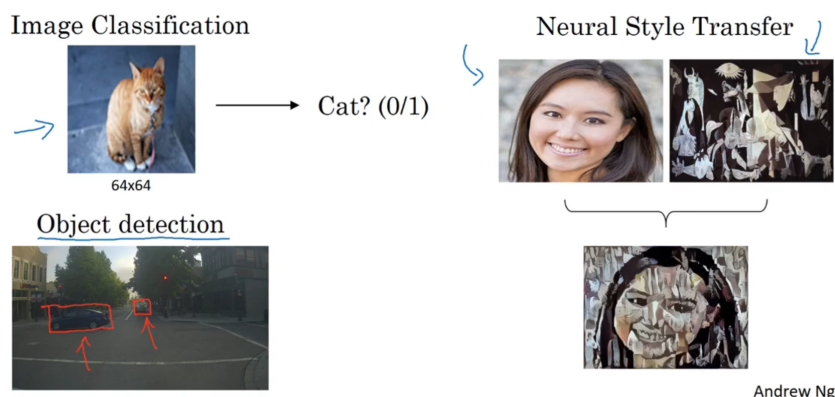


卷积神经网络基础

1. 计算机视觉

Computer Vision Problems



计算机视觉常见的问题：图片分类、目标识别、风格迁移等。

深度神经网络能够很好的处理尺寸较小的图片，然而在大尺寸图片上（比如 $1000 \times 1000 \times 3$ ），输入的规模非常大，使得学习参数的成本大增，这时神经网络便很难满足需要。

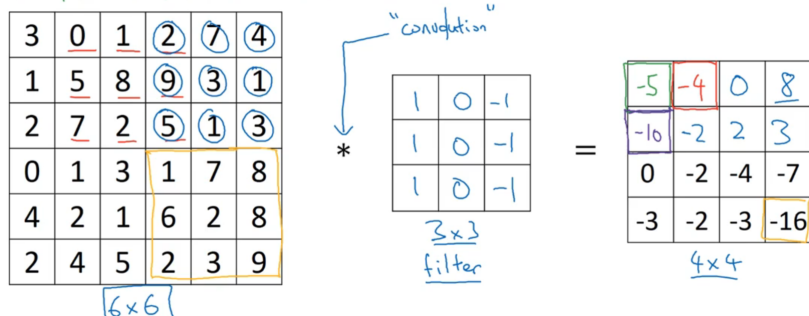
2. 边缘检测

2.1 垂直边缘检测

对于一个 $6 \times 6 \times 1$ 大小的图片，使用一个 3×3 大小的filter（卷积核）进行卷积运算（convolution）卷积运算符号为"*"。最终得到的图片大小为 4×4 ，如下图所示。

Vertical edge detection

$$3 \times 1 + 1 \times 1 + 2 \times 1 + 0 \times 0 + 5 \times 0 + 7 \times 0 + 1 \times -1 + 8 \times -1 + 2 \times -1 = -5$$



2.2 垂直边缘检测的原理

Vertical edge detection



- image-20220516150809781

3. padding

3.1 没有padding的缺点

- 如果一个 $n \times n$ 大小的图片，在利用 $f \times f$ 的filter进行卷积后，图片的大小会变成：

$$n \times n \xrightarrow{f \times f} (n - f + 1) \times (n - f + 1)$$

- 角落和边缘的像素进行卷积的次数较少，会丢掉一些信息。

3.2 加padding

相当于在图像周围填充了一层的像素，使得卷积计算过后图片的大小不变。

如果一个 $n \times n$ 大小的图片，在增加大小为 p 的padding后，再利用 $f \times f$ 的filter进行卷积，图片的大小会变成：

$$n \times n \xrightarrow{p} (n + 2p) \times (n + 2p) \xrightarrow{f \times f} (n + 2p - f + 1) \times (n + 2p - f + 1)$$

3.3 valid convolution和same convolution

- valid：不加padding， $(n - f + 1) \times (n - f + 1)$ 。
- same：加padding，输出图片的尺寸与原图相同。 $p = \frac{f-1}{2}$ ，一般来说 p 为奇数。

4. 卷积步长 (stride)

在之前的例子中，我们设置stride=1。下面是stride=2时的卷积结果：

 image-20220516155655739

用 s 表示stride的大小，进行卷积后，图像大小的变化为：

$$n \times n \rightarrow \left\lfloor \frac{n + 2p - f}{s} + 1 \right\rfloor \times \left\lfloor \frac{n + 2p - f}{s} + 1 \right\rfloor$$

5. 互相关与卷积

在数学上，卷积操作会将卷积核先水平反转后垂直反转，得到一个卷积核的镜像。但是在深度学习中一般不会这么操作。因此严格来说深度学习中的卷积应该被称为互相关。

6. 三维卷积

6.1 RGB图像中的卷积

灰色图像的卷积核一般是二维的。而在RGB图片中，图片有3个通道，此时的卷积核是三维的。卷积核的维度要与图片的通道数相同。

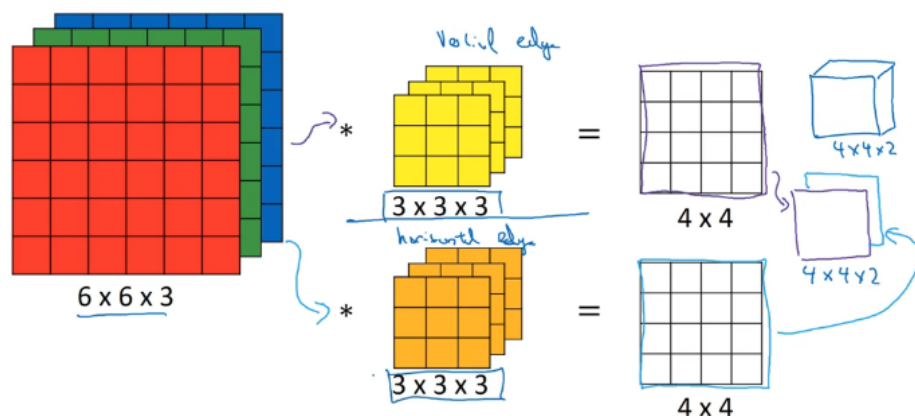
 image-20220516162710795

最终会得到一个 4×4 大小的图片。

6.2 多卷积核

使用垂直边缘检测和水平边缘检测，提取图像的垂直和水平的特征。

Multiple filters



6.3 summary

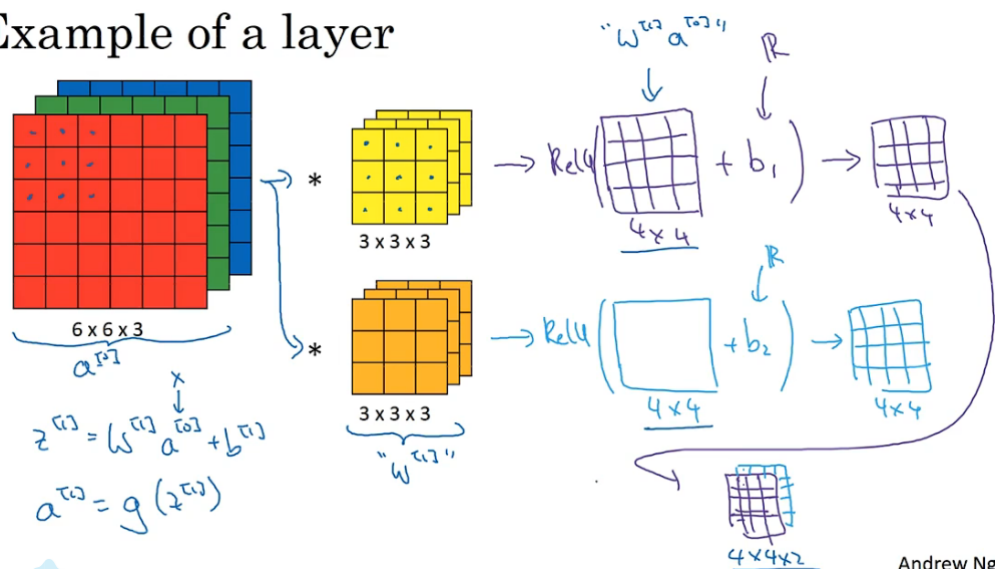
$$n \times n \times n_c * f \times f \times n_c \rightarrow (n - f + 1) \times (n - f + 1) \times n'_c$$

其中 n'_c 表示卷积核的个数（即下一层的通道数）。

7. 单层卷积网络

输入的图像作为 X ，卷积核视为权重 W ，因此与普通的神经网络不同的是，卷积神经网络的权重与数据做的是卷积运算。

Example of a layer



Andrew Ng

7.1 单层卷积的参数个数

10个卷积核，每个卷积核的大小为 $3 \times 3 \times 3$ ，则参数的个数为：

$$(3 \times 3 \times 3 + 1) \times 10 = 280$$

参数个数与input的图片大小无关。

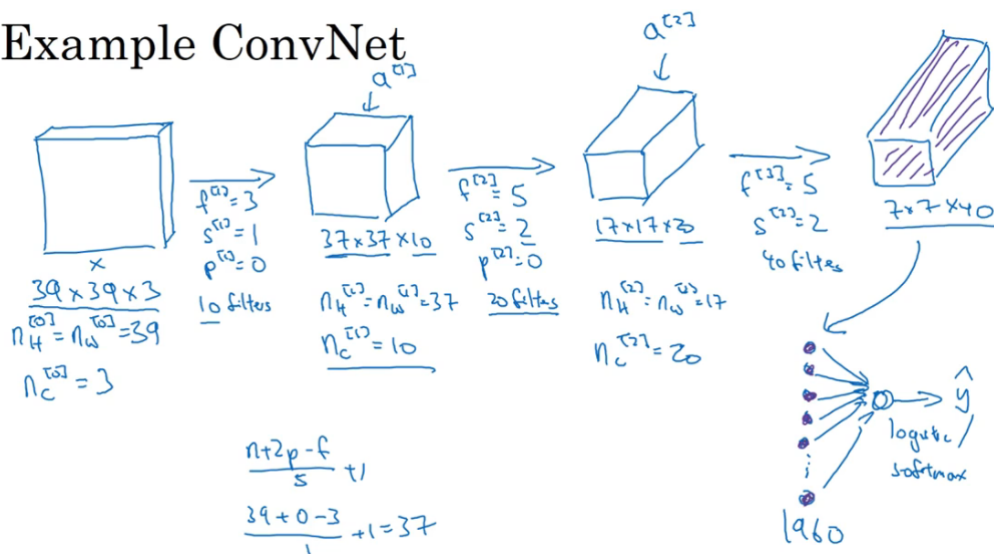
7.2 标记的总结

设 l 为一个卷积层：

- $f^{[l]}$: 卷积核/filter的大小;
- $p^{[l]}$: padding的大小;
- $s^{[l]}$: 步长stride的大小;
- input的图像大小: $n_H^{[l-1]} \times n_W^{[l-1]} \times n_c^{[l-1]}$
- output的大小: $n_H^{[l]} \times n_W^{[l]} \times n_c^{[l]}$, 其中 $n_H^{[l]} = \left\lfloor \frac{n_H^{[l-1]} + 2p^{[l]} - f^{[l]}}{s^{[l]}} + 1 \right\rfloor$, $n_W^{[l]}$ 同理
- $n_c^{[l]}$: 卷积核个数/通道数
- 卷积核大小: $f^{[l]} \times f^{[l]} \times n_c^{[l-1]}$
- Activations: $a^{[l]} \rightarrow n_H^{[l]} \times n_W^{[l]} \times n_c^{[l]}$
- Weights: $f^{[l]} \times f^{[l]} \times n_c^{[l-1]} \times n_c^{[l]}$
- bias: $n_c^{[l]} \rightarrow (1, 1, 1, n_c^{[l]})$

7.3 简单卷积神经网络

Example ConvNet



7.4 卷积网络层的类型

一般来说，卷积网络有三层：

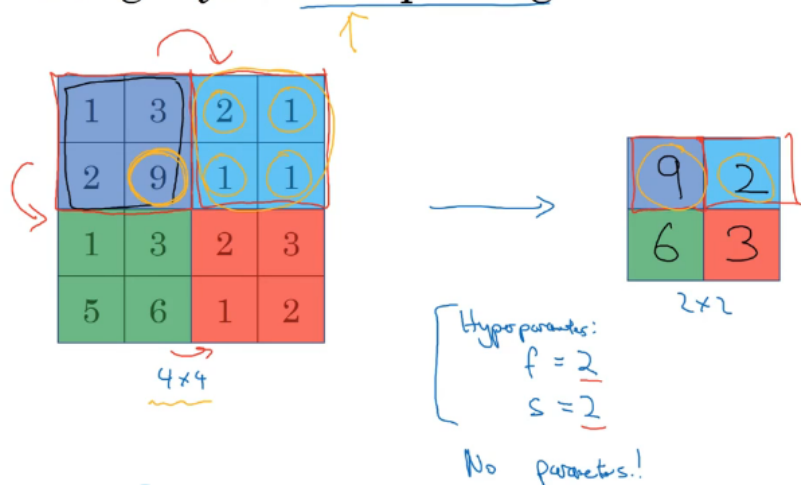
- 卷积层 (convolution)
- 池化层 (pooling)
- 全连接层 (FC, fully connected)

8. 池化层

8.1 Max pooling

对前一层的图像进行区域的划分，每个区域的最大值就是池化后的值。虽然pooling的大小是个超参数，但其实只是个固定运算。

Pooling layer: Max pooling



用 f 表示pooling的大小， s 表示stride大小：

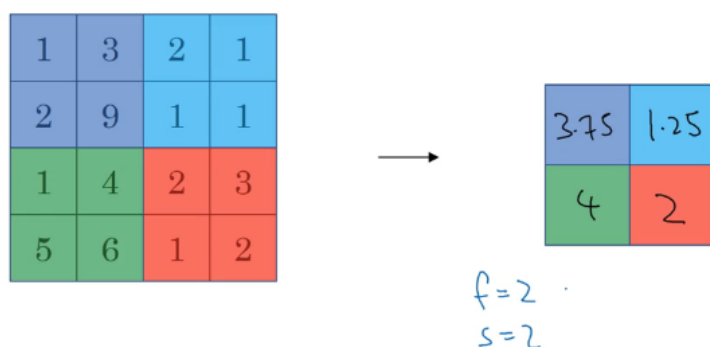
$$n \times n \rightarrow \left\lfloor \frac{n + 2p - f}{s} + 1 \right\rfloor \times \left\lfloor \frac{n + 2p - f}{s} + 1 \right\rfloor$$

注意，pooling后图像的通道数 n_c 不变。

8.2 Average pooling

选择区域的均值代表pooling后的区域。

Pooling layer: Average pooling



8.3 Summary of pooling

超参数（皆不需要学习）：

- f : pooling的filter的大小;
- s : stride的大小
- 选择Max pooling或者是Average pooling;
- p : padding, 基本不使用

pooling后的数据大小: $n_H \times n_W \times n_c \rightarrow \left\lfloor \frac{n-f}{s} + 1 \right\rfloor \times \left\lfloor \frac{n-f}{s} + 1 \right\rfloor \times n_c$

9. 卷积神经网络示例 (LeNet-5)

image-20220518200617692

8.1 深度卷积的常见模式

- 图片的大小逐渐减少, 但是通道数量逐渐增加。
- Conv - Pool - Conv - Pool - Fc - Fc - Fc - Softmax

8.2 深度卷积网络的激活函数

Neural network example

	Activation shape	Activation Size	# parameters
Input:	(32,32,3)	3,072 $a^{[0]}$	0
CONV1 (f=5, s=1)	(28,28,8)	<u>6,272</u>	208 ←
POOL1	(14,14,8)	<u>1,568</u>	0 ←
CONV2 (f=5, s=1)	(10,10,16)	1,600	416 ←
POOL2	(5,5,16)	400	0 ←
FC3	(120,1)	120	48,001 } 10,081 }
FC4	(84,1)	84	
Softmax	(10,1)	10	841

- Conv层的参数较少;
- Pooling层没有参数;
- 全连接层有大量参数;
- Activation Size逐渐减少, 但减少太快会影响整体的性能。

9. 为什么使用卷积?

9.1 卷积神经网络的优势

卷积神经网络的参数较少。如果使用全连接网络，则有 3072×4704 个参数。如果使用卷积神经网络，参数的个数为 $(5 \times 5 + 1) \times 6$ 个参数。

image-20220518201746303

- 参数共享：一个卷积核可以在图片的不同区域提取特征，不仅适用于低阶特征。
- 稀疏连接：每个输出单元仅取决于少量的输入。

9.2 训练卷积神经网络

image-20220518202804117