

深度学习

在计算机视觉中的应用

周登文

控制与计算机工程学院

华北电力大学

What is computer vision?

我们人类很容易感受周围世界的三维结构



What is computer vision?

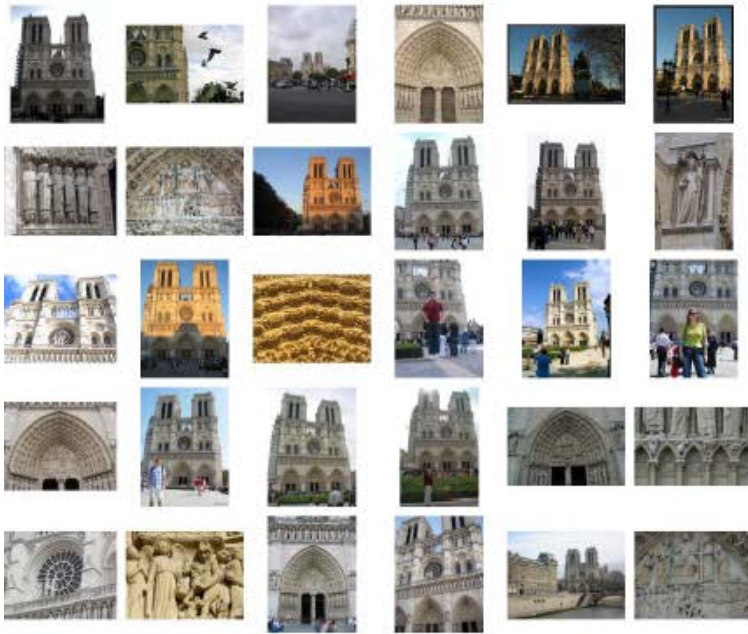
看一幅图像，我们很容易看出照片中有几个人，甚至可以从他们的面部外观，猜测他们的心情。

认知心理学家已经研究了几十年，试图了解人类视觉系统的工作方式，虽然已经知道了其中的一些原理，但是这个问题的完整解决方案仍然是未知的。

What is computer vision?

计算机视觉研究人员已经开发开发出一些数学技术，以恢复图像中物体的三维形状和外观。我们现在拥有可靠的技术，可以从几千个部分重叠的照片中，准确地计算一个环境中的部分3D模型。

What is computer vision?

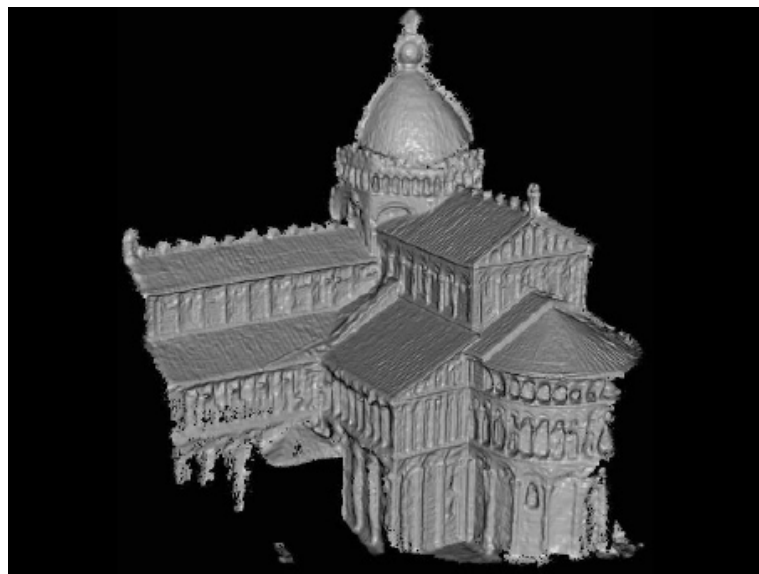


根据数百幅部分重叠的照片，运动算法可以重建大型复杂场景的**稀疏3D点模型的结构**

(Snavely, Seitz和Szeliski 2006) ©2006 ACM

What is computer vision?

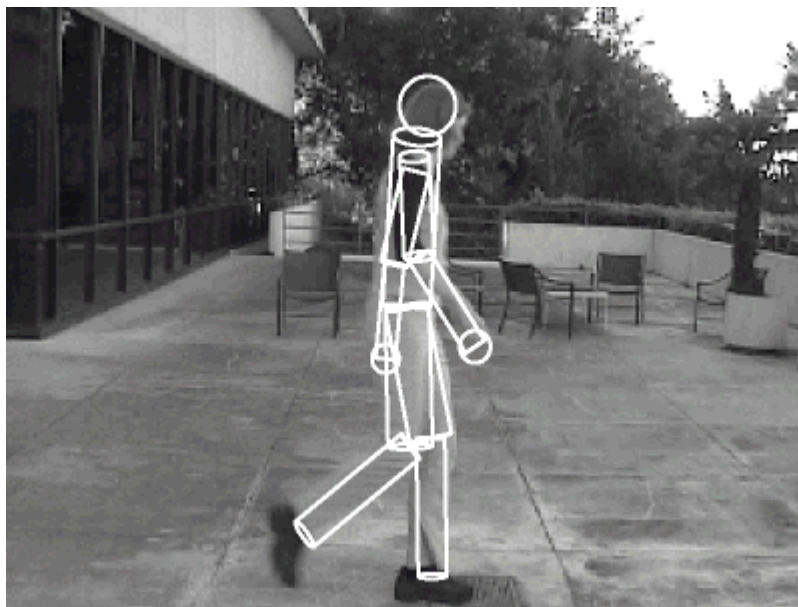
给定某个特定物体足够多的视图，可以使用**立体匹配**技术，精确地创建**稠密的3D表面模型**。



根据互联网上几百幅不同曝光的照片，立体匹配算法可以重建建筑物的**详细3D模型** (Goesele, Snavely, Curless et al. 2007) © 2007 IEEE

What is computer vision?

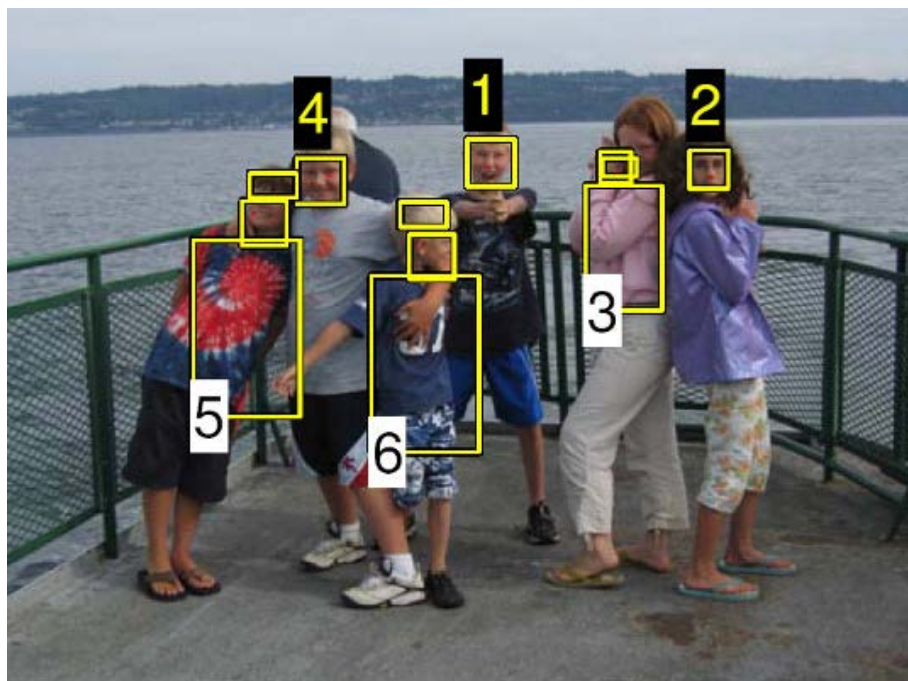
可以在复杂背景下，跟踪一个行人.



行人追踪算法可以跟踪在复杂背景下的行人(Sidenbladh, Black, and Fleet 2000) © 2000 Springer

What is computer vision?

甚至可以组合面部、衣着和头发，找到并命名照片中的所有人员以及他们的名字。



面部检测算法，加上彩色服装和头发检测算法，可以定位和识别图像中的个体(Sivic, Zitnick和Szeliski 2006) © 2006 Springer

What is computer vision?

尽管这些进步，使计算机解释图像，达到两岁小孩的水平，仍然是一个梦想(例如，统计图像中的所有动物)！

为什么计算机视觉如此困难？这是由于计算机视觉是一个**逆问题**：我们试图通过不足的信息，恢复一些未知量，以完全指定解决方案。

What is computer vision?

我们需要基于物理的和概率的模型，去除潜在解决方案中的歧义，而复杂的视觉世界建模非常困难.

在计算机视觉中使用的正向模型，通常在物理(辐射度学、光学和传感器设计)以及计算机图形学中开发.

What is computer vision?

这两个领域均**建模**物体如何运动，光线在物体表面如何反射，在大气中如何散射，通过相机镜头(或人眼)的折射，最后投影到平坦的(或弯曲)的图像平面上。计算机图形学还不完美(还没有能完全替代人类的计算机动画电影)，在有限的领域，如渲染一个由日常物品组成的静态场景，或动画展现灭绝的生物(如恐龙)，是完美的。

What is computer vision?

在计算机视觉中，正好反过来，即描述在一个或多个图像中看到的東西，并重建其属性，如形状、光照和颜色分布。令人吃惊的是，人类和动物做到这一点毫不费力，而计算机视觉算法则容易出错。

计算机视觉应用

◆ 自动驾驶

◆ 医学成像

◆ 物体识别

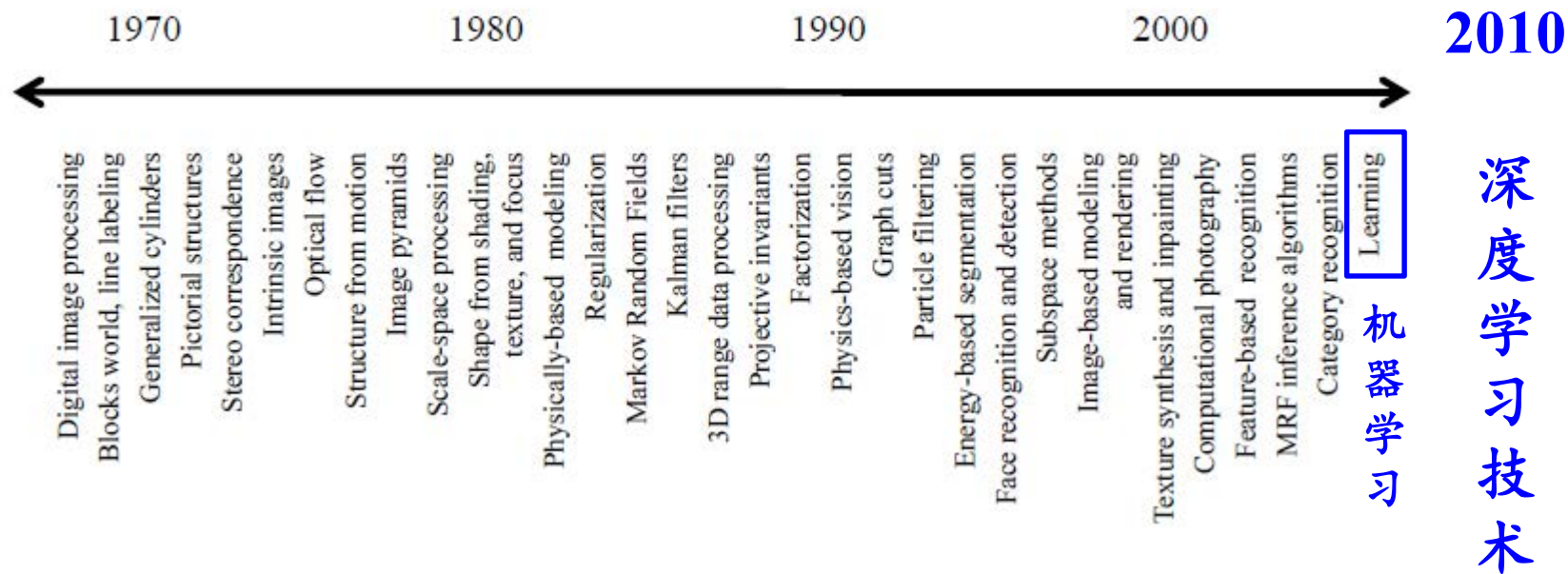
◆ 运动分析

◆ 场景重建

◆ 图像恢复

...

计算机视觉研究中一些最活跃的主题



计算机视觉大致起源于20世纪70年代

计算机视觉研究中一些最活跃的主题

20世纪**80年代**，主要聚焦于复杂的数学技术，以进行定量的图像和场景分析.

20世纪**90年代**，以前提出的主题不断被探索，其中的一些变得更加活跃.

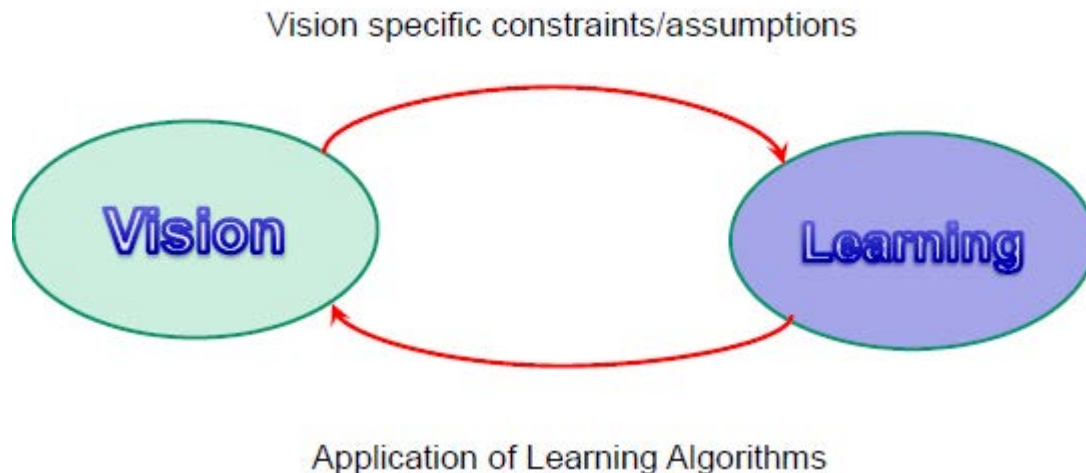
计算机视觉研究中一些最活跃的主题

21世纪头十年，视觉和图形领域的相互作用不断加深。例如，图像缝合(image stitching)，光场捕获和渲染(light-field capture and rendering)，通过多次曝光，捕捉高动态范围(HDR: high dynamic range)图像。一个重要的趋势是：**将尖端的机器学习技术应用于计算机视觉问题**。

计算机视觉研究中一些最活跃的主题

最近几年，已目击了在计算机视觉中应用深度学习技术的爆炸性增长！深度学习几乎成了计算机视觉研究的标配，没有用到深度学习技术的计算机视觉论文，几乎进不了ICCV和CVPR.

计算机视觉研究中一些最活跃的主题



虽然只有短短的几年，深度学习已革新了计算机视觉领域！在人脸识别、图像分类、物体跟踪和检测等几乎所有的主题，都提升了计算机视觉算法的性能！

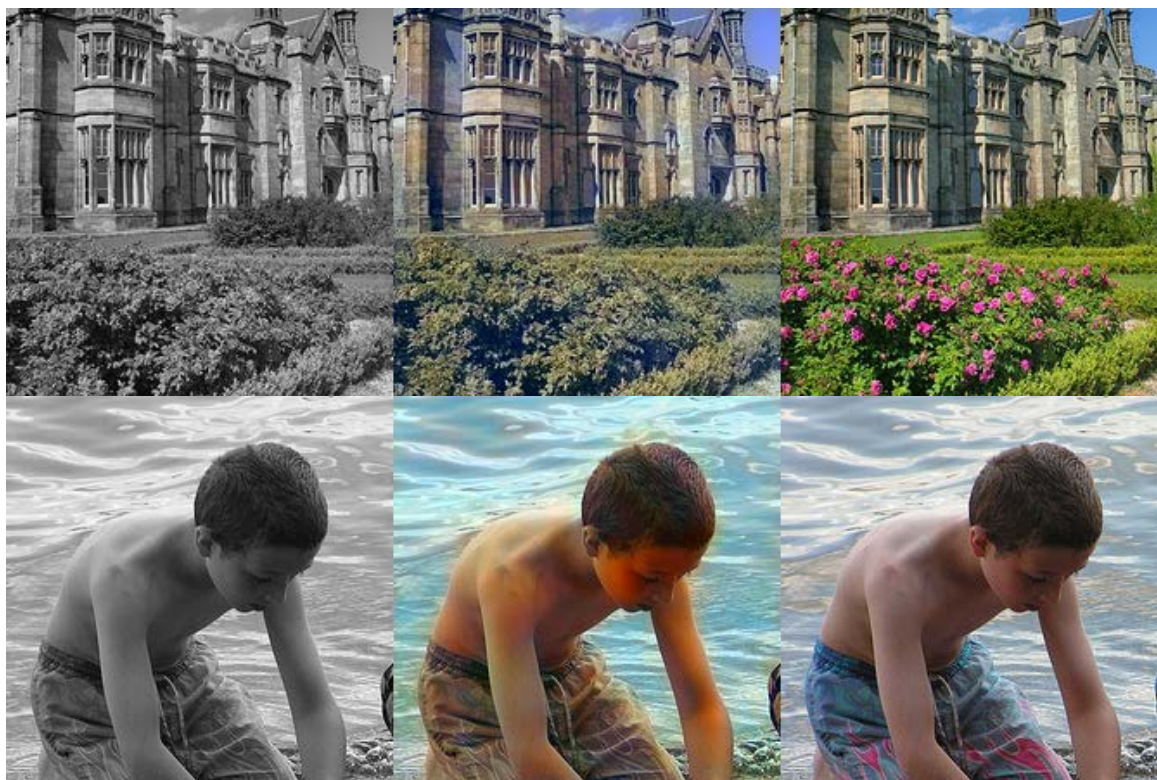
几个有趣的应用

黑白
图
像
自
动
着
色

Grayscale

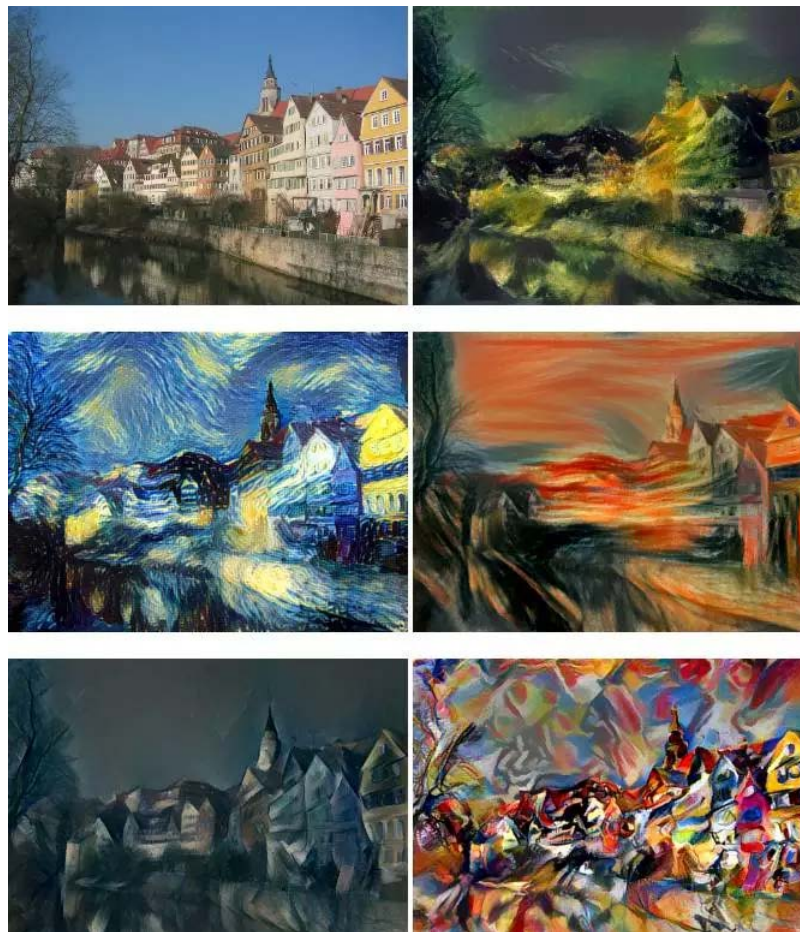
Prediction

Ground Truth



几个有趣的应用

图像风格转换



几个有趣的应用

图像超分辨率



几个有趣的应用

图
像
字
幕
自
动
生
成

A person on a beach flying a kite.



A black and white photo of a train on a train track.



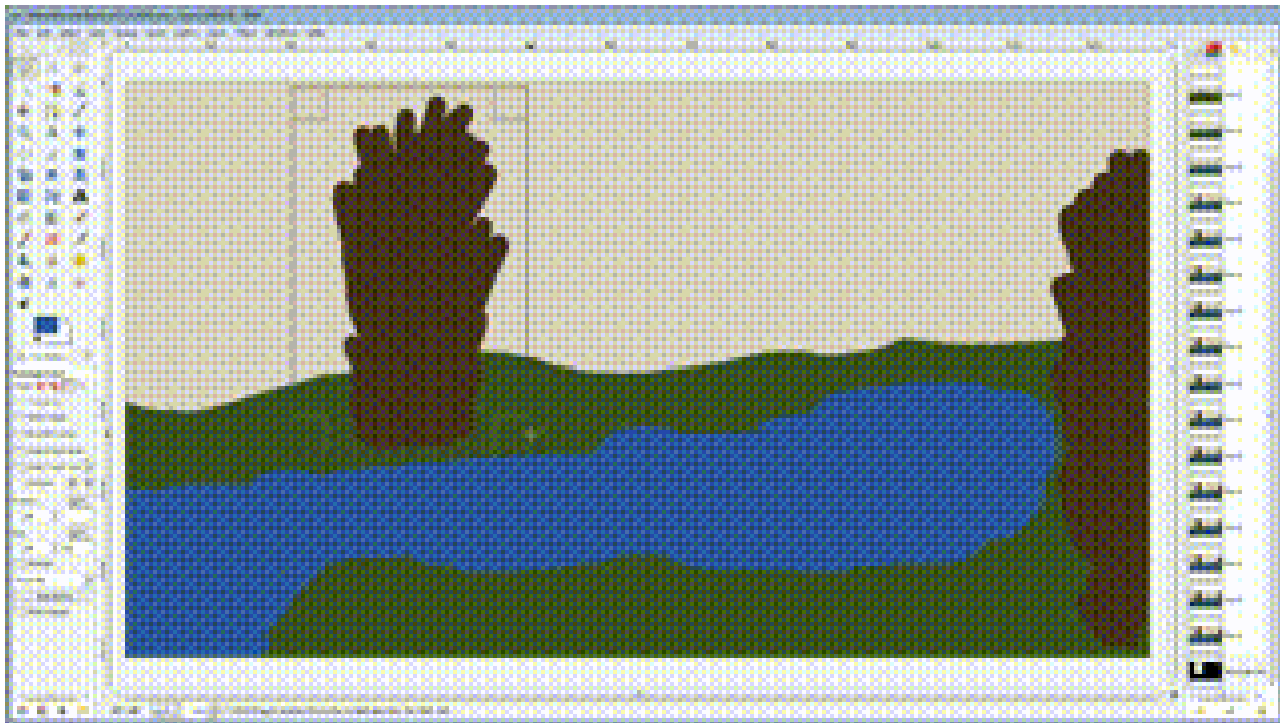
A person skiing down a snow covered slope.



A group of giraffe standing next to each other.

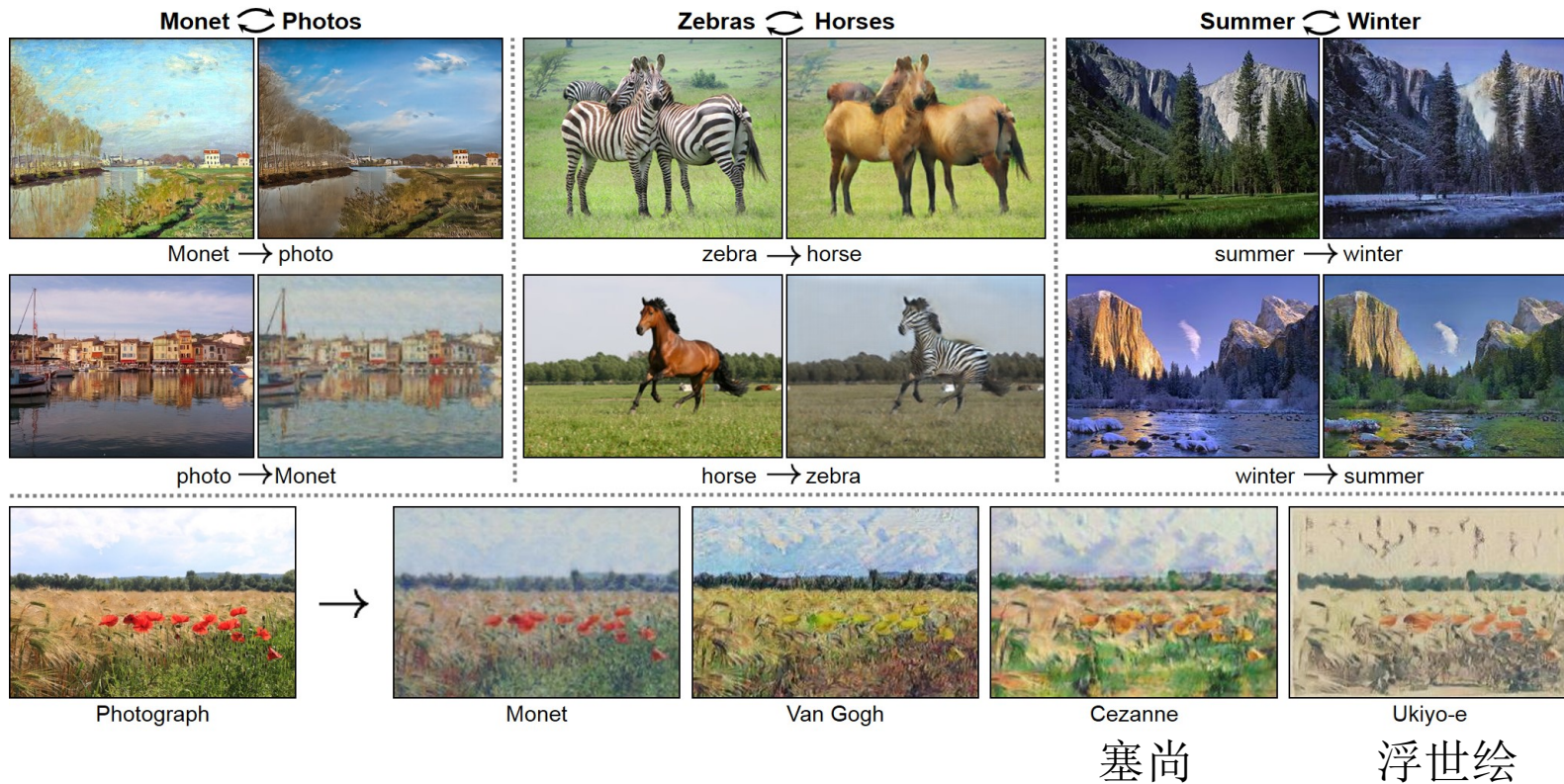


几个有趣的应用



从粗糙的草图自动创建图像

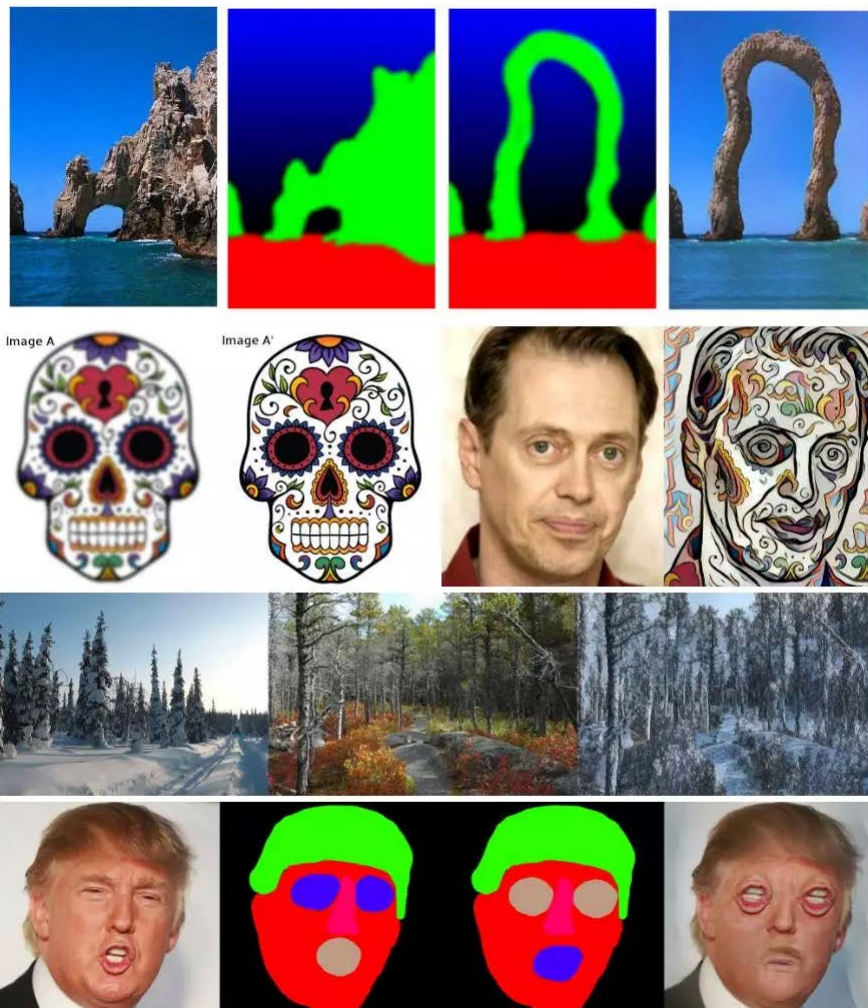
几个有趣的应用



视觉效果转换

几个有趣的应用

图像类推



What is deep learning?

What is deep learning?

“Deep learning allows computational models that are composed of **multiple processing layers** to **learn representations of data with multiple levels of abstraction.**”

– Yann LeCun, Yoshua Bengio and Geoff Hinton

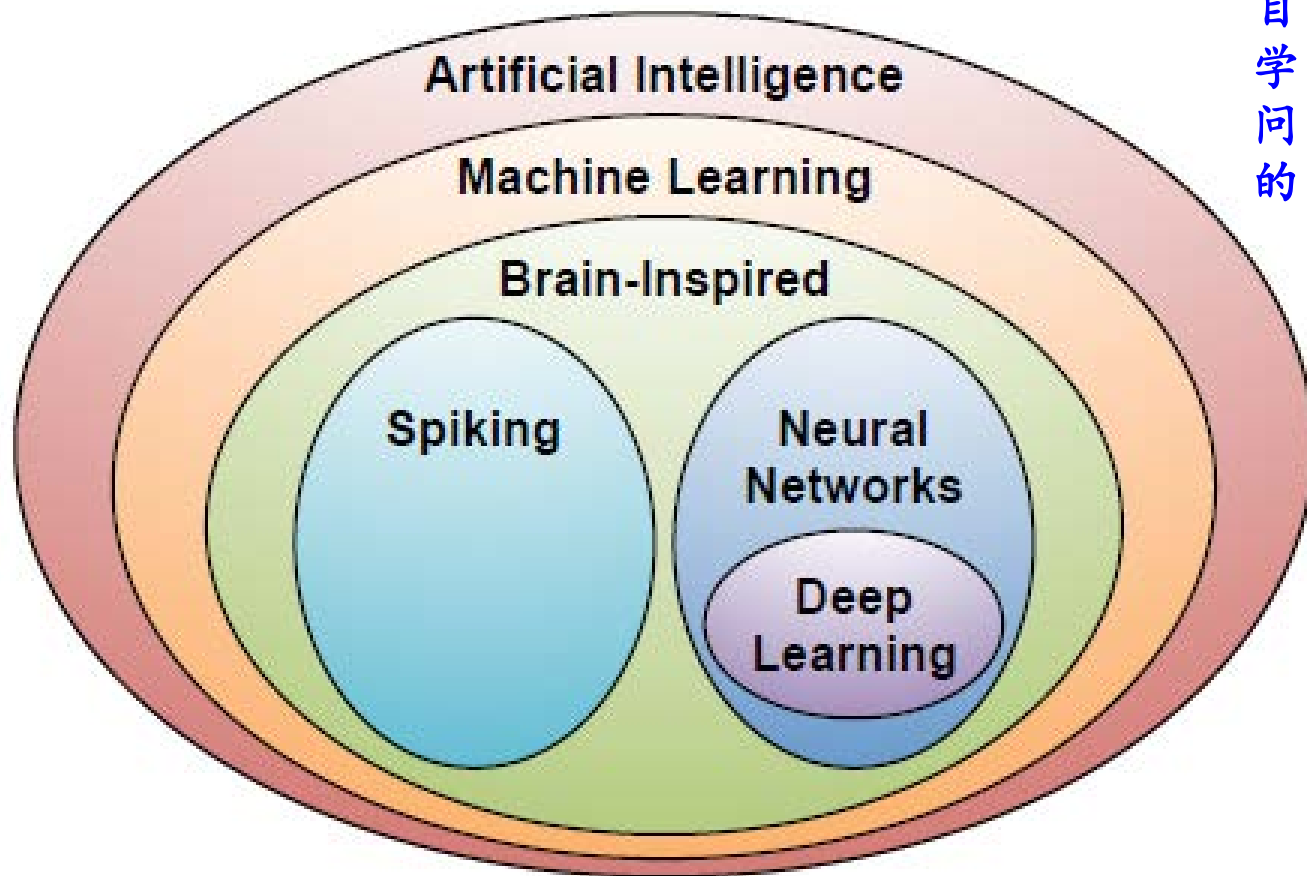
Y. LeCun, Y. Bengio, G. Hinton, "Deep Learning", Nature, Vol. 521, 28 May 2015



深度学习在人工智能中的位置

深度神经网络(也称为深度学习)是人工智能大框架下的一部分. 人工智能的目标是创造具有类似人类能力的智能机器. 人工智能这个术语是计算机科学家John McCarthy在20世纪50年代提出来的.

深度学习在人工智能中的位置



目前大脑是
学习和解决
问题的最好
的“机器”

Deep Neural Networks

虽然科学家们仍然在探索大脑如何工作的细节，但普遍相信：**大脑的主要计算元素是神经元**。人脑平均大约有860亿个神经元。神经元自身与许多进入它们的**树突(dendrites)**和离开它们的**轴突(axon)**元素联系在一起。

Deep Neural Networks

神经元接受通过树突进入的信号，对这些信号进行计算，并在轴突上产生信号。一个神经元的轴突分支出来，并连接到许多其它神经元的树突上。轴突的分支和树突之间的连接称为**突触 (synapse)**。据估计，人脑中平均有 $10^{14} \sim 10^{15}$ 个突触。

Deep Neural Networks

突触的一个关键特征是它可以伸缩通过它的值。伸缩因子称为权，大脑学习的方式被认为是：通过改变与突触相关的权。因而，不同的权重导致对输入的不同响应。学习就是调整权重，作为学习刺激的响应，大脑的组织不会改变。

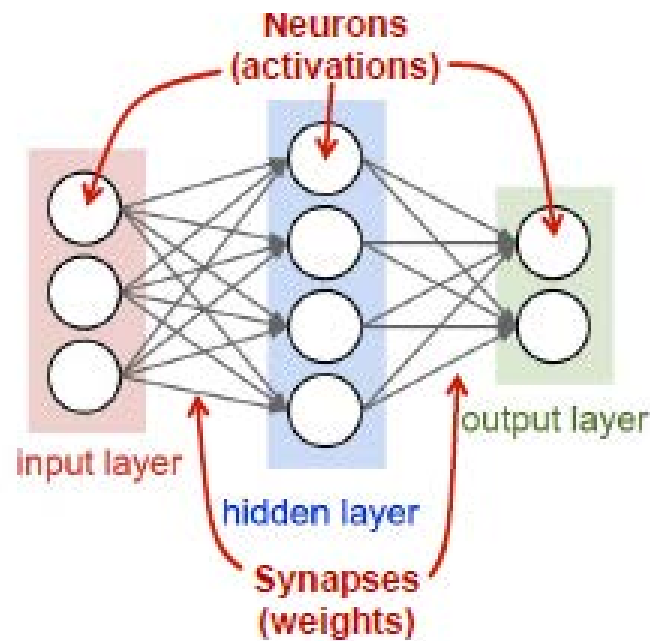
Deep Neural Networks

在大脑启发的计算区域中，有一个称为**尖峰计算**(spiking computing)的子区域。在这个子区域中，启发源自于下面的事实：**树突和轴突上的通讯是尖峰脉冲**，所传递的信息不仅仅是尖峰的幅度，它还依赖于脉冲到达的时间，并且，在神经元中发生的计算不是单个值，而是脉冲宽度和不同脉冲之间的时序关系的函数。

Deep Neural Networks

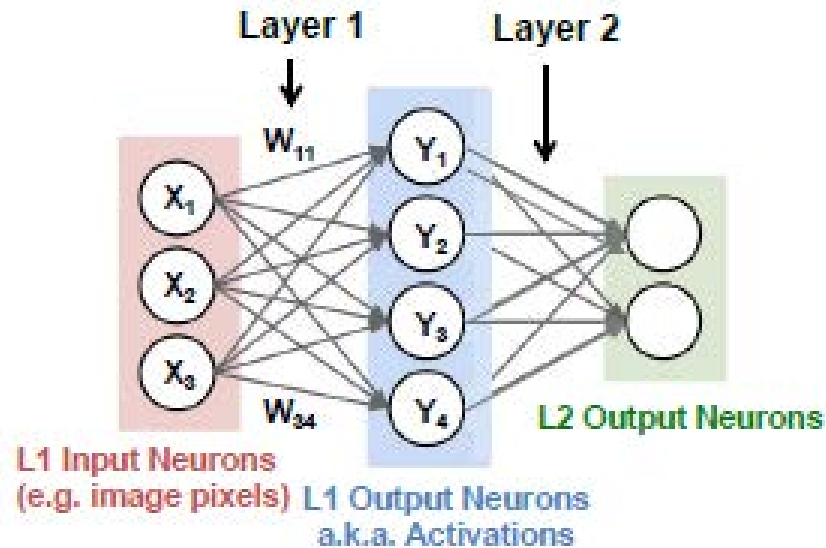
神经网络从下面的概念中获得启发：神经元的计算涉及计算输入值的加权和。这些加权和对应于被突触伸缩的值，以及神经元中这些值的组合。此外，神经元不仅输出加权和，在神经元内组合输入上存在函数运算，这个运算似乎是一个非线性函数：只有当输入跨越某个阈值时，才能使神经元生成输出。

Deep Neural Networks



神经元和突触

Deep Neural Networks

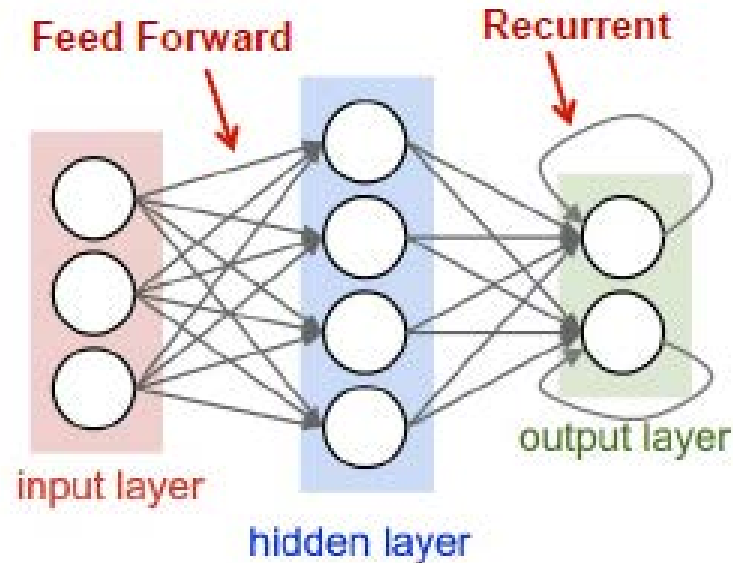


一个计算实例: $y_j = f\left(\sum_{i=1}^3 w_{ij} \times x_i\right)$

w_{ij} : 权

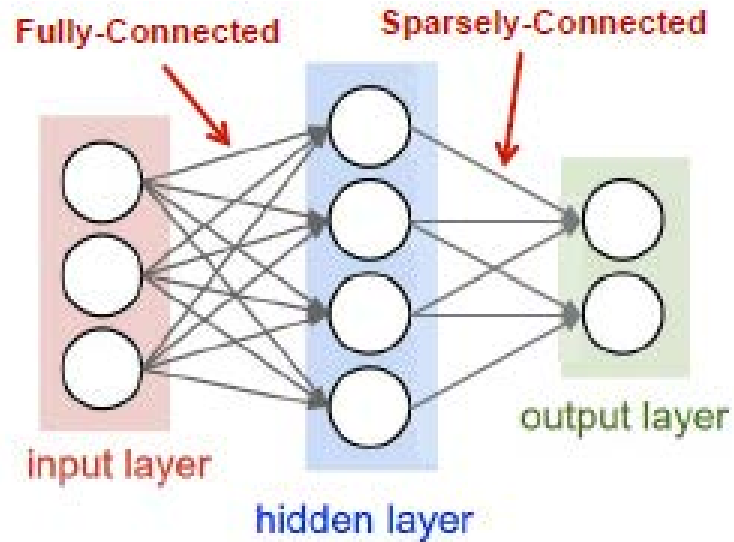
x_i : 输入激活 y_j : 输出激活 $f(\cdot)$: 非线性激活函数

Deep Neural Networks



前馈与反馈(递归)网络

Deep Neural Networks

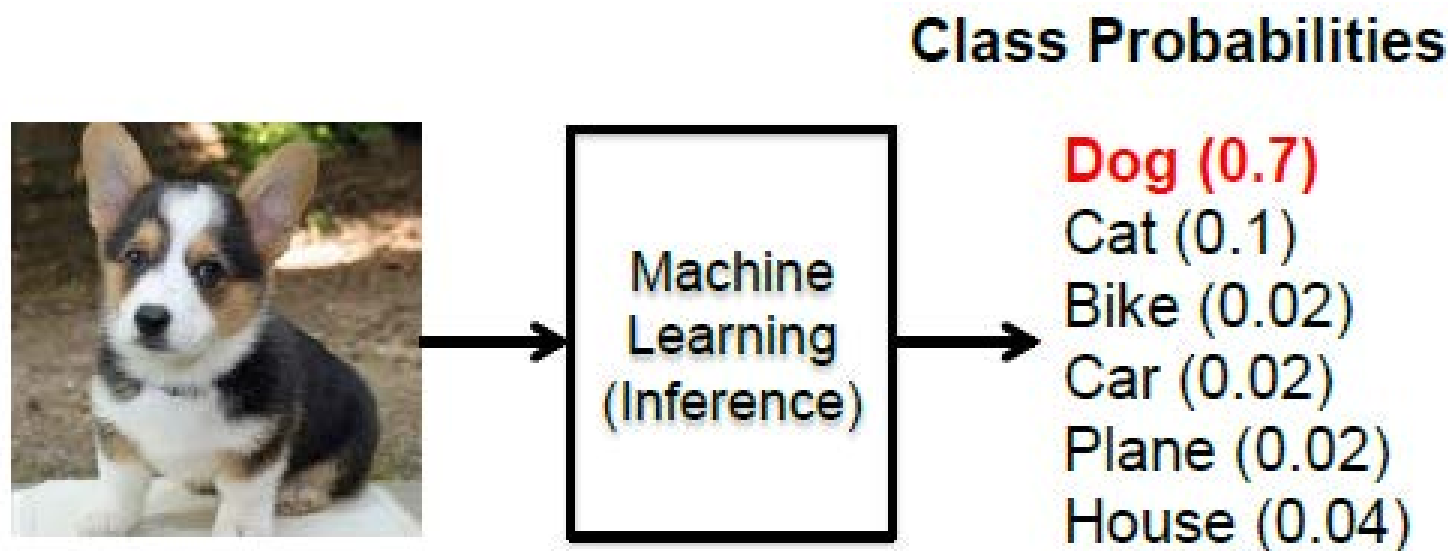


完全连接与稀疏连接网络

推理与训练

在**深度神经网络(DNN)**中，确定网络中权重的值的过程称为**学习**，学习的过程也称为**训练**(training)网络。一旦训练完成，使用训练确定的权重，进行网络计算，使用网络执行任务的计算称为**推断**(inference)。

推理与训练

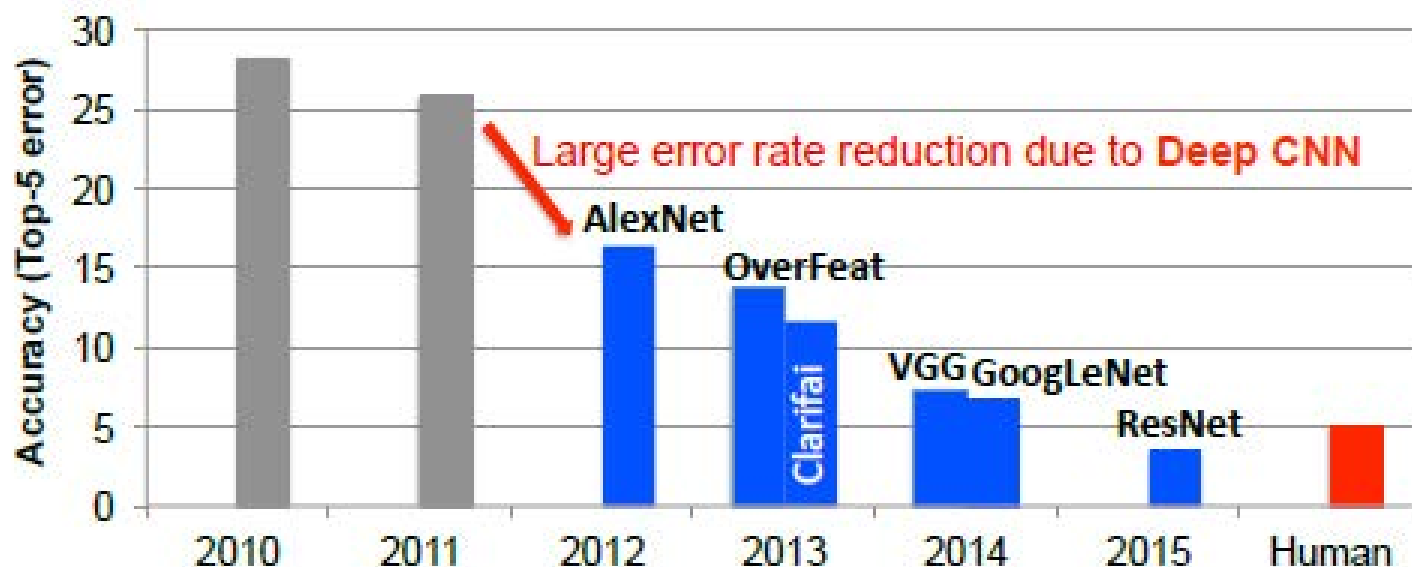


图像分类任务实例. **训练集**由120万个图像组成, 每个图像都标有图像包含的1000个物体类别之一. 然后, 必须准确地识别测试图像中的物体.

DNN发展历史

- 1940年代：神经网络被提出
- 1960年代：深度神经网络被提出
- 1989年：识别数字的神经网络出现(LeNet)
- 1990年代：浅层神经网络硬件出现(Intel ETANN)
- 2011年：基于DNN的语音识别获得突破性进展(Microsoft)
- 2012年：DNN在视觉中，开始的替代了手工制作的方法
(AlexNet)
- 2014+：DNN加速器研究的兴起

DNN发展历史

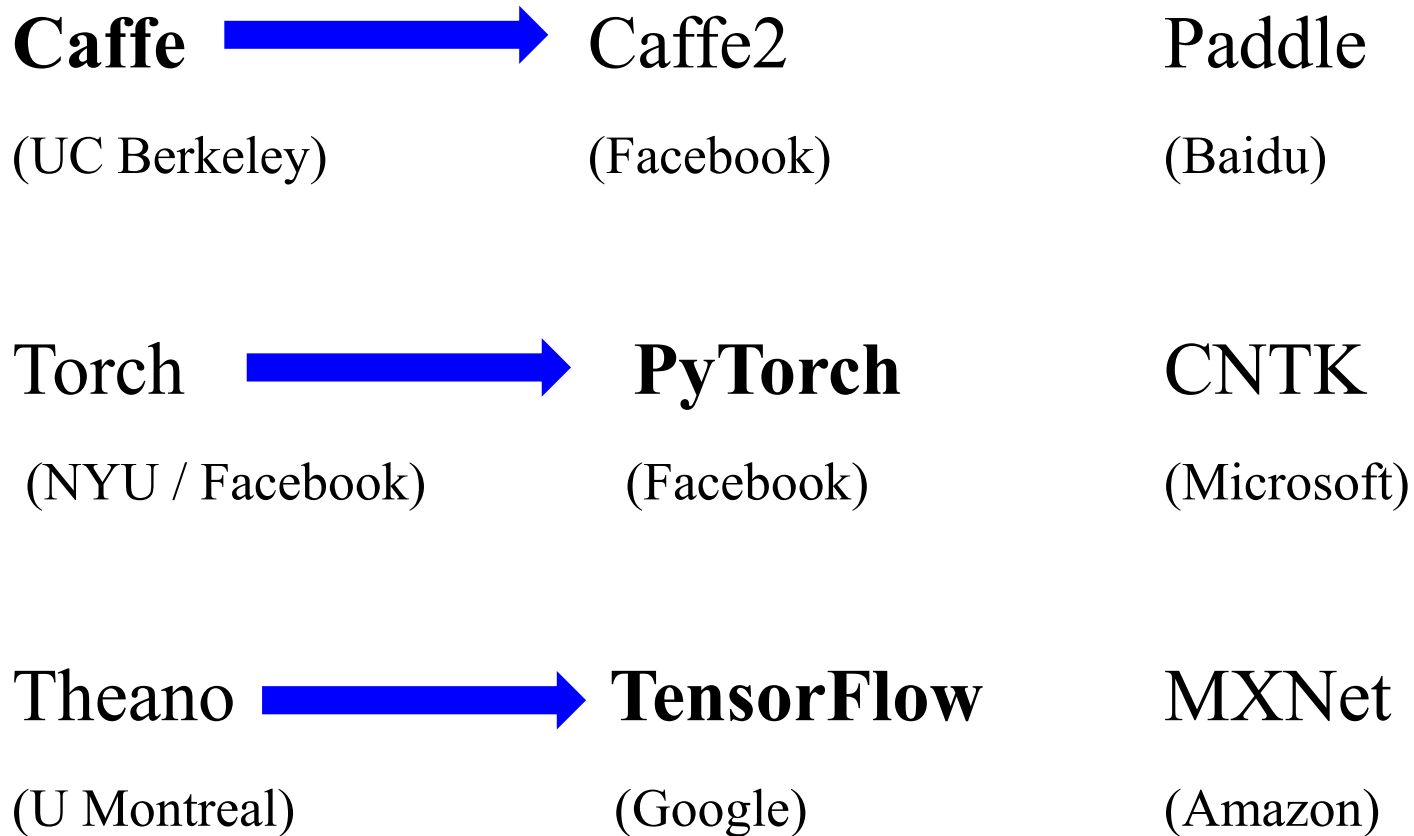


最近几年，**ImageNet**大赛中，最佳参赛者的表现

Frameworks

为了便于DNN的开发，以及共享训练的网络，已经开发出了几个**深度学习框架**。加州大学伯克利分校开发的**Caffe**，在2014年开源。它支持C，C ++，Python和MATLAB接口。Google在2015年发布了**TensorFlow**，它支持C ++和Python接口。另一个流行的框架是**Torch**，由Facebook和NYU开发，支持C、C ++和Lua接口。

Frameworks



Developed by U Washington, CMU, MIT, Hong Kong U, etc but main framework of choice at AWS

Frameworks

- **TensorFlow**对于大多数项目来说，是一个安全的选择. 有巨大的社区，广泛的使用.
- **PyTorch**最适合研究. 不过还是新推出的，不完善
- 考虑**Caffe**，**Caffe2**或**TensorFlow**进行生产部署
- 考虑**TensorFlow**或**Caffe2**用于移动设备

Hardware for DNN Processing



两个关键助推引擎：
大数据和高性能处理器

(单精度浮点运算性能
达到**7 TFLOP/s**)

NVIDIA GeForce GTX TITAN X,
2015年3月推出, 人民币8000元

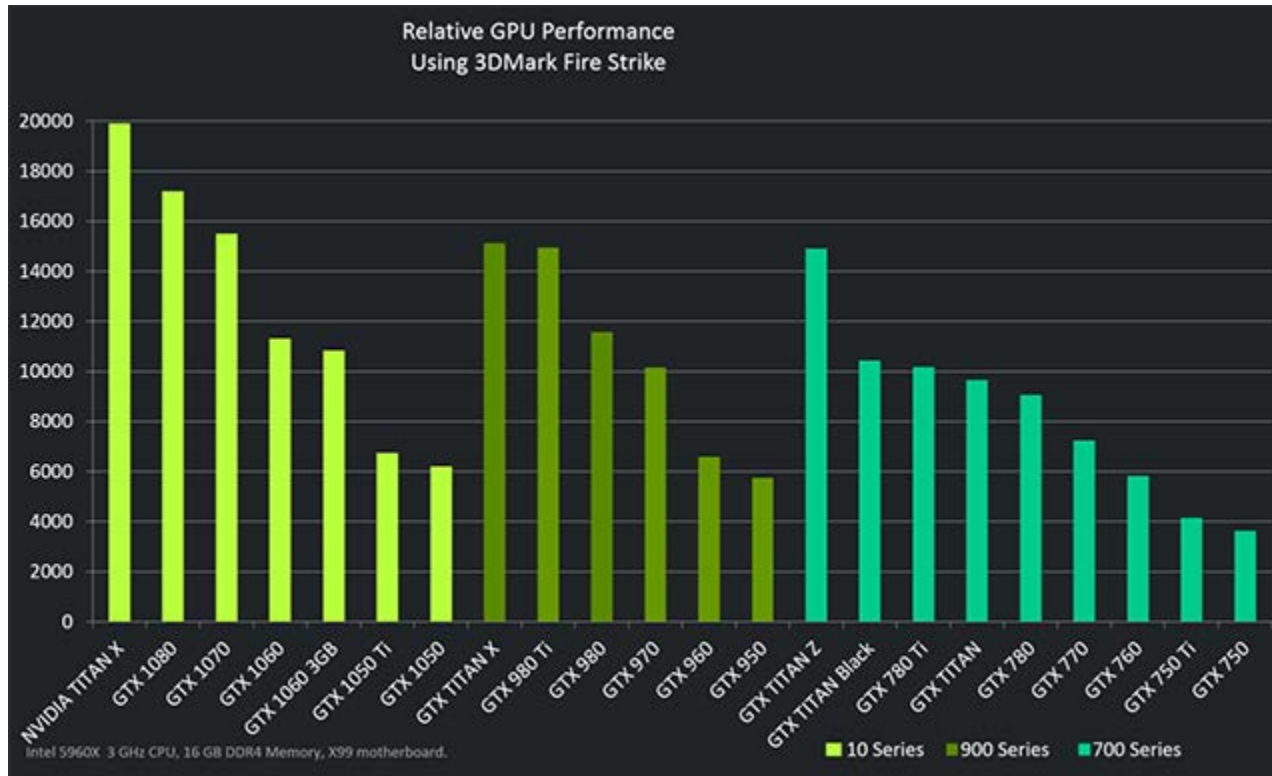
Hardware for DNN Processing



(单精度浮点运算性能达到
11 TFLOP/s)

NVIDIA TITAN X, 2016年7月推出, 人民币10000元

Hardware for DNN Processing



Hardware for DNN Processing



2017年5月16日，股价134.31US\$，市值798.52亿US\$
3月份以来，股价几乎又上涨了30%！

Hardware for DNN Processing

2017年5月10日，在美国加州圣何塞，NVIDIA主持的GPU技术大会(GTC)上宣布：推出新款NVIDIA® **DGX-1V**(配置8块 Tesla V100 GPU)深度学习超级计算机，单台DGX-1V的预订价格是14.9万美元。

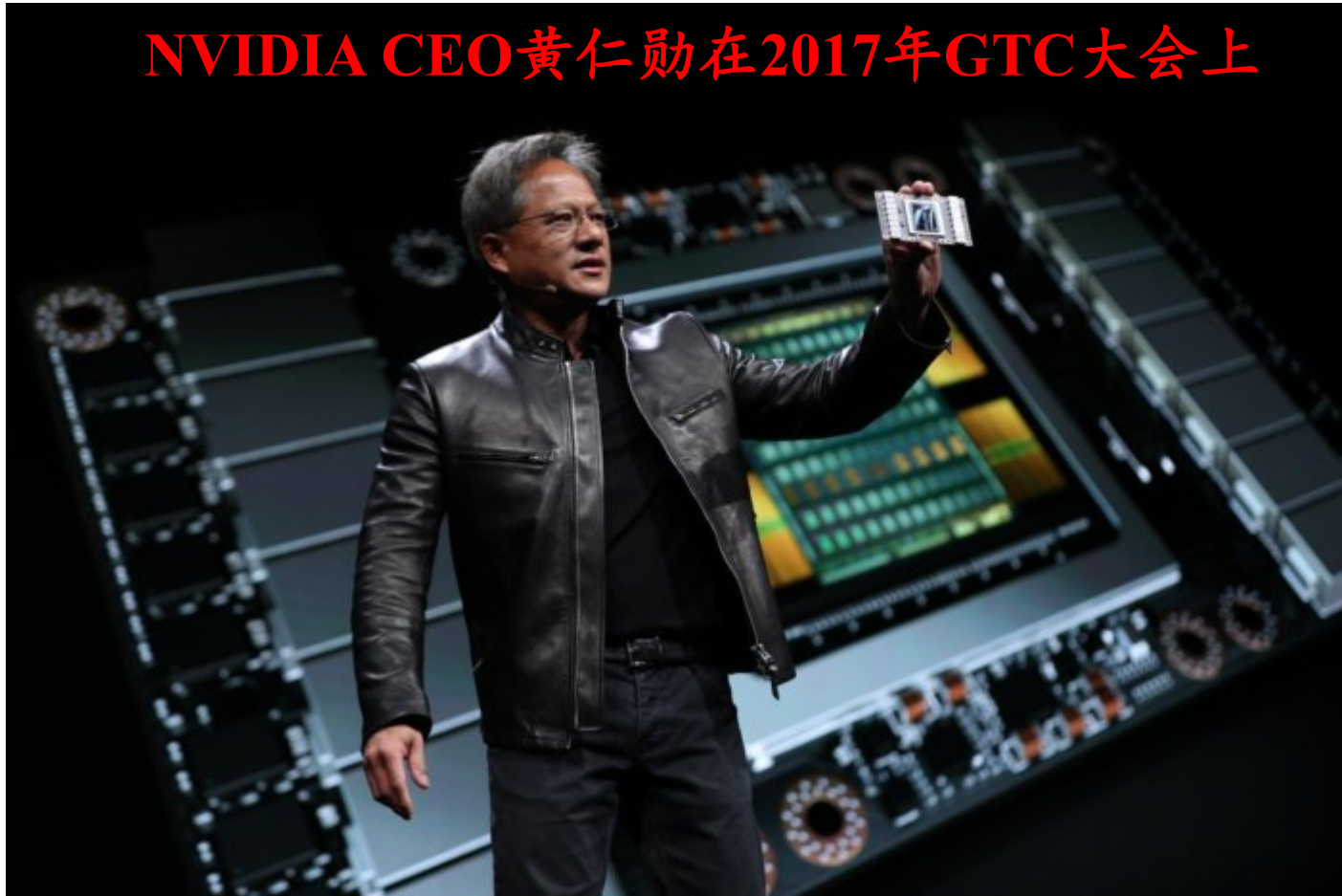
在2016年硅谷的GTC上，NVIDIA发布了深度学习超级计算机**DGX-1**，配置8块**Tesla P100** GPU(16GB 显存)，配置7TB固态硬盘用于储存神经网络训练中的大量数据，整体性能相当于250台普通x86服务器！单台DGX-1的售价是12.9万美元。

Hardware for DNN Processing

DGX-1V深度学习计算机同DGX-1相比，主要是将基于Pascal™ 架构的Tesla P100 GPU，替换为基于 **Volta™ 架构的Tesla® V100** GPU (单精度浮点运算性能达到**15 TFLOP/s**). 实际数据测试：Tesla V100 GPU的计算速度，大约是Tesla P100 GPU的8倍！

Hardware for DNN Processing

NVIDIA CEO黄仁勋在2017年GTC大会上



Hardware for DNN Processing

从**Volta**开始，NVIDIA开始all-in-AI. 这已不再是传统意义上的GPU. 它现在包含一个全新的核心，叫**Tensor core**，完全用于深度学习的计算. 今年发布的基于Volta的**Tesla V100**与上一代Tesla P100相比，单精度浮点运算性能提升了41.5%，由于引入了Tensor单元，在深度学习方面的性能可以达到后者的12倍！老黄太牛了！

华人对深度学习的贡献

- 2011年，**邓力**运用深度学习技术，在语音识别上的巨大成功，助推了深度学习技术的大潮
- 2013年，**余凯**创立深度学习技术研究院，在当时，可谓惊世骇俗
- 2014年**Caffe**开源，使得计算机视觉领域的研究全面深化，至今，仍是最受欢迎的平台之一

华人对深度学习的贡献

- 2015年，**Resnet**使得深度学习网络名副其实，现已成为深度学习的标配
- **MXNet**于2015 年 9 月开源，是目前技术上最为精妙的深度学习平台，无缝支持命令式编程(imperative programming) 和符号编程(symbolic programming)，已成为亚马逊官方深度学习平台