



Столыпинский
вестник

Научная статья

Original article

УДК 004.93'1

DOI 10.55186/27131424_2023_5_2_12

РЕТРОСПЕКТИВА КОМПЬЮТЕРНОГО ЗРЕНИЯ

COMPUTER VISION RETROSPECTIVE

Дударев Д. С., Студент, Псковский государственный университет РФ, г. Псков

Дударев К. С., Студент, Псковский государственный университет РФ, г. Псков

Мотайленко Л. В., доцент кафедры информационно-коммуникационных технологий Псковского государственного университета РФ, г. Псков

Dudarev D. S., Student, Pskov State University of the Russian Federation, Pskov

Dudarev K. S., Student, Pskov State University of the Russian Federation, Pskov

Motailenko L. V., Associate Professor of the Department of Information and Communication Technologies of Pskov State University of the Russian Federation, Pskov

Аннотация: Значительный рост вычислительной мощности компьютеров и появление новых математических моделей и алгоритмов в последние годы позволило добиться значительного прогресса в области компьютерного зрения. Новые технологии, основанные на машинном обучении, позволяют создавать надежные решения для сложных задач распознавания зрительных образов. Многие из этих решений имеют практические применения на рынке, а потому привлекают все больше инвестиций.

Abstract: Significant growth in computer computing power and the emergence of

new mathematical models and algorithms in recent years has made significant progress in the field of computer vision. New machine-based technologies enable you to create reliable solutions for complex visual image recognition tasks. Many of these solutions have practical applications in the market, and therefore attract more and more investment.

Ключевые слова: Компьютерное зрение, машинное обучение, поиск по шаблону, искусственные нейронные сети, глубинное обучение.

Keywords: Computer vision, machine learning, template search, artificial neural networks, deep learning.

Компьютерное зрение (Computer Vision), а также иногда машинное зрение (Machine Vision) – научная область, занимающаяся исследованиями в области автоматической фиксации и разного рода обработки изображений (обнаружение, отслеживание, идентификация) с помощью компьютера.

Одни из первых попыток создать такие системы были предприняты в 1960-х годах, но из-за низкой мощности компьютеров того времени эта область долго не была достаточно исследованной. Современные скоростные процессоры, дешевая память на диске и высококачественные камеры и высокая пропускная способность каналов позволяют добиться значительных результатов в этой области, особенно в связи с последними исследованиями в области машинного обучения и, в частности, нейронных сетей.

Рост качества технологий компьютерного зрения позволил активно их использовать в разных областях промышленности и бизнеса. К примеру, многие современные конвейеры оснащены автоматическими механизмами проверки качества деталей, корректной маркировки, выравнивания деталей и других задач. Примером такого подхода может служить индустрия микрочипов, где камеры автоматически контролируют размещение кремниевых пластин и планировку интегральных схем. Можно перечислить большое количество областей цифровой экономики, в которых применяется современное компьютерное зрение:

- Системы “умный город”.
- Инфраструктурные и транспортные системы с контролем и балансировкой нагрузки.

- Автомобильные системы, в том числе беспилотные автомобили.
- Беспилотные летательные аппараты, в том числе дроны.
- Фармацевтика и медицина.
- Видеоаналитики.
- Распознавание человека и предметов.
- Идентификация лиц.
- Оценка сцены, времени, места.

Основная цель компьютерного зрения – получение полезной информации из изображения или серии изображений. Задачами могут быть такие, как:

- калибровка оптических систем, настройка и синхронизация настроек камер,
- определение движения объектов (Motion Tracking),
- задачи распознавания объектов,
- задачи реконструкции сцены (обычно 3D-сцены из одного или нескольких 2D-кадров),
- задачи сравнения изображений и идентификации изменений.

Искусственные нейронные сети: появление, развитие и современное состояние

Стандартная нейронная сеть состоит из множества простых связанных процессоров, называемых нейронами, каждый из которых производит последовательность активаций (векторов из вещественных чисел). Входные нейроны (нейроны первого, входного, слоя) активируются сенсорами, воспринимающими окружающую среду, остальные активируются с помощью взвешенных соединений с другими нейронами, активировавшимися ранее.

Нейроны могут влиять на окружающую среду, вызывая действия. Обучение (или присвоение доверия) состоит в нахождении таких весов, с которыми нейронная сеть достигает желаемого результата, например, распознавание эталонного образца. В зависимости от задачи и типа связи нейронов, такое поведение может требовать цепи шагов, каждый из которых нелинейно преобразует активацию предыдущих шагов. Глубокое обучение – обучение сетей

с множеством таких шагов.

Неглубокие модели, подобные нейронным сетям, существуют на протяжении многих десятилетий. Метод называется алгоритмом обратного распространения ошибки. Однако, тренировать глубокие нейронные сети этим методом было непрактично из-за недостаточности доступных вычислительных ресурсов. Такие проблемы стали объектами исследований только в 1990-е.

Тогда в основном применялось обучение без учителя, но методы обучения с учителем совершенствовались.

Прямые (ациклические) и рекуррентные нейронные сети успешно применялись во множестве задач. Рекуррентные нейронные сети – самые глубокие, в принципе они являются универсальными вычислительными системами и могут, в теории, обрабатывать и сохранять любые паттерны. В отличие от классических методов, рекуррентные нейронные сети могут обучаться на создание программ, эффективно смешивающих последовательное и параллельное вычисление.

Общепринятая нотация для описания нейронных сетей

Топология нейронной сети может изменяться со временем. В определенный момент времени она может быть записана как конечное множество вершин (нейронов) $N = \{u_1, u_2, u_3, \dots\}$ и конечное множество $H \subseteq N \times N$ направленных взвешенных ребер (связей) между вершинами. Граф прямой нейронной сети – ациклический, граф рекуррентной сети – циклический. Первый (входной слой) – множество входных нейронов. В ациклических нейронных сетях k -й слой ($k \in N, k > 1$) – множество всех вершин $u \in N$ такое, что существует путь длиной $(k - 1)$ между одним из входных нейронов и одним из нейронов u .

Поведение нейронной сети определяется множеством вещественных параметров (весов) w_i ($i = 1 \dots n$). Сосредоточимся пока на одной эпохе распространения информации, не обращая внимание на обучение и изменение весов. То есть опишем то, что происходит при работе обученной нейронной сети.

Входной нейрон получает на вход какое-то вещественное значение. Это значение распространяется, то есть копируется на вход всем нейронам, связанным с исходным, при этом умножается на вес ребра. Каждый из этих нейронов складывает все пришедшие значения, и применяет к сумме нелинейную функцию активации (например, $f(x) = \max(0, x)$) и полученное значение распространяет далее по сети. Значение активации нейронов последнего слоя и есть выходное значение нейронной сети. Существует большое количество дополнительных операций, применяемых в этой процедуре (например, нормализация батча: после выбранного слоя все нейроны линейно изменяются так, чтобы среднее значение всех нейронов было равно нулю, а дисперсия – единице), но они необязательны и применяются в зависимости от задачи.

В обучении с учителем вектор выходных значений x_t нейронной сети можно сравнить с вектором, заданным учителем (вручную, другой сетью или любым другим способом) d_t , при этом разницу между ними можно считать величиной ошибки, например $e_t = \frac{1}{2} (x_t - d_t)^2$. Задача обучения – минимизировать эту ошибку с помощью метода градиентного спуска.

Первые нейронные сети, 1940-е

Ранние архитектуры нейронных сетей не обучались (McCulloch & Pitts, 1943). Нейроны в них обладали дискретными весами и напоминали динамическую логическую схему (Рисунок 1) [4]. Первые исследования о машинном обучении были опубликованы позже (Hebb, 1949). Первые нейронные сети – варианты линейных регрессий, существующих в математике еще с XIX века.

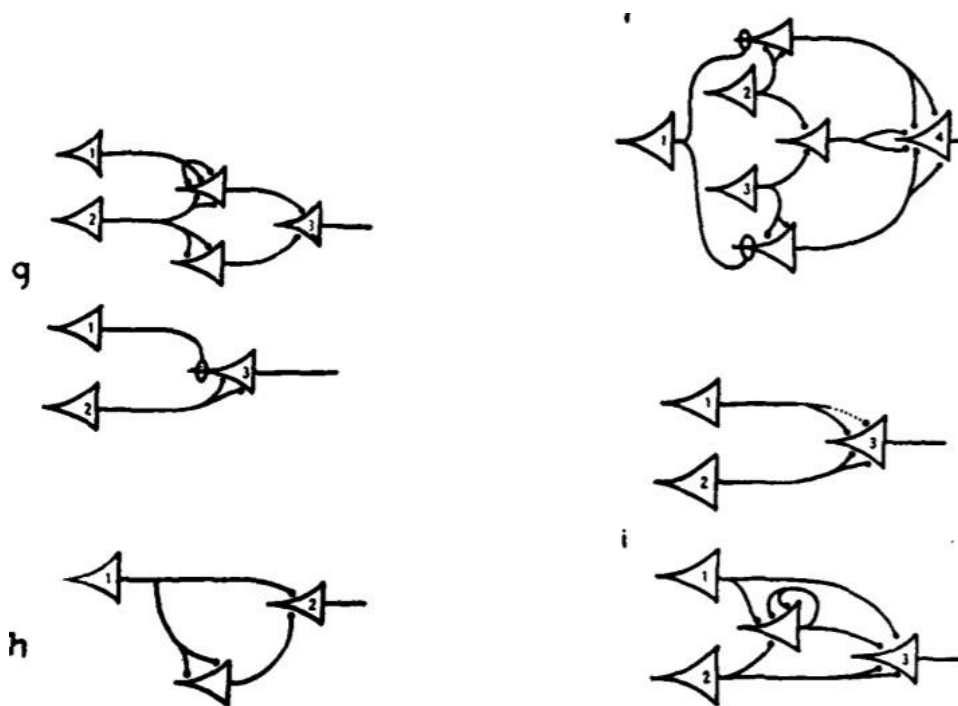


FIGURE 1

Рисунок 1. Возможные схемы работы нейронов, предложенные McCulloch & Pitts

Нейронные сети в 1960-х. Изучение сетчатки глаза

В сетчатке глаза кошки были обнаружены простые и сложные клетки. Эти клетки сигнализировали в ответ на некоторые характеристики зрительного образа, например, вертикальное или горизонтальное расположение границ объектов. Сложные клетки реагировали на более сложные паттерны, чем простые клетки. В будущем эти наблюдения стали основой для сверточных нейронных сетей – архитектуры, до сих пор успешно используемой при решении задач распознавания паттернов.

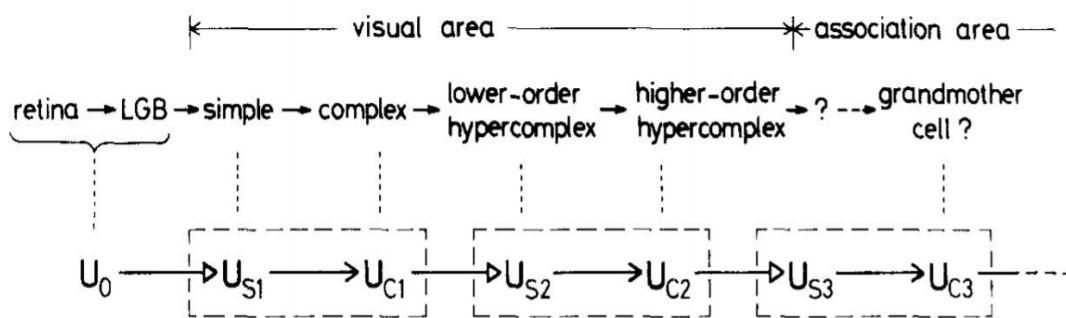
1965: Групповые методы обработки данных. Многослойные сети

Алгоритм группового метода обработки данных (group method of data handling) (Ivakhnenko, 1967) позволил создать первую обучающуюся глубокую нейронную сеть типа многослойный персептрон. Нейроны такой сети имели полиномиальные функции активации, и пошагово модифицировались исходя из результатов регрессионного анализа тренировочных данных, и проверяемых на тестовой выборке, не участвующей в обучении. В таких сетях могло быть несколько слоев. Они широко использовались на протяжении нескольких десятилетий.

1979: Операция свертки. Неокогнитрон Фукусимы

Неокогнитрон (Fukushima, 1979) был первой моделью глубокой нейронной сети, учитывающая нейрофизиологические исследования природы зрения (Рисунок). Тут впервые вводится сверточный слой, входная матрица которого фильтруется матрицей весов с помощью математической операции свертки. Получившаяся в результате матрица может быть передана последующим сверточным слоям. Был также представлен механизм подвыборки (subsampling), который в последующих сетях будут заменен на выборочные слои (pooling layers). Хотя архитектура модели близка к современной, неокогнитрон не обучается с помощью изменения весов алгоритмом обратного распространения ошибки, вместо этого веса задаются алгоритмами обучения без учителя (корреляции и т.д.) и частично вручную.

Рисунок 2. Схема предположительного устройства неокогнитрона Фукусимы.



1960-1981. Разработка алгоритма обратного распространения ошибки

Идея минимизации параметров сложной, нелинейной, дифференцируемой многошаговой модели с помощью градиентного спуска обсуждалась с начала 1960-х, но это часто делалось с помощью якобиана. Такой метод неэффективен для большого количества параметров.

Алгоритм обратного распространения ошибки был впервые опубликован в 1970 году (Linnainmaa, 1970), но без концептуальной связи с нейронными сетями. Превые работы о применении этого метода к нейронным сетям были опубликованы в 1980-х (Werbos, 1981; LeCunn, 1985). Эти исследования сильно повлияли на популярность нейронных сетей как потенциально перспективного метода машинного обучения [4].

1989: Алгоритм обратного распространения ошибки для сверточных сетей

Сверточная нейронная сеть была предложена Яном Лекуном в 1988 году специально для обработки изображений. Лекун ввел два новых типа слоя: сверточный и выборочный, имитирующие действия зрительной коры головного мозга (Рисунок 9-11). Нейронная сеть, содержащая сверточные слои, называется сверточной нейронной сетью. Как и в предыдущем случае, со временем биологические знания и компьютерные технологии разошлись, но сверточные сети доказали свою эффективность, и это один из основных типов современных нейронных сетей.

Сверточный слой использует маску, называемую ядром (kernel), умножаемую последовательно на смежные области предыдущего слоя. Ядро служит фильтром для поиска одного паттерна последовательно во всем слое (Рисунок). Внутри одного слоя несколько фильтров могут применяться независимо, соответственно увеличивая количество матриц на выходе слоя.

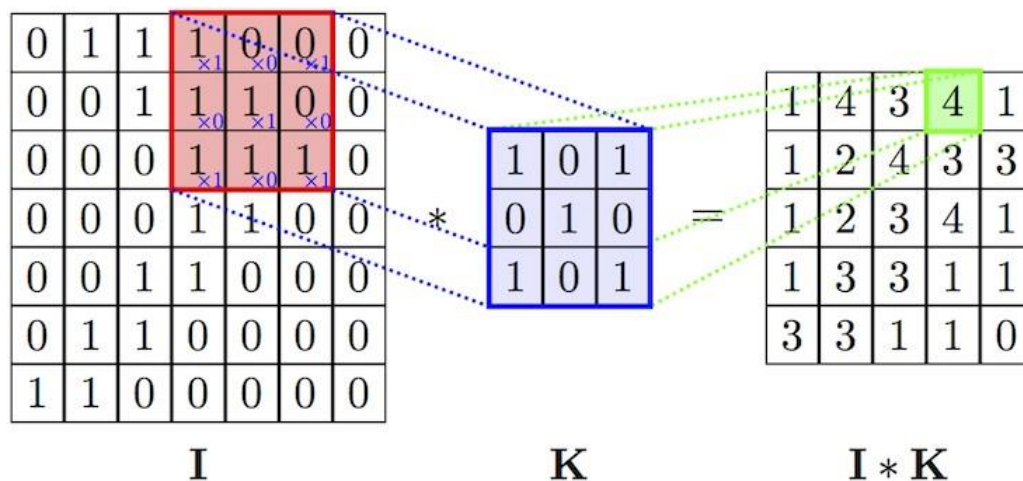


Рисунок 3. Пример работы сверточного слоя с входной матрицей I и фильтром K.

В 1989 алгоритм обратного распространения ошибки был применен (LeCun, 1989) к сверточным нейронным слоям с совместными весами и адаптивными связями, напоминающим неокогнитрон. Эта комбинация, объединенная со слоем максимальной подвыборки (max-pooling) и вычисляемая на графических процессорах, стала ингредиентом многих современных глубоких нейронных сетей. В этом же году был создан MNIST – размеченный набор сканов рукописных цифр, ставший самым известным набором данных для машинного обучения.

Сверточные сети достигли высокого уровня точности в некоторых задачах (в том числе MNIST) и стали использоваться коммерчески.

1991: Фундаментальная задача глубокого обучения градиентным спуском

Работа Хокрейтера (Hochreiter, 1991) стала важным этапом в развитии нейронных сетей. Известно, что глубокие нейронные сети тяжело обучать с помощью алгоритма обратного распространения ошибки. Хокрейтер обнаружил основную причину этого: проблему затухающих градиентов. Стандартные функции активации, примененные несколько раз, экспоненциально уменьшают производную и градиент. Последующие исследования во многом опирались на эти результаты, стремясь найти способы избежать или компенсировать проблему. Одна из основных причин, благодаря которым возможно современное глубокое обучение, состоит в том, что мощные вычислительные ресурсы позволяют частично компенсировать проблему затухающих градиентов, но принципиально ее не решают.

1992: Слой максимальной подвыборки

Неокогнитрон стал базой для создания Крецептрона (Weng, 1992). Одно из его нововведений – смазывающий слой (blurring layer), который сейчас принято называть слоем подвыборки. Этот слой использует пространственные характеристики сигнала, позволяя уменьшить размер слоя при потере небольшого количества информации. Это позволяет сильно сэкономить вычислительные ресурсы. Слои подвыборки используются в большинстве современных сверточных сетей.

1995: Нейронная сеть с долгой краткосрочной памятью

Глубокая рекуррентная нейронная сеть с долгой краткосрочной памятью (LSTM – long short memory unit) (Рисунок 12) позволяет резко уменьшить количество весов, частично компенсируя проблему затухающих градиентов и позволяя создавать очень глубокие рекуррентные нейронные сети.

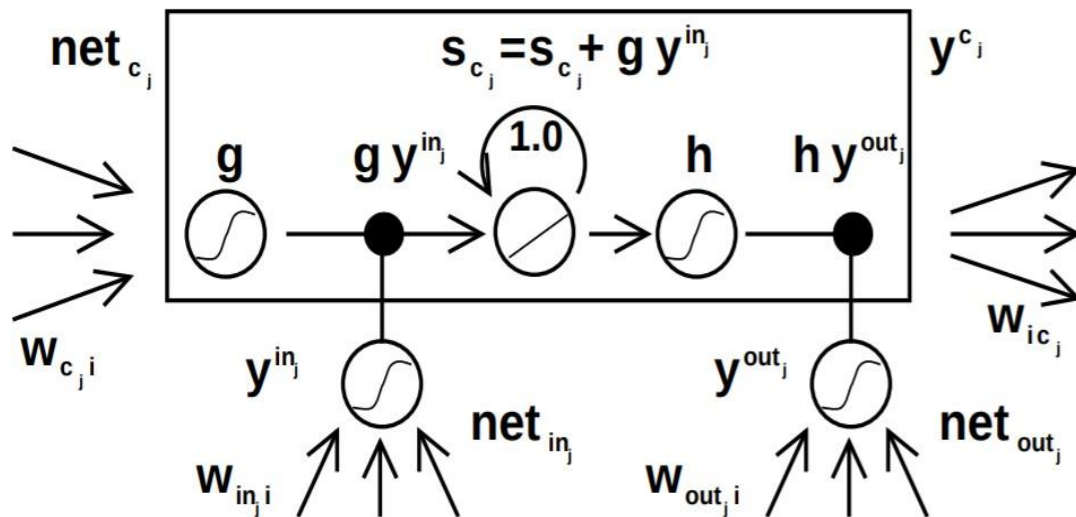


Рисунок 4. Схема устройства блока нейронной сети с долгой краткосрочной памятью

Такие сети могут запоминать информацию на тысячи дискретных временных интервалов обучения, тогда как обычные рекуррентные нейронные сети редко справляются с десятью интервалами. Более того, LSTM можно описать как конечный автомат, самостоятельно обучающийся создавать программу для данной задачи. Многие задачи, решение которых было нереалистичным с помощью глубокого обучения, стали решаемы: идентификация временного порядка сильно разделенных шумом входных сигналов, эффективное сохранение вещественных чисел с высокой точностью на протяжении длительного периода времени (дискретных интервалов, итераций), арифметические операции на непрерывном потоке данных, генерация ритма и траекторий. Например, LSTM RNN способна (Perez-Ortiz, 2003) обучаться справляться с входными последовательностями длиной в миллионы символов, научившись отличать валидные строки вроде $a^{10,000,000}b^{10,000,000}c^{10,000,000}$ от очень похожих невалидных, например, $a^{10,000,000}b^{9,999,999}c^{10,000,000}$. Появившиеся в то же время двунаправленные рекуррентные нейронные сети тут же стали успешно применяться в ряде биологических задач: они предсказывали характеристики малых органических молекул, карты взаимодействия протеина, и т.д..

2009: Первые официальные рекорды точности, установленные рекуррентными и сверточными нейронными сетями

Модель LSTM победила в соревновании по переводу между тремя языками (английским, французским, арабским), не имея никакого априорного знания о задаче. В том же году сложная модель, включавшая в себя сверточную сеть, победила в соревновании TRCVIД 2009 по распознаванию действий людей в видео из камер наблюдения (Yang, 2009).

2010: Графический процессор

Новый рекорд на данных MNIST (99.65%) был поставлен нейронной сетью, обученной с использованием графического процессора (Ciresan, 2010). Реализация алгоритмов обратного распространения ошибки на графическом процессоре была в десятки раз быстрее стандартной. Поскольку все использованные алгоритмы были известными на протяжении многих лет, стало ясно, что развитие вычислительных мощностей компьютеров не менее важно для машинного обучения, чем эволюция алгоритмов. Практически все крупные современные нейронные сети обучаются на графических процессорах.

В определенных узких областях нейронные сети (Ciresan, 2011) (Sermanet, 2011) приближались к человеческой точности распознавания визуальных паттернов. Хотя компьютер выиграл человека в шахматы еще в 1997, в задаче универсального распознавания объектов к 2010 году ни один компьютер не превосходил ребенка.

Заключение: будущее компьютерного зрения

Компьютерное зрение — быстро растущая область исследований, уже трансформирующая многие индустрии, а также повседневную жизнь человека. Точность анализа и распознавания образов неуклонно растет, позволяя создавать все более сложные коммерческие приложения. Основные тенденции развития компьютерного зрения можно обобщить следующим образом:

❖ Расширение применения компьютерного зрения в промышленных системах. Компьютерное зрение все активнее применяется в медицине и фармацевтике, в производстве пищевых продуктов, в конвейерах технически сложных промышленных процессов и позволяет сильно повысить эффективность контроля качества

❖ **Облачные системы.** Рост требований к производительности компьютеров у систем глубокого обучения порождает спрос на аренду облачных серверов для вычислений, что позволяет быстрее проводить необходимые вычисления.

❖ **Робототехника.** Увеличение использования промышленных роботов непременно приводит к повышению спроса на системы компьютерного зрения

Библиографический список:

1. Дэвид Форсайт, Жан Понс. Компьютерное зрение. Современный подход = Computer Vision: A Modern Approach. — М.: «Вильямс», 2004.
2. Л. Шапиро, Дж. Стокман. Компьютерное зрение = Computer Vision.— М.: Бинум. Лаборатория знаний, 2006.
3. Goodfellow, Y. Bengio, A. Courville. “Deep Learning Book”, MIT Press, 2016 <http://deeplearningbook.org>.
4. M. Nielsen, “Neural Networks and Deep Learning”, Determintation Press, 2015 <http://neuralnetworksanddeeplearning.com>.
5. W. Dai, C. Dai, S. Qu, J. Li, S. Das “Very Deep Convolutional Neural Networks for Raw Waveforms”, ICASSP, 2017 <https://arxiv.org/abs/1610.00087>.
6. J. Yosinski “Understanding Neural Networks Through Deep Visualization”, ICML DL Workshop, 2015 <http://yosinski.com/deepvis>.

© Дударев Д. С., Дударев К. С., Мотайленко Л. В., 2023 Научный сетевой журнал «Столыпинский вестник» №2/2023

Для цитирования: Дударев Д. С., Дударев К. С., Мотайленко Л. В. РЕТРОСПЕКТИВА КОМПЬЮТЕРНОГО ЗРЕНИЯ// Научный сетевой журнал «Столыпинский вестник» №2/2023