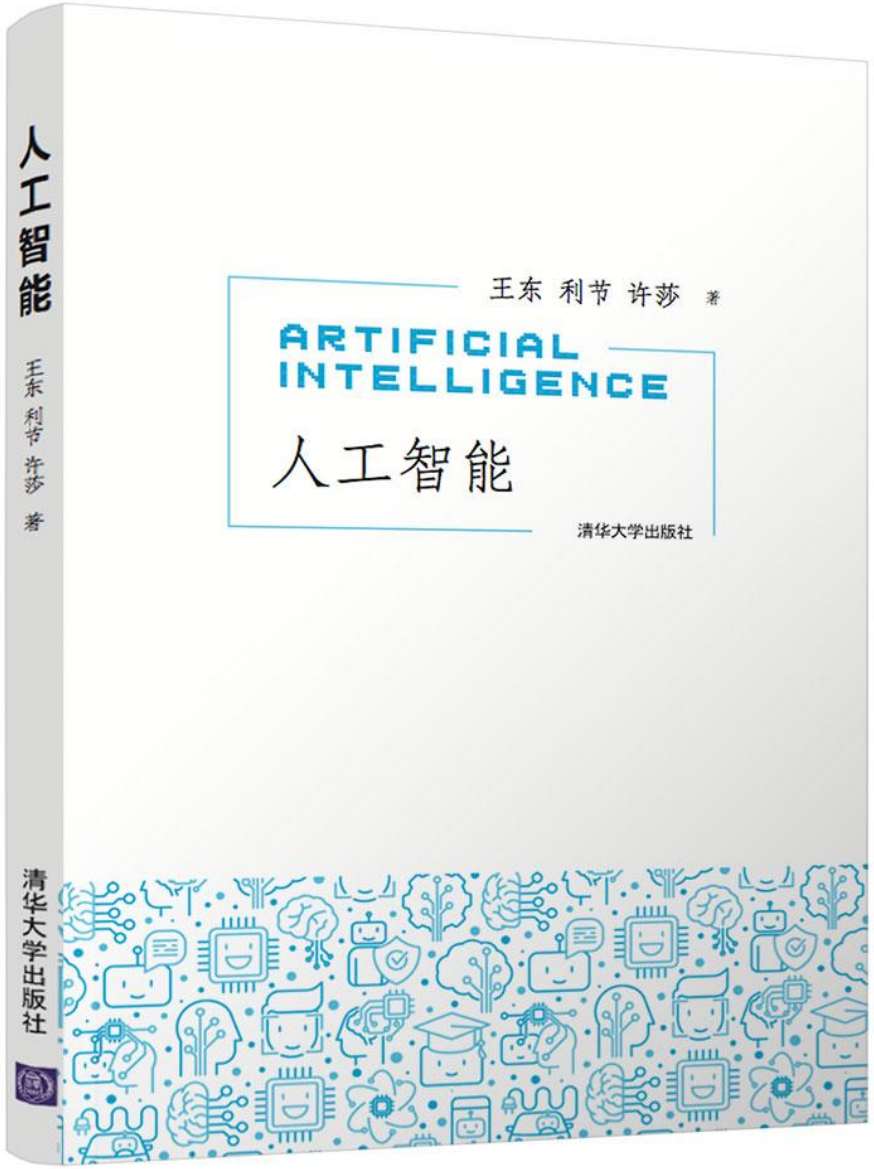
人工智能

王东 & 利节 & 许莎



- 教材名称：人工智能
- 出版社：清华大学出版社
- 作者：王东 利节 许莎

本课程应具备的能力：

- Linux下的基本操作
- python语言基础知识

资源下载：

http://aibook.cslt.org/

模仿你的行为

李蓝天

授课提示

1. 本课件60页，建议授课时间120分钟，每页平均2分钟。重点可放在“基于学习的机器人”和“深度强化学习方法”两部分，以提高学生的学习兴趣。
2. 若授课时间为90分钟，请去掉非课本内容，对“附注”中所提内容尽量简化，并对“基于设计的机器人”中的技术性内容进行缩减。
3. 课件内的图片、视频等素材具有较好的展示效果，建议保留。
4. 建议授课过程中对每个章节末尾的思考问题与学生进行讨论，激发学生思考。

目　录

- 引言
- 现代机器人发展史
- 基于设计的机器人
- 基于学习的机器人
- 深度强化学习方法

目　录

- 引言
- 现代机器人发展史
- 基于设计的机器人
- 基于学习的机器人
- 深度强化学习方法

创造机器的梦想

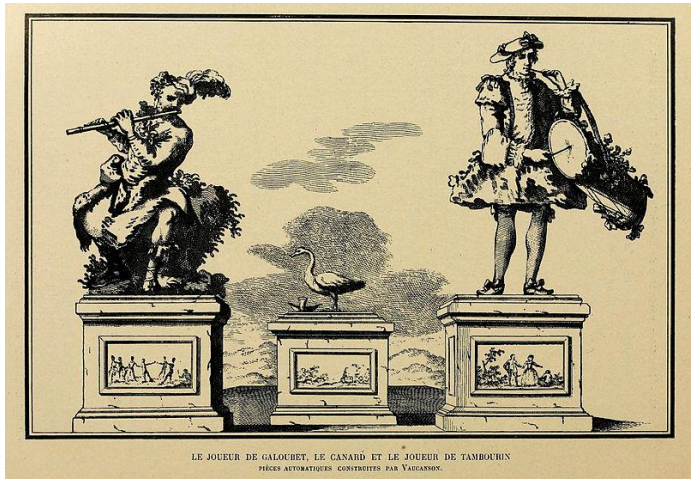
《墨子·鲁问篇》："公输子削竹木以为鹊，成而飞之，三日不下。"
—— 春秋鲁班木鸟

《晋书·舆服志》："记里鼓车，驾四。形制如司南。其中有木人执槌向鼓，行一里则打一槌。"
—— 汉代张衡记里鼓车

创造机器的梦想

1495年，列奥纳多·达·芬奇设计了仿人型的机械骑士，可以完成挥动手臂、摇头等简单动作。

—— 达芬奇机器人



1737年，雅克·迪·沃康松发明了一系列自动演奏机器，可以吹笛子和敲鼓。

18世纪末，田中久重发明了一系列复杂的玩偶，可以奉茶、写字。

梦想成真

机器人经历了模仿、感知和智能的进阶，如今已成为人类的伙伴。

人机对话

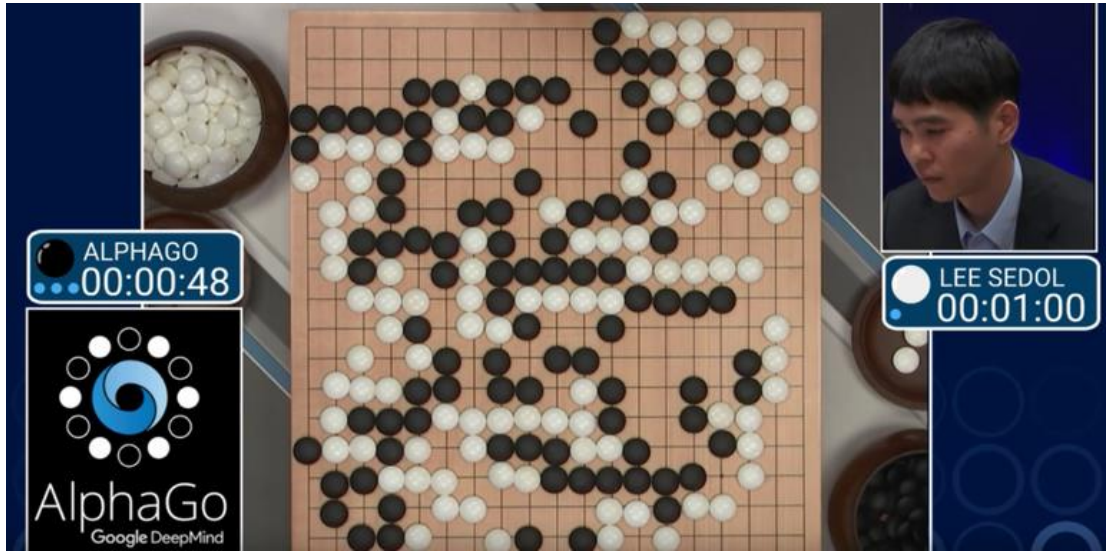
食品加工

消防灭火

自动驾驶

围棋高手 AlphaGo

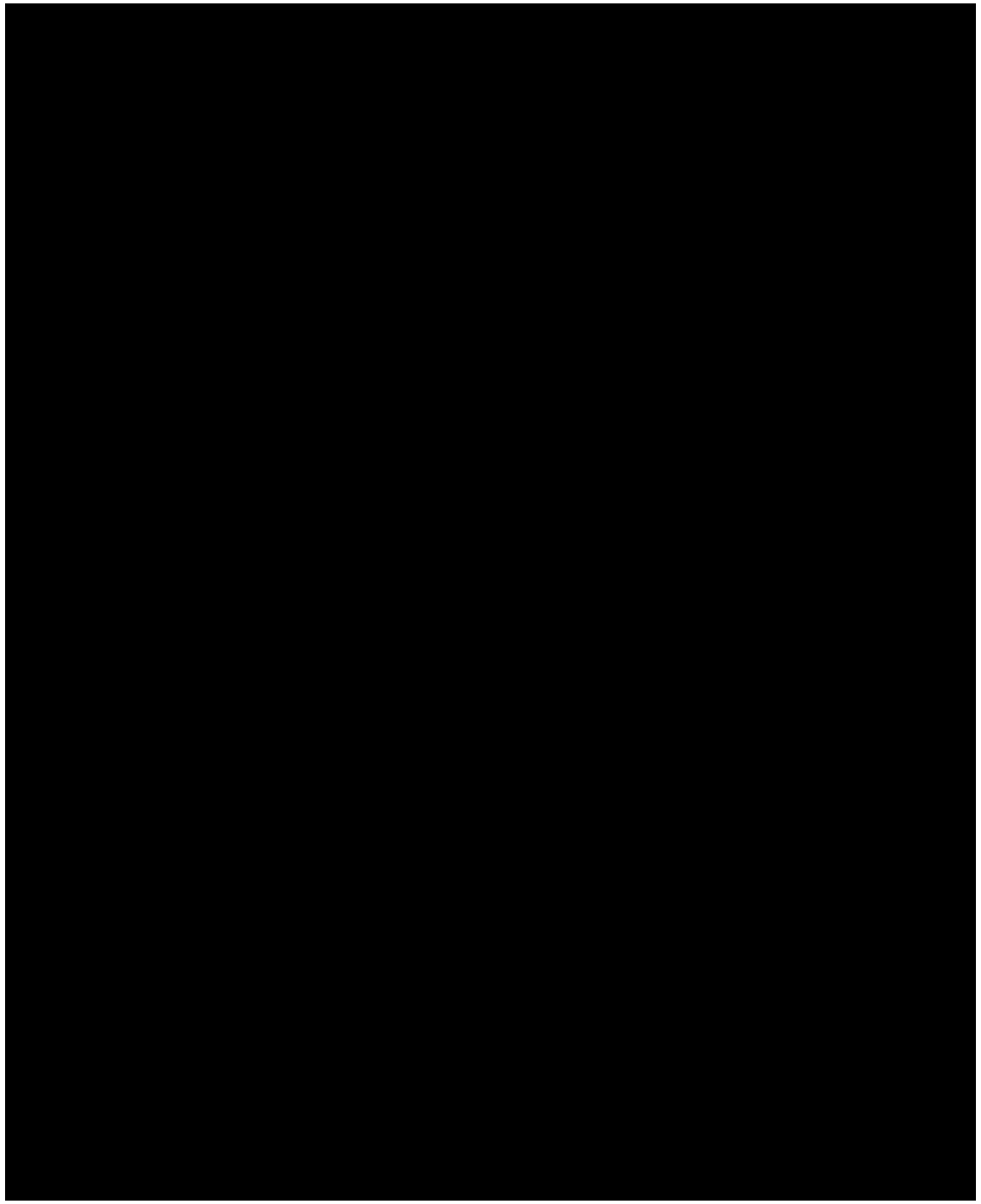
2015年至2017年，AlphaGo 机器人先后战胜了樊麾、李世石、柯洁等世界一流围棋高手。



游戏玩家

2013年11月，初创人工智能公司 DeepMind 开发了计算机玩壁球游戏。

两个月之后，谷歌花了500万欧元买下了 DeepMind 公司。



思考

- 在你生活无处不在的机器人!
- 想象机器人是如何模仿、感知，逐渐智能化的呢?

目 录

- 引言
- **现代机器人发展史**
- 基于设计的机器人
- 基于学习的机器人
- 深度强化学习方法

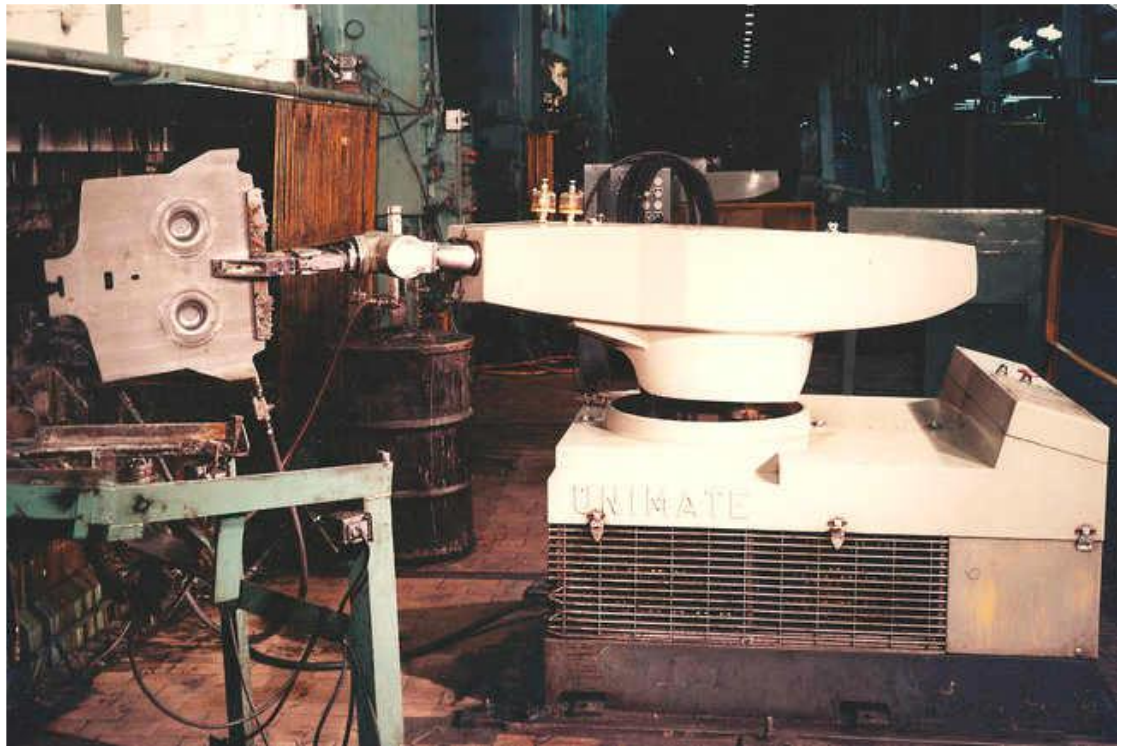
电报箱

1927 年，美国西屋公司的工程师温兹利（Roy J. Wensley）制造了第一个机器人“电报箱”，具有无线电报和简单问答等功能。

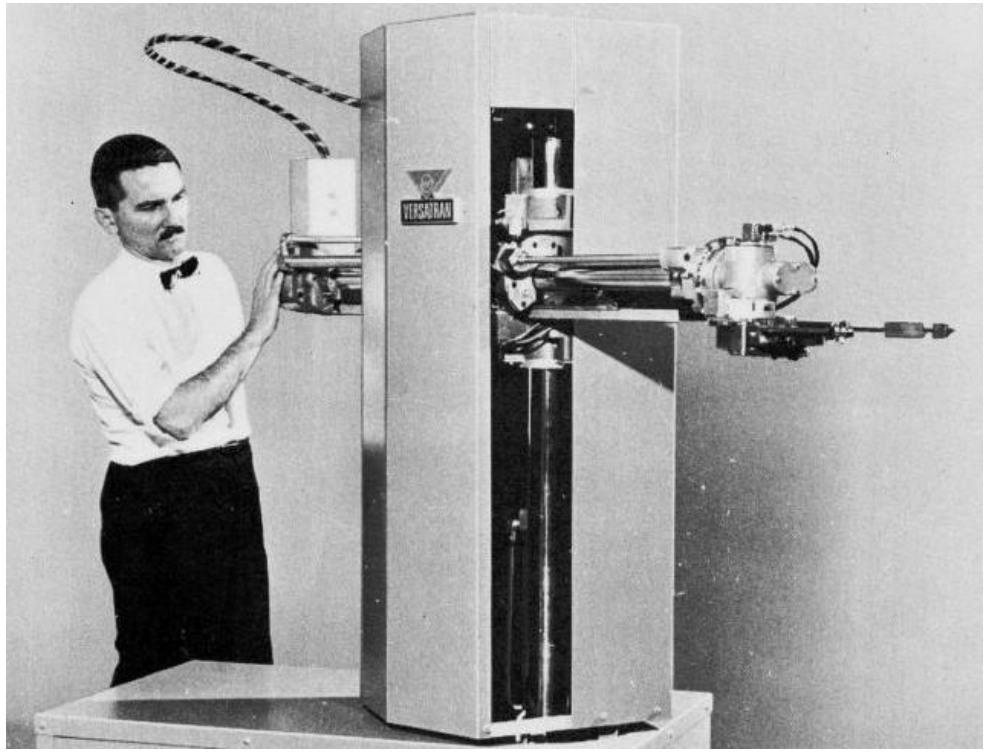
机械机器人

- 1947年，美国原子能委员会的阿尔贡研究所开发了遥控机械手。1948年又开发了机械式的主动机械手。用于代替人类处理危险的放射性物质。
- 1954年，美国工程师德沃尔提出一种用于工业生产的“重复性操作机器人”的设想。
- 1958年，英伯格和德沃尔（“机器人之父”）创立了世界上第一个机器人公司UNIMATION，并于1959年制造出了世界上第一台工业机器人UNIMATE（寓意为“万能自动”）。代替人类完成抓放零件等操作。
- 1962年，AMF公司推出了类似的工业机器人VERSTRAN（寓意为“万能搬动”）。

机械机器人



第一台工业机器人“UNIMATE”

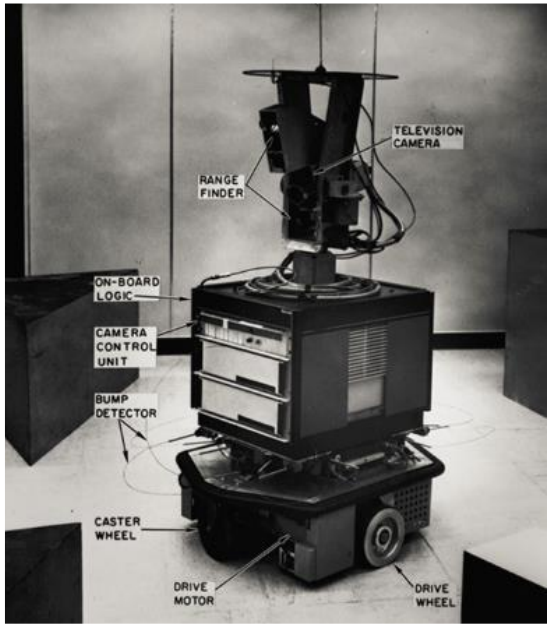


AMF 公司的 “VERSTRAN”

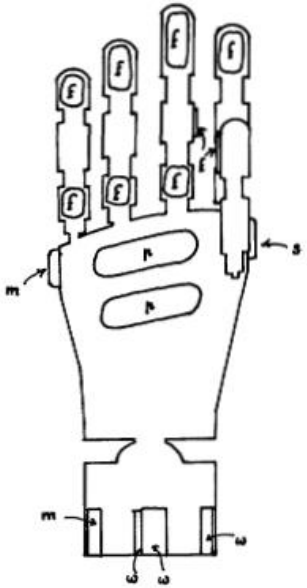
感知机器人

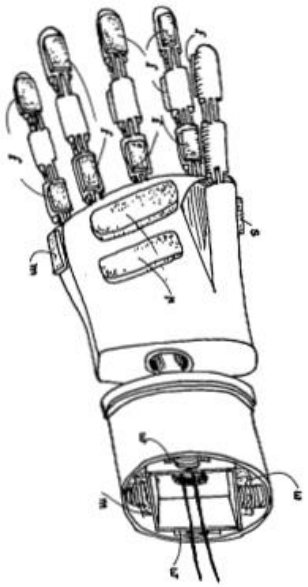
- 1961年，恩斯特采用了触觉传感器来帮助控制机械臂。
- 1962 年，托莫维奇和博尼设计了称为 Belgrade 的“灵巧手”，采用了压力传感器来检测“手”的力度。
- 1965 年，MIT 的 Lawrence Robert 推出了世界上第一个带有视觉传感器，能识别并定位积木的机器人。
- 1969 年，SRI 实验室设计了 Shakey 机器人，利用视觉传感器完成抓取积木的动作，并具有行走能力。
- 1979 年，Unimation 公司推出了 PUMA 系列工业机器人，是一款可编程的通用工业机械臂，具有可配置的视觉、触觉和力觉传感器。PUMA 的推出标志着工业机器人术走向成熟。

感知机器人

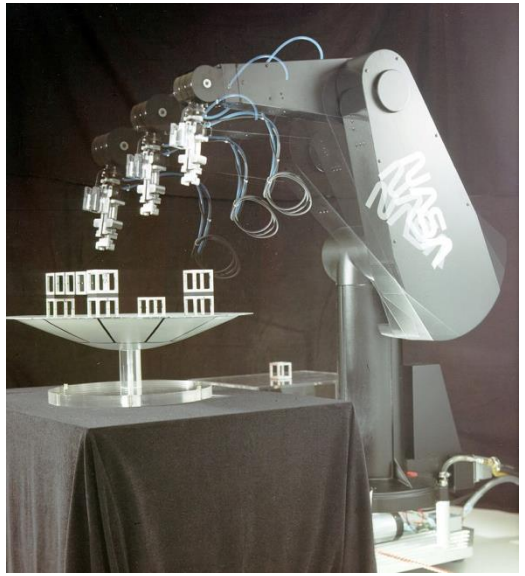


Shakey 机器人





Belgrade（“灵巧手”）



PUMA 系列工业机器人

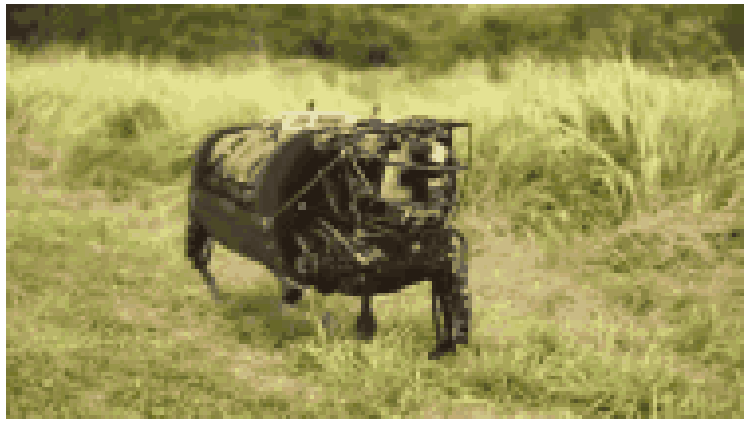
智能机器人

- 人工智能技术使机器人具备了更为强大的感知和学习能力。

- 可以通过视觉和听觉对环境进行感知，做出决策；
- 可以识别身份、情绪；
- 可以快速适应不同的工作环境。

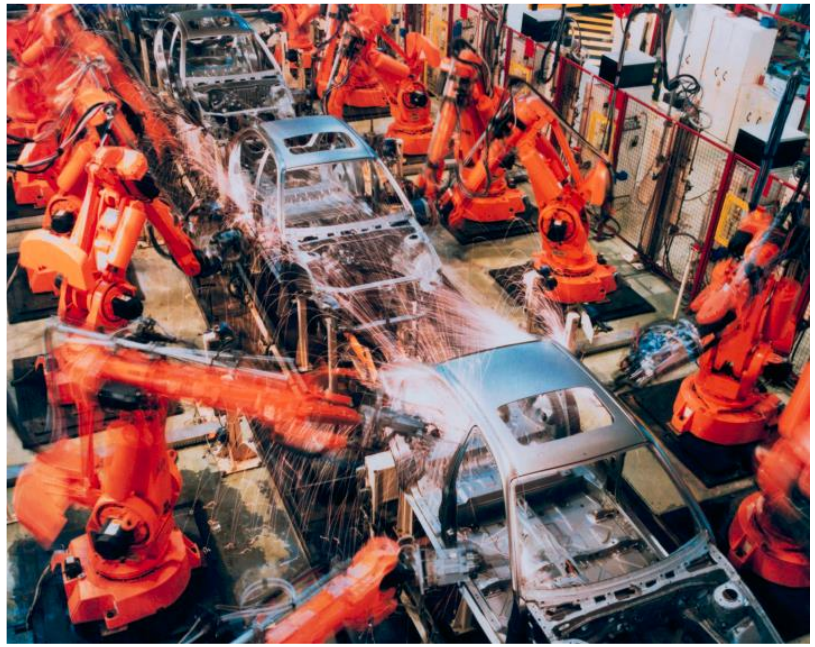
- 应用范围更加广泛、复杂，例如工业制造、精密医疗、高危作业。

智能机器人



波士顿动力公司的大狗（Big Dog）机器人

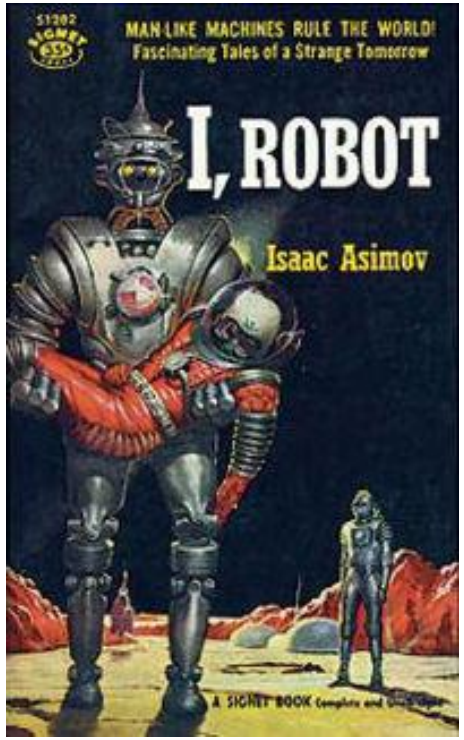
https://www.bilibili.com/video/av19629771/



汽车制造业

手术机器人

机器人与伦理



- 机器人是人工智能技术走向现实世界的载体。
- 如何规范机器人的行为?

- 科幻小说家艾萨克·阿西莫夫在小说《我，机器人》（I, Robot）中订立了著名的“机器人三定律”：
  1. 机器人不得伤害人类，且确保人类不受伤害；
  2. 在不违背第一法则的前提下，机器人必须服从人类的命令；
  3. 在不违背第一及第二法则的前提下，机器人必须保护自己。

思考

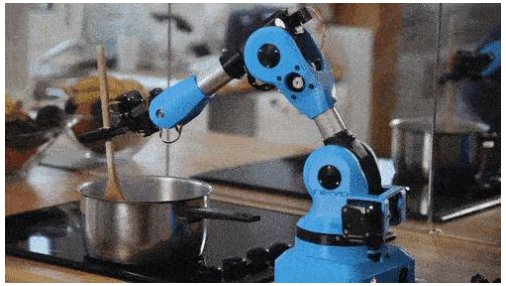
- 回顾机械机器人、感知机器人和智能机器人的发展历史
- 现实中，如何确保机器人不会伤害人类呢?

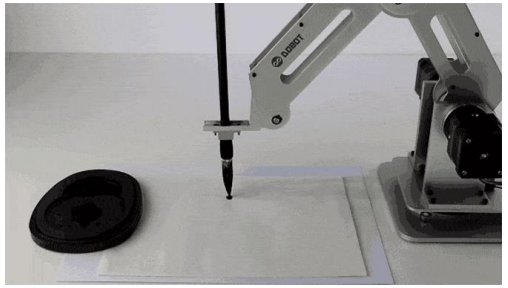
目　录

- 引言
- 现代机器人发展史
- **基于设计的机器人**
- 基于学习的机器人
- 深度强化学习方法

机器人分类

- 操作机器人
  - 关注如何通过关节运动完成某一动作
  - 机械臂

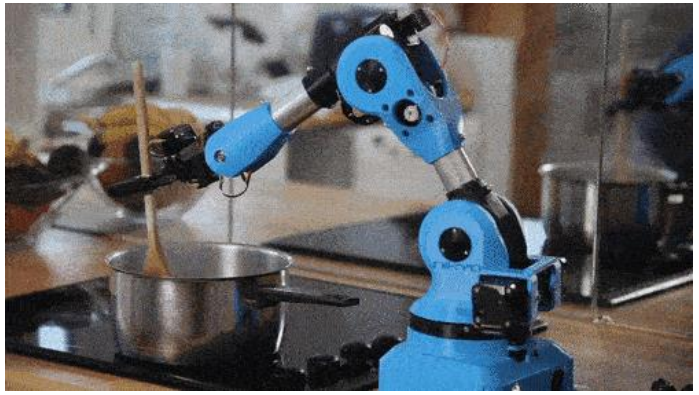




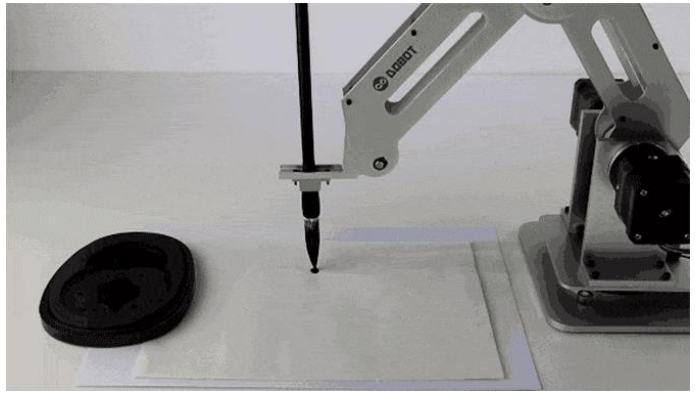
- 移动机器人
  - 关注如何规划路线以到达目的地
  - 无人驾驶汽车

操作机器人

- 目标：完成某种指定的动作，如抓取、焊接等。
- 关键问题：如何将机械臂的**末端（工作点）**放置到预设位置。
- 背景知识：运动学、动力学、控制论

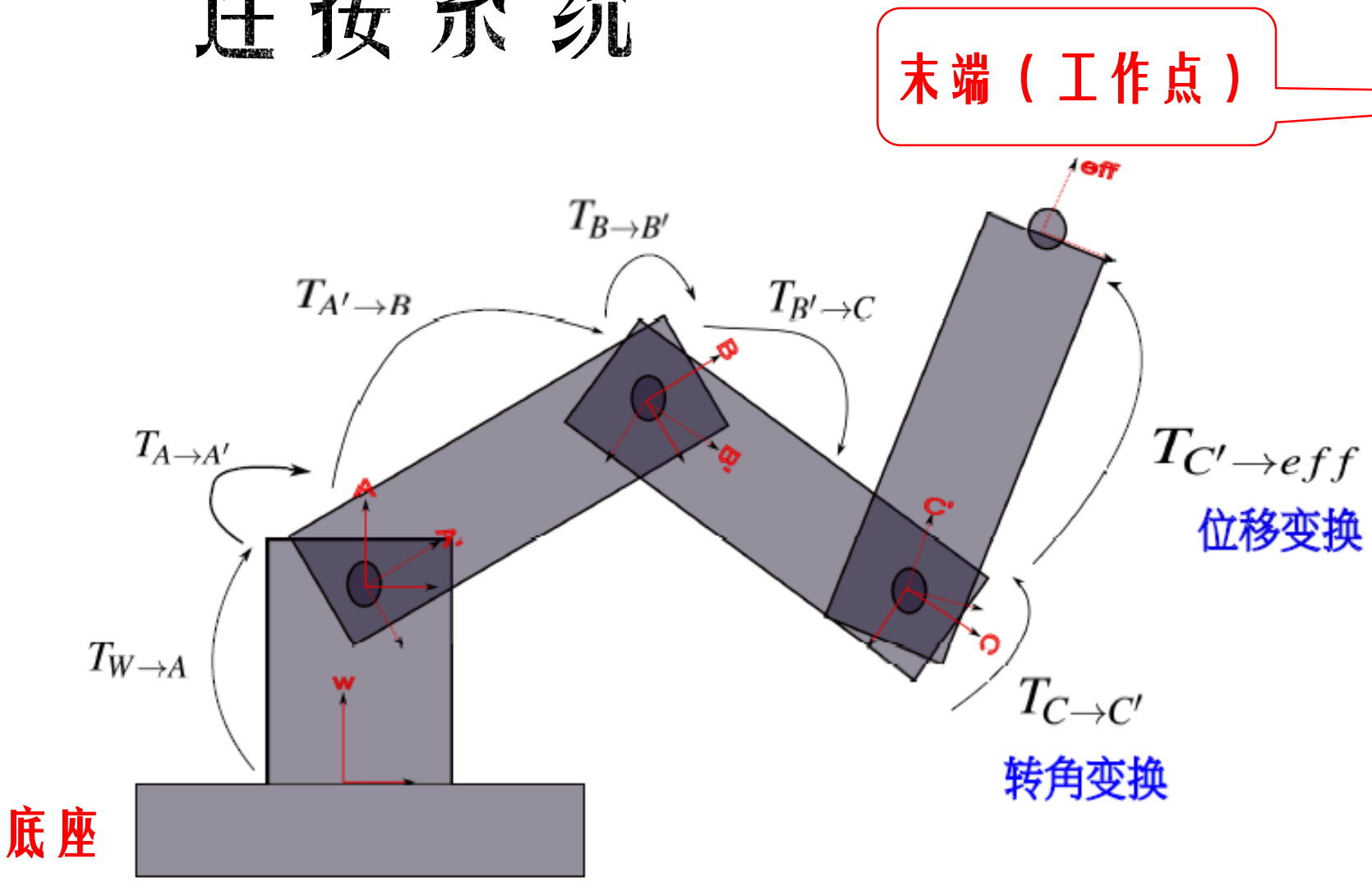


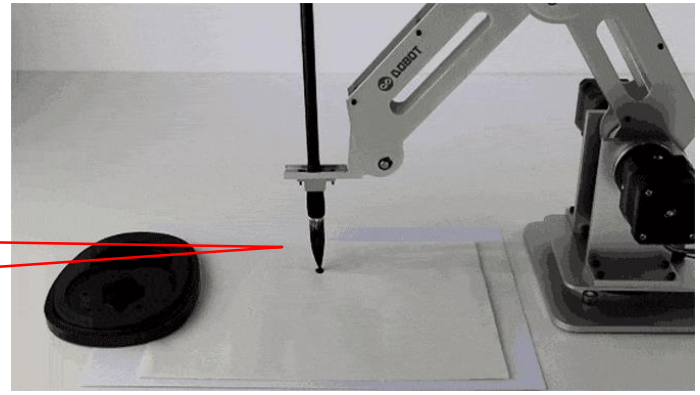
抓住汤勺，搅拌汤汁



抓住毛笔，撰写文字

连接系统





一、如何设置各连接杆的夹角，使末端放到指定位置?

二、为了实现1，如何对每个连接杆施力?

三、如何监控夹角的改变过程，实时调整施力?

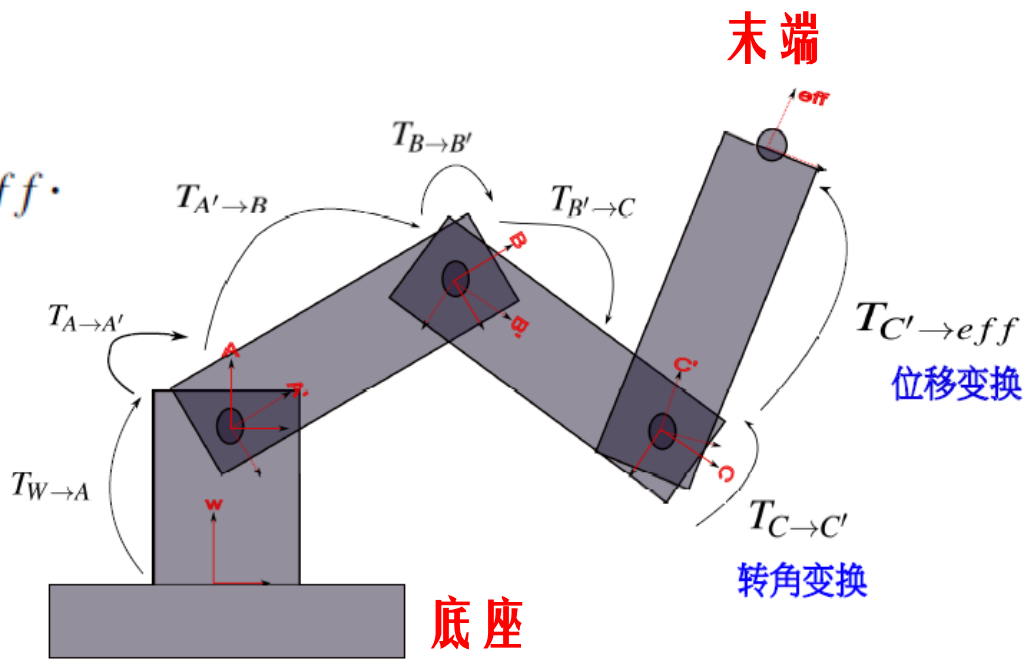
问题一：基于运动学计算目标夹角

- 位形
  - 所有夹角的一组取值称为一个位形
- 变换
  - 一个位形决定了底座到末端的变换

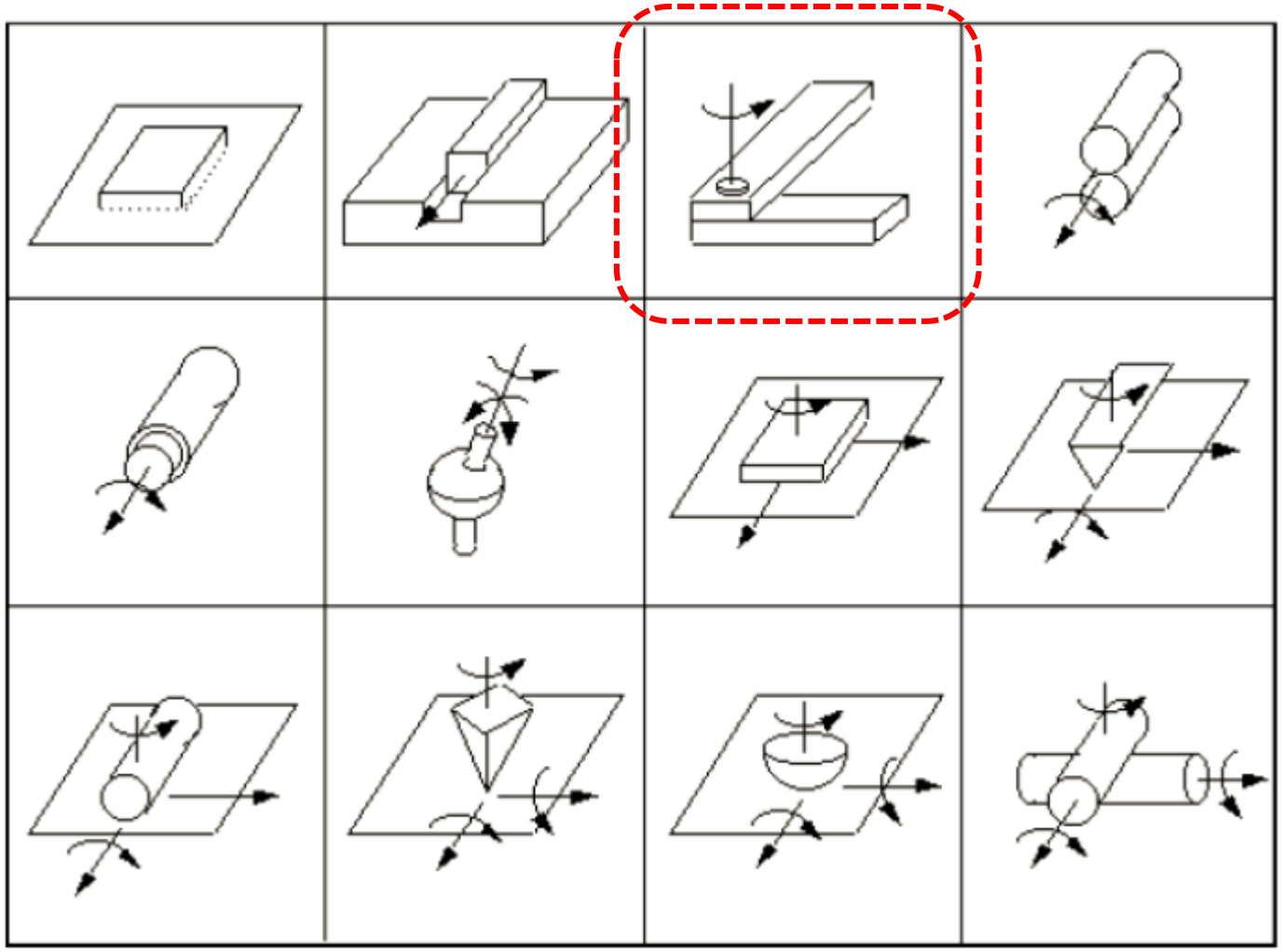
$$T_{W\to eff} = T_{W\to A}T_{A\to A'}T_{A'\to B}T_{B\to B'}T_{B'\to C}T_{C\to C'}T_{C'\to eff}.$$

  - 继而决定了末端位置

总结：计算一个位形，依此位形可以将末端置于目标位置。

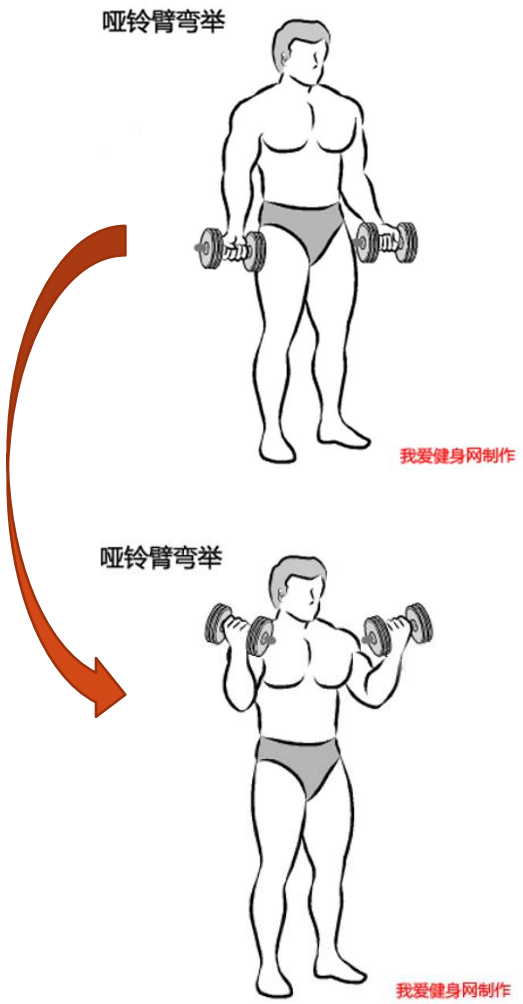


其它连接方式

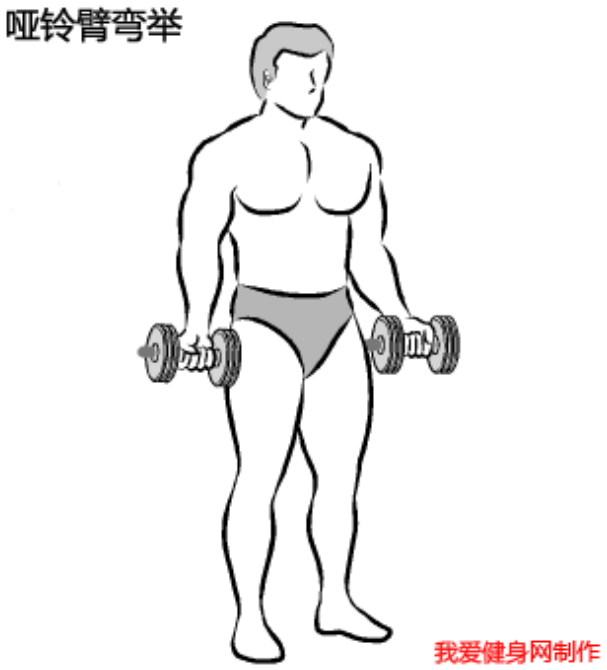


- 连接方式
  - 旋转连接
  - 嵌套连接
  - ....
  - 灵活组合
- 连接个数
  - 自由度

问题二：基于动力学设计力矩



- 理论基础：牛顿力学
- 处理步骤
  - 设计加速-减速（角加速度）方案，完成从起点到终点的变化；
  - 基于角加速度，计算每一时刻施于连接杆（手臂）上的力矩；
  - 基于连接杆（大臂和小臂）之间的连接关系，计算对每个连接杆所施加的力矩。



问题三：基于控制论进行过程调节

- 误差
  - 质量分布、重力、摩擦力等
  - 影响动作精度
- 反馈机制
  - 传感器监视动作完成情况
  - 记录T-1时刻的位置和速度，调整T时刻的力矩大小
- 应用举例
  - 压力传感器应用于擦玻璃机器人
  - 视觉传感器应用于焊接机器人

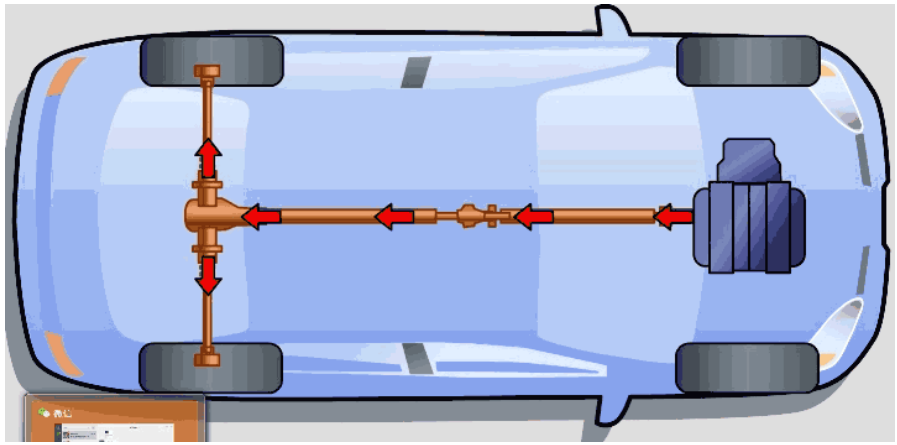
移动机器人

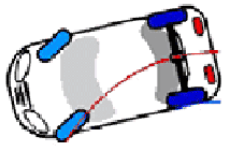
- 任务
  - 实现有效的移动
  - 对环境建立地图，并估计自身位置

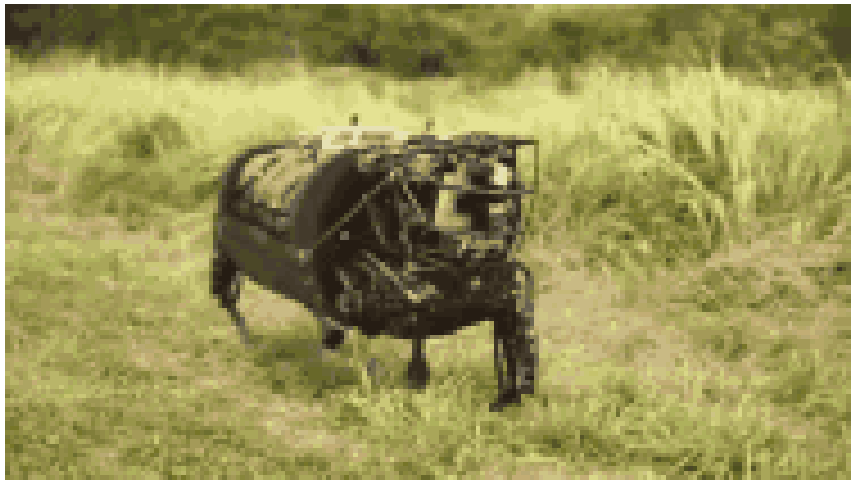
无人驾驶

智能扫地机器人

有效移动

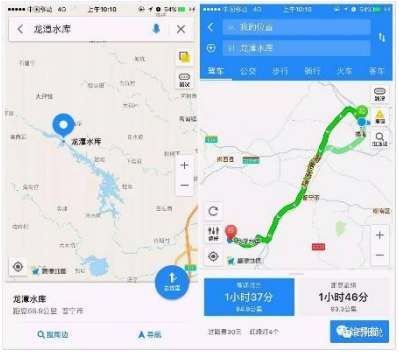




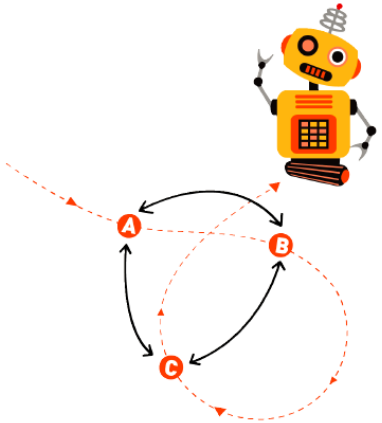


对环境建立地图，估计自身位置

- 方式一：人为预先指定机器人运动路线，并记录下所见的场景
  - 自动驾驶汽车



- 方式二：同步定位与地图构建
  - 家用扫地机器人



思考

- 回顾操作机器人的设计思路
- 移动机器人是如何实时学习地图的?
- 基于规则设计的机器人有何优缺点?

目　录

- 引言
- 现代机器人发展史
- 基于设计的机器人
- **基于学习的机器人**
- 深度强化学习方法

导学

- 行为机器人
  - 基于人为规则设计，可控、稳定；易解释、易预判
  - 不能满足复杂实际需求（难以设计、无法预测）
  - 让它干啥就干啥，复杂一点就白搭 😐
- 学习机器人
  - 行为方式来源于外界经验
  - 制定目标，无需设计动作，自主学习
  - **强化学习**

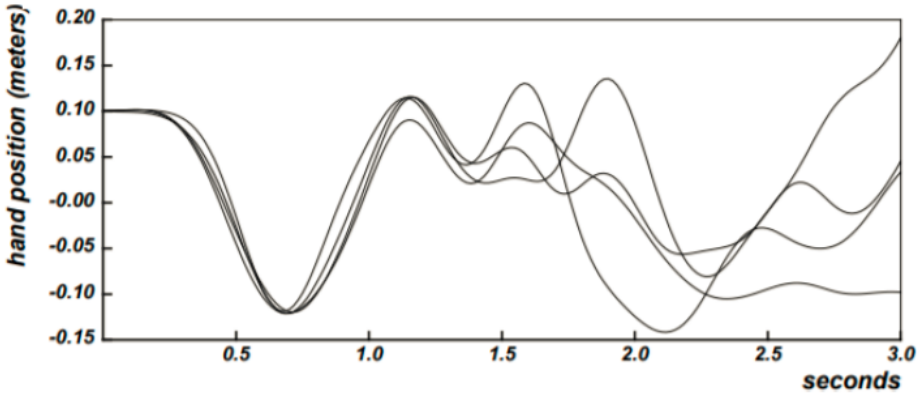
简单的例子：模仿学习

- 示教机器人
  - 人拉着机器人执行任务
  - 机器人重现过程
  - 非自主学习
  - 力矩需计算得到

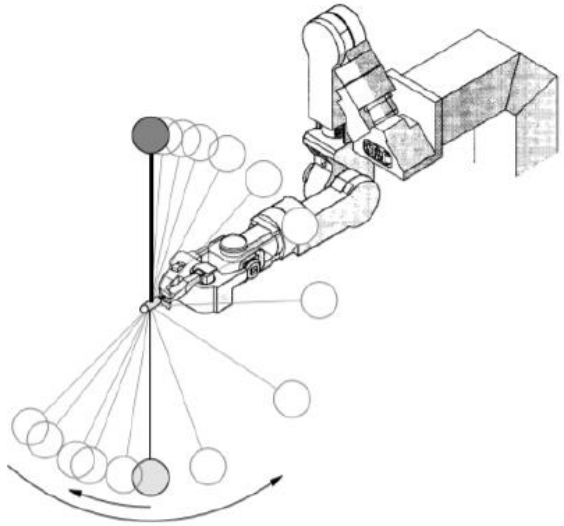
没有人拉着，机器人该怎么学习?

简单的例子：模仿学习

- 模仿学习机器人
  - 主动学习示教过程
  - 不断尝试，获得技巧
  - 力矩是学习得到



人绕小球的几个示范



机器模仿学习人绕小球的示范

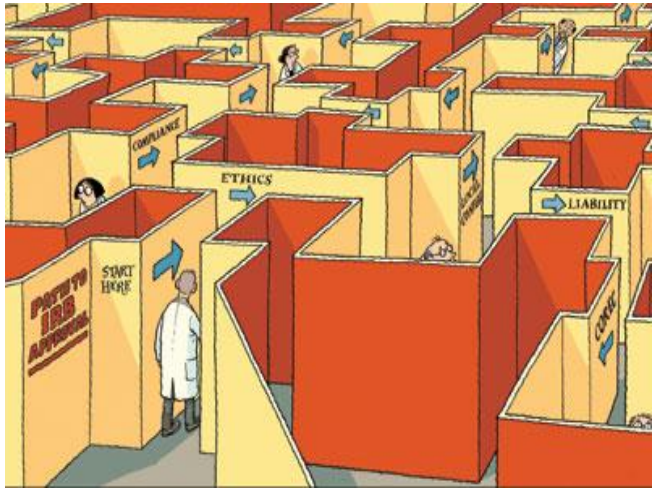
简单的例子：模仿学习

- 模仿学习机器人
  - 主动学习示教过程
  - 不断尝试，获得技巧
  - 力矩是学习得到

没有示范，机器人该怎么学习?

人是怎么学习的?

制定目标



不断尝试

婴儿学习走路

- 学会走路（目标）
- 不断地尝试迈步、抬腿（动作）
- 当偶尔站立或扶着挪动时，得到糖果（奖励）
- 当姿势错误摔倒了，感受到疼痛（惩罚）
- 父母会适当给出一些手势和引导（观察）
- 学会站立和迈步的技巧，不再摔倒

儿童学习骑脚踏车

- 从A顺利地骑到B（目标）
- 不断地尝试脚蹬车前行（动作）
- 平稳正确地前行（奖励）
- 摔倒、停滞不前（惩罚）
- 看其它小伙伴们的骑行（观察）
- 学会骑行的技巧，不再摔倒

学生学习知识

- 学会知识（目标）
- 不断地学习（动作）
- 学的好，分数高（奖励）
- 学不会，分数低（惩罚）
- 与老师同学交流（观察）
- 掌握高效学习知识的方法

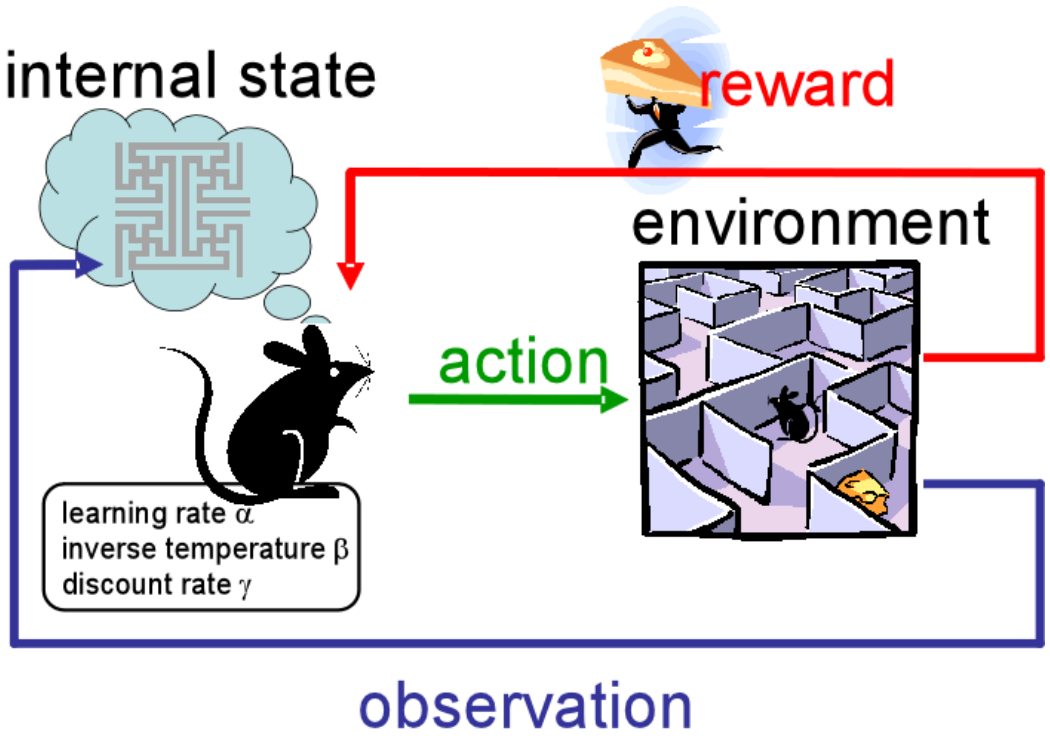
小结

- 人类学习是通过**目标、动作、奖励、惩罚、观察**等来实现的。
- 这类学习方法通过主动和环境进行**交互**，利用环境给出的**反馈**（奖励或惩罚）学习**行为策略**，使得最终**收益**最大化，称之为**强化学习**。
- 强化学习是人类的学习方式，将其应用至机器人学习中。

强化学习的概念

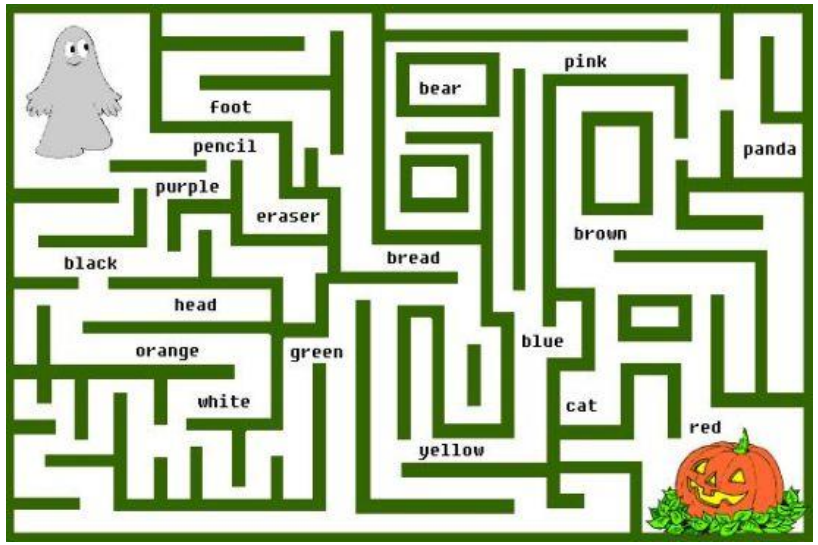
- 强化学习是指在一系列**环境**中，通过多步恰当的**决策**来达到一个**目标**的学习过程，是一种序列多步决策的问题。

- 强化学习的目标就是要寻找一个能使得我们获得最大累积**收益**的**策略**。



强化学习的分类

- **多轮任务**：成本低、时间短，收益最后显现
  - 象棋、围棋等游戏

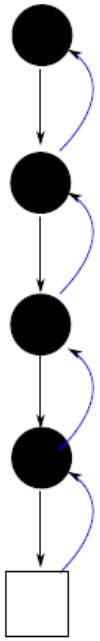


- **连续任务**：成本高、时间长，奖惩随时可见
  - 股票操作，行走机器人

强化学习的方法

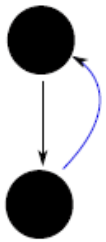
- **蒙特卡洛（MC）算法：**完整模拟一次任务执行过程，通过任务结束后的收益调整行为策略。（用于多轮任务）

一系列行为（全路径）



- **时序差分（TD）算法：**基于当前行为策略生成一个动作，这个动作通常会立即得到一个反馈，基于这一反馈对行为策略进行即时调整。（用于连续任务）

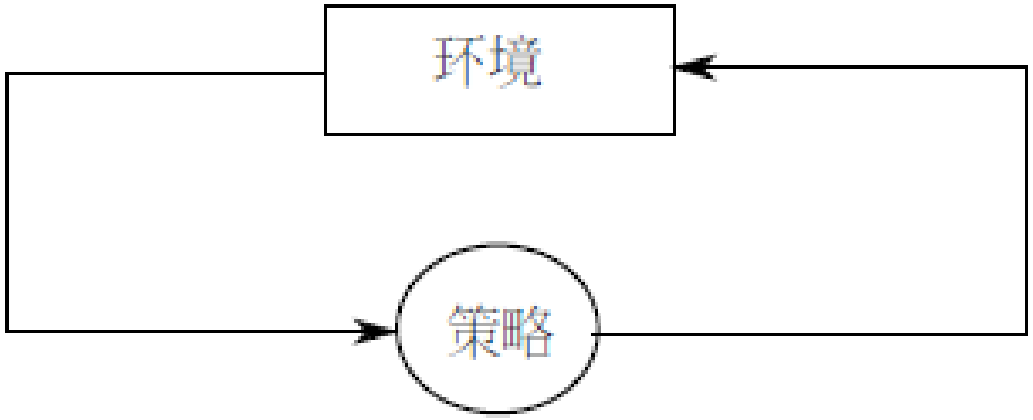
一步或多步行为



强化学习的学习对象

1、行为策略 — 策略学习

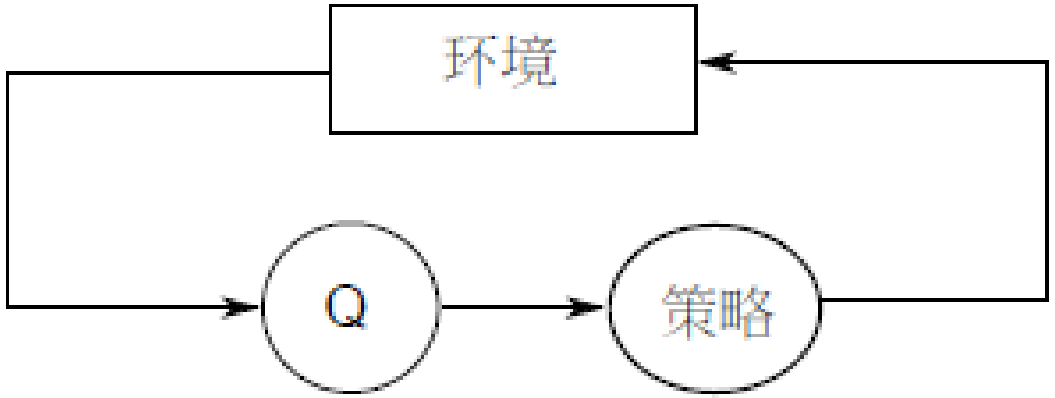
- 不考虑环境动态性，直接对策略函数学习 $\pi(s)$
- 在学习时，通过对 $\pi(s)$ 的参数进行调整，使得 $\pi(s)$ 与环境交互所产生的总体收益最大化。



强化学习的学习对象

2、价值函数 — 值函数学习

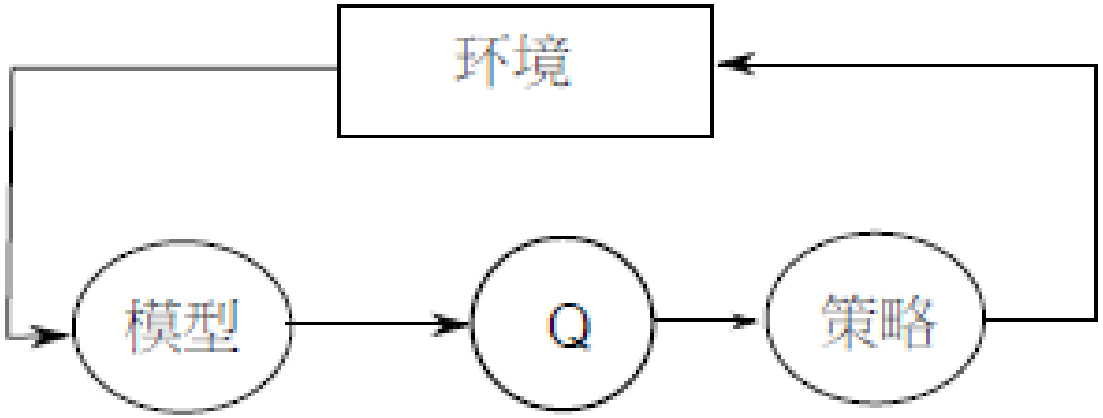
- 不考虑环境环境动态性
- 学习某一状态下采取某一动作的长远收益 $Q(s,a)$，基于此导出行为策略 $\pi(s)$



强化学习的学习对象

3、环境模型 — 模型学习

- 首先学习环境反馈 $r(s,a)$ 和状态跳转 $T(s,a)$ 规律
- 然后计算动作值函数 $Q(s,a)$，依此生成行为策略 $\pi(s)$



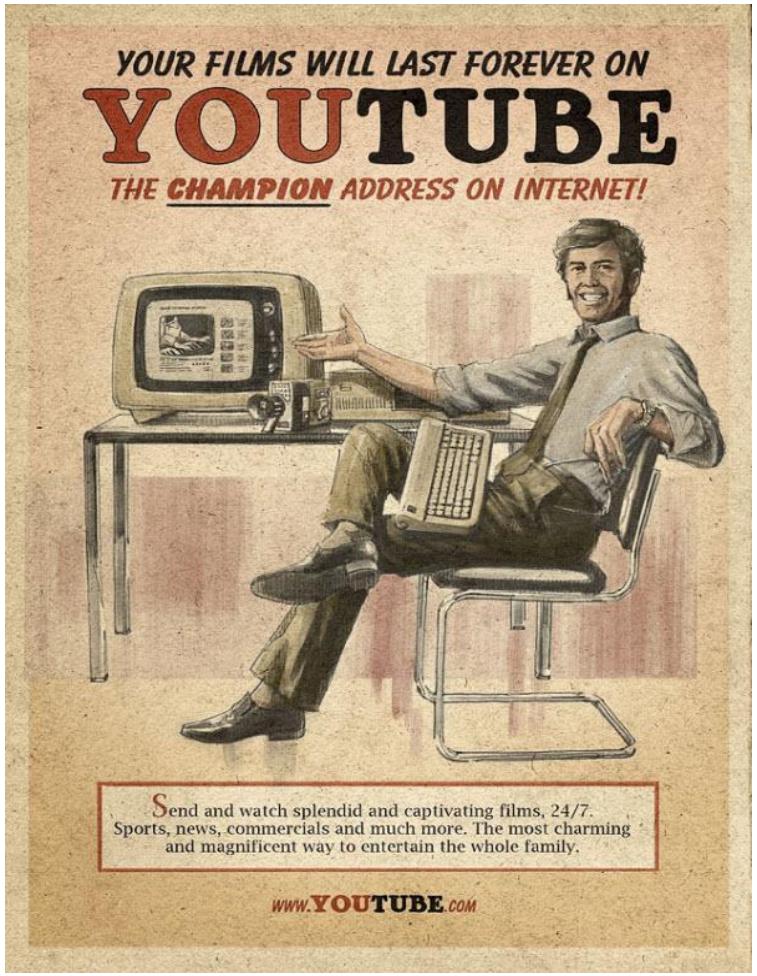
强化学习的应用（1）

机器人：自主学习如何行动，如何完成制定目标

强化学习的应用（2）

生产与生活

- 如何摆放食品销量最大?
- 如何投放广告吸引更多客户?



强化学习的应用（3）

电子游戏

思考

- 强化学习还可以用于哪些场景中?
- 假如不是强化学习，婴儿该如何学走路呢?
- 除了强化学习，人类还有哪些学习方法?

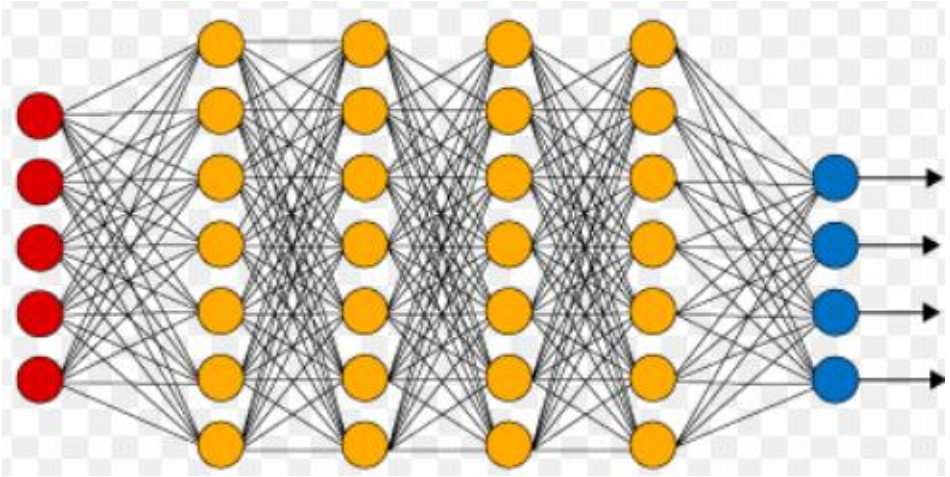
目　录

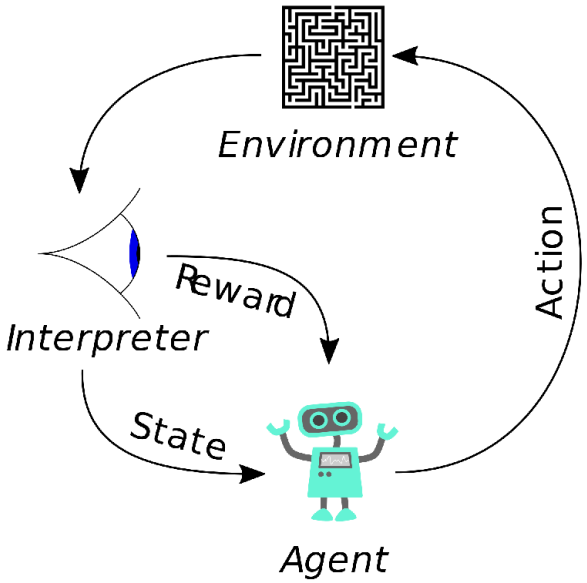
- 引言
- 现代机器人发展史
- 基于设计的机器人
- 基于学习的机器人
- **深度强化学习方法**

环境状态

- 对环境状态 s 的判断直接影响到最终的学习效果
- 对于简单搜索任务，s 是离散的、确定的
- 对于实际任务，s 是连续的、原始的
- 需要寻找更为复杂的函数近似方法，实现从原始复杂的环境输入中提取状态信息

什么是深度强化学习

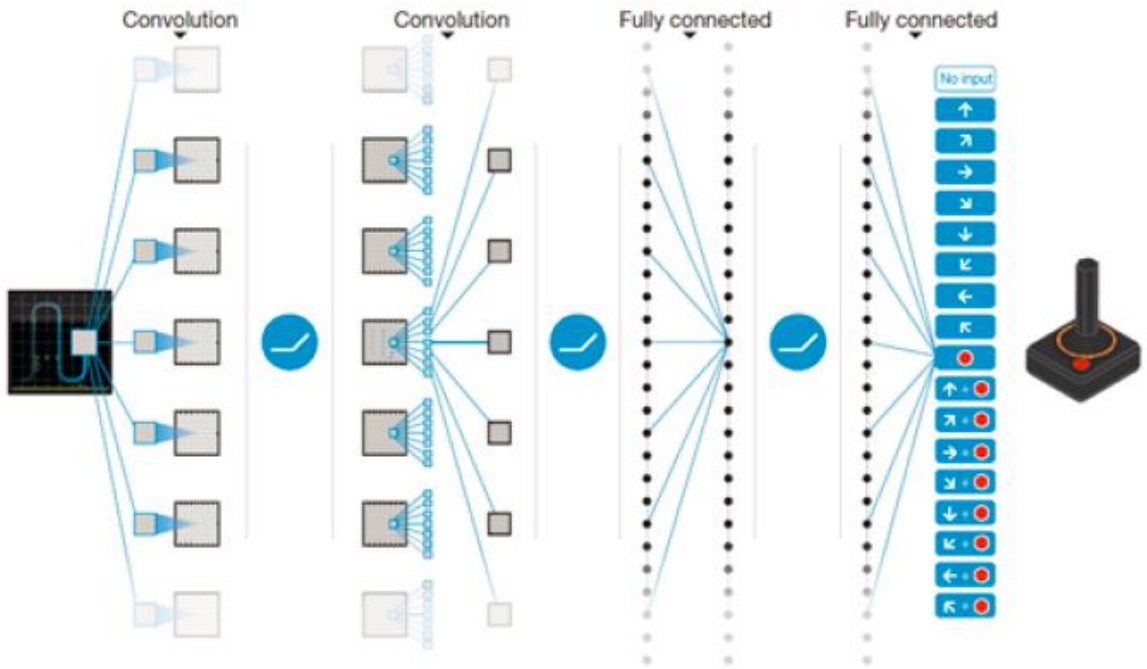




- **第一部分：深度学习**，将原始感知信号逐层处理，提取环境状态。
- **第二部分：强化学习**，利用强化学习目标更新深度神经网络参数。
- 深度强化学习是当前机器学习乃至整个人工智能领域研究的热点之一，推动了近年来人工智能的热潮。

Atari 游戏

- Deep Mind 利用深度 Q-learning 网络（DQN）教会机器玩游戏
- 输入为游戏画面，经过多层卷积神经网络（CNN）来提取状态信息，得到状态空间；
- 基于该状态空间，学习不同状态下采取某一操作的价值 $Q(s,a)$。

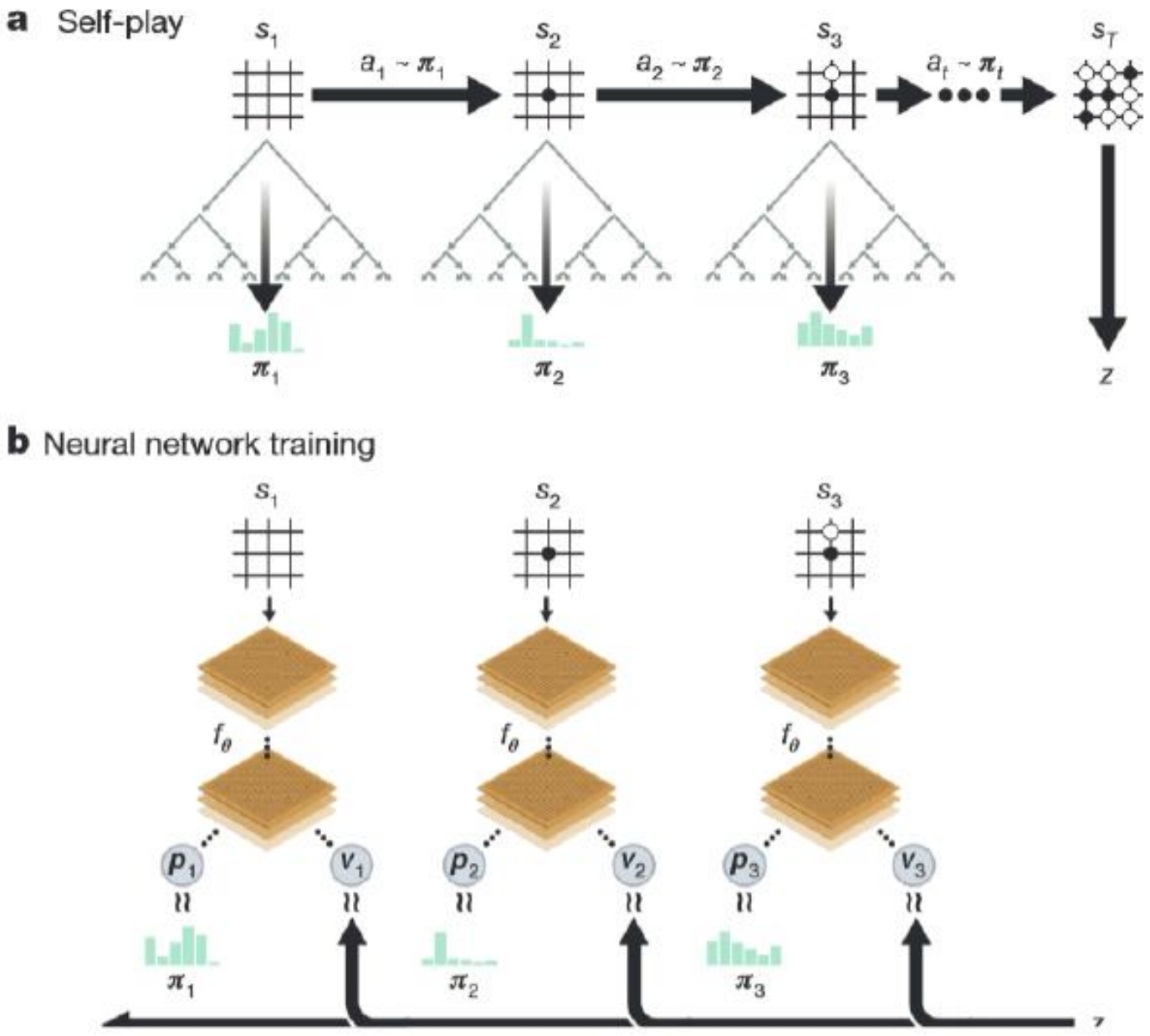


AlphaGo Zero

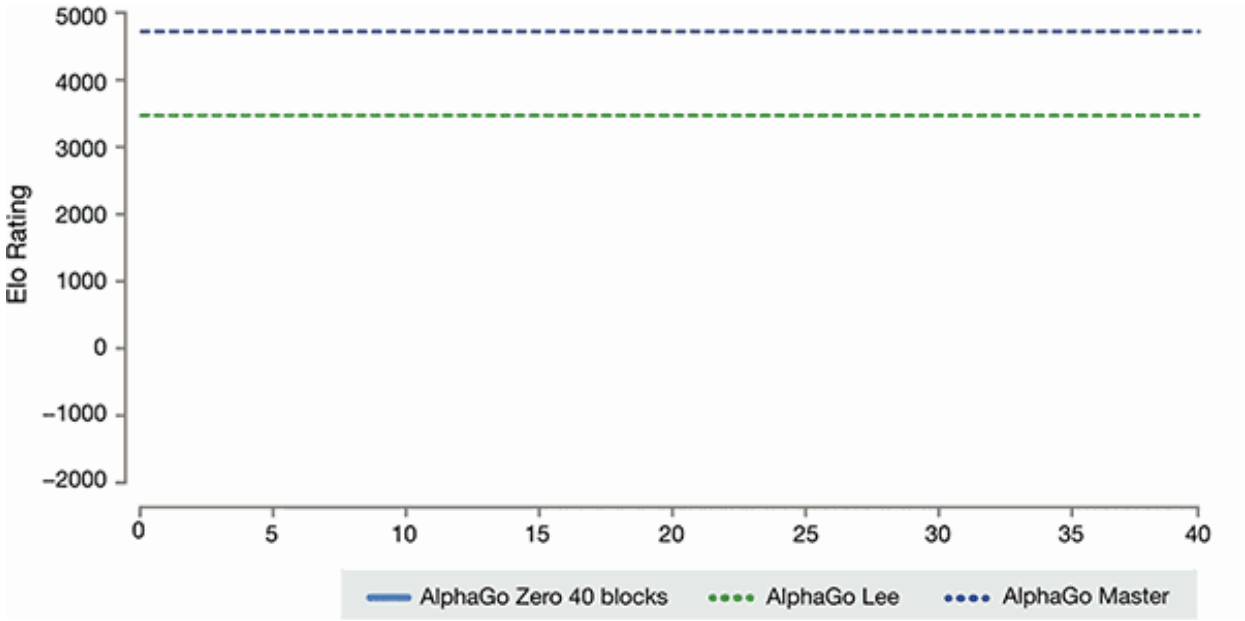
- 不再学习人类的棋谱，而是完全依赖深度强化学习，通过自我对奕学习机器自己的走棋方式。
- 利用深度CNN网络和蒙特卡洛树（MCT）计算出（1）模拟棋盘下的走棋概率；（2）当前模拟棋盘的胜率。
- 训练目标是使CNN对走棋概率的预测与MCT的预测尽可能接近，同时使CNN输出的胜率与自我对奕的结果一致。
- 多轮自我对弈，CNN和MCT的预测能力同步增强，成为无敌的国手！

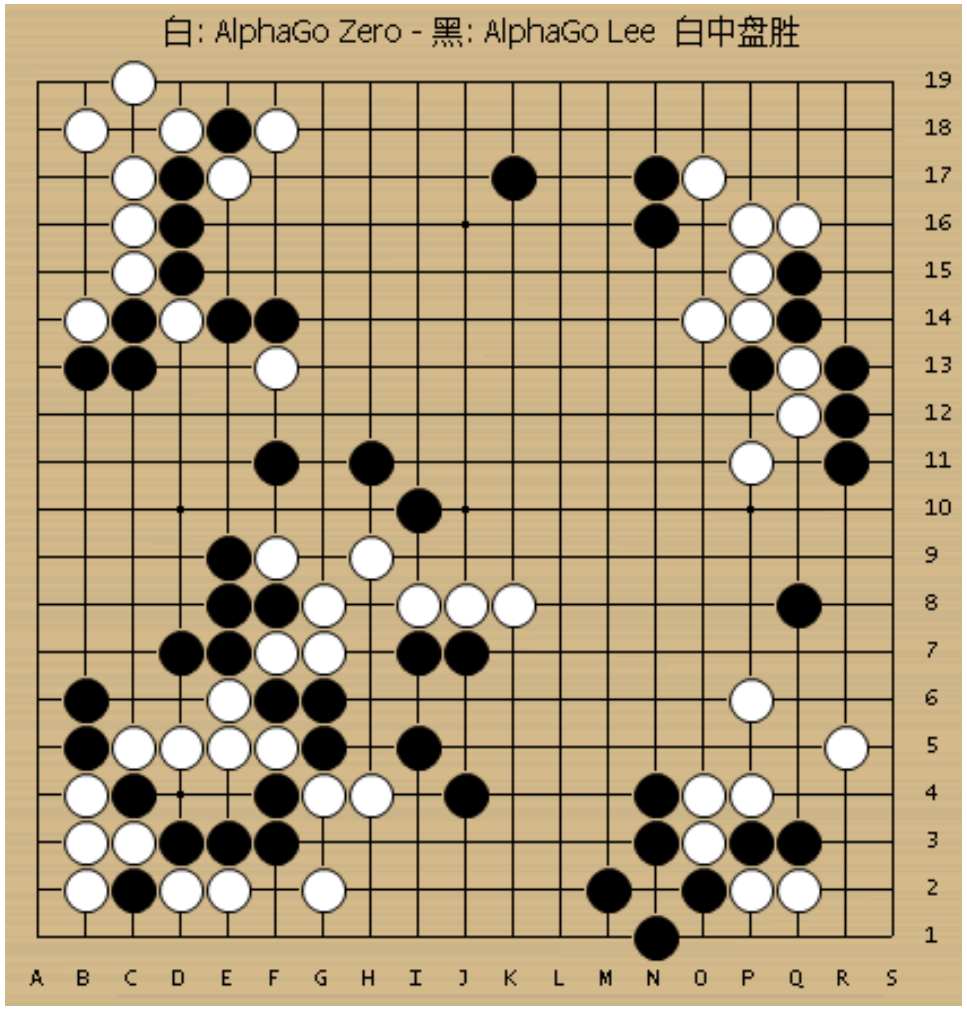
AlphaGo Zero

- 使用 64 个 GPU 和 19 个 CPU，AlphaGo Zero 用三天时间完成了自我对弈 490 万局。
- 几天之内它就发展出击败人类顶尖棋手的技能。



AlphaGo Zero

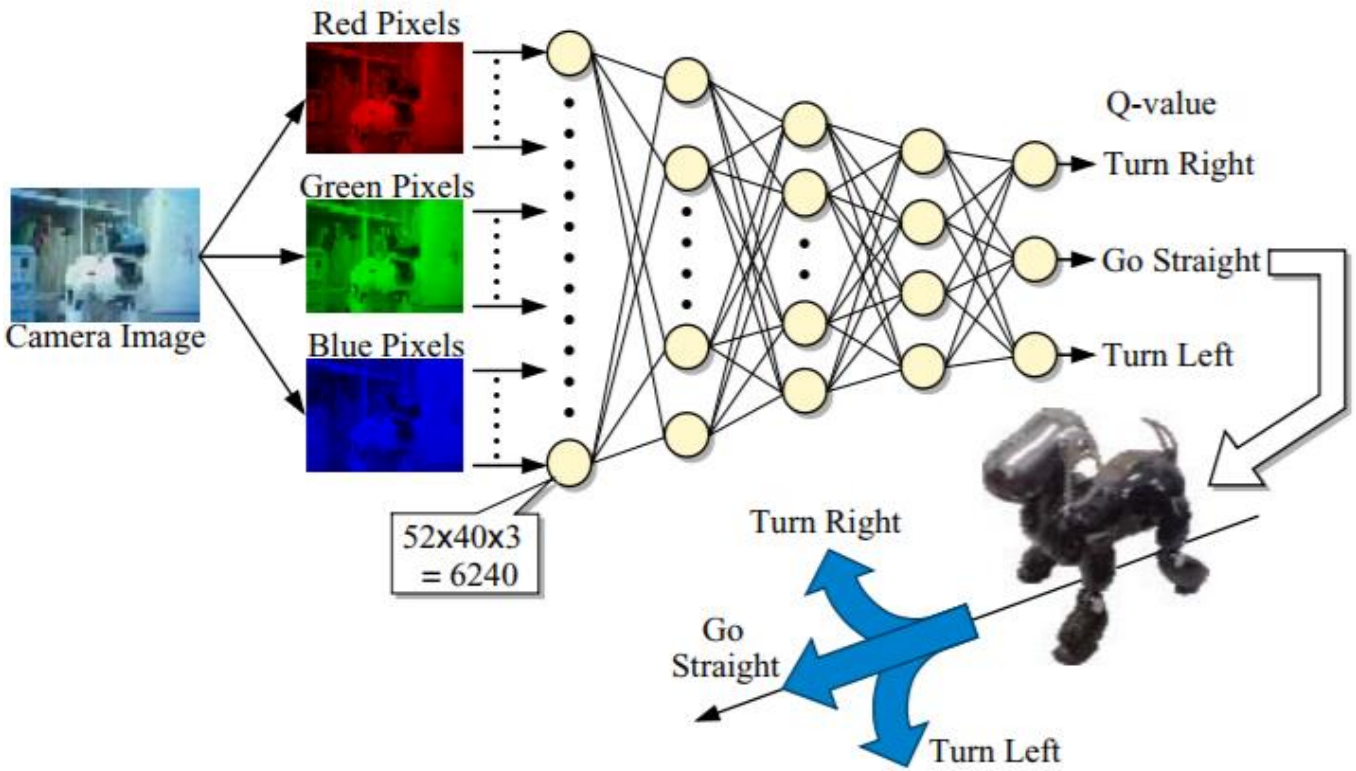




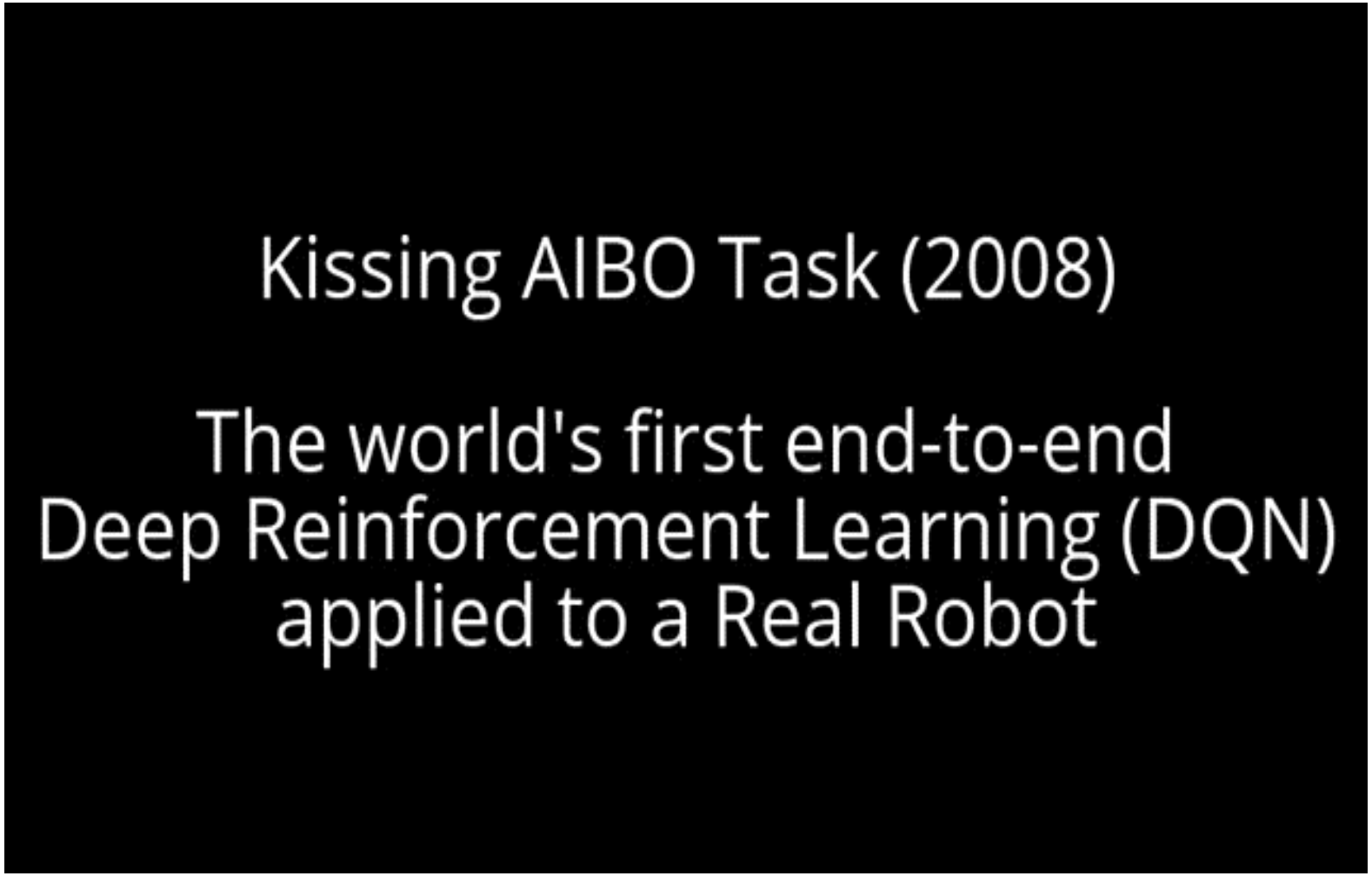
实体机器人：AIBO 接吻机器狗

初始状态

目标状态



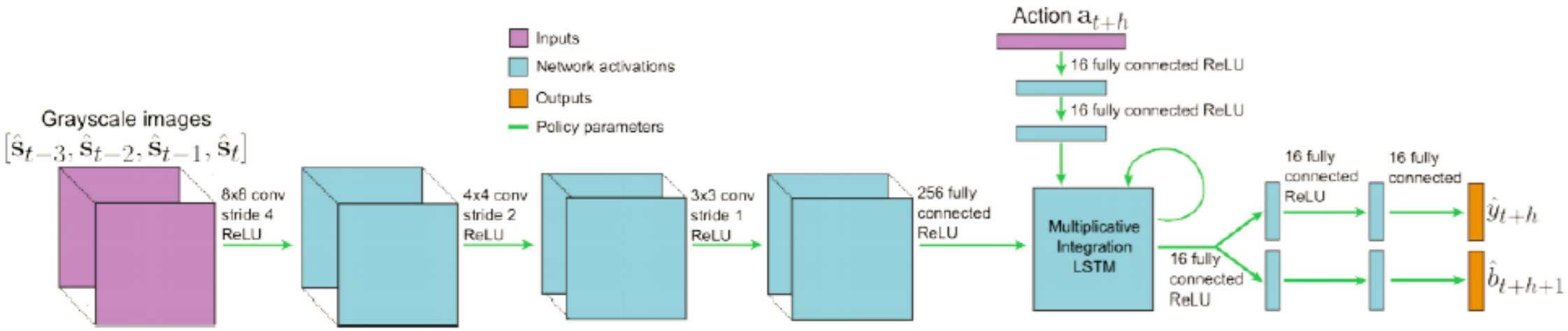
实体机器人：AIBO 接吻机器狗



2008 年日本 Oita 大学

实体机器人：自动驾驶小车

- 输入包括（1）摄像头在一段时间内拍摄的图像；（2）采取的动作
- 输出为该动作可能产生的后果，包括发生碰撞的概率和系统的安全状态。



实体机器人：自动驾驶小车

Self-supervised Deep Reinforcement Learning with Generalized Computation Graphs for Robot Navigation

Gregory Kahn, Adam Villaflor, Bosen Ding, Pieter Abbeel, Sergey Levine

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

2017年加州伯克利

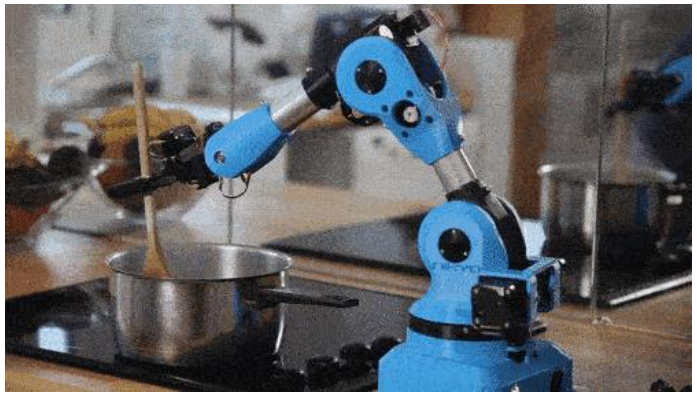
两种机器人的比较

- 基于设计的机器人

安全可靠

处理复杂任务困难

实体机器人

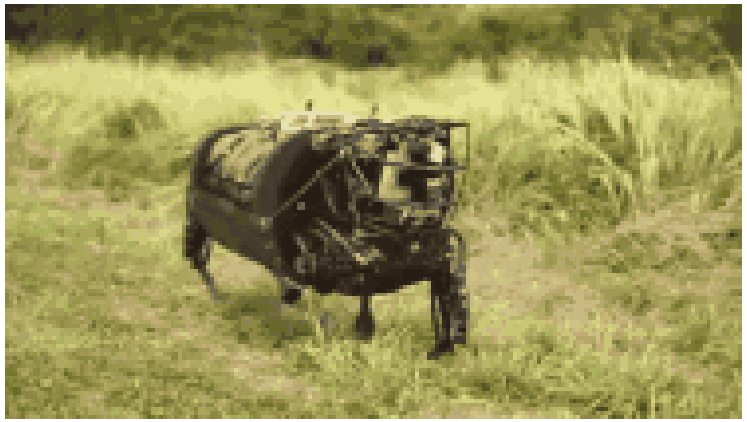


- 基于学习的机器人

可适应复杂环境

需要大量尝试

虚拟机器人



思考

- 回顾深度强化学习的特点
- 展望未来的机器人，能做什么？如何实现？

The end !