

Операция свертки и ее вариации

Рассматривается дискретное пространство. Операция свертки имеет вид:

$$S(x, y) = \sum_{m=-a}^a \sum_{n=-b}^b k(n, m) I(x - n, y - m) \quad (1)$$

Корреляция имеет вид:

$$R(n, m) = \sum_{m=-a}^a \sum_{n=-b}^b k(n, m) I(x + n, y + m) \quad (2)$$

В нейросетях, используется операция корреляции с использованием bias или же кросс корреляция. В частности в библиотеке pytorch используется функция взаимной-корреляции

$$R(n, m) = bias + \sum_{m=-a/s}^{a/s} \sum_{n=-b/s}^{b/s} k(sn, sm) I(x + sn, y + sm) \quad (3)$$

Где s-stride

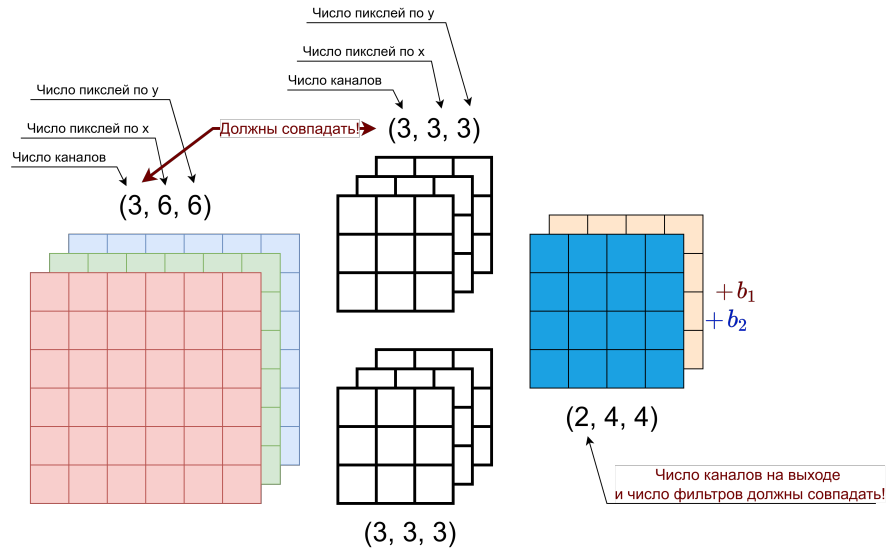


Figure 1: Описание размерностей для сверток в нейросети

Функция `torch.nn.conv2d(in_channels, out_channels, kernel_size, stride=1, padding=0, dilation=1, groups, bias=True, padding_mode=device=None, dtype=None)`

Рассмотрим операцию свертки. Обратим внимание на пулинг `pooling=(1,1)` и `stride=2`.

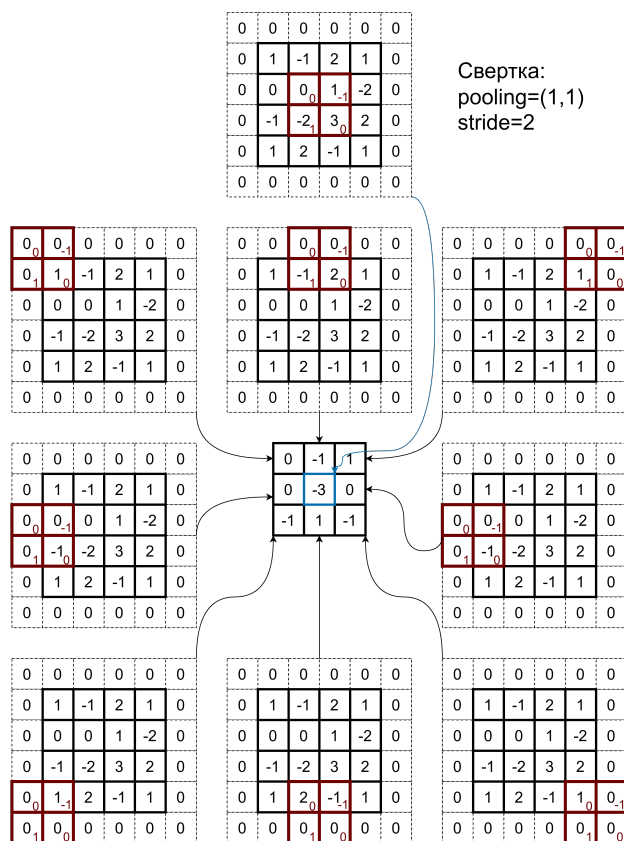


Figure 2: Операция свертки(вернее будет взаимная корреляция, но принято называть сверткой)

Transposed convolution

Transposed convolution или расширенную свертку применяют, когда необходимо из меньшего размера изображения получить большее.

`torch.nn.ConvTranspose2d(in_channels, out_channels, kernel_size, stride=1, padding=0, output_padding=0, groups=1, bias=True, dilation=1, padding_mode='zeros', device=None, dtype=None)`

1. `in_channels` - число входных каналов
2. `out_channels` - число выходных каналов
3. `kernel_size` - размерность ядра
4. `stride` - шаг расширенной свертки(расстояние от любого левого края левого пикселя до левого края следующего правого пикселя) по всем осям
5. `padding` - вставка нулей по окраине
6. `output_padding` - добавляется только справа сверху с одной стороны результата расширенной свертки
7. `dilation` - шаг между ячейками ядра при свертке https://github.com/vdumoulin/conv_arithmetic/blob/master/README.md

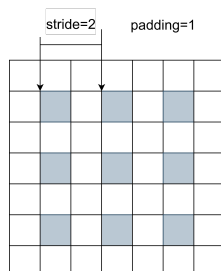


Figure 3: Некоторые обозначения входа расширенной свертки

$$H_{out} = (H_{in} - 1)s - 2p_1 + d(k - 1) + p_2 + 1 \quad (4)$$

- H_{in} - размерность входа вдоль оси
- s - расстояние
- p_1 - паддинг с обеих сторон исходного изображения
- k - размер ядра
- p_2 - односторонний паддинг исходного изображения