

РАЗВИТИЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ КЛАССИФИКАЦИИ И ЛОКАЛИЗАЦИИ

DEVELOPMENT OF CONVOLUTIONAL NEURAL NETWORKS FOR SOLVING CLASSIFICATION AND LOCALIZATION PROBLEMS

A. Kochnev

Summary. Over the past few years, the use of approaches to the recognition of images, objects, optoelectronic devices designed for visual control or automatic image analysis in combination with computer vision methods based on the use of neural networks have proven their effectiveness in solving various problems. Convolutional networks, being a key element of most data mining systems, are able to influence the processes occurring in various systems due to the relationship between data about a specific event and the ability to predict future events. However, the development in the field of adaptation and use of neural networks for solving localization and classification problems is very slow. And if some of the problems in this area related to object recognition have already been overcome by training neural networks based on elementary algorithms, then such problems as the performance of neural networks and the choice of optimal network training algorithms for these purposes are still not solved.

In the article, the author examines the main theoretical problems of the development of convolutional neural networks for solving classification and localization problems and comes to the conclusion that it is necessary to develop approaches to faster and more accurate training of networks. At the same time, such approaches should be comprehensive and focused not only on solving the problem of increasing the amount of processed information without losing the quality of the network, but also on increasing the number of layers of the neural network without losing the accuracy of the neural network and its performance.

Keywords: convolutional neural networks; classification and localization tasks, neural network training, recognition tasks, computer vision.

Кочнев Александр Александрович

Старший бэкэнд разработчик

Your Next Agency

drdispool@gmail.com

Аннотация. За последние несколько лет использование подходов к распознаванию образов, объектов, опико-электронных устройств, предназначенных для визуального контроля или автоматического анализа изображений в сочетании с методами компьютерного зрения, основанных на использовании нейронных сетей, доказало свою эффективность в решении различных задач. Сверточные сети, являясь ключевым элементом большинства систем интеллектуального анализа данных, способны влиять на процессы, происходящие в различных системах за счет взаимосвязи между данными о конкретном событии и возможностями прогнозирования будущих событий. Однако развитие в области адаптации и использования нейронных сетей для решения задач локализации и классификации происходит очень медленно. И если часть проблем в указанной сфере, связанных с распознаванием объектов, уже удалось преодолеть путем обучения нейронных сетей на основе элементарных алгоритмов, то такие проблемы, как производительность нейронных сетей и выбор оптимальных алгоритмов обучения сетей для указанных целей, по-прежнему не решены.

В статье автор рассматривает основные теоретические проблемы развития сверточных нейронных сетей для решения задач классификации и локализации и приходит к выводу о том, что в настоящее время назрела необходимость разработки подходов к более быстрому и точному обучению сетей. При этом такие подходы должны быть комплексными и ориентированными не только на решение проблемы увеличения количества обрабатываемой информации без потери качества работы сети, но и на увеличение количества слоев нейронной сети без потери точности ее работы и производительности.

Ключевые слова: сверточные нейронные сети, задачи классификации и локализации, обучение нейронных сетей, задачи распознавания, компьютерное зрение.

Направление, связанное с решением задач классификации и локализации с помощью нейронных сетей, в современном мире только начинает активно развиваться. При этом первоначально искусственные нейронные сети появились вовсе не для решения задач в различных областях, а для моделирования работы нейронов человеческого мозга [1, 4, 5]. Только спустя много лет было доказано, что искусственные нейроны точно также, как и нейроны челове-

ского мозга способны обучаться при подборе соответствующих алгоритмов обучения [8]. Вопросы подбора соответствующих алгоритмов обучения искусственных нейронных сетей для решения различных задач являются одними из самых актуальных и одновременно самыми неизученными в современной науке.

Такая неизученность алгоритмов обучения обусловлена сложностью задач, которые ставятся перед

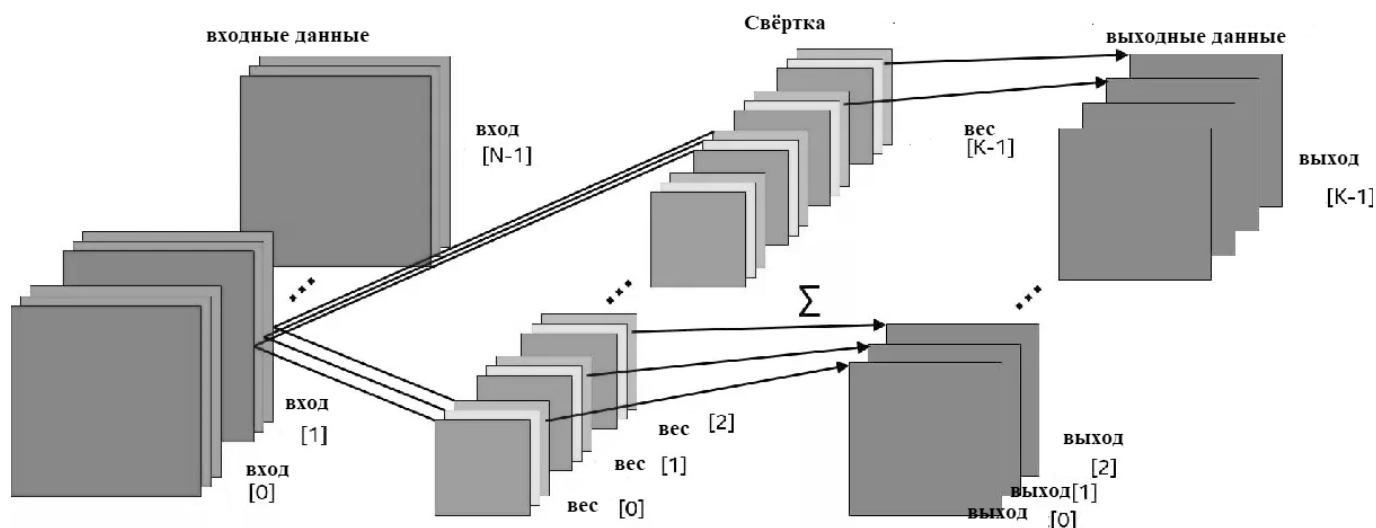


Рис. 1. Принцип работы сверточной нейронной сети

нейронными сетями по мере развития науки и техники. И если еще тридцать лет назад нейронные сети решали простые задачи классификации и локализации, к примеру, такие, как распознавание изображений и объектов, то теперь задачи, которые ставятся перед нейронными сетями, стали гораздо сложнее — полная автоматизация отдельных процессов, в том числе процессов, связанных с управлением без контроля за процессом управления со стороны человека.

Вместе с тем, для решения сложных задач необходимо повышать производительность нейронных сетей, собирать большие наборы данных, изучать более мощные модели и использовать более эффективные методы обучения нейронных сетей. Длительное время наборы данных помеченных изображений для целей обучения нейронных сетей были относительно небольшими — порядка десятков тысяч изображений, что позволяло решать простые задачи классификации и локализации. Так, например, частота ошибок в задаче распознавания цифр базы данных MNIST (объемная база данных образцов рукописного написания цифр) оставляет меньше, чем 0,3%, что очень близко к производительности человека [7, 8].

Однако, объекты в реалистичных условиях, когда нейронная сеть работает, решая конкретную задачу управления, могут проявлять изменчивость, а происходящие вокруг объекта события могут быть не прогнозируемы, поэтому чтобы научиться распознавать такие объекты, необходимы гораздо более объемные наборы данных и более сложные алгоритмы обучения.

В настоящее время опубликовано достаточно количество работ, посвященных проблематике обучения

нейронных сетей и подбору нужного количества обучающих данных (в том числе изображений) [2, 6]. При этом акцентируется внимание на том, что эффективное решение задач классификации возможно исключительно при правильном подходе к глубокому обучению нейронной сети с учетом правильного выбора структуры сети, позволяющей решать сложные задачи.

Эффект глубокого обучения заключается в том, что нейронные сети работают аналогично 3D моделям и на основе обработки огромных объемов данных (как структурированных, так и неструктурированных), способны делать точные прогнозы. При этом чем больше объем информации, который предоставляется нейронной сети, тем точнее прогноз. Вместе с тем, несмотря на наличие возможностей глубокого обучения нейронных сетей с использованием большого массива разрозненных данных, до настоящего момента универсального механизма обучения нейронной сети, позволяющего полностью сделать процессы в области классификации и локализации автономными, не разработано. Основная причина существующих проблем заключается в невозможности одномоментного учета всех возможных факторов риска, с которыми может столкнуться нейронная сеть при решении соответствующих задач.

Из всех известных структур нейронных сетей решение задач классификации и локализации наиболее эффективно осуществляется с помощью сверточной нейронной сети. Основной особенностью сверточной нейронной сети является то, что она содержит трехмерное расположение нейронов вместо стандартного двумерного массива. Первый слой называется сверточным. Каждый нейрон в сверточном слое обрабатывает информацию только из небольшой части поля зрения.

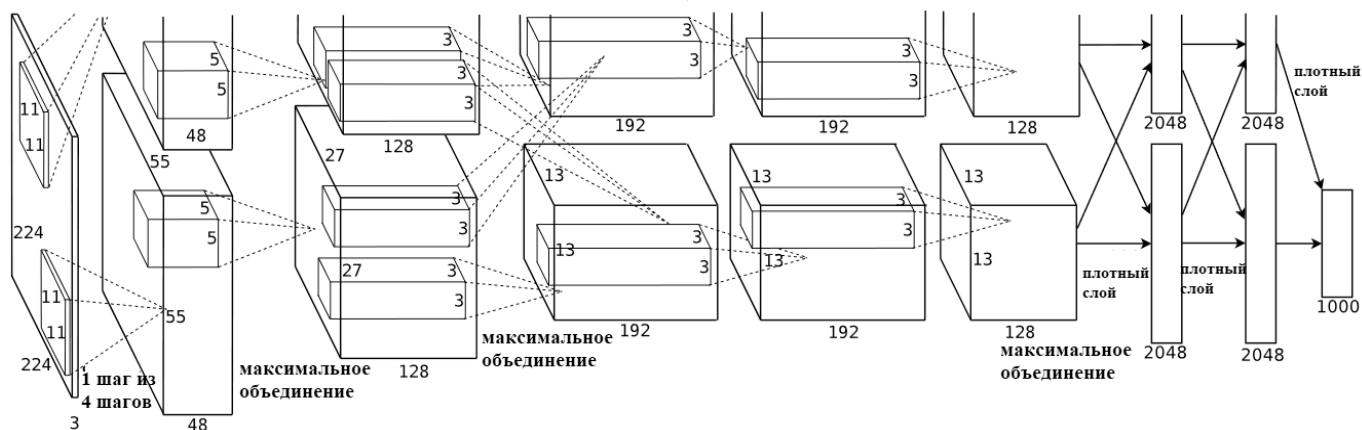


Рис. 2. Работа сверточной нейронной сети, состоящей из восьми слоев

Входные характеристики берутся пакетно, как фильтр [10]. Сеть понимает изображения по частям и может вычислять эти операции несколько раз, чтобы завершить полную обработку изображений. Обработка включает в себя преобразование изображения из RGB (Red, Green, Blue) или HSI (Horizontal Situation Indicator) масштаба в серый масштаб. Дальнейшее изменение значения пикселя поможет обнаружить края и изображения могут быть классифицированы по разным категориям (рис. 1).

В свете приведенной выше информации целью данной работы является критическая оценка различных подходов с использованием сверточной нейронной сети для выполнения задач классификации и локализации.

Распространение является однонаправленным, когда сверточная сеть содержит один или несколько сверточных слоев, за которыми следует объединение, и двунаправленным, когда выходной сигнал сверточного слоя поступает в полностью связанную нейронную сеть для классификации изображений. Фильтры используются для извлечения определенных частей изображения. Именно поэтому сверточные нейронные сети показывают очень эффективные результаты в распознавании изображений и видео, семантическом анализе и обнаружении парафраз в текстовых документах. Их емкость можно регулировать, изменяя их глубину и ширину. Сверточные нейронные сети также после обучения делают правильные предположения о природе изображений (а именно стационарность статистики и локальность зависимостей пикселей), их легче обучать по сравнению со всеми остальными видами нейронных сетей.

Теперь представим, как работает сверточная нейронная сеть, описав ее общую архитектуру.

Как показано на рис. 2, сеть имеет восемь слоев с весами. Первые пять слоев — это сверточные слои, а остальные три слоя полностью связаны. Выходные данные последнего полностью подключенного слоя подаются на 1000-полосный Softmax (Функция Softmax применяется в машинном обучении для задач классификации, когда количество возможных классов больше двух), который производит распределение по 1000 меткам классов. Сверточная сеть при этом максимизирует цель мультиномиальной логистической регрессии, что будет равняться максимизации среднего значения по обучающим случаям логарифмической вероятности правильной метки при распределении прогноза.

Ядра второго, четвертого и пятого сверточных слоев подключены только к тем картам ядра в предыдущем слое, которые находятся на том же графическом процессоре (рис. 2). Ядра третьего сверточного слоя связаны со всеми картами ядра во втором слое. Нейроны в полностью связанных слоях соединены со всеми нейронами в предыдущем слое.

Исходя из данных, представленных на рисунке 2, получается следующая схема работы нейронной сети. Первый сверточный слой фильтрует входное изображение размером 224 × 224 × 3 с 96 ядрами размером 11 × 11 × 3 с шагом 4 пикселя (это расстояние между центрами рецептивного поля соседних нейронов на карте ядра). Далее второй сверточный слой принимает в качестве входных данных выходной сигнал первого сверточного слоя и фильтрует его с помощью уже 256 ядер размером 5 × 5 × 48. Третий, четвертый и пятый сверточные слои соединены друг с другом без каких-либо промежуточных слоев. Третий сверточный уровень имеет 384 ядра размером 3 × 3 × 256, подключенных к выходам второго сверточного уровня. Четвертый сверточный слой имеет 384 ядра размером 3 × 3 × 192, а пятый сверточный слой имеет 256 ядер

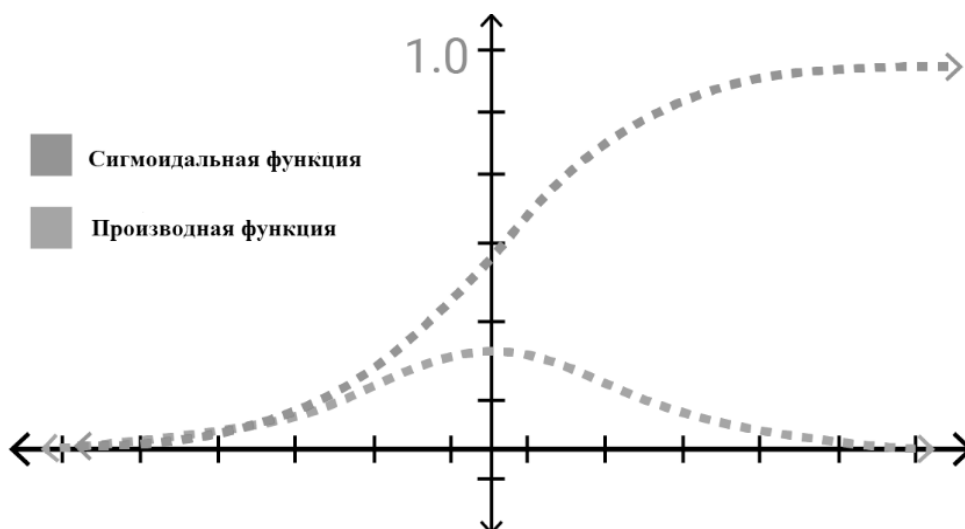


Рис. 3. График сигмоидальной и производной функций



Рис. 4. График функции активации и производной функций

размером $3 \times 3 \times 192$. Полностью соединенные слои содержат 4096 нейронов в каждом.

Представленная на рисунке 2 модель обучения сверточных нейронных сетей, как уже было отмечено ранее, имеет только 8 сверточных слоев и является элементарной. Учитывая общую теорию развития сверточных нейронных сетей для решения задач локализации и классификации, можно предположить, что увеличение количества слоев нейронной сети увеличивает качество обучения нейронной сети [3, 9]. Однако, чем больше слоев нейронной сети (глубже сеть), тем больше проблем возникает с ее обучением из-за проблемы исчезновения градиентов.

Таким образом, получается, что развитие сверточных нейронных сетей в части увеличения количества слоев сети для решения задач классификации и локали-

зации с одной стороны является необходимым элементом, направленным на более эффективную работу сети, а с другой стороны увеличение глубины сети порождает проблемы усложнения технологии обучения по причине исчезающего градиента. Для наиболее лучшего представления проблемы исчезающего градиента необходимо рассмотреть график сигмоидальной функции (функция активации) и ее производной функции (рис. 3).

Как можно увидеть из данных, представленных на рис. 3, при увеличении количества скрытых слоев градиенты в более ранних слоях сети будут очень низкими из-за умножения производных каждого слоя. По указанной причине обучение сверточной нейронной сети на начальных уровнях будет осуществляться гораздо медленнее, чем обучение на более высоких уровнях. Вышеобозначенная проблема порождает проблему снижения производительности нейронной сети.

В то же время проблему исчезающего градиента можно решить за счет обнуления градиента для положительных чисел. В специализированной литературе предлагается для решения обозначенной проблемы использовать функцию активации Rectified Linear Unit (ReLU) [6, 10]. Рассмотрим, как это работает на примере (рис. 4).

Если предположить, что производная функции активации будет равна либо нулю, либо равна единице, то при умножении производных для большинства слоев никакого ухудшения процесса не произойдет, и в то же время, если вес сети будет постоянно ниже нуля, то градиент будет исчезать, но гораздо медленнее по сравнению с процессом, показанным на рис. 3.

Такой подход позволит частично решить проблему исчезающих градиентов, тормозящую работу сверточной нейронной сети с большим количеством слоев, и, соответственно, увеличить точность работы нейронной сети.

Необходимо отметить, что обучение сверточных нейронных сетей с большим количеством слоев на ос-

новании данных с миллионами изображений с высоким разрешением стало возможным всего несколько лет назад и по указанной причине исследований в данной области недостаточно.

Сверточные сети, являясь ключевым элементом большинства систем интеллектуального анализа данных, способны влиять на процессы, происходящие в различных системах за счет взаимосвязи между данными о конкретном событии и возможностями прогнозирования будущих событий, и именно поэтому анализ специфики развития сверточных нейронных сетей для решения задач классификации и локализации является важным направлением научных исследований. В связи со сказанным, необходима разработка подходов к более быстрому и точному обучению сетей. При этом такие подходы должны быть комплексными и ориентированными не только на решение проблемы увеличения количества обрабатываемой информации без потери качества работы сети, но и на увеличение количества слоев нейронной сети без потери точности работы нейронной сети и ее производительности.

ЛИТЕРАТУРА

1. Каляев И.А., Гайдук А.Р., Капустян С.Г. Модели и алгоритмы коллективного управления в группах роботов. — М.: ФИЗМАТЛИТ, 2009. — 280 с.
2. Качагина К.С., Сафарова А.Д. Нейронные сети — перспективы развития // E-Scio. — 2021. — № 2 (53). — С. 34–39.
3. Ковалиева Е.Р., Алексеева В.П. Причины стремительного развития нейронных сетей // E-Scio. — 2021. — № 5 (56). — С. 122–129.
4. Мак-Каллок У.С., Питтс В. Логическое исчисление идей, относящихся к нервной активности. В сб.: «Автоматы» под ред. К.Э. Шеннона и Дж. Маккарти, 1943. — С. 115–133.
5. Усков А.А. Кузьмин А.В. Интеллектуальные технологии управления. Искусственные нейронные сети и нечеткая логика. — М.: Горячая линия-Телеком, 2004. — 143 с.
6. Cebollada, S., Payá, L., Jiang, X. et al. Development and use of a convolutional neural network for hierarchical appearance-based localization. *Artif Intell Rev* 55, 2847–2874 (2022).
7. Cebollada S, Payá L, Flores M, Román V, Peidró A, Reinoso O. A deep learning tool to solve localization in mobile autonomous robotics. In: ICINCO 2020, 17th International Conference on Informatics in Control, Automation and Robotics (Liesaint-Paris, France, 7–9 July, 2020), Ed. INSTICC.
8. Lin T.Y., Goyal P., Girshick R., He K., and Dollar P. //Focal loss for dense object detection. — Feb. 2018. — P. 1–10.
9. Mikolov T., Karafiat M., Burget L., Cernocky J., Khudanpur S. Recurrent neural network-based language model // 11th Annual Conference of the International Speech Communication Association. Japan, 2010. — P. 1045–1048.
10. Turaga S.C., Murray J.F., Jain V., Roth F., Helmstaedter M., Briggman K., Denk W., and Seung H.S. Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Computation*, 22(2):511–538, 2010.

© Кочнев Александр Александрович (drdispool@gmail.com).

Журнал «Современная наука: актуальные проблемы теории и практики»