用深度学习解决图像识别问题，从直观上讲是一个从**细节到抽象**的过程。所谓细节，就是指输入图像的每个像素点，甚至像素点构成的边也可以理解为是细节。假设我们大脑接收到一张动物图，大脑最先反应的是该图的点和边。然后由点和边抽象成各种形状，比如三角形或者圆形等，然后再抽象成耳朵和脸等特征。最后由这些特征决定该图属于哪种动物。深度学习识别图像也是同样的道理。这里关键的就是抽象。何为抽象呢？抽象就是把图像中的各种零散的特征通过某种方式汇总起来，形成新的特征。而利用这些新的特征可更好区分图像类别。如刚才这个例子，点和边就是零散的特征，通过将边进行汇总我们就得到了三角形或圆形等新的特征，同理，将三角形这个特征和一些其他零散的特征汇总成耳朵这个新特征。显而易见，耳朵这个新特征会比三角形特征更利于识别图像。

深度学习正是通过**卷积操作实现从细节到抽象**的过程。因为**卷积的目的就是为了从输入图像中提取特征，并保留像素间的空间关系。**何以理解这句话？我们输入的图像其实就是一些纹理，此时，可以将卷积核的参数也理解为纹理，我们目的是使得卷积核的纹理和图像相应位置的纹理尽可能一致。当把图像数据和卷积核的数值放在高维空间中，纹理等价于向量，卷积操作等价于向量的相乘，相乘的结果越大，说明两个向量方向越近，也即卷积核的纹理就更贴近于图像的纹理。因此，卷积后的新图像在具有卷积核纹理的区域信号会更强，其他区域则会较弱。这样，就可以实现从细节（像素点）抽象成更好区分的新特征（纹理）。每一层的卷积都会得到比上一次卷积更易区分的新特征。

而池化目的主要就是为了减少权重参数，但为什么可以以Maxpooling或者MeanPooling代表这个区域的特征呢？这样不会有可能损失了一些重要特征吗？这是因为**图像数据在连续区域具有相关性，一般局部区域的像素值差别不大**。比如眼睛的局部区域的像素点的值差别并不大，故我们使用Maxpooling或者MeanPooling并不会损失很多特征。