

行为克隆

2021. 8. 3 曹俊

模仿学习

- 模仿学习介绍
- 行为克隆介绍
- 存在问题

模仿学习

模仿是人类学习的关键一环，即使没有任何奖励。

强化学习一直围绕着奖励Reward展开，但是存在以下两个问题：

- 在实际生活中Reward总是很稀疏，甚至没有Reward；
- 手工设计的Reward可能会导致无法控制的行为

在强化学习中，模仿学习的核心思想是：

专家指导agent如何解决问题。

总共分类两个方法：

- 行为克隆 Behavior Cloning
- 反向强化学习 Inverse Reinforcement Learning

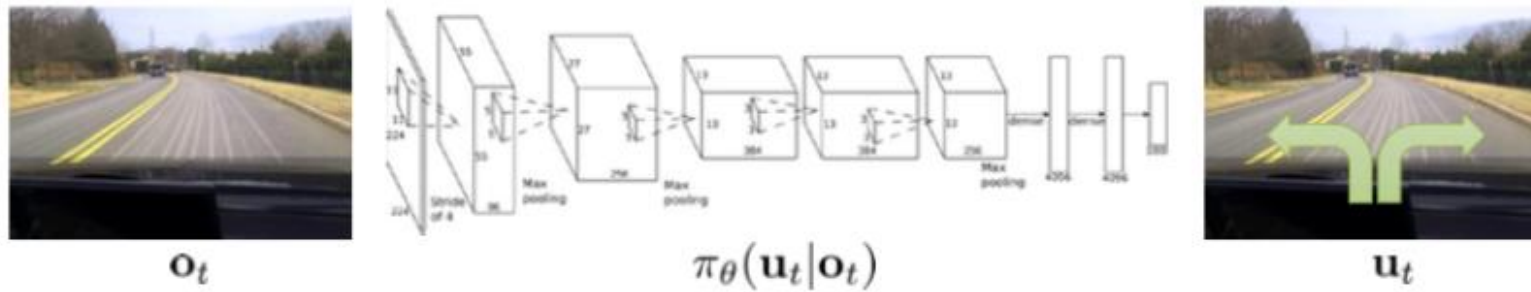
行为克隆

人类专家的决策数据 $\{\tau_1, \tau_2, \dots, \tau_m\}$

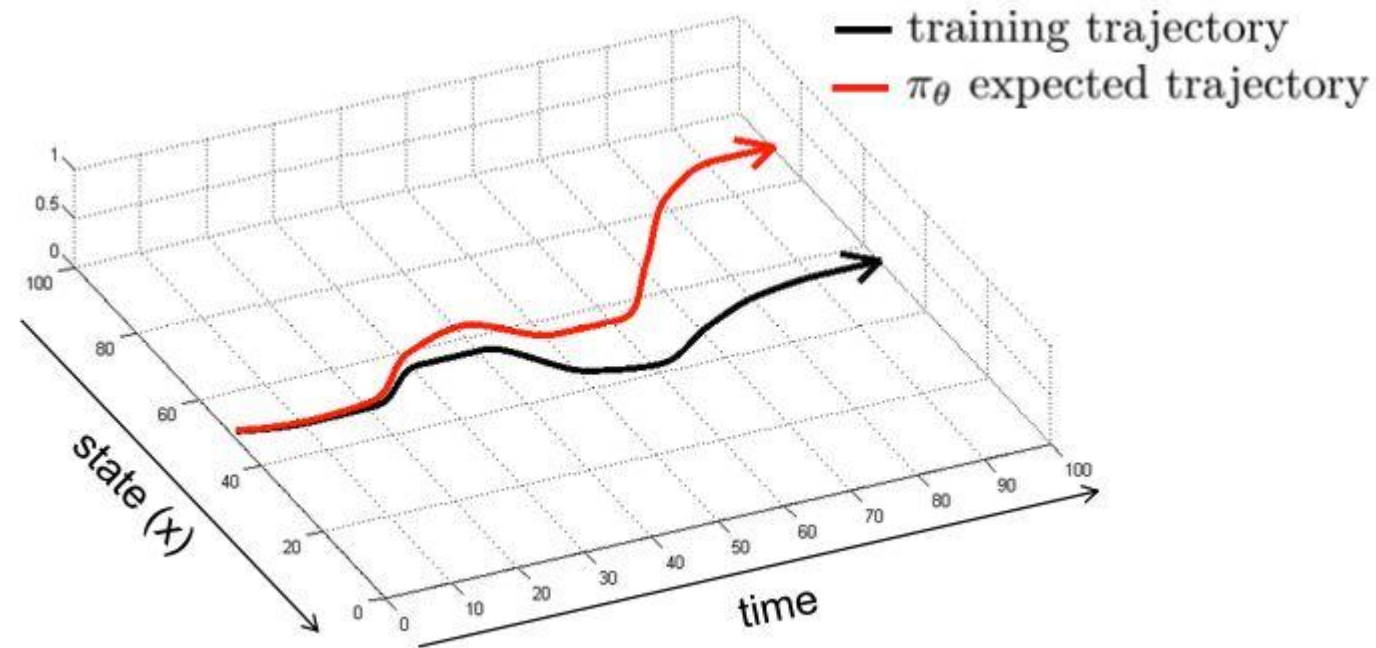
$$\tau_i = \langle s_1^i, a_1^i, s_2^i, a_2^i, \dots, s_{n_{\tau_i}+1}^i \rangle$$

状态-动作对 $\mathcal{D} = \{(s_1, a_1), (s_2, a_2), (s_3, a_3), \dots\}$

行为克隆

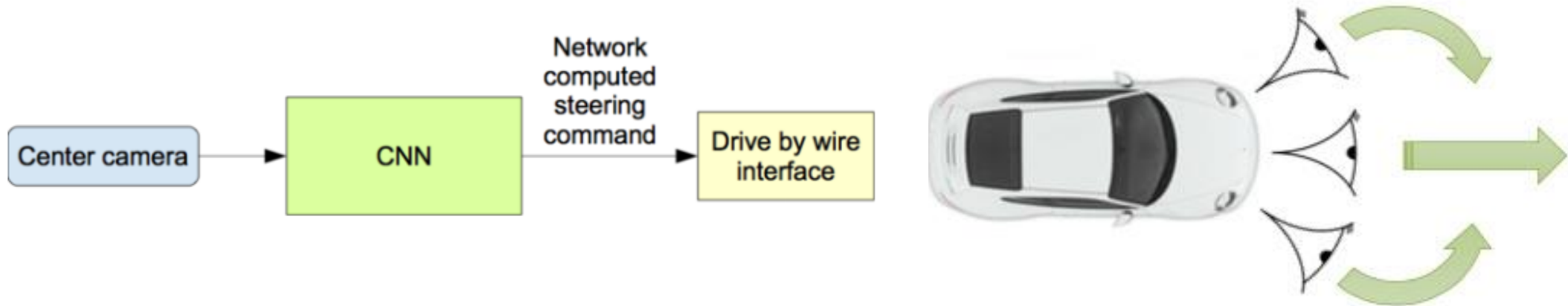
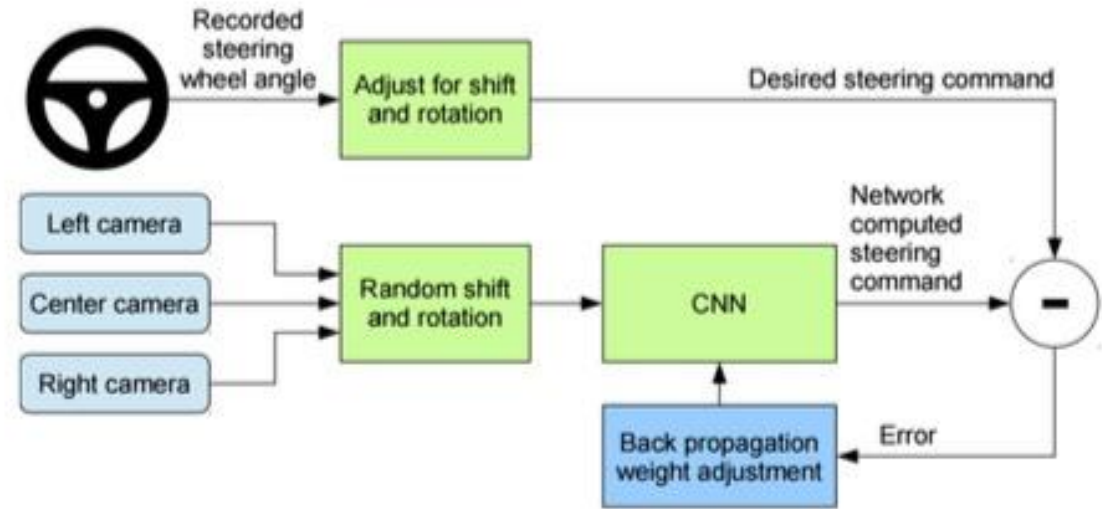


泛化问题

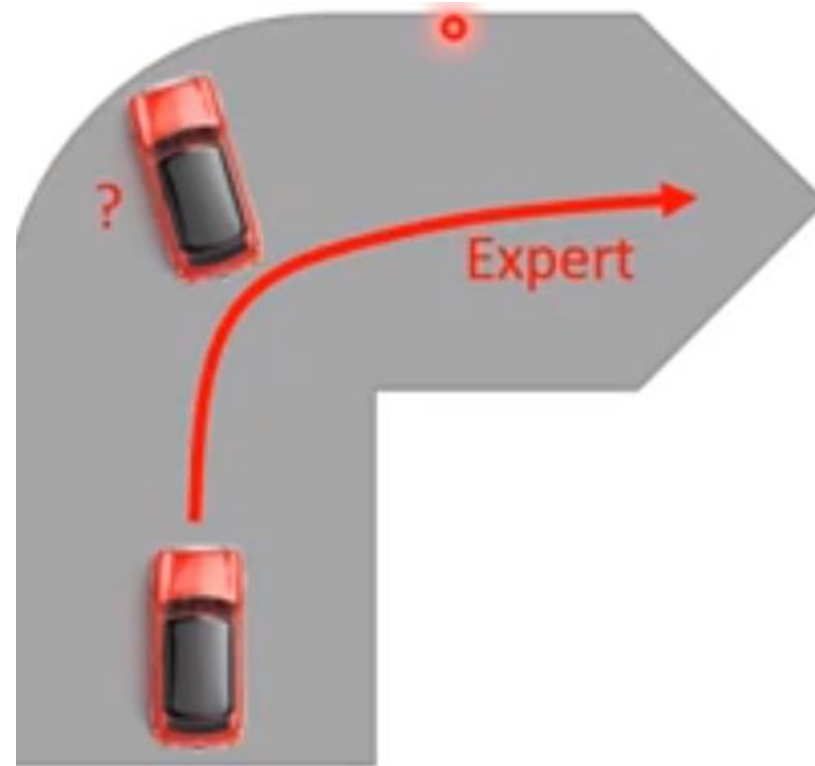


数据增强

End to End Learning for Self-Driving Cars



Dataset Aggregation



Dataset Aggregation

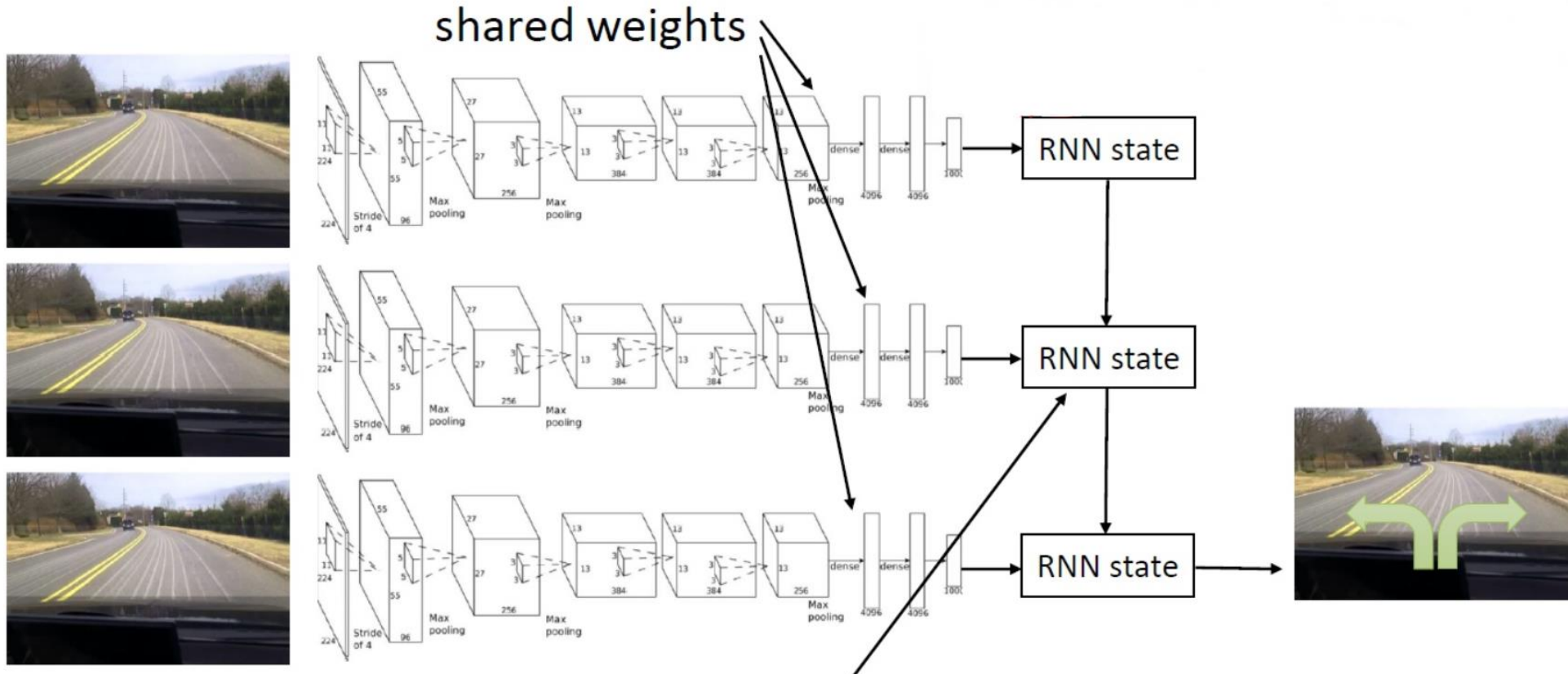
1. 通过数据集 $\mathcal{D} = \{(o_1, a_1), (o_2, a_2), (o_3, a_3), \dots\}$ 训练出策略 $\pi_\theta(u_t|o_t)$
2. 执行 $\pi_\theta(u_t|o_t)$ 得到一个新的数据集 $\mathcal{D}_\pi = \{o_1, o_2, o_3, \dots\}$
3. 人工给 \mathcal{D}_π 中的状态标上动作 (action) u_t
4. 聚合 (Aggregate) : $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{D}_\pi$
5. 跳到步骤 1

模仿学习存在问题

- Non-Markovian behavior
- Multimodal behavior
- Limited Data

Non-Markovian behavior

How can we use the whole history?



Non-Markovian behavior

Causal Confusion in Imitation Learning

NIPS 2019

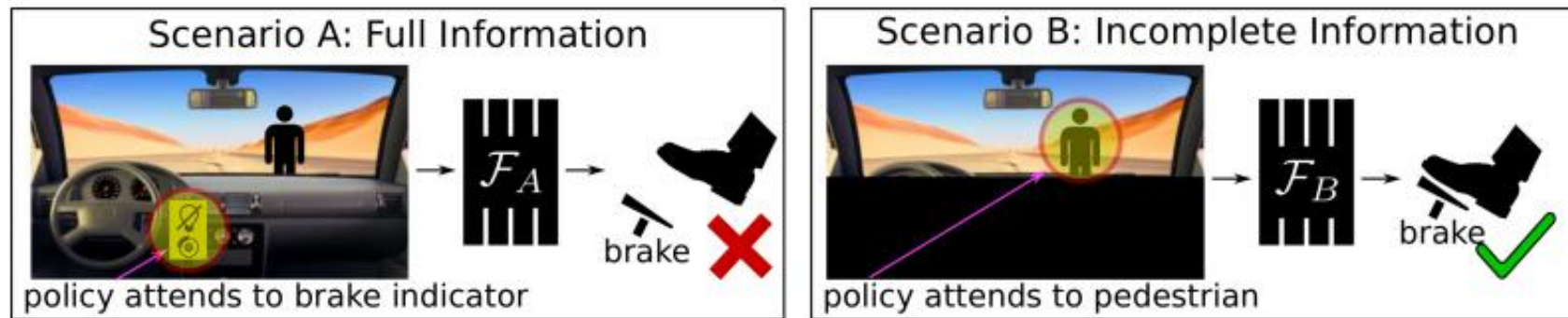


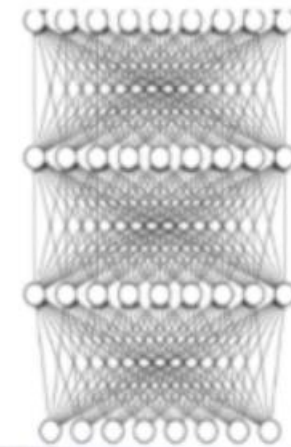
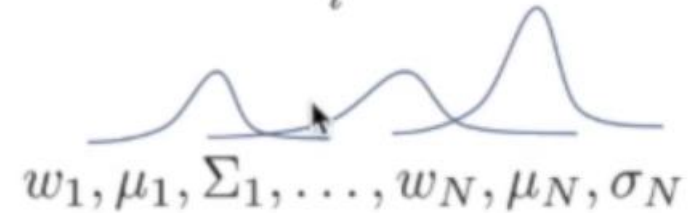
Figure 1: Causal misidentification: *more* information yields worse imitation learning performance. Model A relies on the braking indicator to decide whether to brake. Model B instead correctly attends to the pedestrian.

Multimodal behavior

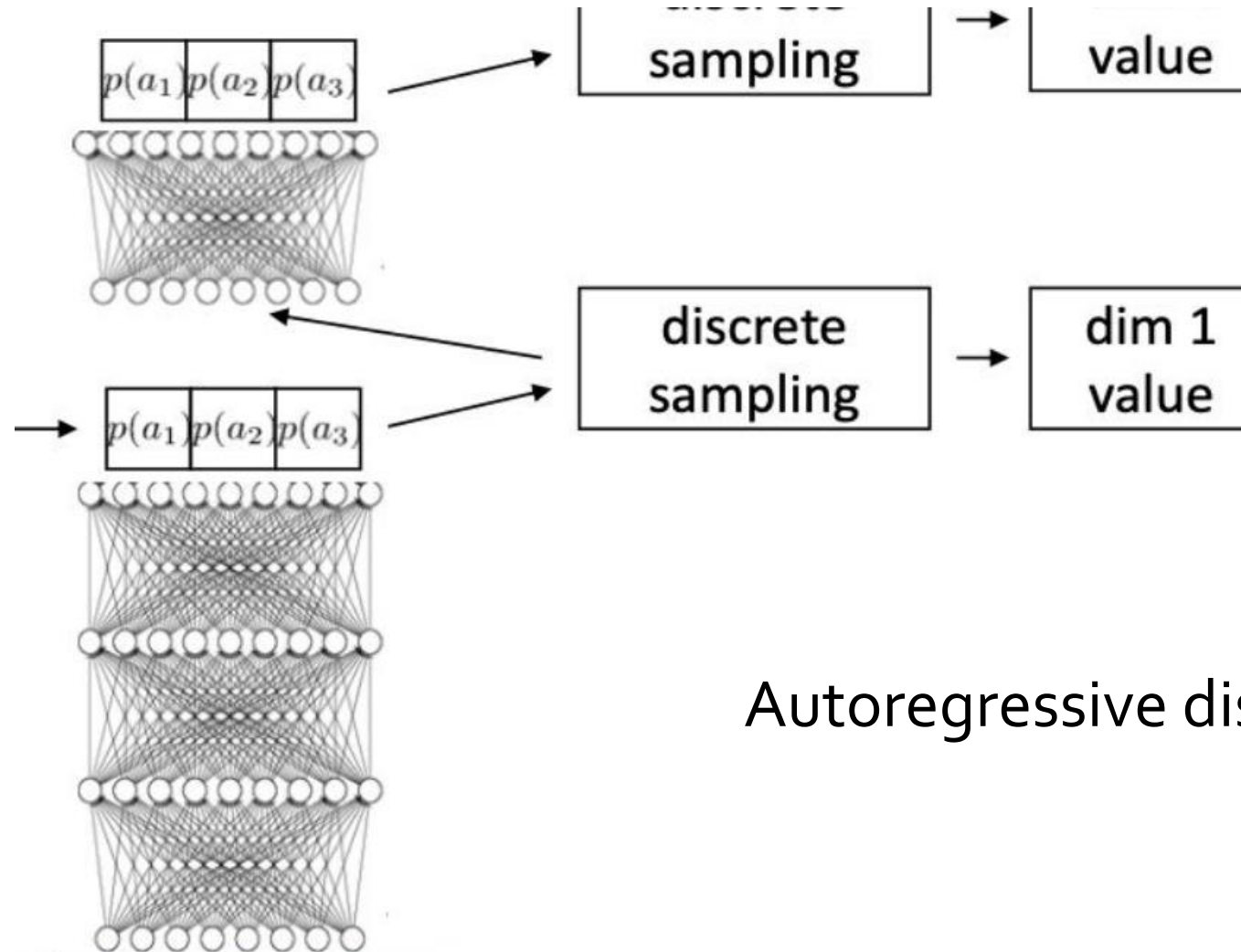
- Handling multimodal behavior
 - Output mixture of Gaussians
 - Latent variable models
 - Autoregressive discretization



$$\pi(\mathbf{a}|\mathbf{o}) = \sum_i w_i \mathcal{N}(\mu_i, \Sigma_i)$$



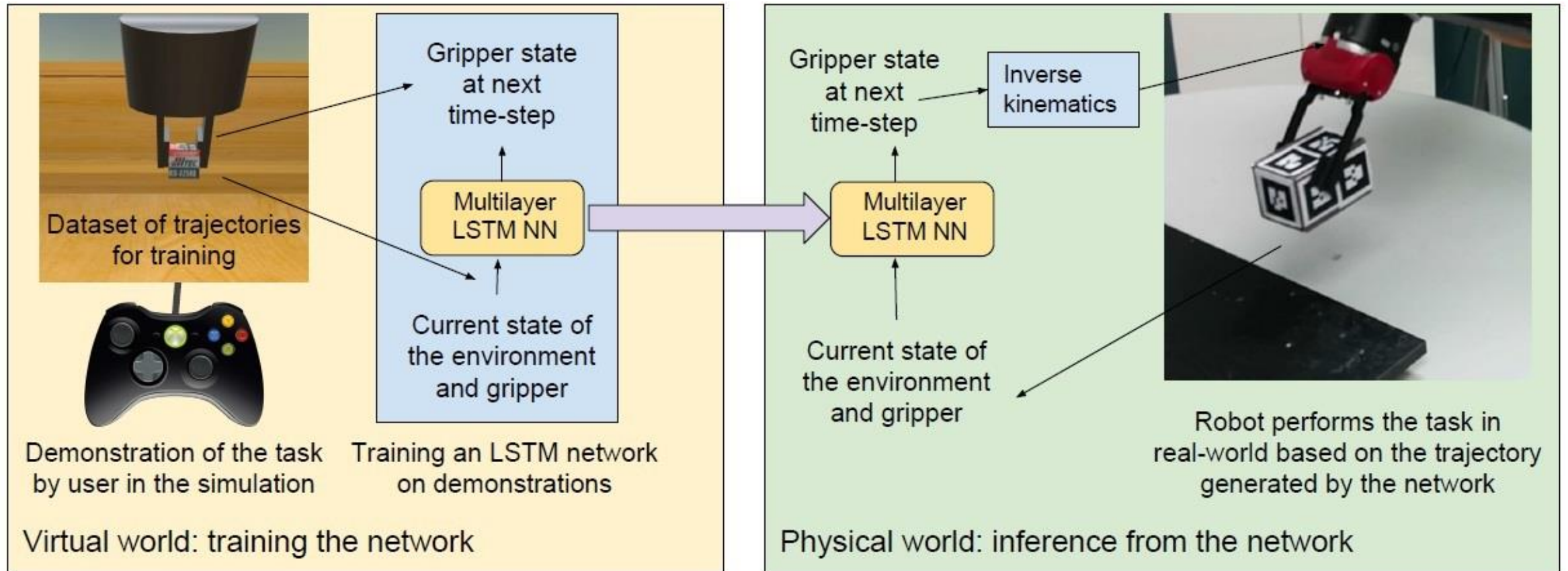
Multimodal behavior



Autoregressive discretization

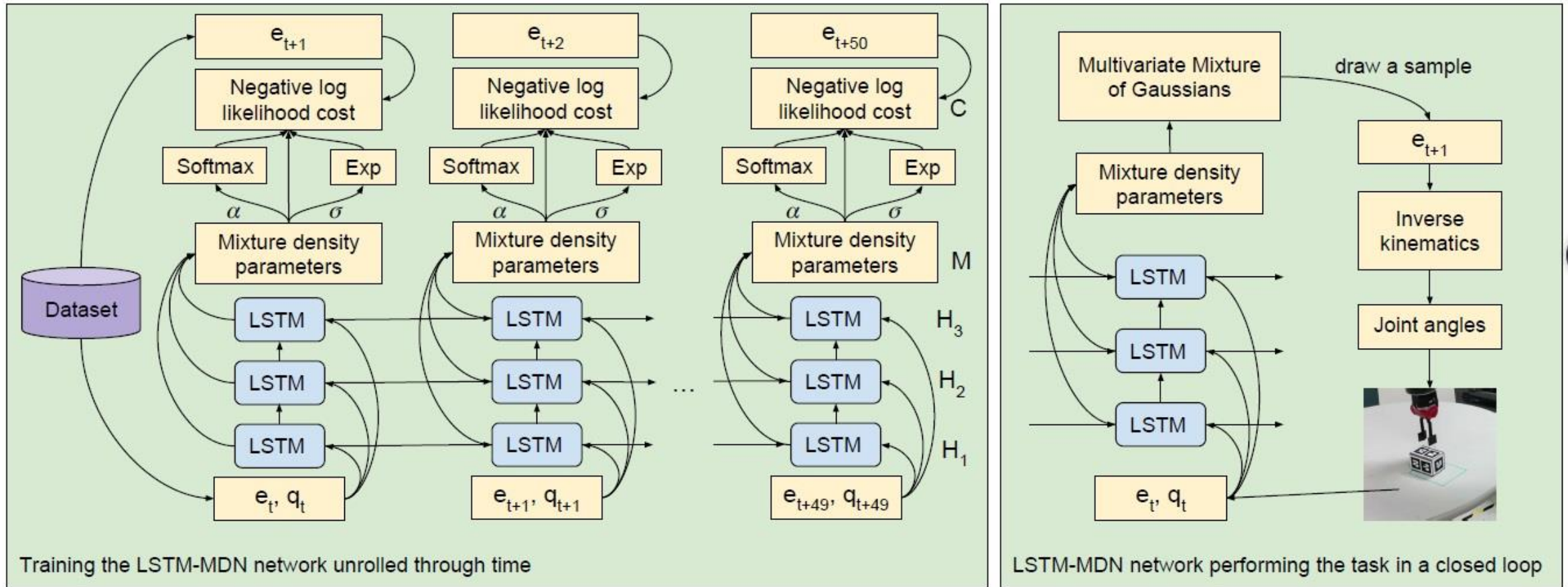
Multimodal behavior

learning real manipulation tasks from virtual demonstrations using LSTM



Multimodal behavior

learning real manipulation tasks from virtual demonstrations using LSTM



Limited Data



Thanks!