

*Горячкин Б. С., кандидат технических наук, доцент, Московский
государственный технический университет им. Н.Э. Баумана*

*Китов М. А., магистр, Московский государственный технический
университет им. Н.Э. Баумана*

КОМПЬЮТЕРНОЕ ЗРЕНИЕ

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.

Данная статья представляет собой обзор основных подходов к решению задач компьютерного зрения, с акцентом на глубинное машинное обучение и искусственные нейронные сети. Также в статье приведен обзор рыночных показателей в данной области и показано, что рынок компьютерного зрения быстро развивается и имеет серьезные перспективы.

Ключевые слова: Компьютерное зрение, машинное обучение, поиск по шаблону, искусственные нейронные сети, глубинное обучение, рынок систем компьютерного зрения в России.

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.

This article is an overview of the basic approaches to solving computer vision problems, with a focus on deep machine learning and artificial neural networks. The article also provides an overview of market indicators in this area and shows that the computer vision market is developing rapidly and has serious prospects.

Keywords: Computer vision, machine learning, template search, artificial neural networks, deep learning, computer vision systems market in Russia.

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.

Технологии компьютерного зрения

На физическом уровне системы компьютерного зрения состоят из средств захвата изображения (камера или несколько камер) и компьютера общего назначения, используемого для обработки изображений. При этом применяются специальные программные средства. Многие программные средства могут работать “в облаке”, удаленно, что позволяет масштабировать системы компьютерного зрения и централизовать управление ими.

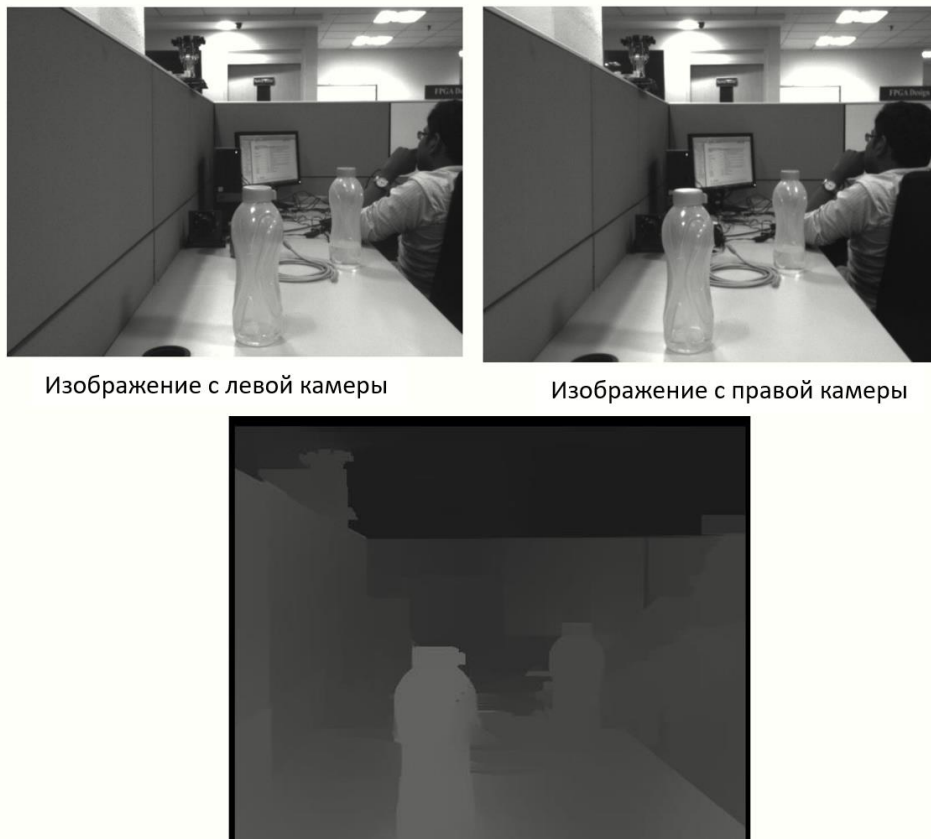
В некоторых случаях система также может включать средства для искусственной подсветки интересующих объектов, тогда компонент, объединяющий камеры и подсветку, называется оптической системой. От качества оптической системы во многом зависит эффективность работы, так как

многие современные методы используют анализ малых деталей на изображении, для чего требуется его высокая четкость.

Типы оптических систем

Оптические системы обычно классифицируют на одномерные, двумерные и объемные (трехмерные), при этом могут быть исключения, например, различные системы панорамной съемки не всегда легко отнести к одному из этих типов [2].

Классический метод для получения трехмерного изображения – стереозрение (рис. 1).



Изображение с левой камеры

Изображение с правой камеры

Рис. 1. Пример работы системы стереозрения

Это метод извлечения информации о дальности объектов от камеры при помощи двух кадров (как правило, снятых с разных камер). Сопоставление таких кадров позволяет оценить расстояние до объектов. Панорамные системы (рис. 2) используются в основном для видеонаблюдения в условиях, где

необходим широкий обзор (например, обнаружение объектов в небе) и в геоинформационных системах (Google Maps и т.д.).



Рис. 2. Панорамная съемка

Еще один способ получить 3D-изображение – многокамерные оптические системы, в которых камеры сфокусированы на общем объекте. Такие технологии часто используются, например, для моделирования трехмерной виртуальной реальности.

Существует ряд распространенных программных платформ, позволяющих эффективно реализовывать системы компьютерного зрения. Перечислим некоторые из них:

- ❖ OpenCV – самый популярный пакет, распространяющийся бесплатно и с открытым исходным кодом. Это библиотека математических алгоритмов для анализа изображений, реализованная на C++, но имеющая API для многих популярных языков программирования, таких как Python, Java, Matlab и других.

- ❖ PCL – также открытая платформа, позволяющая обрабатывать двумерные и трехмерные изображения. Содержит множество реализованных

алгоритмов разной направленности: оценку характеристик, реконструкцию поверхности, восстановление сцены, поиск соответствий в изображениях и многое другое.

❖ ROS – специализированная платформа для управления робототехникой, включает богатый набор алгоритмов, применяемых в этой области.

❖ MATLAB – высокоуровневый язык программирования общематематического назначения, популярный в академической среде. Содержит большое количество инструментов, упрощающих проведение исследований: от анализа данных и визуализации до создания и внедрения моделей.

❖ CUDA – проприетарная программная разработка компании NVIDIA, специально заточенная для работы с графическими процессорами производства этой компании. Реализует крайне эффективную архитектуру для параллельных вычислений, что делает ее практически незаменимой в любой высоконагруженной системе, обрабатывающей изображения.

Алгоритмы и методы компьютерного зрения можно условно разделить на две группы:

1. “Классическое” компьютерное зрение, применяемое, когда требуется получить некоторую количественную информацию об изображении (связанную с цветом, формой, количеством объектов и т.д.). Наиболее надежно работает в задачах, поддающихся формализации и разбиению на подзадачи. Большинство методов из этой группы сначала извлекают полезные характеристики из изображения, а затем работают с ними для решения задачи.

2. Системы машинного обучения, в том числе глубокого машинного обучения (нейронные сети). Это сложные системы, гораздо более требовательные к вычислительным ресурсам и объемам данных, но бурно развивающиеся в последние годы. Системы такого рода частично имитируют человеческие способности к восприятию изображения, и поэтому позволяют извлекать из изображения гораздо более сложную информацию. Такие системы

имеет смысл применять в случаях, когда разбиение задачи и составление пошагового алгоритма крайне трудозатратно или вовсе невозможно.

Подходы классического компьютерного зрения

Существуют следующие основные подходы к решению классических задач [1]:

- I. Анализ контуров и градиентов
- II. Сопоставление по шаблону
- III. Сопоставление по характеристикам (обычно ключевым точкам)

Рассмотрим некоторые самые распространенные подходы.

I. Контурный анализ.

Контур – кривая, описывающая границы объекта или его фрагмента. Метод контурного анализа полагается в основном на контуры изображения. Самые распространенные контурные методы – это методы математической морфологии, позволяющие эффективно решать простые задачи (отделение объекта от фона, создание масок объектов, восстановление плохо читаемого текста из отсканированных фотографий и т.д.) (рис. 3).

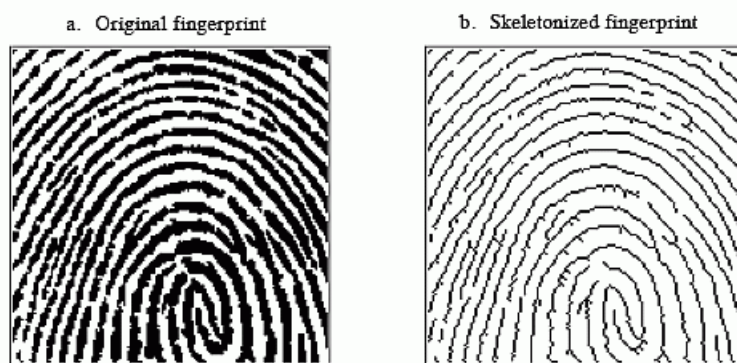


FIGURE 25-11
Binary skeletonization. The binary image of a fingerprint, (a), contains ridges that are many pixels wide. The skeletonized version, (b), contains ridges only a single pixel wide.

Рис. 3. Обработка отпечатка пальца с помощью контурного анализа

Эти методы имеют ряд важных ограничений:

✓ они не могут быть применены, если на изображении невозможно различить контур (из-за природы объектов или просто из-за плохого цветового контраста),

✓ они не способны адекватно обработать перекрытие одних объектов другими,

✓ они неустойчивы к шуму, плохая видимость или плохо качество изображение сильно ухудшает качество их работы.

II. Поиск по шаблону, поиск по ключевым точкам

Поиск по шаблону – самый распространенный метод классического компьютерного зрения. Существует несколько его подвидов, например, при поиске простого соответствия исходное изображение сканируется пошагово, и при каждом сканировании происходит расчет доли совпадения данного участка с шаблоном согласно выбранной метрике.

Наиболее важный в этой задаче подход – выделение и сравнение ключевых точек (рис. 4) [3]. Существует несколько классических алгоритмов (например, SIFT, SURF), которые:

1. Выбирают наиболее “важные” точки на изображении, оценивая их положение, контраст, с соседними точками и т. д.

2. Создают “дескрипторы” ключевых точек, то есть идентификаторы достаточно устойчивые к основным типам преобразований: изменение освещения, изменение масштаба, изменение расстояния до объекта, поворот объекта.

3. Позволяют сравнивать дескрипторы на разных изображениях и, тем самым, сопоставлять их фрагменты.

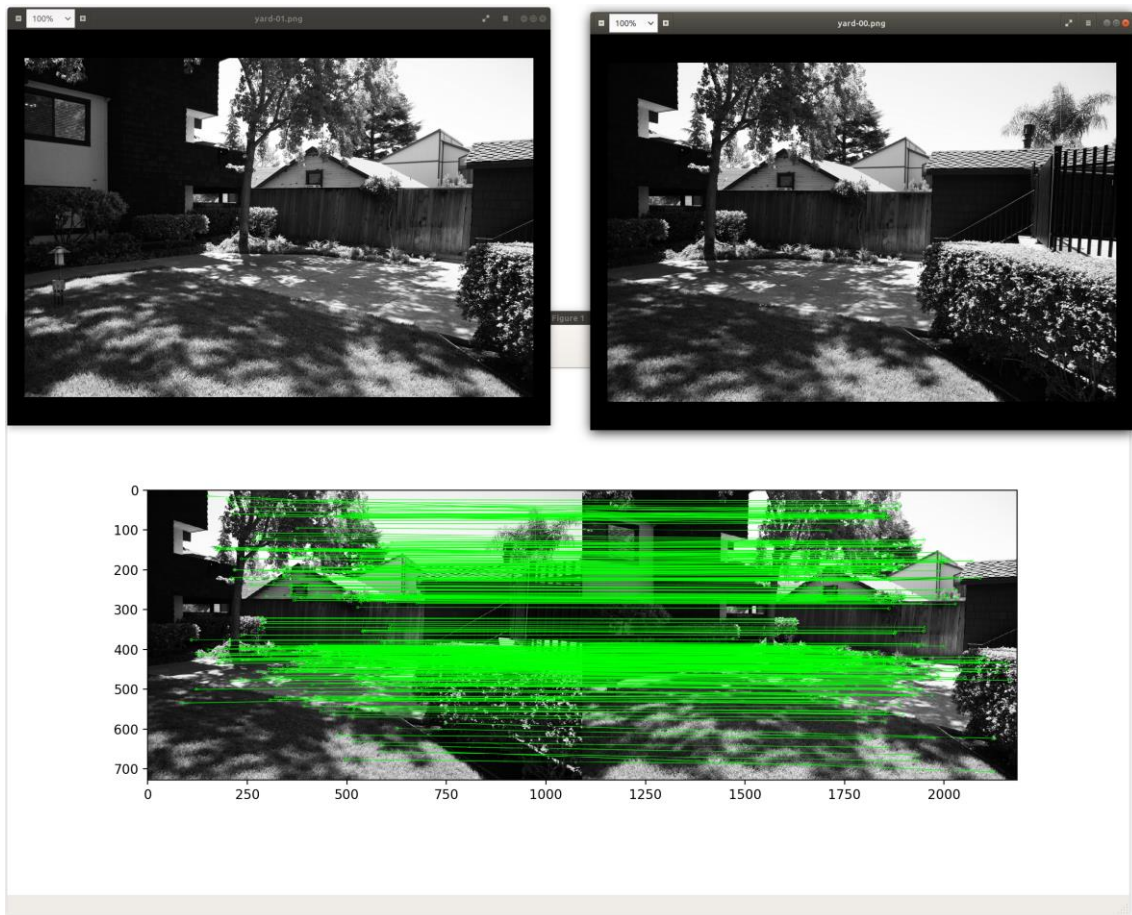


Рис. 4. Сопоставление ключевых точек с помощью метода SIFT

На рисунке 5 можно увидеть пример использования метода сопоставления по ключевым точкам для склеивания панорамы из нескольких изображений.



Рис. 5. Склеивание панорамы с помощью метода SIFT

Еще один популярный способ – поиск соответствия на базе областей (корреляционные методы), основанные на смешивании двух указанных выше методов (рис. 6). Это эффективно, когда ключевых точек на изображениях недостаточно, и приходится пользоваться дополнительно пиксельным сравнением.

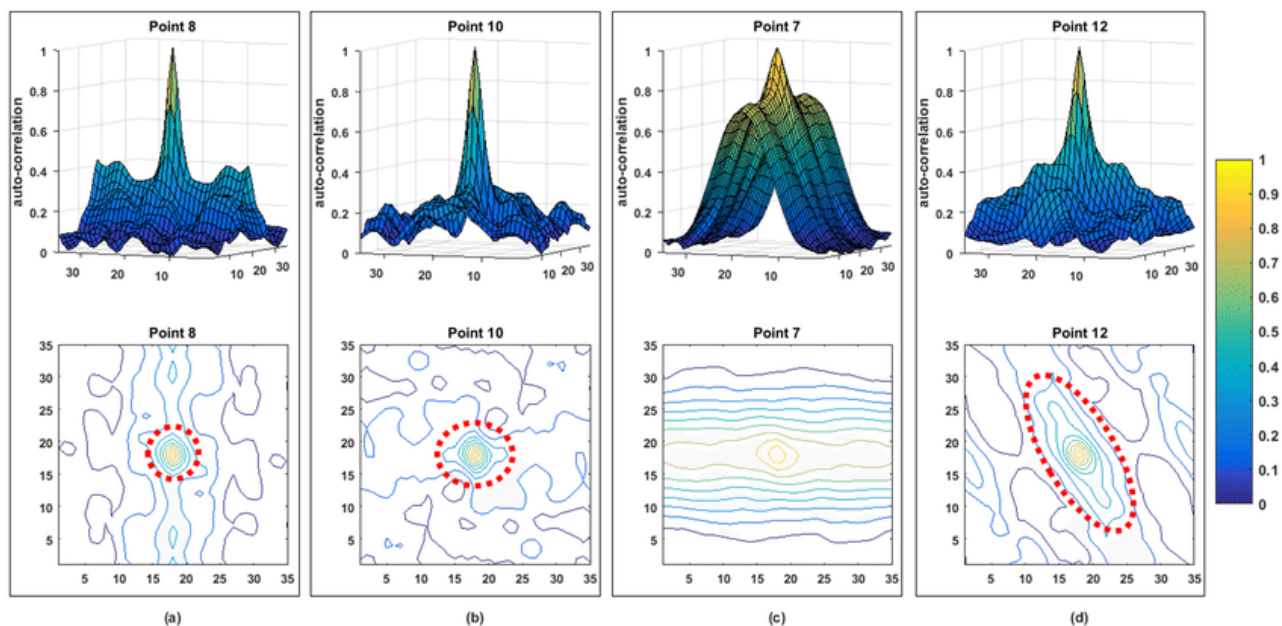


Рис. 6. Результаты анализа изображений в геодезии с помощью корреляционных методов

Перечисленные подходы являются ключевыми для классического компьютерного зрения.

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. Модели с несколькими нелинейными слоями нейронов появились в 1960-х - 1970-х годах. Эффективный метод обучения нейронных сетей любой глубины с учителем разрабатывался в 1970-х и был применен к нейронным сетям в 1981 году [5]. 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.

В новом тысячелетии нейронные сети привлекли большое внимание, став в некоторых задачах более точными, чем классические алгоритмы, например, машины опорных векторов. С 2009 года нейронные сети стали активно применять в различных задачах по распознаванию паттернов, и, именно, они впервые достигли сверхчеловеческой точности в ограниченных областях.

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) [4]. Первые исследования о машинном обучении были опубликованы позже (Hebb, 1949). Первые нейронные сети – варианты линейных регрессий, существующих в математике еще с XIX века.

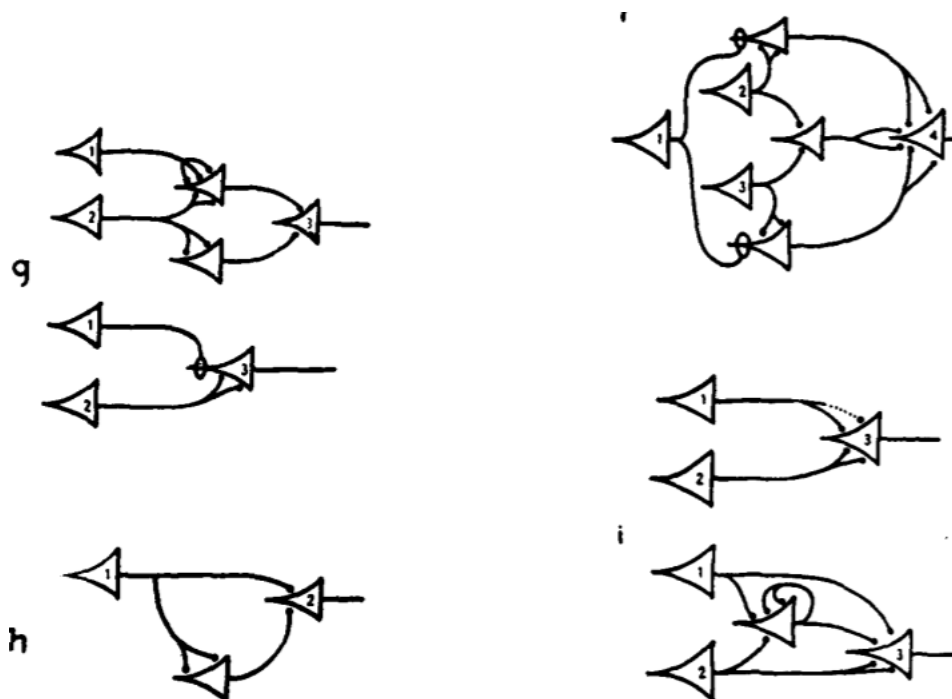


FIGURE 1

Рисунок 7. 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.

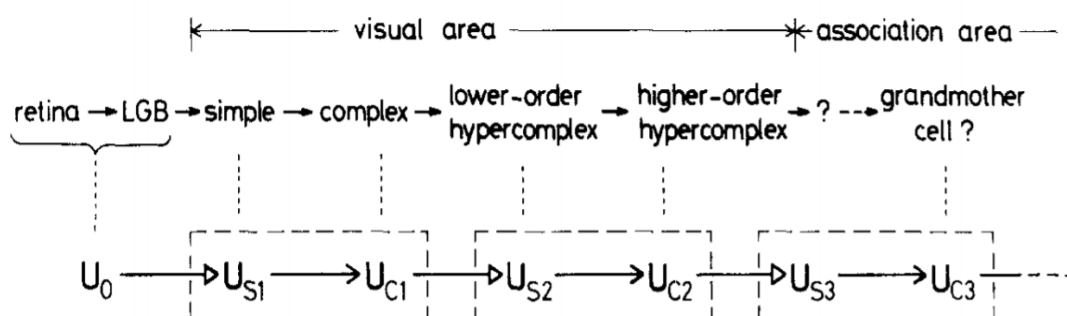


Рисунок 8. Схема предположительного устройства неокогнитрона Фукусимы.

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. Первые работы о применении этого метода к нейронным сетям были опубликованы в 1980-х (Werbos, 1981; LeCun, 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 Лекунем 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.

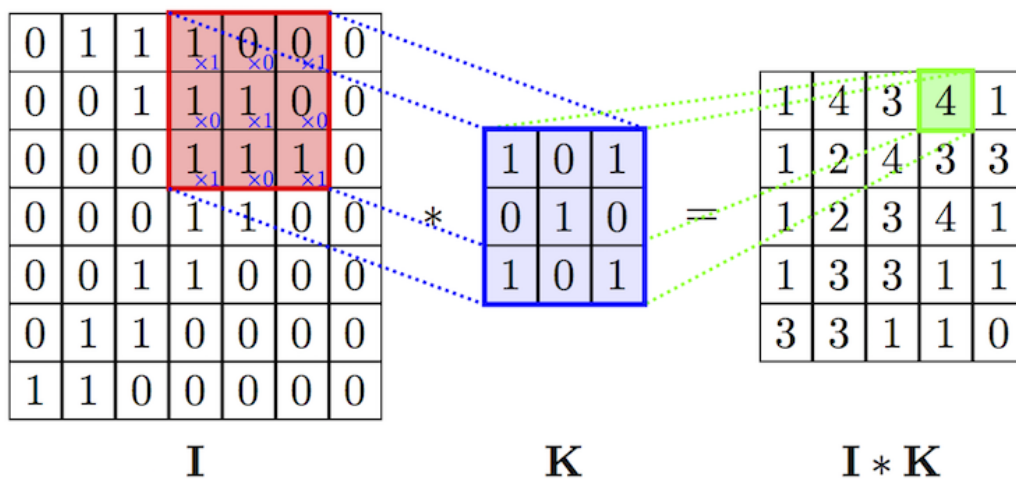


Рисунок 9. Пример работы сверточного слоя с входной матрицей I и фильтром K.

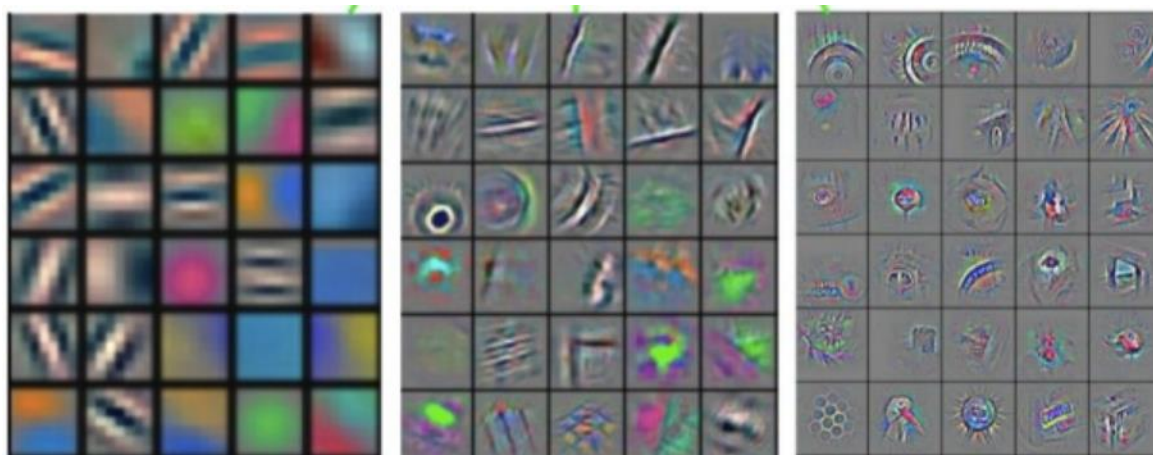


Рисунок 10 Визуализация весов нескольких слоев GoogleNet

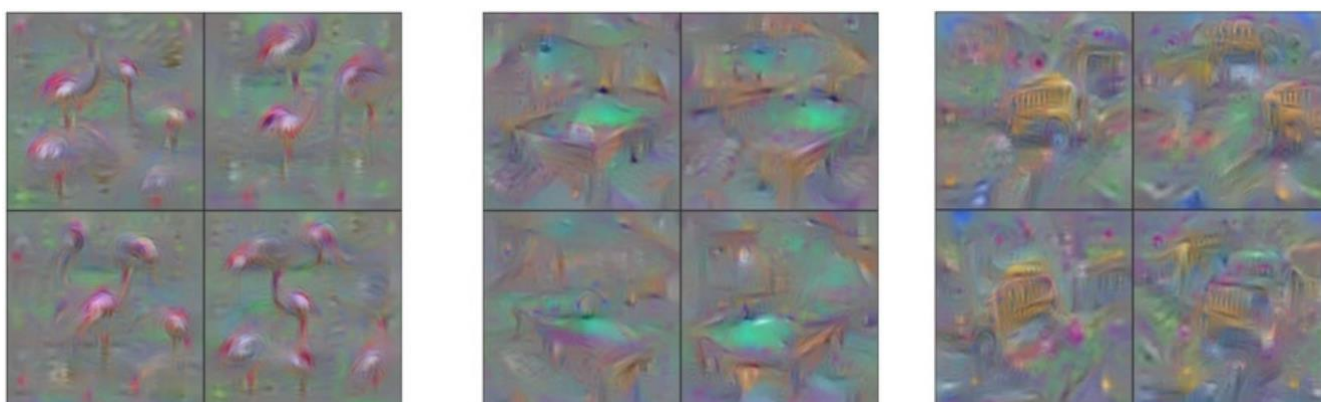


Рисунок 11. Визуализация фильтров для классов «фламинго», «бильярдный стол» и «школьный автобус» из последнего свертогого слоя GoogleNet.

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.

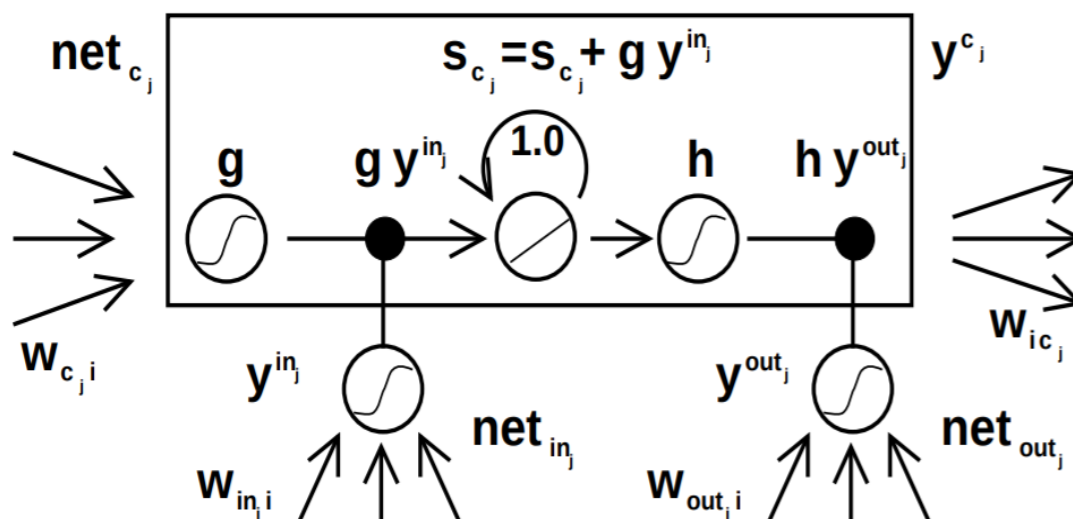


Рисунок 12. Схема устройства блока нейронной сети с долгой краткосрочной памятью

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.

На протяжении 2000-х годов LSTM постепенно улучшали точность распознавания речи, в конце концов обойдя производительность всех других методов при большом словаре (Geiger, 2014): скрытых марковских моделей, гауссовский смешанных моделей и т.д. LSTM применяются во многих задачах, связанных с преобразованием одних последовательностей в другие, доминируя, например, в области машинного перевода.

2003: Новые рекорды и коммерчески успешные глубокие нейронные сети

В начале двухтысячных в большинстве коммерческих продуктов для машинного обучения использовались не нейронные сети, а другие методы, например, метод опорных векторов (SVM). Но в некоторых узких областях нейронные сети обгоняли в точности другие алгоритмы. Например, не очень глубокая сверточная нейронная сеть (Simard, 2003), обученная алгоритмом обратного распространения ошибки, поставила рекорд в MNIST: точность 99.6%.

В то же время LSTM в распознавании речи начали получать результаты, сравнимые с результатами скрытых марковских моделей.

2006: Глубокая сеть доверия

В 2006 году Смоленски опубликовал модель сети глубокого доверия (Deep Belief Network), состоящую из наложенных поверх друг друга ограниченных машин Больцмана (Restricted Boltzman Machine), каждый из которых воспринимает более высокоуровневый паттерн. Сеть показала хорошие результаты на MNIST без предобработки (точность 98.8%), а также в задаче

распознавания фонем (TIMIT, точность 63.3%). Это вызвало волну интереса к глубокому обучению (и сам термин “глубокое обучение” стал широко употребляться только в 2006 году).

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 TRCVID 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.

2012: Нейронные сети и ImageNet

В 2012 году связка сверточных нейронных сетей Alexnet (Рисунок 13) достигла рекордного результата в точности распознавания объектов ImageNet – самом крупном соревновании по распознаванию тысяч различных объектов на изображениях.

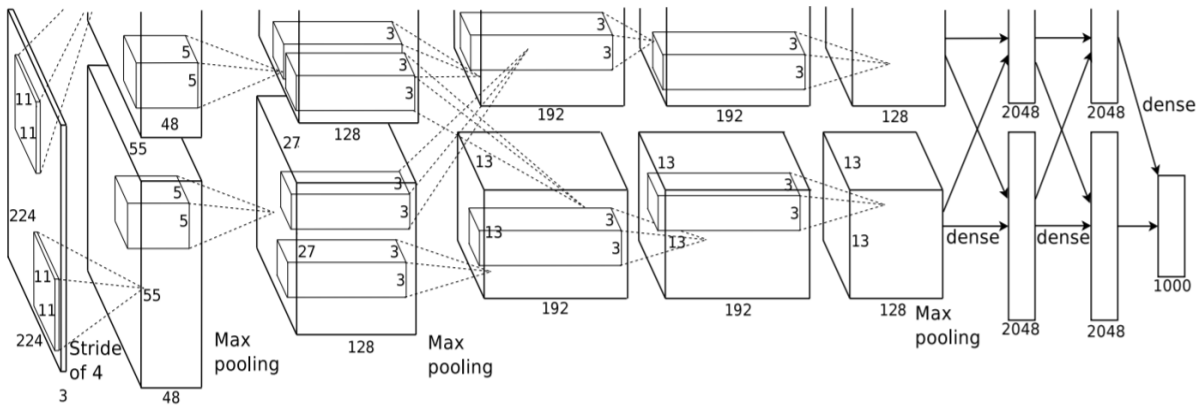


Рисунок 13. Архитектура AlexNet

Все последующие состязания Imagenet тоже выигрывали нейронные сети, как и многие другие соревнования по распознаванию сложных паттернов в большом объеме данных (Рисунок 4) [6]. Из-за недостаточного объема памяти в графическом процессоре автора разделили сеть на две части, что видно на рисунке 13.

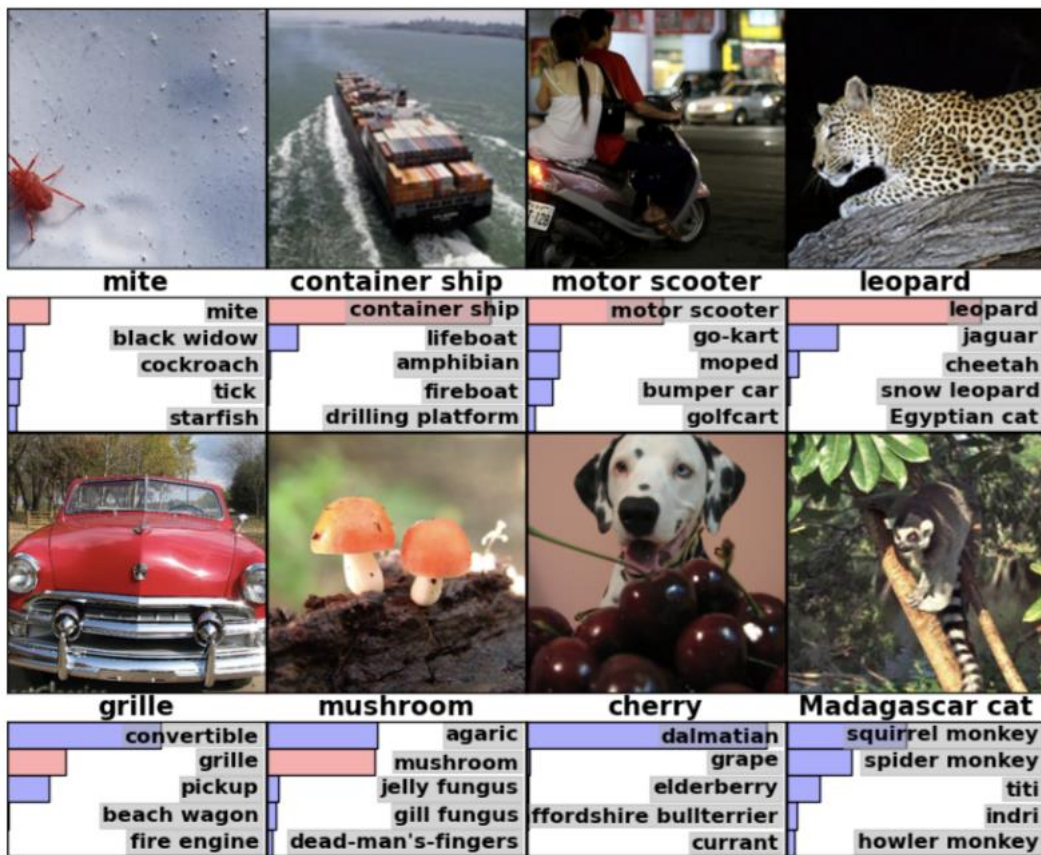


Рисунок 14 Пример работы сети AlexNet из статьи на изображениях ImageNet

Пример анализа системы компьютерного зрения

В этой главе на примере будут проанализированы основные принципы архитектуры систем компьютерного зрения и их возможные преимущества, и недостатки.

Целью данной системы является обработка информации, поступающей с видеокамер в помещениях и в окрестности. Многофункциональная система позволяет пользователю настроить персональный алгоритм действий, включающий следующие этапы:

- Подключение новых видеопотоков к системе через удобный веб-интерфейс. Для этого достаточно лишь знать адрес потока в локальной сети, далее система сама распознает данные о потоке и подстроится под необходимый формат обработки данных. Также возможно удаление и изменение потоков в системе.

- Выбор интересующих модулей распознавания. К системе могут подключаться различные модули, что позволяет гибко работать с разными требованиями пользователей. Например, возможно подключить модуль распознавания лиц или подсчета людей в помещении. Другие примеры модулей и их архитектура будут описаны ниже.

- Составление задачи: в графическом интерфейсе пользователь может указать необходимые действия, которые можно составить из последовательностей модулей. Например, чтобы идентифицировать грузовые автомобили, необходимо подключить модуль, обнаруживающий автомобили, отфильтровать грузовые, и к результатам фильтра применить модуль чтения автомобильного номера. Для одного видеопотока можно формулировать несколько задач. Система будет автоматически балансировать нагрузку.

- Отчет о выполнении: Выполняя задачи, система регулярно отправляет сигнал с информацией о результате в базу данных. Впоследствии в базе данных можно формировать отчеты и проводить аналитику (например, оценить, в какое время помещения наиболее загружены). Поскольку это обычная событийная база данных, пользователи могут устанавливать свои

модули обработки данных. Это не требует экспертного знания в компьютерном зрении, так как результаты представлены в универсальном удобном виде.

Архитектура такой системы может состоять из следующих основных блоков:

- Модуль управления потоками, приводящий видео из разных камер к стандартному виду
- Сервер балансировки нагрузки, разделяющий задачи по модулям и запускающий дополнительные копии модулей при необходимости.
- Сервера модулей
- Веб-интерфейс, через который пользователь управляет системой
- База данных с архивом событий и система управления этой базой данных.

Сервера модулей обмениваются информацией с системой посредством стандартного API и имеют унифицированное API для их автоматического запуска. В остальном они могут быть устроены по-разному: например, базовая сеть распознавания объектов — это сверточная нейронная сеть на базе архитектура Inception, а модуль распознавания лиц — сама по себе многомодульная система с нейронной сетью для выделения векторов, базой лиц, сохраненных пользователем и веб-интерфейсом для управления ей.

У такой архитектуры системы есть ряд преимуществ:

- Модульность позволяет разработчику легко добавлять и удалять новые алгоритмы и архитектуры компьютерного зрения, а пользователю включать или выключать их по необходимости.
- Разделение работы с данными и работы с потоками через стандартный перекодированием видео в универсальный формат с помощью ffmpeg позволяет легко добавлять видео разного вида.
- Гибкость работы с задачами и событиями позволяет разным пользователям самим настраивать систему под себя, не беспокоясь о технической совместимости и оптимизации работы.

При этом можно выделить следующие недостатки, над исправлением которых ведется работа:

- Сложность разграничения прав доступа разных пользователей из-за отсутствия внутри системы универсального механизма аутентификации
- Из-за большого количества разделенных модулей время на передачу информации между ними становится проблемой при работе с системами реального времени. Для оптимизации этого необходимо более «плотное» размещение модулей в сети и уменьшение количества посредников между источником и приемником информации.

Стоит также отметить, что в связи с бурным развитием нейросетевых технологий, модули необходимо часто дорабатывать и обновлять, чтобы их производительность оставалась на высоком уровне, и система продолжала быть актуальной. Такие обновления могут быть незаметны для пользователя, но занимают значительную часть ресурсов, выделенных на поддержание системы.

Рынок компьютерного зрения в мире и в России

Диапазон применения систем компьютерного зрения стремительно расширяется. Так, одна из ведущих групп в области IT-аналитики, компания Tractica, только за период с 2014 по 2016 г.г. отчиталась о развитии нескольких новых областей применения этих технологий (рис. 15) [7]. Например, розничная торговля (retail): в последние годы стало возможно производить «умное» видеонаблюдение в супермаркетах, позволяющее проводить аналитику о поведении клиентов в помещении, их предпочтениях и склонностях. Согласно прогнозам, рост будет только ускоряться.

Computer Vision Revenue by Application Market, World Markets: 2015-2022

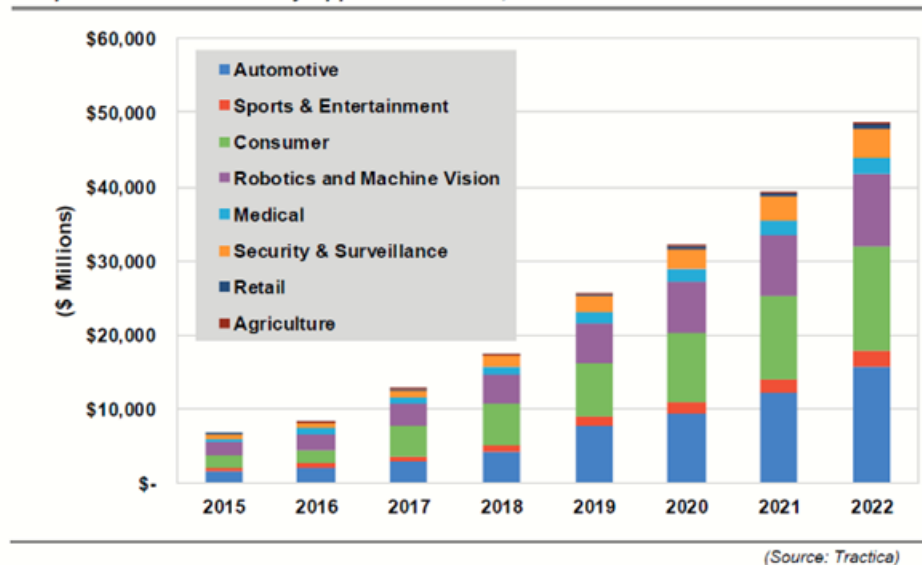


Рис. 15. Отчет Tractica о мировом рынке компьютерного зрения

При этом надо отметить, что разные компании по-разному оценивают объем рынка. Это во многом связано с нестрогостью определения «компьютерное зрение», особенно в случаях, когда оно пересекается с предметной областью машинного обучения. Некоторые компании могут объединять их в категорию «искусственный интеллект», а само компьютерное зрение считать сопутствующим направлением, не выделяя его отдельно. Так, аналитическая компания McKinsey в своем отчете об искусственном интеллекте подчеркивает неточность разделения областей и затрудненность объективной оценки рынка (рис. 16). Тем не менее, агентство оценивает ежегодный объем инвестиций в компьютерное зрение в 2.5 — 3.5 миллиарда долларов.

С другой стороны, компания Market Research Future пишет о цифре в 12 миллиардов долларов и предсказывает бурный рост аж до 50 миллиардов к 2023 году.



Рис. 16. Отчет McKinsey о мировом рынке компьютерного зрения

Оценка объема рынка компьютерного зрения в России была проведена компанией TAdviser в 2018 году с использованием многофакторной сравнительной оценки (рис. 17). В результате анализа тенденций развития рынка компьютерного зрения в России данным методом и сравнения их с мировыми трендами оценка объема рынка в России в 2018 составила около 8 миллиардов рублей. При этом прогнозируется рост до 37 миллиардов рублей к 2023 году. Основными факторами, способствующими росту рынка в России, были названы национальная программа цифровой экономики, рынок робототехники, розничная торговля и интеллектуальные системы «Умный город» и «Безопасный город». Основными факторами, препятствующими распространению технологий в России, были названы несовершенное законодательство и проблемы при разработке бизнес-моделей.



Рис. 17. Отчет TAdviser о рынке компьютерного зрения в России

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.: Бином. 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>.
7. «Компьютерное зрение: технологии, рынок, перспективы», отчет TAdviser, 2018.