背景

模型的学习表达能力、数据集的大小和过拟合手段是影响物体识别任务的三个重要因素。

从模型角度考虑，图像识别任务的复杂性要求模型需要具有一定的先验知识，而 CNN 可以对自然图像整体统计特性的平稳性以及像素之间局部依赖关系做出很好的假设，且与其它前向传播的神经网络相比结构简单、易于训练，因此作者选择 CNN 作为网络模型。

当前 GPU 算力使得使用 CNN 训练高分辨率的图像成为可能，所以作者利用 GPU 进行网络加速训练，且使用足够大的数据集可以避免了严重的过拟合现象。

方法

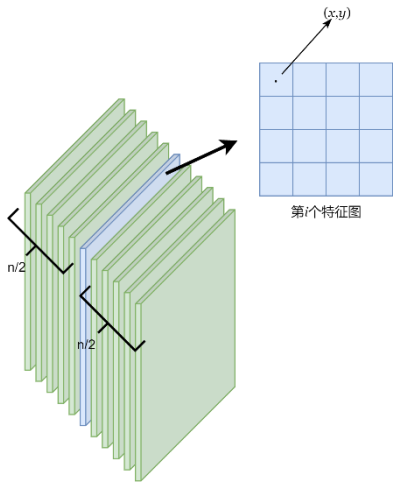
(1)采用 ReLU 非线性非饱和函数作为神经元激活函数加速训练。

因为非饱和性可以保证在进行梯度反向传播的时候没有梯度消失现象，且 ReLU 函数求导简单所以在使用梯度下降法时收敛的更快。

(2)采用局部响应归一化

增加了泛化能力，使输出中响应比较大的值变得相对更大，并抑制其他响应较小的神经元。

$$b_{x,y}^{i}=\frac{a_{x,y}^{i}}{(k+\alpha\sum_{j=\max(0,i-\frac{n}{2})}^{\min(N-1,i+\frac{n}{2})}(a_{x,y}^{j})^{2})^{\beta}} \tag{1}$$



$a_{x,y}^{i}$ 表示当前第 i 个特征图在(x,y)坐标处经过 ReLU 激活函数后的值，$b_{x,y}^{i}$ 表示经过局部响应归一化后的值。k,α,β 为超参数，由验证集决定。

此操作将第 i 个特征图及其步长为 $\frac{n}{2}$ 内的特征图，(x,y)位置上的值进行归一

化操作。

(3)重叠的池化操作

一般的池化操作 kernel_size 和 stride 一般是相等的，如果 kernel_size 大于 stride 就会发生重叠池化。加上重叠池化操作后 top-1 和 top-5 错误率可以分别降低 0.4%和 0.3%。加上该操作后不容易过拟合。

(4)label-preserving 的数据增强操作

①训练阶段对输入图像(256×256)进行随机裁剪(224×224)以及水平翻转；测试阶段将输入的图像裁剪五组（四组从角落裁剪，一组从中心位置进行裁剪）再加上对应的水平翻转后的五组，在这十组增强过后的数据进行预测，最后取其 softmax 的平均值。

②将训练集图像的 RGB 空间进行主成分分析，将得到的特征向量和一个大小与其特征值成比例的随机变量（均值为 0，标准差为 0.1 的高斯扰动）相加，由此可以得到自然图像中一些不随光照和颜色而改变的重要特性。

(5)Dropout

结合多个网络的预测结果可以减小误差，但是其成本很大。因此提出了 Dropout 来随机“失活”网络隐藏层中的某些神经元，这些神经元不参与前向和后向传播，再下一次迭代中又重新随机“失活”某些隐藏层中的神经元，直到训练结束。

在测试阶段，使用所有的神经元但是将其输出都乘 0.5，以此对多个 dropout 产生的多个网路做合理的近似。

实验

训练时使用的是随机梯度下降，所有的层都使用了相同的学习率，在训练过程中手动调整学习率，学习率初始化为 0.01，当验证误差率在当前学习率下不再提高时就将学习率除以 10，在终止前降低三次。

实验部分先和传统的图像分类技术 Sparse coding（ILSVRC2010 冠军）、SIFT+FVs 进行了对比，结果表明使用 CNN 进行图像分类的效果更好。

再使用论文中提到的 CNN 模型得到的结果、平均 5 个相似 CNN 得到的结果、在最后一个池化层上加入第六个卷积层、平均 5 个相似的使用预训练的 CNN 得到的结果进行对比。

个人理解

文章整个理解下来就是一个基础且简单的网络加了一些防止过拟合的 trick 来减小误差。采用 ReLU 函数作为激活函数可以有效避免梯度消失。