

СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

Романов А.А.

*Романов Александр Андреевич – аспирант,
кафедра общей физики, физико-технический факультет,
Тверской государственный университет, г. Тверь*

Аннотация: специализированная искусственная нейронная сеть, основным направлением которой является эффективное распознавание изображений, входит в состав технологий глубокого обучения.

Ключевые слова: распознавание изображений, технология глубокого обучения, операция свертки, полносвязная нейронная сеть.

УДК 004.032.26

С большим развитием информационных технологий и большого распространения камер видеонаблюдения одной из основных задач становится классификация полученного изображения. Человек без промедления безошибочно распознает объекты, которые он видит. При этом мы не только называем предметы, за которыми мы наблюдаем, но также можем определить их размерные характеристики, определить их контуры и отделить предмет от окружающего мира. Некоторым образом наши глаза видят объемные пиксели, а мозг преобразует эти пиксели в различные категории – кривые, окружности, изгибы и другие формы – которые дают нам понимание, на что мы смотрим.

Программные методы, подражающие человеческому зрению, основаны на использовании нейронных сетях. Сегодня нейронные сети лежат в основе многих сервисов: Facebook использует нейронные сети для алгоритмов автоматического проставления тегов, Google — для поиска среди фотографий пользователя, Amazon — для генерации рекомендаций товаров, Pinterest — для персонализации домашней страницы пользователя, а Instagram — для поисковой инфраструктуры [1].

Изображение для компьютера это массив пикселей (смотрите рисунок 1). В зависимости от качества изображения, его разрешения, например, может быть $32 \times 32 \times 3$ (где 3 — это количество цветовых каналов RGB). Если рассматривать изображения в формате JPG, и взять размер 480×480 . Из этого следует, что массив будет $480 \times 480 \times 3$. Каждому элементу массива соответствует значение от 0 до 255, которое дает понять насыщенность этого пикселя. Для человека этот набор цифр является бессмысленным, когда как эта матрицы являются единственными вводными данными, с помощью которых компьютер может сделать выводы. Основная идея в том, что при получении компьютером этих данных, на выходе получается вероятность класса изображений, к которому можно его отнести.

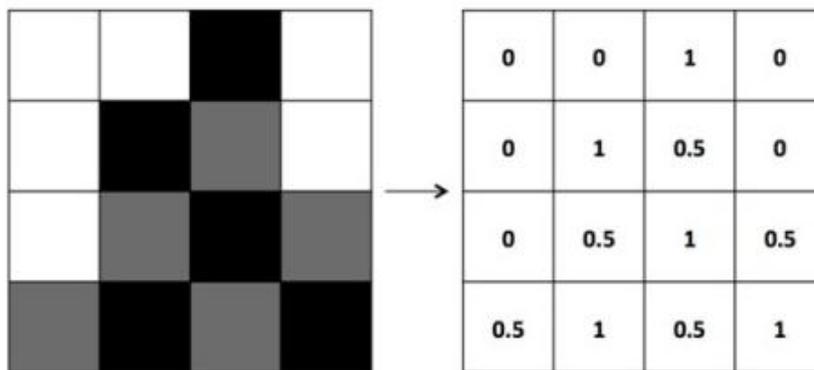


Рис. 1. Пример изображения, как его «видит» машина

Для решения данных задач, нахождения сложных признаков из большого массива пикселей используются тип нейронных сетей, называемый сверточным.

Сверточная нейронная сеть (СНС) очень похожа на зрительную кору головного мозга. На зрительной коре имеются небольшие участки клеток нейронов, которые связаны с определенными местами зрительного поля. За это открытие Дэвид Хьюбел и Торстен Визель удостоились Нобелевской премии по медицине 1981 года. Хьюбел и Визель в 1962 году провели эксперимент, в котором показали, что отдельные клетки нейронов откликались исключительно при наблюдении границ конкретной ориентации. В частности, отдельные клетки нейронов откликались, когда принимали вертикальные грани, а некоторые — горизонтальные или диагональные. Хьюбел и Визель определили, что все эти нейроны сконцентрированы в образе стержневой построения и все вместе основывают человеческое визуальное восприятие. Этот тезис специальных элементов внутри системы, которые решают конкретную задачу (как нейроны зрительной коры головного мозга, которые отыскивают специальные

характеристики) и используются в компьютерных алгоритмах машинного зрения, и на этом основана СНС [2].

Если рассматривать СНС более подробно, то она состоит из серии слоев. Берётся изображение, пропускается через чередование свёрточных, нелинейных слоев, и с помощью полносвязного слоя порождается вывод. В качестве вывода может выступать класс или вероятность класса, которое лучше всего описывает изображение.

В качестве первого слоя всегда выступает сверточный слой. Как уже отмечалось ранее, вводимое изображение представляет матрицу некоторого размера, например $32 \times 32 \times 3$ с пиксельными значениями. Фильтр представляет собой матрицу (её ещё называют матрицей весов или матрицей параметров) размером, например $5 \times 5 \times 3$. Если взять позицию, в которой находится фильтр, верхний левый угол (смотрите рисунок 2). Фильтр производит операцию свёртки т.е умножает элементы фильтра на значения пикселей исходного изображения (поэлементное умножение). После умножения все числа суммируются (для фильтра $5 \times 5 \times 3$ всего 75 умножений). И в итоге получается одно число (См. формулу 1).

$$\sum_i w_i x_i \quad (1)$$

И этот фильтр движется по всей области вводного изображения, а области, на которые взаимодействует фильтр, называются рецептивным полем (полем восприятия). После прохода фильтра по всей области (движение с шагом один) в итоге получается новая матрица размера $28 \times 28 \times 1$ (можно получить и другую размерность, это зависит от применимости граничных условий при движении фильтра по изображению). Если использовать несколько фильтров размерностью $5 \times 5 \times 3$ вместо одного. Тогда выходным значением будет $28 \times 28 \times N$, где N количество фильтров.

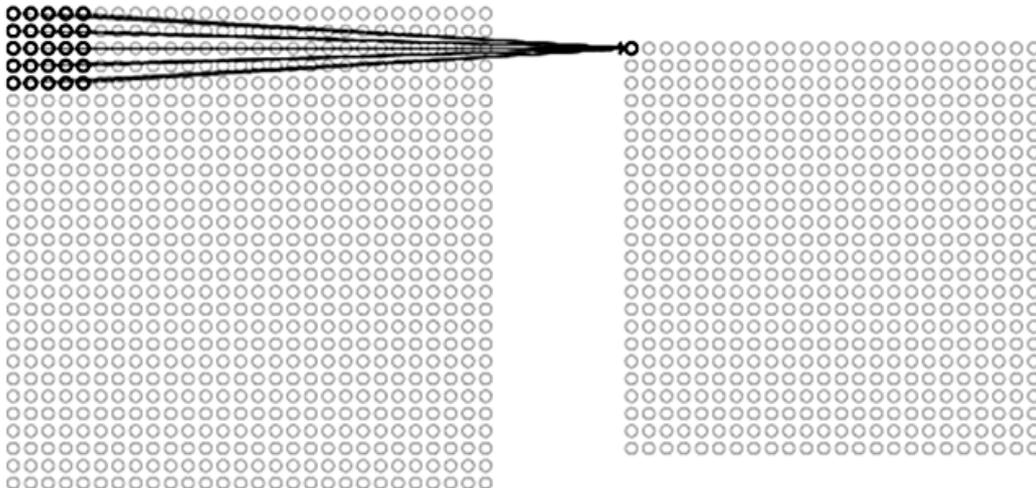


Рис. 2. Пример операции свёртки изображения с помощью фильтра размером 5×5

В архитектуре СНС обычно применяется слой пулинга (подвыборки) между последовательности свёрточных слоев. Основная задача состоит в последовательном уменьшении пространственных габаритов (разрешение) изображения с намерением уменьшения количества входных параметров для следующего слоя и, соответственно, вычислительных операций в сети, а также контроля обучаемости (смотрите рисунок 3). Слои пулинга работают независимо от глубины данных на входе и масштабируют весь объем пространственно. Среди функций можно выделить функцию максимума и функцию среднего, но также есть и другие.

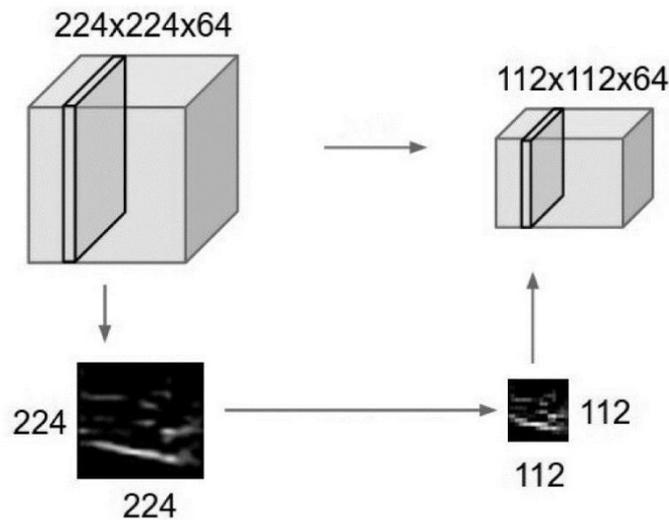


Рис. 3. Пример операции пулинга изображения с уменьшением его в четыре раза

Обычно применяется слой с фильтром размера 2×2 и шагом 2. Такие слои снижают размерность каждого среза глубины входных данных в два раза по ширине и по высоте, отбрасывая при этом $\frac{3}{4}$ элементов активация. При каждой операции максимума в таком случае (смотрите рисунок 4) выбирается максимальная величина из 4 чисел. Глубине при этом остается неизменным.

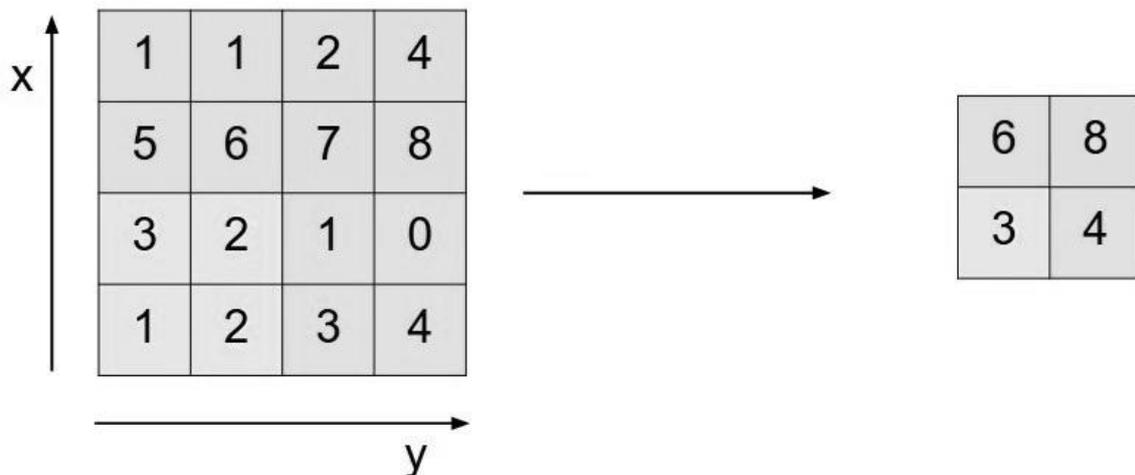


Рис. 4. Пример операции пулинга изображения с помощью фильтра размером 2×2 с применением функции максимума

Помимо функции максимума подвыборки, слой пулинга может выполнять и другие функции, например, усредняющую функцию подвыборки или, например, L2-нормированную функцию. Исторически усредняющая функция подвыборки применялась достаточно часто, но в последнее время больше применяется функция с максимизационной подвыборкой, которая на практике работает лучше.

Еще одним слоем, иногда применяемым в последовательности преобразований, является линейная ректификация. Его смысл состоит в использовании поэлементной функции, такой как $f(x) = \max(0, x)$, определяющий нулевой порог. Другими словами, эта функция выполняет следующие преобразования: если $x > 0$, то значение остается прежним, а если $x < 0$, то значение заменяется на 0.

Последним слоем СНС является полносвязная нейронная сеть. Входными данными для полносвязной нейронной сети являются предыдущие слои (которые выводят высокоуровневые карты свойств) и определение свойства, которые больше связаны с определенным классом. Скрытый слой может состоять из нескольких скрытых слоев (обычно два), что позволяет сократить общее количество нейронов в полносвязном слое. Слой полносвязной нейронной сети наблюдают за тем, как высокоуровневые карты свойств сильно связаны с каким-либо классом и содержит конкретные веса, поэтому, когда вычисляются взаимодействие весов с предыдущими слоями, то получаются верные вероятности для разных классов.

На выходе получаем N-пространственный вектор, где N соответствует числу классов. Каждое значение в этом N-пространственном векторе представляет собой вероятность конкретного класса [3].

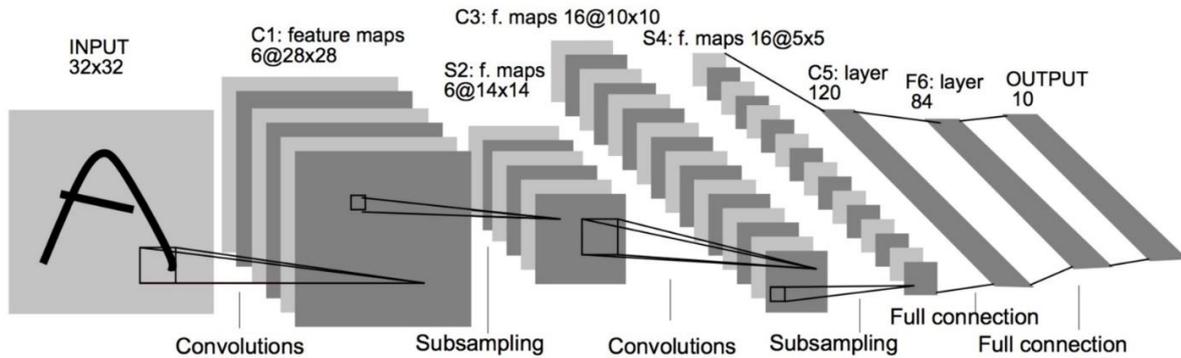


Рис. 5. Пример СНС, состоящей из слоев: 1 - слой свертки, 2 - слой пулинга, 3 - слой свертки, 4 - слой пулинга, 5 и 6 – полносвязные слои

В итоге СНС в каждом слое преобразования трансформирует данное изображение (смотрите рисунок 5). Преобразование начинается с первоначальных значений исходного изображения и заканчивается определением класса изображения.

Список литературы

1. *Deshpande A.* A Beginner's Guide To Understanding Convolutional Neural Networks / A. Deshpande // [Электронный ресурс]. Режим доступа: <https://adeshpande3.github.io/> (дата обращения: 10.01.2018).
2. *Hubel D.H.* Brain mechanisms of vision. / D. Hubel, T. Wiesel // Scientific American, 1979. P. 150-162.
3. *Le Cun Y.* Backpropagation Applied to Handwritten Zip Code Recognition / Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel // Neural Computation. 1 (4). P. 541-551.