

Deep Learning. Сверточные нейронные сети (CNN)

Егор Конягин

20 августа 2024 г.

1. Недостатки полносвязных нейросетей
2. Понятие свёртки функции/изображения
3. Обучаемая свёртка - основа CNN
4. LeNet - 5
5. Гиперпараметры сверточных нейронных сетей
6. Summary

Недостатки полносвязных нейросетей

Рассмотрим задачу, связанную с бинарной классификацией цветных изображений размера 256×256 . Положим, что слоев в такой нейросети от 5 до 10.

Вопрос оцените по порядку величины кол-во параметров данной нейросети.

Подсказка: $256 \cdot 256 \cdot 3 = 196608$.

Если предположить, что на первом слое размерность уменьшается в 1000 раз, то матрица весов будет иметь порядка 40 000 000 параметров! Это только первый слой. Поэтому такая архитектура будет слишком энергозатратной даже для несложных задач.

Анализ соседних пикселей

Рассмотрим фрагмент изображения, имеющий следующий вид:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{pmatrix} \quad (1)$$

При преобразовании этой матрицы в вектор элементы, "стоящие рядом" окажутся далеки друг от друга. Таким образом, снизится возможность учитывать информацию не только о конкретном признаке, но и об окружающих его пикселях.

Понятие свёртки функции/изображения

Свертка функции. Свертка изображения

В математике под сверткой двух функций подразумевают следующую величину:

$$(f * g)(x) = \int_{\mathcal{R}^n} f(y)g(x - y)dy \quad (2)$$

В дискретном двумерном случае:

$$g(x, y) = (\mathcal{K} * f)(x, y) = \sum_{s=0}^{n_k} \sum_{t=0}^{n_k} \mathcal{K}_{st} \cdot f_{x-s, y-t}. \quad (3)$$

Таким образом, свертка от изображения - это тоже изображение, но каждое его значение было получено из линейной комбинации пикселей на исходном изображении.

Свертка изображения. Иллюстрация

Рассмотрим, как же получается свертка и какие детали она может выявлять на изображении:

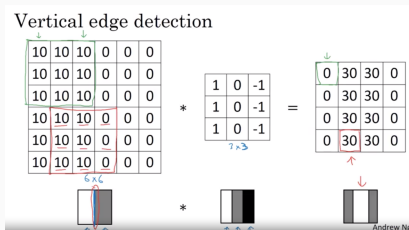


Рис. 1: Вычисление свертки изображения. Источник: Andrew Ng's classes

Как видно из данного примера, свертка с таким фильтром позволяет находить вертикальные линии на изображении.

ВОПРОС: как будет выглядеть свертка для поиска горизонтальных линий?

Для решения этой задачи необязательно использовать именно такую свертку. Достаточно популярными в анализе изображения являются Sobel filter и Scharr filter:

$$\begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} \quad \begin{pmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{pmatrix} \quad (4)$$

Обучаемая свёртка - основа CNN

Рассмотрим матрицу фильтров как совокупность параметров, значения которых будут подбираться:

$$W = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{pmatrix} \quad (5)$$

Для того чтобы эти параметры обучались, надо написать для них шаги back propagation. В 1989 году эту задачу решил математик, специалист по компьютерному зрению Yann LeCun. Мы не будем рассматривать уравнения back propagation для сверток, поскольку они имеют слишком много выкладок.

Max pooling. Average pooling

Помимо операции свертки, сверточные нейронные сети часто выполняют еще одну операцию - пулинг. Она служит для понижения размерности изображения.

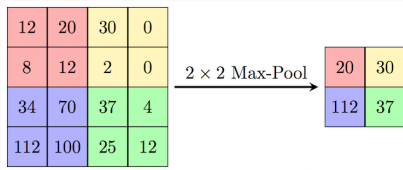


Рис. 2: Вычисление max pooling в каждом из 4-ех фрагментов. Источник: Andrew Ng's classes

Average pooling - это взятие не максимального значения пикселя из фрагмента изображения, а вычисление и взятие среднего от фрагмента.

LeNet - 5

LeNet - это сверточная нейронная сеть, спроектированная тем же специалистом (Yann LeCun) в 1998 году. Она предназначалась для распознавания символов (а именно цифр) по фотографиям.

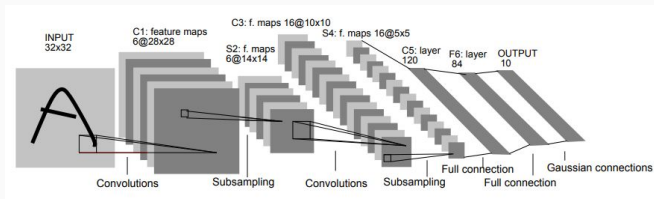


Рис. 3: Архитектура LeNet-5 Источник: Yann LeCun's original paper

Данная нейронная сеть выполняет уже не задачу бинарной классификации, а задачу многоклассовой классификации!

Функция Softmax

Функция softmax - это ещё одна популярная функция активации, используемая в нейронных сетях. Отличается она от других тем, что в качестве аргумента эта функция принимает не число (как было с функциями ReLU, tanh и сигмной), а вектор. Тогда i -ая координата softmax-а задается следующим уравнением:

$$\sigma(\mathbf{z})_i = \frac{\exp z_i}{\sum_{j=1}^{n_c} \exp z_j}. \quad (6)$$

Заметим, что сумма значений всех координат равна единице. Тогда i -ую координату можно интерпретировать как вероятность принадлежности объекта к i -ому классу.

Softmax используется в задаче многоклассовой классификации.

Гиперпараметры сверточных нейронных сетей

Сверточный слой

Под сверточным слоем обычно подразумевают композицию непосредственно операции свертки и пулинга.

Гиперпараметры свёрточной части:

- кол-во фильтров ($n_{filters}$);
- размер фильтров (kernel size);
- значение padding;
- stride;
- функция активации.

Гиперпараметры pooling-части:

- вид пулинга;
- размер фрагмента.

Гиперпараметры сверточной части

В одном сверточном слое обычно производится свертка не одним фильтром, а несколькими. В таком случае кол-во каналов на выходе будет равно как раз кол-ву сверток в слое. Каждая свёртка в слое обязана иметь одинаковый размер (размерность матрицы фильтра).

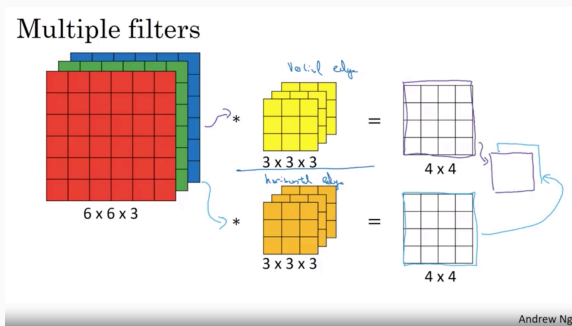


Рис. 4: Свертка над многоканальным изображением. Источник: Andrew Ng's classes

Понятие padding

Padding - это операция, при которой вокруг изображения "рисуются рамка из нулей":

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{21} & a_{22} & a_{23} & 0 \\ 0 & a_{31} & a_{32} & a_{33} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (7)$$

Padding нужен, чтобы контролировать размерность изображения на слоях. Чаще всего он не указывается в явном виде, указывается лишь т.н. режим:

- valid - при данном режиме padding не применяется;
- same - в данном случае применяется такой padding, чтобы размерность изображения на последующем сверточном слое не изменилась.

Stride - это шаг, через который свертка может "перешагивать".

Точно так же, как и в полносвязных нейронных сетях, в сверточных нейронных сетях применяются функции активации:

$$K_{ij} = \sum_{n_c} \sum_{k=0}^{size} \sum_{l=0}^{size} (W_{kl} \cdot I_{i-k, l-j}) + b \quad (8)$$

$$A_{ij} = \sigma(K_{ij}), \quad (9)$$

где A_{ij} - значение ij -ого пикселя в полученной после сверток матрице.

Следующие формулы используются для вычисления размерности слоя с учетом всех гиперпараметров:

$$n_{new} = n - f + 1, \quad (10)$$

где f - размерность фильтра (по договоренности ее всегда выбирают нечетной!).

С учетом stride и padding:

$$n_{new} = \left\lceil \frac{n + 2p - f}{s} + 1 \right\rceil. \quad (11)$$

Summary

Мы начали разговор о сверточных нейронных сетях. Мы поговорили о следующих понятиях:

- понятие свертки изображения;
- понятие обучаемого фильтра;
- познакомились с операцией свертки и pooling-a, являющихся основой CNN.
- познакомились с гиперпараметрами CNN;
- поговорили о задаче многоклассовой классификации и функции активации softmax.