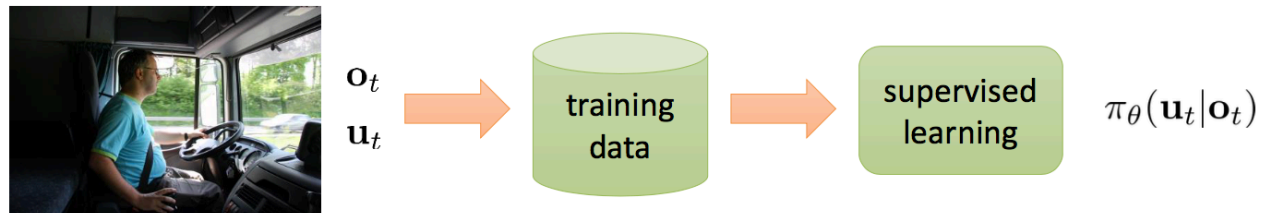
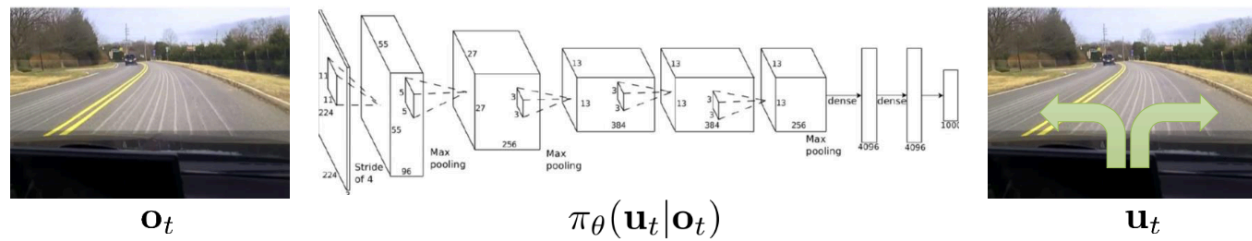


# Imitation Learning

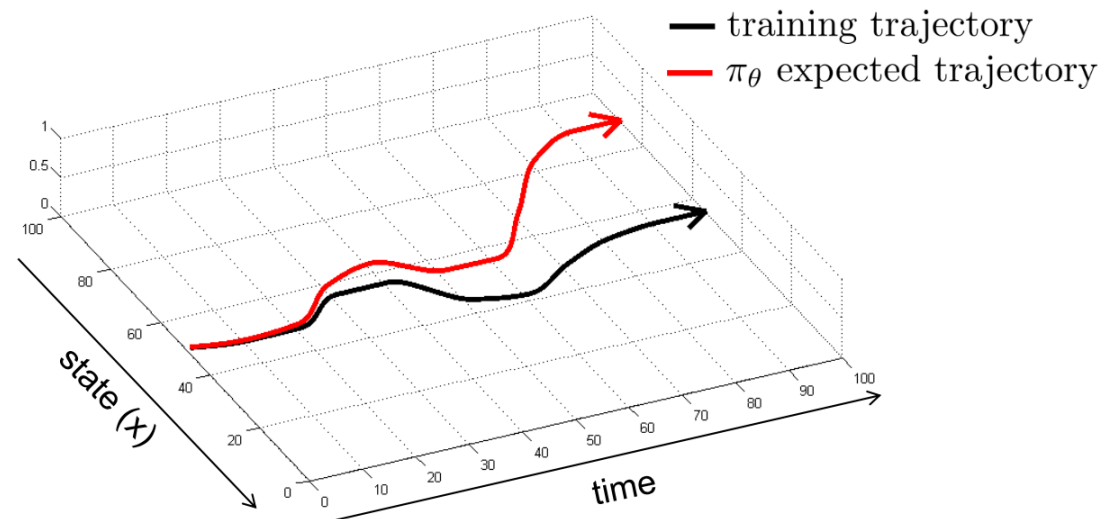


Images: Bojarski et al. '16, NVIDIA

[slides from Levine]

Does it work?

No!



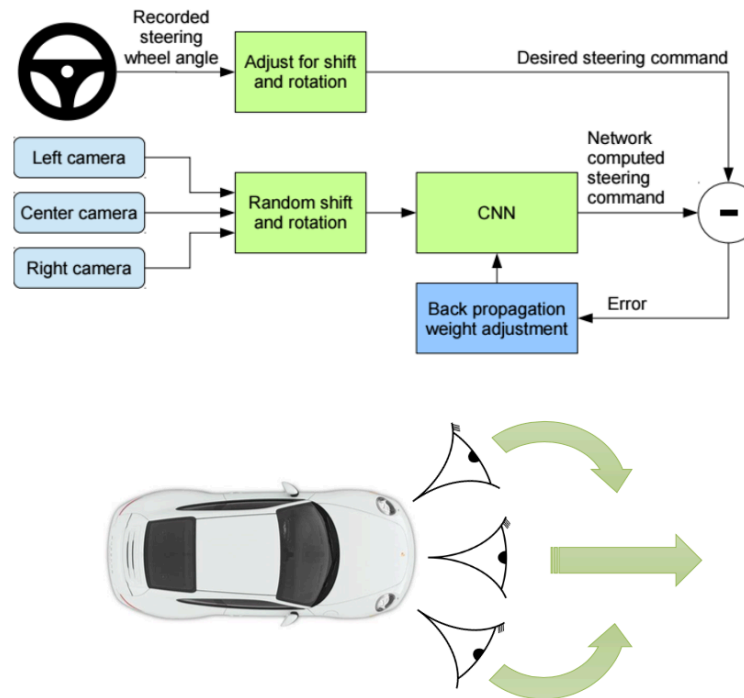
Does it work?

Yes!



Video: Bojarski et al. '16, NVIDIA

# Why did that work?



# Can we make it work more often?

can we make  $p_{\text{data}}(\mathbf{o}_t) = p_{\pi_\theta}(\mathbf{o}_t)$ ?


idea: instead of being clever about  $p_{\pi_\theta}(\mathbf{o}_t)$ , be clever about  $p_{\text{data}}(\mathbf{o}_t)$ !

## **D**Agger: Dataset Aggregation

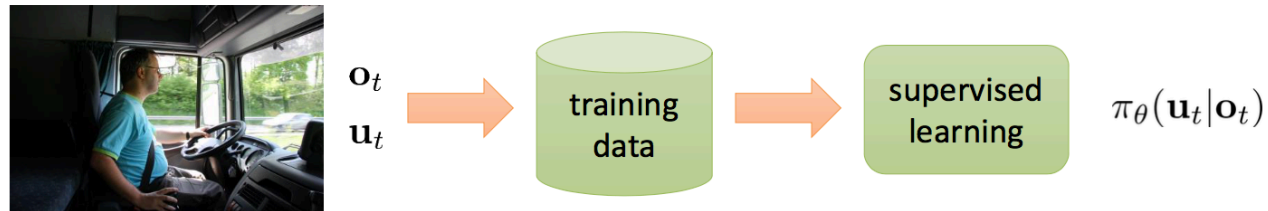
goal: collect training data from  $p_{\pi_\theta}(\mathbf{o}_t)$  instead of  $p_{\text{data}}(\mathbf{o}_t)$

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

# Imitation learning: recap



- Often (but not always) insufficient by itself
  - Distribution mismatch problem
- Sometimes works well
  - Hacks (e.g. left/right images)
  - Samples from a stable trajectory distribution
  - Add more **on-policy** data, e.g. using DAgger

