

问题

一提到计算机视觉，第一反应就是 CNN，但是大家有没有想过，为什么图像识别领域的网络结构都是使用的 CNN 呢，或者说 CNN 网络有哪些特点可以使其在图像识别领域表现良好？这个问题对于我们而言都习惯到理所当然了，所以面试官要是突然问这种问题，估计很多同学都得懵逼一会。

分析

网上没有找着相关的问题答案，所以下面的分析仅仅是我个人的观点，大家交流交流，有其他见解的话非常欢迎提出来！

为什么图像识别领域要使用 CNN，其实潜在意思是在问 CNN 中的卷积层与全连接层相比好在哪里？为什么这么说，因为在卷积神经网络之前，一般的网络都采用的是全连接的方式，前一层的每一个单元都对下一层的每一个单元有影响，而 CNN 中虽然存在全连接层，但是更核心的是前面用于提取特征的卷积层，所以这个问题便转换成了卷积层和全连接层的比较问题。

卷积层和全连接层的区别

卷积层相比于全连接层，主要有两个特点：

1. **局部连接**：全连接层是一种稠密连接方式，而卷积层却只使用卷积核对局部进行处理，这种处理方式其实也刚好对应了图像的特点。在视觉识别中，关键性的图像特征、边缘、角点等只占据了整张图像的一小部分，相隔很远的像素之间存在联系和影响的可能性是很低的，而局部像素具有很强的相关性。
2. **共享参数**：如果借鉴全连接层的话，对于1000×1000大小的彩色图像，一层全连接层便对应于三百万数量级维的特征，即会导致庞大的参数量，不仅计算繁重，还会导致过拟合。而卷积层中，卷积核会与局部图像相互作用，是一种稀疏连接，大大减少了网络的参数量。另外从直观上理解，依靠卷积核的滑动去提取图像中不同位置的相同模式也刚好符合图像的特点，不同的卷积核提取不同的特征，组合起来后便可以提取到高级特征用于最后的识别检测了。

所以 CNN 应用的条件一般是要求卷积对象有局部相关性，这正是图像所具备的，因此图像领域都是使用 CNN 就解释地通了。

参考资料

[CNN为什么能使用在NLP？](#)

[\(二\)计算机视觉四大基本任务\(分类、定位、检测、分割\)](#)