

УДК 621.391

ВЫБОР АРХИТЕКТУРЫ НЕЙРОННОЙ СЕТИ ДЛЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ФИЗИЧЕСКОЙ АКТИВНОСТИ

П.С. МАРИНИЧ, А.А. БОРИСКЕВИЧ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 8 марта 2021

Аннотация. Выбрана оптимальная архитектура нейронной сети для распознавания физической активности. Описана архитектура работы сверточной нейронной сети. Обоснован выбор среды разработки.

Ключевые слова: нейронные сети, изображение.

Введение

Одним из основных навыков, которыми человек овладевает в первые годы жизни, является распознавание образов, которые предстают перед ним. Эволюция заложила в человека эту функцию для моментального распознавания уже знакомых образов, будь то убежище, пища или угроза. Чем больше мы сталкиваемся с тем или иным явлением или объектом, тем быстрее мы сможем охарактеризовать их, запомнить определенные паттерны и дать соответствующие ярлыки происходящему вокруг. Однако мир становится все сложнее, объемы информации увеличиваются, а задачи, которые ставит перед собой человечество – (сложнее), и, следовательно, требуются все более точные системы, которые не только помогают естественным чувствам, но и зачастую их заменяют.

Выбор и описание архитектуры

Наилучшей архитектурой искусственных нейронных сетей, направленных на распознавание образов на изображениях, является архитектура сверточных нейронных сетей. Она использует некоторые особенности работы человеческого мозга, в частности, зрительной коры. В 1962 году в ходе экспериментов группы американских нейрофизиологов выяснилось, что отдельные мозговые нервные клетки реагировали (или активировались) только при визуальном восприятии границ определенной ориентации. Например, некоторые нейроны активировались, когда воспринимали вертикальные границы, а некоторые – горизонтальные или диагональные. Нейроны, ответственные за это, сосредоточены в виде стержневой архитектуры и вместе формируют визуальное восприятие. Самый простой пример из повседневной жизни – выявление отличительных признаков и представление на его основе выводов об объекте: мы видим две круглые детали внизу объекта – это велосипед. Эта идея легла в основу работы французского ученого Яна Лекуна, который в 1988 году и представил архитектуру сверточных нейронных сетей.

На вход сверточной нейронной сети подается матрица чисел (если используется изображение в оттенках серого), которые характеризуют каждый пиксель изображения, либо набор матриц (изображение, разделенное на три слоя – R-, G- и B-каналы).

За распознавание определенных характеристик изображения отвечает фильтр – также матрица, но уже с искомыми значениями.

Выбор архитектуры сверточной нейронной сети для распознавания активности должен определяться сложностью обработки слоев, их количества, а также скоростью получения результата, поскольку на вход системы поступает видеопоток, состоящий из отдельных изображений, с частотой порядка 20–30 кадров в секунду. Исходя из этого следует выбрать оптимальный в количестве и скорости выполняемых шагов алгоритм с минимальными потерями в качестве результата.

Сверточные нейронные сети состоят из нескольких слоев, количество которых зависит от ее архитектуры и решаемой задачи [1]. Первым слоем всегда выступает сверточный слой, от которой эта сеть и получила название. Он представляет собой так называемую карту признаков – результирующую матрицу, являющуюся результатом операции свертки матрицы фильтра и части матрицы изображения, сопоставимой размерами с матрицей фильтра (рис. 1). Каждый элемент результата вычисляется как скалярное произведение матрицы фильтра и подматрицы такого же размера (части изображения) с помощью выражения:

$$C_{i,j} = \sum_{u=0}^{m_x-1} \sum_{v=0}^{m_y-1} A_{i+u,j+v} B_{u,v},$$

где A и B – матрицы размера $n_x \times n_y$ и $m_x \times m_y$ соответственно; C – матрица размера $(n_x - m_x + 1) \times (n_y - m_y + 1)$.

Большее числовое значение результирующей матрицы говорит о большем подобии части исходного изображения к фильтру (рис. 1).

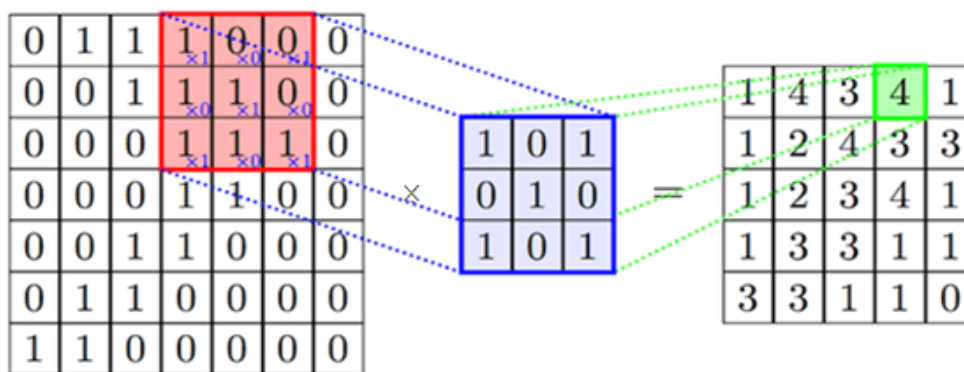


Рис. 1. Пример свертки двух матриц размером 7×7 и 3×3

Важно понимать, что количество каналов фильтра должно соответствовать количеству каналов исходного изображения. То есть при работе с тремя матрицами изображения необходимо иметь три матрицы фильтра для каждого из каналов.

Количество карт определяется требованиями к задаче, если взять большое количество карт, то повысится качество распознавания, но увеличится вычислительная сложность [2]. В большинстве случаев предлагается брать соотношение один к двум: каждая карта предыдущего слоя (например, у первого сверточного слоя, предыдущим является входной слой) связана с двумя картами сверточного слоя (рис. 2). Количество карт – 6, их размер рассчитывается в соответствии с вышеприведенной формулой.

Чем больше сверточных слоев проходит изображение и чем дальше оно движется по сети, тем более сложные характеристики выводятся в картах активации.

Подвыборочный слой также, как и сверточный, имеет карты, но их количество совпадает с предыдущим (сверточным) слоем – таких карт также 6. Цель слоя – уменьшение размерности карт предыдущего слоя. Если на предыдущей операции свертки уже были выявлены некоторые признаки, то для дальнейшей обработки настолько подробное изображение уже не нужно, и оно уплотняется до менее подробного. К тому же фильтрация уже ненужных деталей помогает не переобучаться.

В процессе сканирования ядром подвыборочного слоя карты предыдущего слоя, сканирующее ядро не пересекается в отличие от сверточного слоя. Зачастую каждая карта имеет ядро размером 2×2 , что позволяет уменьшить предыдущие карты сверточного слоя в 2 раза. Вся карта признаков разделяется на ячейки 2×2 элемента, из которых выбираются максимальные по значению. Также чаще всего в подвыборочном слое применяется функция активации ReLU [3]. Благодаря ней картина, формируемая с помощью операции свертки, получает некоторое искажение, позволяющее нейронной сети более ясно оценивать ситуацию.

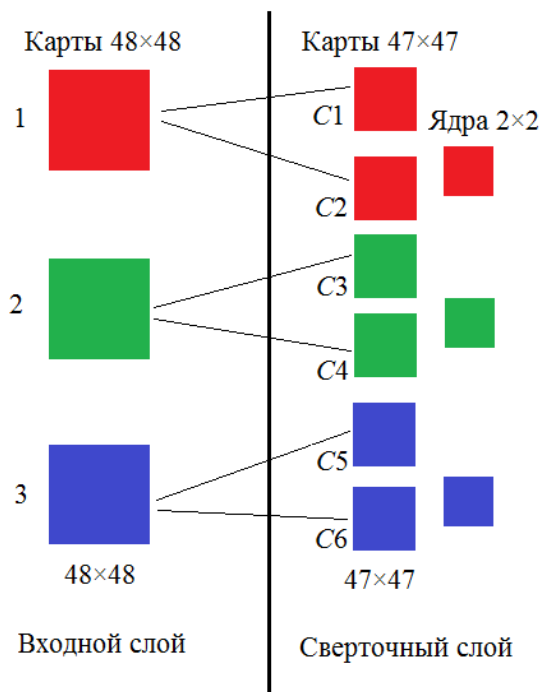


Рис. 2. Связь количества карт между сверточным и входным слоями

Последний из типов слоев – полносвязный слой. Чаще всего его представляет слой многослойного перцептрона. Цель слоя – классификация, которая моделирует сложную нелинейную функцию, оптимизируя которую улучшается качество распознавания. После нескольких проходов свертки изображения и уплотнения с помощью подвыборки система перестраивается от конкретной сетки пикселей с высоким разрешением к более абстрактным картам признаков, как правило на каждом следующем слое увеличивается число каналов и уменьшается размерность изображения в каждом канале. В конце концов остается большой набор каналов, хранящих небольшое число данных (даже один параметр), которые интерпретируются как самые абстрактные понятия, выявленные из исходного изображения.

Эти данные объединяются и передаются на обычную полносвязную нейронную сеть, которая тоже может состоять из нескольких слоев. При этом полносвязные слои уже утрачивают пространственную структуру пикселей и обладают сравнительно небольшой размерностью (по отношению к количеству пикселей исходного изображения).

На рис. 3 представлена обобщенная схема работы нейронной сети с многослойным перцептроном.

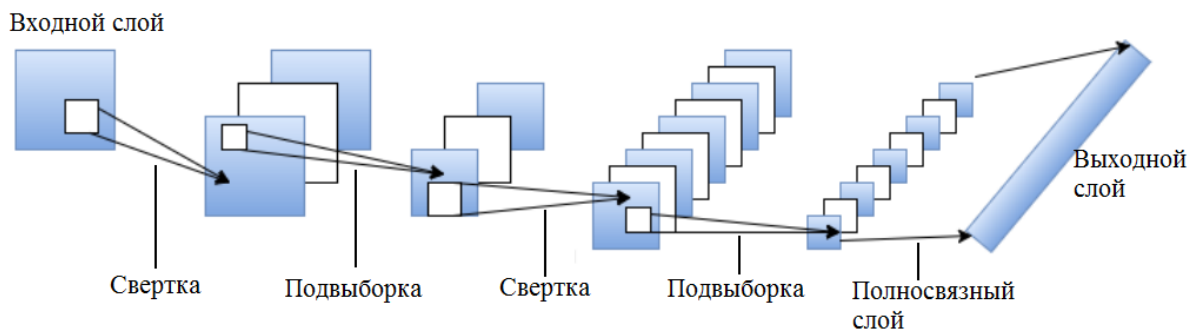


Рис. 3. Обобщенная схема работы сверточной нейронной сети с использованием многослойного перцептрона

Выходной слой отвечает за формирование вероятностей принадлежности входного образа тому или иному классу (некоторому числу). Чтобы добиться этого, выходной слой должен

содержать количество нейронов, соответствующих количеству классов. Взвешенные и просуммированные сигналы далее модифицируются с помощью функции активации. Обучение проходит классическим способом обратного распространения ошибки.

Заключение

Начиная с 2010 года, на основе базы данных проекта ImageNet ежегодно проводится тестирование методов распознавания образов и машинного зрения. Данное испытание считается одним из самых авторитетных в сфере машинного обучения, на результаты которого равняются многие разработчики по всему миру. В большинстве случаев его лидером становится нейронная сеть, разработанная на принципах свертки, показатель точности которых на сегодняшний день составляет не менее 95 %.

Оптимальным выбором для решения задачи распознавания активности выбрана нейронная сеть PNASNet-5, считающаяся на данный момент самой быстрой в принятии решений сверточной нейронной сетью [4]. Нейронная сеть PNASNet-5 реализована на языке программирования Python с использованием библиотеки TensorFlow, вошедшей в дистрибутив Anaconda, зарекомендовавшей себя как универсальная среда, помогающая достигать качества человеческого восприятия.

SELECTING NEURAL NETWORK ARCHITECTURE FOR TASK OF PHYSICAL ACTIVITY RECOGNIZING

P.S. MARYNICH, A.A. BORISKEVICH

Abstract. The optimal neural network architecture for recognizing physical activity is selected. The architecture of the convolutional neural network is described. The choice of the development environment is justified.

Keywords: neural networks, image.

Список литературы

1. Yann LeCun. Gradient-Based Learning Applied to Document Recognition, 1998.
2. Alejandro Escontrela. Convolutional Neural Networks from the ground up, 2018.
3. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks, University of Toronto, 2010.
4. Chenxi Liu. Progressive Neural Architecture Search, 2018.