Introduction of Reduced Convolutional Networks

精简卷积神经网络简介

唐呈俊 桂林电子科技大学

August 9, 2019

Outline

Preliminaries

Preliminaries 1: 自学内容

我们假设你在学习本课程前已经了解了如下概念的内容：

FC Fully Connected Layer，全连接层

Activations Activation Functions 激活函数

当然，基本的高等数学、线性代数、概率论知识也是必要的。信息论也许也会有一定的帮助。

Preliminaries 2: 卷积层的计算

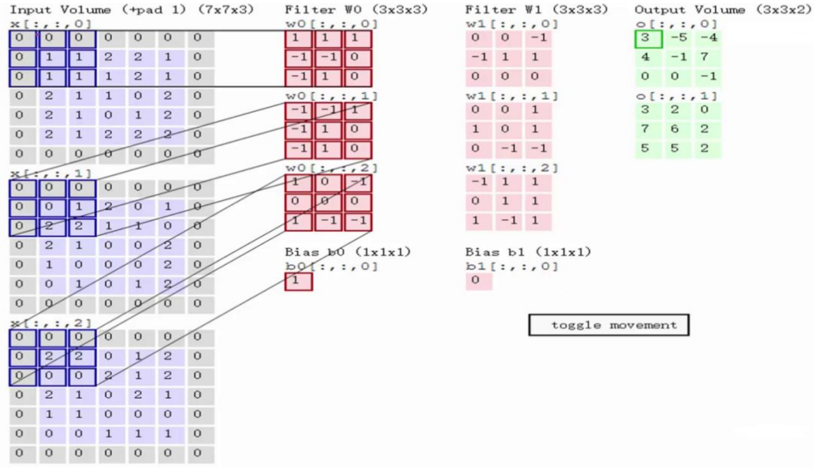


Figure. 卷积计算过程

Preliminaries 3: 卷积层的参数

上图讲解的是一个3*3的2D卷积，那么卷积层还有什么样的参数呢：

Filter 滤波器数量
Kernel Size 卷积核大小
Strides 步长
Padding 填充

Preliminaries 4: 池化层

虽然今天没有涉及池化的优化，但卷积总是和池化分不开的：

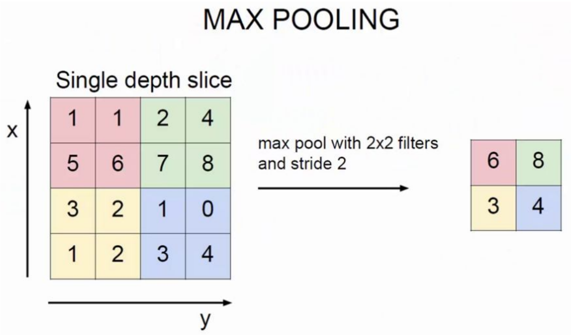


Figure. 最大池化

Methodology

Methodology 1: AlexNet

虽然AlexNet并非轻量化神经网络，但其中的两个方法，却是如今轻量化网络的重要基石：

Group Conv AlexNet中，为了在两个GPU上并行训练，其将卷积分解为两个Group，从而在两个GPU上完成。

ReLU AlexNet中，采用了ReLU激活函数，相比先前的tanh和sigmoid函数，ReLU函数的计算速度，和训练时的收敛速度都更快。

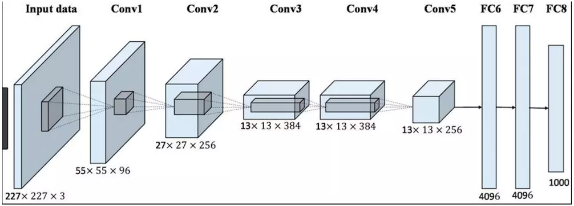


Figure. AlexNet网络结构

Methodology 2: Group Conv

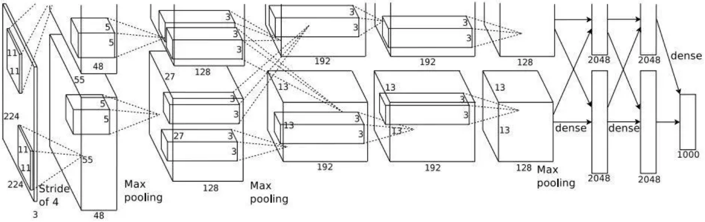


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input 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.

Figure. AlexNet的Group Conv结构

你能从中找到对应Group Conv的部分吗？

Methodology 3: Depthwise Conv

虽然AlexNet的Group Conv的目的是将计算分配给两个GPU并行运行，但是这也同样减少了计算量。（想想看为什么）

如果说将每个通道，都分到其对应的Group中，这种Group Conv的特殊情况，就将其运算量减少到了最少的情况。但这样做的代价是什么？（可以想想互信息的概念）

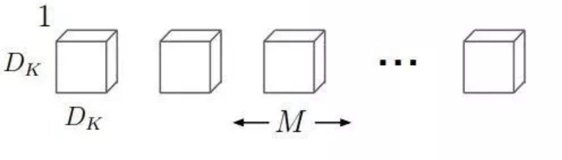


Figure. Depthwise Conv，其中M是通道数，也是滤波器数

Methodology 4: Depthwise Conv

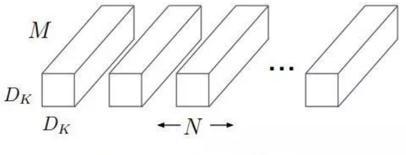


Figure. 标准的Conv（相对于 Depthwise Conv）

$$G_{k,l,n} = \sum_{i,j,m} \boldsymbol{K}_{i,j,m,n} \cdot \boldsymbol{F}_{k+i-1,l+j-1,m} \text{普通卷积}$$

$$\hat{G}_{k,l,m} = \sum_{i,j} \hat{\boldsymbol{K}}_{i,j,m} \cdot \boldsymbol{F}_{k+i-1,l+j-1,m} \text{Depthwise Conv}$$

想想看，两种卷积的计算复杂度。

Applications

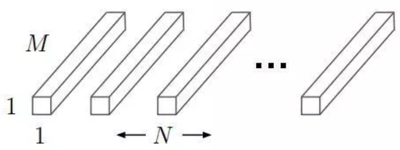


Figure. Pointwise Conv，即1 结构的普通卷积

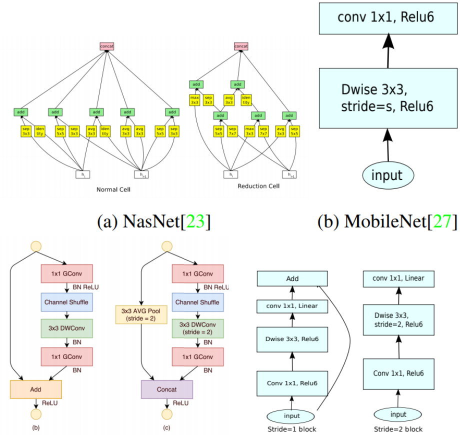
计算复杂度：

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \text{普通卷积}$$

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \text{Depthwise + Pointwise Conv}$$

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2}$$

Applications 2: MobileNet V2



(a) NasNet[23] (b) MobileNet[27]

(c) ShuffleNet [20] (d) Mobilenet V2

Figure. MobileNet V2与其它网络的对比

Applications 3: MobileNet V3

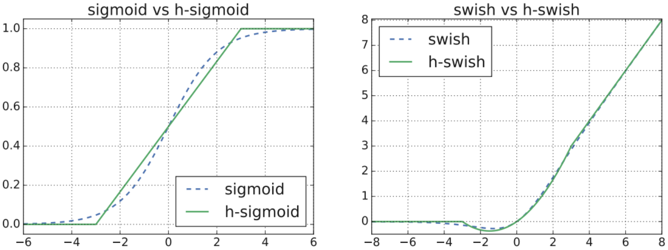


Figure. Sigmoid 和 Swish 的非线性形式及其hard近似

$$\text{swish } x = x \cdot \sigma(x)$$

$$\mathrm{h}- \text{ swish } [x] = x \frac{\mathrm{Re\,LU6}(x+3)}{6}$$

Discussions

Discussions 1: 网络性能对比实验

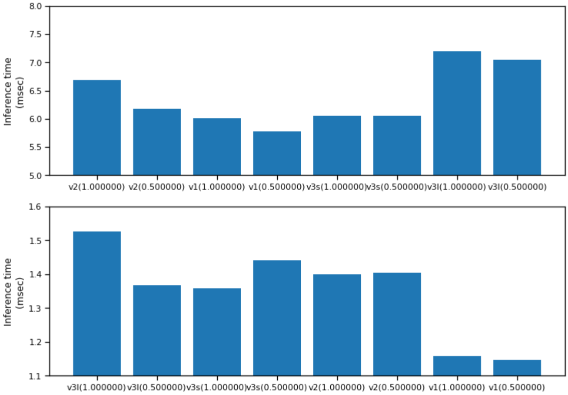


Figure. 各版本MobileNet在CPU和GPU上性能的表现

Discussions 2: 如何改进现有网络

大家可以思考，根据今天讲到的网络特性，对于不同的实际使用环境，如何改进现有网络？

今天的内容结束了，谢谢大家。