

Применение сверточных нейронных сетей для распознавания динамического изображения

Карачанская Елена Викторовна, канд. физ.-мат. наук, доцент
Дальневосточный государственный университет сообщений (г. Хабаровск),
Тихоокеанский государственный университет (г. Хабаровск)
Бочкарева Юлия Сергеевна
Тихоокеанский государственный университет (г. Хабаровск)

Представлена модель сверточной нейронной сети для распознавания динамического изображения. Описаны основные понятия и преимущества использования данного вида нейронных сетей. На основе результатов проведенного эксперимента по распознаванию динамических изображений сделаны выводы о преимуществах и недостатках использования сверточных нейронных сетей.

Ключевые слова: искусственный интеллект, сверточная нейронная сеть, глубинное обучение, каскад Хаара, метод Виолы–Джонса.

Введение

В последнее десятилетие машинное обучение стало набирать все большую и большую популярность и сегодня занимает одну из лидирующих позиций в сфере информационных технологий [1]. Объем цифровых данных, используемый человечеством, увеличивается в разы, в связи с чем возникает необходимость в автоматическом анализе смарт-данных, чтобы и в дальнейшем продолжалось развитие технологического прогресса. Наряду с активным развитием технологий, разработкой новых способов защиты информации от несанкционированного доступа, растут и технологии взлома. Особенно ценны методы интеллектуальной защиты. К их числу относятся искусственные нейронные сети.

Искусственная нейронная сеть состоит из большого числа тесно связанных между собой элементов обработки информации, так называемых нейронов, работающих одновременно [2]. Обучение нейронной сети схоже с обучением человека: основной идеей является запоминание образов и примеров чего-либо.

Сверточная нейронная сеть обладает тем свойством, что она акцентирует внимание на наиболее ярко выраженные вертикальные или горизонтальные узкие прямоугольные области - «стержни». Именно эти области несут наибольшую информацию об изображении.

Целью статьи является применение сверточной нейронной сети для распознавания динамического изображения.

На сегодняшний день искусственные нейронные сети нашли свое применение практически в любой сфере. Появилось большое количество нейронных сетей, которые специализируются на различных задачах распознавания изображений: от нейронных сетей, занимающихся идентификацией объекта на изображении и его стилизации, до нейронных сетей, способных определять смысл изображенного на картине или рисунке. Данная область машинного обучения развивается стремительно, но стоит отметить тот факт, что подобные алгоритмы предназначены, в основном, для статических изображений. Распознавание нейронной сетью динамического изображения – энергоемкий и трудный процесс. Поэтому, данная тема актуальна и требует проведения необходимых исследований.

Архитектура сверточной нейронной сети

На данный момент существует немалое количество искусственных нейронных сетей – нейронная сеть прямого распространения, нейронная сеть Хопфилда, цепи Маркова, машина Больцмана, сети Кохонена, автокодировщик, сеть типа «deep belief», развёртывающая нейронная сеть [3]. Однако для распознавания и классификации изображений целесообразно использовать особый вид нейронных сетей – сверточный.

Сверточная нейронная сеть была предложена Я. Лекуном и Й. Бенгуа во время исследований зрительной коры головного мозга кошки [4]. После ряда экспериментов, ученые пришли к выводу, что, благодаря своему устройству, сверточная нейронная сеть используется для решения сложных графических задач распознавания и классификации изображений, а в дальнейшем, после систематического анализа и исследований – для защиты и анализа ранее используемых методов.

Сверточная нейронная сеть обычно представляет собой чередование нескольких видов слоев – сверточных, субдискретизирующих и полносвязных слоев. Пример сверточной нейронной сети представлен на рисунке 1 [5].

Особым слоем представленной нейронной сети является сверточный слой (рис. 2). В сверточном слое нейроны, которые используют одни и те же веса, объединяются в карты признаков, каждый нейрон из карты признаков связан с частью нейронов предыдущего слоя. При вычислении сети получается, что каждый нейрон выполняет свертку некоторой области предыдущего слоя.

Субдискретизирующий слой также является особым слоем в сверточной нейронной сети и располагается между последовательностью сверточных слоев. Основной его задачей является последовательное уменьшение разрешения изображения с целью сокращения количества входных параметров для последующих слоев и дальнейших вычислительных операций в сети. Среди методов данного слоя выделяется метод выбора максимального элемента – вся карта признаков разделяется на ячейки, из которых выбираются максимальные по значению (рис. 3).

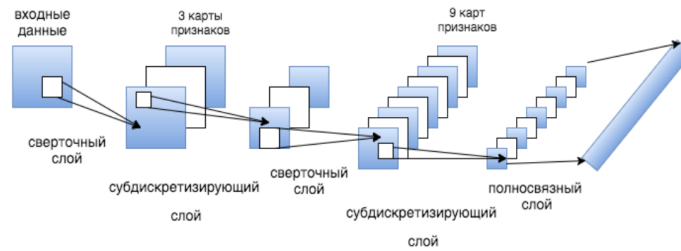


Рисунок 1 – архитектура сверточной нейронной сети

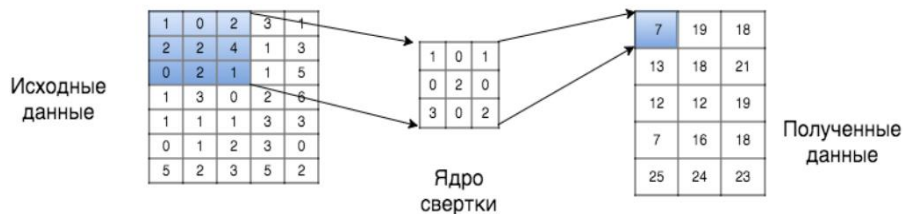


Рисунок 2 – Сверточный слой

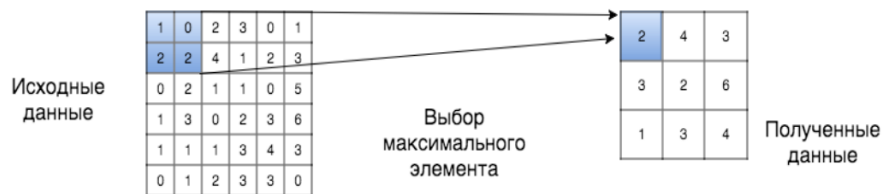


Рисунок 3 – Субдискретизирующий слой

Основными достоинствами данной сети являются:

- сверточный слой данной сети, в котором происходят преобразования изображений, использует ядра. Его наличие значительно уменьшает время и объем вычислительных ресурсов на обучение по сравнению с полносвязной нейронной сетью;

- использование ядра свертки способствует обобщению полученной информации. Восприятие входного изображения по областям позволяет учесть все его свойства, что увеличивает качество распознавания изображений в несколько раз;

- устойчивость к различного рода поворотам и сдвигам;

- частичная неизменность к масштабу за счет сжатия изображения;

- обучение нейронной сети может происходить как с помощью различных методов (например, метод обучения «без учителя»), так и с помощью других технологий глубинного обучения (свёрточный авто-ассоциатор, свёрточная версия разреженного кодирования).

К недостаткам сверточной нейронной сети относятся:

- продолжительное время на обучение, от нескольких часов до нескольких дней, иногда недель. Также, при отсутствии необходимого оборудования и технического оснащения, время обучения значительно увеличивается;

- большое количество обучающего материала;

- эффект дообучения. В случае недостаточного количества обучающей выборки, распознавание будет происходить не всегда. Для повышения качества распознавания потребуются дополнительное количество изображений и, следовательно, дообучение нейронной сети на добавленном материале. Во-вторых, во время тестирования системы происходило распознавание нескольких объектов в кадре. Этот момент расценивается как дополнительная защита персональных данных от несанкционированного доступа – защита содержимого рабочего стола ПК при работе в личном профиле, а также, возможен отказ системы в доступе к персональному компьютеру из-за отсутствия идентификации пользователя.

- узкая сфера применения данного вида нейронной сети (распознавание изображений);

- большое количество параметров (количество слоёв, размер ядра свертки для каждого из слоёв, количество ядер для каждого из слоёв, шаг сдвига и т.д.). Каждый из параметров существенно влияет на результат работы нейронной сети, поэтому для каждой новой задачи они подбираются эмпирически.

Математическая база сверточной нейронной сети

В основе теории и построения нейронной сети лежит результат, основанный на работах А. Н. Колмогорова [6] и В. И. Арнольда [7, 8].

Суперпозиционная теорема Колмогорова [6]. При любом целом $n \geq 2$ существуют такие определенные на единичном отрезке $E^1 = [0,1]$ непрерывные действительные функции $\varphi^{pq}(x)$, что каждая определенная на n -мерном единичном кубе E^n непрерывная действительная функция $f(x_1, x_2, \dots, x_n)$ представима в виде

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} \chi_q \left[\sum_{p=1}^n \varphi^{pq}(x_p) \right],$$

где функции $\chi_q(y)$ действительны и непрерывны.

Таким образом, любая непрерывная функция n переменных всегда может быть представлена в виде суммы непрерывных функций *одного* переменного. При доказательстве теоремы для определения базисных функций φ^{pq} была построена система кубиков, покрывающих n -мерный единичный куб E^n так, что любая точка E^n может быть покрыта не менее, чем $n+1$ раз. Из теоремы следует, что n -мерный куб может быть базисно вложен в пространство \mathbf{R}^{2n+1} . Данная идея о «расслоении» точки лежит в основе нейронной сети.

В конце 80-х годов Р. Хехт-Нильсен, профессор Калифорнийского Университета, переработал теорему Арнольда–Колмогорова непосредственно для нейронных сетей. Была сформулирована теорема о том, что любая многомерная функция нескольких переменных представима с помощью нейронной сети фиксированной размерности [9]. Было доказано следующее: «...для любого множества пар произвольной размерности (X^k, Y^k) , где Y^k - скаляр, существует двуслойная однородная (с одинаковыми функциями активации) нейронная сеть первого порядка с последовательными связями и с конечным числом нейронов, которая выполняет отображение $X \rightarrow Y$, выдавая на каждый входной сигнал правильный выходной сигнал Y^k » [10]. В результате данного доказательства было установлено, что универсальной нейронной сетью является двуслойный перцептрон.

Следствие из теоремы Арнольда–Колмогорова–Хехт-Нильсена определяет необходимое количество нейронов в скрытых слоях [10]:

$$\frac{N_y Q}{1 + \log_2(Q)} \leq N_w \leq N_y \left(\frac{Q}{N_x} + 1 \right) (N_x + N_y + 1) + N_y,$$

где N_y – размерность выходного сигнала, Q – число элементов множества обучающих примеров, N_w – необходимое число синаптических связей, N_x – размерность входного сигнала.

Произведя с помощью этой формулы оценку необходимого количества синаптических связей N_w , можно произвести расчет необходимого числа нейронов в скрытых слоях. К примеру, число нейронов скрытого слоя в двуслойном перцептроне представляется так:
$$N = \frac{N_w}{N_x + N_y}.$$

Теорема, доказанная Р. Хехт-Нильсеном, устанавливает разрешимость задачи, однако не показывает способ построения такой сети.

Обучение сверточной нейронной сети. Метод Виолы–Джонса

Метод Виолы–Джонса на сегодняшний день является одним из актуальных методов поиска объектов на изображении в реальном времени. Данный метод был разработан и предложен в 2001 году П. Виолой и М. Джонсом и стал первым методом, демонстрирующим отличные результаты в обработке изображений в режиме реального времени [11]. Достоинствами данного метода является высокая скорость распознавания за счет использования каскадного классификатора и высокая точность обнаружения с низким процентом ложных срабатываний по сравнению с более медленными алгоритмами.

Алгоритм использует набор признаков Хаара совместно с алгоритмом AdaBoost.

Признаки Хаара. Каскад Хаара

Признак Хаара является набором прямоугольных областей изображения, примыкающих друг к другу и разделенных на две группы [12]. На рисунке 2 представлены возможные прямоугольники, разделенные на области черного (группа 1) и белого (группа 2) цветов.

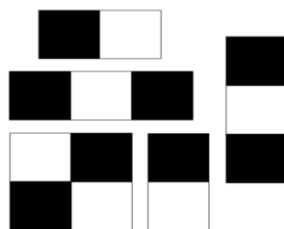


Рисунок 4 – возможные виды прямоугольных областей и их расположение на плоскости

Чтобы вычислить значение конкретного признака Хаара для какого-либо изображения, надо сложить яркости пикселей изображения в первой и второй группах прямоугольных областей по отдельности, а затем вычесть из первой полученной суммы вторую.

В этой форме значение каждого пикселя является суммой яркостей этого пикселя и всех пикселей, что находятся выше и левее него, при условии, что пиксель с начальными координатами находится в левом верхнем углу изображения: $S_{yx} = \sum_{i=1}^y \sum_{j=1}^x x_{ij}$, где S_{yx} – значение пикселя (интегральной формы) с координатами $[y; x]$, x_{ij} – значение пикселя (исходной формы) с координатами $[i; j]$.

Каскад Хаара является одним из способов обнаружения объектов на изображении [13]. Каскад Хаара, обученный в течение продолжительного времени, принимает на вход изображение и определяет, есть ли на этом изображении искомый объект. Иными словами, каскад выполняет задачу классификации и разделяет данные на входе на два класса (есть ли искомый объект или нет искомого объекта).

Нормализация изображения

Прежде чем классифицировать изображение, его следует нормализовать, то есть привести к стандартному виду. При использовании каскадов Хаара это означает, что надо перевести изображение из текущей цветовой схемы в черно-белую, а также нормализовать (сделать равным единице) его среднеквадратичное отклонение.

Шаг. 1. Следует начать с перевода изображения в черно-белую цветовую схему. Так как обычно изображения изначально загружаются в цветовой схеме RGB, будет приведена формула только для нее:

$$i[x; y] = \frac{I_r[x; y] + I_g[x; y] + I_b[x; y]}{3},$$

где $i[x; y]$ – пиксель для черно-белого изображения; $I_r[x; y]$, $I_g[x; y]$, $I_b[x; y]$ – компоненты изображения (красная, синяя и зеленая соответственно) [13].

Шаг 2. Нормирование среднего квадратического отклонения (СКО). Данная процедура необходима для минимизации влияния разницы в освещенности на результат классификации.

Шаг 2.1. Для начала следует вычислить среднее арифметическое изображения: $a = \frac{1}{s \cdot h} \sum_{i=1}^s \sum_{j=1}^h x_{ij}$,

а затем СКО по такой формуле: $\sigma = \left(\frac{1}{s \cdot h} \sum_{i=1}^s \sum_{j=1}^h (2x_{ij} - a)^2 \right)^{1/2}$.

Шаг 2.2. Нормализация значения, соответствующего изображению, происходит делением на СКО значения каждого пикселя в изображении: $x_{ij} = \frac{x_{ij}}{\sigma}$.

Полученное изображение теперь можно использовать для обучения каскада Хаара. Подобные операции необходимо выполнять с изображением перед его классификацией.

Интегральная форма представления изображения

Для вычисления значений каждого признака Хаара для изображений требуется большое количество операций сложения. Чтобы упростить данную задачу, необходимо использовать интегральную форму изображения. На рисунке 3 представлена схема вычисления интегральной формы представления изображения.

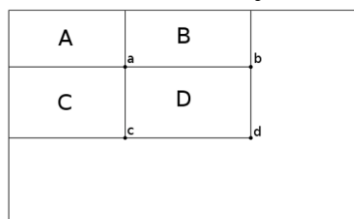


Рисунок 5 – иллюстрация для вычисления значения признака Хаара с помощью интегральной формы изображения

Значение a в точке a является суммой значений пикселей в прямоугольнике A , значение \tilde{b} в точке b является суммой значений пикселей в прямоугольниках A и B , значение \tilde{c} в точке c является суммой значений пикселей в прямоугольниках A и C , а значение d в точке d является суммой значений пикселей в прямоугольниках A, B, C и D . Для вычисления суммы значений пикселей в прямоугольнике B применяется формула: $\sum_{x_{ij} \in B} x_{ij} = d - \tilde{c} - \tilde{b} + a$.

Алгоритм AdaBoost

Алгоритм AdaBoost (Adaptive Boosting) представляет собой алгоритм усиления классификаторов путем их объединения в каскад [14].

Он построен на том, что из огромного числа простых способов классификации (в дальнейшем будем их называть «слабыми классификаторами») можно составить новый, более эффективный, способ. Под слабым классификатором понимают некоторые алгоритмы вычисления минимального расстояния, каждый из которых предназначен для решения конкретных задач и проверке условий.

В качестве слабого классификатора принимается функция, которая на вход принимает изображение, вычисляет значение признака Хаара для данного изображения и проводит сравнение получившегося значения с порогом, возвращая или нуль, или единицу [15]:

$$h_i(x) = \begin{cases} 1, & \text{если } p_i f_i(x) < p_i \theta_i, \\ 0, & \text{иначе,} \end{cases}$$

где θ_i – порог, x – входное изображение, $f_i(x)$ – значение признака Хаара для изображения x , p_i – направление знака неравенства (если он равен единице, то знак неравенства не меняется, если минус единице, то знак неравенства меняется в противоположную сторону), $h_i(x)$ – слабый классификатор.

Представленный алгоритм проводит перебор всех возможных слабых классификаторов и выбирает те, которые допускают меньше ошибок.

На рисунке 4 представлена структура каскада со слабыми классификаторами (цифры в кругах обозначают порядковые номера), а ряды кругов – это стадии каскада.

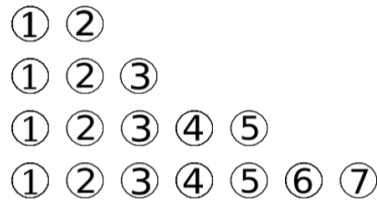


Рисунок 6 – структура каскада

Каскад строится следующим образом – к текущей стадии добавляются слабые классификаторы до тех пор, пока число ложных срабатываний с «позитивом» не будет больше первоначально заданного значения. Как только число ложных срабатываний с «позитивом» станет меньше заданного значения, необходимо переходить к следующей стадии и точно также добавлять к ней слабые классификаторы.

Результат работы сверточной нейронной сети

Для тестирования было загружено 10 классов лиц. Для каждого класса в директории было добавлено по 10 различных изображений. Нулевая директория содержала изображения лиц людей, взятых из сети интернет. Данный класс лиц необходим для оценки качества работы классификатора на большом массиве изображений. Остальные директории содержали изображения лиц родных и знакомых людей.

Для удачного распознавания лиц в видеопотоке реализуемой программы необходимо было установить порог ложных срабатываний. Эмпирическим путем было выявлено значение порога равное 0,005. По достижению данного значения, обучение нейронной сети считается законченным.

Процесс распознавания каждого лица проходил в несколько этапов. Первый этап распознавания происходил еще в процессе обучения нейронной сети – обучение приостанавливалось, и веб-камера считывала лицо человека. В большинстве случаев считывание давало отрицательный результат: для каждого человека потребовалось несколько тестов (от 3 до 5), чтобы результат был удовлетворительным. В общей сложности, для 10 тестируемых людей потребовалось около 40-55 тестов, из них удачно прошедшими оказались 12 (21,8%). В данном случае требовалось дообучение нейронной сети. После процесса дообучения наступал второй этап распознавания динамических изображений. Результаты этого этапа были более успешными: количество тестов было увеличено до 8, что способствовало неплохому выводу в 56,7%. Заключительный этап подразумевал завершение процесса обучения нейронной сети и проведение финальных тестов по распознаванию. Точность работы классификатора на заключительном этапе эксперимента равна 80%. Из 10 тестируемых людей полностью распознаны были 8 человек с 1 попытки, 1 человек – со 2 попытки, 1 человек – с 3 попытки. Стоит отметить, что условия, в которых проводилось тестирование системы были максимально приближены к условиям создания выборки изображений для будущего обучения.

Проведенные эксперименты показали, что нейронная сеть распознавала лица со всеми «особенностями». Фотографии лиц людей в анфас были сделаны как при естественном освещении, в дневное время суток, так и при рассеянном свете настольной лампы в ночное время суток. Также на лицах некоторых людей присутствовали аксессуаров в виде пирсинга и очков.

Проведенные эксперименты помогли вычислить ряд достоинств данного алгоритма:

- 1) высокая скорость обучения;
- 2) универсальность алгоритма, которая позволяет проводить обучение на любом количестве слоев нейронной сети;

3) распараллеливание. Алгоритм реализуется на параллельных вычислительных устройствах, что способствует сокращению времени обучения и повышению точности распознавания;

4) отличное распознавание объектов.

Заключение

Использование сверточных нейронных сетей для машинного обучения показало их эффективность для решения задачи распознавания динамических объектов.

Разработанная система может использоваться при решении различных задач видеоаналитики, и, в первую очередь, имеет непосредственное применение в системах контроля доступа и идентификации личности.

Литература:

1. Murphy K. P. Machine Learning: A Probabilistic Perspective. Cambridge : Massachusetts Institute of Technology, 2012. 1067 p.
2. Хайкин С. Э. Нейронные сети: полный курс. 2-е изд., Москва : Вильямс. 2008. 1103 с.
3. Каллан Р. Основные концепции нейронных сетей. Москва : Вильямс. 2001. 287 с.
4. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge : The MIT Press. 2016. 800 p.
5. Галушкин А. И. Нейронные сети. Основы теории. Москва : Горячая линия—Телеком. 2010. - 496 с.
6. Колмогоров А. Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения. // Докл. АН СССР, 1959. Т. 114. № 5. С. 953—956.
7. Арнольд В. И. О представлении функций нескольких переменных в виде суперпозиции функций меньшего числа переменных // Матем. просв. 1958. Вып. 3. С.41—61.
8. Арнольд В. И. О представлении непрерывных функций трех переменных суперпозициями непрерывных функций двух переменных // Матем. сб. 1959. Т.48(90). №1. С. 3—74.
9. Hecht-Nielsen R. Kolmogorov's Mapping Neural Network Existence Theorem // IEEE First Annual Int. Conf. on Neural Networks, San Diego, 1987, Vol. 3, pp. 11-13.
10. Круглов В.В., Борисов В. В. Искусственные нейронные сети. Теория и практика. - 2-е изд., стереотип. М.: Горячая линия-Телеком, 2002. 382 с.
11. Квасов Д. С., Зиновьев И. И., Кокорин И. Г., Коробко С. С. Обнаружение и идентификация лиц на растровом изображении с применением метода Виолы и Джонса // Труды XVII Всероссийской научно-методической конференции "Телематика". Санкт-Петербург. 2010. С. 253—254.
12. Мазуров В. Д. Математические методы распознавания образов : учебное пособие. 2-е изд., испр. и доп. Екатеринбург : Изд-во Урал. ун-та. 2010. 101 с.
13. Зиновьев И. И., Шамин, П. Ю. Обнаружение лиц людей в системах видеонаблюдения. // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. 2011. Т.2. С. 70-75.
14. Рассе С., Норвиг П. Искусственный интеллект: современный подход. 2-е изд. Москва : Вильямс. 2006. 1408 с.
15. Барский, А. Б. Нейронные сети: распознавание, управление, принятие решений. Москва : Финансы и статистика. 2004. 176 с.