

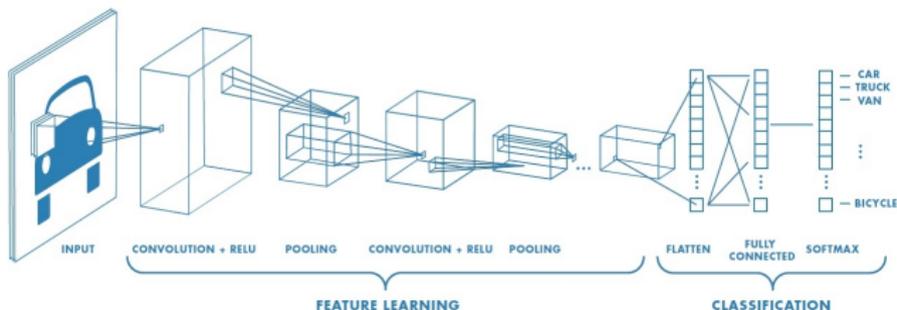
Сверточные нейросети

Виктор Китов

v.v.kitov@yandex.ru

Сверточные нейросети

- Сверточная нейросеть:
 - используется для обработки изображений
 - операции свертки (с нелинейностями) и слои пулинга.
 - MLP в конце, если решается задача классификации или регрессии.

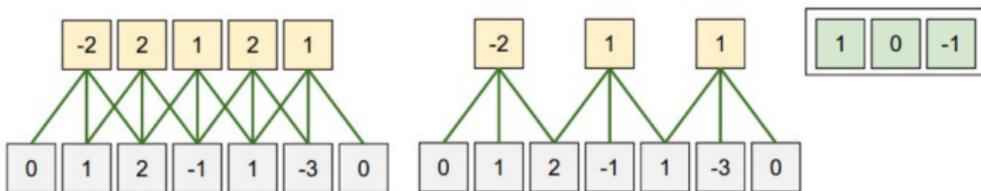


Содержание

1 Операция свертки

2 Пулинг

Одномерная свертка



$$out1D(x, y) = \sum_{i=-n}^n K(i+n+1)in(x+i)+b, \quad Kernel \in \mathbb{R}^{2n+1}, \quad b \in \mathbb{R}$$

Параметры¹:

- W - ширина входа; $2n + 1$ - размер ядра
- P - расширение (padding) для увеличения размера выхода
- Тип расширения (valid [отсутствует], zero [нулями], same [повтор], mirror [отражение])
- S - шаг (stride) - при применении; D - смещение (dilation) - внутри свертки

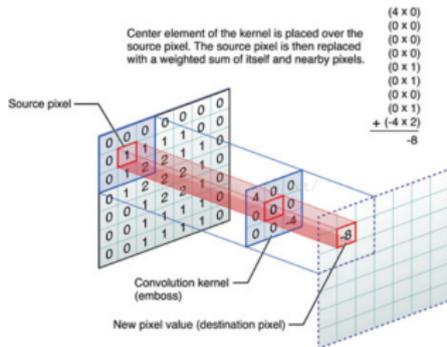
¹Какой будет размер выхода свертки при данных параметрах?

Примеры сверток

- $K = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$ - равномерное усреднение.
- $K = (0.1, 0.2, 0.4, 0.2, 0.1)$ - усреднение с убывающими весами.
- $K = (-1, 0, +1)$ - величина изменения ($f' \approx z_{t+1} - z_{t-1}$)
- $K = (+1, -2, +1)$ - динамика величины изменения ($f'' \approx (z_{t+1} - z_t) - (z_t - z_{t-1})$)

Двумерная свертка

Двумерная свертка



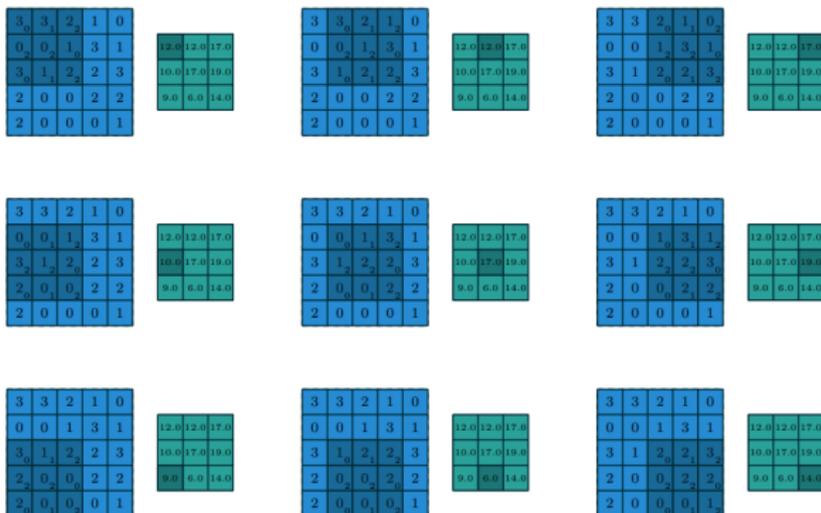
$$out2D(x, y) =$$

$$\sum_{i=-n}^n \sum_{j=-n}^n K(i+n+1, j+n+1) in(x+i, y+j) + b,$$

$$K \in \mathbb{R}^{(2n+1) \times (2n+1)}, b \in \mathbb{R}$$

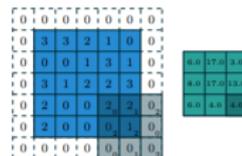
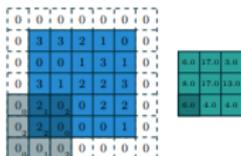
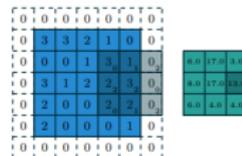
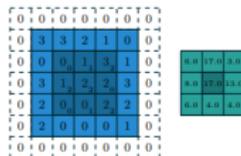
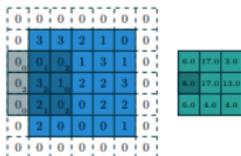
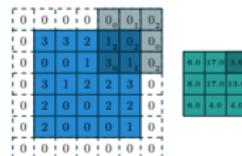
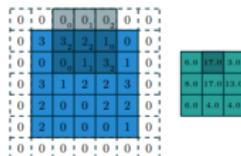
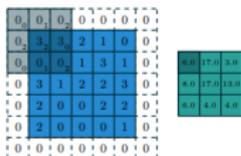
Пример: применение двумерной свертки²

Применение двумерной свертки:

²Иллюстрации: Dumoulin et al. 2018.

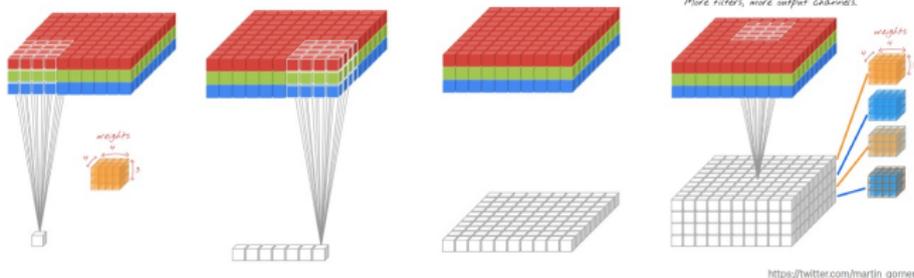
Пример: zero padding и stride=2

Свертка с расширением (zero padding=1) и смещением (stride=2):



Трехмерная свертка

Трехмерная свертка



$$out3D(x, y, c) =$$

$$\sum_{i=-n}^n \sum_{j=-n}^n \sum_{c=1}^C K(i+n+1, j+n+1, c) in(x+i, y+j, c) + b,$$

$$K \in \mathbb{R}^{(2n+1) \times (2n+1)}, b \in \mathbb{R}$$

Комментарии

Комментарии:

- #параметров↓:
 - нейрон имеет связь только с пространственно близкими
 - применяется одинаковое преобразование к разным позициям
- #параметров= $((2n + 1) \times (2n + 1) C_{in} + 1) C_{out}$ для набора C_{out} сверток, примененных к тензору с C_{in} каналами.
- Область видимости свертки $\pm n$ относительно предыдущего слоя.
 - увеличивается для более ранних слоев.
- Stride>1: уменьшение пространственного разрешения.
- После свертки должно следовать нелинейное преобразование (напр. ReLU, LeakyReLU).
- Часто используют набор предобученных сверточных слоев (например, по ImageNet)
- Есть физиологические подтверждения сверток в мозге.

Визуализация сверток 1го слоя сети AlexNet

Визуализация сверток 1го слоя сети AlexNet

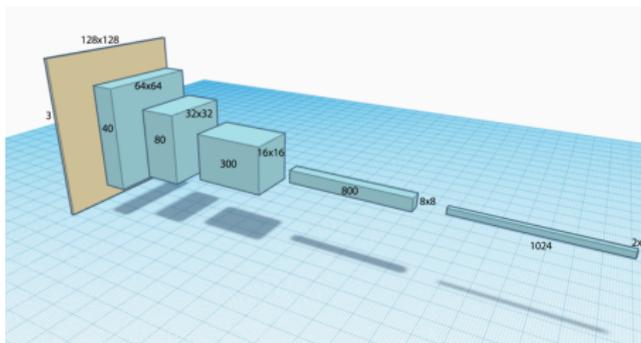


- Свертки 1го слоя используют ядра $\in \mathbb{R}^{(2n+1) \times (2n+1) \times 3}$, поэтому м. быть визуализированы как картинки.
- Последующие свертки содержат больше каналов, поэтому способы визуализации:
 - подбор существующих патчей, максимизирующих активацию свертки

Пирамида свертки

- Последующие свертки извлекают все более сложные признаки:
 - пример: цвета->границы->линии, углы, изгибы->геометрические фигуры->глаза, нос, рот, брови
- Типичная схема: постепенно ↓ пространственное разрешение и ↑ число каналов.
 - ↓ пространственное разрешение: пилинги или свертки с $\text{stride} > 1$

Пирамида свертки



Содержание

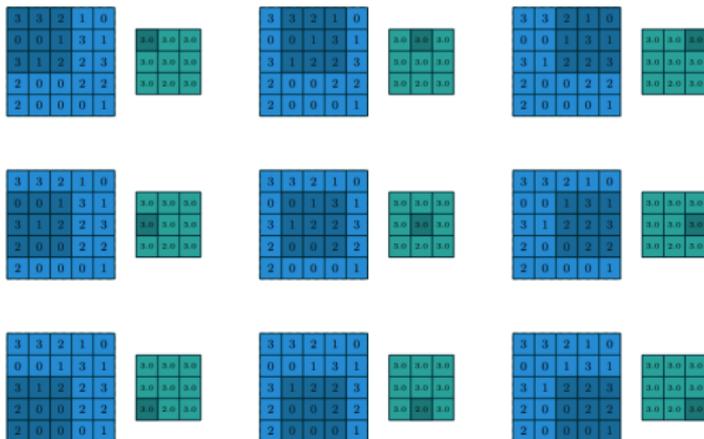
1 Операция свертки

2 Пулинг

Max pooling

- Max пулинг (pooling): "признак присутствует где-то в области" (например, детектор угла)
- Добавляет инвариантность к небольшим сдвигам изображения.
- $\text{Stride} > 1$: уменьшение пространственного разрешения.

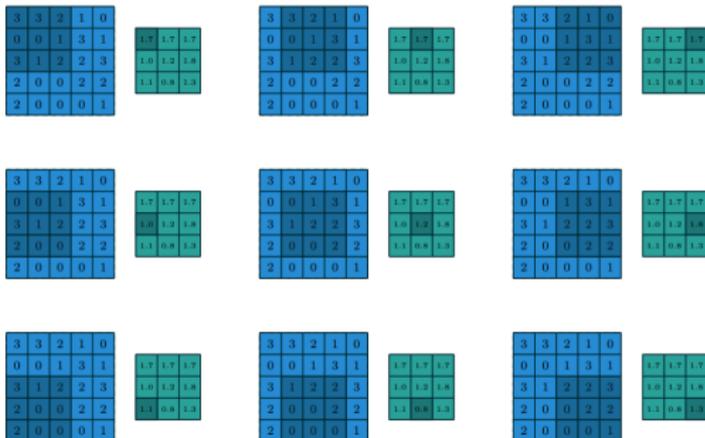
Max 3x3 pooling:



Average pooling

- Average pooling: "средняя представленность признака" (например, для детектора изолированных точек).
- Добавляет инвариантность к небольшим сдвигам изображения.
- $\text{Stride} > 1$: уменьшение пространственного разрешения.

Average 3x3 pooling:



Повышение пространственного разрешения (upscaling)

Повышение пространственного разрешения (upscaling)

возможно за счет расширения входа для свертки:

- транспонированная свертка (transposed convolution) - принцип "bed of nails", заполнение нулями

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a & 0 & b & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & c & 0 & d & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- перемасштабирование (ближайшим соседом, с линейной/квадратичной интерполяцией)

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \rightarrow \begin{pmatrix} a & a & b & b \\ a & a & b & b \\ c & c & d & d \\ c & c & d & d \end{pmatrix}$$

Пространственный тензор \rightarrow вектор фикс. размера

- Классификация, регрессия - в конце MLP.
 - MLP ожидает вход фикс. размера
- Решения:
 - привести входное изображение к фикс. размеру
 - перемасштабировать & обрезать
 - приводит к потере информации и искажению
 - глобальный пулинг (global pooling): $\mathbb{R}^{C \times W \times H} \rightarrow \mathbb{R}^C$
 - информация сокращается слишком резко
 - полностью исчезает пространственная информация
 - пирамидальный пулинг (pyramid pooling): $\mathbb{R}^{C \times W \times H} \rightarrow \mathbb{R}^{Cm^2}$
 - 1 разбиваем выход пространственной сеткой $m \times m$.
 - 2 применяем глобальный пулинг отдельно к каждому элементу сетки
 - 3 конкатенируем результаты

Заключение

- Полносвязные слои содержат много параметров
- Свертки решают это:
 - используя только локальные связи
 - применяя одинаковое преобразование
- Average и max pooling:
 - снижение пространственного разрешения
 - инвариантность к небольшим сдвигам
- Последующие свертки извлекают более сложные и глобальные признаки.