

第 8 章 深度学习

教材： 王万良 《人工智能导论》（第4版）

<https://www.icourse163.org/course/ZJUT-1002694018>

社区资源： <https://github.com/Microsoft/ai-edu>

参考资料： 《深度学习入门》

Outline

- 深度学习的小历史
- 深度学习的高速化
- 深度学习的未来

深度学习的小历史

- 深度学习之所以受到大量关注，其契机是 2012 年举办的大规模图像识别大赛 ILSVRC（ImageNet Large Scale Visual Recognition Challenge）
- 基于深度学习的方法（通称 **AlexNet**）以压倒性的优势胜出，彻底颠覆了以往的图像识别方法

ImageNet

- 拥有超过 100 万张图像的数据集
 - 它包含了各种各样的图像，并且每张图像都被关联了标签（类别名）

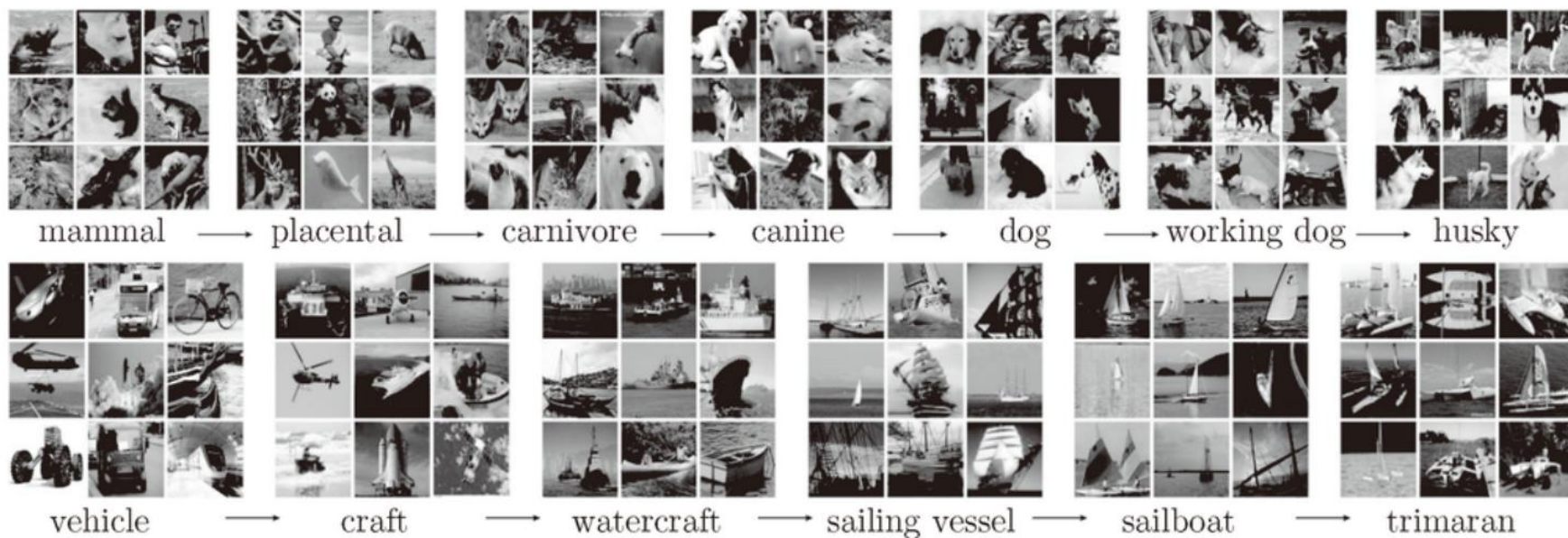


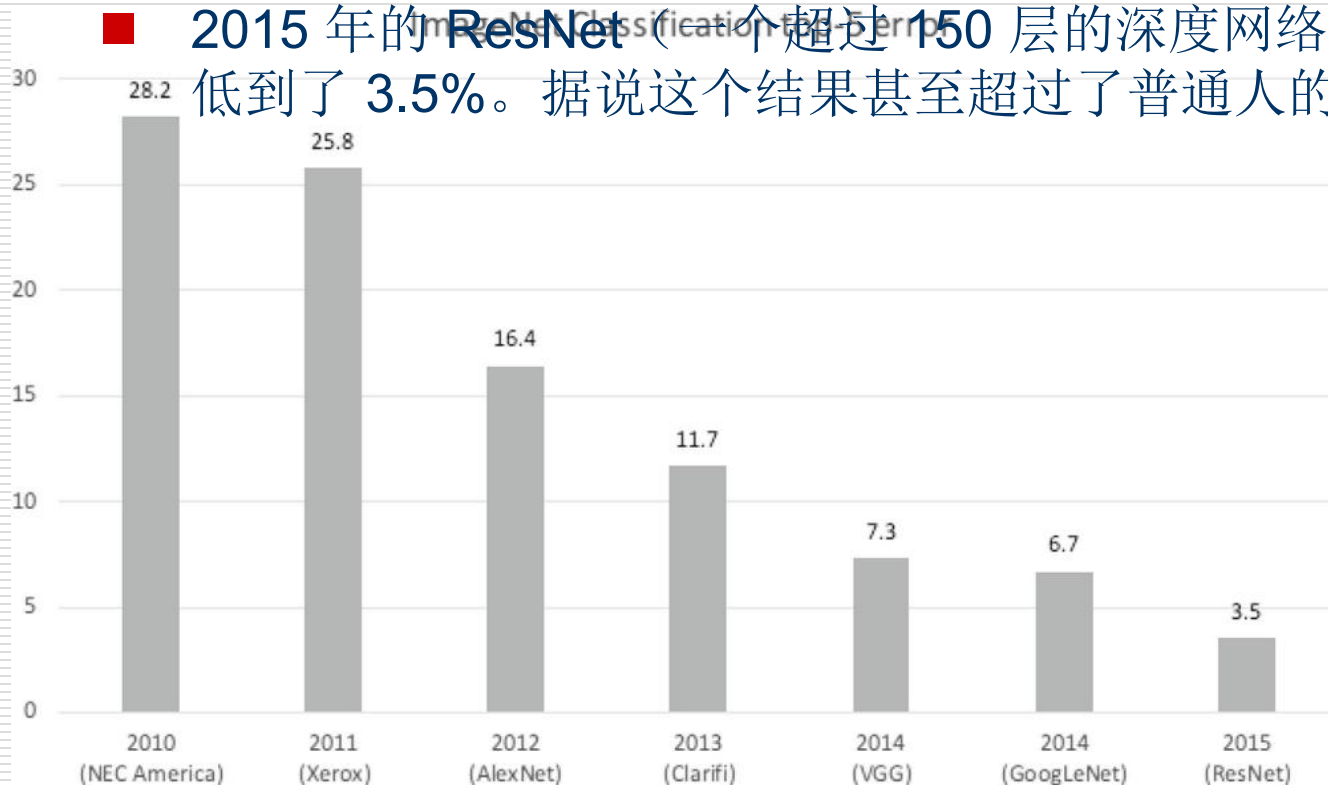
图 8-7 大规模数据集 ImageNet 的数据例（引用自文献 [25]）

ImageNet

❑ ILSVRC 大赛有多个测试项目，其中之一是“类别分类”（**classification**），在该项目中，会进行 1000 个类别的分类，比试识别精度

■ 从 2010 年到 2015 年的优胜队伍的成绩。这里，将前 5 类中出现正确解的情况视为“正确”，此时的错误识别率用柱形图来表示

■ 2015 年的 ResNet（一个超过 150 层的深度网络）将错误识别率降低到了 3.5%。据说这个结果甚至超过了普通人的识别能力

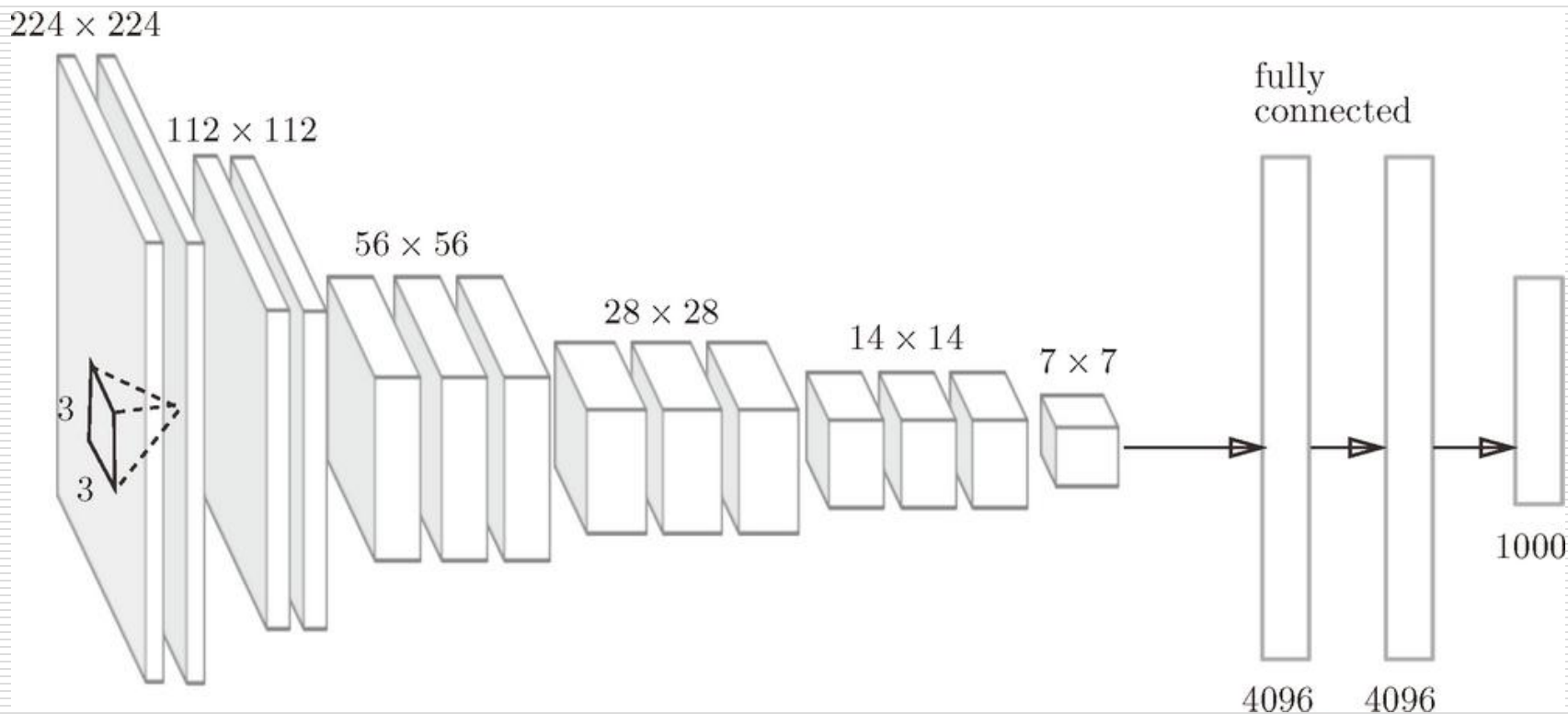


□ VGG 是由卷积层和池化层构成的基础的 CNN

- 特点在于将有权重的层（卷积层或者全连接层）叠加至 16 层（或者 19 层），具备了深度（根据层的深度，有时也称为“VGG16”或“VGG19”）
- 基于 3×3 的小型滤波器的卷积层的运算是连续进行的。
- 重复进行“卷积层重叠 2 次到 4 次，再通过池化层将大小减半”的处理，最后经由全连接层输出结果

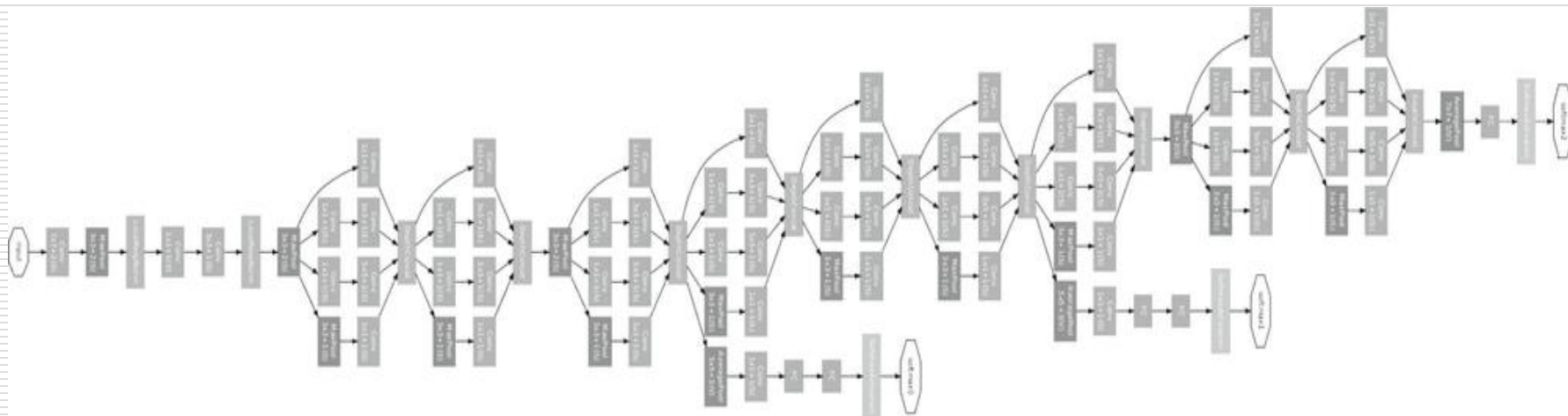
VGG

- VGG 是由卷积层和池化层构成的基础的 CNN
 - VGG 在 2014 年的比赛中最终获得了第 2 名的成绩。虽然在性能上不及 GoogleNet，但因为 VGG 结构简单，应用性强，所以很多技术人员都喜欢使用基于 VGG 的网络



GoogLeNet

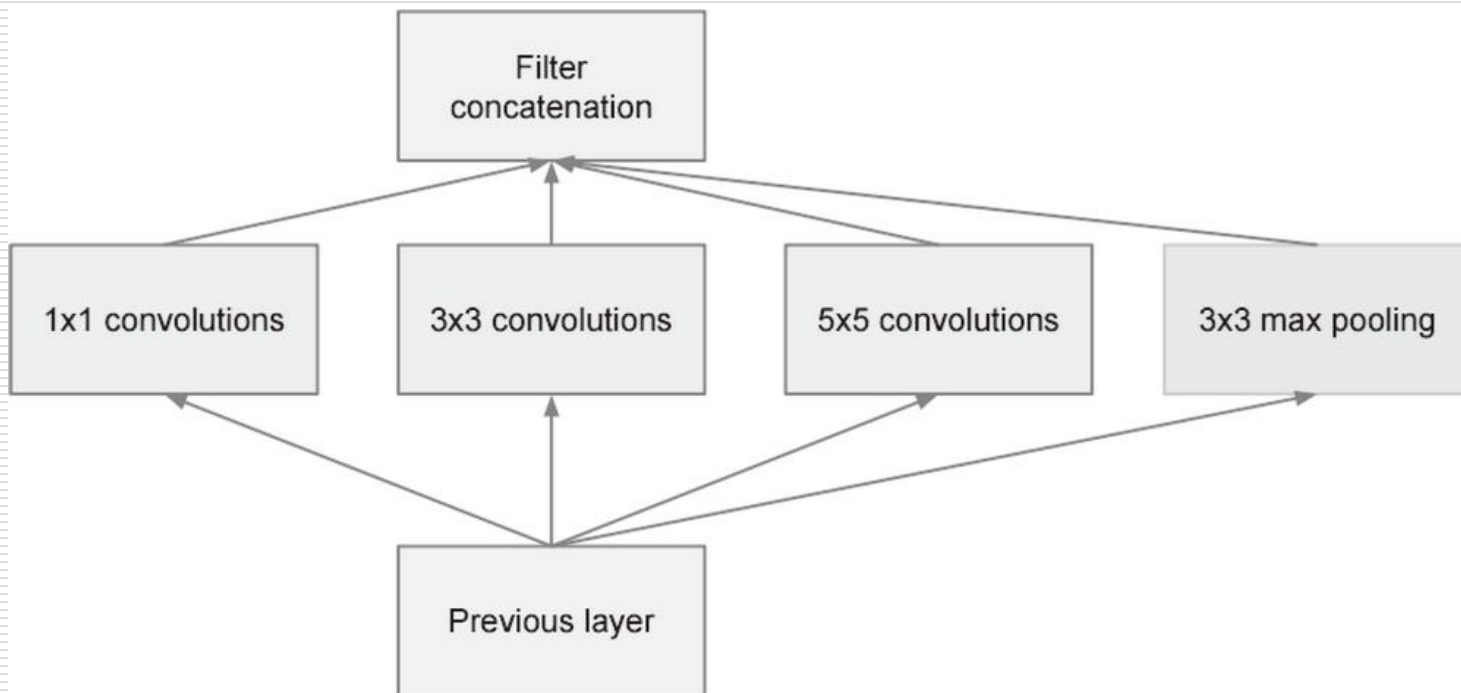
□ 图中的矩形表示卷积层、池化层等



■ 网络不仅在纵向上有深度，在横向上也有深度（广度）

GoogLeNet

- GoogLeNet 在横向上有“宽度”，这称为“Inception 结构”
 - 使用了多个大小不同的滤波器（和池化），最后再合并它们的结果
 - 很多地方都使用了大小为 1×1 的滤波器的卷积层
 - 1×1 的卷积运算通过在通道方向上减小大小，有助于减少参数和实现高速化处理



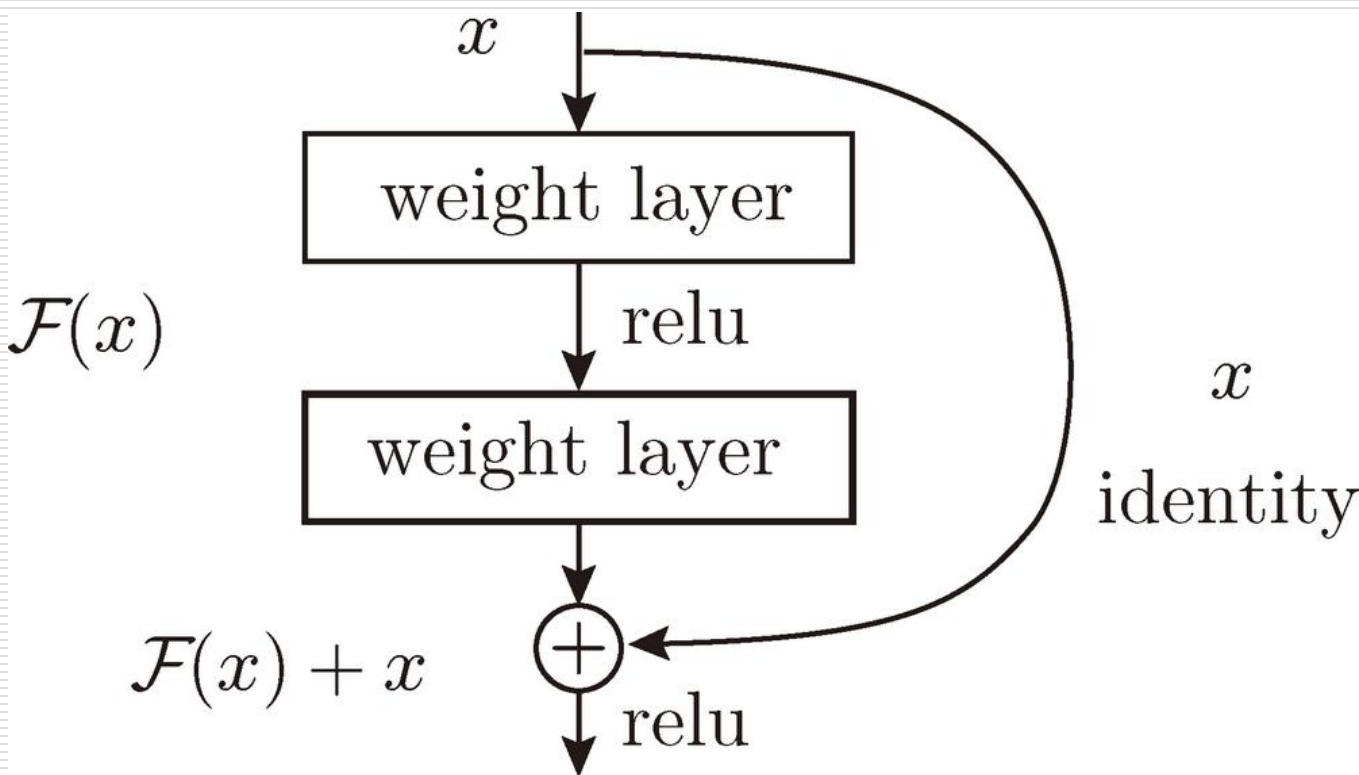
ResNet

- 微软团队开发的网络。它的特征在于具有比以前的网络更深的结构
 - 加深层对于提升性能很重要
 - 导入了“快捷结构”（也称为“捷径”或“小路”），可以随着层的加深而不断提高性能

ResNet

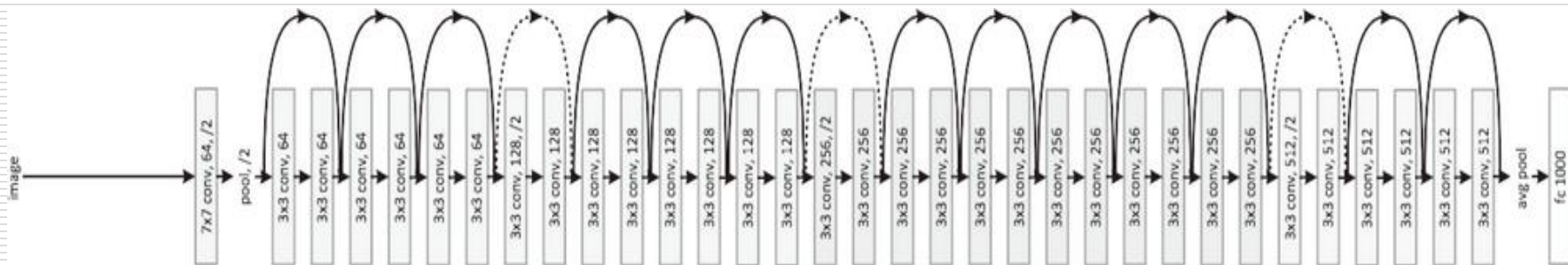
□ “快捷结构”

- 在连续 2 层的卷积层中，将输入 x 跳着连接至 2 层后的输出。
- 通过快捷结构，原来的 2 层卷积层的输出 $F(x)$ 变成了 $F(x)+x$ 。
 - 通过引入这种快捷结构，即使加深层，也能高效地学习。这是因为，通过快捷结构，反向传播时信号可以无衰减地传递



ResNet

- 以 VGG 网络为基础，引入快捷结构以加深层
 - 通过以 2 个卷积层为间隔跳跃式地连接来加深层
 - 根据实验的结果，即便加深到 150 层以上，识别精度也会持续提高
 - 迁移学习
 - 实践中经常会灵活应用使用 ImageNet 这个巨大的数据集学习到的权重数据，将学习完的权重（的一部分）复制到其他神经网络，进行再学习（fine tuning）



深度学习的高速化

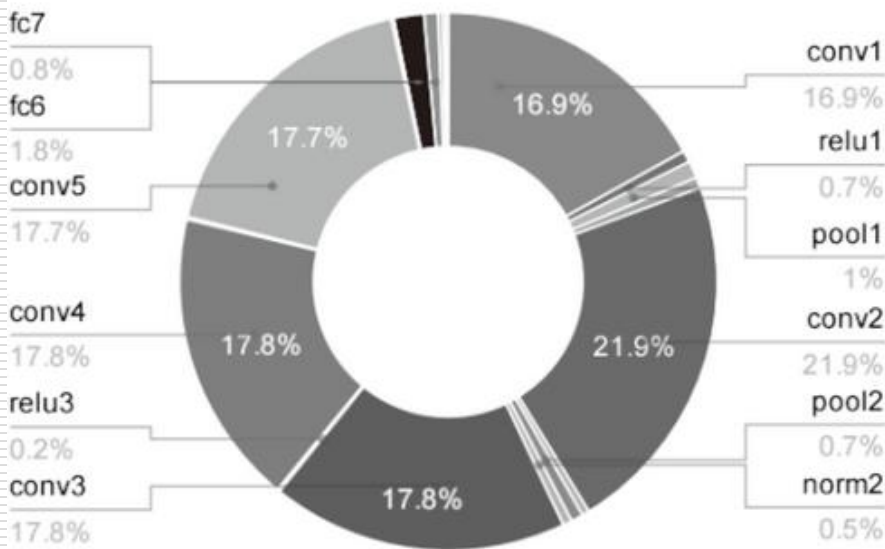
- 随着大数据和网络的大规模化，深度学习需要进行大量的运算
- 大多数深度学习的框架都支持 GPU（Graphics Processing Unit），可以高速地处理大量的运算
- 最近的框架也支持多个 GPU 或多台机器上的分布式学习

需要努力解决的问题

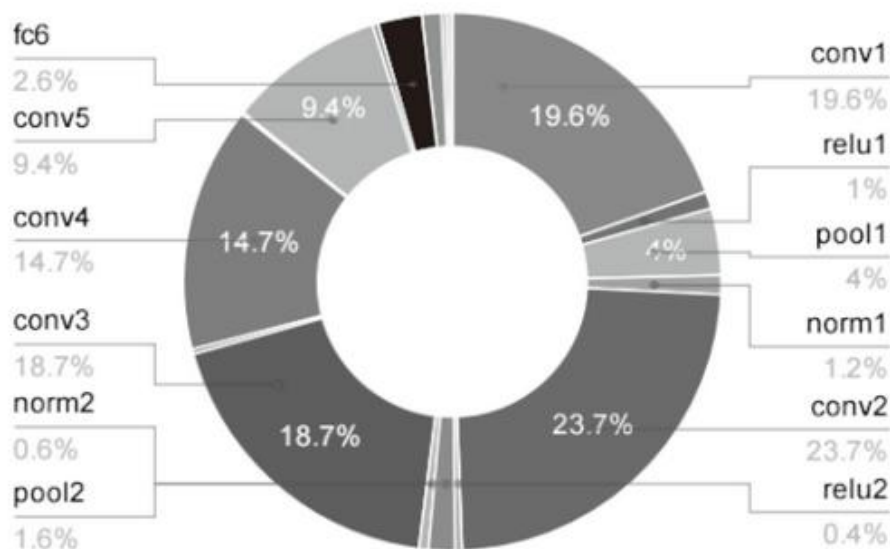
□ 以 AlexNet 的 forward 处理为对象，用饼图展示了各层所耗费的时间。

■ AlexNet 的 forward 处理中各层的时间比：左边是使用 GPU 的情况，右边是使用 CPU 的情况。图中的“conv”对应卷积层，“pool”对应池化层，“fc”对应全连接层，“norm”对应正规化层

GPU Forward Time Distribution



CPU Forward Time Distribution



需要努力解决的问题

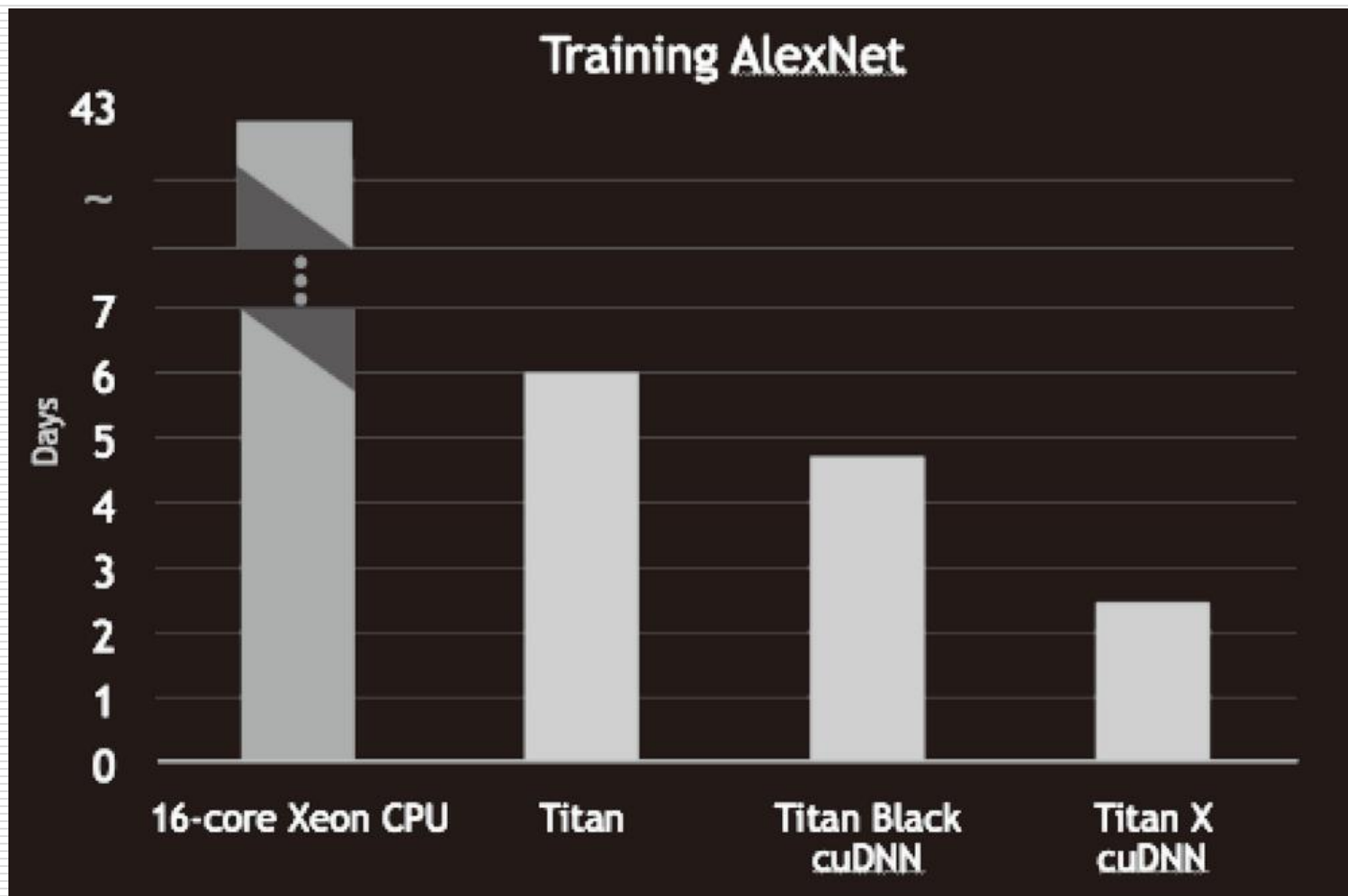
- AlexNet 中，大多数时间都被耗费在卷积层上。实际上，卷积层的处理时间加起来占 GPU 整体的 95%，占 CPU 整体的 89% ！
- 如何高速、高效地进行卷积层中的运算是深度学习的一大课题。

基于 GPU 的高速化

- GPU 可以高速地进行并行数值计算
- GPU 计算的目标就是将这种压倒性的计算能力用于各种用途
 - GPU 计算，是指基于 GPU 进行通用的数值计算的操作
 - 深度学习中需要进行大量的乘积累加运算（或者大型矩阵的乘积运算）深度学习中需要进行大量的乘积累加运算（或者大型矩阵的乘积运算）

基于 GPU 的高速化

- 基于 CPU 和 GPU 进行 AlexNet 的学习时分别所需的时间



基于 GPU 的高速化

- GPU 主要由 NVIDIA 和 AMD 两家公司提供。
 - 虽然两家的 GPU 都可以用于通用的数值计算，但与深度学习比较“亲近”的是 NVIDIA 的 GPU。
 - 实际上，大多数深度学习框架只受益于 NVIDIA 的 GPU。
 - 深度学习的框架中使用了 NVIDIA 提供的 CUDA 这个面向 GPU 计算的综合开发环境。
 - cuDNN 是在 CUDA 上运行的库，它里面实现了为深度学习最优化过的函数等。

主流芯片应用



CPU

中央处理器



GPU

图形加速处理器



FPGA

可编程逻辑门阵列



ASIC

专用定制芯片

全球芯片产业应用



CPU+GPU



CPU+GPU / ASIC



CPU+GPU / ASIC / DSP



CPU / GPU / TPU
ASIC / FPGA

全球芯片产业巨头



透析芯片产业链条

□ 设计、仿真验证（IP）



□ 流片、生成/代工



□ 应用



透析芯片产业链条

- ❑ 多数企业只涉足芯片的设计
- ❑ 芯片设计验证复杂，常基于ARM的参考设计
- ❑ 芯片生产投入大、风险高，时间周期长
- ❑ 掌握核心技术，或掌握应用生态——功夫在芯片外
- ❑ “缺芯少魂”背景下的芯片霸主——台积电

什么是AI芯片

□ AI算法+芯片=AI芯片？

■ 只要能是AI算法跑起来？

■ CPU/GPU/FPGA/DSP.....

■ AI，并非指某种特定算法或理论



什么是AI芯片

- 为AI算法量身定制的芯片=AI芯片
 - 能使AI算法高效的运行起来
 - 功耗/速度/体积/成本
 - 深度学习/神经网络

分布式学习

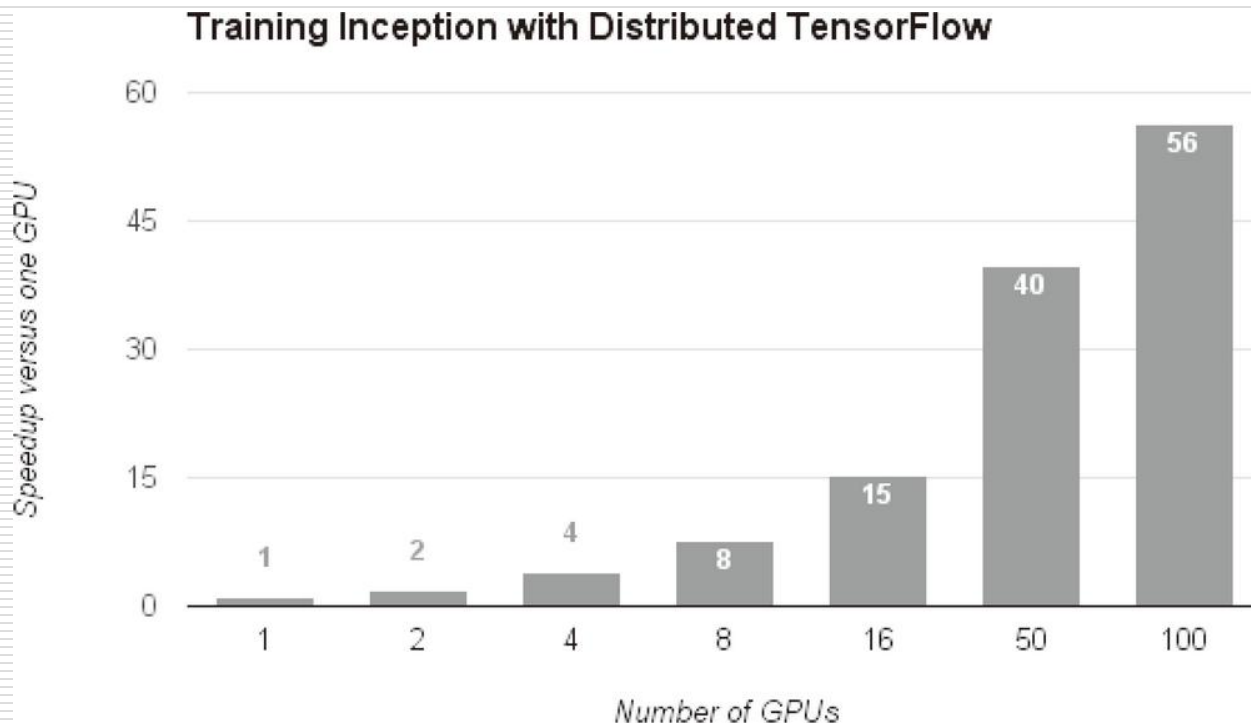
- 深度学习伴随着很多试错。
- 为了创建良好的网络，需要反复进行各种尝试，产生尽可能地缩短一次学习所需的时间的要求
 - 考虑在多个 GPU 或者多台机器上进行分布式计算
 - 支持多 GPU 或者多机器的分布式学习的框架
 - Google 的 TensorFlow
 - 微软的 CNTK (Computational Network Toolkit)

分布式学习

□ 基于 TensorFlow 的分布式学习的效果

■ 随着 GPU 个数的增加，学习速度也在提高

- 与使用 1 个 GPU 时相比，使用 100 个 GPU（设置在多台机器上，共 100 个）似乎可以实现 56 倍的高速化
- 之前花费 7 天的学习只要 3 个小时就能完成



运算精度的位数缩减

- 在深度学习的高速化中，除了计算量之外，内存容量、总线带宽等也有可能成为瓶颈
 - 考虑将大量的权重参数或中间数据放在内存中
 - 总线带宽，当流经 GPU（或者 CPU）总线的数据超过某个限制时，就会成为瓶颈
 - 实数
 - 主要使用 64 位或者 32 位的浮点数
 - 通过使用较多的位来表示数字，虽然数值计算时的误差造成的影响变小了，但计算的处理成本、内存使用量却相应地增加了，还给总线带宽带来了负荷。

运算精度的位数缩减

- 深度学习并不那么需要数值精度的位数
 - 基于神经网络的健壮性而产生的
 - 比如，即便输入图像附有一些小的噪声，输出结果也仍然保持不变
 - 流经网络的数据即便有所“劣化”，对输出结果的影响也较小
- 为了实现深度学习的高速化，位数缩减是今后必须关注的一个课题，特别是在面向嵌入式应用程序中使用深度学习时，位数缩减非常重要。

深度学习的未来

□ 图像风格变换

■ 使用深度学习来“绘制”带有艺术气息的画

- 两个输入图像中，一个称为“内容图像”，另一个称为“风格图像”



深度学习的未来

- 图像风格变换，使用深度学习来“绘制”带有艺术气息的画
 - 在学习过程中使网络的中间数据近似内容图像的中间数据。
 - 使输入图像近似内容图像的形状。
 - 为了从风格图像中吸收风格，导入了风格矩阵的概念
 - 通过在学习过程中减小风格矩阵的偏差，就可以使输入图像接近梵高的风格。

深度学习的未来

□ 图像的生成

■ 基于深度学习，可以实现从零生成“卧室”的图像

□ 基于 DCGAN (Deep Convolutional Generative Adversarial Network) 方法生成的卧室图像



深度学习的未来

□ 图像的生成

- 能画出以假乱真的图像的 DCGAN 会将图像的生成过程模型化
 - 使用大量图像（比如，印有卧室的大量图像）训练这个模型，学习结束后，使用这个模型，就可以生成新的图像
 - 使用了 **Generator**（生成者）和 **Discriminator**（识别者）
 - **Generator** 生成近似真品的图像，**Discriminator** 判别它是不是真图像（是 **Generator** 生成的图像还是实际拍摄的图像）

深度学习的未来

□ 自动驾驶

■ 正确识别周围环境的技术尤其重要

□ 基于 CNN 的神经网络 SegNet

■ 道路、车辆、建筑物、人行道等被高精度地识别了出来



小结

- 对于大多数的问题，都可以期待通过加深网络来提高性能。
- 在最近的图像识别大赛 **ILSVRC** 中，基于深度学习的方法独占鳌头，使用的网络也在深化。
- **VGG**、**GoogLeNet**、**ResNet** 等是几个著名的网络。
- 基于 **GPU**、分布式学习、位数精度的缩减，可以实现深度学习的高速化。
- 深度学习（神经网络）不仅可以用于物体识别，还可以用于物体检测、图像分割。
- 深度学习的应用包括图像标题的生成、图像的生成、强化学习等。最近，深度学习在自动驾驶上的应用也备受期待



THE END