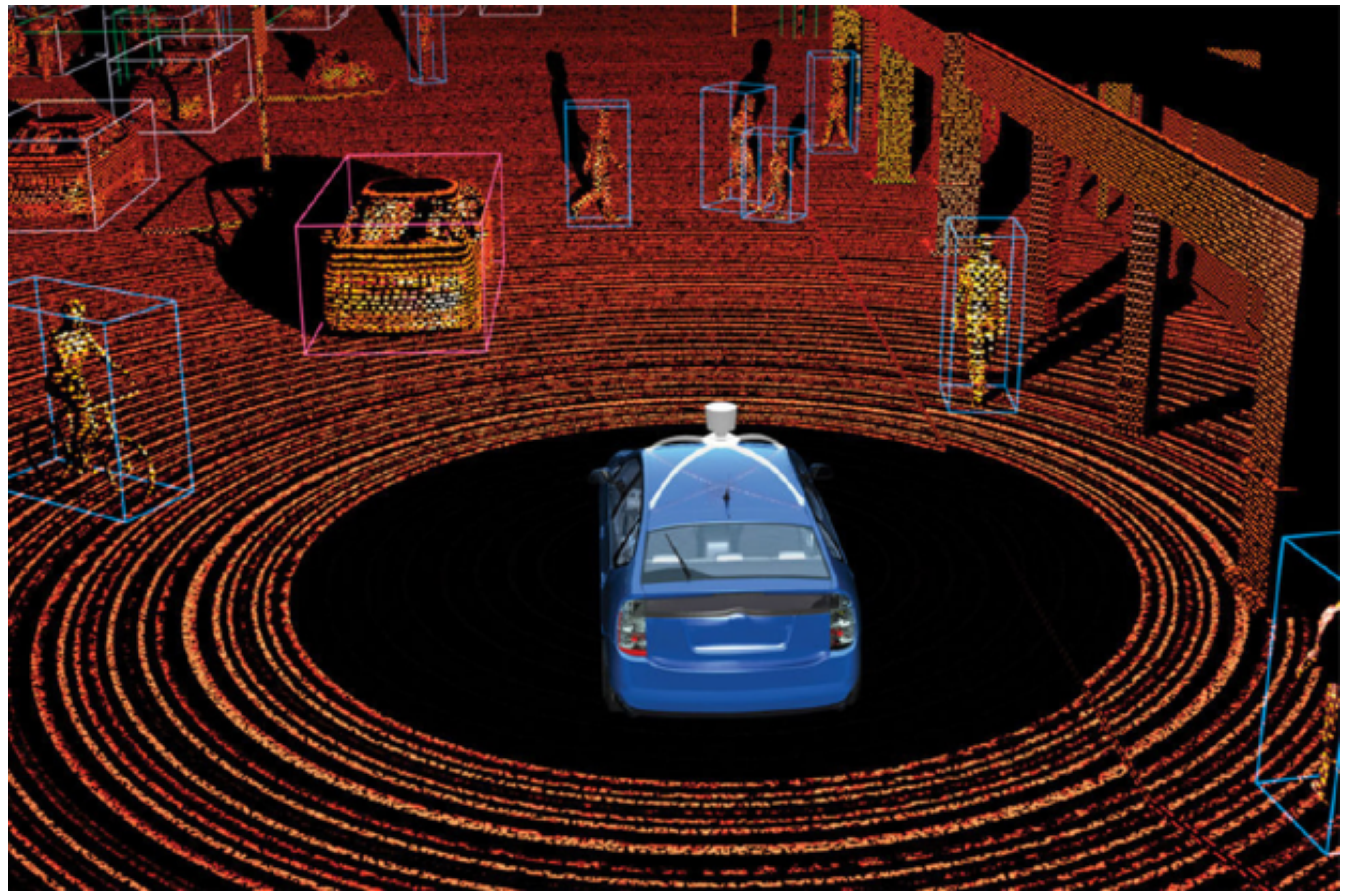
Reinforcement learning: An example

- Suppose you want to learn how to drive a car such that it follows a lane and minimises the number of collisions

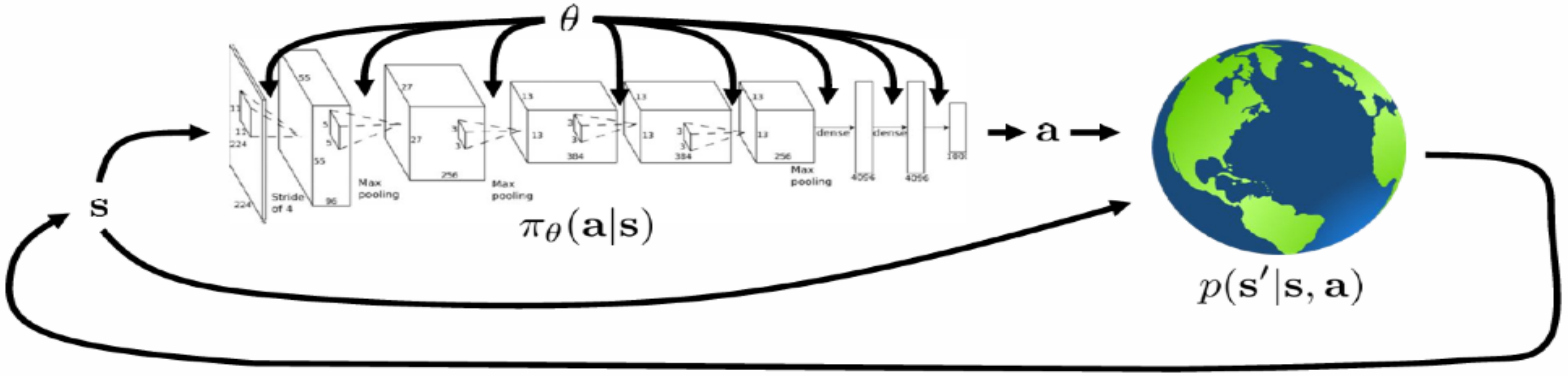


- You observe the state s of the surroundings: which can consist of very high dimensional data like LIDAR, images, etc.

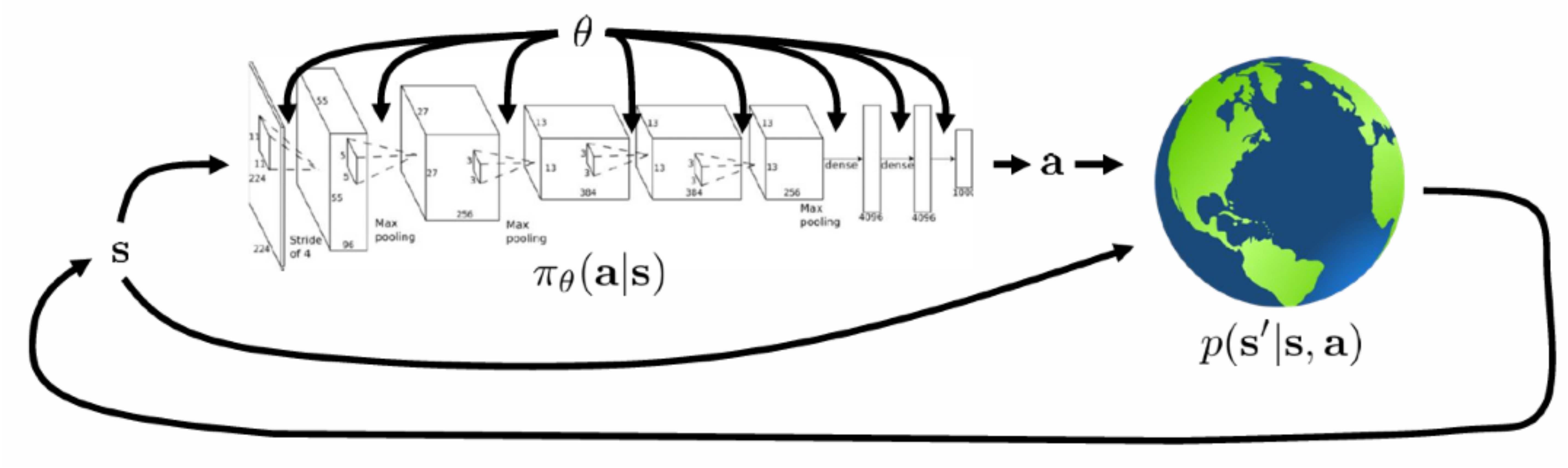
- You take an action a = [steering angle, acceleration]

- Observe the next state $s' \sim \mathbb{P}(\,\cdot\,|\,s, a)$ and reward $= \mathbb{I}\{\text{on the lane}\} - \#\text{no of collisions}$

- Objective: Design a policy, which is a mapping from state to action e.g. a neural network, such that the reward over T rounds is maximized



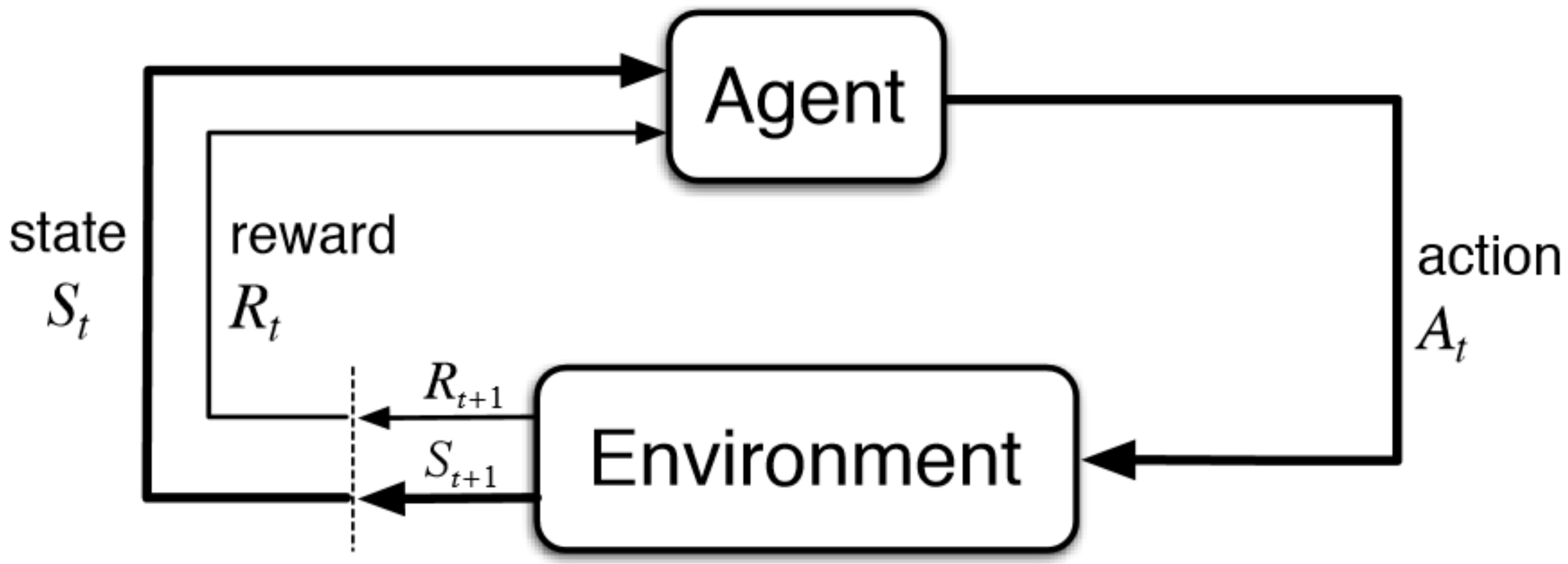
Reinforcement learning


θ
s
a
$\pi_\theta(\mathbf{a}|\mathbf{s})$
$p(\mathbf{s}'|\mathbf{s},\mathbf{a})$
Stride of 4
Max pooling
Max pooling
Max pooling
dense
dense

- Observe initial state s_0
- For every round $t = 1,\ldots,T$
  - Take action a_t
  - Observe next state $s_{t+1} \sim \mathbb{P}(\cdot \mid s_t, a_t)$ and reward r_t

Goal: Find policy π which is a mapping from state to action such that expected reward over T rounds is maximized

$$\sum_{t=1}^{T} \mathbb{E}_{a_t\sim\pi(\cdot|s_t),s_{t+1}\sim\mathbb{P}(\cdot|s_t,a_t)}[\gamma^{t-1} r_t(s_t, a_t) \mid s_0]$$


Agent
Environment
state
S_t
reward
R_t
action
A_t
R_{t+1}
S_{t+1}