

# 1. Introduction 导言

---

WU Xiaokun 吴晓堃

xkun.wu [at] gmail

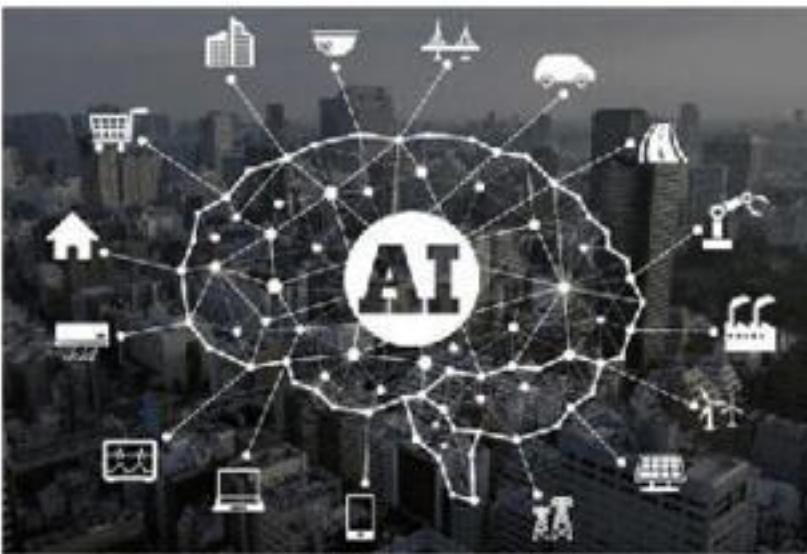
2023/02/21

# Contents

- AI：科幻与现实
- 学习、机器学习
- 机器学习基本流程
- 为什么是深度学习？
- 为什么是现在？

# AI：科幻与现实

# AI无所不在



# 底特律：化身为人



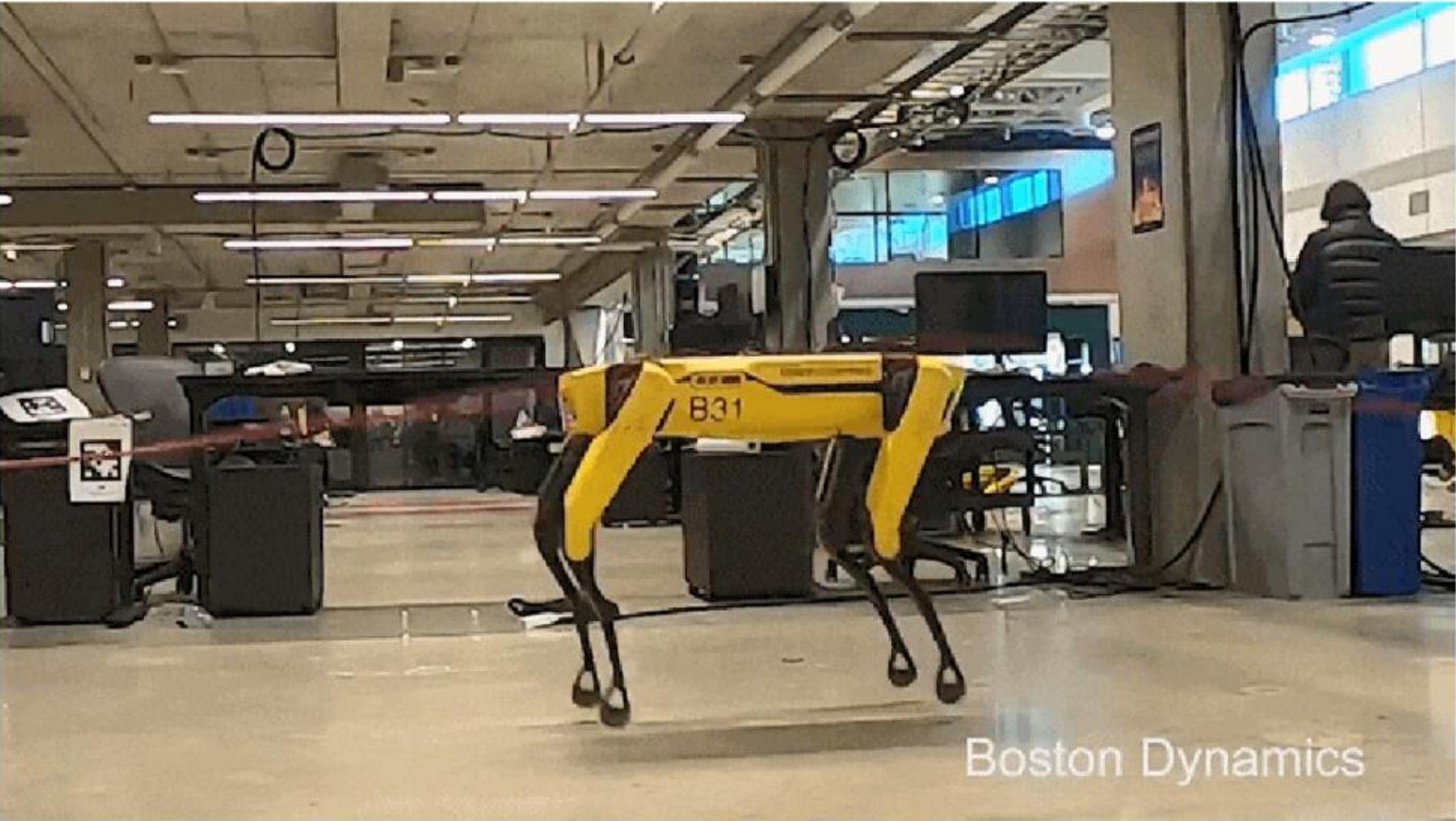
# BostonDynamics



# BostonDynamics - 拾取



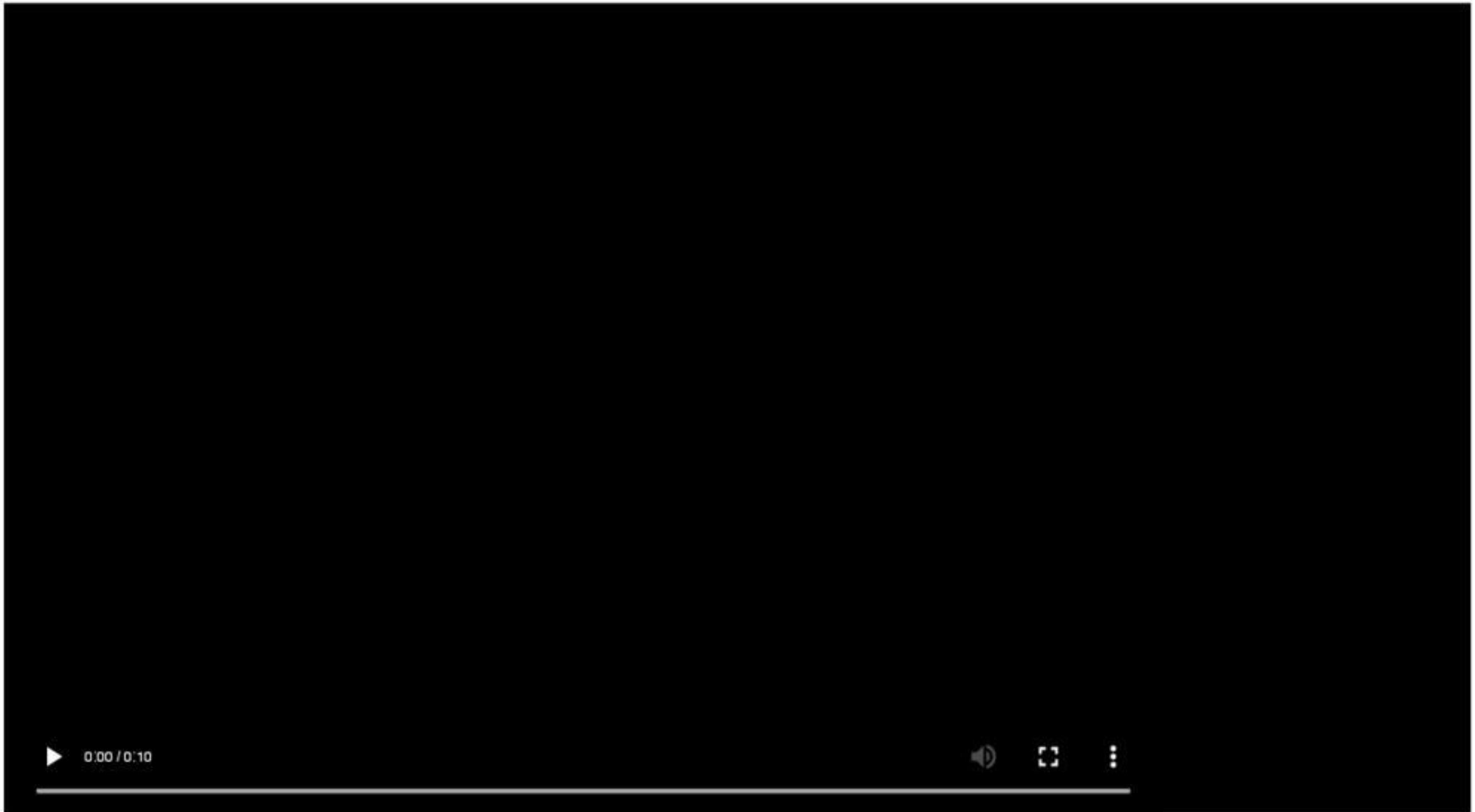
# BostonDynamics - 跳绳



# BostonDynamics - 后空翻



# BostonDynamics - 跑酷



# 课前调研

1. 你学习此课程的动机是什么？
2. 你期望从此课程中学到什么？
3. 你考虑从事与此课程相关的工作吗？具体是哪些方面？
4. 你有良好的学习习惯吗？（课前预习、课后复习等）
5. 你有与任课教师讨论专业问题的习惯吗？
6. 你有自学的习惯吗？（大量阅读相关文献、查看专业视频，练习编程等）
7. 你有独立完成作业的习惯吗？

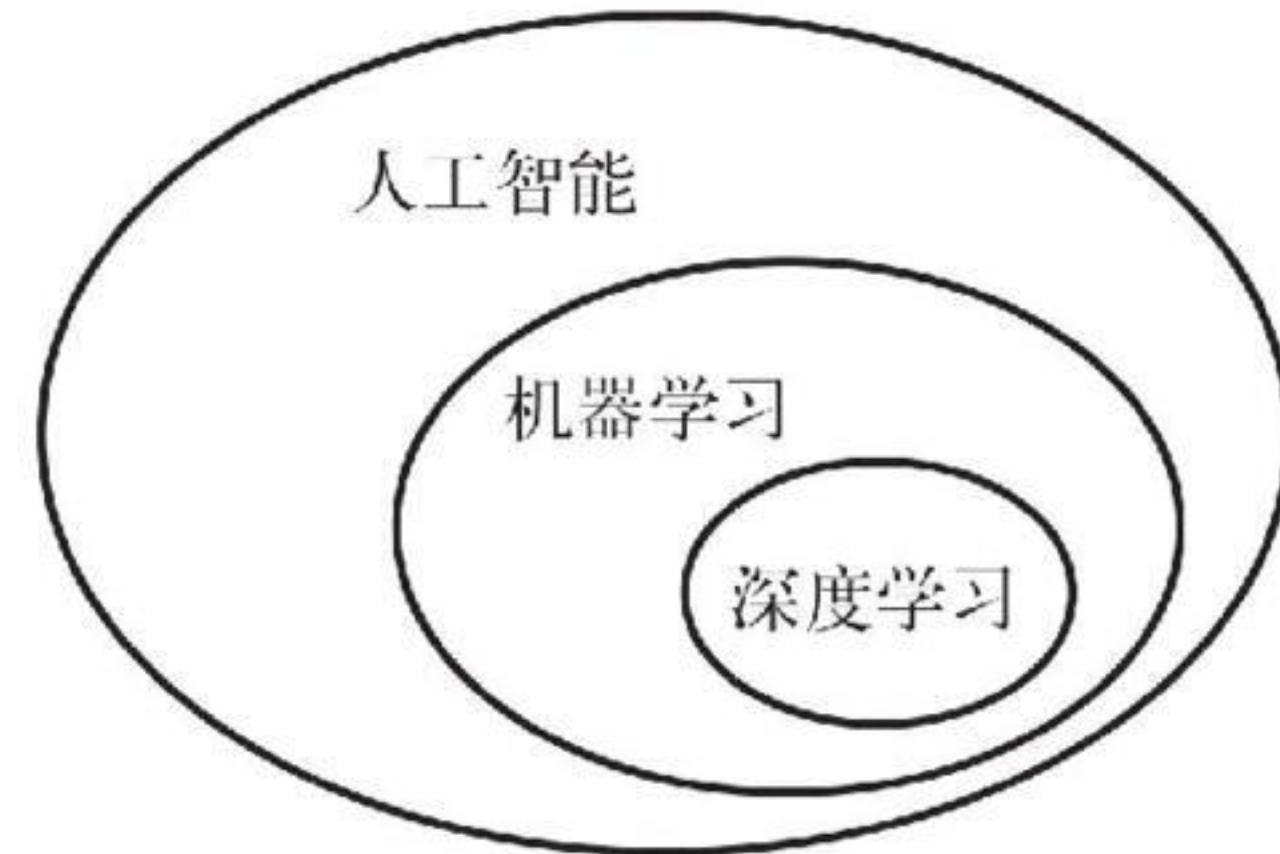
# 学习、机器学习

# 人工智能、机器学习、深度学习

人工智能：具备人类智能的机器

机器学习：经验、模型、预测

深度学习：多层的机器学习方法



# 人工智能

人工智能诞生于20世纪50年代。

## 人工智能的简洁定义

将通常由人类完成的智力任务自动化。

# 人工智能

人工智能诞生于20世纪50年代。

## 人工智能的简洁定义

将通常由人类完成的智力任务自动化。

在相当长的时间内，许多专家相信，只要程序员精心编写足够多的明确规则来处理知识，就可以实现与人类水平相当的人工智能。

- 符号主义人工智能 **symbolic AI**: 从 20 世纪 50 年代到 80 年代末
- 专家系统 **expert system**: 20 世纪 80 年代

# 机器学习与Turing

人工智能先驱Alan Turing在 1950 年发表论文《计算机器和智能》。

## Turing测试

测试者与被测试者（一个人和一台机器）隔开的情况下，通过一些装置（如键盘）向被测试者随意提问。进行多次测试后，如果机器让平均每个参与者做出超过30%的误判，那么这台机器就通过了测试，并被认为具有人类智能。

# 机器学习与Turing

人工智能先驱Alan Turing在 1950 年发表论文《计算机器和智能》。

## Turing测试

测试者与被测试者（一个人和一台机器）隔开的情况下，通过一些装置（如键盘）向被测试者随意提问。进行多次测试后，如果机器让平均每个参与者做出超过30%的误判，那么这台机器就通过了测试，并被认为具有人类智能。

机器学习的概念来自于Turing引述 Ada 的问题后的进一步思考：

## 通用计算机是否能够学习与创新？

除了“我们命令它做的任何事情”之外，它能否自我学习执行特定任务的方法？

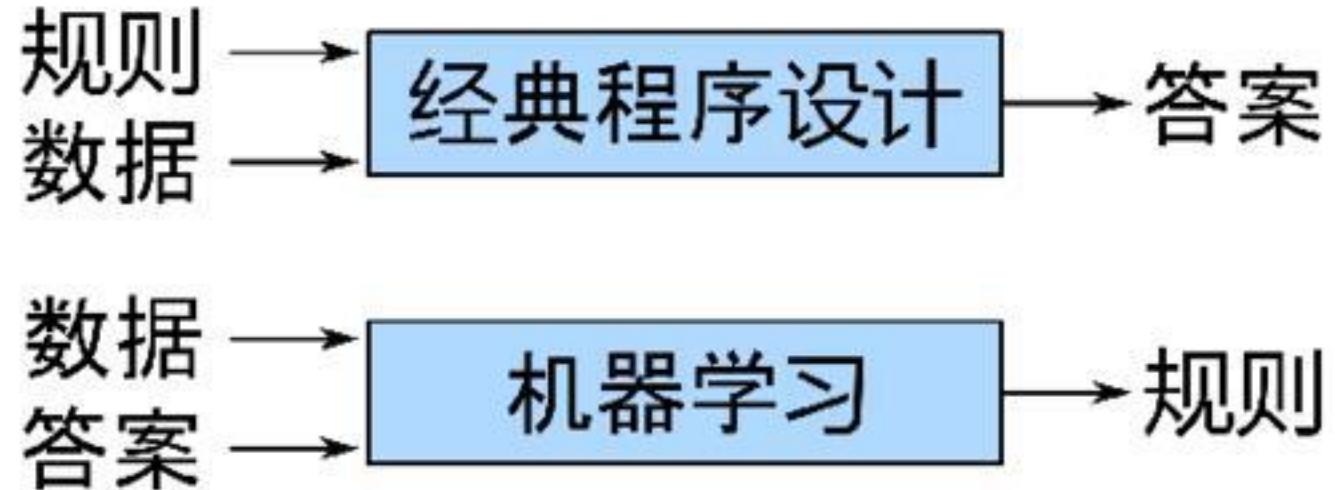
如果没有程序员精心编写的数据处理规则，计算机能否通过观察数据自动学会这些规则？

# 新的编程范式

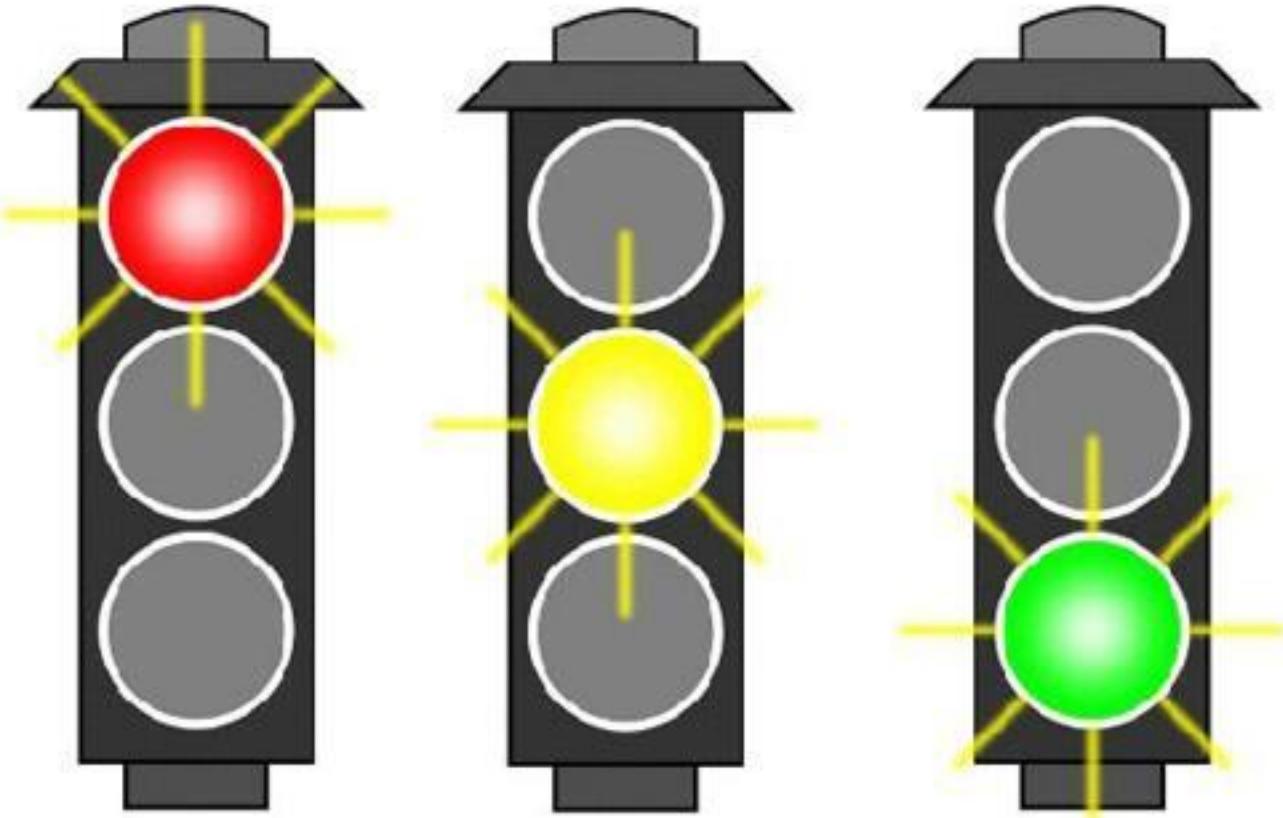


注意：机器学习系统是训练出来的，而不是明确地用程序编写出来的。

# 新的编程范式：举例



注意：机器学习系统是训练出来的，而不是明确地用程序编写出来的。



舉例：原始人穿越到现在，如何学会交通规则？观察与试错。

# 机器学习算法三要素

给定包含预期结果的示例，机器学习将会发现执行一项数据处理任务的规则。

- 输入数据的可计算形式。
- 预期输出的示例。
- 衡量算法效果好坏的方法。

# 机器学习算法三要素

给定包含预期结果的示例，机器学习将会发现执行一项数据处理任务的规则。

- 输入数据的可计算形式。
- 预期输出的示例。
- 衡量算法效果好坏的方法。

## “学习”的简单理解

用反馈信号衡量结果、调节算法的工作方式。这个调节过程就是我们所说的学习。

# 机器学习算法三要素

给定包含预期结果的示例，机器学习将会发现执行一项数据处理任务的规则。

- 输入数据的可计算形式。
- 预期输出的示例。
- 衡量算法效果好坏的方法。

## “学习”的简单理解

用反馈信号衡量结果、调节算法的工作方式。这个调节过程就是我们所说的学习。

思考：人是如何学习的？

- 人能理解其他人的思维过程吗？

# 机器学习算法的核心问题

机器学习模型将输入数据变换为有意义的输出

- 从已知的输入和输出示例中进行“学习”，将经验应用到未知数据。

# 机器学习算法的核心问题

机器学习模型将输入数据变换为有意义的输出

- 从已知的输入和输出示例中进行“学习”，将经验应用到未知数据。

机器学习和深度学习的核心问题在于有意义地变换数据

- 输入数据的有效表示 **representation** 及变换方法

# 机器学习算法的核心问题

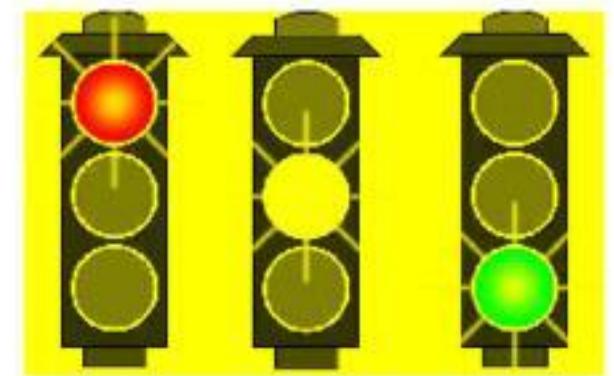
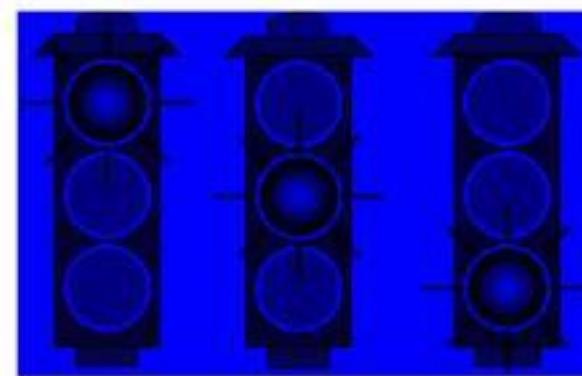
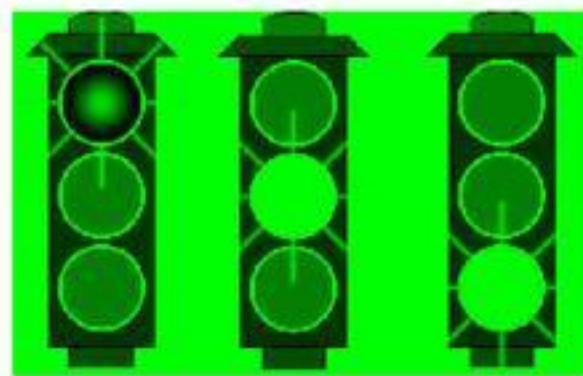
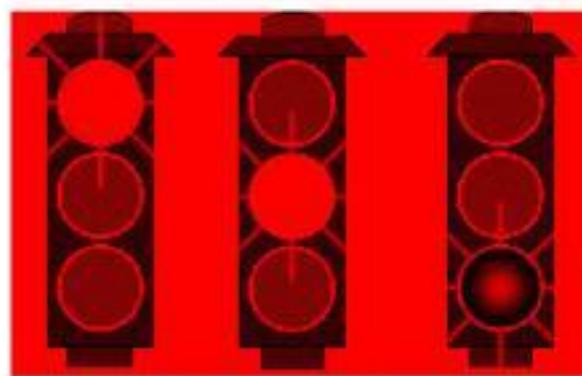
机器学习模型将输入数据变换为有意义的输出

- 从已知的输入和输出示例中进行“学习”，将经验应用到未知数据。

机器学习和深度学习的核心问题在于有意义地变换数据

- 输入数据的有效表示 **representation** 及变换方法

举例：信号灯图像的颜色响应 **activation**



# 机器学习 - 学习与假设空间

## 学习

寻找更好数据表示的自动搜索过程。

寻找更好数据变换的自动搜索过程。

# 机器学习 - 学习与假设空间

## 学习

寻找更好数据表示的自动搜索过程。

寻找更好数据变换的自动搜索过程。

## 假设空间

机器学习算法在解决问题时通常没有什么创造性，而仅仅是遍历一组预先定义好的操作集合，这个操作集合叫作假设空间 **hypothesis space**。

# 机器学习 - 学习与假设空间

## 学习

寻找更好数据表示的自动搜索过程。

寻找更好数据变换的自动搜索过程。

## 假设空间

机器学习算法在解决问题时通常没有什么创造性，而仅仅是遍历一组预先定义好的操作集合，这个操作集合叫作假设空间 **hypothesis space**。

## 机器学习 - 技术定义

在预先定义好的可能性空间中，利用反馈信号的指引来寻找输入数据的有效表示。

# 机器学到了什么？

- 实践派：什么都没有，“黑盒”
- 理论派：可能有，但我们还不理解
  - 人也很难理解其他人（甚至自己）的思维过程
  - 也许机器理解其工作原理，但表达不出来

# 机器学到了什么？

- 实践派：什么都没有，“黑盒”
- 理论派：可能有，但我们还不理解
  - 人也很难理解其他人（甚至自己）的思维过程
  - 也许机器理解其工作原理，但表达不出来

(人们希望) 机器学习是对人脑这个黑盒函数进行模拟。

*Machine learned almost nothing - instead, humans (researchers) learned a lot during the fine-tuning.*

*– Manning*

# 深度学习之“深度”

## 深度学习

从数据中学习表示的新方法：强调从一系列称作层 layer的中间模块中逐步学习，不断产生越来越有意义的表示。

# 深度学习之“深度”

## 深度学习

从数据中学习表示的新方法：强调从一系列称作层 layer的中间模块中逐步学习，不断产生越来越有意义的表示。

## 注意

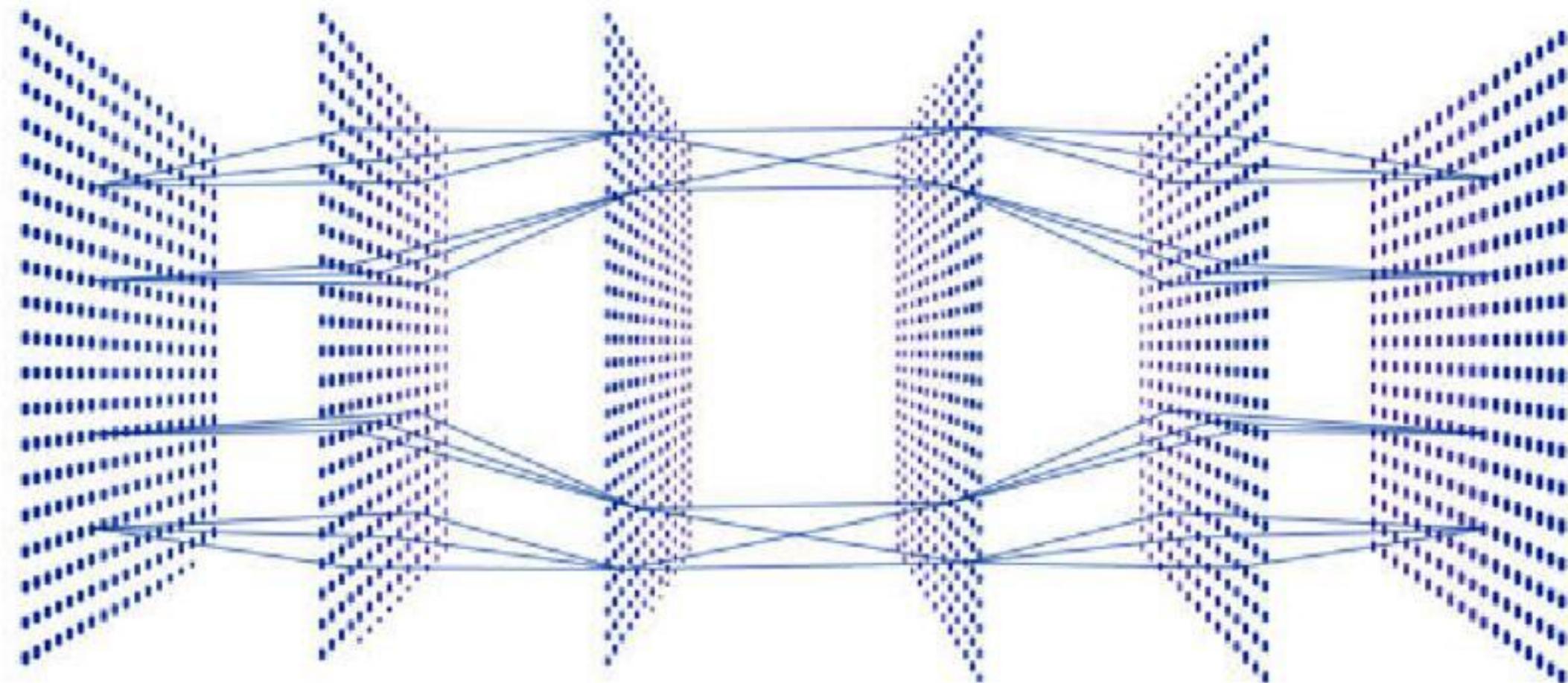
“深度”并不是说机器能够获取的“更深层次的理解”，而仅指一系列顺序表示的层。

就像展开纸团需要多个步骤一样。

# 深度学习 - 层、深度

## 层、深度

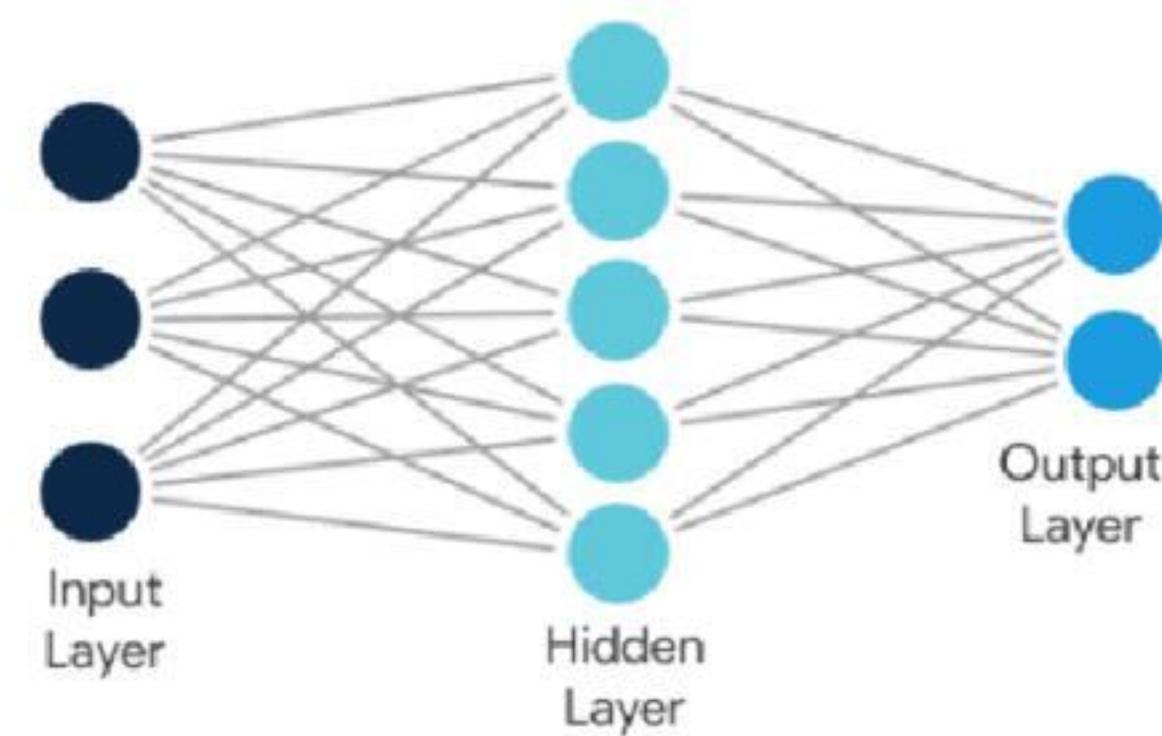
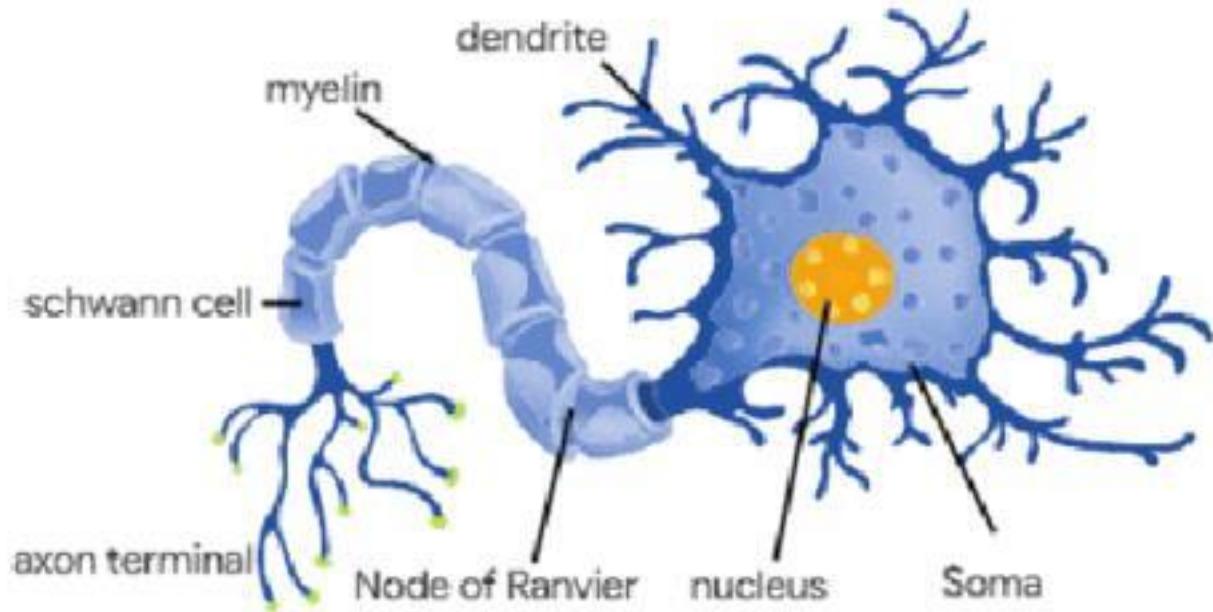
数据模型中包含多少层，这被称为模型的深度 **depth**。



# 深度学习 - 神经网络

神经网络 neural network 这一术语来自于神经生物学

- 深度学习的一些核心概念是从人们对大脑的理解中汲取部分灵感而形成的
- 但没有证据表明大脑的学习机制与现代深度学习模型所使用的相同。



就表现形式而言，深度学习是从数据中学习表示的一种数学框架。

# 小结

机器学习：将输入（比如图像）映射到目标（比如标签“猫”）

- 这一过程是通过观察许多输入和目标的示例来完成的。

深度神经网络通过一系列简单**的数据变换**（层）来逐步实现这种输入到目标的映射

# 小结

机器学习：将输入（比如图像）映射到目标（比如标签“猫”）

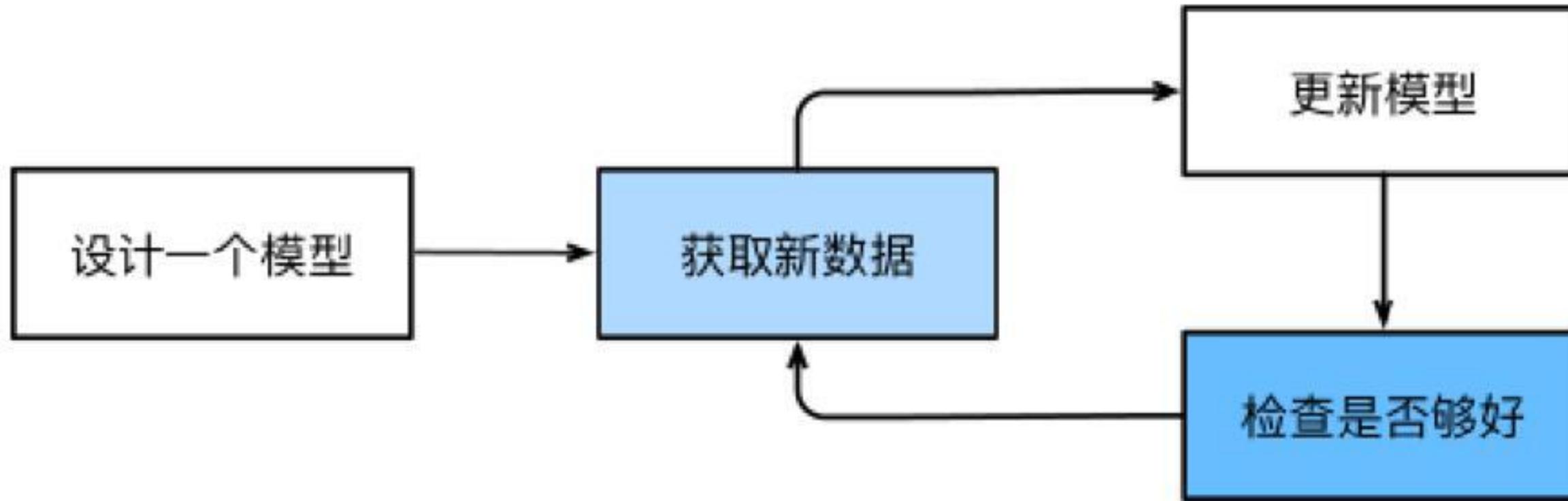
- 这一过程是通过观察许多输入和目标的示例来完成的。

深度神经网络通过一系列简单**的数据变换**（层）来逐步实现这种输入到目标的映射

下面用三张图来具体看一下机器学习的过程是如何发生的。

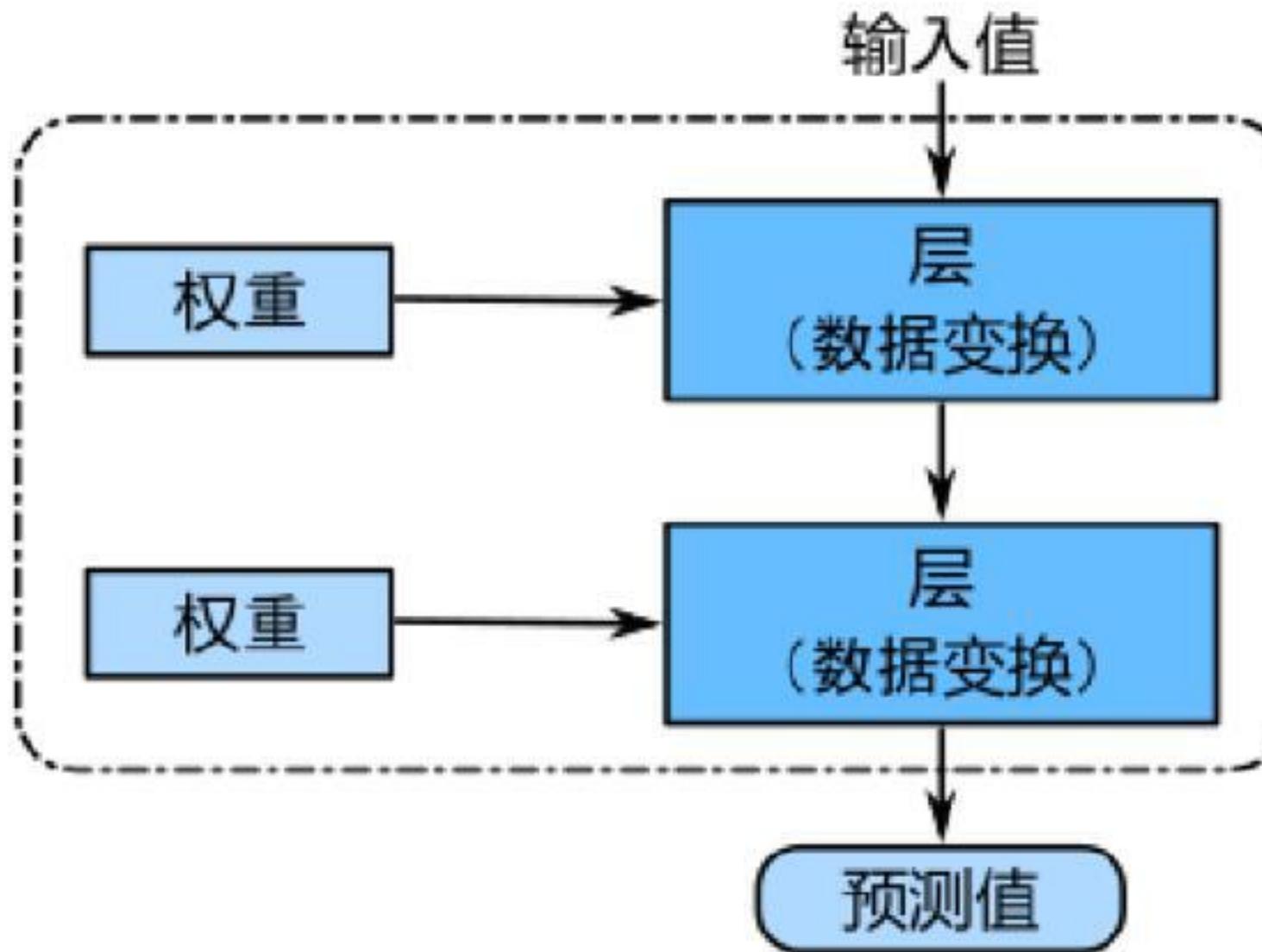
# 机器学习基本流程

# 典型学习过程



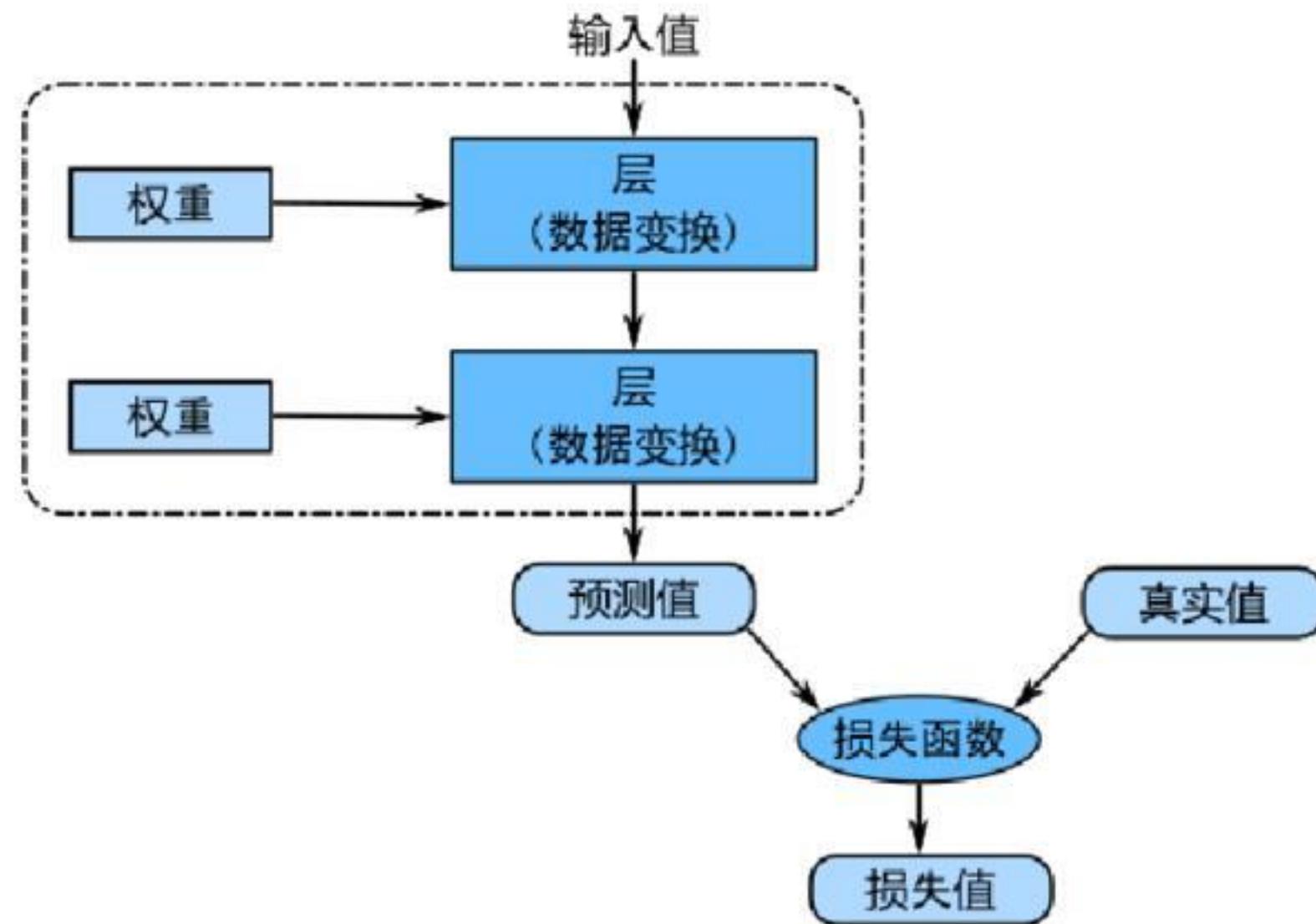
# 神经网络是由其权重来参数化

神经网络中每层对输入数据所做的具体操作保存在该层的权重 weight 中，其本质是一串数字，有时也被称为该层的参数 parameter。



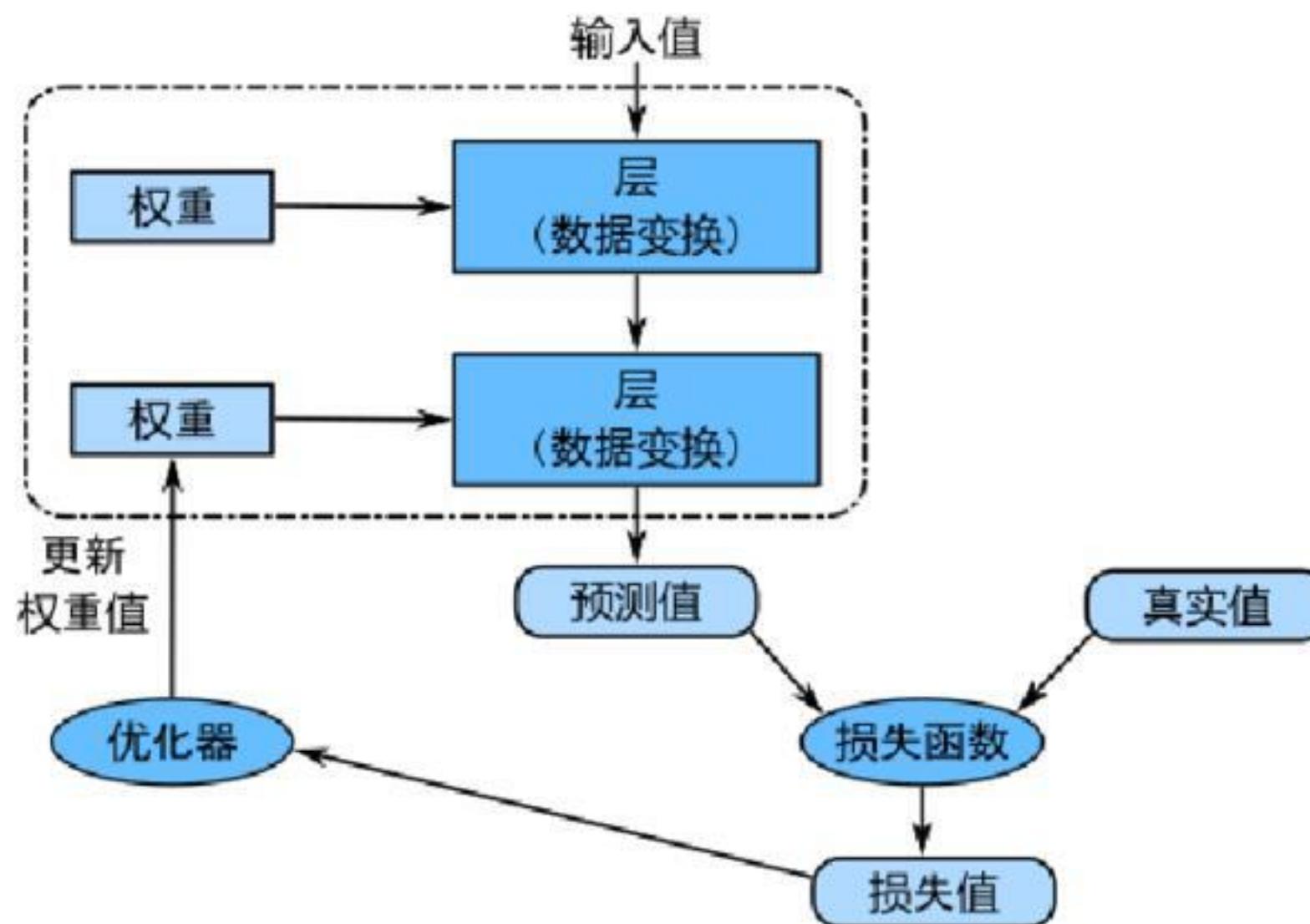
# 损失函数用来衡量网络输出结果的质量

想要控制神经网络的输出，就需要能够衡量该输出与预期值之间的距离。这是神经网络损失函数 **loss function** 的任务，该函数也叫目标函数 **objective function**。



# 将损失值作为反馈信号来调节权重

深度学习的基本技巧是利用这个距离值作为反馈信号来对权重值进行微调，以降低当前示例对应的损失值。这种调节由优化器 **optimizer** 来完成，它实现了所谓的反向传播 **backpropagation** 算法，这是机器学习的核心算法。



# 量变导致质变

一开始对神经网络的权重随机初始化，因此网络只是实现了一系列随机变换。

- 其输出结果自然也和理想值相去甚远，相应地，损失值也很高。

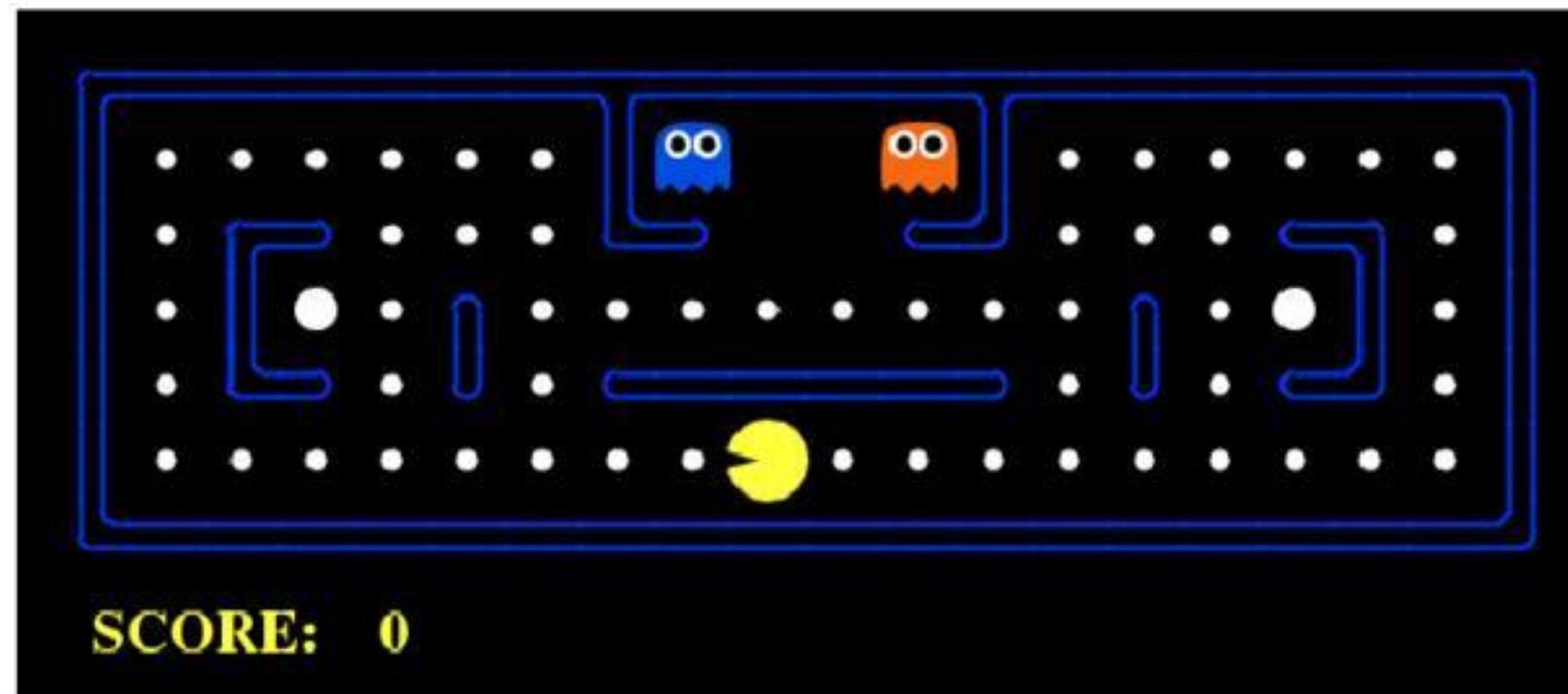
# 量变导致质变

一开始对神经网络的权重随机初始化，因此网络只是实现了一系列随机变换。

- 其输出结果自然也和理想值相去甚远，相应地，损失值也很高。

随着处理的示例越来越多，权重值向正确的方向逐步微调，损失值也逐渐降低。

- 这就是训练循环 **training loop**：将这种循环重复足够多的次数，得到的权重值可以使损失函数最小。



# 为什么是深度学习？

# 回到神经网络

自 2012 年以来，深度卷积神经网络已成为所有计算机视觉任务的首选算法

- 更一般地说，在所有感知任务上都有效

# 回到神经网络

自 2012 年以来，深度卷积神经网络已成为所有计算机视觉任务的首选算法

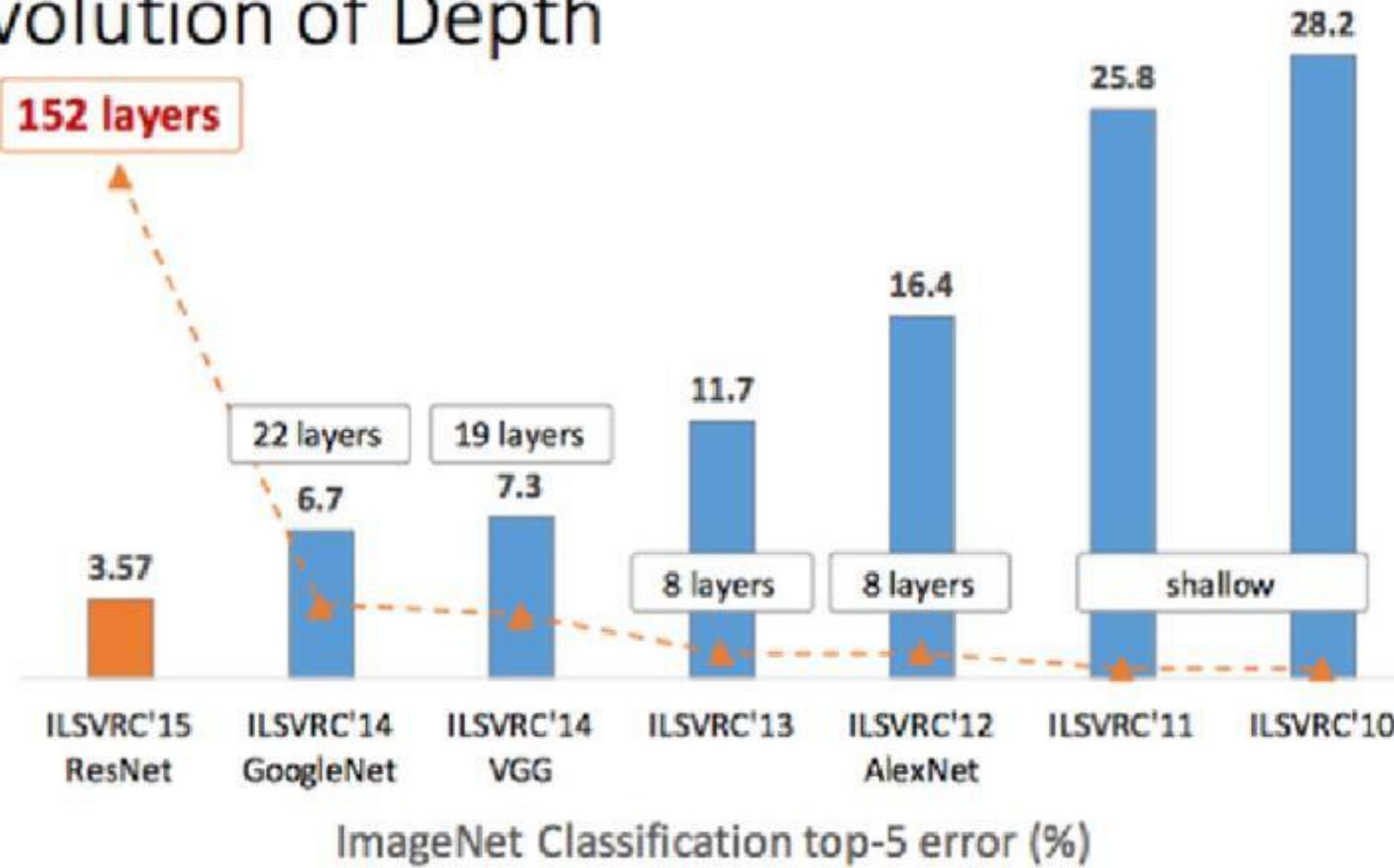
- 更一般地说，在所有感知任务上都有效

深度学习在许多其他类型的问题上成为首选算法，比如自然语言处理

- 已经在大量应用中完全取代了传统机器学习算法，如 SVM 与决策树

# 大型神经网络

## Revolution of Depth



# 深度学习已经取得的进展

- 接近（或超过）人类水平的图像分类
- 接近人类水平的语音识别
- 接近人类水平的手写文字转录
- 更好的机器翻译
- 更好的文本到语音转换
- 数字助理，比如谷歌即时（Google Now）和亚马逊 Alexa
- 接近人类水平的自动驾驶
- 更好的广告定向投放，Google、百度、必应都在使用
- 更好的网络搜索结果
- 能够回答用自然语言提出的问题
- 在围棋上战胜人类

# 神经网络的巅峰



**Turing Award 2018:** Yann LeCun, Geoffrey Hinton, Yoshua Bengio

# 向优秀的人学习

要想了解机器学习算法和工具的现状，一个好方法是看一下线上机器学习竞赛

- 需要招聘、解决问题的公司；学习、爱好者
- 哪种算法能够可靠地赢得竞赛？顶级参赛者使用哪些工具？

Kaggle 竞赛：<https://www.kaggle.com/>

# 向优秀的人学习

要想了解机器学习算法和工具的现状，一个好方法是看一下线上机器学习竞赛

- 需要招聘、解决问题的公司；学习、爱好者
- 哪种算法能够可靠地赢得竞赛？顶级参赛者使用哪些工具？

Kaggle 竞赛：<https://www.kaggle.com/>

Kaggle 上主要有两大方法：梯度提升机和深度学习。

- 梯度提升机用于处理结构化数据的问题，大多使用 **XGBoost**。
- 深度学习用于图像分类等感知问题，大多使用 **TensorFlow, PyTorch**。

# 深度学习有何不同？

深度学习发展得如此迅速，主要原因在于表现出更好的性能。

- 受到工业界的青睐，有极高商业价值

# 深度学习有何不同？

深度学习发展得如此迅速，主要原因在于表现出更好的性能。

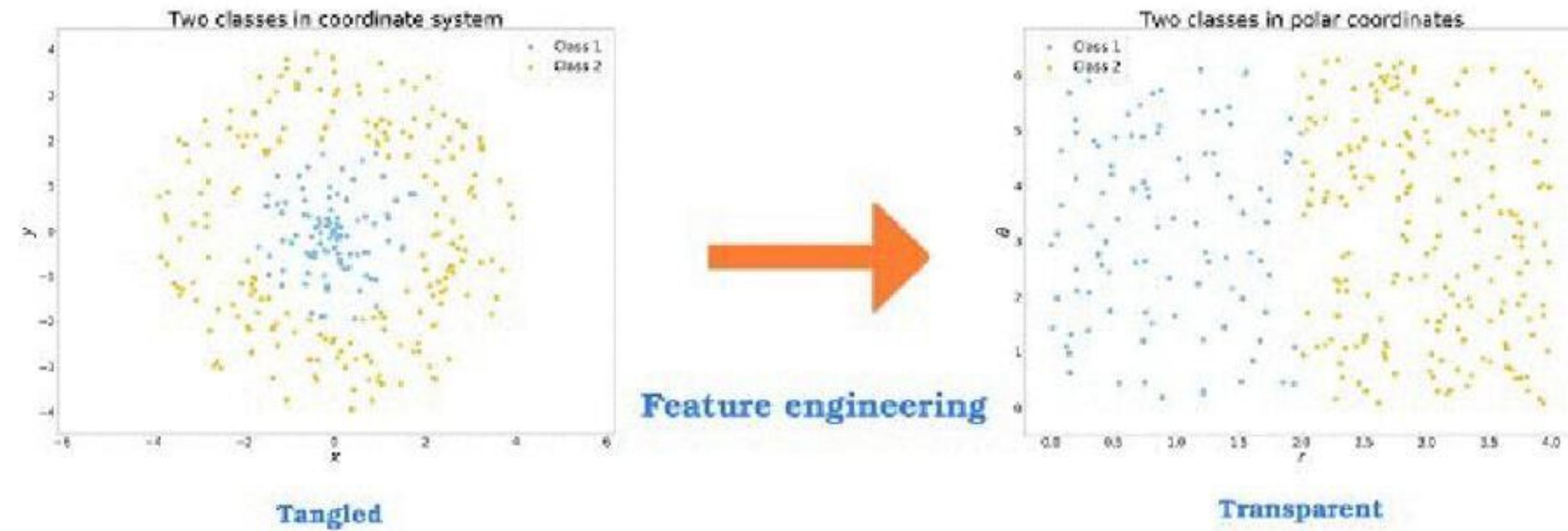
- 受到工业界的青睐，有极高商业价值

将特征工程完全自动化，让解决问题变得更加简单

- 曾经是机器学习算法流程中最关键的一步
- 学术界专家以外的从业者也能解决复杂问题。

# 什么是特征工程？

现阶段的人工智能可以理解为人机合作



人工干预的多少取决于模型的性能

- 早期机器学习技术或浅层学习：需要手动设计表示空间之间的数据变换。
- 深度学习完全将这个步骤自动化；在集成框架内一次性学习所有特征。

# 深度学习不是浅层学习的简单叠加

在实践中，如果连续应用浅层学习方法，其收益会随着层数增加迅速降低

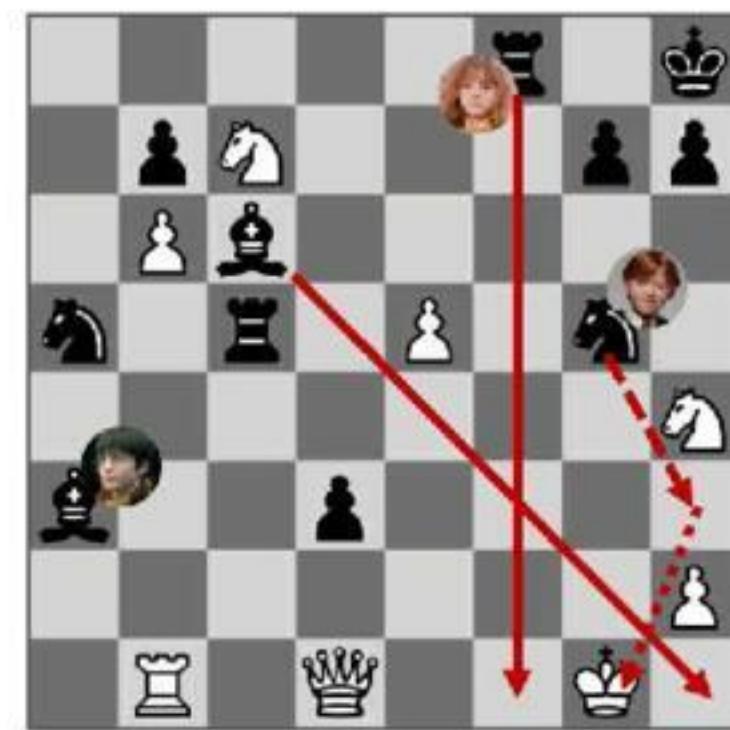
- 三层模型中最优的第一表示层并不是单层或双层模型中最优的第一表示层。

# 深度学习不是浅层学习的简单叠加

在实践中，如果连续应用浅层学习方法，其收益会随着层数增加迅速降低

- 三层模型中最优的第一表示层并不是单层或双层模型中最优的第一表示层。

深度学习的变革性在于，模型同时学习所有表示层，由单一反馈信号来监督。



# 为什么是现在？

# 从技术成熟到广泛应用

深度学习用于计算机视觉的两个关键思想，即卷积神经网络和反向传播，在1989年就已经为人们所知。

长短期记忆 LSTM, *long short-term memory* 算法是深度学习处理时间序列的基础，它在1997年就被开发出来了，而且此后几乎没有发生变化。

为什么深度学习在2012年之后才开始取得成功？这二十年间发生了什么变化？

# 从技术成熟到广泛应用

深度学习用于计算机视觉的两个关键思想，即卷积神经网络和反向传播，在1989年就已经为人们所知。

长短期记忆 LSTM, *long short-term memory* 算法是深度学习处理时间序列的基础，它在1997年就被开发出来了，而且此后几乎没有发生变化。

为什么深度学习在2012年之后才开始取得成功？这二十年间发生了什么变化？

总的来说，三种技术力量在推动着机器学习的进步：

- 硬件。
- 数据集和测试基准。
- 算法上的改进。

# 机器学习是一门工程科学

只有当合适的数据和硬件可用于尝试新想法时（或者将旧想法的规模扩大，工程实践往往如此），才可能出现算法上的改进。

- 在 20 世纪 90 年代和 21 世纪前十年，真正的瓶颈在于数据和硬件。
  - 算法没有可以实践的条件

# 机器学习是一门工程科学

只有当合适的数据和硬件可用于尝试新想法时（或者将旧想法的规模扩大，工程实践往往如此），才可能出现算法上的改进。

- 在 20 世纪 90 年代和 21 世纪前十年，真正的瓶颈在于数据和硬件。
  - 算法没有可以实践的条件

但在这段时间内发生了下面这些事情

- 互联网高速发展
- 针对游戏市场的需求开发出了高性能图形芯片。

# 数据，硬件

年代	数据规模	内存	每秒浮点运算
1970	100 (鸢尾花卉)	1 KB	100 KF (Intel 8080)
1980	1 K (波士顿房价)	100 KB	1 MF (Intel 80186)
1990	10 K (光学字符识别)	10 MB	10 MF (Intel 80486)
2000	10 M (网页)	100 MB	1 GF (Intel Core)
2010	10 G (广告)	1 GB	1 TF (Nvidia C2050)
2020	1 T (社交网络)	100 GB	1 PF (Nvidia DGX-2)

# 硬件

从 1990 年到 2010 年，非定制 CPU 的速度提高了约 5000 倍。

# 硬件

从 1990 年到 2010 年，非定制 CPU 的速度提高了约 5000 倍。

在 20 世纪前十年里，NVIDIA 和 AMD 等公司投资数十亿美元来开发快速的大规模并行芯片（图形处理器，GPU），以便为越来越逼真的视频游戏提供图形显示支持。

# 硬件

从 1990 年到 2010 年，非定制 CPU 的速度提高了约 5000 倍。

在 20 世纪前十年里，NVIDIA 和 AMD 等公司投资数十亿美元来开发快速的大规模并行芯片（图形处理器，GPU），以便为越来越逼真的视频游戏提供图形显示支持。

2007 年，NVIDIA 推出了 CUDA，作为其 GPU 系列的编程接口。

- 游戏中实时渲染复杂的3D场景需要大量的并行化矩阵计算。
- 非常巧合，深度神经网络也需要大量的并行化矩阵计算。

# 硬件

从 1990 年到 2010 年，非定制 CPU 的速度提高了约 5000 倍。

在 20 世纪前十年里，NVIDIA 和 AMD 等公司投资数十亿美元来开发快速的大规模并行芯片（图形处理器，GPU），以便为越来越逼真的视频游戏提供图形显示支持。

2007 年，NVIDIA 推出了 CUDA，作为其 GPU 系列的编程接口。

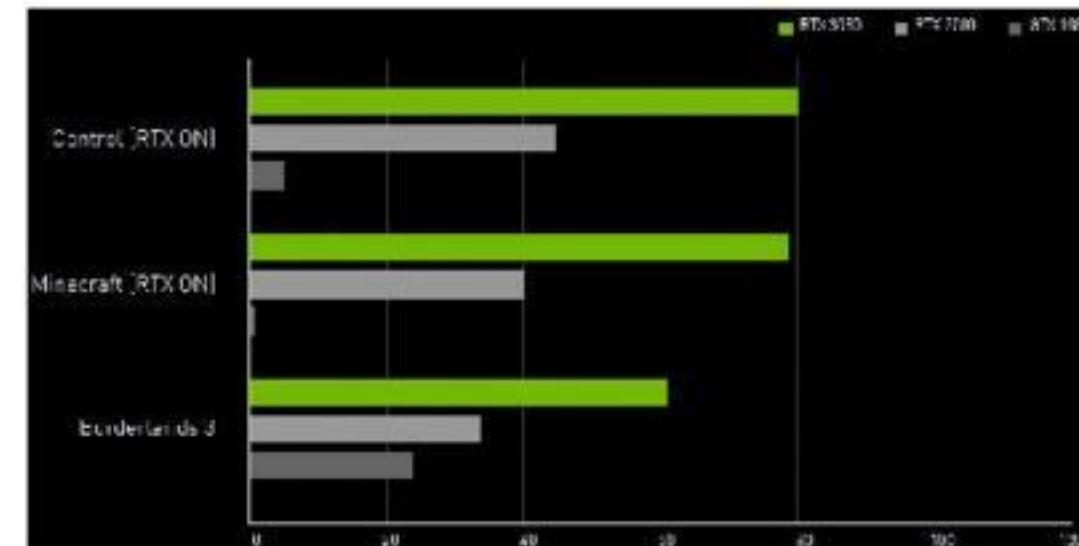
- 游戏中实时渲染复杂的3D场景需要大量的并行化矩阵计算。
- 非常巧合，深度神经网络也需要大量的并行化矩阵计算。
- 同样巧合，加密货币计算也需要大量的并行化矩阵计算。

# 游戏玩家的力量

## 游戏市场推动了人工智能应用必需的超级计算能力

NVIDIA 专业显卡比一台现代笔记本电脑的速度要快几百倍。

与此同时，大公司还在包含数百个 GPU 的集群上训练深度学习模型，这种类型的 GPU 是专门针对深度学习的需求开发的。

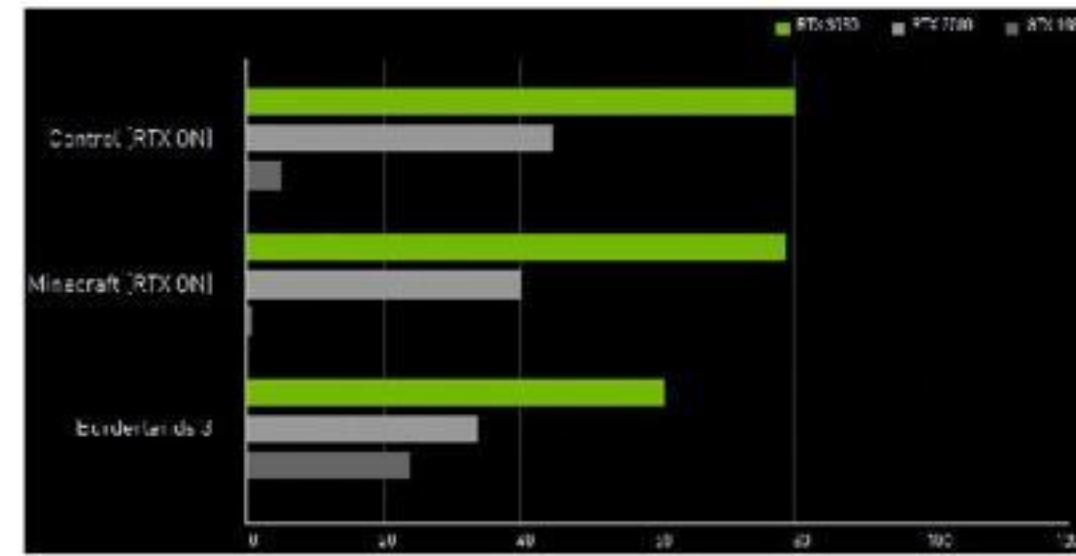


# 游戏玩家的力量

游戏市场推动了人工智能应用必需的超级计算能力

NVIDIA 专业显卡比一台现代笔记本电脑的速度要快几百倍。

与此同时，大公司还在包含数百个 GPU 的集群上训练深度学习模型，这种类型的 GPU 是专门针对深度学习的需求开发的。



2016 年，Google 展示的张量处理器 TPU 比最好的 GPU 还要快 10 倍。

# 数据

人工智能有时被称为新的工业革命。

- 如果深度学习是这场革命的蒸汽机，那么数据就是煤炭，即驱动智能机器的原材料，没有煤炭一切皆不可能。

# 数据

人工智能有时被称为新的工业革命。

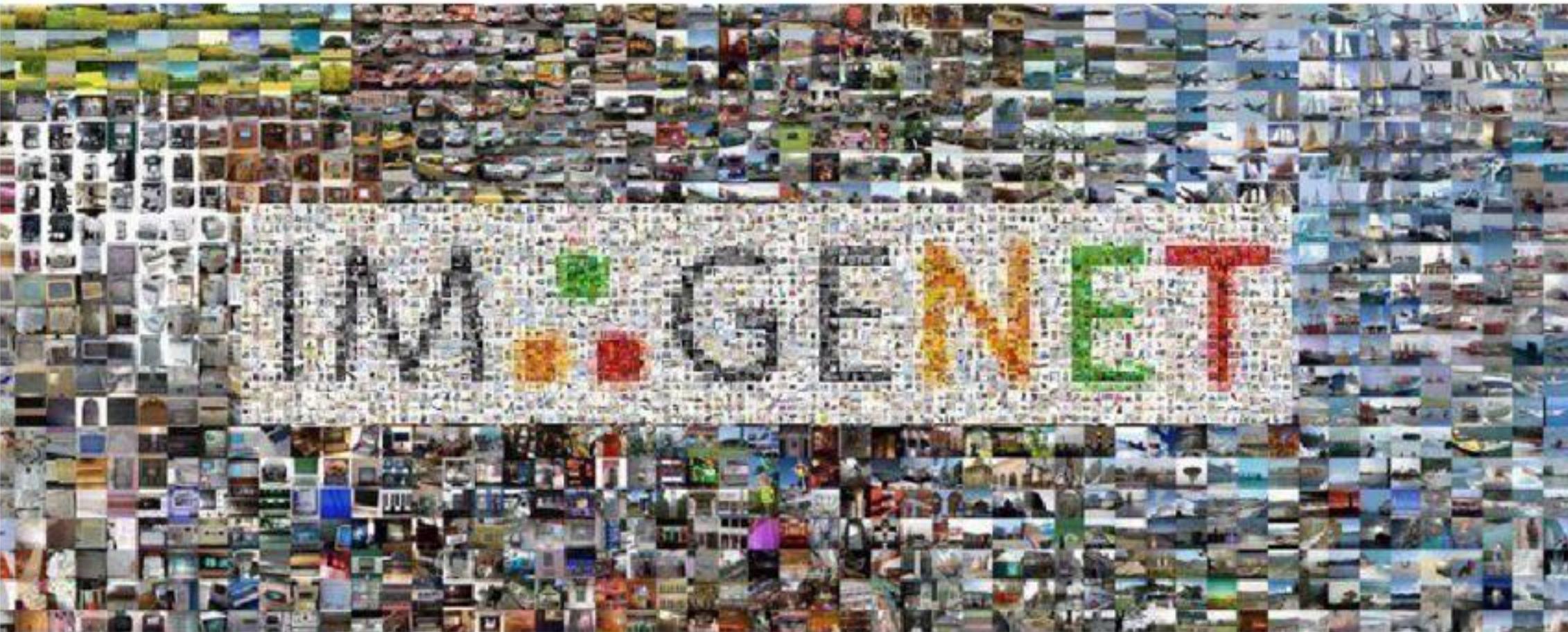
- 如果深度学习是这场革命的蒸汽机，那么数据就是煤炭，即驱动智能机器的原材料，没有煤炭一切皆不可能。

就数据而言，除了过去 20 年里存储硬件的指数级增长（遵循摩尔定律）

- 最大的变革来自于互联网的兴起，它使得收集与分发用于机器学习的超大型数据集变得可行。

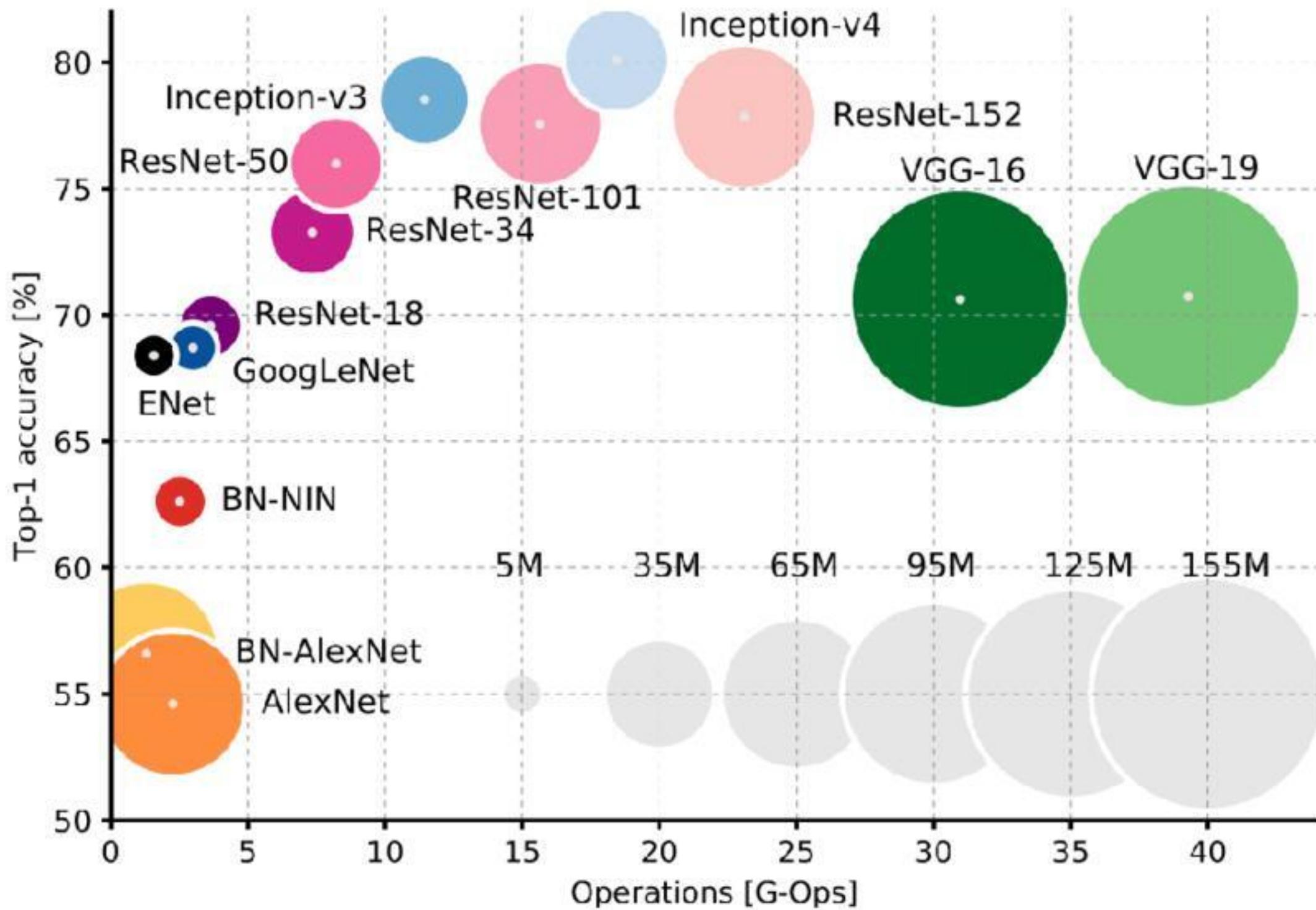
# 测试基准

深度学习兴起的催化剂：ImageNet 数据集。



ImageNet 的特殊之处不仅在于数量之大，还在于与它相关的年度竞赛：ILSVRC。

# 复杂度-精度



# 深度学习的大众化

## 技术层面

- 在早期，从事深度学习需要精通 C++ 和 CUDA，只有少数人才能掌握。
- TensorFlow, PyTorch 等用户友好型框架使深度学习变得像搭积木一样简单。

# 深度学习的大众化

## 技术层面

- 在早期，从事深度学习需要精通 C++ 和 CUDA，只有少数人才能掌握。
- TensorFlow, PyTorch 等用户友好型框架使深度学习变得像搭积木一样简单。

## 教育层面

- 我国正在试图在初等教育中推广人工智能和Python编程课程。
- 从2020年起大量高校开设人工智能学院或专业。

# 这种趋势会持续吗？

深度神经网络是否只是难以持续的昙花一现？

- 短期前所未见的资金投入：投资带来就业岗位
- 但到目前为止，这些资金很少能够转化为改变世界的产品和流程。

# 这种趋势会持续吗？

深度神经网络是否只是难以持续的昙花一现？

- 短期前所未见的资金投入：投资带来就业岗位
- 但到目前为止，这些资金很少能够转化为改变世界的产品和流程。

波士顿动力

- 2020年6月，机器狗Spot开始向公众开放，售价7.5万美元。



# 投资驱动的必要性

训练一个大型模型需要多大开销? (以训练 BERT-large 模型为例)

- 16 Cloud TPUs =  $16 * 4.5 = 72$  USD / hour
- One-day cost =  $72 * 24 = 1,728$  USD
- Four-day cost =  $1,728 \text{ USD} * 4 = 6,912$  USD

# 投资驱动的必要性

训练一个大型模型需要多大开销? (以训练 BERT-large 模型为例)

- $16 \text{ Cloud TPUs} = 16 * 4.5 = 72 \text{ USD / hour}$
- $\text{One-day cost} = 72 * 24 = 1,728 \text{ USD}$
- $\text{Four-day cost} = 1,728 \text{ USD} * 4 = 6,912 \text{ USD}$

注意: 以上开销计算对前提是训练成功!

- GPT-3 训练一次的费用时460万美元
  - 总训练成本大概1200万美元
- ChatGPT 每天的运行成本约10万美元

# 投资驱动的反噬

七年三次易主，估值仅剩三分之一

- 波士顿动力（1992年从MIT拆分出来，军用）
- 谷歌（2014年，30亿美元，研究）
- 软银（2017年，商业化）

# 投资驱动的反噬

七年三次易主，估值仅剩三分之一

- 波士顿动力（1992年从MIT拆分出来，军用）
- 谷歌（2014年，30亿美元，研究）
- 软银（2017年，商业化）
- 现代（2022年，11亿美元，娱乐？？？）



# 现代狗团

狗子易主狗团献舞：<https://www.bilibili.com/s/video/BV1m44y1i7oT>



# 不要相信短期炒作

警惕把达到人类水平的通用智能 **human-level general intelligence** 的讨论。

- 在短期内期望过高的风险是，一旦技术上没有实现，那么研究投资将会停止，而这也导致在很长一段时间内进展缓慢。

# 不要相信短期炒作

警惕把达到人类水平的通用智能 **human-level general intelligence** 的讨论。

- 在短期内期望过高的风险是，一旦技术上没有实现，那么研究投资将会停止，而这也导致在很长一段时间内进展缓慢。

两次人工智能冬天 **AI winter**:

- 20 世纪 60 年代的符号主义人工智能
- 20 世纪 80 年代的符号主义人工智能——专家系统 **expert system**

# 不要相信短期炒作

警惕把达到人类水平的通用智能 **human-level general intelligence** 的讨论。

- 在短期内期望过高的风险是，一旦技术上没有实现，那么研究投资将会停止，而这也导致在很长一段时间内进展缓慢。

两次人工智能冬天 **AI winter**:

- 20 世纪 60 年代的符号主义人工智能
- 20 世纪 80 年代的符号主义人工智能——专家系统 **expert system**

目前可能在炒作与失望的第三次循环，而且很多人过于乐观。

- 最好的做法是降低短期期望，确保对这一技术领域不太了解的人能够清楚地知道深度学习能做什么、不能做什么。

# 展望未来

不要怀疑：人工智能的时代即将到来。

- 人工智能最终将应用到社会和日常生活的所有方面，正如之前的互联网一样。
- 20 年后我们可能不再使用神经网络，但我们那时所使用的工具都是直接来自于现代深度学习及其核心概念。

# 展望未来

不要怀疑：人工智能的时代即将到来。

- 人工智能最终将应用到社会和日常生活的所有方面，正如之前的互联网一样。
- 20 年后我们可能不再使用神经网络，但我们那时所使用的工具都是直接来自于现代深度学习及其核心概念。

在一次科学革命之后，科学发展的速度通常会遵循一条 **S 形曲线**：首先是一个快速发展时期，接着随着研究人员受到严重限制而逐渐稳定下来，然后进一步的改进又逐渐增多。深度学习仍然处于这条 S 形曲线的前半部分，在未来几年将会取得更多进展。

# Review

# Summary

- Known: basic concepts, background
- Understand: machine learning pipeline
- Master: development environment configuration

