

AlexNet Notes

1. 网络结构

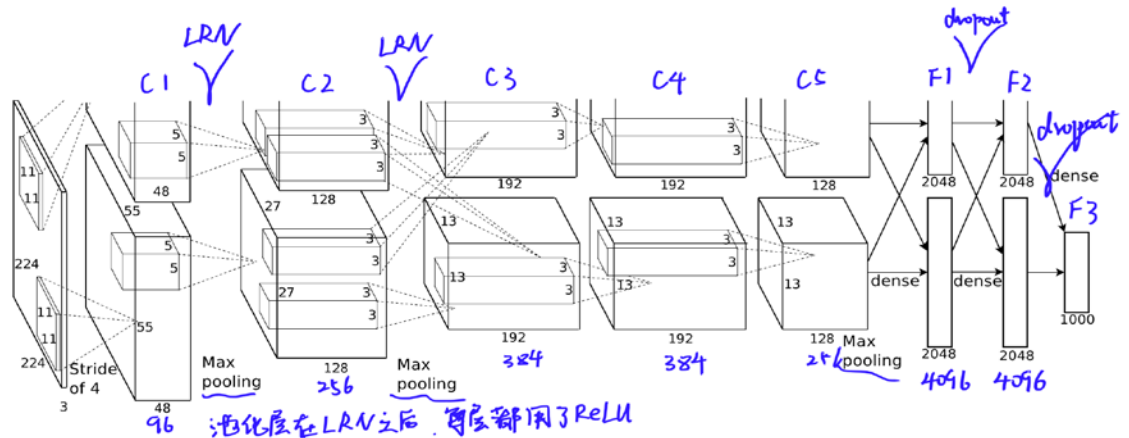


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

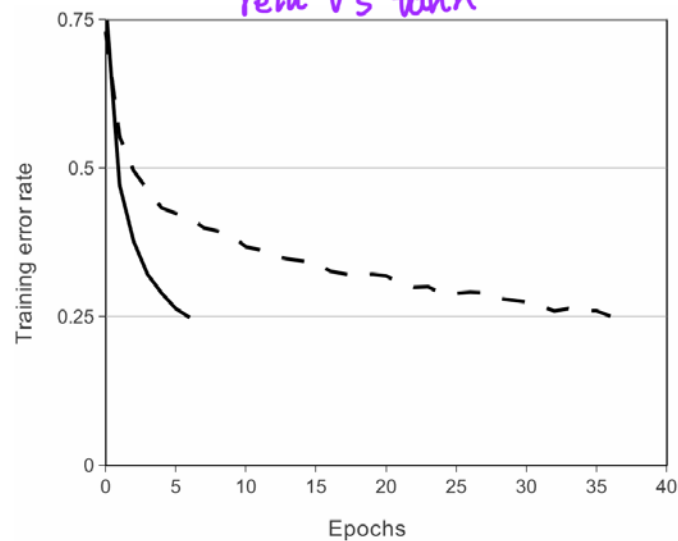
5 个卷积层+三个全连接层

2. AlexNet 中一些网络结构特征

2.1. ReLU

$$f(x) = \max(0, x)$$

达到25%错误率所用epoch的对比
relu vs tanh



2.2. Local Response Normalization (局部响应归一化)

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

这个 VGG 论文里说没什么用，目前很少用了

2.3. Overlapping Pooling

之前的池化 stride=卷积核大小，这篇论文里 stride<卷积核大小

3. 避免过拟合的一些方法

3.1. 数据增强

作者在进行图片转换时，使用 CPU 转换，使用 GPU 训练，转换与训练同时进行，不用进行转换图片的存储。

第一种转换方法：训练阶段，从 256*256 的图片中随机提取 224*224 大小的图片和它们的水平翻转图片，这使训练集扩大了 2048 倍；测试阶段，从测试图片的四个角和中心各裁剪一张 224*224 大小的图片，并将它们水平翻转，并将这十张图片的预测结果平均。

第二种转换方法：PCA，没看懂。

3.2. Dropout

Dropout 使模型收敛所需的迭代次数增加了一倍。

4. Some take-aways

4.1. 当验证集上的错误率不再下降，将学习率除以 10

-
- 4.2. 作者提到用欧式距离衡量图片送入网络得到的向量之间的相似度，用来做 Image Retrival。