滴滴内部
学习资料
请勿外传

机器学习

深 度 强 化 学 习

滴滴学院 CTO支持中心 胡海檬

课程简介

深度学习是近十年来人工智能领域取得的重要突破。它在语音识别、自然语言处理、计算机视觉等诸多领域的应用取得了巨大成功。强化学习是关于序列决策的一种工具。深度学习的突破性进展也促使强化学习重新获得了产业各界的关注。2016年DeepMind开发的AlphaGo程序利用深度强化学习算法击败世界围棋高手李世石，至此深度强化学习算法引起更多学者的关注，并正在改变和影响着这个世界。滴滴为提升整个公司的深度强化学习能力规划了一系列课程，旨在培养更多员工具备这项能力，并不断应用到实际问题解决中，用技术推动业务发展。

这个系列课程会系统介绍深度强化学习基本原理，代表性算法以及应用。目前深度和强化学习已在滴滴很多应用场景取得了较好的成果，本系列课程也将阶段性的对这些实践成果进行介绍。

目标学员及课程收益

	人群	收益
1	所有想了解深度强化学习的同学	了解和理解主要理论知识、方法论
2	工作中有实际应用场景的同学	了解相关工具和开源代码可以解决的问题
3	需要深入使用和具体实操的同学	获取具体使用的方法

课程安排

第1讲，深度强化学习综述及应用　叶杰平
第2讲，马尔可夫决策过程与强化学习基础/OpenAI Gym　徐哲
第3讲，动态规划求解方法/GridWorld　徐哲
第4讲，Q学习/Pacman　徐哲
第5讲，深度Q学习入门与实践　Tony Qin
第6讲，Contextual bandits算法总览　Tony Qin
第7讲，神经网络简介　龚平华
第8讲，深度学习工具简介和入门(Keras/Tensorflow)　龚平华
第9讲，卷积神经网络的原理和应用　王征
第10讲，序列神经网络的原理和应用　王征
第11讲，深度学习进展　王征
第12~17讲，强化学习　徐哲
第18讲，深度强化学习应用之AlphaGo　Tony Qin
第19~23讲，生成对抗网络(GAN)　陈学文

深度强化学习在滴滴及业界的实践应用分享将在课程体系中穿插进行

从9月15日起每隔周五
10:00-11:00
数字山谷2号楼5层分享厅

讲师团队

- 资深大数据技术专家
- 业界的最佳实践者
- 滴滴内部应用实操项目负责人

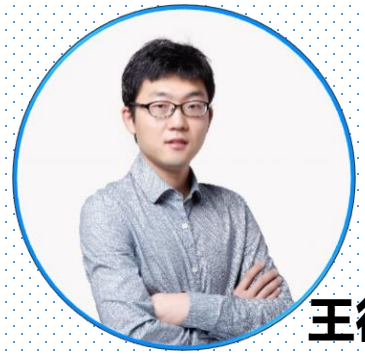
部分讲师


叶杰平


陈学文


秦志伟(Tony)


刘亚书


王征


徐哲


龚平华


王杰

学习攻略

- ✓ 现场参训，从通知中了解具体地点
- ✓ 观看直播，直播链接随开课通知邮件发布
- ✓ 观看录播，每次课程结束后两个工作日内，将课程视频上传至木吉学堂

- ■ 关注通知和提醒，针对所有报名的同学，我们将定期发送calendar
- ■ 获取学习资料，提前预习，我们将在钉钉群中提前1-2天发送课程ppt供预习
- ■ 针对积极学习，并进行实践应用的同学，将有机会赢得机智奖学金

桔子TV

木吉学堂

机器学习

深 度 强 化 学 习

综述及应用

学习方法

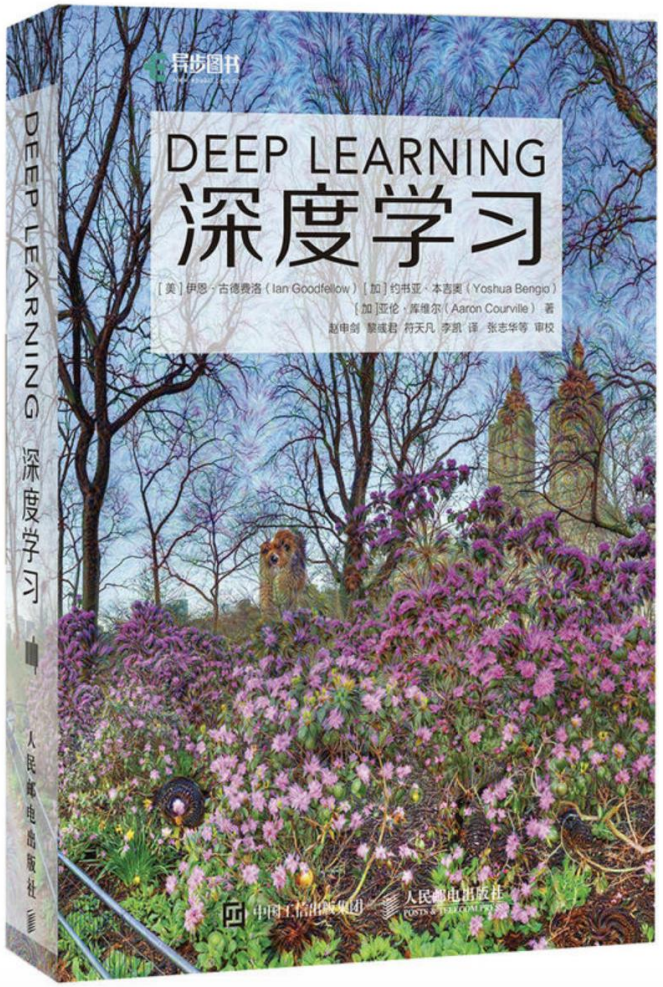
WHO，适合谁来学？

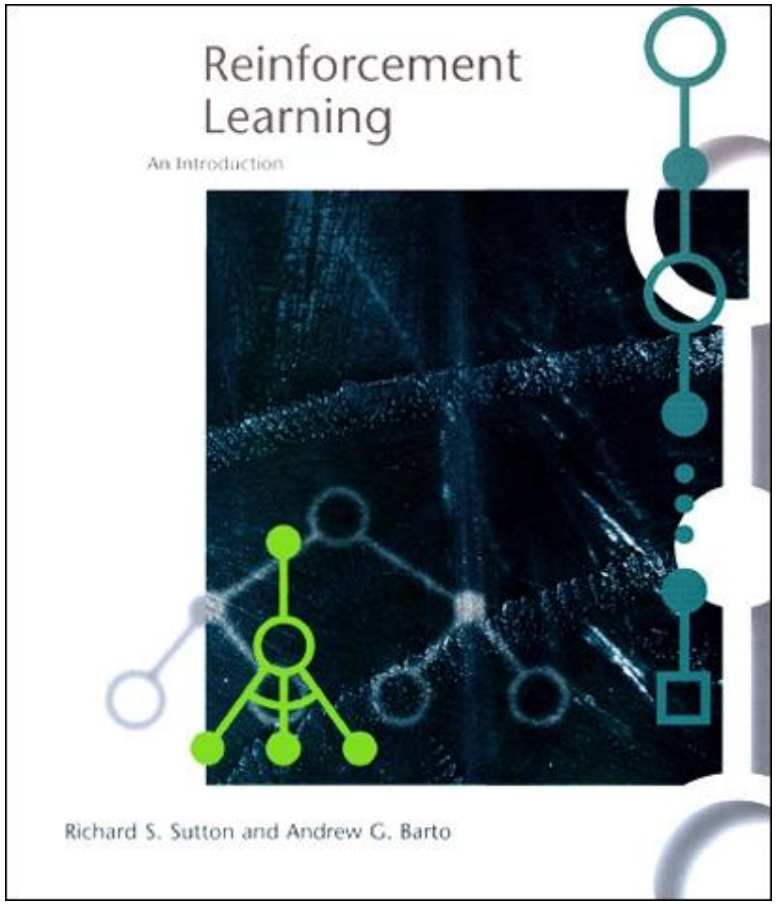
WHAT，能学到什么？

HOW，怎么学？

课程特点

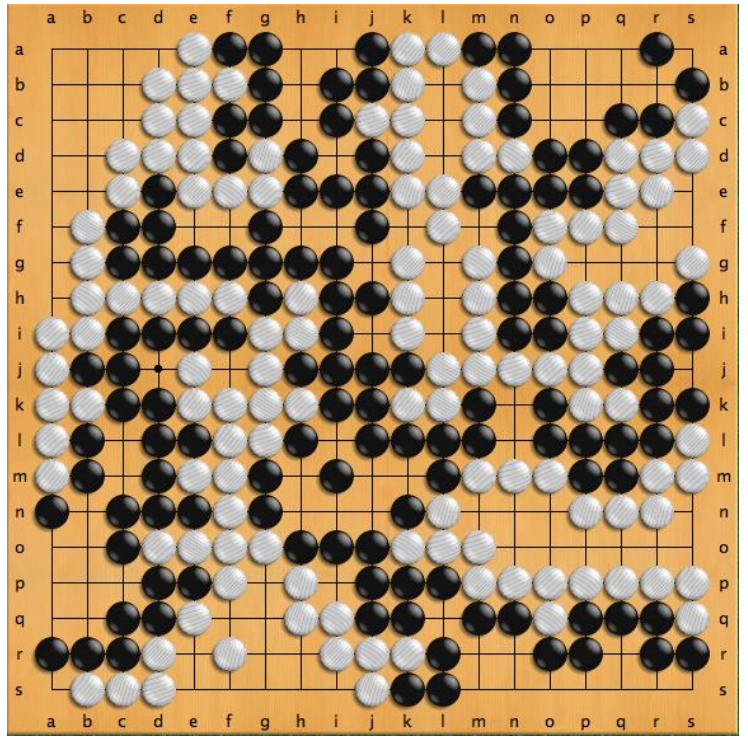
推荐书籍





http://incompleteideas.net/sutton/book/bookdraft2017june19.pdf

AlphaGo



国务院印发《新一代人工智能发展规划》

2017年7月8日，国务院印发《新一代人工智能发展规划》，人工智能在国家层面进行系统布局，成为国家战略，人工智能作为新一轮科技革命和产业变革的核心驱动力，将深刻改变人类社会生活，改变世界。

滴滴的愿景是成为引领汽车和交通行业变革的世界顶级科技公司，人工智能技术的发展和应用必将促进我们更快地迈向愿景。

国务院关于印发
新一代人工智能发展规划的通知

国发〔2017〕35号

各省、自治区、直辖市人民政府，国务院各部委、各直属机构：

现将《新一代人工智能发展规划》印发给你们，请认真贯彻执行。

国务院

2017年7月8日

（此件公开发布）

《麻省理工科技评论》世界十大技术突破

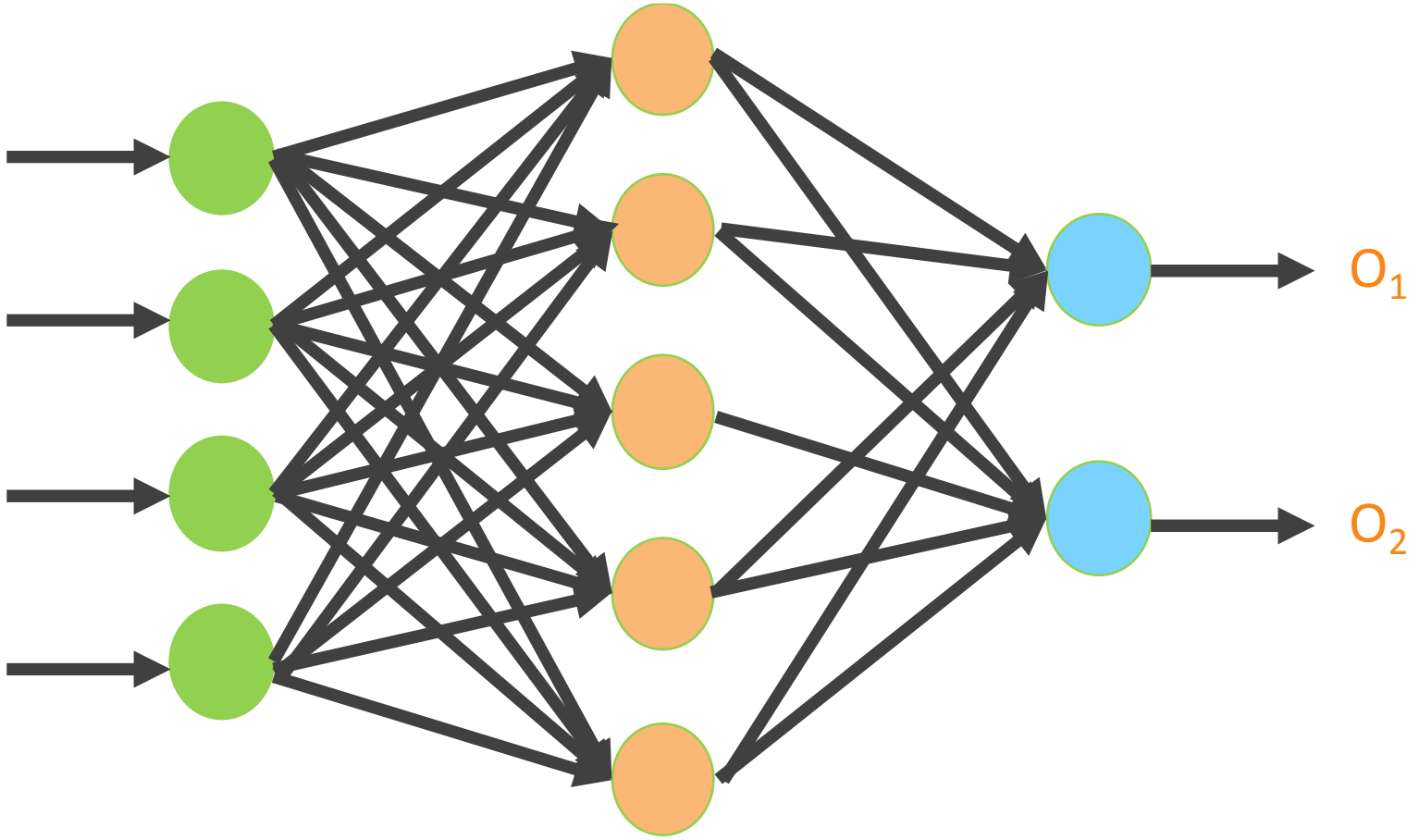
- 鉴于**深度学习**在学术界和工业界的巨大影响力，2013年，《麻省理工科技评论》(MIT Technology Review)将其列为世界十大技术突破之首。

- 2017年，《麻省理工科技评论》(MIT Technology Review)将**强化学习**列为世界十大技术突破之首。

机器学习

深度学习简介

神经网络

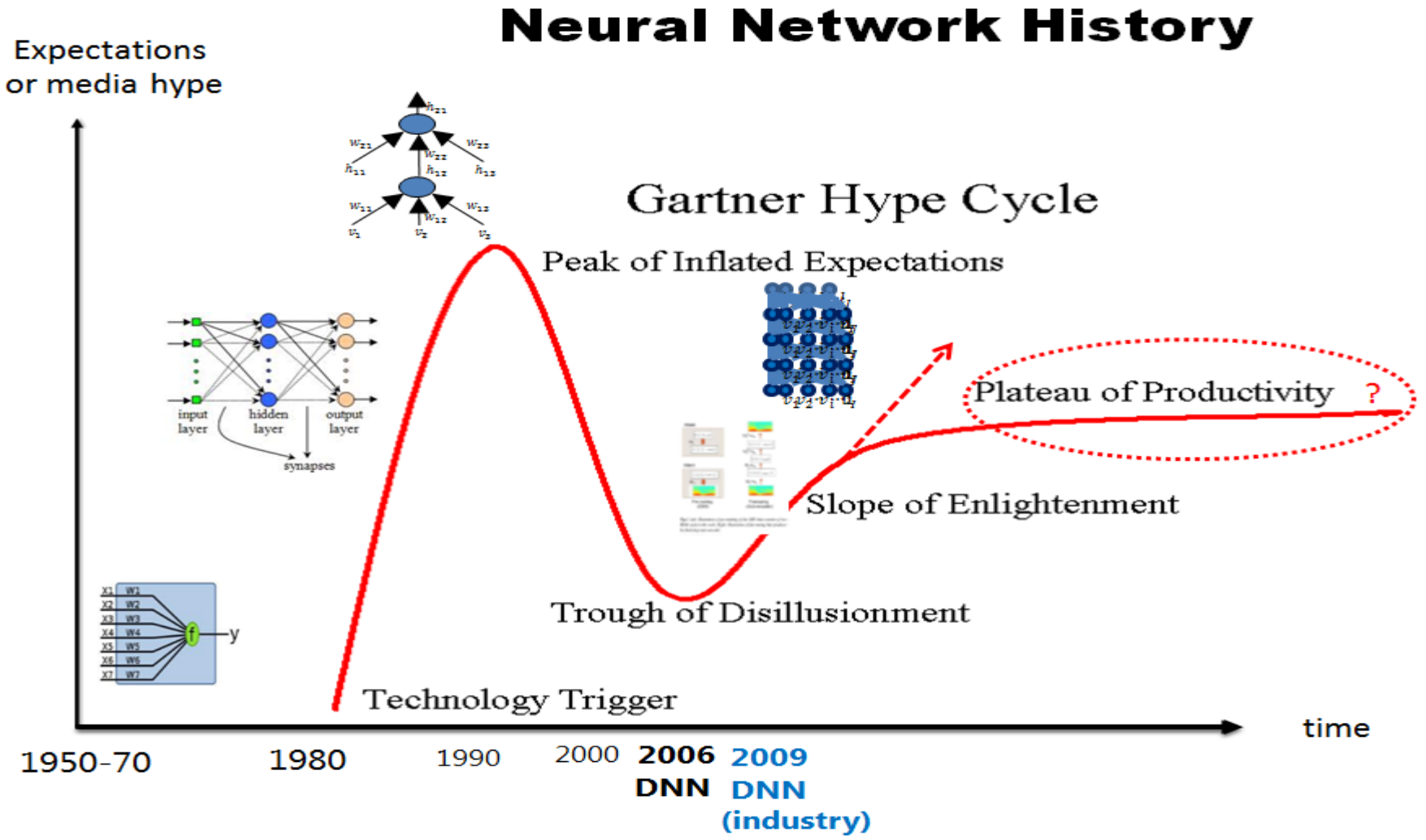


输入层　　隐层　　输出层

神经网络简介

- 神经网络的起源可追溯到20世纪40年代，曾经在八九十年代流行
  - 1986年，Rumelhart，Hinton 和 Williams 发表了著名的反向传播算法用于训练神经网络，该算法直到今天仍被广泛应用。
  - 神经网络有大量参数，经常发生过拟合问题。这是因为当时的训练数据集规模都较小，加之计算资源有限。与其他模型相比，神经网络并未在识别准确率上体现出明显的优势。
- 支持向量机、Boosting等分类器九十年代开始流行
  - 称为浅层机器学习模型。在这种模型中，往往是针对不同的任务设计不同的系统，并采用不同的手工设计的特征。

神经网络发展史


Neural Network History
Expectations or media hype
Gartner Hype Cycle
Peak of Inflated Expectations
Plateau of Productivity
?
Slope of Enlightenment
Trough of Disillusionment
Technology Trigger
input layer
hidden layer
output layer
synapses
time
1950-70
1980
1990
2000
2006 DNN
2009 DNN (industry)

深度学习的兴起

- 2006年, Hinton提出了深度信念网络(DBN), 通过 “预训练+微调” 使得深度模型的最优化变得相对容易
- 2012年, Hinton 组参加ImageNet 竞赛, 使用CNN 模型以超过第二名10个百分点的成绩夺得当年竞赛的冠军
- 深度学习模型在计算机视觉、自然语言处理、语音识别等众多领域都取得了较大的成功

深度学习兴起的原因

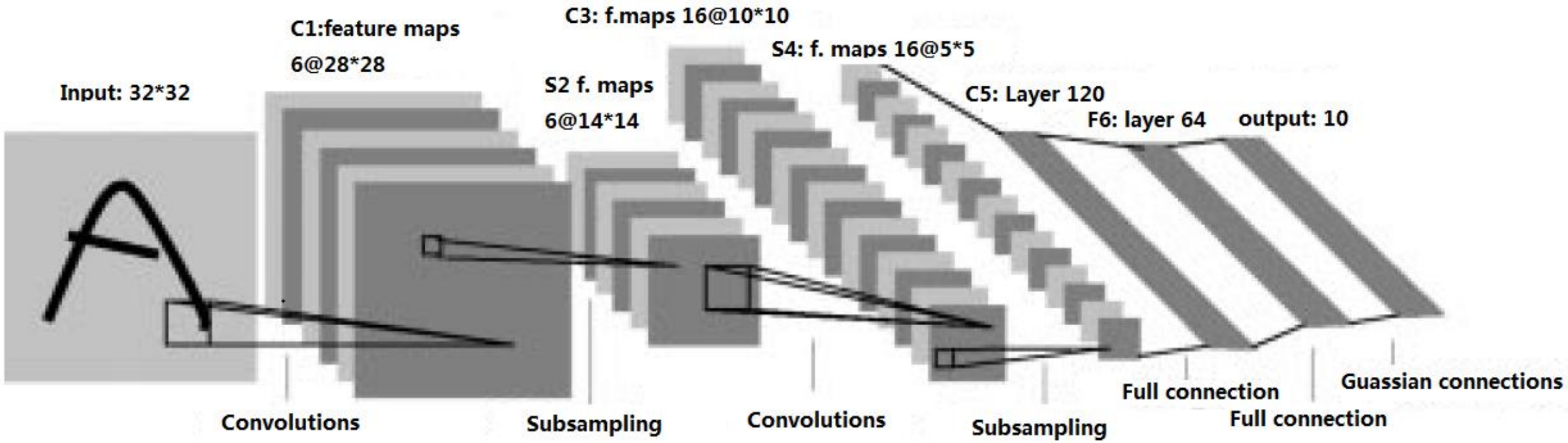
- 大规模训练数据的出现在很大程度上缓解了训练过拟合的问题。例如，ImageNet训练集拥有上百万个有标注的图像。
- 计算机硬件的飞速发展为其提供了强大的计算能力，一个GPU芯片可以集成上千个核。这使得训练大规模神经网络成为可能。
- 神经网络的模型设计和训练方法都取得了长足的进步。例如，为了改进神经网络的训练，学者提出了非监督和逐层的预训练，使得在利用反向传播算法对网络进行全局优化之前，网络参数能达到一个好的起始点，从而在训练完成时能达到一个较好的局部极小点。

深度学习@业界

- 在Hinton的科研小组赢得ImageNet比赛冠军之后的6个月，谷歌和百度都发布了新的基于图像内容的搜索引擎。他们采用深度学习模型，应用在各自的数据上，发现图像搜索准确率得到了大幅度提高。
- 脸谱于2013年12月在纽约成立了新的人工智能实验室，聘请深度学习领域的著名学者、卷积网络的发明人雅恩·乐昆(Yann LeCun)作为首席科学家。
- 2014年1月，谷歌抛出四亿美金收购了深度学习的创业公司DeepMind。

- 鉴于深度学习在学术界和工业界的巨大影响力，2013年，《麻省理工科技评论》(MIT Technology Review)将其列为世界十大技术突破之首。

卷积神经网络（CNN）


C3: f.maps 16@10*10
C1:feature maps
6@28*28
S4: f. maps 16@5*5
S2 f. maps
6@14*14
Input: 32*32
C5: Layer 120
F6: layer 64
output: 10
Convolutions
Subsampling
Convolutions
Subsampling
Full connection
Full connection
Guassian connections

卷积神经网络简介（1）

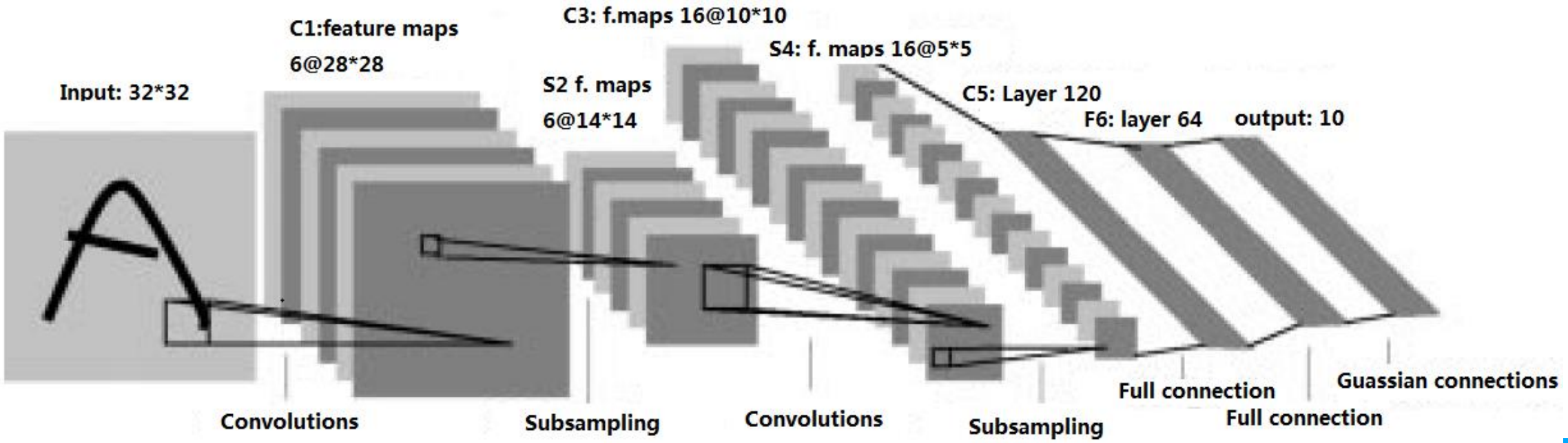
- 卷积神经网络（Convolutional Neural Network，CNN）最初是为解决图像识别等问题设计的，也可用于时间序列信号，比如音频信号、文本数据等。
- 在早期的图像识别研究中，最大的挑战是如何组织特征，因为图像数据不像其他类型的数据那样可以通过人工理解来提取特征。
  - 在深度学习出现之前，我们必须借助SIFT、HoG等算法提取具有良好区分性的特征，再集合SVM等机器学习算法进行图像识别。
  - SIFT对一定程度内的缩放、平移、旋转、视角改变、亮度调整等畸变，都具有不变性。然而SIFT这类算法在ImageNet比赛的最好结果的错误率也有26%以上，而且难以产生突破。

卷积神经网络简介（2）

- 卷积神经网络提取的特征则可以达到更好的效果，同时它不需要将特征提取和分类训练两个过程分开，它在训练时就自动提取了最有效的特征。
- CNN的最大特点在于卷积的权值共享结构，可以大幅减少神经网络的参数量，防止过拟合的同时又降低了神经网络模型的复杂度。
- 20世纪80年代，日本科学家提出神经认知机（Neocognitron）的概念，可以算作是卷积网络最初的实现原型。
  - 神经认知机中包含两类神经元，用来抽取特征的S-cells，还有用来抗形变的C-cells，其中S-cells对应我们现在主流卷积神经网络中的卷积核滤波操作，而C-cells则对应激活函数、最大池化（Max-Pooling）等操作。

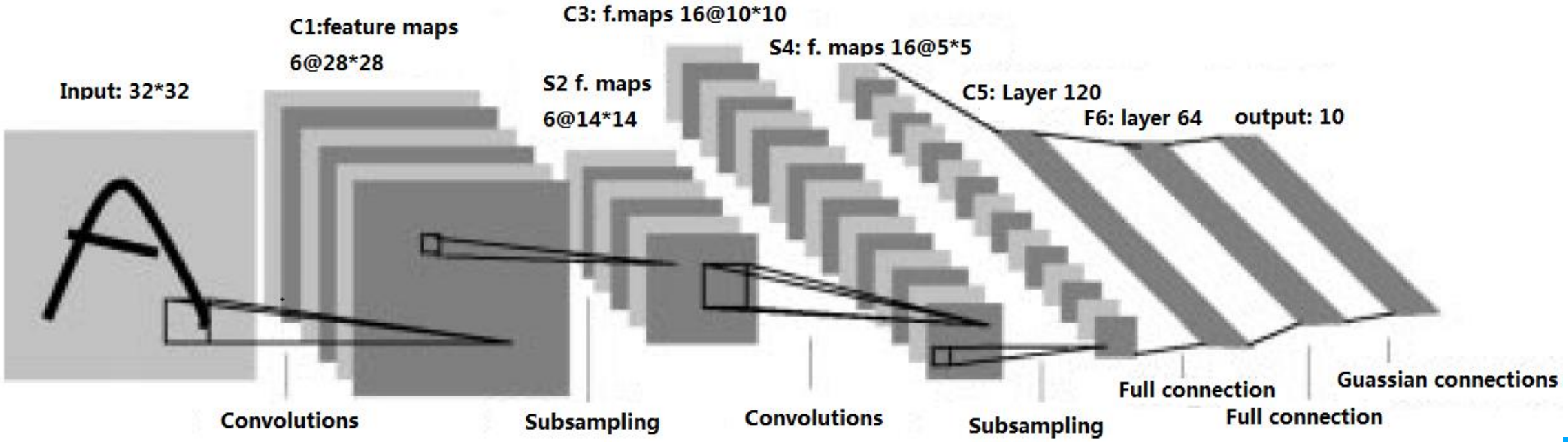
卷积神经网络简介（3）

- 一般的卷积神经网络由多个卷积层构成，每个卷积层中通常会进行如下几个操作。
  - 图像通过多个不同的卷积核的滤波，并加偏置，提取出局部特征，每一个卷积核会映射出一个新的2D图像。
  - 将前面卷积核的滤波输出结果，进行非线性的激活函数处理。目前最常见的是使用ReLU函数，而以前Sigmoid函数用得比较多。
  - 对激活函数的结果再进行池化操作（即降采样，比如将2×2的图片降为1×1的图片）。



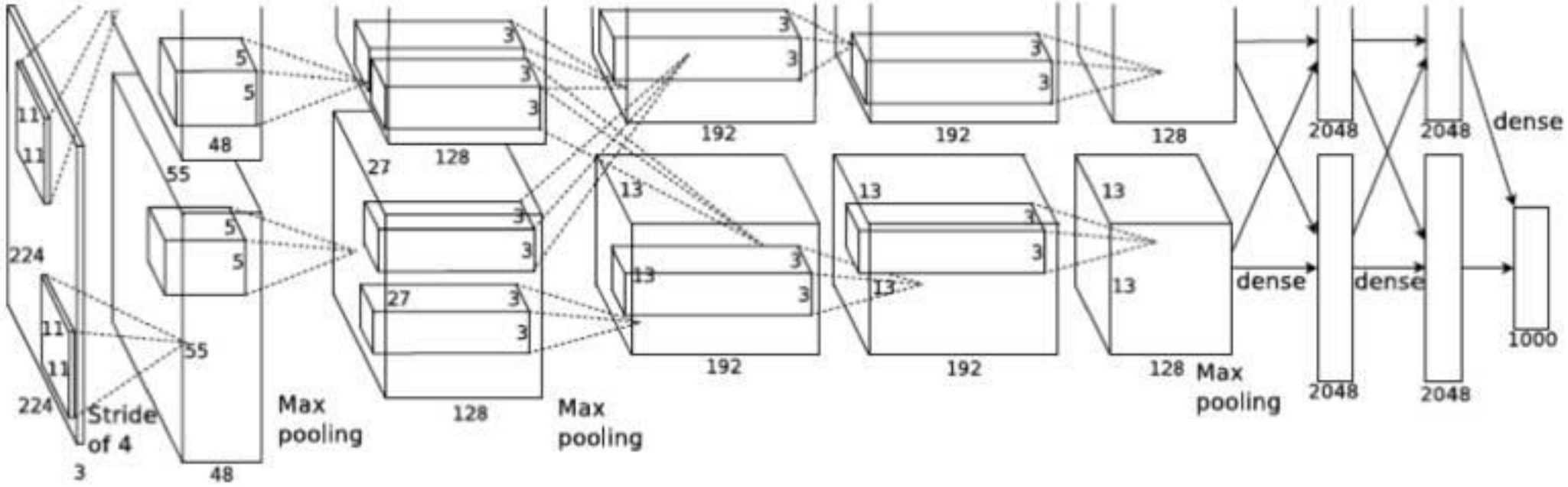
卷积神经网络简介（4）

- 卷积神经网络的要点就是局部连接（Local Connection）、权值共享（Weight Sharing）和池化层（Pooling）中的降采样（Down-Sampling）。
  - 局部连接和权值共享降低了参数量，使训练复杂度大大下降，并减轻了过拟合，权值共享还赋予了卷积网络对平移的容忍性
  - 池化层降采样则进一步降低了输出参数量，并赋予模型对轻度形变的容忍性，提高了模型的泛化能力。



AlexNet简介（1）

- 2012年，Hinton的学生Alex Krizhevsky提出了深度卷积神经网络模型AlexNet，它可以算是LeNet的一种更深更宽的版本。AlexNet中包含了几个比较新的技术点，也首次在CNN中成功应用了ReLU激活函数、Dropout等Trick。同时AlexNet也使用了GPU进行运算加速。



AlexNet简介（3）

- AlexNet包含了6亿3000万个连接，6000万个参数和65万个神经元，拥有5个卷积层，其中3个卷积层后面连接了最大池化层，最后还有3个全连接层。AlexNet以显著的优势赢得了竞争激烈的ILSVRC 2012比赛，top-5的错误率降低至了16.4%，相比第二名的成绩26.2%错误率有了巨大的提升。

- AlexNet确立了深度学习（深度卷积网络）在计算机视觉的统治地位，同时也推动了深度学习在语音识别、自然语言处理、强化学习等领域的拓展。

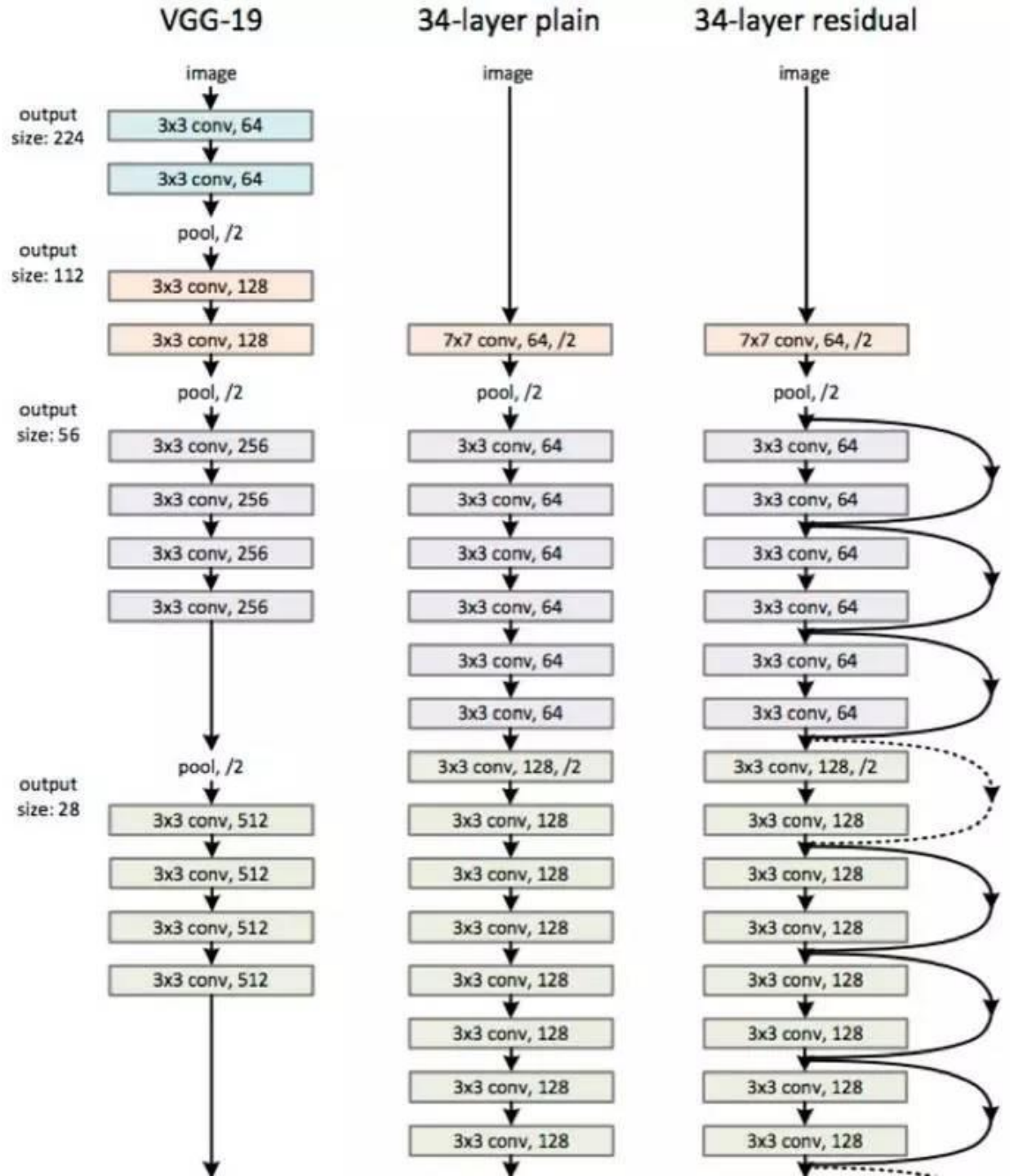
ImageNet（1）

ImageNet（2）

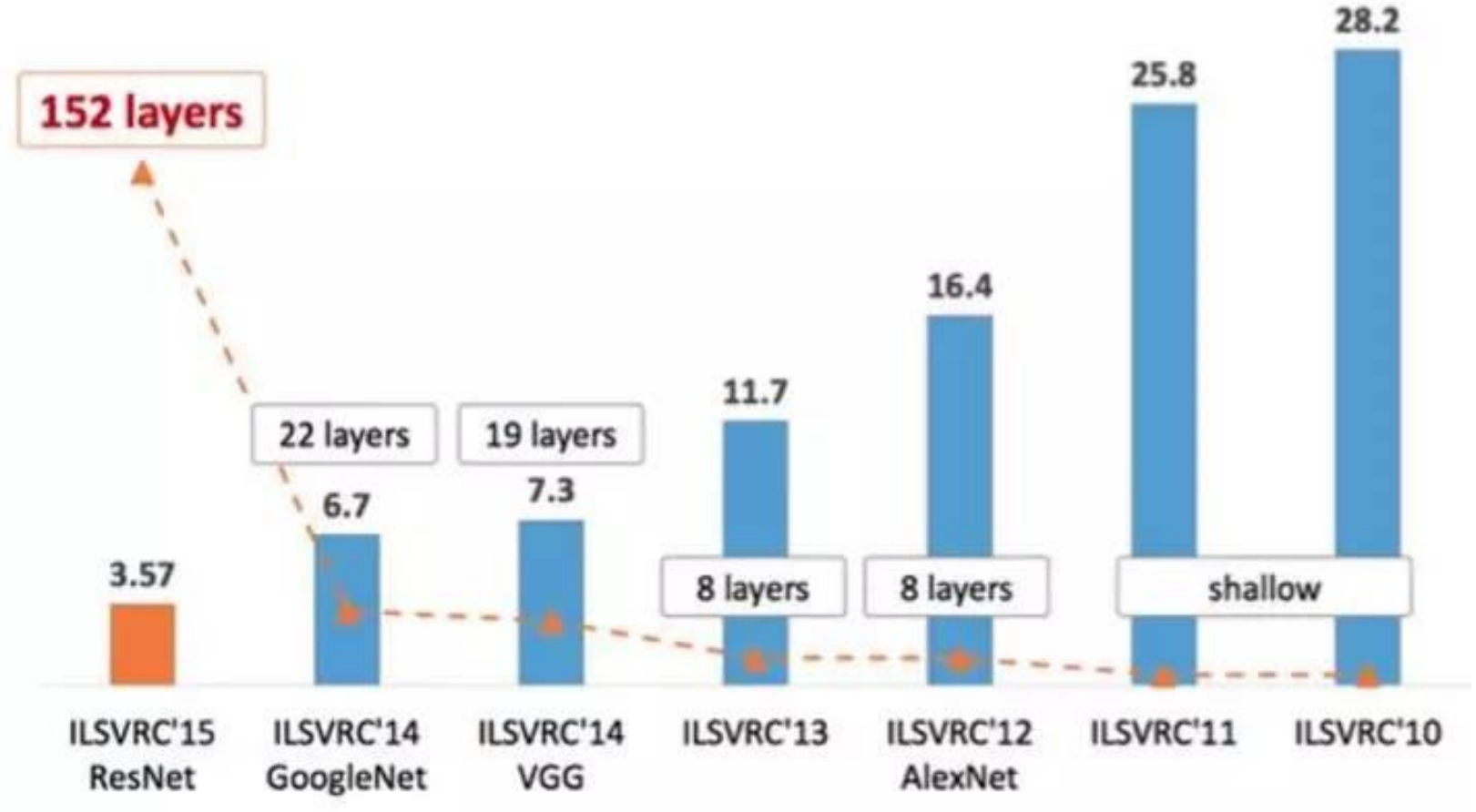

mite
container ship
motor scooter
leopard
mite
black widow
cockroach
tick
starfish
container ship
lifeboat
amphibian
fireboat
drilling platform
motor scooter
go-kart
moped
bumper car
golfcart
leopard
jaguar
cheetah
snow leopard
Egyptian cat

其它CNN模型

- VGG by Simonyan and Zisserman 发表在他们2014 文章 《Very Deep Convolutional Networks for Large Scale Image Recognition》.
- ResNet by He et al. 发表在他们2015文章《Deep Residual Learning for Image Recognition》。



ImageNet比赛结果



深度学习小结

- Perceptron（感知机）于**1957年**由Frank Resenblatt提出，而Perceptron不仅是卷积网络，也是神经网络的始祖。Neocognitron（神经认知机）是一种多层级的神经网络，由日本科学家Kunihiko Fukushima于20世纪80年代提出，具有一定程度的视觉认知的功能，并直接启发了后来的卷积神经网络。
- LeNet-5由CNN之父Yann LeCun于**1997年**提出，首次提出了多层级联的卷积结构，可对手写数字进行有效识别。
- **2012年**，Hinton的学生Alex依靠8层深的卷积神经网络一举获得了IMAGENET 2012比赛的冠军，瞬间点燃了卷积神经网络研究的热潮。

深度学习@滴滴

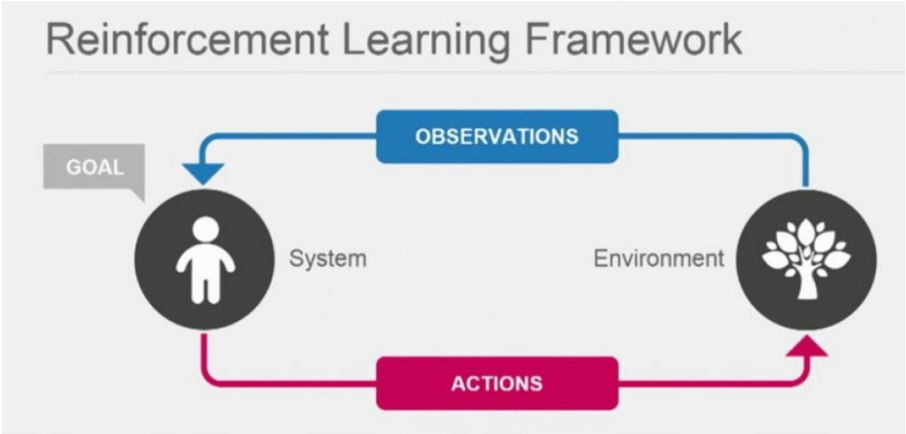
- 图像识别
- 语音识别
- 供需预测
- ETA
- …

机器学习

强化学习简介

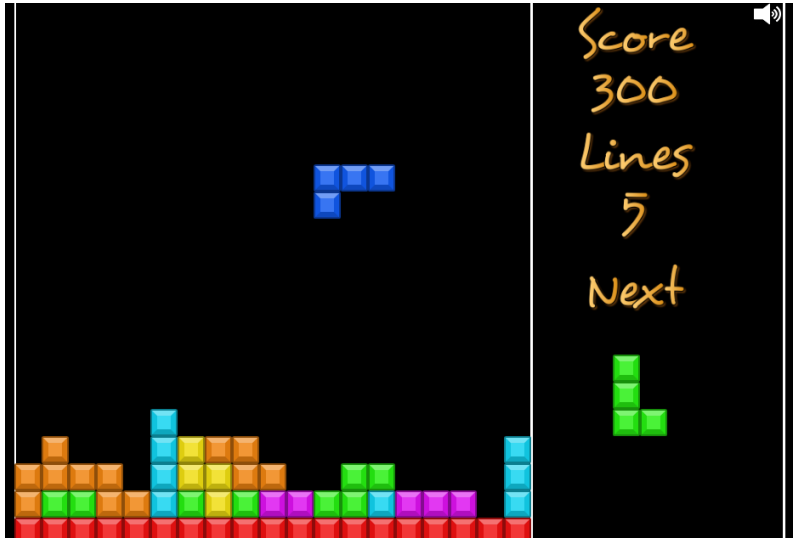
简介

- 怎样学习一项新的技能？或者在一个全新的环境里做最优的决策？
  - 例如小孩学走路
- 我们不一定能一步达到最后的目标
- 和环境交互，获取反馈，循序进步
  - 增强学习是一种模仿人类学习新技能的方法



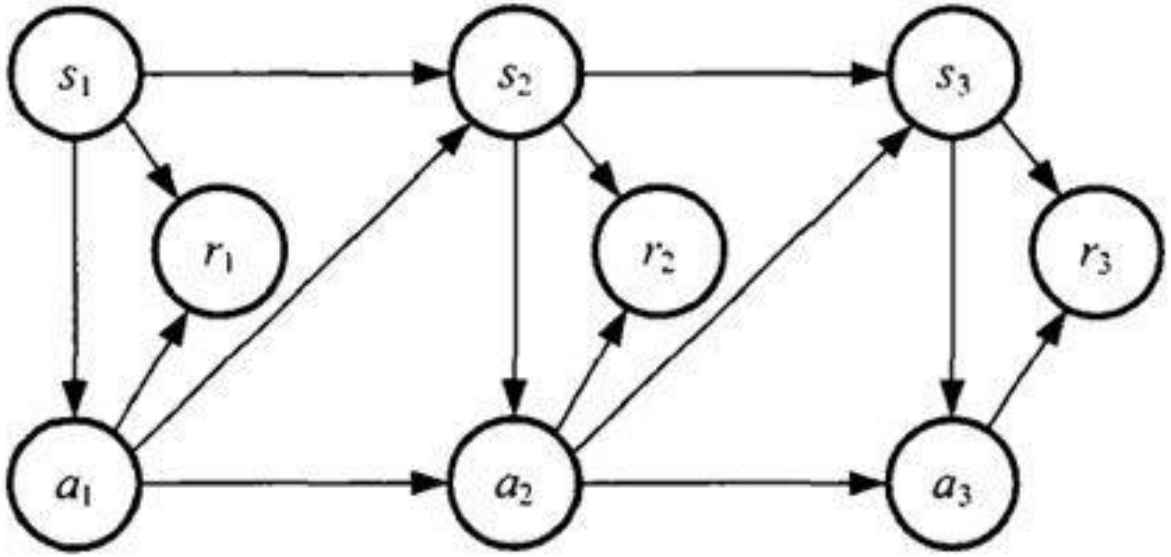
问题构建

- 马可夫决策过程：连续决策过程
  - 当前的决策对未来有深远影响
  - 例子：俄罗斯方块
  - 状态：当前的砖块堆积情况和下落砖块以及下一个砖块的形状
  - 动作：设定下落砖块的方向和位置
  - 奖励：一行或多行砖块被清除
  - 目标：清除尽可能多行的砖块
- 马可夫性质
  - 下一个状态和奖励只取决于当前状态和动作



马可夫决策过程 (MDP)

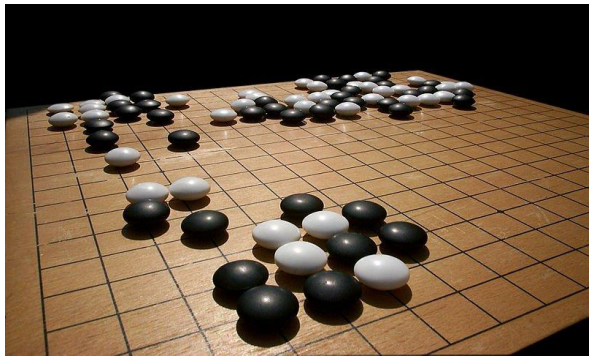
- $\max E[R_1 + R_2 + R_3 + \cdots]$ 最大化整个决策过程中的综合奖励
  - R ~ P(r | s, a)：奖励的获得遵循依赖于状态和动作的分布
- 字符表示
  - s: 状态
  - a: 动作
  - r: 奖励



其他例子

- 多臂老虎机
  - 每次只拉一个手杆
  - 最大化总赢

- 围棋
  - 每次在棋盘上放置一个子
  - 比对手完全包围棋盘上的一个更大的区域

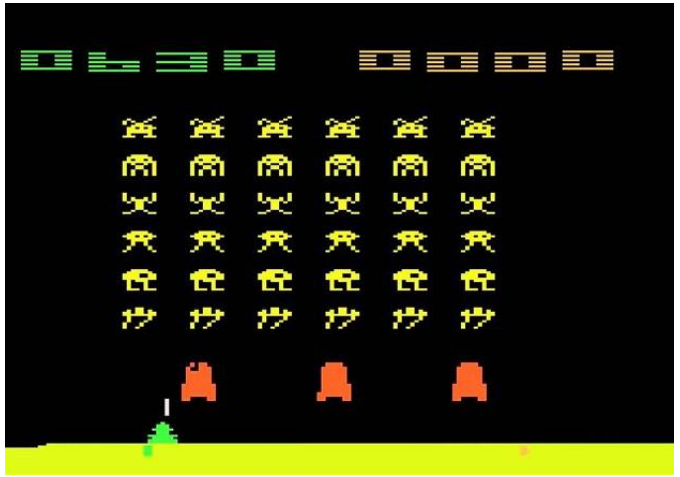


和其他机器学习范式的不同

- 监督学习
  - 监督学习中的标签是完全事实：一个订单是或不是欺诈
  - 增强学习中观测到的奖励只是部分含噪音的事实：俄罗斯方块中一个能消除一行砖的动作在将来多步后并不一定显得好
- 非监督学习
  - 非监督学习中没有任何标签信息：试图从数据本身发现数据下的模式
- 第三种范式
  - 从过往的经验中学习做对长远有益的策略

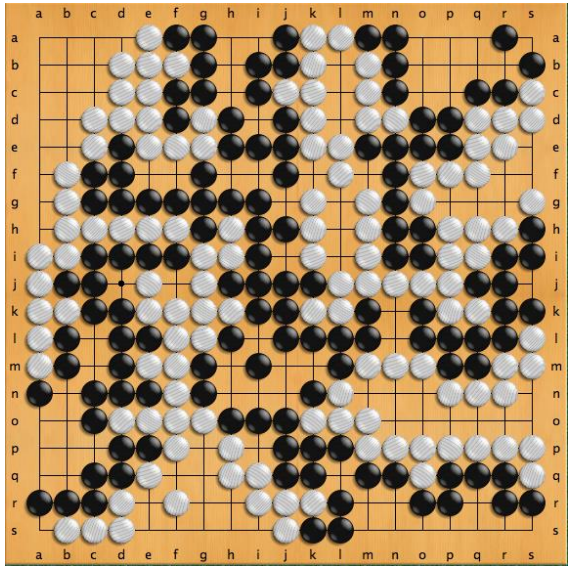
应用：亚塔利游戏

- 状态
  - 原始的游戏视频画帧，没有任何其他特征工程
- 动作
  - 手杆方向和开火键
- 奖励
  - 屏幕上的得分
- DQN
  - 状态作为网络输入，每一动作的价值函数作为一个输出
- 当前成绩
  - 在多数游戏中能打平或打败人类水平



应用: AlphaGo

- 状态：棋盘上子的位置
- 动作：下一子放置的位置
- 奖励：赢棋
- 执行-批判算法
  - 用深度神经网络同时维护价值网络和策略网络
- 蒙特卡洛树搜索
  - 一种以混合蒙特卡洛模拟方法和价值函数来评估和选择走子的方法
- 当前成绩
  - 打败了来自中国和韩国的世界冠军



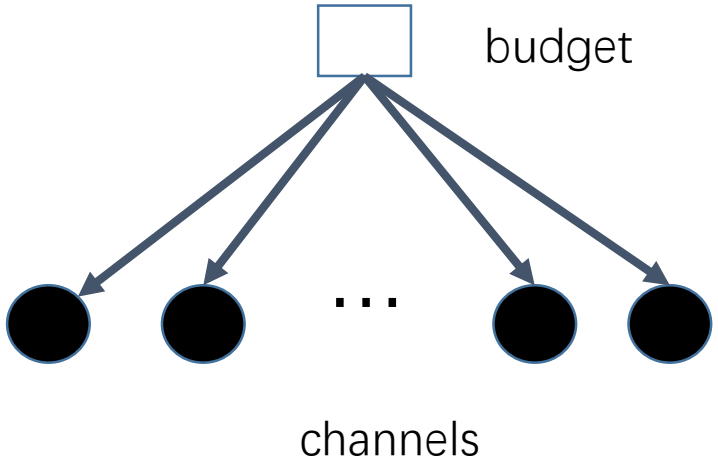
滴滴应用：司机激励机制

- 背景
  - 通过短信，发圈，最大化司机的生命周期价值
- 状态
  - 司机当前在平台的活动指标数据
- 动作
  - 发短信，发指定面值的券
- 奖励
  - 首单完成，后续订单的完成
- 批量RL，DQN

ROI提升超过40%

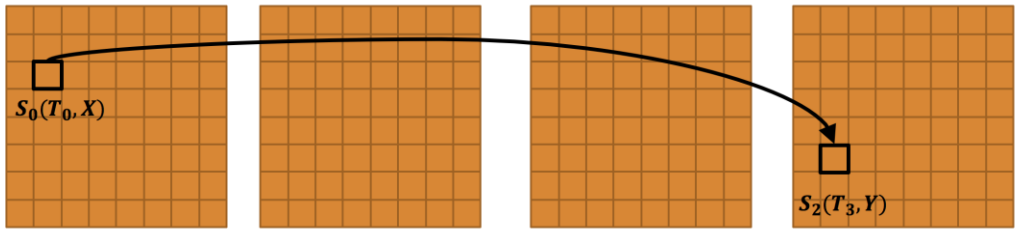
滴滴应用：市场渠道预算分配优化

- 背景
  - 多个市场营销（广告）渠道做投资，如百度，谷歌，360
  - 优化每天的预算分配，最大化长期的ROI
- 状态
  - 每个渠道的广告反馈指标
- 动作
  - 每个渠道投入的金额
- 奖励
  - ROI: 注册数量 / 投入金额
- 预算约束：所有渠道里的投入有上限
  - 各渠道上的决策是绑住的



滴滴应用：分单

- 智能体：司机
- 状态
  - 离散时空状态：（空间格子，时间格子）
- 动作：等待或接指定的定单
- 奖励：订单金额
- 策略
  - 一次对于多个司机同时指派动作
  - KM匹配算法：Q价值作为基本匹配系数



$$V(S_0) \leftarrow V(S_0) + \alpha(R + \gamma V(S_2) - V(S_0))$$

机器学习

Thank You

www.xiaojukeji.com

滴滴
滴滴一下 美好出行