AI不同，择与为谋：MathWorks助你跨学科深度创新

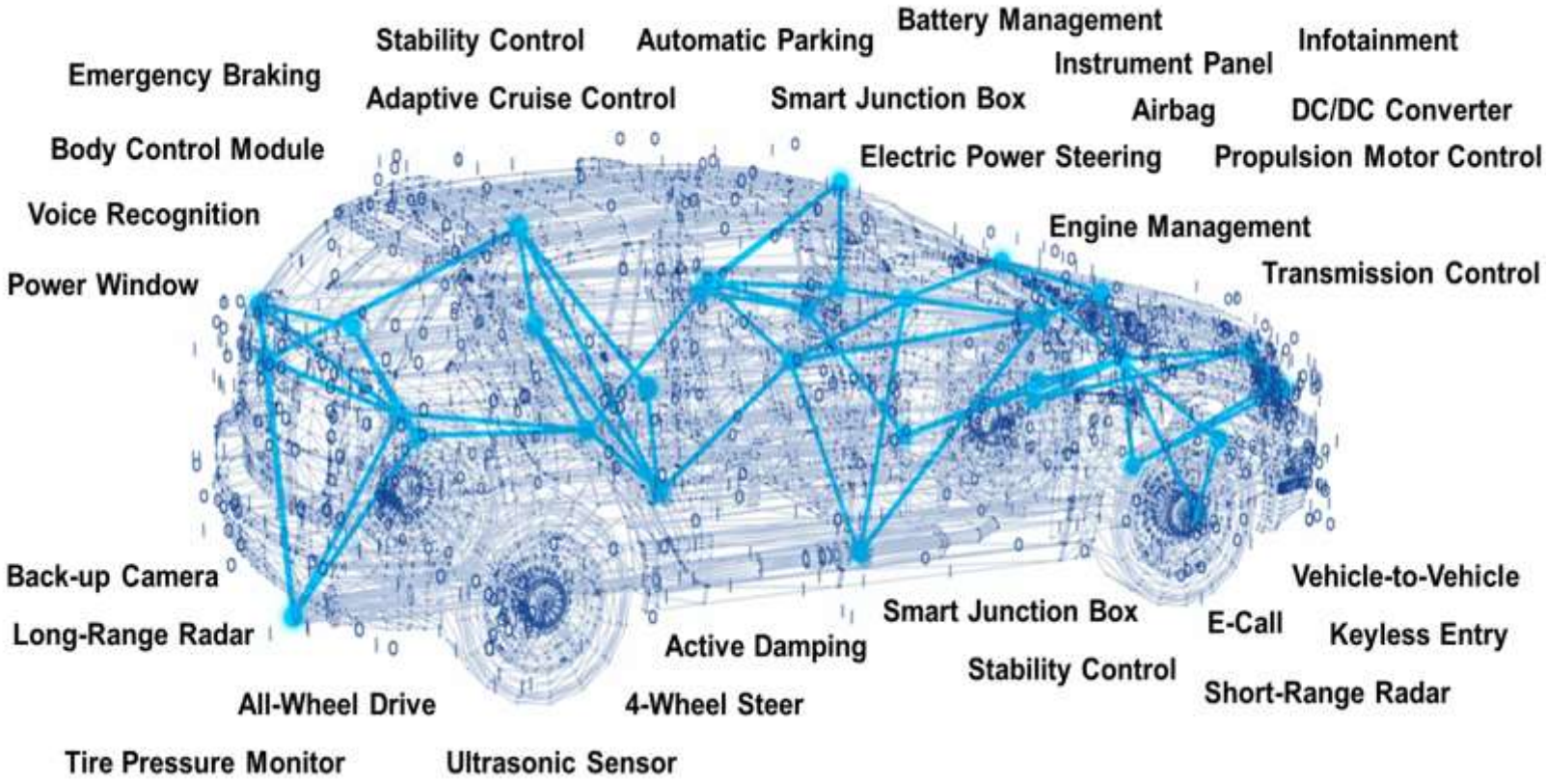
袁航，MathWorks应用工程师

内容概要

- 机械、电子、自动化与人工智能的跨学科应用场景
- 在 MATLAB 中实现基于强化学习的自主系统设计

数字变革

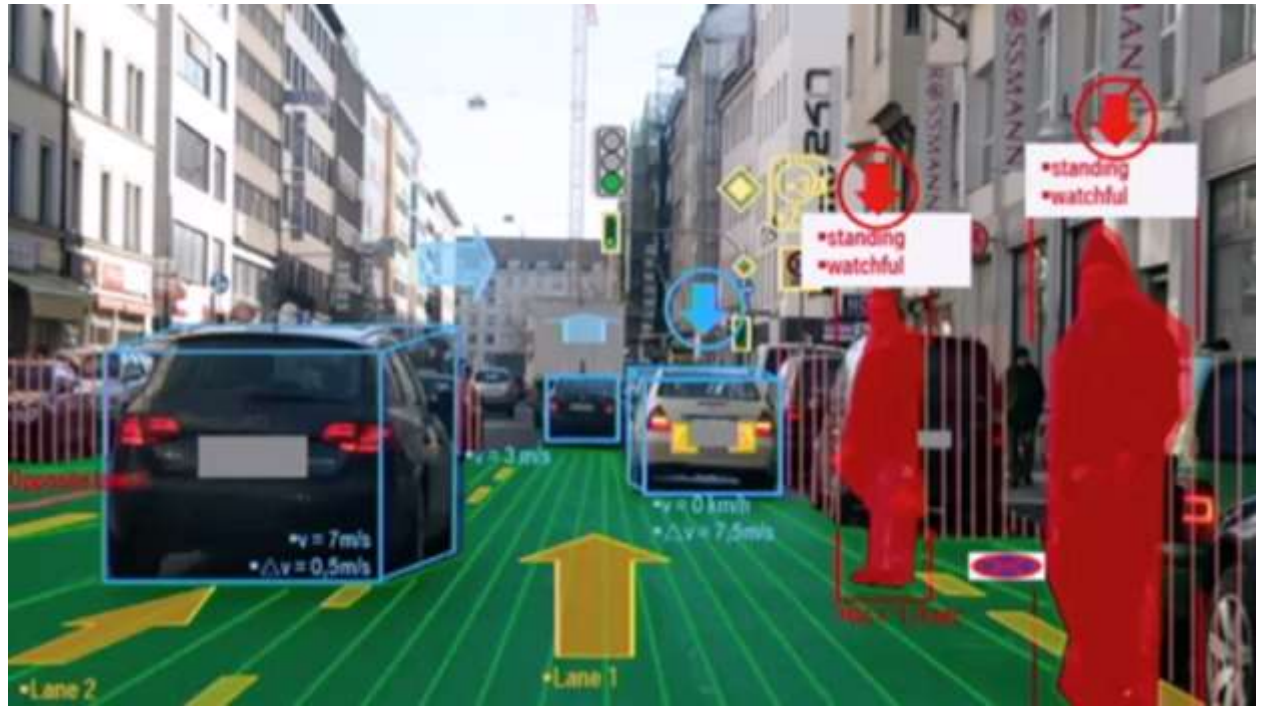
嵌入式软件的普及



数字变革

人工智能的渗透

数据驱动算法
机器学习和深度学习
自主系统

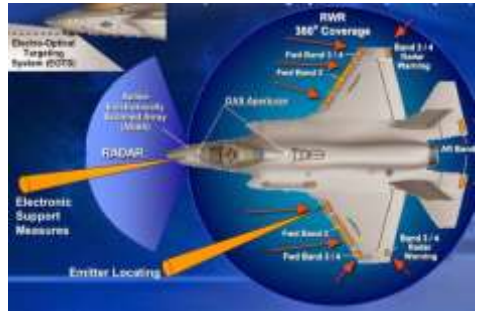


下一代无线通信

5G 和下一代WiFi

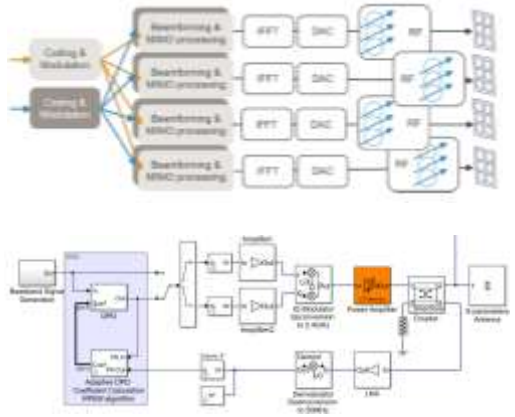
新标准、新技术
研发和测试

新一代雷达和卫星通信



多功能航空国防系统
(EW, Radar, Comms)

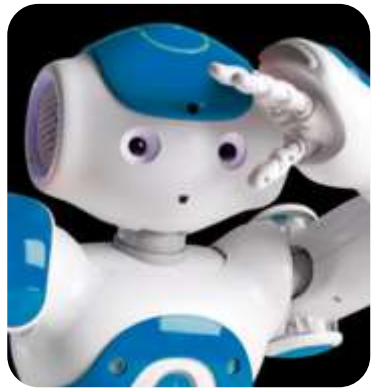
集成的射频、天线和DSP设计



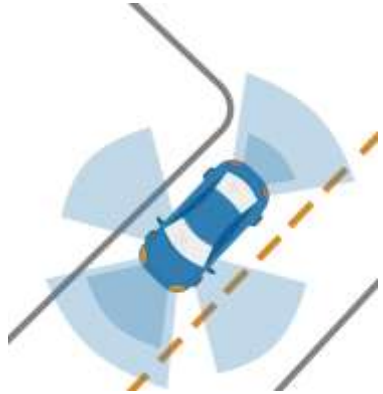
设备复杂度
多域设计

未来物联网、车联网

新兴应用：
V2X, 大规模物联网

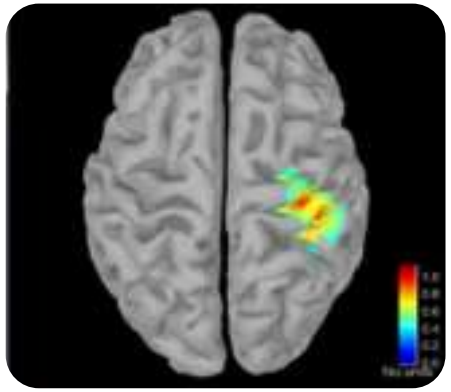


机器人



自动驾驶

机器学习



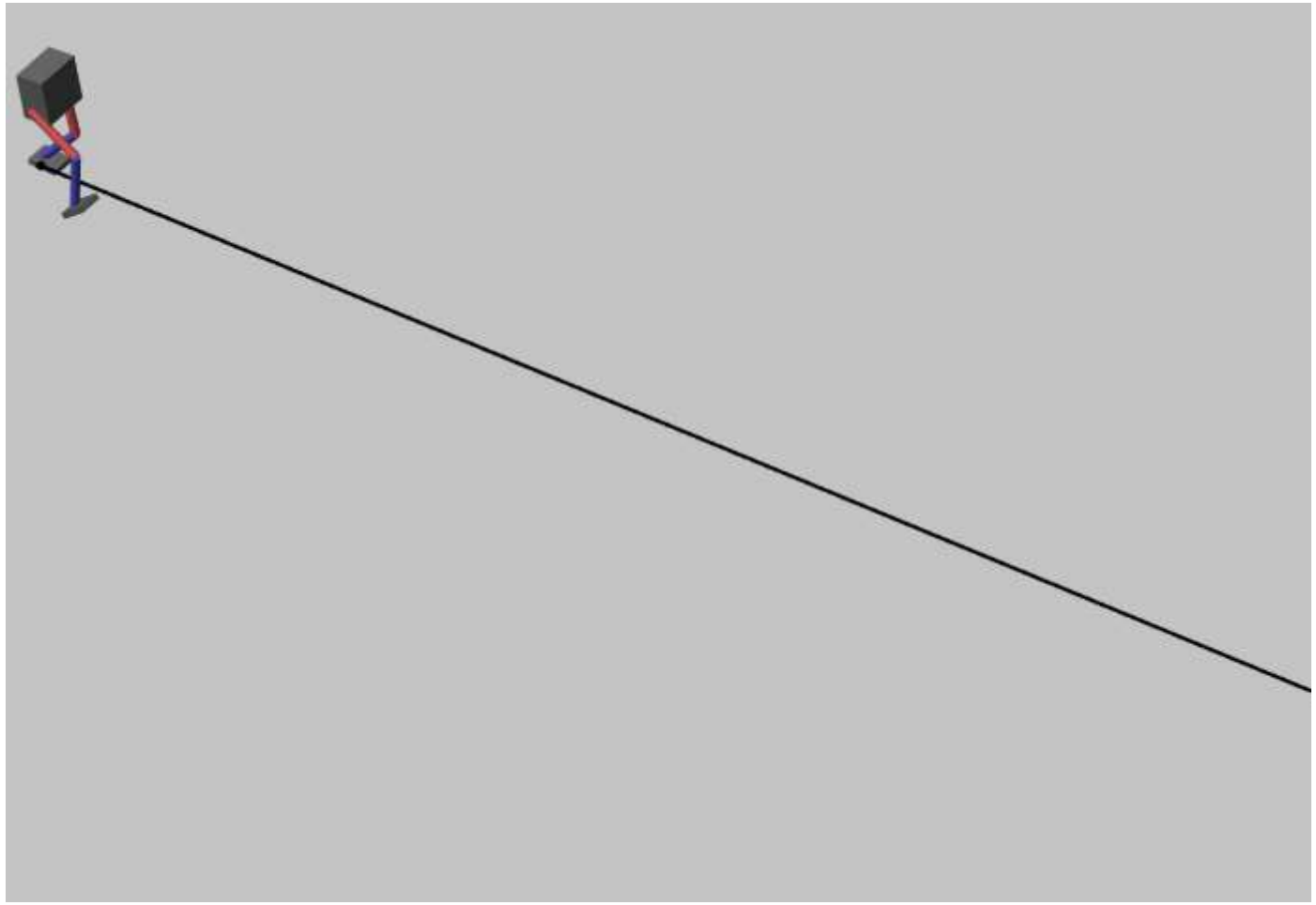
医药创新

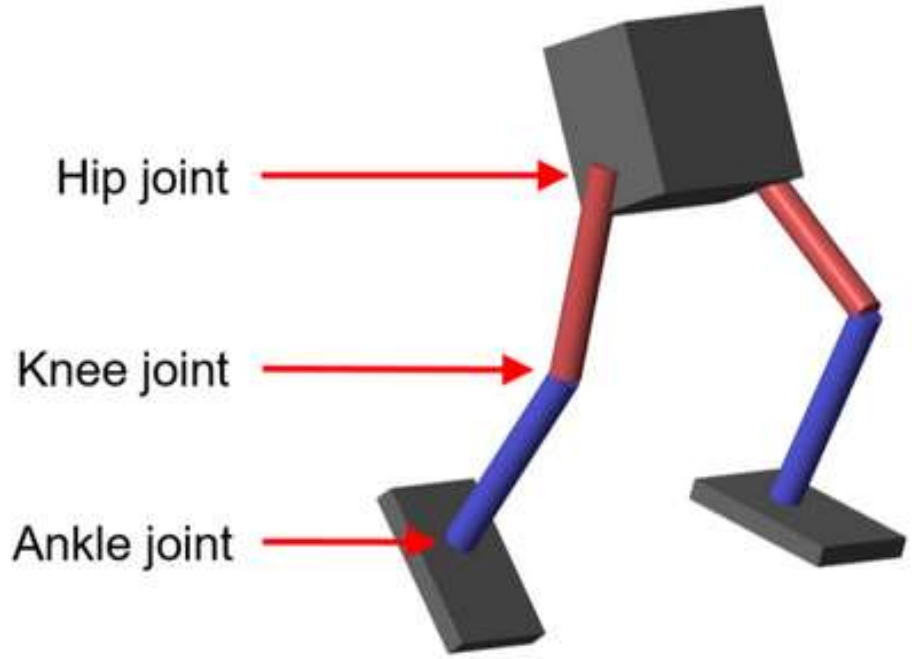
物联网

如何培养未来的科学家和工程师？

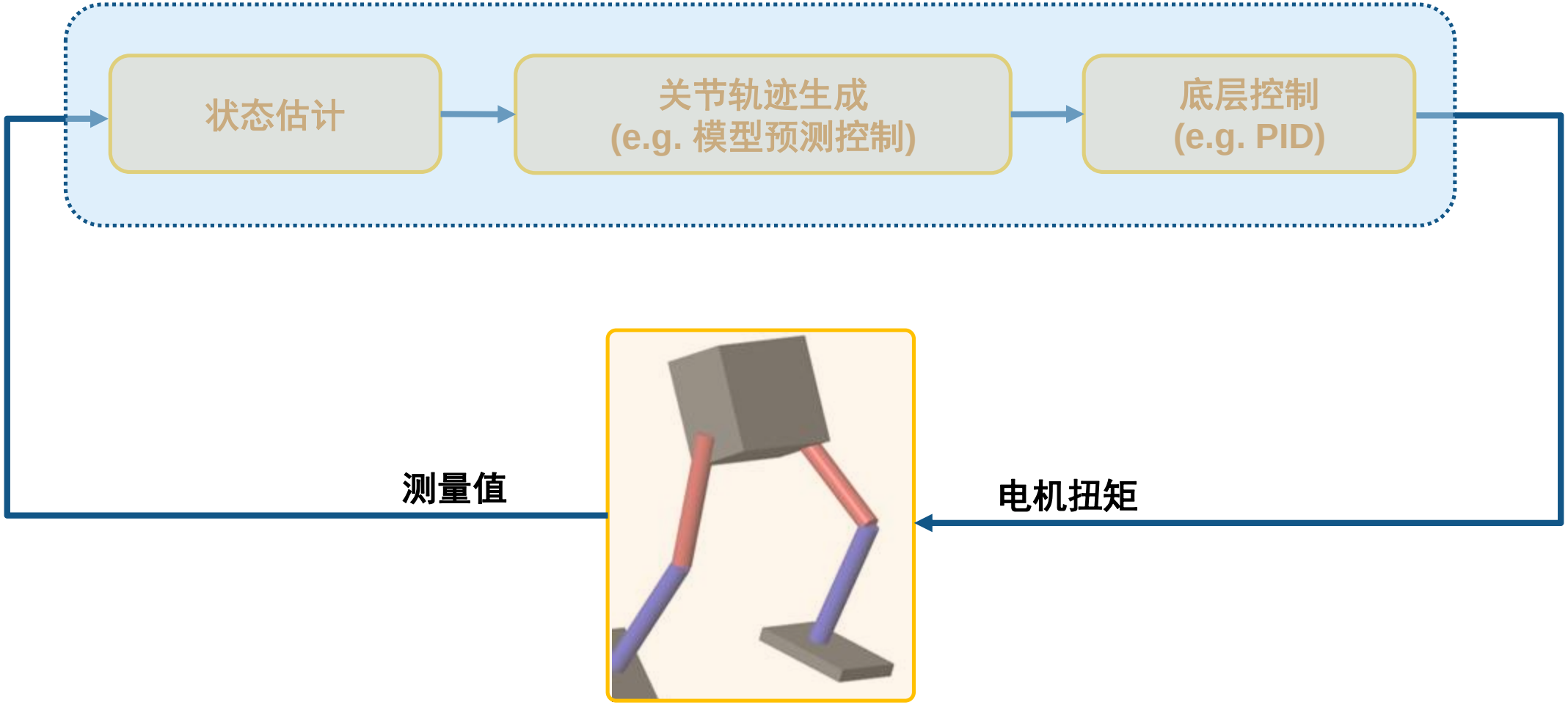
机械、电子、自动化与人工智能的跨学科应用场景

人工智能的跨学科应用场景 —— 双足机器人行走控制

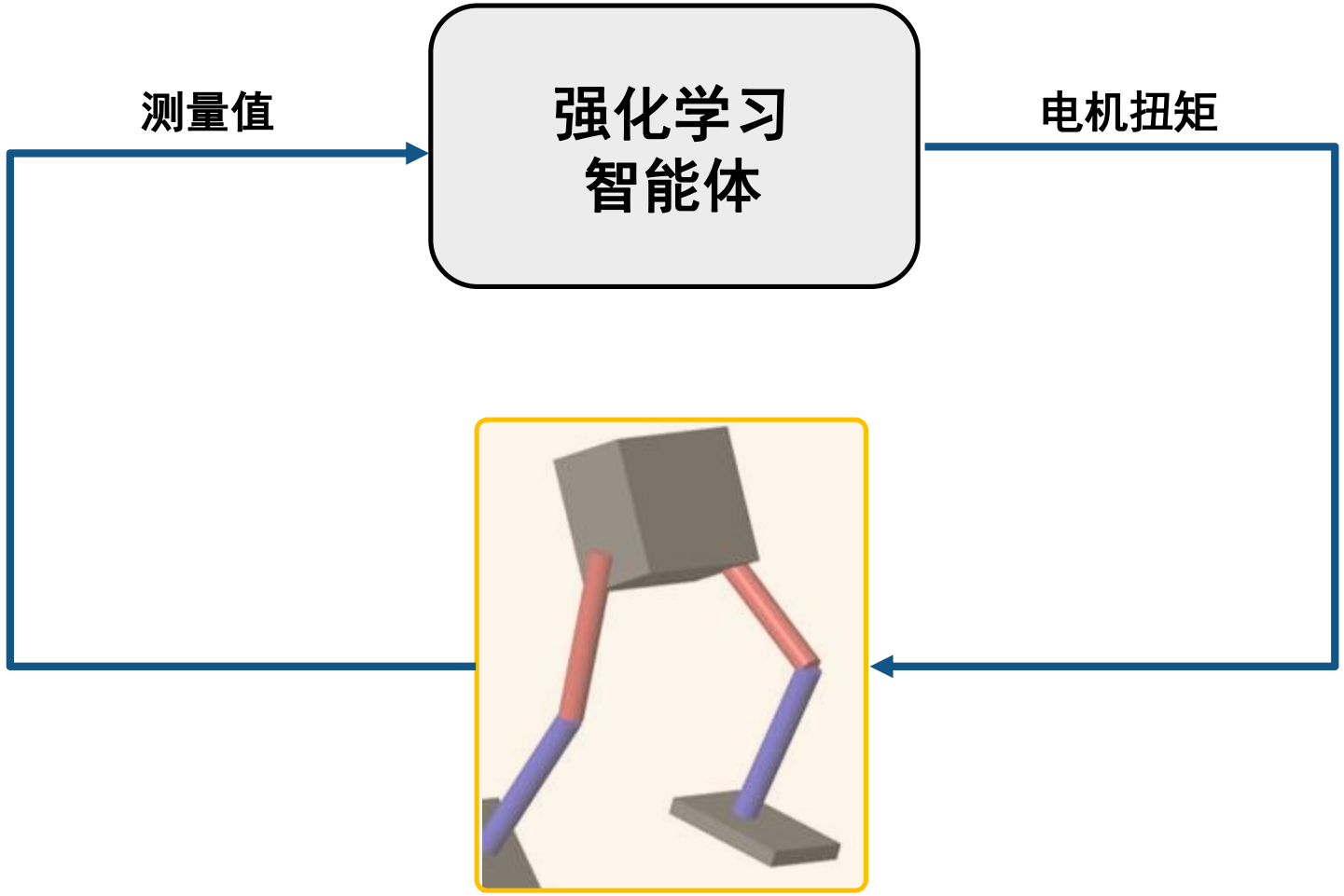




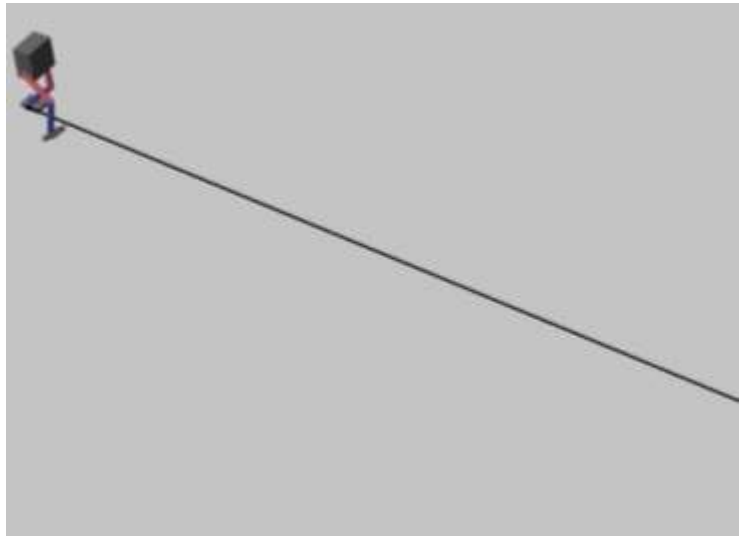
传统学科研究方向

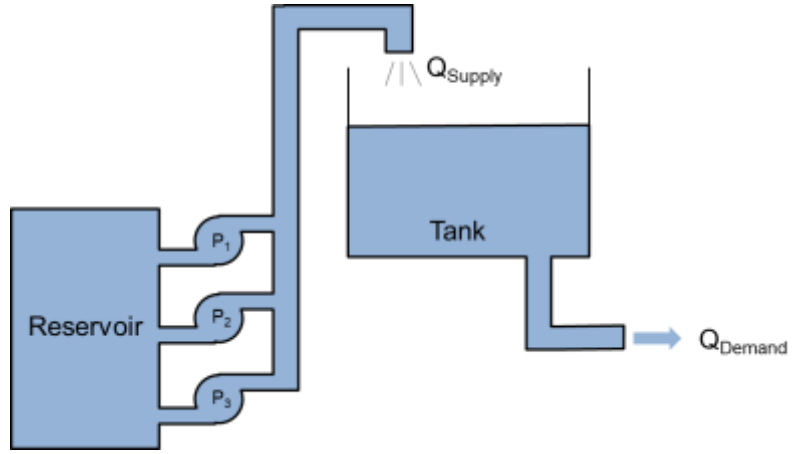


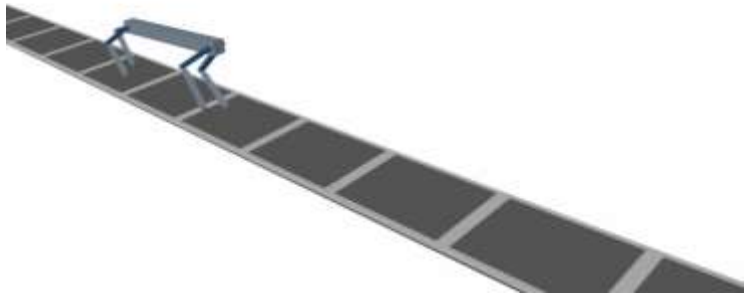
基于强化学习的双足机器人行走控制

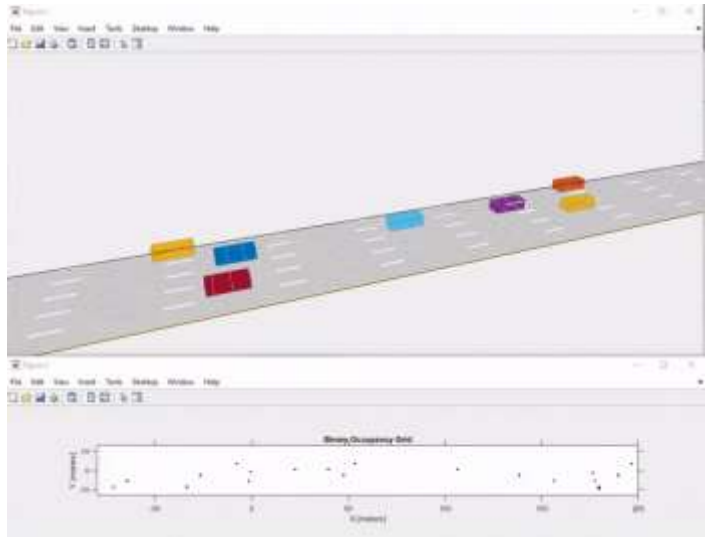


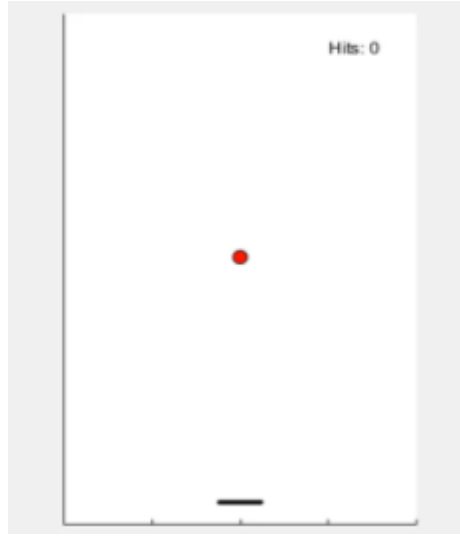
强化学习应用领域



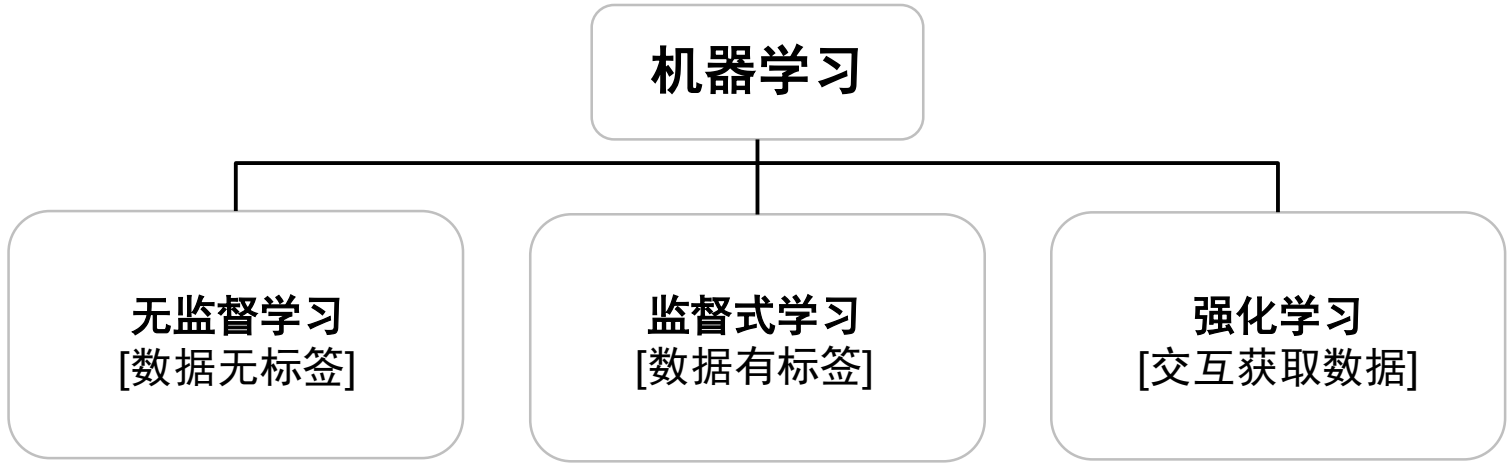


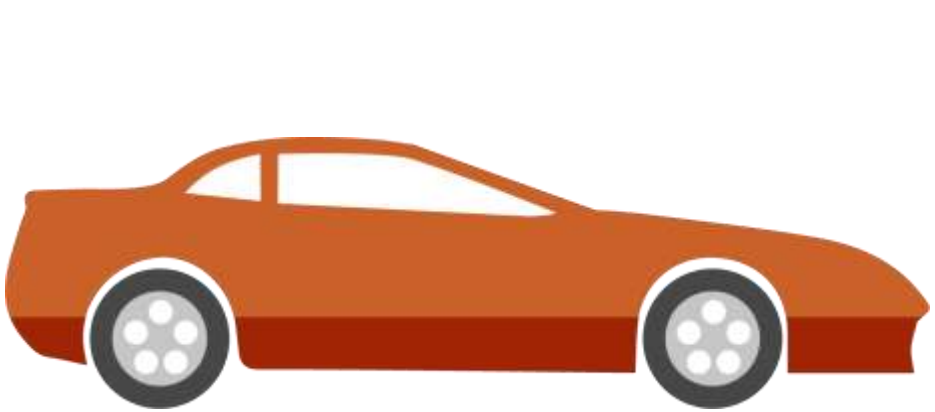




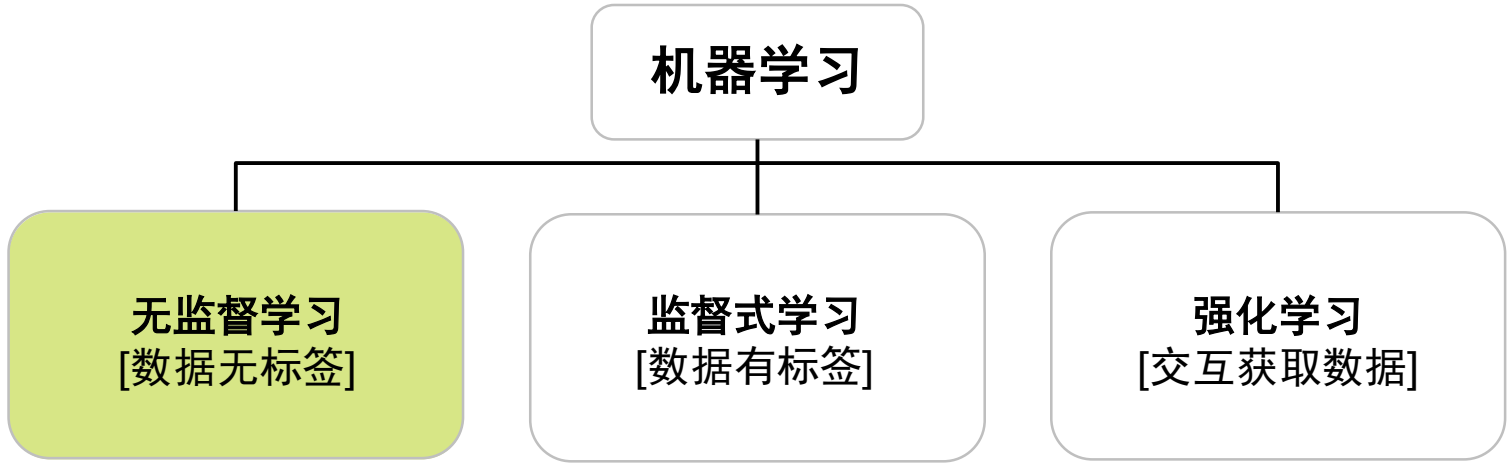


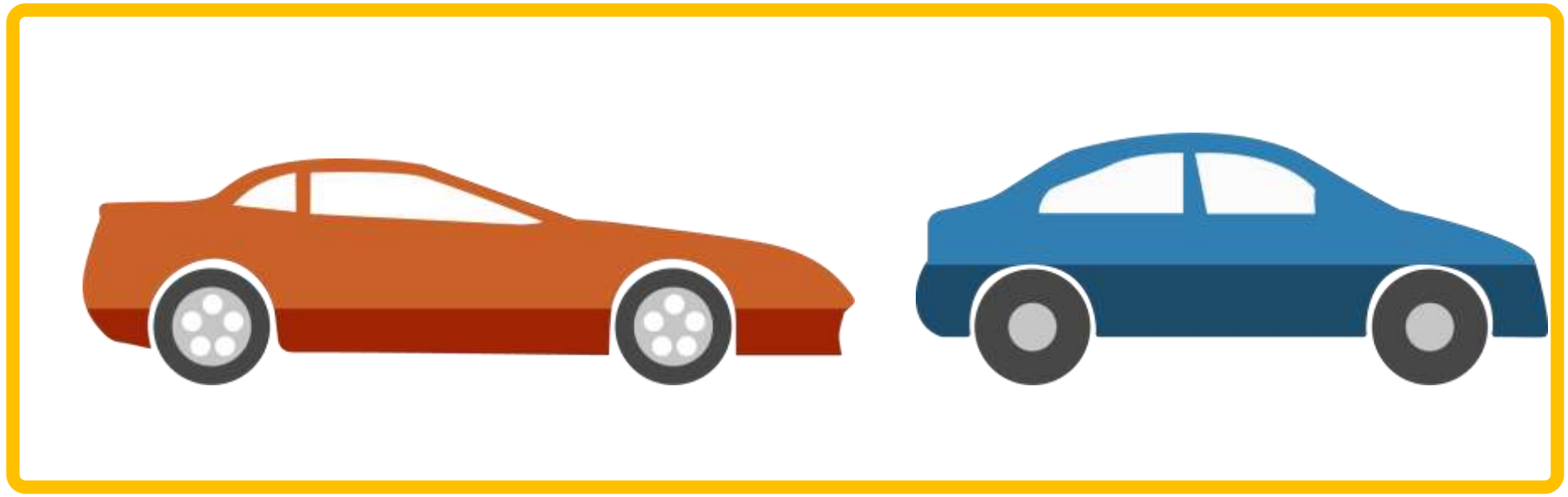
强化学习、机器学习、深度学习



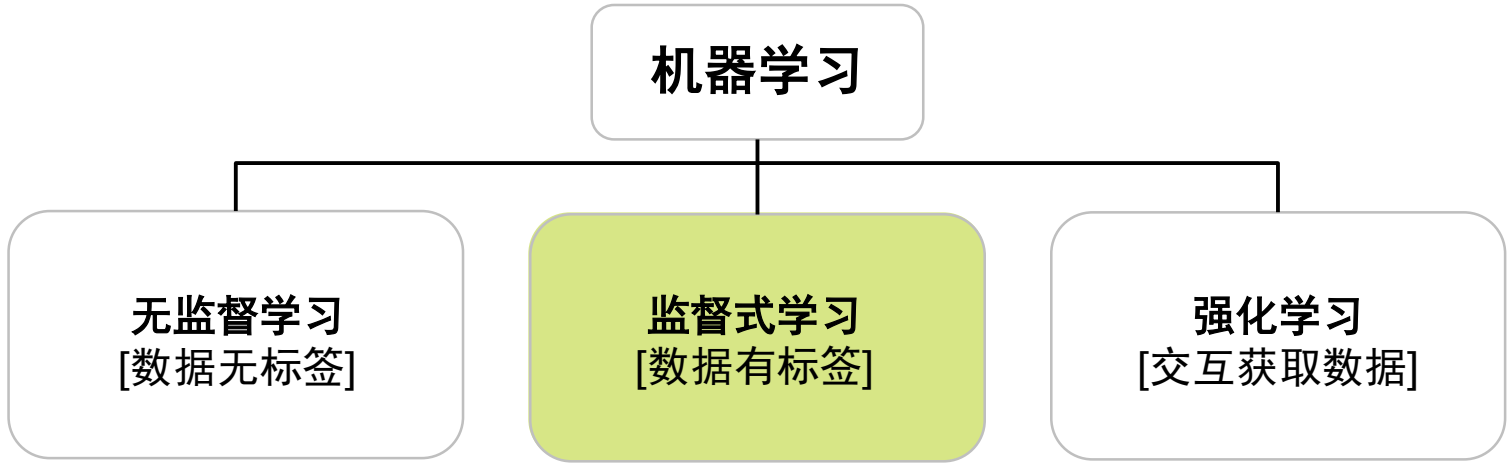


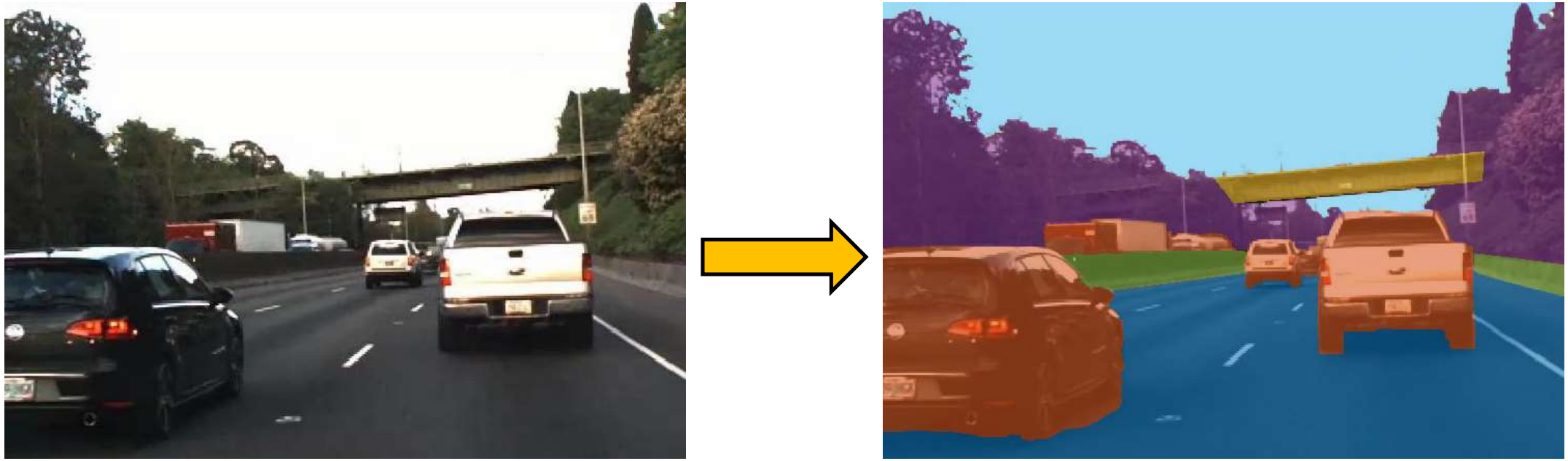
强化学习、机器学习、深度学习



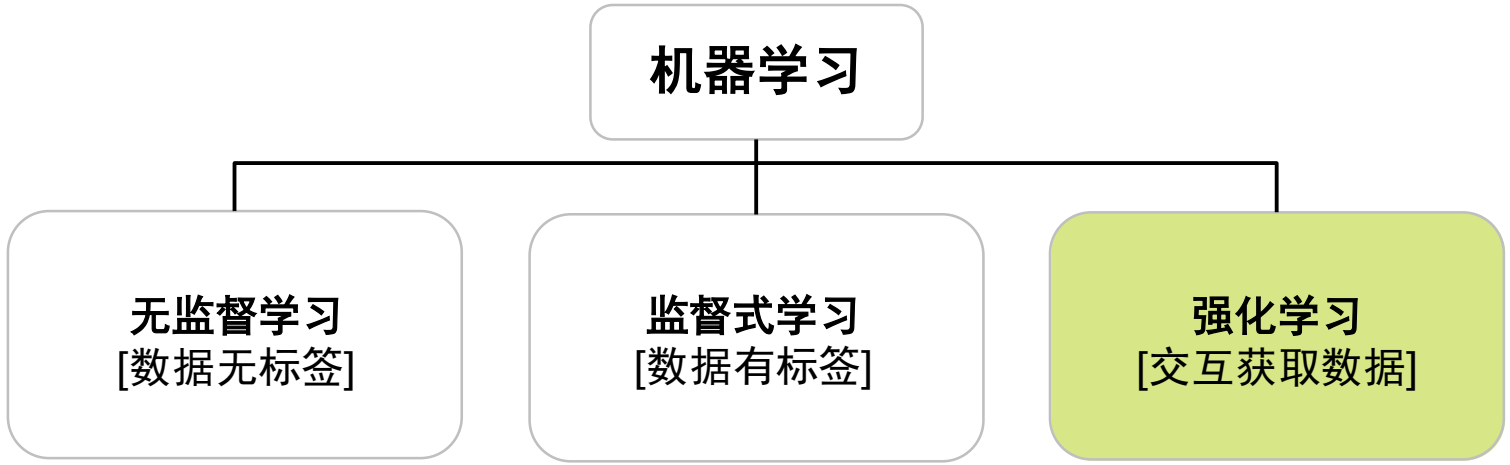


强化学习、机器学习、深度学习

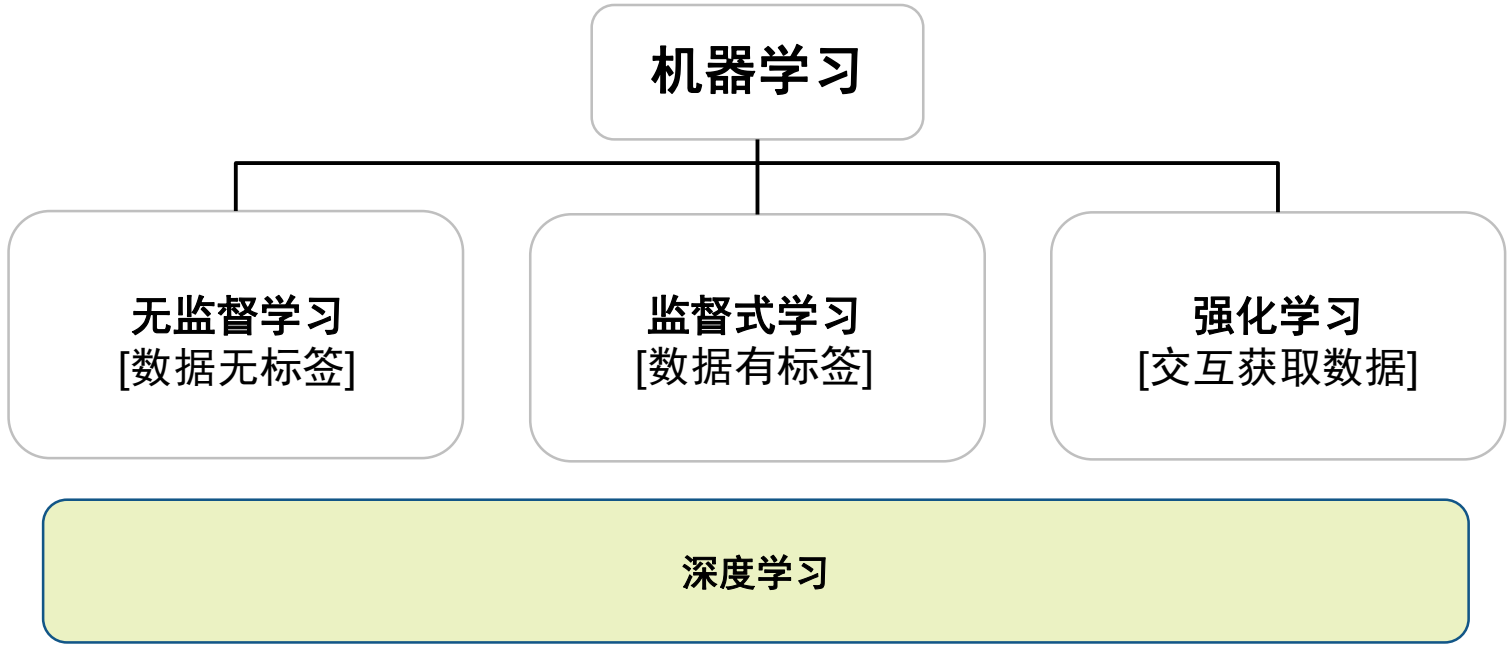




强化学习、机器学习、深度学习



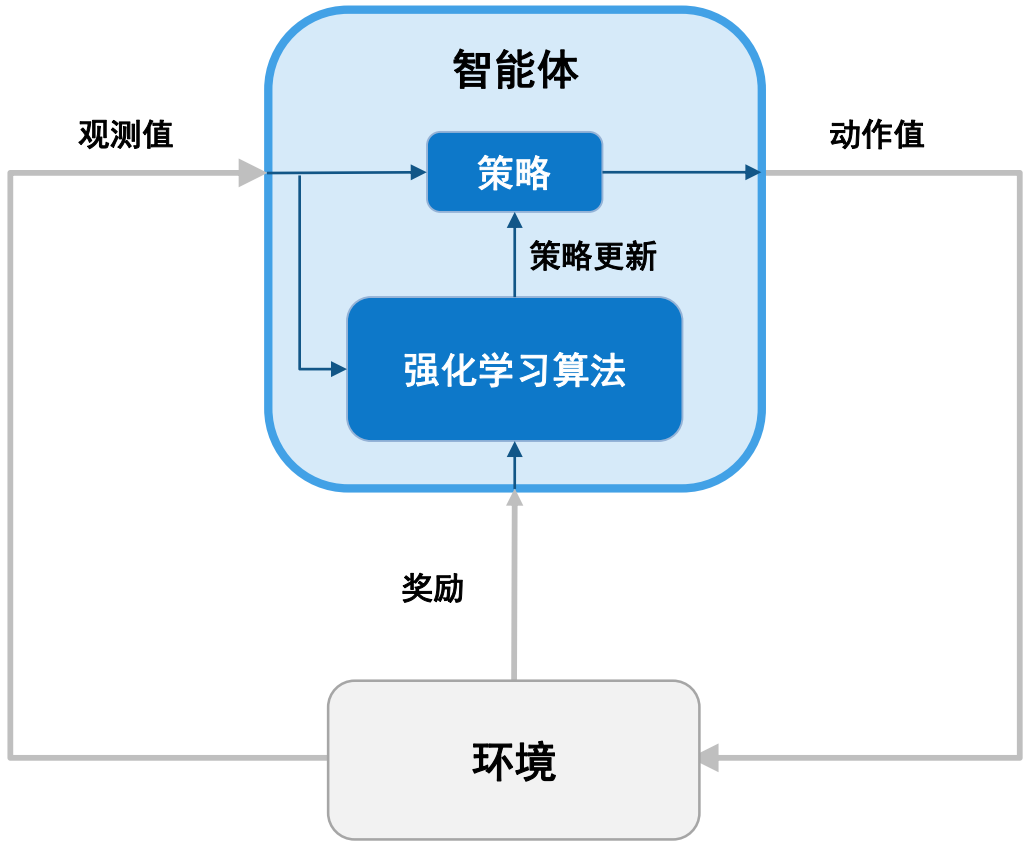
强化学习、机器学习、深度学习



深度学习扮演着什么角色？

复杂的强化学习问题通常需要借助深度神经网络
[深度强化学习]

强化学习基本概念
自动驾驶场景

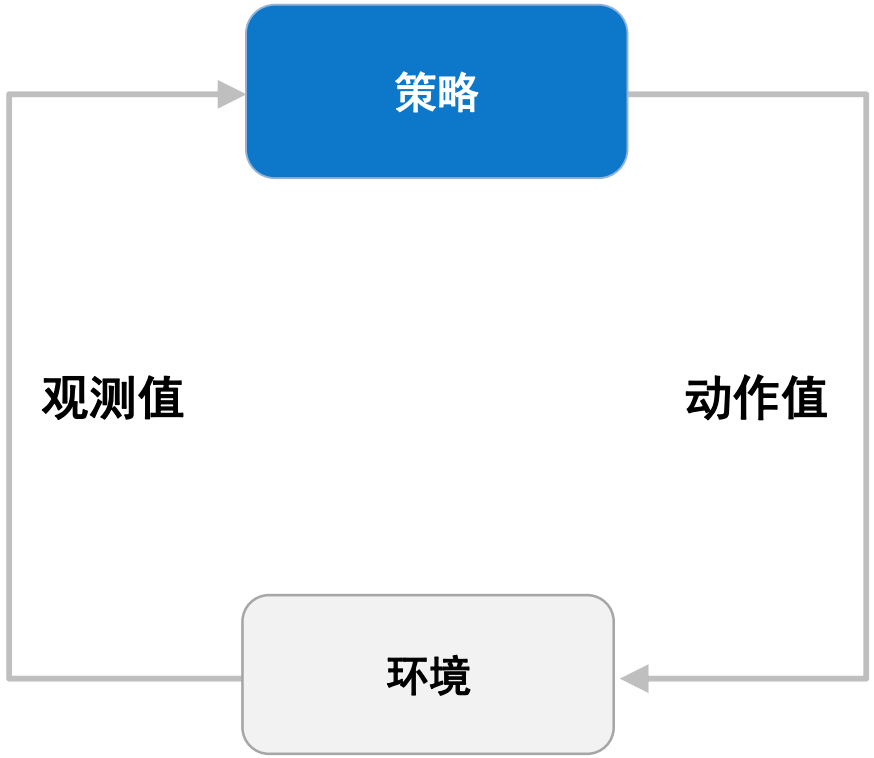


- 车载计算机…
  (**智能体**)
- 从激光雷达、摄像头等读取传感器数据…
  (**观测值**)
- 以表征道路状况、汽车位置等…
  (**环境**)
- 并生成转向、刹车和油门指令…
  (**动作值**)
- 根据内置的状态-动作映射关系…
  (**策略**)
- 试图优化设计目标，例如单圈时间、燃油经济性…
  (**奖励**)

- 通过不断试错，**强化学习算法**将更新策略

强化学习基本概念
自动驾驶场景

训练完成后，只需要根据策略进行控制

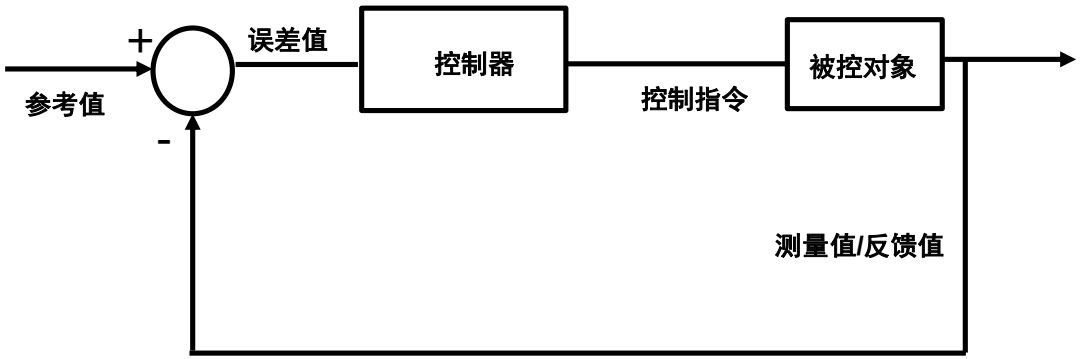


- 车载计算机使用最后更新好的状态-动作映射…
  (策略)
- 生成转向、刹车和油门指令…
  (动作)
- 根据激光雷达、摄像头等传感器数据…
  (观测值)
- 用于表征道路状况、车辆位置…
  (环境)

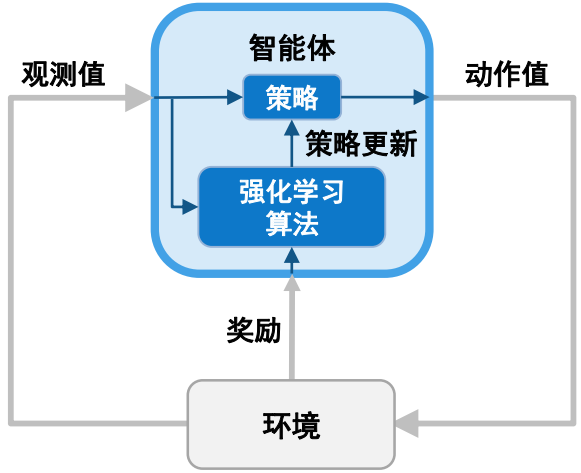
根据任务定义，训练好的策略将优化单圈时间和燃油经济性

强化学习 vs 控制论

控制系统



强化学习系统



调整机制	强化学习算法
误差/损失函数	奖励
控制指令	动作值
测量值/反馈值	观测值
被控对象	环境
控制器	策略

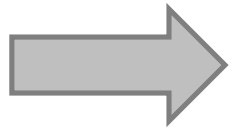
强化学习和**控制系统设计**的横向类比

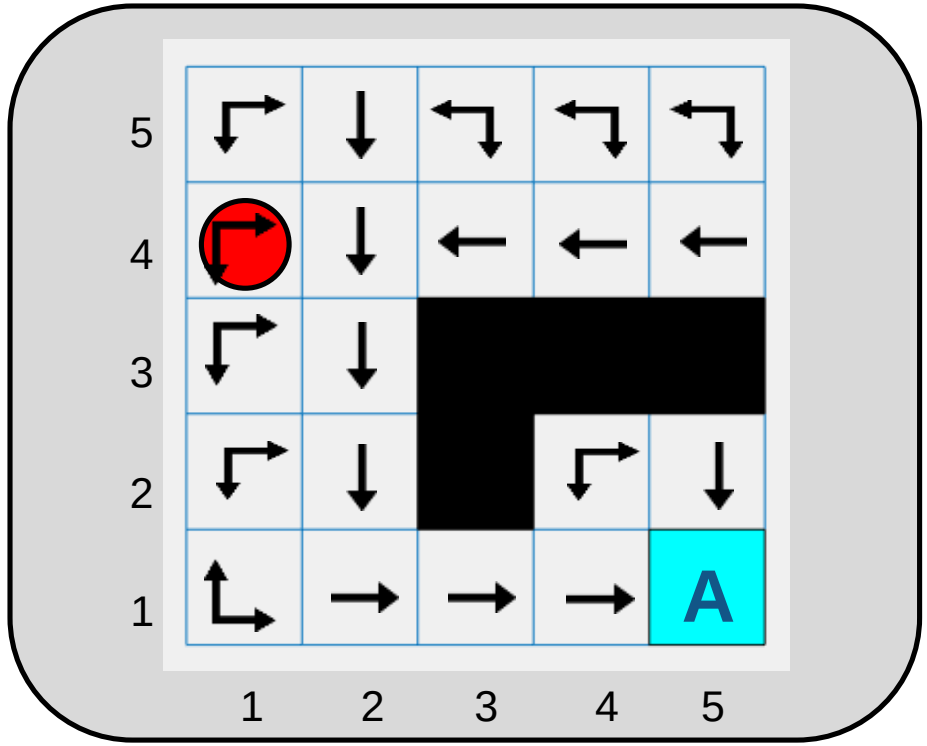
策略表达和深度学习

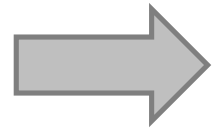
表达形式

- 查找表
- 多项式

观测值







下一个动作值

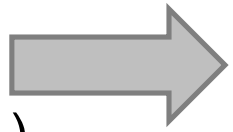
查找表 **无法很好地扩展到高维空间**

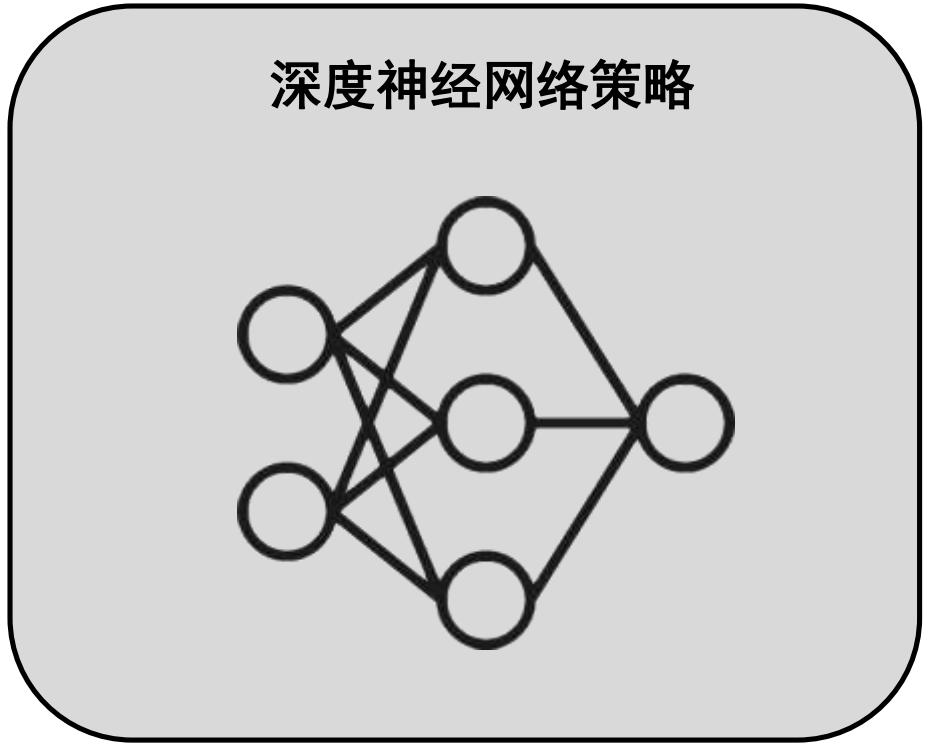
策略表达和深度学习

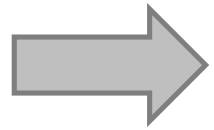
表达形式

- 查找表
- 多项式
- **(深度) 神经网络**

观测值
(摄像头，传感器 …)







下一个动作值

神经网络支持**复杂策略**的表达

如何在MATLAB中实现
基于强化学习的自主系统设计工作流

示例：机器人、自动驾驶

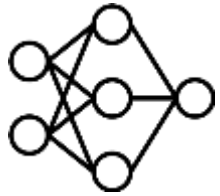
强化学习工作流

- 仿真模型或物理硬件
- 虚拟模型更安全/成本更低

环境

奖励

- 数值反馈用于评估控制性能
- 奖励函数定义和重塑难度高



策略表达

- 神经网络？查找表？多项式？

智能体

- 选择训练算法
- 超参数调优

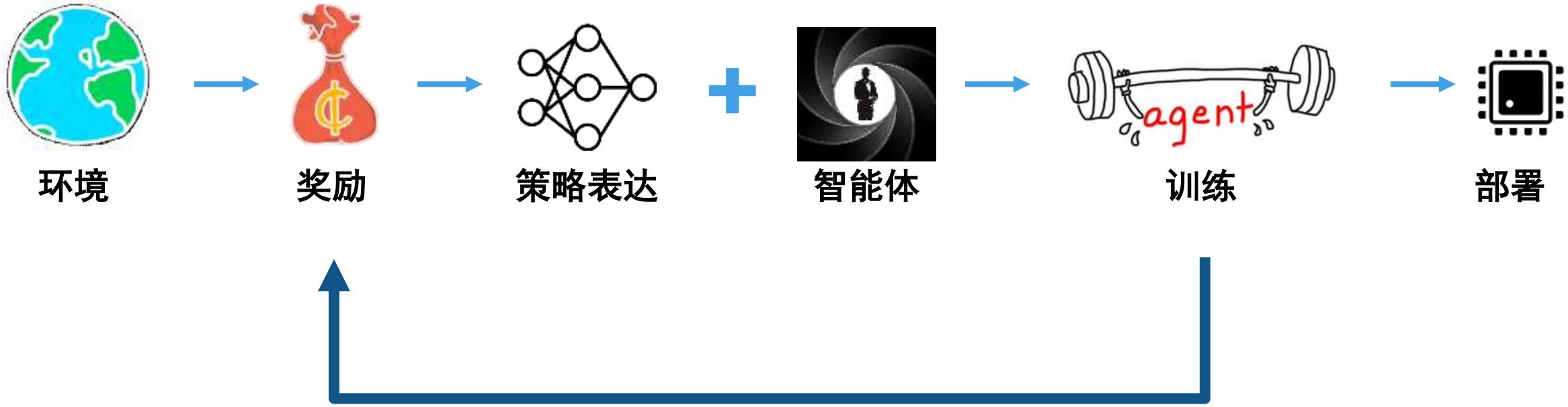
训练

- 需要进行大量仿真数据生成
- 并行运算/GPU计算加速训练
- 训练过程耗时长

部署

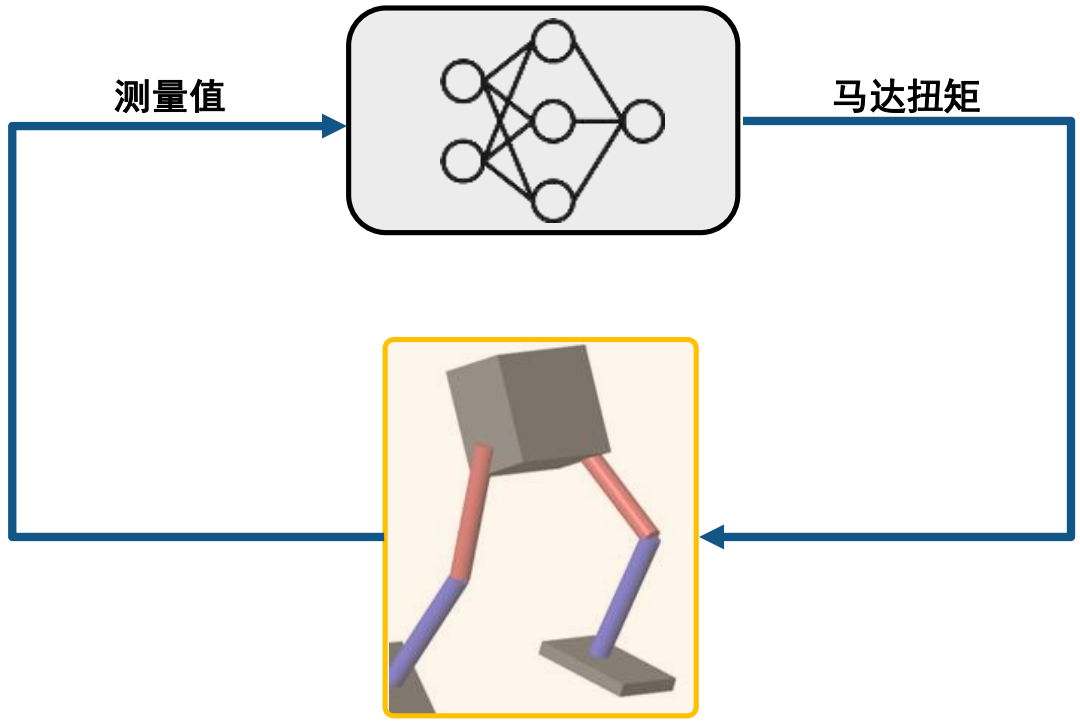
- 训练好的策略是一个独立函数

强化学习工作流

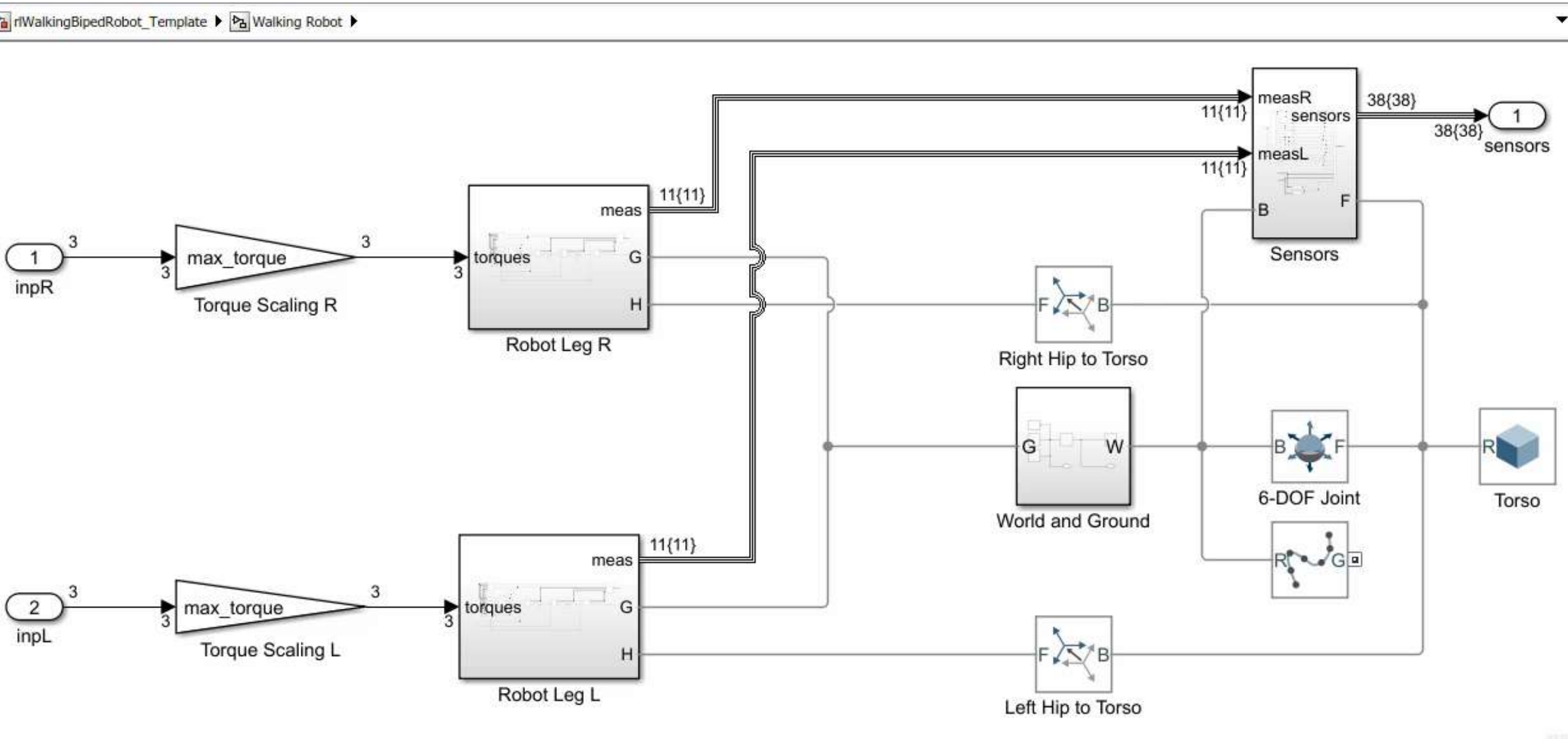


示例：双足机器人

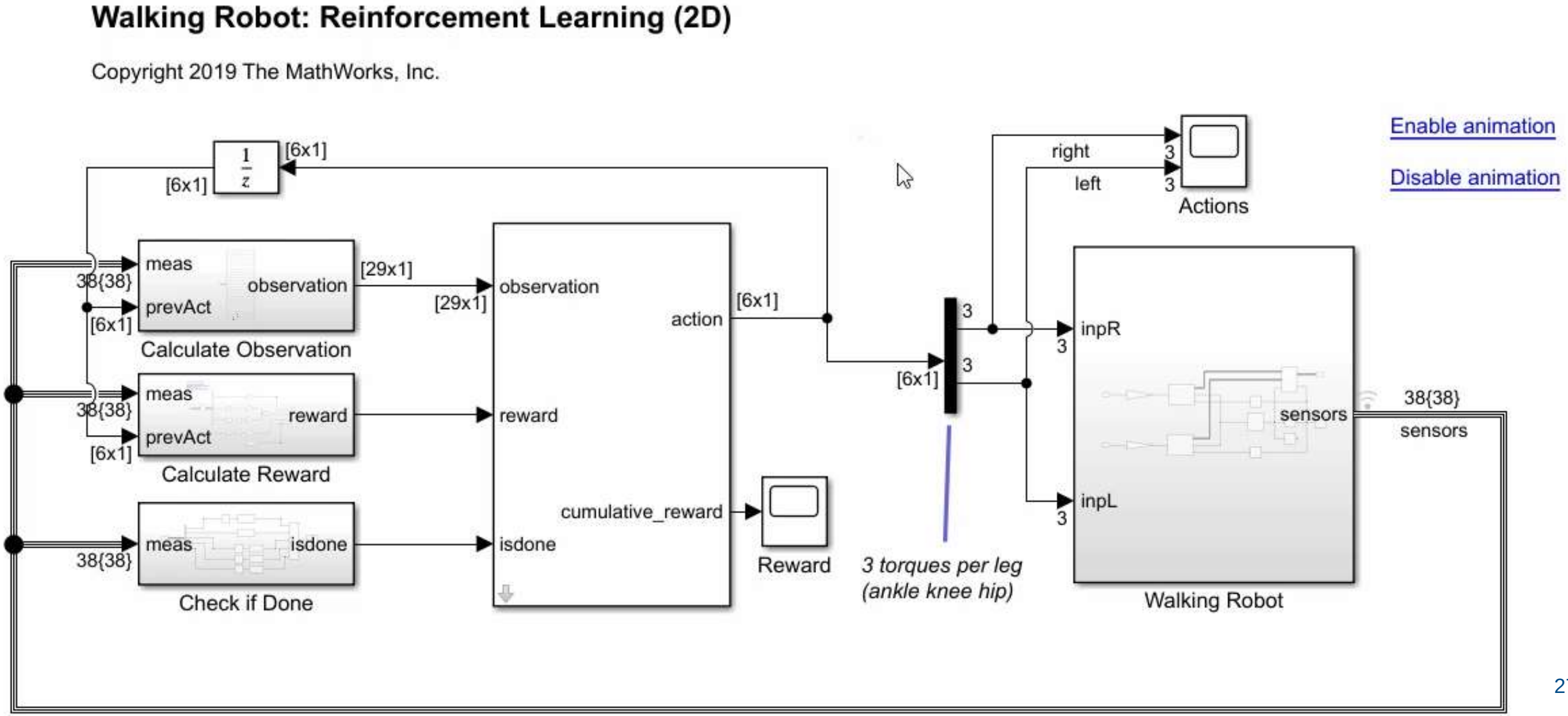
控制目标：沿直线行走



创建环境


rlWalkingBipedRobot_Template
Walking Robot
1
inpR
3
max_torque
Torque Scaling R
torques
meas
G
H
Robot Leg R
11{11}
2
inpL
Torque Scaling L
Robot Leg L
measR
measL
sensors
B
F
38{38}
Sensors
1
sensors
F
B
Right Hip to Torso
G
W
World and Ground
6-DOF Joint
R
Torso
Left Hip to Torso

塑造奖励

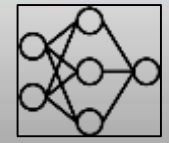

Walking Robot: Reinforcement Learning (2D)
Copyright 2019 The MathWorks, Inc.
Enable animation
Disable animation
1/z
[6x1]
[6x1]
right
left
3
3
Actions
meas
prevAct
observation
38{38}
[6x1]
[29x1]
[29x1]
Calculate Observation
observation
action
reward
cumulative_reward
isdone
[6x1]
meas
prevAct
reward
38{38}
[6x1]
Calculate Reward
meas
isdone
38{38}
Check if Done
Reward
3
3
[6x1]
3 torques per leg
(ankle knee hip)
inpR
inpL
3
3
sensors
38{38}
sensors
Walking Robot

Deep Network Designer

DESIGNER

New Import | Duplicate Cut Copy Paste | Fit to View Zoom In Zoom Out | Auto Arrange | Analyze | Export

FILE BUILD NAVIGATE LAYOUT ANALYSIS EXPORT



LAYER LIBRARY

Filter layers...

INPUT

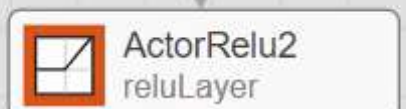
- imageInputLayer
- image3dInputLayer
- sequenceInputLayer
- roiInputLayer

CONVOLUTION AND FULLY CONNECTED

- convolution2dLayer
- convolution3dLayer
- groupedConvolution2dLayer
- transposedConv2dLayer
- transposedConv3dLayer
- fullyConnectedLayer

SEQUENCE

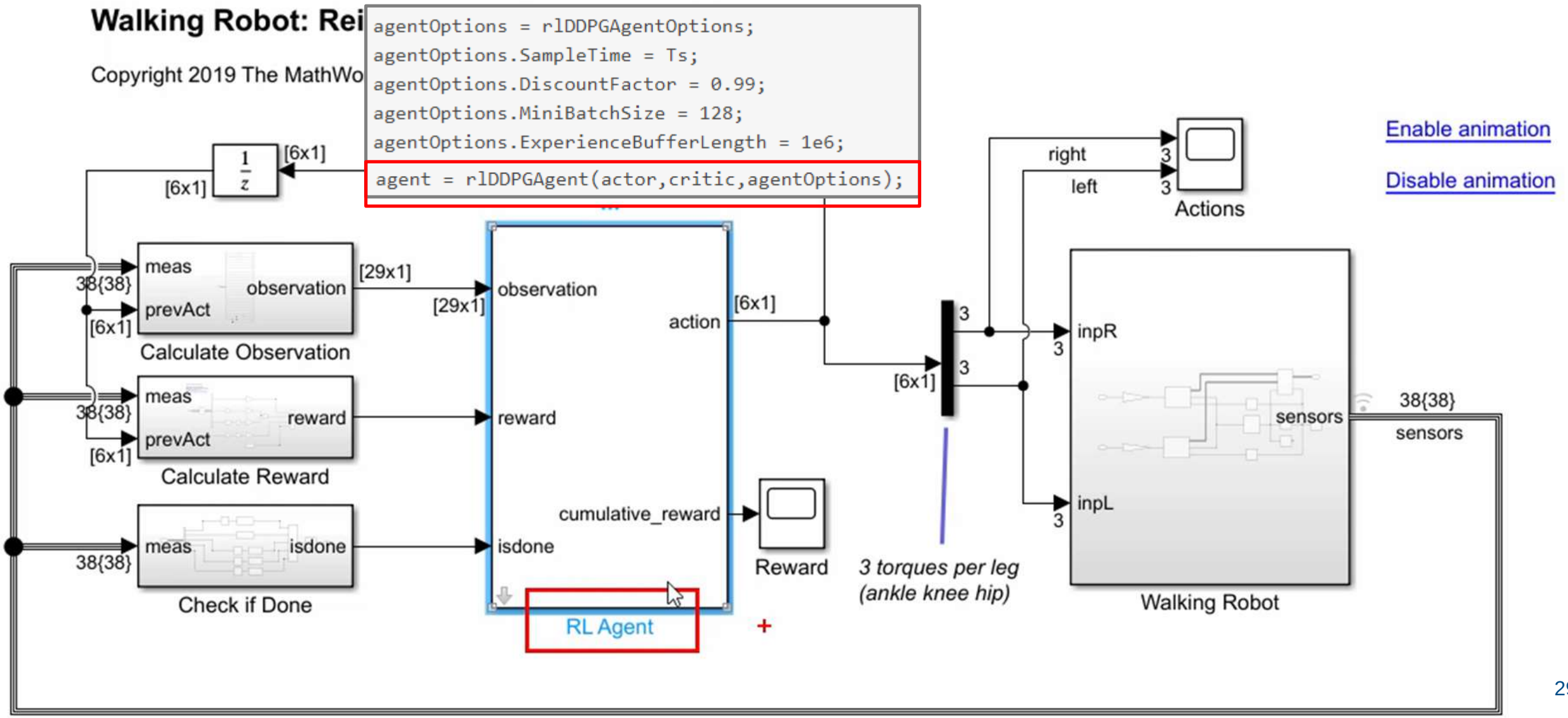
- lstmLayer
- bilstmLayer



PROPERTIES

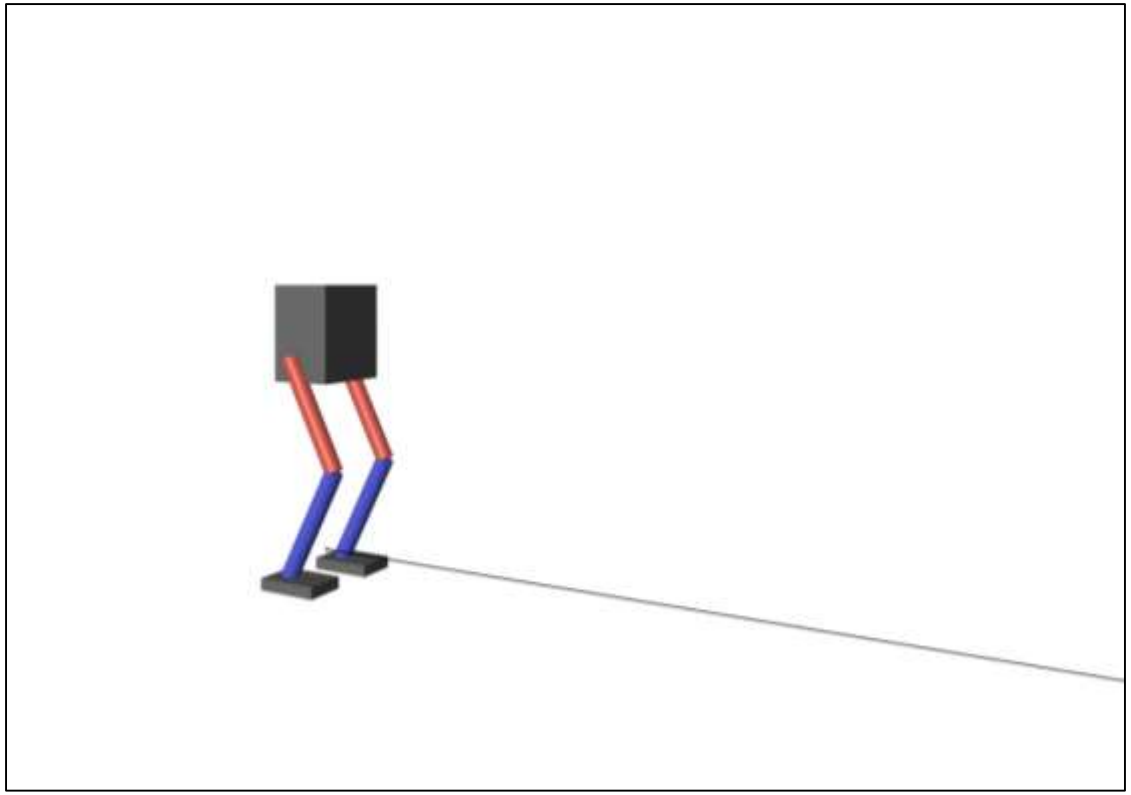
Number of layers	7
Number of connections	6
Input type	Image
Output type	None

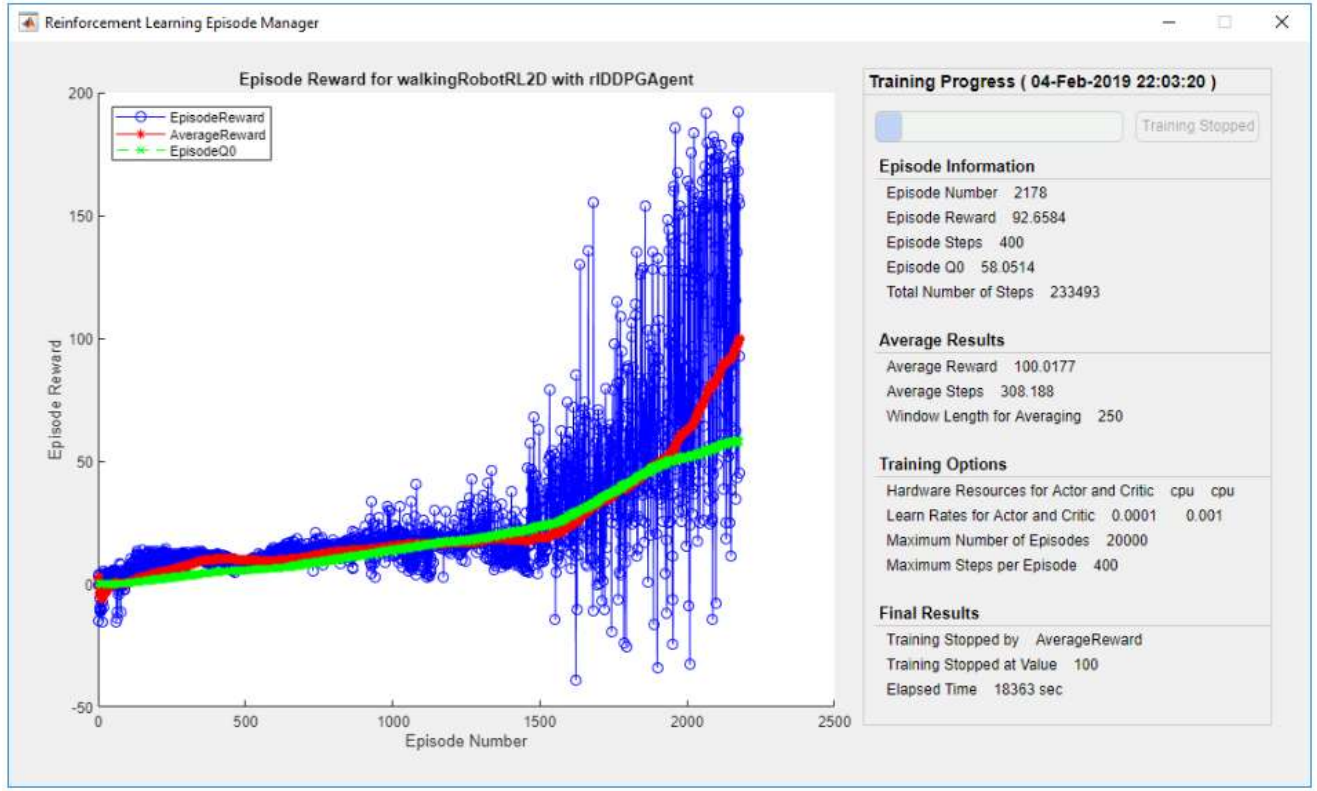
创建智能体

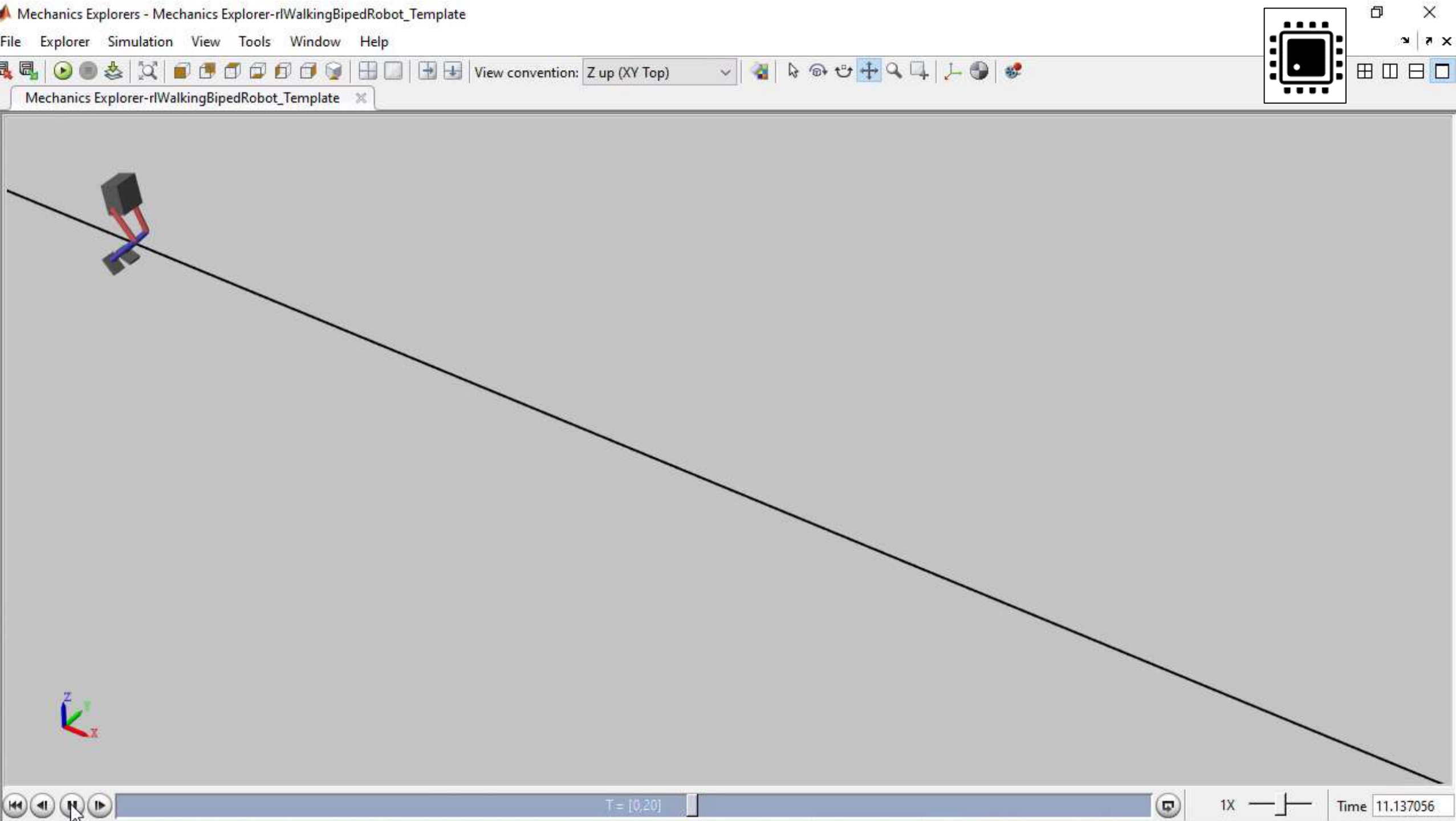

Walking Robot: Rei
Copyright 2019 The MathWo
agentOptions = rlDDPGAgentOptions;
agentOptions.SampleTime = Ts;
agentOptions.DiscountFactor = 0.99;
agentOptions.MiniBatchSize = 128;
agentOptions.ExperienceBufferLength = 1e6;
agent = rlDDPGAgent(actor,critic,agentOptions);
Enable animation
Disable animation
right
left
Actions
[6x1]
1/z
meas
prevAct
observation
[29x1]
Calculate Observation
reward
Calculate Reward
isdone
Check if Done
action
cumulative_reward
RL Agent
Reward
3 torques per leg
(ankle knee hip)
inpR
inpL
sensors
Walking Robot
38{38}

训练智能体

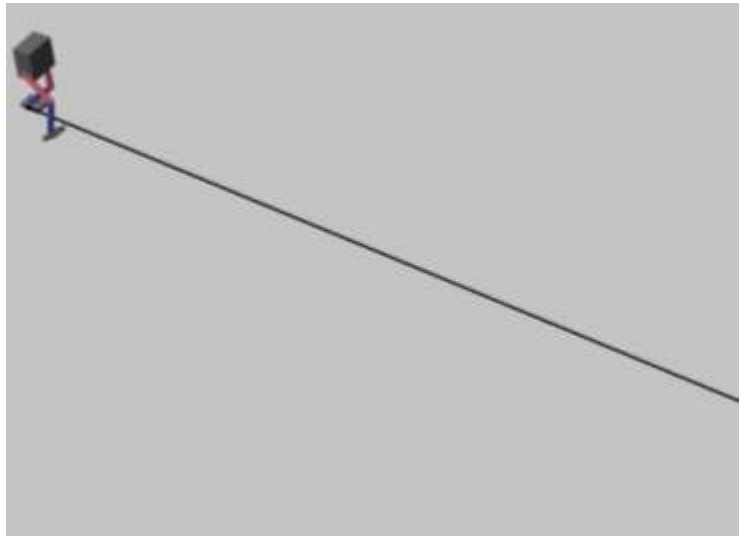

```
trainOpts.UseParallel = true;
trainOpts.ParallelizationOptions.Mode = 'async';
```

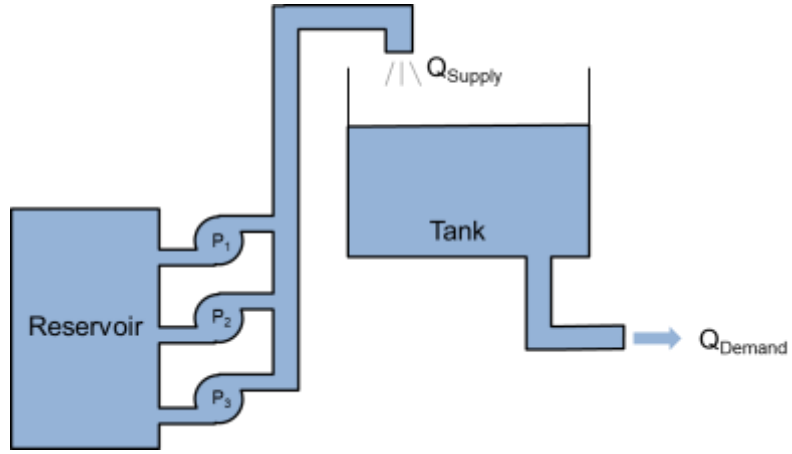


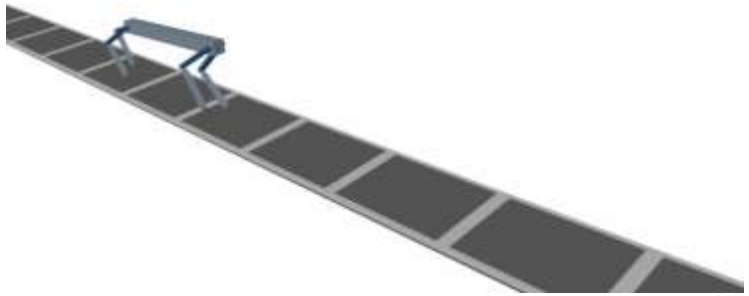


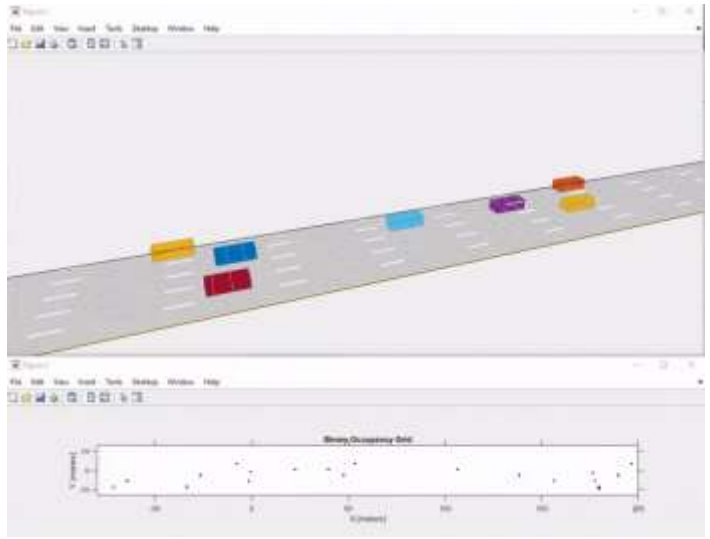

Mechanics Explorers - Mechanics Explorer-rlWalkingBipedRobot_Template
File
Explorer
Simulation
View
Tools
Window
Help
View convention:
Z up (XY Top)
Mechanics Explorer-rlWalkingBipedRobot_Template
T = [0,20]
1X
Time
11.137056

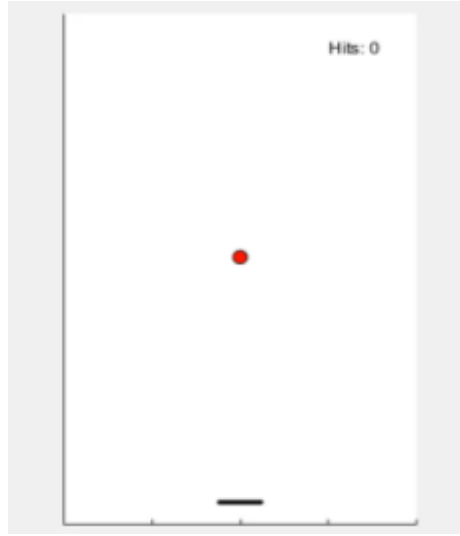
强化学习应用领域



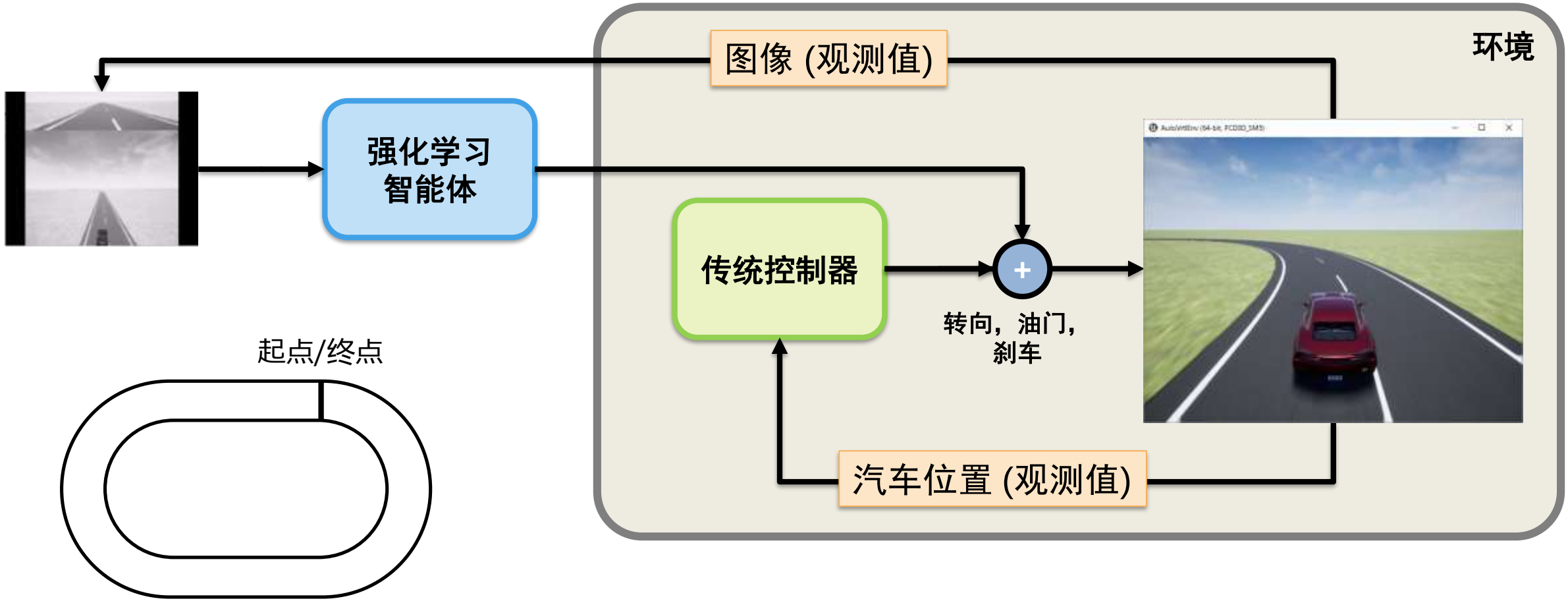




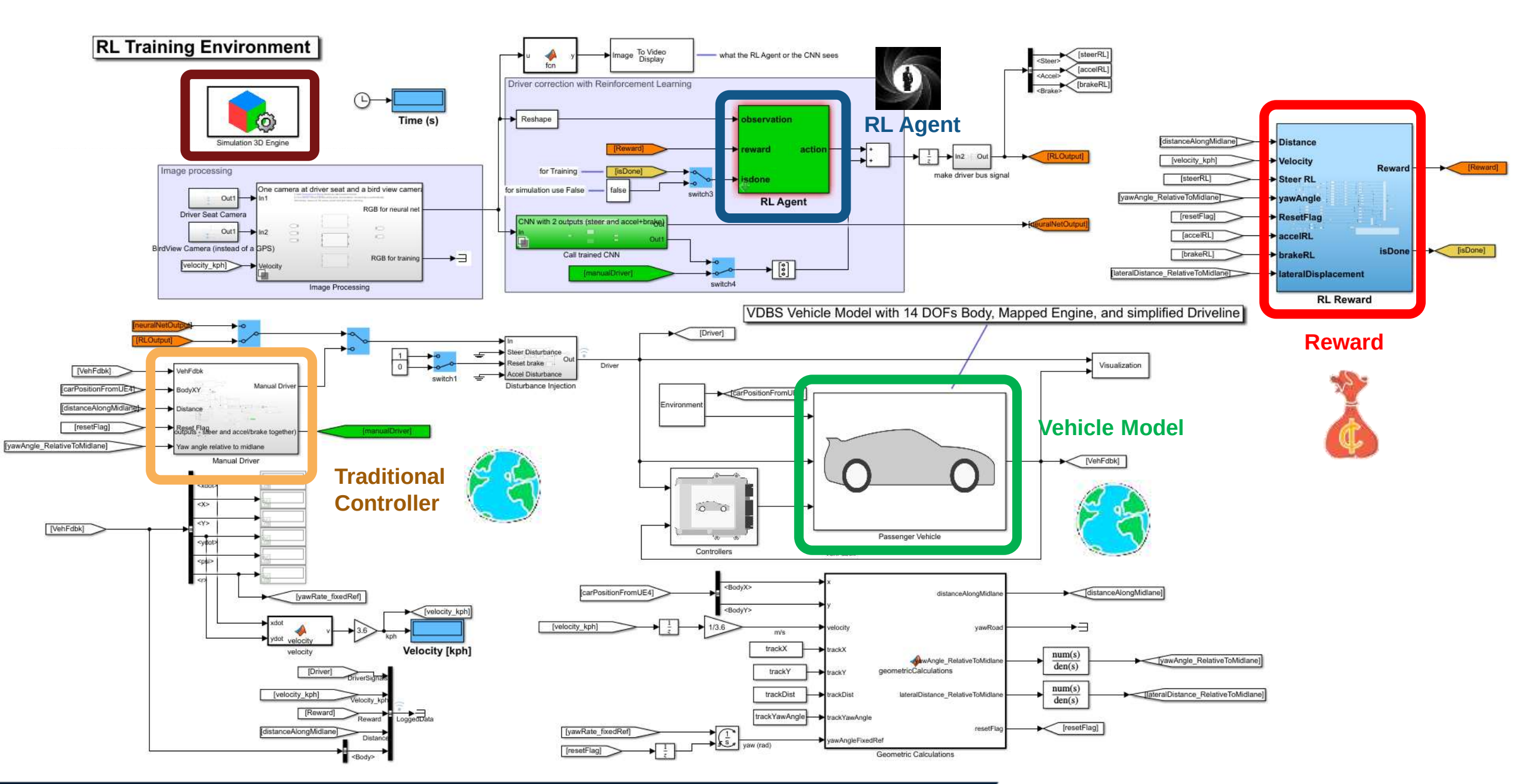




自动驾驶示例



目标：通过强化学习
增强传统控制器以优化单圈时间


RL Training Environment
Simulation 3D Engine
Time (s)
Image processing
One camera at driver seat and a bird view camera
Driver Seat Camera
BirdView Camera (instead of a GPS)
RGB for neural net
RGB for training
Image Processing
fcn
Image To Video Display
what the RL Agent or the CNN sees
Driver correction with Reinforcement Learning
Reshape
[Reward]
for Training
[isDone]
for simulation use False
false
switch3
observation
reward
action
isdone
RL Agent
CNN with 2 outputs (steer and accel+brake)
Call trained CNN
[manualDriver]
switch4
RL Agent
<Steer>
<Accel>
<Brake>
[steerRL]
[accelRL]
[brakeRL]
make driver bus signal
[RLOutput]
[neuralNetOutput]
[distanceAlongMidlane]
[velocity_kph]
[steerRL]
[yawAngle_RelativeToMidlane]
[resetFlag]
[accelRL]
[brakeRL]
[lateralDistance_RelativeToMidlane]
Distance
Velocity
Steer RL
yawAngle
ResetFlag
accelRL
brakeRL
lateralDisplacement
Reward
isDone
RL Reward
Reward
[neuralNetOutput]
[RLOutput]
[VehFdbk]
[carPositionFromUE4]
[distanceAlongMidlane]
[resetFlag]
[yawAngle_RelativeToMidlane]
VehFdbk
BodyXY
Distance
Reset Flag
Yaw angle relative to midlane
Manual Driver
[manualDriver]
Manual Driver
Traditional Controller
switch1
In
Steer Disturbance
Reset brake
Accel Disturbance
Out
Disturbance Injection
Driver
VDBS Vehicle Model with 14 DOFs Body, Mapped Engine, and simplified Driveline
[Driver]
Visualization
Environment
[carPositionFromUE4]
Controllers
Passenger Vehicle
Vehicle Model
[VehFdbk]
[VehFdbk]
<xdot>
<X>
<Y>
<ydot>
<psi>
<r>
[yawRate_fixedRef]
xdot
ydot
velocity
3.6
kph
[velocity_kph]
Velocity [kph]
[Driver]
[velocity_kph]
[Reward]
[distanceAlongMidlane]
<Body>
DriverSignals
Velocity_kph
Reward
Distance
LoggedData
[carPositionFromUE4]
[velocity_kph]
[yawRate_fixedRef]
[resetFlag]
<BodyX>
<BodyY>
1/3.6
m/s
yaw (rad)
trackX
trackY
trackDist
trackYawAngle
x
y
velocity
trackX
trackY
trackDist
trackYawAngle
yawAngleFixedRef
geometricCalculations
Geometric Calculations
distanceAlongMidlane
yawRoad
yawAngle_RelativeToMidlane
lateralDistance_RelativeToMidlane
resetFlag
num(s)
den(s)
[distanceAlongMidlane]
[yawAngle_RelativeToMidlane]
[lateralDistance_RelativeToMidlane]
[resetFlag]

Simulation 3D Engine

Clock

0.75

Time (s)

Reshape

observation

[Reward]

reward

action

for Training

[isDone]

for simulation use False

false

switch3

isdone

RL Agent

make driver bus signal

[RLOutput]

[neuralNetOutput]

Image processing

Out1

Driver Seat Camera

One camera at driver seat and a bird view camera

In1

RGB for neural net

Out1

In2

BirdView Camera (instead of a GPS)

RGB for training

[velocity_kph]

Velocity

Image Processing

CNN with 2 outputs (steer and accel+brake)

[distanceAlongMidlane]

[velocity_kph]

[steerRL]

[yawAngle_RelativeToMidlane]

[resetFlag]

[accelRL]

[brakeRL]

[lateralDistance_RelativeToMidlane]

Distance

Velocity

Steer RL

yawAngle

ResetFlag

accelRL

brakeRL

lateralDisplacement

Reward

isDone

RL Reward

AutoVrtlEnv (64-bit, PCD3D_SM5)

Reinforcement Learning Episode Manager

Episode Reward for OvalTrackVDBS$_{RL_1}$9a with rlDDPGAgent

Episode Reward

Episode Number

0

0.1

0.2

0.3

0.4

0.5

0.6

0.7

0.8

0.9

1

Training Options

Hardware Resources for Actor and Critic	cpu	cpu
Learn Rates for Actor and Critic	0.001	0.0001
Maximum Number of Episodes	5000	
Maximum Steps per Episode	167	

Final Results

Training Stopped by	...
Training Stopped at Value	...
Elapsed Time	...

BS Vehicle Model with 14 DOFs Body, Mapped Engine, and simplified Driveline

Visualization

Passenger Vehicle

VehFdbck

To Video Display2

m/s

velocity

trackX

trackY

trackDist

trackYawAngle

yaw (rad)

yawAngleFixedRef

distanceAlongMidlane

yawRoad

geometricCalculations

lateralDistance_RelativeToMidlane

resetFlag

[resetFlag]

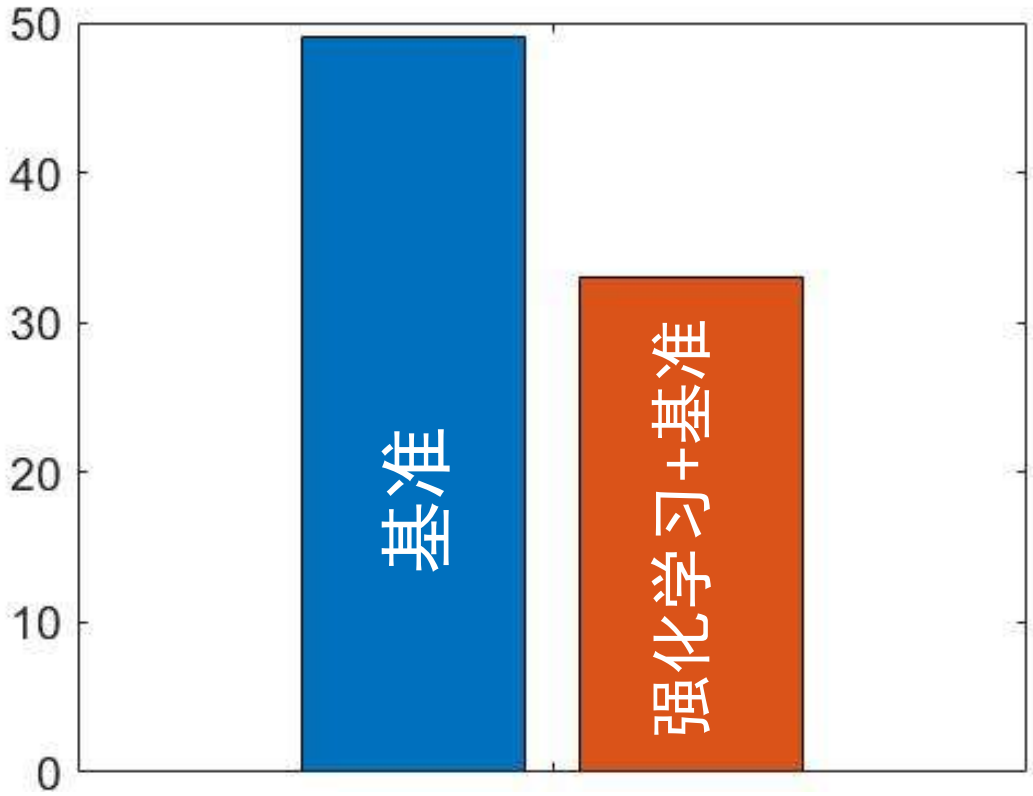
num(s)
den(s)

num(s)
den(s)

Geometric Calculations

结果比较

单圈时间 (s)



30% 性能提高

传统控制器 + 强化学习

强化学习的优势与技术挑战

优势

- 训练前**无需数据**
- 为难以解决的问题提供了AI方向**新的可能性**
- 可处理**不确定和非线性**环境问题
- 可开发复杂的端到端解决方案

挑战

- 训练好的策略**难以验证**(无法保障性能)
- 需要大量试错数据(样本效率低)
  - 使用真实物理硬件进行训练成本高且具有危险性
- 大量设计参数
  - 奖励信号
  - 网络架构
  - 训练超参数

仿真是强化学习的关键

MATLAB和Simulink如何帮助您加快强化学习工作流？

挑战

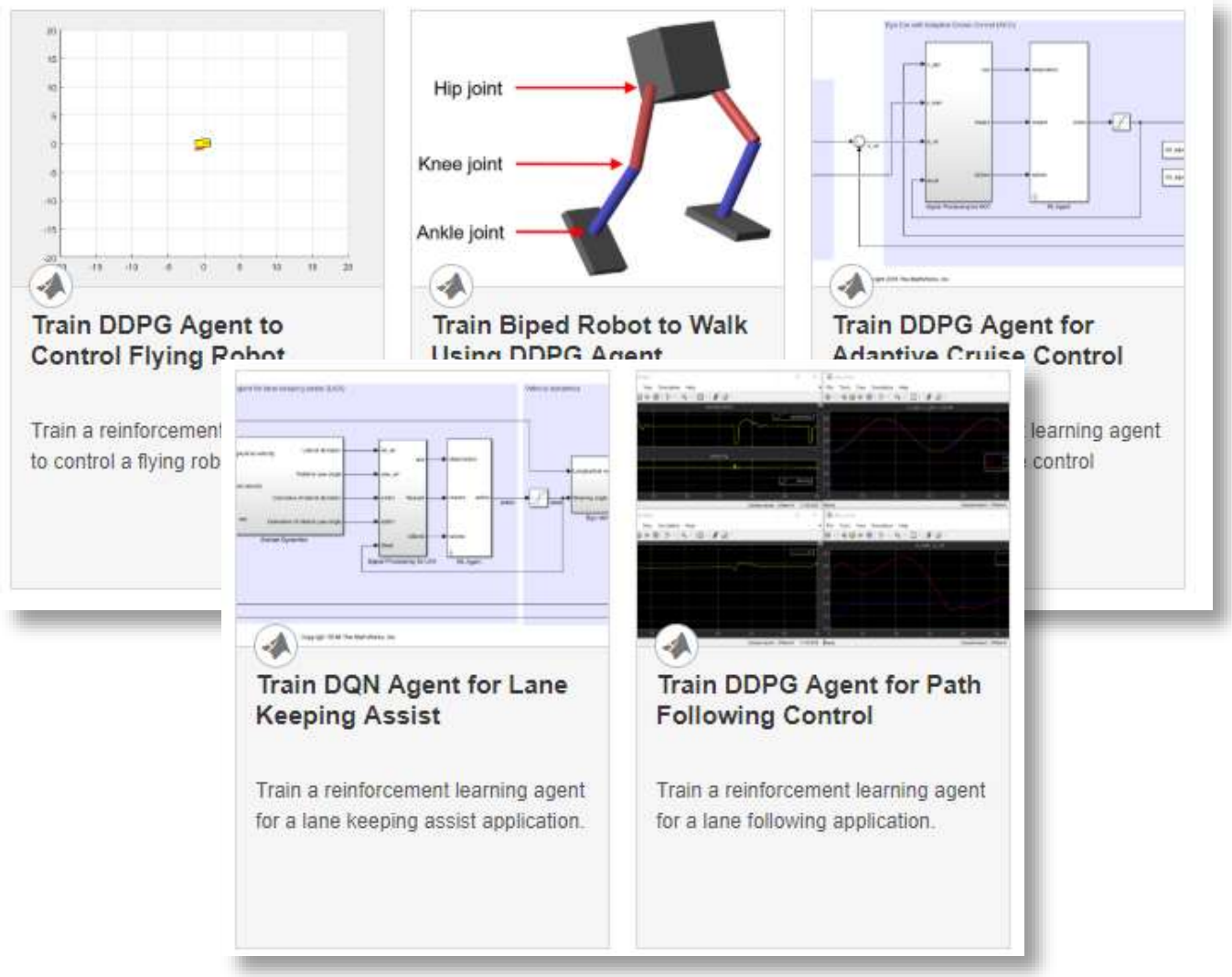
- 训练好的策略**难以验证**
  (无法保障性能)
- 需要大量试错数据
  (样本效率低)
  - 使用真实物理硬件进行训练成本高且具有危险性
- 大量**设计参数**
  - 奖励信号
  - 网络架构
  - 训练超参数

MATLAB&SIMULINK

- **复用现有的代码和模型创建环境**
- 通过仿真**验证策略**
  - 仿真极端场景
- **并行**运行仿真加速训练
- 参考强化学习工具箱**示例**
  - 通过仿真迭代调试

更多参考示例

- 控制器设计
- 机器人行走控制
- 车道保持辅助驾驶
- 自适应巡航控制
- 模仿学习



了解更多

- 控制、自动驾驶、机器人等领域的参考示例
- 为工程师和领域专家撰写的专业文档
- 强化学习系列技术讲座视频
- MathWorks官网下载强化学习系列电子书