CS182 Introduction to Machine Learning

Recitation 8

2025.4.16

Outline

- Correlation & Convolution
- CNN
- Upsampling & Downsampling

Correlation 相关 & Convolution 卷积

Correlation

$$w(s,t) \star f(x,y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t) f(x+s, y+t)$$

Convolution

$$w(s,t) \bigstar f(x,y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t) f(x-s, y-t)$$

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

TABLE 3.5
Some fundamental properties of convolution and correlation. A dash means that the property does not hold.

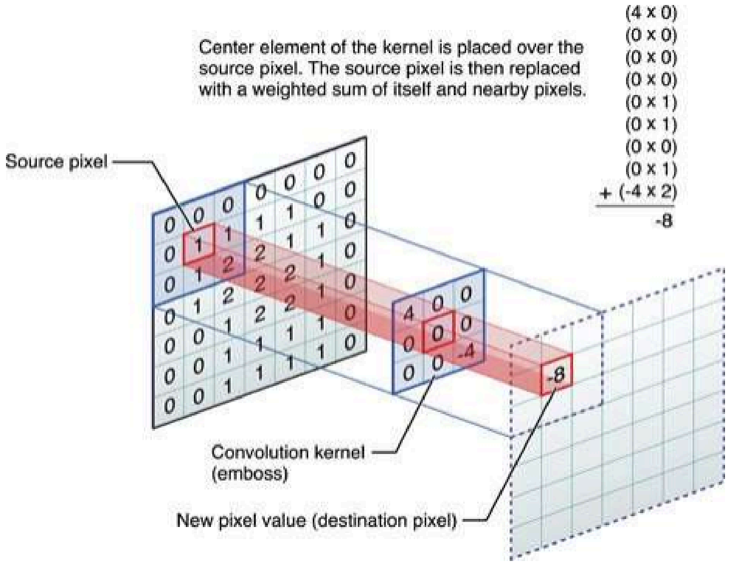
Property	Convolution	Correlation
Commutative	$f \bigstar g = g \bigstar f$	—
Associative	$f \bigstar (g \bigstar h) = (f \bigstar g) \bigstar h$	—
Distributive	$f \bigstar (g+h) = (f \bigstar g) + (f \bigstar h)$	$f \star (g+h) = (f \star g) + (f \star h)$

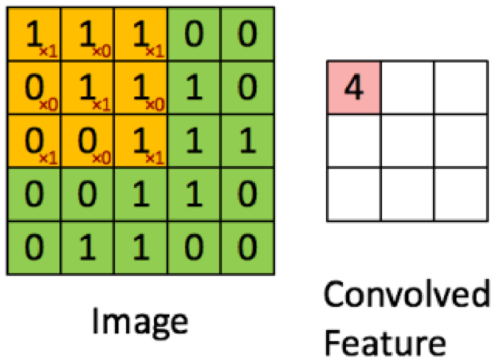
卷积 = 卷积核反转一下之后做相关

2D Convolution

If A and B are two 2-D arrays, then:

$$(A * B)_{ij} = \sum_s \sum_t A_{st} B_{i-s,j-t}.$$



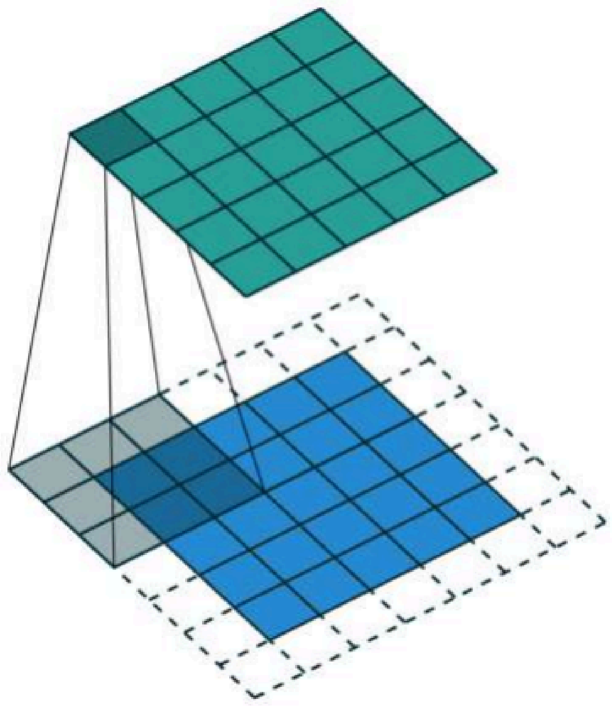


卷积核反转一下就是做一个相关, 因为卷积核的参数是可训练的, 所以在CNN里具体是做的卷积还是做的相关其实并不重要. 实际上torch就是用correlation算的

https://pytorch.org/docs/stable/generated/torch.nn.Conv2d.html#torch.nn.Conv2d

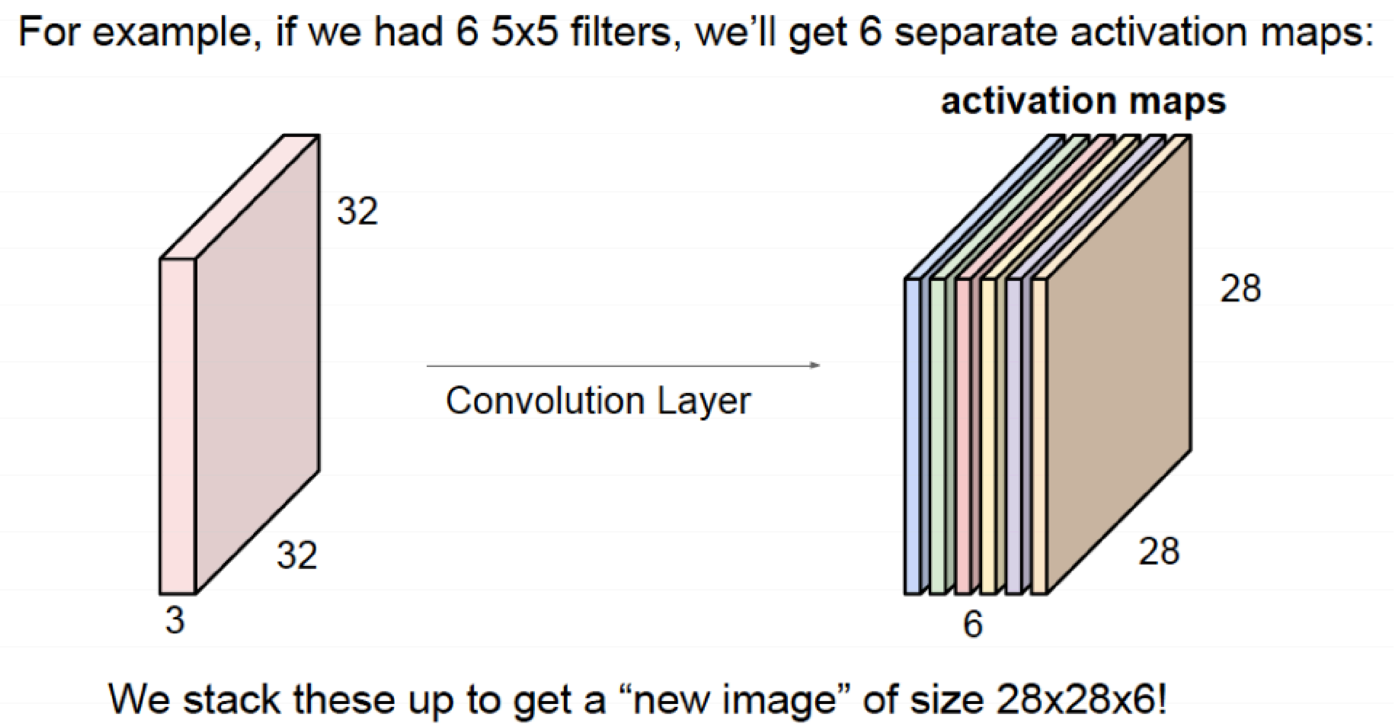
Padding 填充

Zero padding to handle non-integer cases or control the output sizes



padding_mode='zeros' / 'reflect' / 'replicate' / 'circular', default='zeros'

Dimension 维度



一个32*32*3的图像, 经过6个5*5*3的卷积核, 没有padding, 步长为1: 变成28*28*6的特征图.

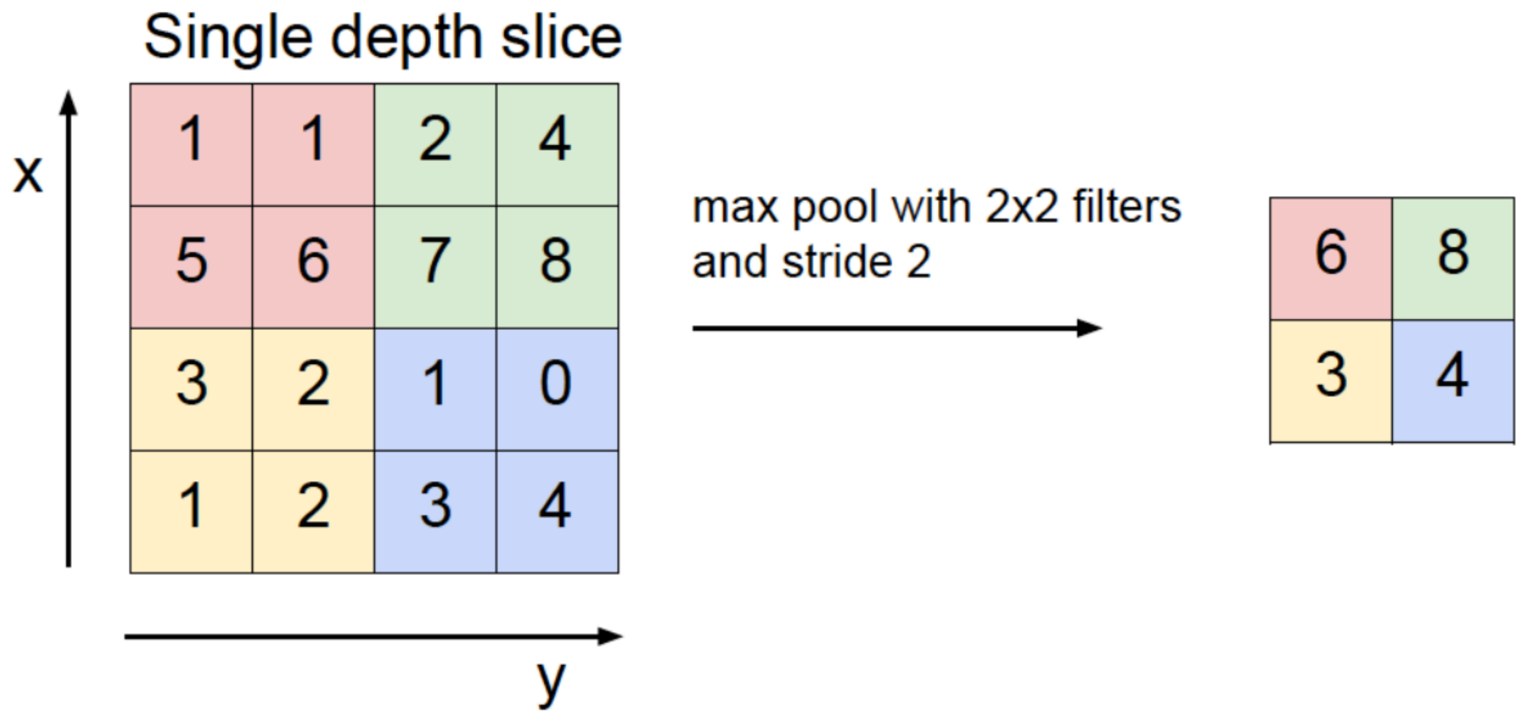
> special: $W*H*D$的图像经过n个$1*1*D$的卷积核, 得到$W*H*n$的特征图.
> 作用: 改变维度数 (feature层面的pooling)

Dimension of convolution operation 卷积操作的维度

- Accepts a volume of size $W_1 \times H_1 \times D_1$
- Requires four hyperparameters:
  - Number of filters K,
  - their spatial extent F,
  - the stride S,
  - the amount of zero padding P.
- Produces a volume of size $W_2 \times H_2 \times D_2$ where:
  - $W_2 = (W_1 - F + 2P)/S + 1$
  - $H_2 = (H_1 - F + 2P)/S + 1$ (i.e. width and height are computed equally by symmetry)
  - $D_2 = K$
- With parameter sharing, it introduces $F \cdot F \cdot D_1$ weights per filter, for a total of $(F \cdot F \cdot D_1) \cdot K$ weights and K biases.
- In the output volume, the d-th depth slice (of size $W_2 \times H_2$) is the result of performing a valid convolution of the d-th filter over the input volume with a stride of S, and then offset by d-th bias.

一个$W_1 * H_1 * D_1$的图像, 经过K个$F * F * D_1$的卷积核, padding的大小为P, 步长为S最终得到一个$W_2 * H_2 * D_2$的特征图

Upsampling 上采样 & Downsampling 下采样



- upsample: 时域 / 频域 补零
- downsample: max pooling / average pooling / ...

Dimension of pooling operation 池化操作的维度

- Accepts a volume of size $W_1 \times H_1 \times D_1$
- Requires three hyperparameters:
  - their spatial extent F,
  - the stride S,
- Produces a volume of size $W_2 \times H_2 \times D_2$ where:
  - $W_2 = (W_1 - F)/S + 1$
  - $H_2 = (H_1 - F)/S + 1$
  - $D_2 = D_1$
- Introduces zero parameters since it computes a fixed function of the input
- Note that it is not common to use zero-padding for Pooling layers

> pooling操作只改变图像的大小, 不改变特征的维度
> 一个$W_1 * H_1 * D_1$的图像, 经过**感受野**$F * F$池化, 步长为S
> 最终得到一个$W_2 * H_2 * D_1$的特征图

Have a try!

- LeNet-5