

## cs231n-2: More common architectures of CNN

### 1. AlexNet

1. 参数很多，大多集中在Fully Connection Layer，所以affine层要少用

### 2. ZFNet

### 3. VGGNet

1. much more layers and smaller filter size
2. effective receptive field（感受野）
  1. 某一层对前面某层的相关面积
  2. 比如第二层对第一层， $3 \times 3$  filter，那么receptive field就是3，相应的第三层（也是 $3 \times 3$ ）对第一层是 $5 \times 5$ （注意overlap）
3. 因此，当层数增加，我们用较小的filter得到了相同的effective receptive field，而且相对于 $7 \times 7$ 的filter，我们用3层 $3 \times 3$ 就可以得到一样的effective receptive field，hyper parameter从 $45 \times C_{in} \times C_{out}$ 降到了 $27 \times C_{in} \times C_{out}$ ！！ $C_{in}$ 是输入信道， $C_{out}$ 是输出信道。
4. 我们只计算可以计算权重的layer，pooling这种我们不计算，见VGG16.
5. softmax在分类问题往往做的不错，当然我们也可以用任何其他的损失函数。
6. VGG没用local response normalization，因为他们发现没什么特别大的帮助。
7. 同样的，我们也可以用ensemble model来得到最佳效果。
8. FC7可以很好的表示这些类的特征。

### 4. GoogleNet

#### 1. inception module

1. Apply parallel filter operations on the input from previous layer
2. Then concatenate all of the outputs depth-wise，也就是在信道上加起来，注意我们要用zero padding来维持原状
3. 计算复杂度：
  1. 每层三个layer导致的更昂贵的计算代价。
  2. 可以用bottleneck layer来解决：用1 by 1 filter来将input reduce到更低的维度（信道减少）。

2. It removes a FC layers because it works well without them. And by doing this, we can reduce many parameters.

3. Auxiliary classification outputs

### 5. ResNet - very deep network using residual connections（残差网络）

1. 对VGG之类的模型，我们发现单纯的增加层数并不能改进它的效果，甚至会更差，连training error都变差了，这意味着甚至连overfitting都没有发生！ResNet对此的解释是deeper导致optimize变得更难。
2. 在原来的 $output = H(x)$ 的基础上，我们令 $output = F(x) + x$ ， $H$ 和 $F$ 都是神经网络学习出来的东西，由于 $H(x)$ 作为一个预测或者说Loss，和input的关系大概是一致的，所以用了ResNet以后，我们从学习 $H(x)$ 变成了学习 $F(x)$ ， $F$ 就是那个残差，这个残差的意思是，我们固定了 $x$ 这一项，只要求出增加的或者说减少的残差项。
3. ResNet基于一个假设（!!!）：学习这个残差会更加容易。但是效果很好...

4. 还可以用bottleneck layer来控制conv层的size, 在conv结束以后再bottleneck回去。

Training ResNet in practice:

- Batch Normalization after every CONV layer
  - Xavier/2 initialization from He et al.
  - SGD + Momentum (0.9)
  - Learning rate: 0.1, divided by 10 when validation error plateaus
  - Mini-batch size 256
  - Weight decay of  $1e-5$
5. - No dropout used
  6. ResNet有很多变种: Wide ResNet、ResNeXT等等。
    1. 有些人注意到深度不是最重要的, Residual才是重要的, 只要增加width即可达到ResNet一样的效果, 计算更简单

#### 6. 其他CNN:

1. Network in Network
2. Stochastic Depth
3. DenseNet
4. FractalNet
5. SqueezeNet

#### 7. 暂时ResNet是最好的, 现在deep learning研究主要的方面是

1. 越来越deep的网络的趋势
2. layer和connection方式的设计
3. Depth (deeper)、width (每一层的filter数量)、residual之间的问题