

有名的卷积网络结构

图像识别一直是机器学习中非常重要的一个研究内容, 2010年开始举办的[ILSVRC](#)比赛更是吸引了无数的团队. 这个比赛基于一个百万量级的图片数据集, 提出一个图像1000分类的挑战. 前两年在比赛中脱颖而出的都是经过人工挑选特征, 再通过 `SVM` 或者 `随机森林` 这样在过去十几年中非常成熟的机器学习方法进行分类的算法。

CIFAR 10

cifar 10 这个数据集一共有 50000 张训练集, 10000 张测试集, 两个数据集里面的图片都是 png 彩色图片, 图片大小是 $32 \times 32 \times 3$, 一共是 10 分类问题, 分别为飞机、汽车、鸟、猫、鹿、狗、青蛙、马、船和卡车。这个数据集是对网络性能测试一个非常重要的指标, 可以说如果一个网络在这个数据集上超过另外一个网络, 那么这个网络性能上一定要比另外一个网络好, 目前这个数据集最好的结果是 95% 左右的测试集准确率。

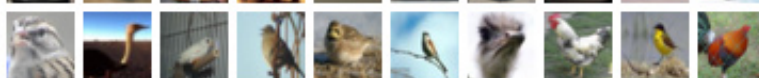
airplane



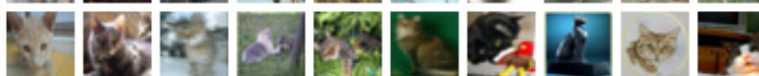
automobile



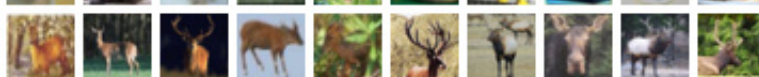
bird



cat



deer



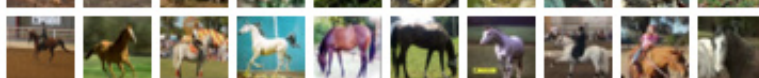
dog



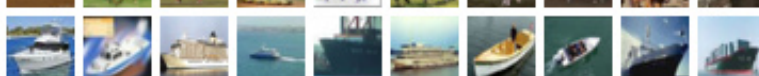
frog



horse



ship



truck

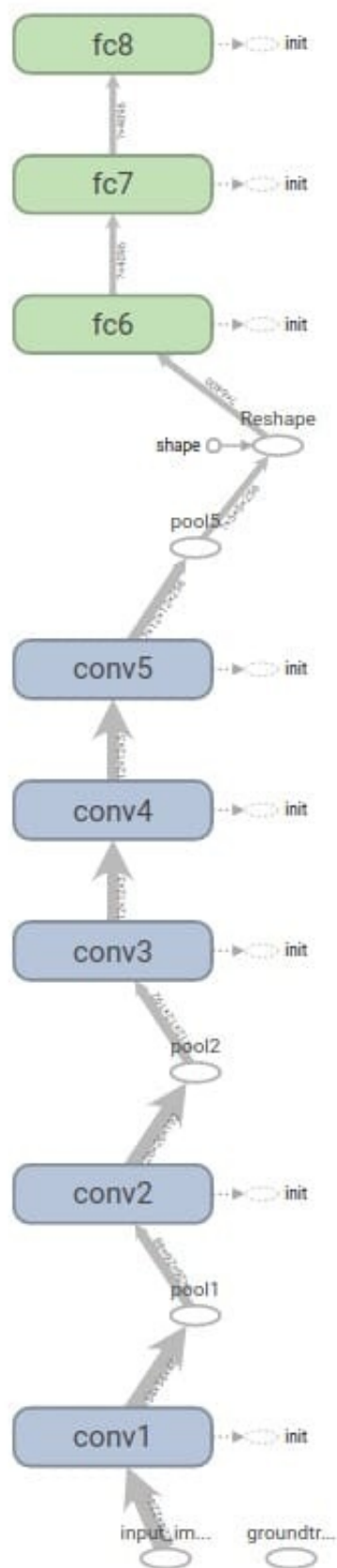


你能用肉眼对这些图片进行分类吗？

AlexNet

在2012年, 由 [Alex Krizhevsky](#), [Ilya Sutskever](#), [Geoffrey Hinton](#)提出了一种使用卷积神经网络的方法, 以 [0.85](#) 的 `top-5` 正确率一举获得当年分类比赛的冠军, 超越使用传统方法的第二名10个百分点, 震惊了当时的学术界, 从此开启了人工智能领域的新篇章.

这次的课程我们就来复现一次 `AlexNet`, 首先来看它的网络结构

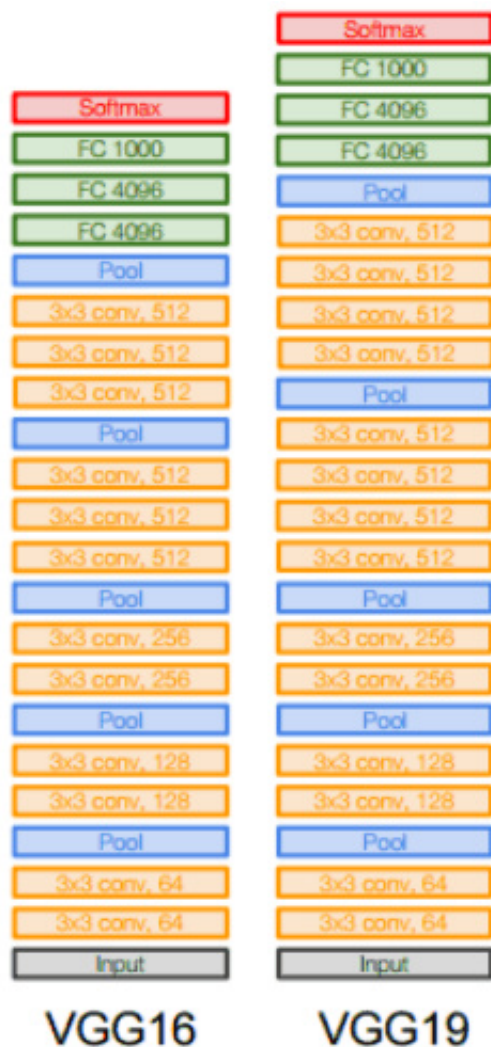


可以看出 AlexNet 就是几个卷积池化堆叠后连接几个全连接层。

VGG

vggNet 是第一个真正意义上的深层网络结构，其在 ImageNet2014年的冠军，得益于 python 的函数和循环，我们能够非常方便地构建重复结构的深层网络。

vgg 的网络结构非常简单，就是不断地堆叠卷积层和池化层，下面是一个简单的图示



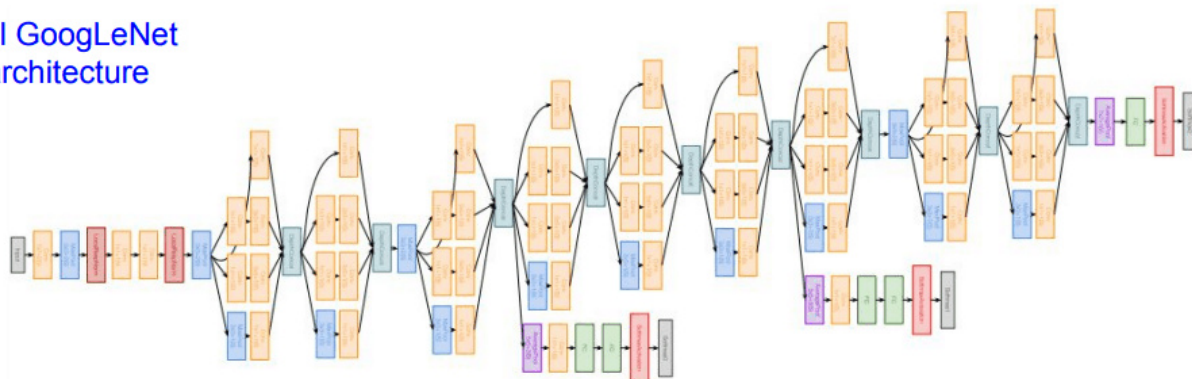
vgg 几乎全部使用 3×3 的卷积核以及 2×2 的池化层，使用小的卷积核进行多层的堆叠和一个大的卷积核的感受野是相同的，同时小的卷积核还能减少参数，同时可以有更深的结构。

vgg 的一个关键就是使用很多层 3×3 的卷积然后再使用一个最大池化层，这个模块被使用了很多次，下面我们照着这个结构来写一写

GoogLeNet

前面我们讲的 VGG 是 2014 年 ImageNet 比赛的亚军，那么冠军是谁呢？就是我们马上要讲的 GoogLeNet，这是 Google 的研究人员提出的网络结构，在当时取得了非常大的影响，因为网络的结构变得前所未有的，它颠覆了大家对卷积网络的串联的印象和固定做法，采用了一种非常有效的 inception 模块，得到了比 VGG 更深的网络结构，但是却比 VGG 的参数更少，因为其去掉了后面的全连接层，所以参数大大减少，同时有了很高的计算效率。

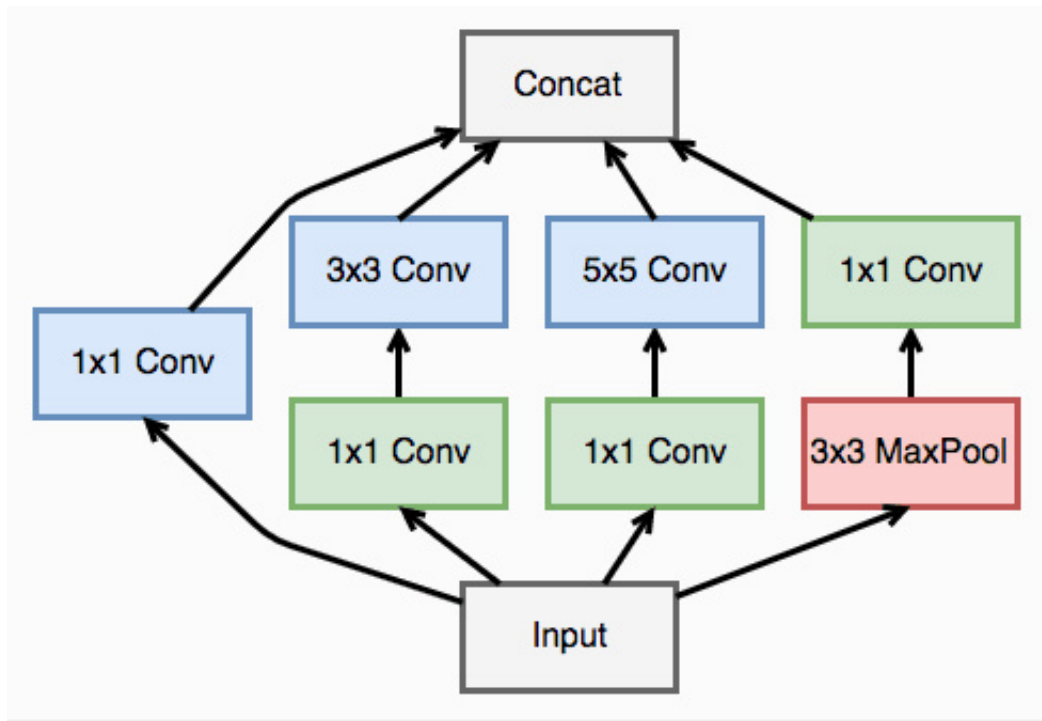
Full GoogLeNet architecture



这是 googlenet 的网络示意图，下面我们介绍一下其作为创新的 inception 模块。

Inception 模块

在上面的网络中，我们看到了多个四个并行卷积的层，这些四个卷积并行的层就是 inception 模块，可视化如下



一个 inception 模块的四个并行线路如下：1. 一个 1×1 的卷积，一个小的感受野进行卷积提取特征 2. 一个 1×1 的卷积加上一个 3×3 的卷积， 1×1 的卷积降低输入的特征通道，减少参数计算量，然后接一个 3×3 的卷积做一个较大感受野的卷积 3. 一个 1×1 的卷积加上一个 5×5 的卷积，作用和第二个一样 4. 一个 3×3 的最大池化加上 1×1 的卷积，最大池化改变输入的特征排列， 1×1 的卷积进行特征提取

最后将四个并行线路得到的特征在通道这个维度上拼接在一起。

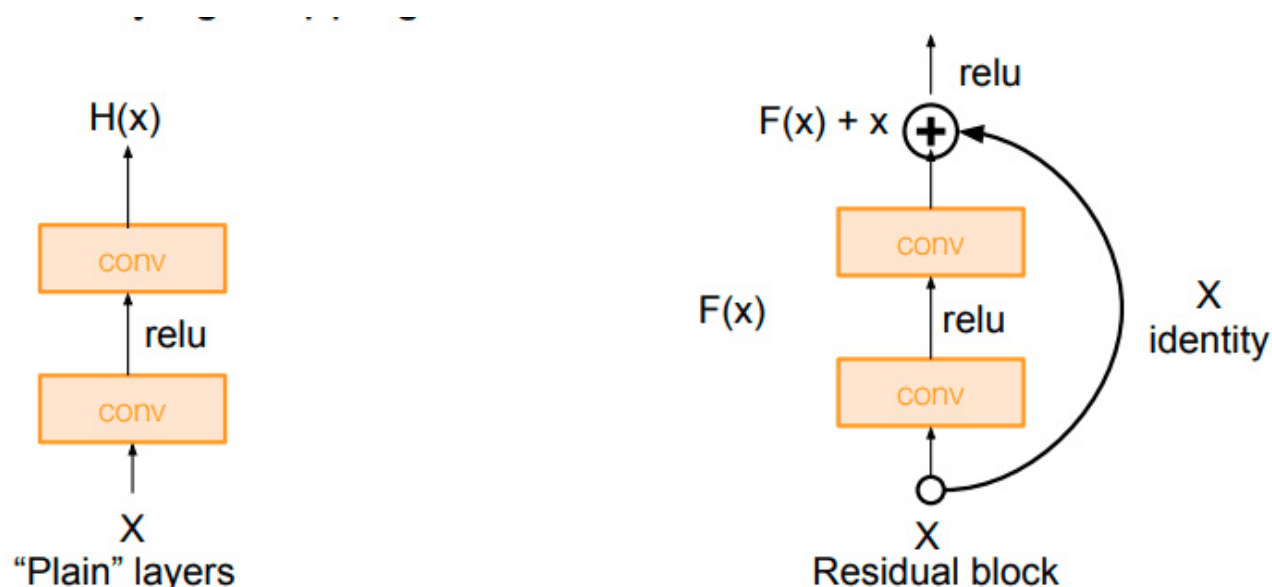
ResNet

当大家还在惊叹 GoogLeNet 的 inception 结构的时候，微软亚洲研究院的研究员已经在设计更深但结构更加简单的网络 ResNet，并且凭借这个网络子在 2015 年 ImageNet 比赛上大获全胜。

ResNet 有效地解决了深度神经网络难以训练的问题，可以训练高达 1000 层的卷积网络。网络之所以难以训练，是因为存在着梯度消失的问题，离 loss 函数越远的层，在反向传播的时候，梯度越小，就越难以更新，随着层数的增加，这个现象越严重。之前有两种常见的方案来解决这个问题：

- 1.按层训练，先训练比较浅的层，然后在不断增加层数，但是这种方法效果不是特别好，而且比较麻烦
- 2.使用更宽的层，或者增加输出通道，而不加深网络的层数，这种结构往往得到的效果又不好

ResNet 通过引入了跨层链接解决了梯度回传消失的问题。



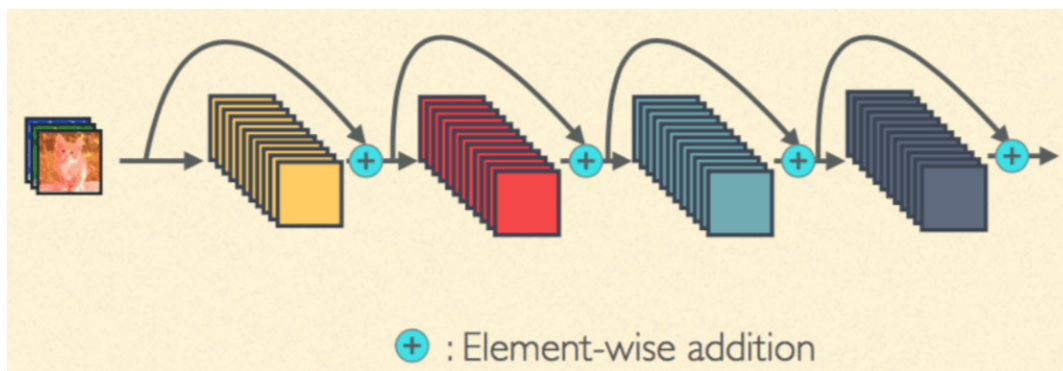
这就普通的网络连接跟跨层残差连接的对比图，使用普通的连接，上层的梯度必须要一层一层传回来，而是用残差连接，相当于中间有了一条更短的路，梯度能够从这条更短的路传回来，避免了梯度过小的情况。

假设某层的输入是 x ，期望输出是 $H(x)$ ，如果我们直接把输入 x 传到输出作为初始结果，这就是一个更浅层的网络，更容易训练，而这个网络没有学会的部分，我们可以使用更深的网络 $F(x)$ 去训练它，使得训练更加容易，最后希望拟合的结果就是 $F(x) = H(x) - x$ ，这就是一个残差的结构。

DenseNet

因为 ResNet 提出了跨层链接的思想，这直接影响了随后出现的卷积网络架构，其中最著名的就是 cvpr 2017 的 best paper，DenseNet。

DenseNet 和 ResNet 不同在于 ResNet 是跨层求和，而 DenseNet 是跨层将特征在通道维度进行拼接，下面可以看看他们两者的图示



第一张图是 ResNet，第二张图是 DenseNet，因为是在通道维度进行特征的拼接，所以底层的输出会保留进入所有后面的层，这能够更好的保证梯度的传播，同时能够使用低维的特征和高维的特征进行联合训练，能够得到更好的结果。