**学习进度**

本周学习任务：

学习吴恩达深度学习课程v5.72，关于卷积神经网络模块的内容。

1. 初识卷积

相对传统神经网络，卷积神经网络CNN的特殊之处，在于卷积计算。如图1.1所示，**卷积计算**本质上是一种基于矩阵的计算方式，它依赖能够提取特定特征的卷积核，按照预设步长，在矩阵表示的特征图中从左到右、由上及下的分步进行矩阵乘法运算，从而卷积计算出新的特征图。

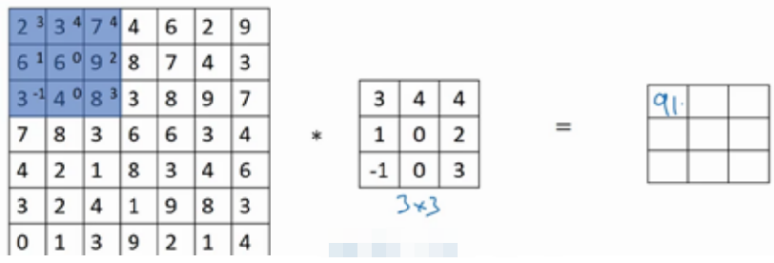


图1.1 卷积计算示例

为什么使用卷积？在较为复杂的计算情景中，和只用全连接层实现计算相比，使用卷积能够显著减少参数，有助于提高计算效率。这是由于卷积计算具有参数共享和稀疏连接的特性，前者是指一个卷积核可以适用于特征图中所有的局部区域，后者是指每个卷积结果都只与特征图中对应位置的特征值有关。

有哪些常用的卷积核？如图1.2所示，常用的过滤器（卷积核）有：垂直边缘检测过滤器、水平边缘检测过滤器、Scharr过滤器及Sobel过滤器。经验是：不一定非要使用给定的过滤器，可以把过滤器中的数字看做可学习参数，从而通过反向传播算法获得特定的过滤器。

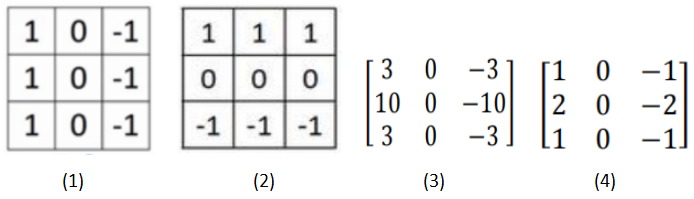


图1.2 垂直边缘检测、水平边缘检测、Scharr及Sobel过滤器

值得注意的是，每次卷积计算都会使得特征图的维度缩小，并且极易丢失图像边缘位置信息。为此，可在卷积前引入**Padding**操作，在图像外围填充圈值全为零的填充层，从而达到维持图像维度及充分利用边缘信息的目的。填充圈数的计算公式为：，其中是指过滤器的维度（为保证对称填充，一般设定为奇数）。

推广到**一次使用个不同卷积核进行三维卷积**的情况，特征图和过滤器都将有高、宽及通道数维度，且参与卷积运算的特征图和过滤器的通道数必须一致。举个例子，假设有维度的输入图像，个维度的过滤器，在无填充且步长为一的情况下，卷积后的特征图维度应是。

1. 卷积神经网络CNN

卷积神经网络由卷积层、池化层及全连接层等基本模块构成，网络结构并不固定。因此，应用CNN解决实际问题的关键在于：“如何整合基本模块，构建出更高效的卷积神经网络。”

基本模块介绍如下：

1. 卷积层

假设层输出的特征图维度为，则层输出的特征图维度为,,。此外，**卷积层**中的前向传播过程和普通神经网络殊途同归，以第一层卷积层为例，其激活值

1. 池化层

引入**池化层**，主要是为了降维提取特征图中的显著特征，以提高计算速度及提取特征的鲁棒性。常用的池化方法有：最大池化及平均池化。如图2.1所示，以最大池化为例，同卷积计算一样，池化计算按照预设步长及框选维度，在特征图中一步一步地提取局部区域内的最大值。此外，池化前后的特征图维度计算方法，同卷积层一致，此处省略。

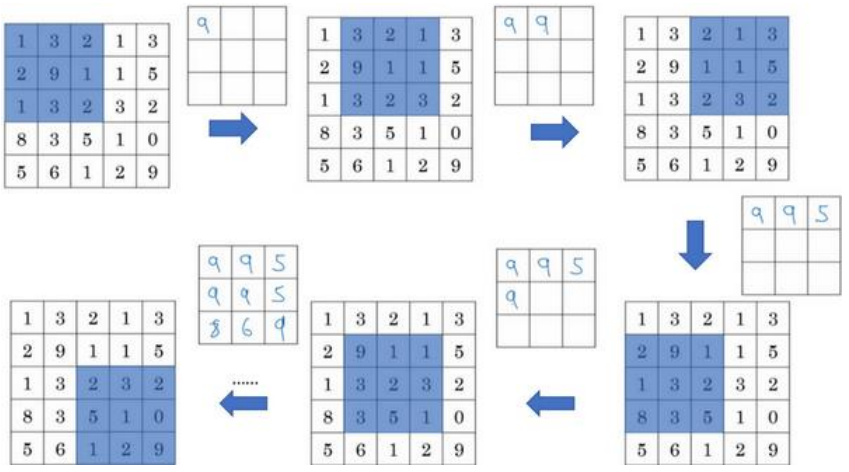


图2.1 max pooling过程

1. 全连接层

**全连接层**的作用是，综合卷积和池化层中提取到的特征，从而帮助网络模型实现对输入数据的高级特征提取和信息表示。值得注意的是，由于卷积原理，卷积及池化层的参数相对较少，而全连接层的参数相对较多。此外，在卷积神经网络中，不一定只有一层全连接层。

1. 卷积神经网络架构

卷积神经网络常用于处理计算机视觉领域的任务，为了更好地解决不同情景下的问题，随之发展出了许多不同的CNN网络架构，诸如：LeNet-5、AlexNet、VGGNet、ResNet及Inception等卷积神经网络。

如图3.1所示，**AlexNet网络**由n个卷积池化组合层、n个卷积层、1个池化层、n个全连接层及1个softmax输出层构成。相较于传统的LeNet-5网络，AlexNet网络拥有更复杂的网络结构，能够学习到更多层次的抽象特征表示，从而提高模型的理解与表达能力。此外，AlexNet网络使用了ReLU激活函数，引入了数据增强及Dropout正则化，极大地提高了模型的泛化性能。值得注意的是，AlexNet网络还具有并行计算能力，使其在ImageNet大规模视觉识别比赛中取得了突破性的成果。

graphic

图3.1 AlexNet网络架构

然而，更深的神经网络，关于梯度消失及梯度爆炸的风险也愈加明显。**ResNet网络**通过跳跃连接的方式，将浅层激活值作为残差块传递到更深层，可以帮助梯度更好地传播，从而避免信息的丢失。如图3.2所示，相较于传统神经网络架构，ResNet网络通过每两层额外增加一个传递通道的方式，实现了浅层激活值的跳跃传递。

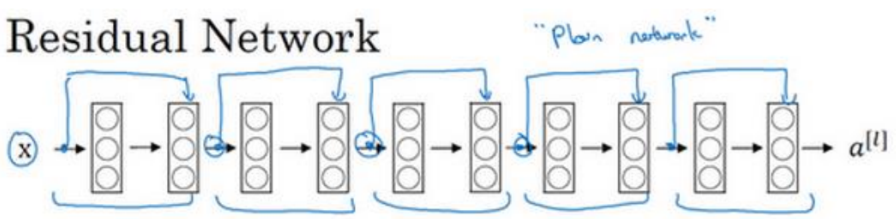


图3.2 ResNet网络架构

此外，ResNet能够准确完成跳跃连接的前提是，每个卷积层都使用same卷积，通过padding的方式保证了各层激活值的维度一致性。那么问题是：一定要遵守ResNet的标准架构吗？ResNet实际上并不限制每两层之间必须有一个跳跃连接。根据实际需求，可以根据网络的复杂性和任务的要求，决定在哪些层之间添加跳跃连接以及连接的方式。这种灵活性使得ResNet能够适应不同的网络深度和任务需求。

最后，回到一开始提出的关于“使用卷积神经网络的关键在于：如何整合基本模块”的问题上，诸如AlexNet、ResNet和Inception等卷积神经网络及其变体，都是应对一些具体情景中部分问题的经典整合方案。给予的启示是：使用前人创造的网络模型架构时，不应拘于形式，应该考虑具体的问题情景及模型架构的优缺点，理性引用甚至是优化网络模型架构。

1. 卷积神经网络在计算机视觉领域的应用

相较于序列数据，卷积神经网络能够在：目标定位、特征识别、目标检测、人脸识别及神经风格迁移等图像处理问题中，发挥地更为出色。

其中，**目标定位**不仅要判断图片中是否存在目标物体，还要在图片中标记出目标的位置。**特征识别**通过输出图片中目标物体的若干个特征点坐标，从而实现目标特征的识别判定，比如人体姿态检测及笑脸检测。如图4.1所示，**神经风格迁移**将风格图片的特征引入到内容图片中，从而创作出具有特定风格的新图片。

graphic

图4.1 神经风格迁移

在**人脸识别**领域，关键在于解决一次学习问题，需要仅通过一张图片，就能完成人脸识别任务。常规做法是定义并学习similarity函数，即：将两幅图片依次投入到同一参数下的神经网络模型中，分别计算出相应的图片特征编码，通过两幅图片特征编码之差的范数来计算它们之间的距离，从而实现访问对象人脸与数据库中合法用户人脸的匹配判定。因此，怎样获得图片中人脸的高质量特征编码，成了人脸识别任务的关键目标，常见的做法有三元组法及二分类法。以三元组法为例，通常需要同时看三张图片：Anchor图片、Positive图片和Negative图片，优化目标是。值得注意的是，三元组法需要用到同一个人的多张图片。

在**目标检测**领域，传统做法是：通过卷积计算的方式，在整张图片上进行滑动窗口目标检测，分步判定每个局部区域是否存在目标。然而，此类方法无法准确给出，目标物体的边界框位置。为此，**YOLO算法**应运而生，它是一种通过单次前向传播过程，直接预测目标类别及边界框精确位置的实时目标检测算法。此类算法使用交并比作为性能度量指标，其中A、B指代不同的边框。如图4.2所示，YOLO算法首先将图片分割为个网格，并基于网格依次卷积实现目标检测；然后每个网格都预测出由n个anchor box对象构成的向量，其中指预测为类的概率，分别指代边界框的中心点坐标及边框宽高，指检测物体的类别（若，则说明预测该物体为类）；最后利用极大值抑制及类别判定，消除冗余的anchor box对象。

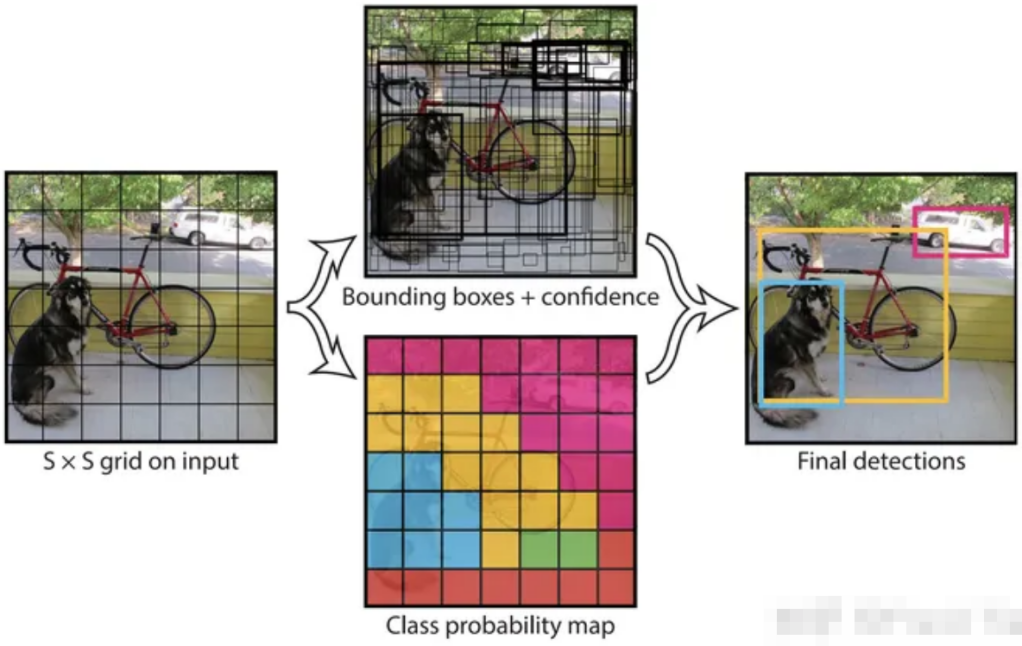


图4.2 YOLO算法工作过程

那么问题是：YOLO算法与传统目标检测方法都是基于卷积实现的，区别在哪？YOLO算法能够更精确地显式输出边界框坐标，而且输出的边界框可以具有任意宽高比，甚至可以超出所属局部网格的范围，并不会受到滑动窗口分类器的维度及步长限制。此外，YOLO算法为了保证模型在不同大小物体上的精确框定能力，需要对预测的边界框指标数据进行归一化处理，即：