

MCDNN

数据先被预处理然后被扭曲

preprocessing: 对整个数据集的操作, 非随机处理, Image Adjustment、Histogram Equalization等

distortion: 小幅度 translation、scaling、rotation

通过bilinear interpolation得到最终固定大小的图片数据

网络结构:

$$input \rightarrow Conv \rightarrow MP \rightarrow Conv \rightarrow MP \rightarrow Conv \rightarrow MP \rightarrow FC \rightarrow FC$$

引入max-pooling layers

设计了多个DNN columns, 在训练前所有DC被随机初始化(服从不同正态分布), 最后对所有DC的输出结果做平均, 得到输出

VGG

数据预处理: 每个像素减去训练集上的RGB均值

思想: 用较小卷积核的叠加来取代较大的卷积核

结构: 卷积+最大池化

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

优势：网络结构简单，便于设置

缺点：全连接层过多导致参数过多

ResNet

Motivation: degradation

当网络的层数增加到一定程度时，继续增加反而会导致准确率的降低。但是理论上更深的网络不会比更浅的网络准确率低，因为可以将多余的层学习为identity mappings，出现这种问题说明了用非线性层表示identity mapping是很困难的。基于这种想法，令 $\mathcal{H}(x)$ 为需要学习的目标函数，该网络不学习 $\mathcal{H}(x)$ ，改而学习 $\mathcal{H}(x) - x$ ，令 $\mathcal{F}(x) = \mathcal{H}(x) - x$ ，构造如下结构：

$$y = \mathcal{F}(x, W_i) + x$$

其中 x, y 分别表示输入输出， $\mathcal{F}(x, W_i)$ 表示需要被学习的映射。这样在遇到之前的问题时该网络只需将 $\mathcal{F}(x)$ 学习为0，降低了学习的难度和资源消耗。

x 是通过“shortcut connections”直连到输出的，这里假设 x 与 \mathcal{F} 维数相同。如果维数不同就对 x 作投影变换。

$\mathcal{F}(x)$ 在本网络中为两层或三层网络需要拟合的函数

实验结果

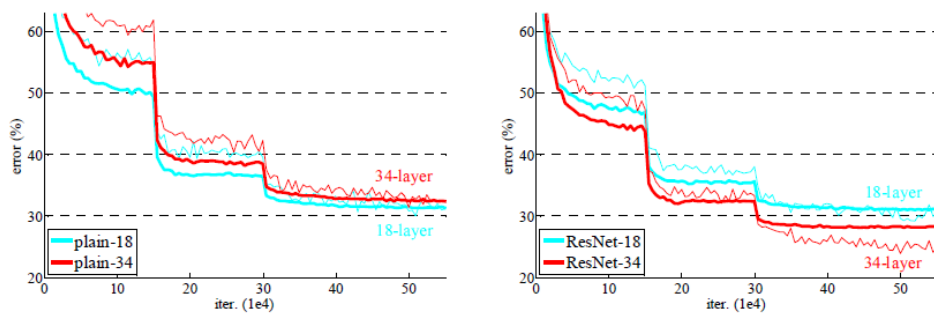


Figure 4. Training on ImageNet. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

baseline在网络层数为34层时的错误率要高于18层的网络，说明出现了degradation；而ResNet在34层时的错误率要低于18层，说明degradation被比较好地解决了，即可以通过增加网络深度来提高准确率了。同时发现ResNet的收敛速度也要更快。

不过并不能无限制地增加深度：

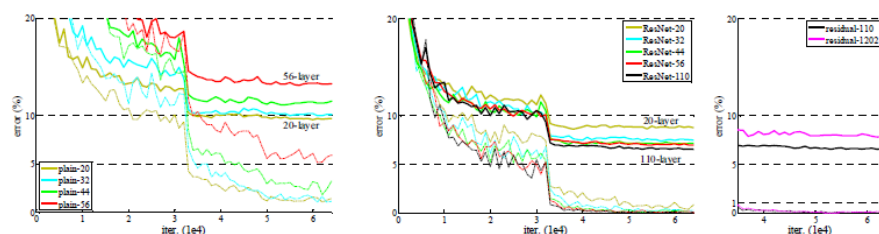


Figure 6. Training on CIFAR-10. Dashed lines denote training error, and bold lines denote testing error. Left: plain networks. The error of plain-110 is higher than 60% and not displayed. Middle: ResNets. Right: ResNets with 110 and 1202 layers.

发现在网络过深(1000余层)时，仍然会因为过拟合而出现准确率降低的情况。