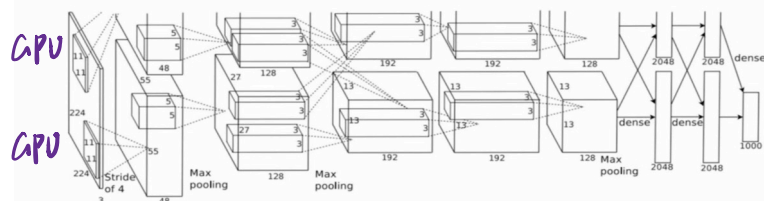


2013 AlexNet

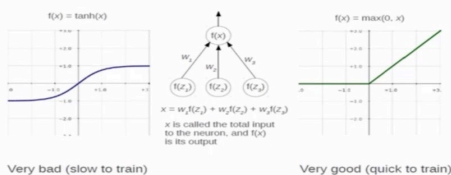
A. Krizhevsky, I. Sutskever and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing 25, MIT Press, Cambridge, MA, 2012.



ImageNet包含超过120万张彩色图片，属于1000个不同类别，这是目前为止最大的图像识别数据库。Alex Krizhevsky等人构建了一个包含65万多个神经元，待估计参数超过6000万的大规模网络，这一网络被称为AlexNet

2013 AlexNet 改进

Neurons



$$\frac{\partial \text{ReLU}}{\partial x} = 1 \quad x > 0$$

(1) 以ReLU函数代替公式 (3-10) 中的sigmoid或tanh函数

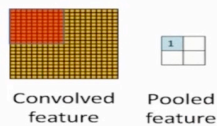
$$\text{ReLU}(x) = \max(0, x)$$

实践证明，这样做能使网络训练以更快速度收敛。

在每一层中每次激活的神经元减少
收敛性变好

2013 AlexNet 改进

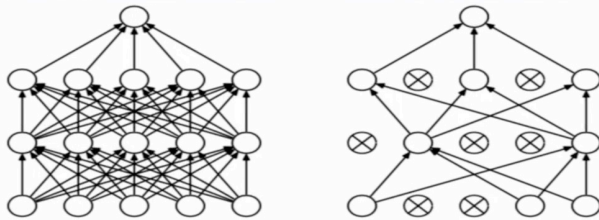
(2) 为降采样操作起了一个新的名字—池化 (Pooling)，意思是把邻近的像素作为一个“池子”来重新考虑。如图 3.31所示，左边所有红色的像素值可以看做是一个“池子”，经过池化操作后，变成右边的一个蓝色像素。



在AlexNet中，提出了最大池化(Max Pooling)的概念，即对每一个邻近像素组成的“池子”，选取像素最大值作为输出。在LeNet中，池化的像素是不重叠的；而在AlexNet中进行的是有重叠的池化。实践表明，有重叠的最大池化能够很好的克服过拟合问题，提升系统性能。

(Pooling 与 Sub Sampling)
取最大值 取平均值
BP: 其边梯度0 每位置
取max梯度 取梯度/个数
放原位置

(3) 随机丢弃 (Dropout)。为了避免系统参数更新过快导致过拟合，每次利用训练样本更新参数时候，随机的“丢弃”一定比例的神经元，被丢弃的神经元将不参加训练过程，输入和输出该神经元的权重系数也不做更新。这样每次训练时，训练的网络架构都不一样，而这些不同的网络架构却分享共同的权重系数。实验表明，随机丢弃技术减缓了网络收敛速度，也以大概率避免了过拟合的发生。



Dropout做法是，对每一层，每次训练时以概率 p 丢弃一些神经元，这样每次训练的网络都不一样。

训练结束后的测试流程，要用完整的网络结构，同时对该层的所有参数 (W, b) 都要乘以 $(1-p)$ 。

测试时:

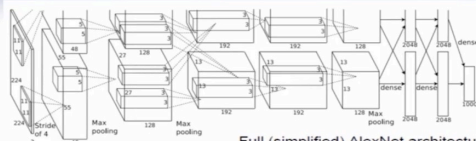
使用整个神经网络
同时 $(W, b) (1-p)$

(每一次使有限个神经元激活)

(4) 增加训练样本。尽管ImageNet的训练样本数量有超过120万幅图片，但相对于6亿待估计参数来说，训练图像仍然不够。Alex等人采用了多种方法增加训练样本，包括：1. 将原图水平翻转；2. 将 256×256 的图像随机选取 224×224 的片段作为输入图像。运用上面两种方法的组合可以将一幅图像变为2048幅图像。还可以对每幅图片引入一定的噪声，构成新的图像。这样做可以较大规模增加训练样本，避免由于训练样本不够造成的性能损失

(5) 用GPU加速训练过程。采用2片GTX 580 GPU对训练过程进行加速，由于GPU强大的并行计算能力，使得训练过程的时间缩短数十倍，哪怕这样，训练时间仍然用了六天。

Spot the GPUs!
(graphics processing unit)



思考题：请计算一下
ALEXNET的参数个数

Full (simplified) AlexNet architecture:
[227x227x3] INPUT
[55x55x96] CONV1: 96 11x11 filters at stride 4, pad 0
[27x27x96] MAX POOL1: 3x3 filters at stride 2
[27x27x96] NORM1: Normalization layer
[27x27x256] CONV2: 256 5x5 filters at stride 1, pad 2
[13x13x256] MAX POOL2: 3x3 filters at stride 2
[13x13x256] NORM2: Normalization layer
[13x13x384] CONV3: 384 3x3 filters at stride 1, pad 1
[13x13x384] CONV4: 384 3x3 filters at stride 1, pad 1
[13x13x256] CONV5: 256 3x3 filters at stride 1, pad 1
[6x6x256] MAX POOL3: 3x3 filters at stride 2
[4096] FC6: 4096 neurons
[4096] FC7: 4096 neurons
[1000] FC8: 1000 neurons (class scores)