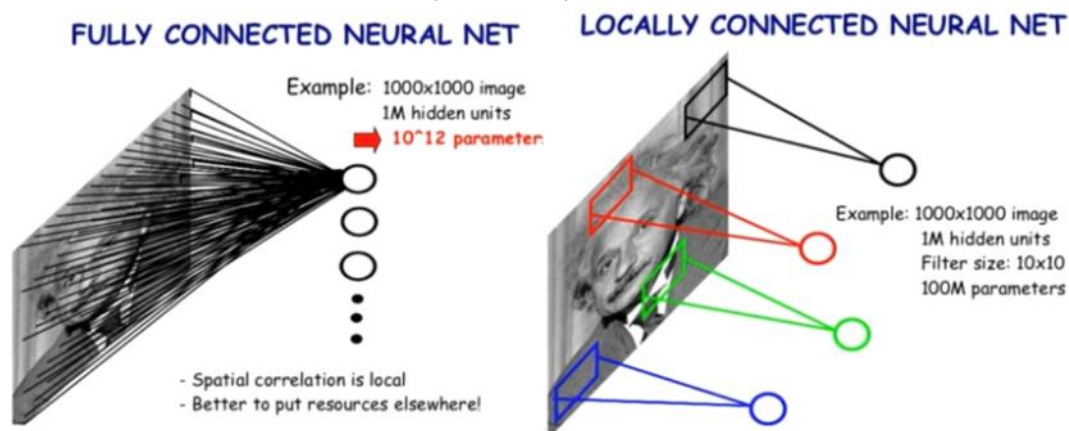


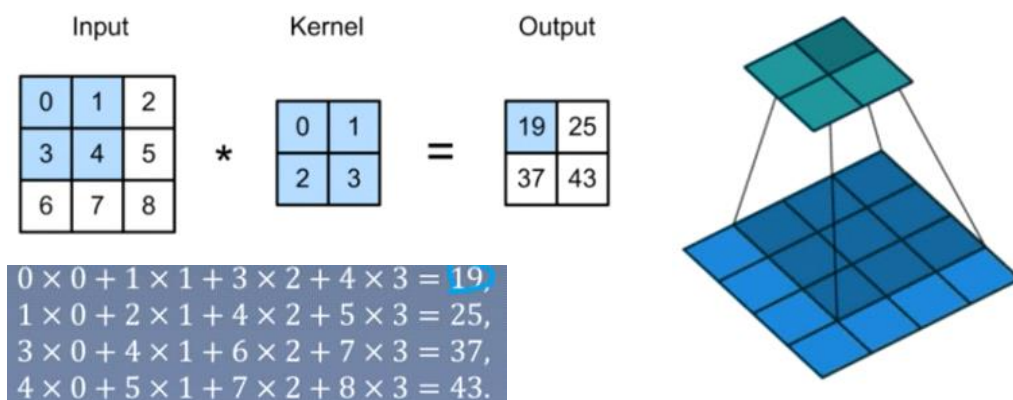
Convolutional Neural Network

2023年8月2日 18:45

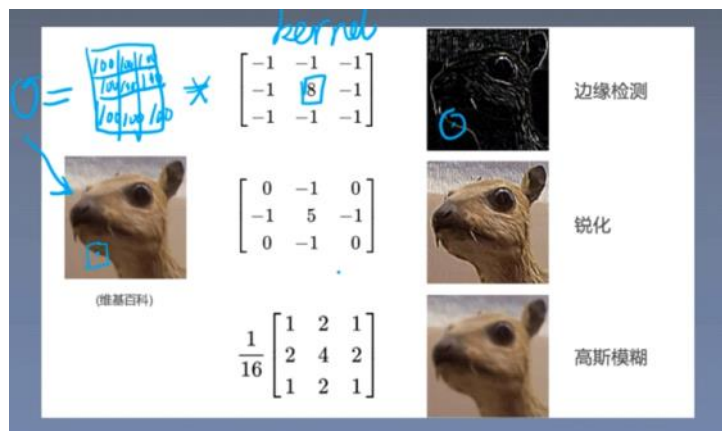
- 猫的视觉神经实验对CNN的启发：
 - a. 视觉系统是分层,分级的进行处理, 从低级到高级的抽象过程→堆叠使用卷积和池化
 - b. 神经元实际上是存在局部的感受区域的, 具体说来, 它们是局部敏感→神经元局部连接
- 第一个卷积神经网络雏形: 新认知机Neocognitron 1980 Kuniyiko Fukushima福岛邦彦
 - 缺点: 没有反向传播算法更新权值, 模型性能有限
- 第一个大规模商用卷积神经网络: Lenet-5 Lecun等人1989年, 成功应用在美国邮政系统中大规模手写邮政编码识别
 - 缺点: 无大量数据和高性能计算资源
- 第一个技惊四座的卷积神经网络: AlexNet 2012年赢得ILSVRC分类任务冠军, 从此拉开卷积神经网络统治图像领域的序幕
 - 成功要素: 算料: ImageNet
 - 算力: GPU (Nvidia GTX 5*2)
 - 算法: AlexNet
- 图像识别特点:
 1. 特征具有局部性: 卷积核每次仅连接K*K (卷积核尺寸) 区域



2. 特征可能出现在任何位置: 卷积核参数重复使用 (参数共享), 在图像上滑动

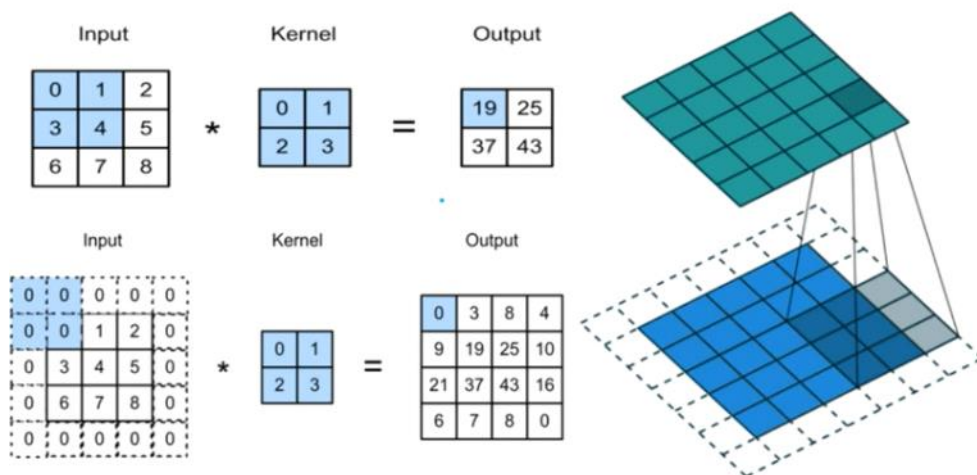


- 1) 卷积核Kernel: 具可学习参数的算子, 用于对输入图像进行特征提取, 输出通常称为特征图Feature maps

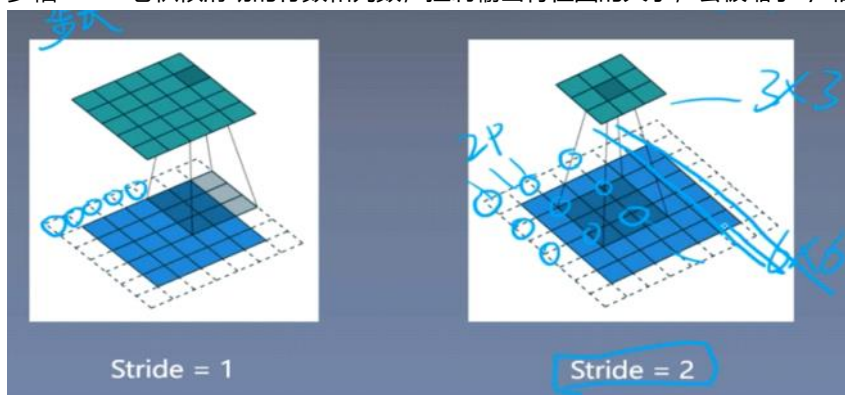


2) 填充 (Padding): 在输入图像的周围添加额外的行/列

- i) 使卷积后图像分辨率不变, 方便计算特征图尺寸的变化 ($p=1$, 原来是 5×5 池化后还是 5×5)
- ii) 弥补边界信息/像素丢失(边界贡献较少, 池化后弥补了这个缺点)

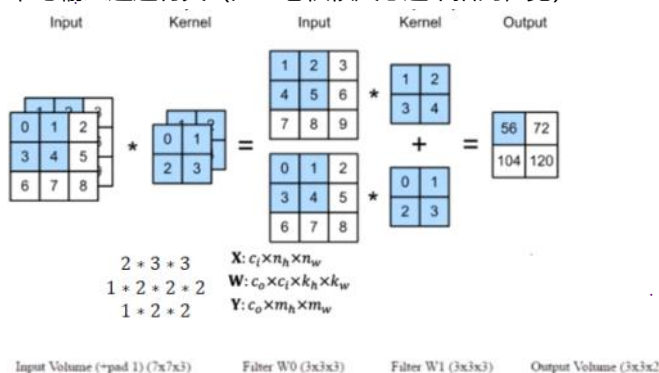


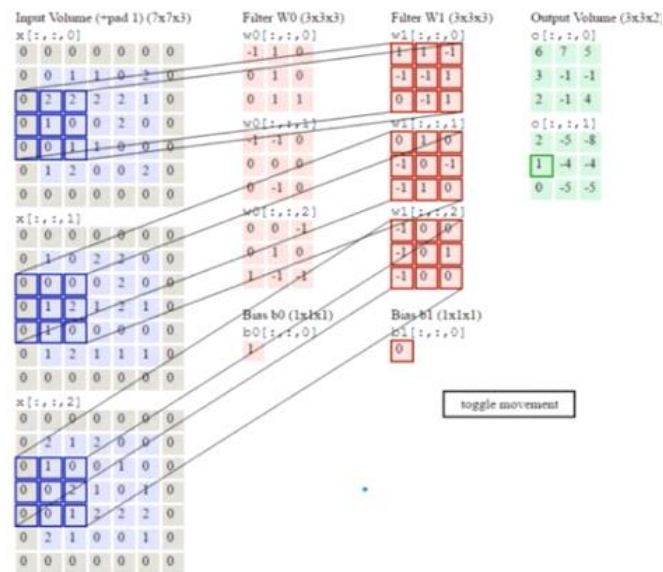
3) 步幅stride: 卷积核滑动的行数和列数, 控制输出特征图的大小, 会被缩小 $1/s$ 倍



4) 卷积层Convolutional Layer: 输出特征图尺寸计算: $F_{output} = \left\lfloor \frac{F_{input} - k + 2p}{s} \right\rfloor + 1$, s : stride

- i) 多通道卷积: RGB图像是 $3 \times h \times w$ 的三维数据, 第一个维度3, 表示channel, 通道数一个卷积核是3-D张量, 第一个维与输入通道有关 (注: 卷积核尺寸通常指高, 宽)





3. 下采样图像不改变图像目标:降低计算量, 减少特征冗余

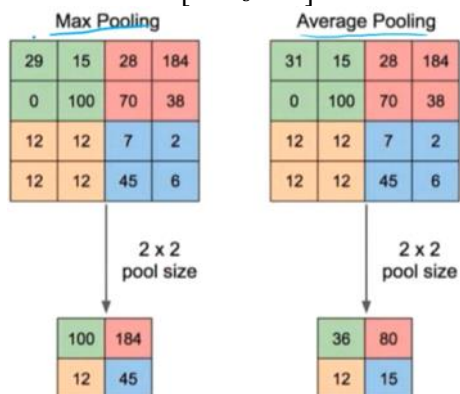
1) Pooling layer(池化层): 一个像素表示一块区域的像素值, 降低图像分辨率

方法:

- i) Max pooling:取最大值
- ii) Average pooling: 取平均值

输出尺寸计算与卷积操作类似, 注意: pooling layer无可学习参数。现在池化用的比较少, 一般用卷积设置 stride=2来代替, 池化是一种特殊的卷积。

$$F_{output} = \left\lfloor \frac{F_{input} - k + 2p}{s} \right\rfloor + 1$$



三个作用:

- 1) 缓解卷积层对位置的过度敏感: 与c细胞功能相似
- 2) 减少冗余
- 3) 降低图像分辨率, 从而减少参数量

Lenet-5: 堆叠使用卷积池化

