Imitation Learning (Behavior Cloning)

Fabian Caba Heilbron

IVUL – February 16

http://rll.berkeley.edu/deeprlcourse/docs/week_2_lecture_1_behavior_cloning.pdf

Imitation vs Reinforcement

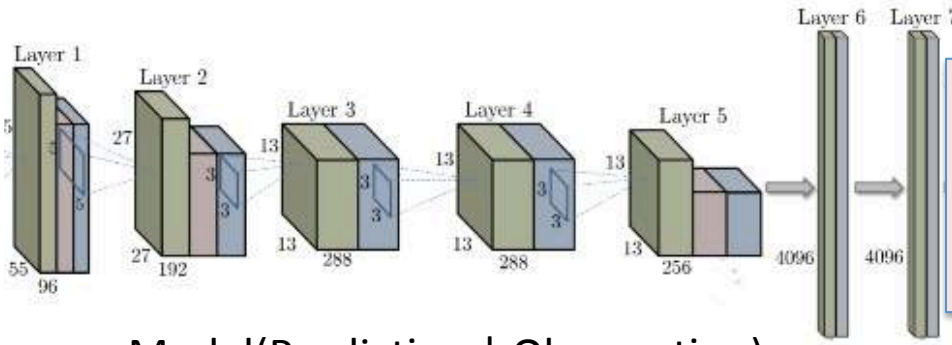
Define a reward for each task!

Today's talk

- Definition of sequential problem
- Imitation learning: supervised learning for decision making
  - Does imitation learning works?
  - How can we make it work more often?
- Case studies of recent work in (deep) imitation learning

Supervised Learning

Observation



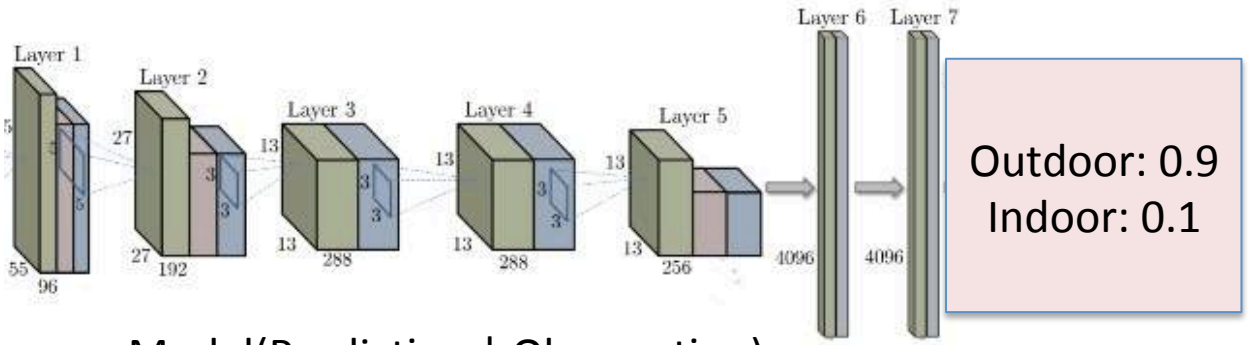
Model(Prediction | Observation)

Tiger: 0.85
Cat: 0.1
…
Fish: 0.001

Animal Prediction

Supervised Learning

Observation

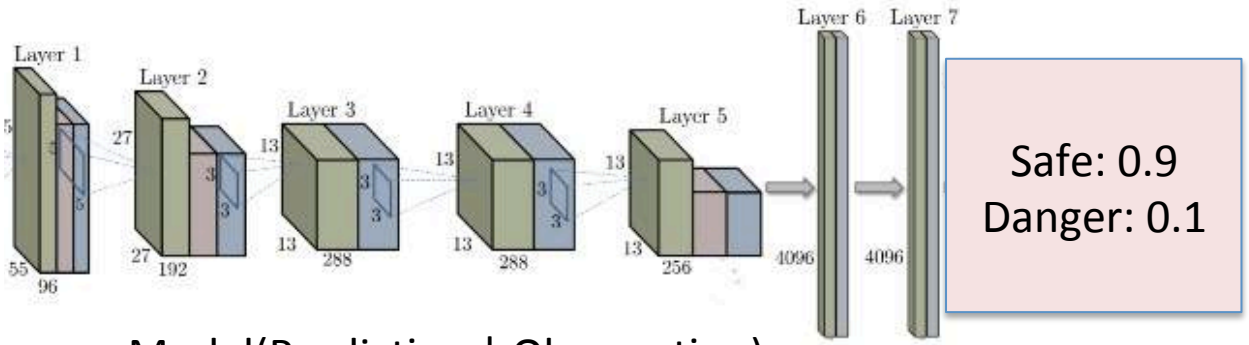


Model(Prediction | Observation)

Scene Prediction

Supervised Learning

Observation

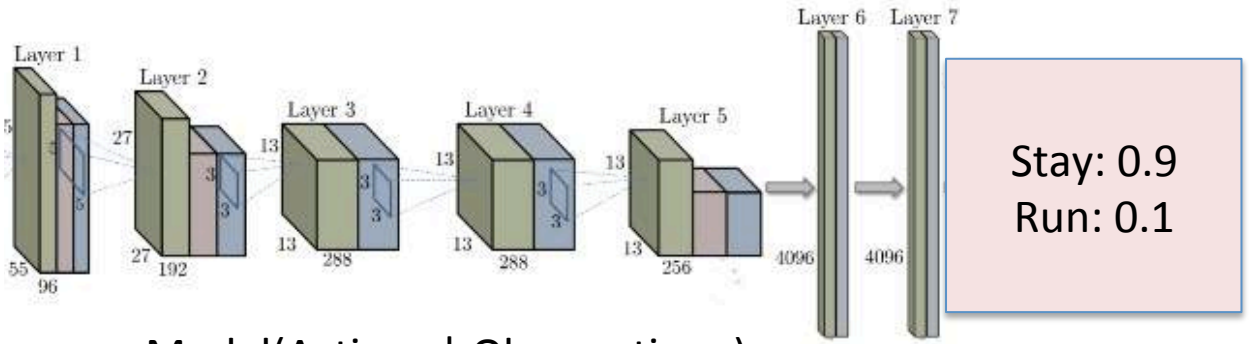


Model(Prediction | Observation)

Safety Prediction

Sequential decision problem

Observation at time 0

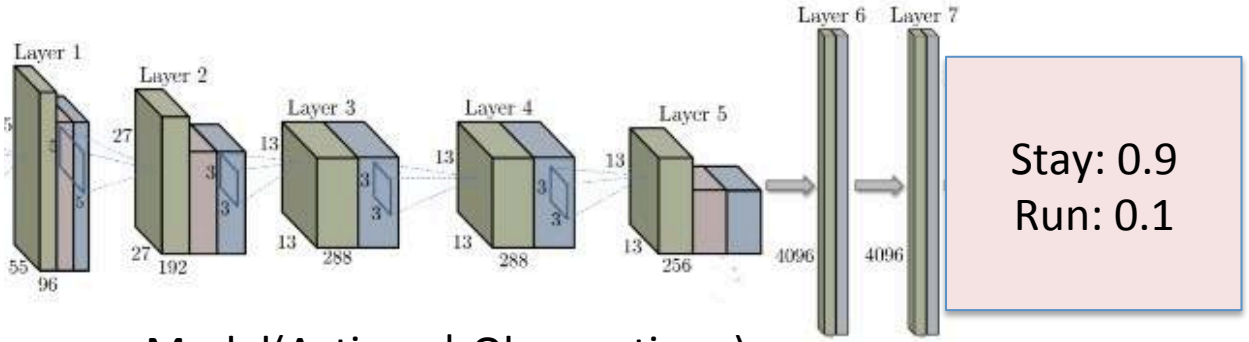


Model(Action$_0$ | Observation$_0$)

Action at time 0

Sequential decision problem

Observation at time 0

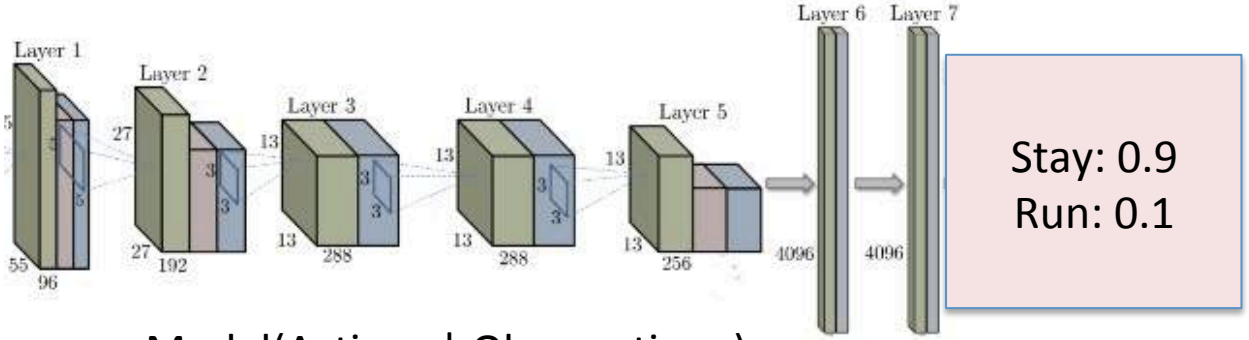


Model(Action$_0$ | Observation$_0$)

Stay: 0.9
Run: 0.1

Action at time 0

Sequential decision problem



Stay: 0.9
Run: 0.1

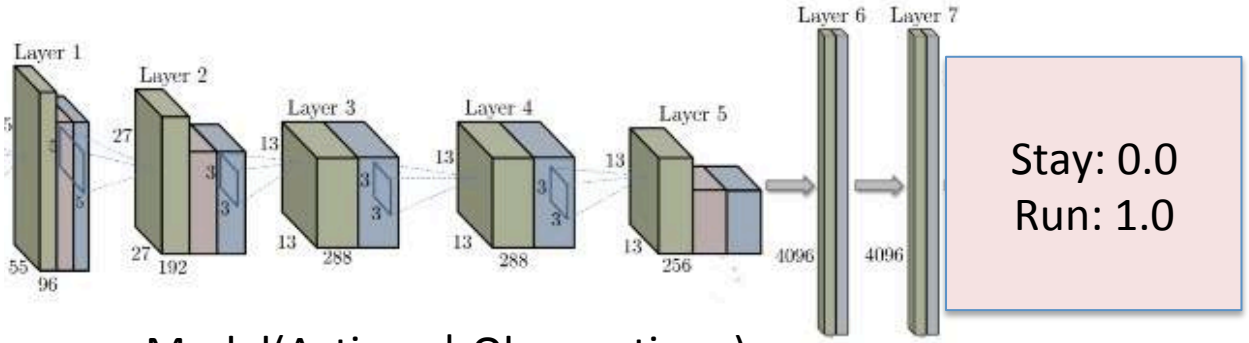
Model(Action$_0$ | Observation$_0$)

Observation at Time 1

Action at Time 0

Sequential decision problem

Observation at Time 1

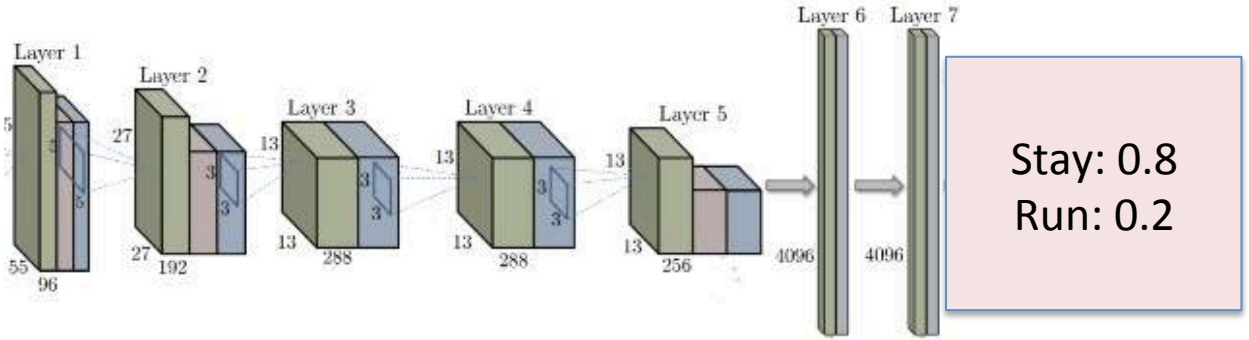


Model(Action$_1$ | Observation$_1$)

Action at Time 1

Sequential decision problem

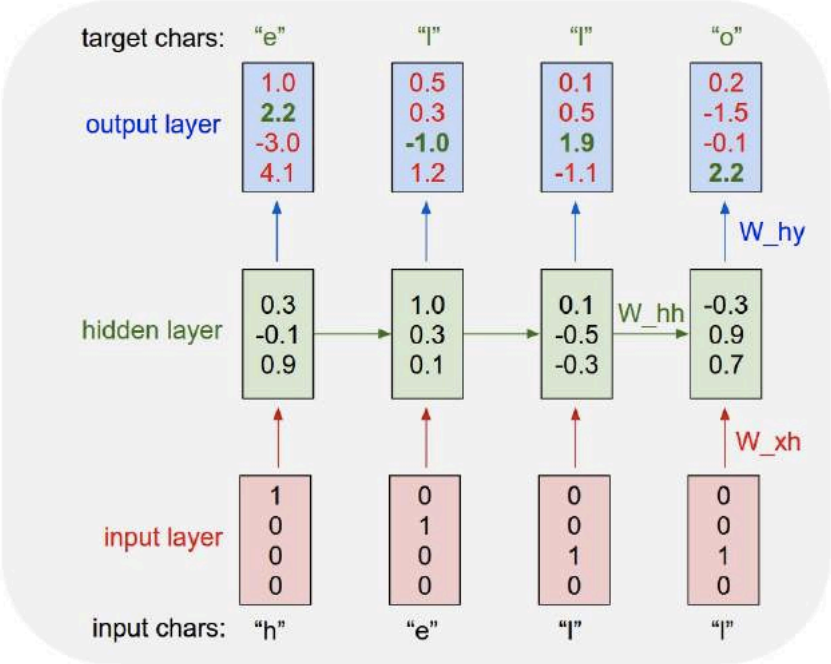
Observation at Time t



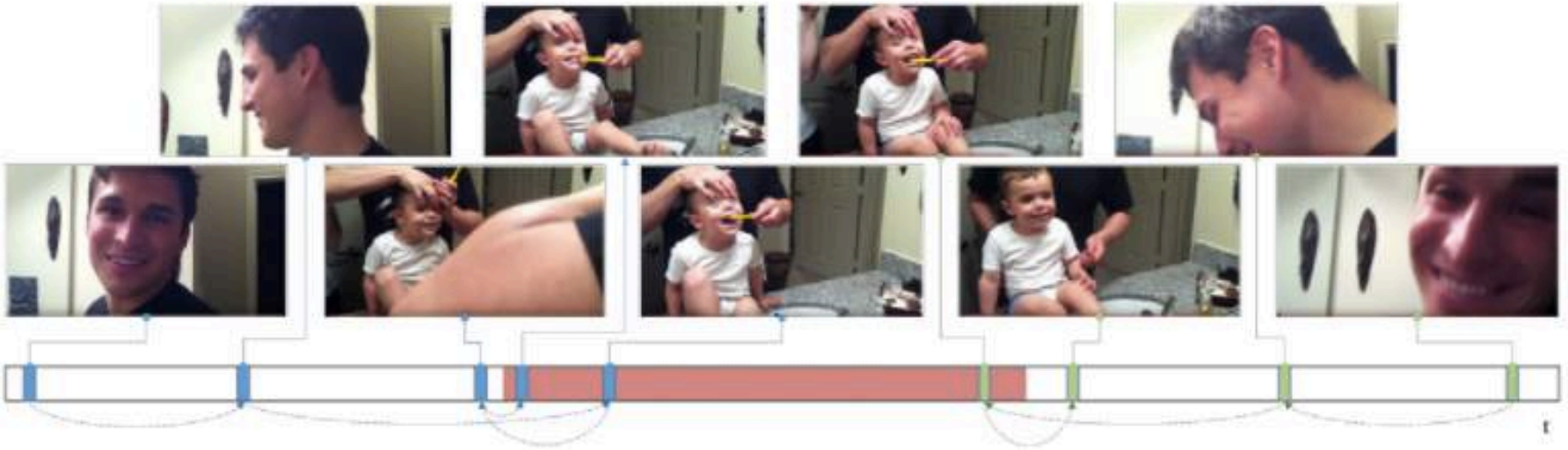
Model(Action$_t$ | (Observation$_t$, , Action$_{t-1}$))

Action at Time t

Sequential decision problem examples



Sequential decision problem examples

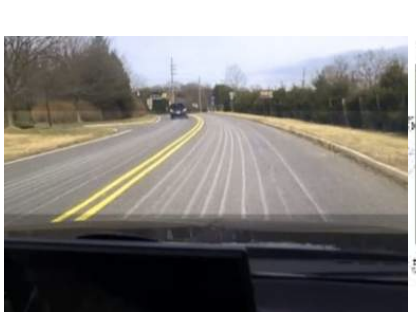


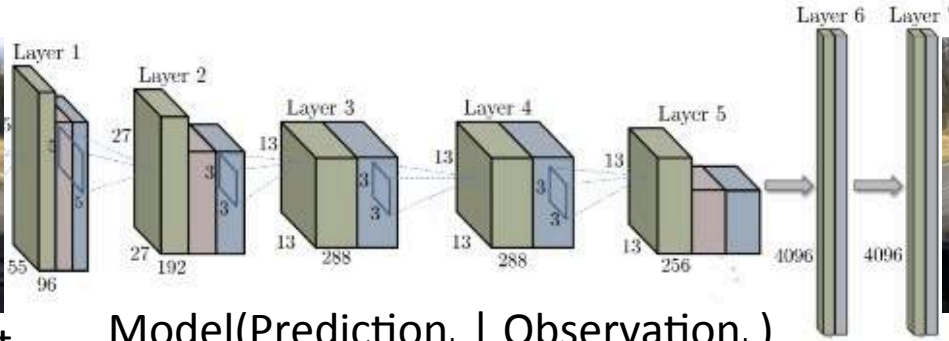
Learning to search actions

Today's talk

- ~~Definition of sequential problem~~
- Imitation learning: supervised learning for decision making
  - Does imitation learning works?
  - How can we make it work more often?
- Case studies of recent work in (deep) imitation learning

Steering wheel prediction



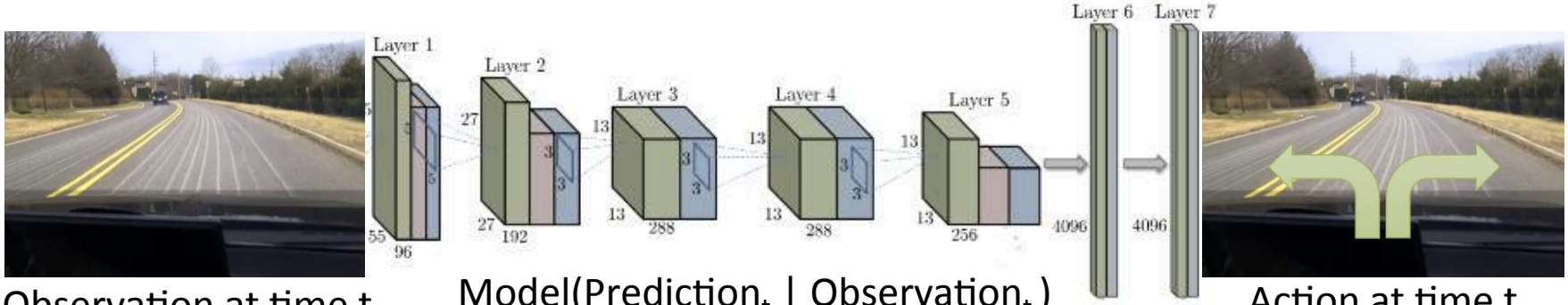


Observation at time t

Model(Prediction$_t$ | Observation$_t$)

Action at time t

Steering wheel prediction



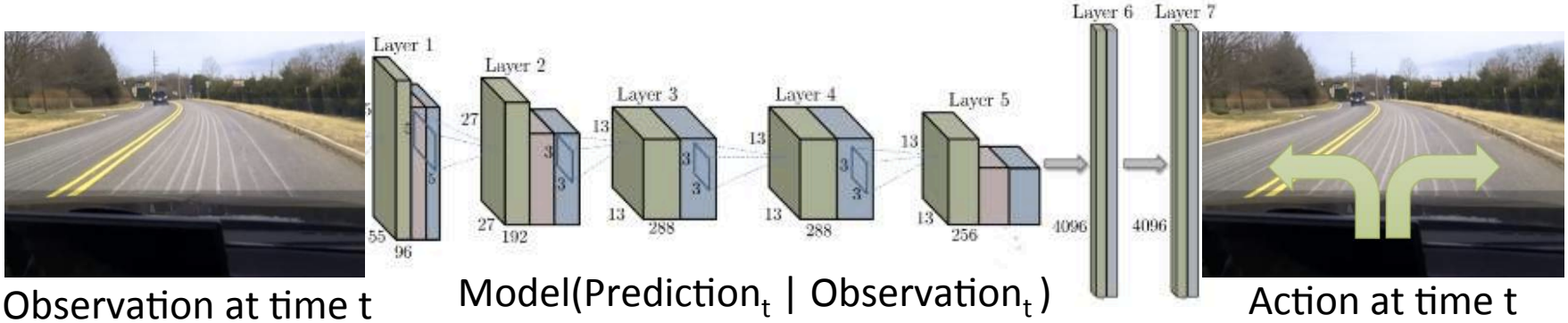
Observation at time t

Model(Prediction$_t$ | Observation$_t$)

Action at time t

Collect observation and actions

Steering wheel prediction



Observation at time t

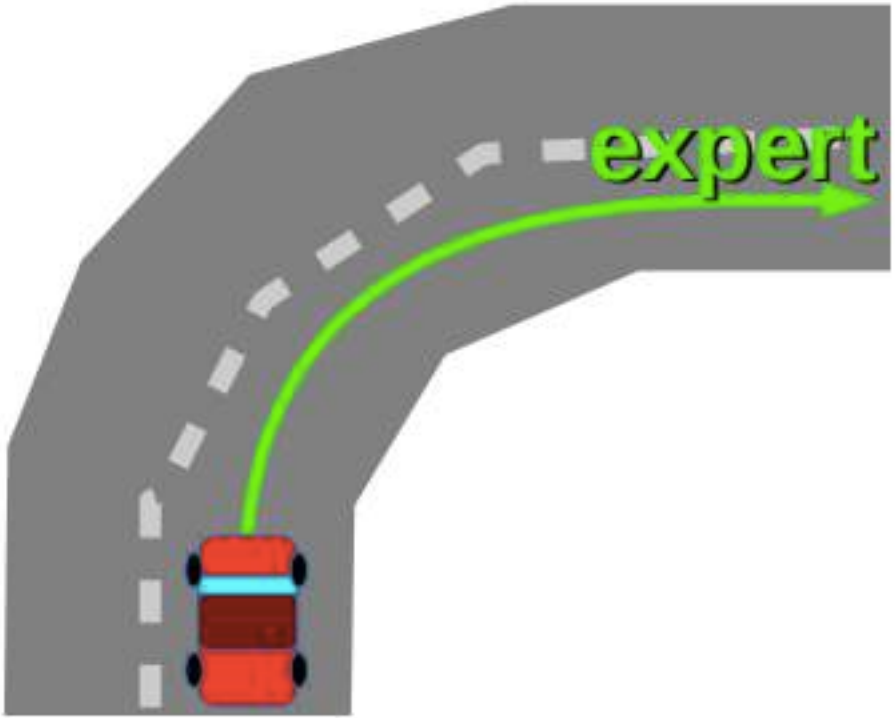
Model(Prediction$_t$ | Observation$_t$)

Action at time t

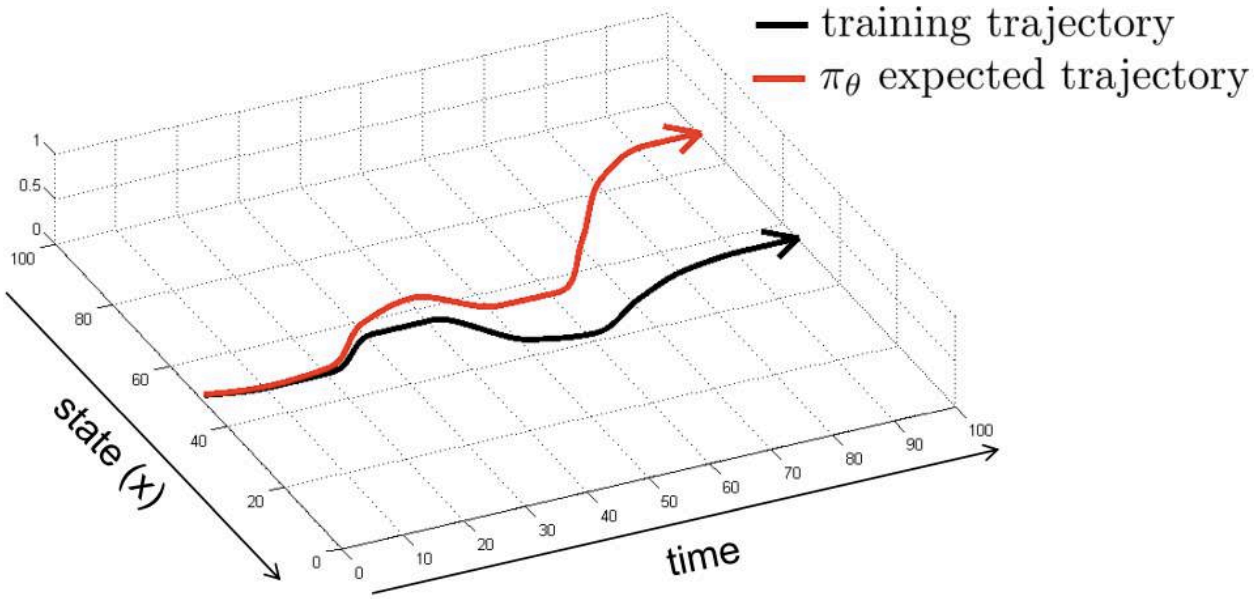
Collect observation and actions

Should it work?

Should it work? **NO**

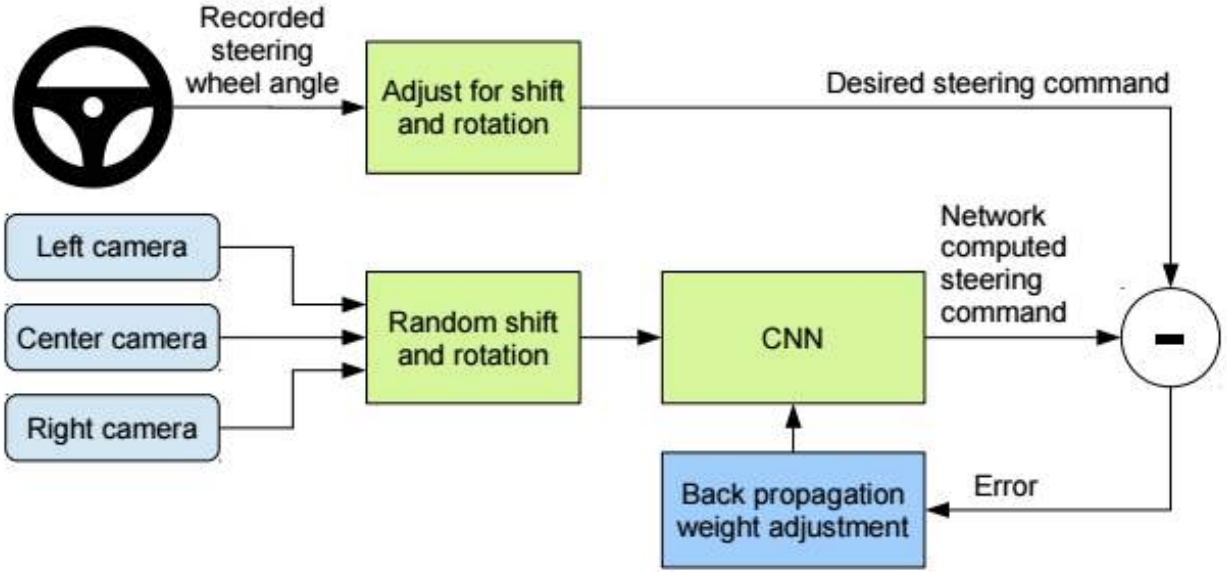


Should it work? **NO**

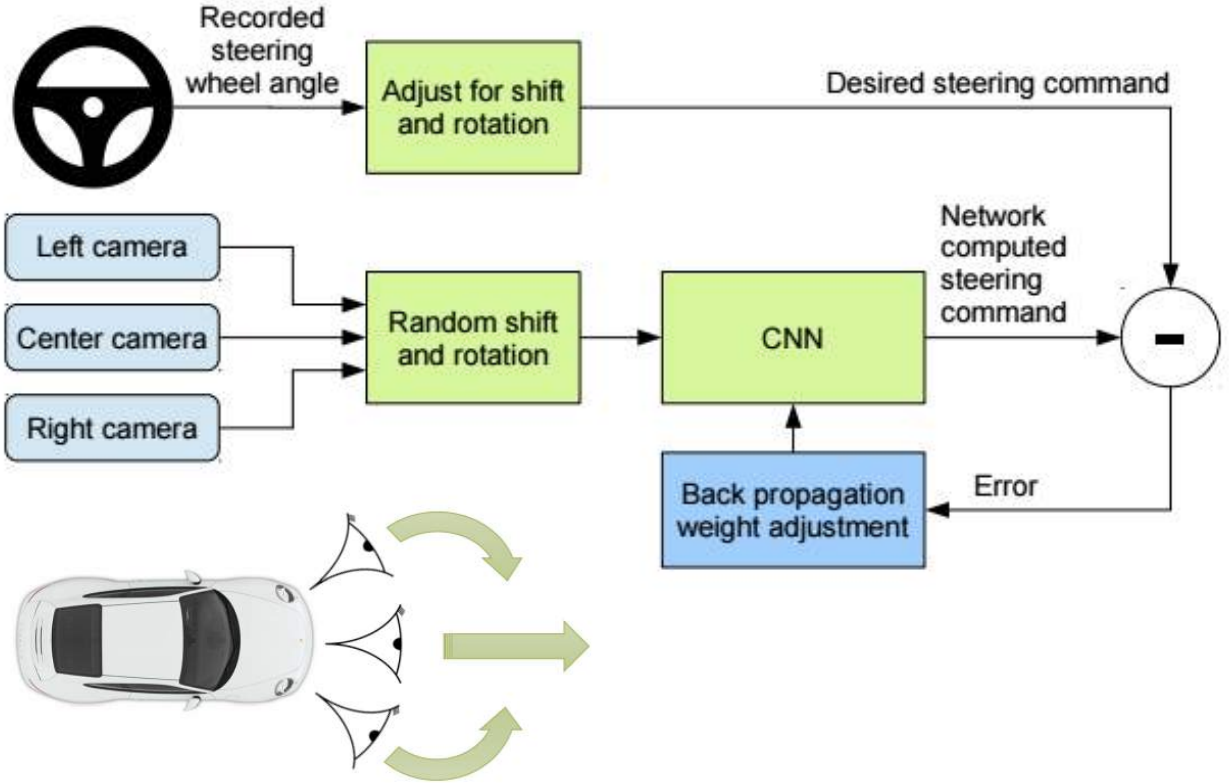


Does it work? **Yes!** **NVidia**

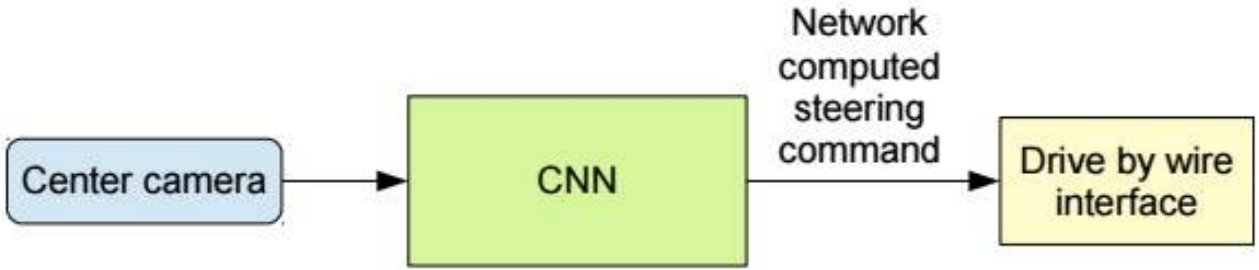
Why it works?


Recorded steering wheel angle
Adjust for shift and rotation
Desired steering command
Left camera
Center camera
Right camera
Random shift and rotation
CNN
Network computed steering command
-
Back propagation weight adjustment
Error

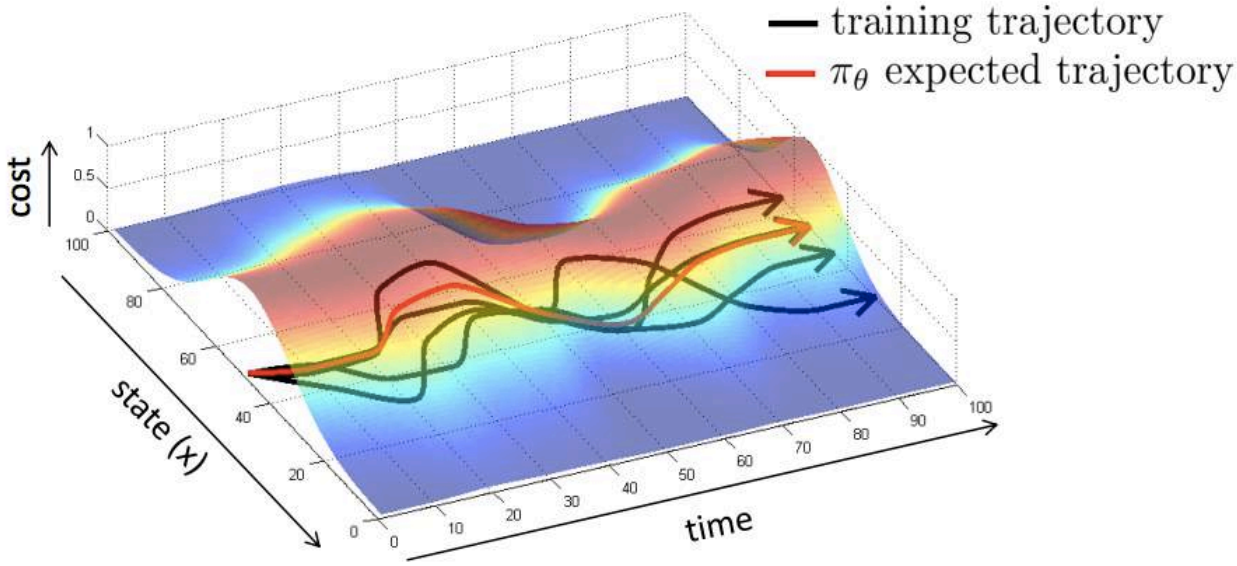
Why it works? It's a Hack!



Why it works? It's a Hack!



Can we make it work more often?



Can we make it work more often?

DAgger: Dataset Aggregation

goal: collect training data from $p_{\pi_\theta}(\mathbf{o}_t)$ instead of $p_{\text{data}}(\mathbf{o}_t)$

how? just run $\pi_\theta(\mathbf{u}_t|\mathbf{o}_t)$

but need labels $\mathbf{u}_t$!

1. train $\pi_\theta(\mathbf{u}_t|\mathbf{o}_t)$ from human data $\mathcal{D} = \{\mathbf{o}_1, \mathbf{u}_1, \ldots, \mathbf{o}_N, \mathbf{u}_N\}$
2. run $\pi_\theta(\mathbf{u}_t|\mathbf{o}_t)$ to get dataset $\mathcal{D}_\pi = \{\mathbf{o}_1, \ldots, \mathbf{o}_M\}$
3. Ask human to label $\mathcal{D}_\pi$ with actions $\mathbf{u}_t$
4. Aggregate: $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}_\pi$

DAgger: Watch!

Can we make it work more often?

DAgger: Dataset Aggregation

goal: collect training data from $p_{\pi_\theta}(\mathbf{o}_t)$ instead of $p_{\text{data}}(\mathbf{o}_t)$

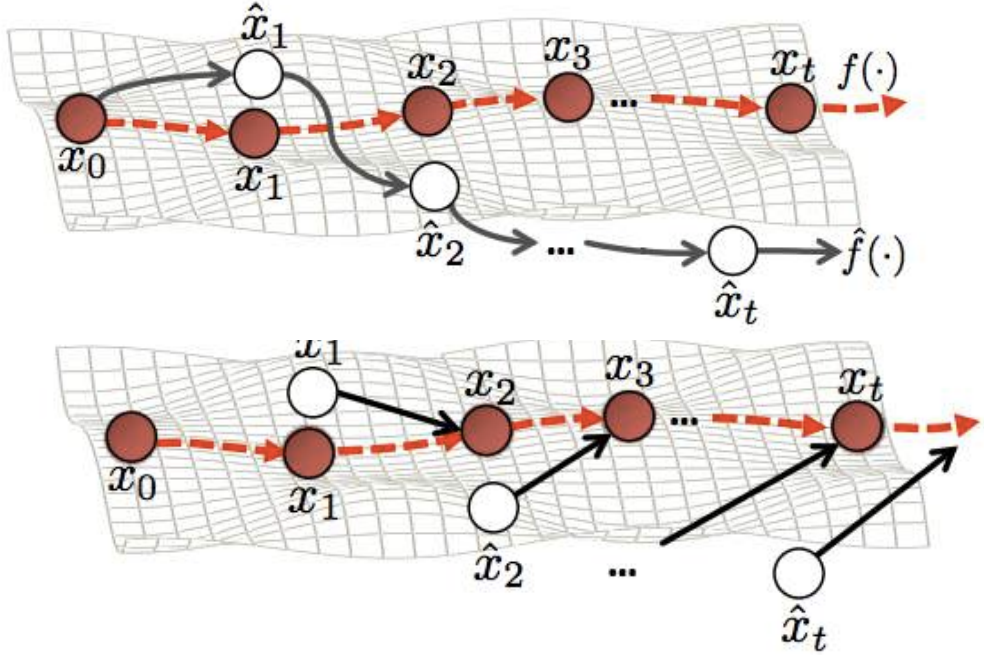
how? just run $\pi_\theta(\mathbf{u}_t|\mathbf{o}_t)$

but need labels $\mathbf{u}_t$!

1. train $\pi_\theta(\mathbf{u}_t|\mathbf{o}_t)$ from human data $\mathcal{D} = \{\mathbf{o}_1, \mathbf{u}_1, \ldots, \mathbf{o}_N, \mathbf{u}_N\}$
2. run $\pi_\theta(\mathbf{u}_t|\mathbf{o}_t)$ to get dataset $\mathcal{D}_\pi = \{\mathbf{o}_1, \ldots, \mathbf{o}_M\}$
3. Ask human to label $\mathcal{D}_\pi$ with actions $\mathbf{u}_t$
4. Aggregate: $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}_\pi$

AFP

Data as Demonstrator



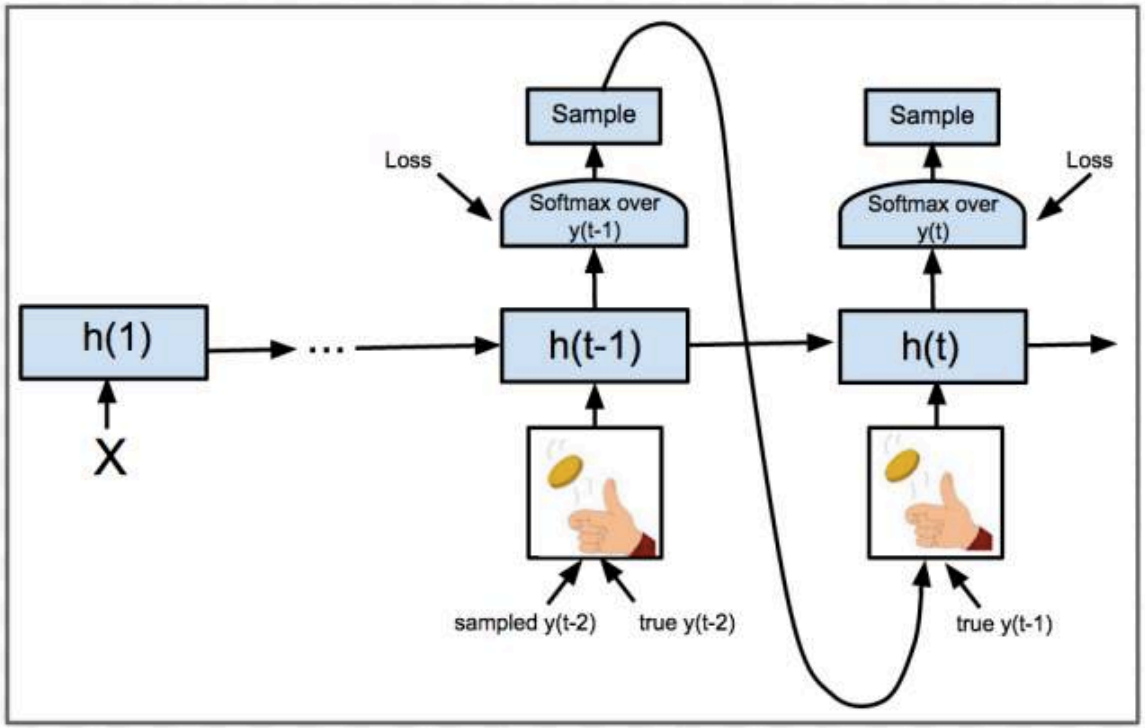
Today's talk

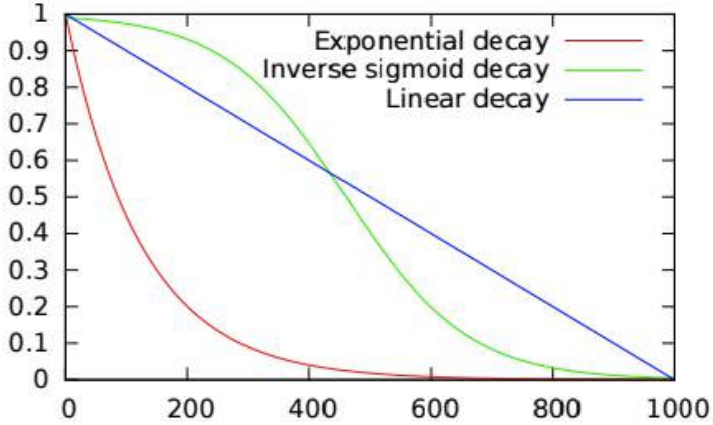
- ~~Definition of sequential problem~~
- ~~Imitation learning: supervised learning for decision making~~
  - ~~Does imitation learning works?~~
  - ~~How can we make it work more often?~~
- Case studies of recent work in (deep) imitation learning

Case Study 1: Trail following as classification

- Watch!

Case Study 2: Scheduled Sampling





Imitation Learning Summary

- Some times works well
  - Hacks
  - Intelligent way of getting training samples
  - Really diverse dataset that summarizes the data distribution
- Problems
  - Humans needs to provide data, which is usually finite
  - Humans are not good at provided some kind of actions

That's all Folks!