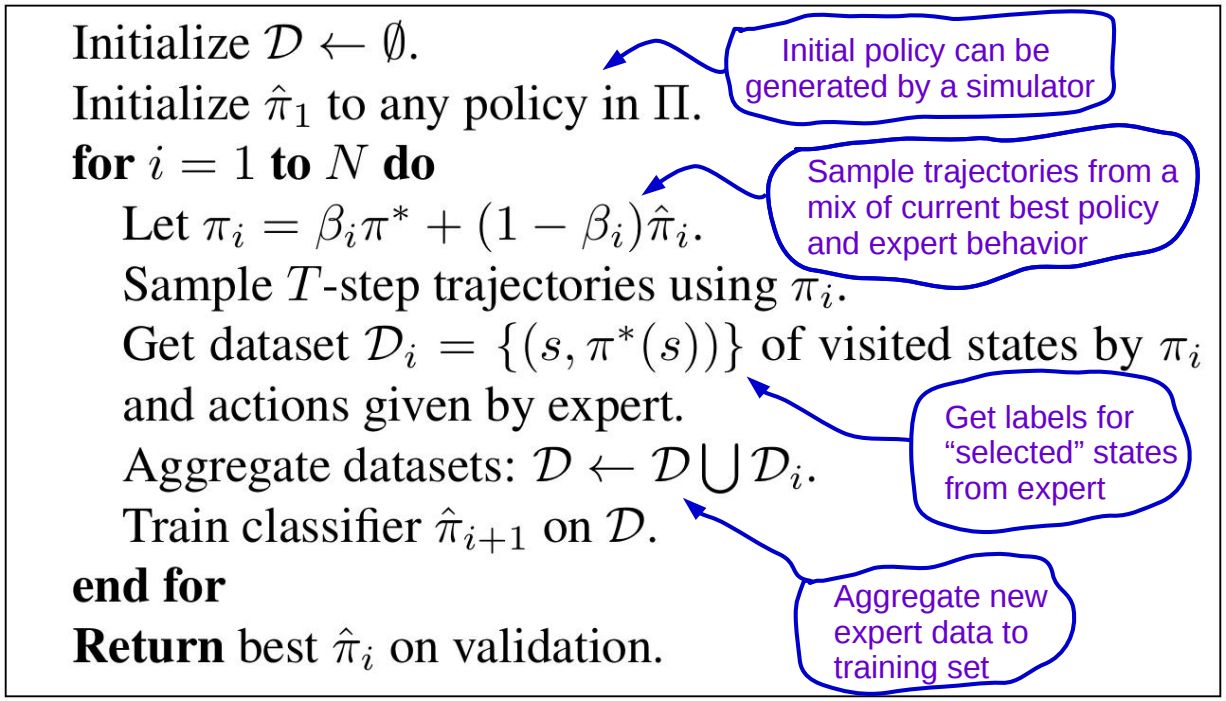
Solution 4: Mix of simulation and human feedback

Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning *S. Ross et al., 2011*

DAGGER (Dataset Agregation) Algorithm



Key idea: run current policy to obtain state visits. Then get feedback from expert to know what to do on those states. Update policy and repeat.



- DAGGER proceeds by collecting a dataset at each iteration under the current policy and trains the next policy under the aggregate of all collected datasets.
- At the beginning $\beta_i = 1$, i.e., DAGGER queries the expert to choose controls (initial uninformed policy visit states that are usually irrelevant).
- Over the iterations, β_i get closer to zero, i.e., DAGGER collects inputs that the learned policy is likely to encounter during its execution based on previous experience.

Conclusions

- Nature has selected RL and Imitation learning as some of its main tools to develop intelligent individuals. They are poweful!.

Conclusions

- Nature has selected RL and Imitation learning as some of its main tools to develop intelligent individuals. They are poweful!.
- However, in the context of IA, it is not so easy to apply them. People are not so forgiveness with machine mistakes.

Conclusions

- Nature has selected RL and Imitation learning as some of its main tools to develop intelligent individuals. They are poweful!.
- However, in the context of IA, it is not so easy to apply them. People are not so forgiveness with machine mistakes.
- Artificial and simulated environment help, but they still do not close the gap with a real environment.

Conclusions

- Nature has selected RL and Imitation learning as some of its main tools to develop intelligent individuals. They are poweful!.
- However, in the context of IA, it is not so easy to apply them. People are not so forgiveness with machine mistakes.
- Artificial and simulated environment help, but they still do not close the gap with a real environment.
- Stay tune!, there are tons of applications in decision systems, robotics, and softbots.