**卷积神经网络**是一种带有卷积结构的深度神经网络，卷积结构可以减少深层网络占用的内存量，其三个关键的操作，其一是局部感受野，其二是权值共享，其三是pooling层，有效的减少了网络的参数个数，缓解了模型的过拟合问题。

**局部感受野与权值共享**

卷积神经网络的核心思想就是局部感受野、是权值共享和pooling层，以此来达到简化网络参数并使得网络具有一定程度的位移、尺度、缩放、非线性形变稳定性。

**局部感受野**：由于图像的空间联系是局部的，每个神经元不需要对全部的图像做感受，只需要感受局部特征即可，然后在更高层将这些感受得到的不同的局部神经元综合起来就可以得到全局的信息了，这样可以减少连接的数目。

**权值共享**：不同神经元之间的参数共享可以减少需要求解的参数，使用多种滤波器去卷积图像就会得到多种特征映射。权值共享其实就是对图像用同样的卷积核进行卷积操作，也就意味着第一个隐藏层的所有神经元所能检测到处于图像不同位置的完全相同的特征。其主要的能力就能检测到不同位置的同一类型特征，也就是卷积网络能很好的适应图像的小范围的平移性，即有较好的平移不变性（比如将输入图像的猫的位置移动之后，同样能够检测到猫的图像

**卷积层、下采样层、全连接层**

卷积层：因为通过卷积运算我们可以提取出图像的特征，通过卷积运算可以使得原始信号的某些特征增强，并且降低噪声。卷积核可以从输入数据中提取出特定的特征，例如边缘、角点、纹理等。卷积操作可以将卷积核在输入数据上进行滑动，每次滑动都会生成一个特征图，特征图中的每个元素都是卷积核与输入数据对应位置的乘积之和。通过多个卷积核的组合，可以提取出更加复杂的特征。

下采样层：下采样层有两个作用，一是减少计算量，防止过拟合；二是增大感受野，使得后面的卷积核能够学到更加全局的信息。因为对图像进行下采样，可以减少数据处理量同时保留有用信息，采样可以混淆特征的具体位置，因为某个特征找出来之后，它的位置已经不重要了，我们只需要这个特征和其他特征的相对位置，可以应对形变和扭曲带来的同类物体的变化。

全连接层：采用softmax全连接，得到的激活值即卷积神经网络提取到的图片特征。

填充（ Padding ）是指在输入数据的边缘添加一定数量的像素，使得输出数据的尺寸能够与输入数据相匹配。也就是在矩阵的边界上填充一些值，以增加矩阵的大小，通常用 0 或者复制边界像素来进行填充。

步长（Stride）是指卷积核在每一次卷积操作中滑动的距离。步长的大小可以影响输出数据的大小，也可以影响特征提取能力和计算复杂度。当步长增大时，输出数据的尺寸会减小，特征提取能力会变弱，但计算速度会加快。

通道数（Channel）也称为深度或特征图数量，是指卷积神经网络中每一层输出的特征图数量。通道数的大小直接影响了卷积神经网络的特征提取能力和计算复杂度。通过增加通道数，可以增强卷积神经网络的特征提取能力，但也会增加计算复杂度

AlexNet

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| layer\_name | kernel\_size | kernel\_num | padding | stride | input\_size | output\_size |
| Conv1 | 11 | 96 | [1, 2] | 4 | [224, 224, 3] | [55, 55, 96] |
| Maxpool1 | 3 | None | 0 | 2 | [55, 55, 96] | [27, 27, 96] |
| Conv2 | 5 | 256 | [2, 2] | 1 | [27, 27, 96] | [27, 27, 256] |
| Maxpool2 | 3 | None | 0 | 2 | [27, 27, 256] | [13, 13, 256] |
| Conv3 | 3 | 384 | [1, 1] | 1 | [13, 13, 256] | [13, 13, 384] |
| Conv4 | 3 | 384 | [1, 1] | 1 | [13, 13, 384] | [13, 13, 384] |
| Conv5 | 3 | 256 | [1, 1] | 1 | [13, 13, 384] | [13, 13, 256] |
| Maxpool3 | 3 | None | 0 | 2 | [13, 13, 256] | [6, 6, 256] |
| FC1 | 2048 |  |  |  |  |  |
| FC2 | 2048 |  |  |  |  |  |
| FC3 | 1000 |  |  |  |  |  |

(1)成功使用ReLU作为CNN的激活函数，并验证其效果在较深的网络超过了Sigmoid，成功解决了Sigmoid在网络较深时的梯度弥散问题。虽然ReLU激活函数在很久之前就被提出了，但是直到AlexNet的出现才将其发扬光大。

(2)训练时使用Dropout随机忽略一部分神经元，以避免模型过拟合。Dropout虽有单独的论文论述，但是AlexNet将其实用化，通过实践证实了它的效果。在AlexNet中主要是最后几个全连接层使用了Dropout。

(3)在CNN中使用重叠的最大池化。此前CNN中普遍使用平均池化，AlexNet全部使用最大池化，避免平均池化的模糊化效果。并且AlexNet中提出让步长比池化核的尺寸小，这样池化层的输出之间会有重叠和覆盖，提升了特征的丰富性。

(4)提出了LRN层，对局部神经元的活动创建竞争机制，使得其中响应比较大的值变得相对更大，并抑制其他反馈较小的神经元，增强了模型的泛化能力。LRN通过在相邻卷积核生成的feature map之间引入竞争，从而有些本来在feature map中显著的特征在A中更显著，而在相邻的其他feature map中被抑制，这样让不同卷积核产生的feature map之间的相关性变小。

**VGG网络**

VGG很好的继承了Alexnet的衣钵同时拥有着鲜明的特点。相比Alexnet ，VGG使用了更深的网络结构，证明了增加网络深度能够在一定程度上影响网络性能。

2.VGG的优点

(1)小卷积核组:作者通过堆叠多个3\*3的卷积核(少数使用1\*1）来替代大的卷积核，以减少所需参数；

(2)小池化核:相比较于AlexNet使用的3\*3的池化核，VGG全部为2\*2的池化核；

(3)网络更深特征图更宽:卷积核专注于扩大通道数，池化专注于缩小高和宽，使得模型更深更宽的同时，计算量的增加不断放缓；

(4)将卷积核替代全连接:作者在测试阶段将三个全连接层替换为三个卷积，使得测试得到的模型结构可以接收任意高度或宽度的输入。

(5)多尺度:作者从多尺度训练可以提升性能受到启发，训练和测试时使用整张图片的不同尺度的图像，以提高模型的性能。

(6)去掉了LRN层:作者发现深度网络中LRN（Local Response Normalization，局部响应归一化）层作用不明显。