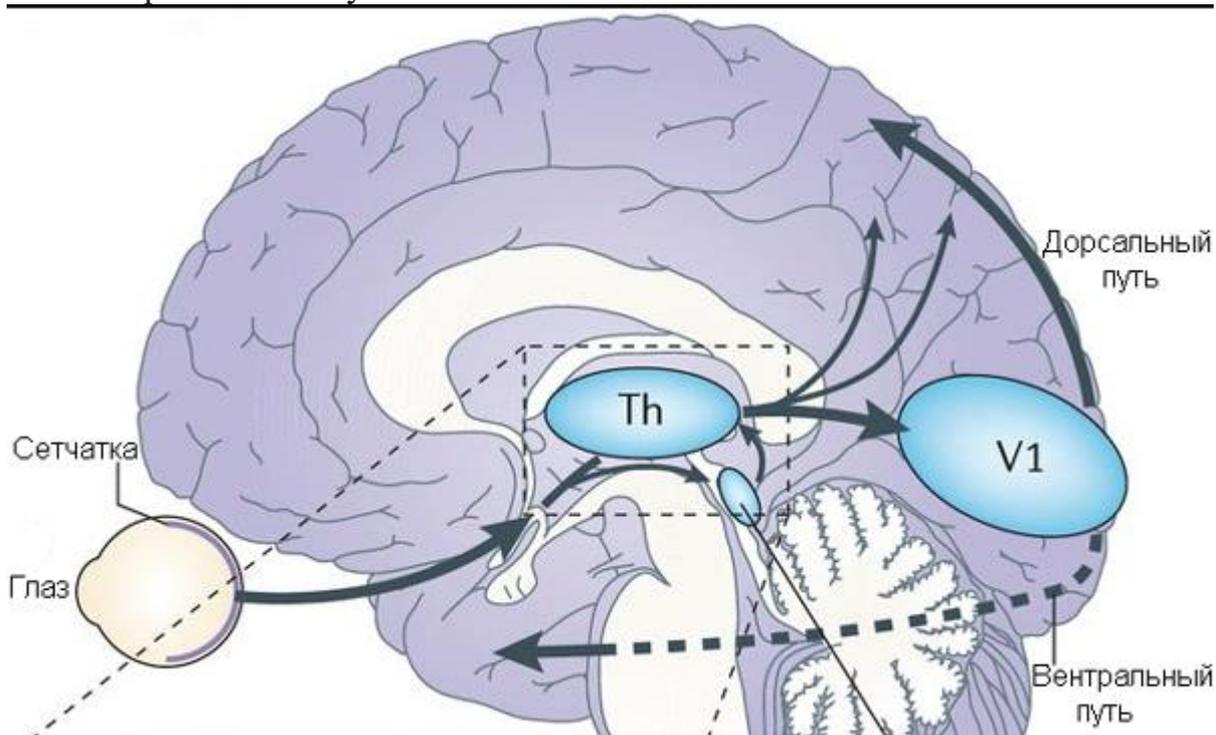


Применение нейронных сетей в распознавании изображений

Прототипом для создания нейронных сетей послужили, как это ни странно, биологические нейронные сети. Две трети всей сенсорной информации, которая к нам попадает, приходит с зрительных органов восприятия. Более одной трети поверхности нашего мозга заняты двумя самыми главными зрительными зонами — дорсальный зрительный путь и вентральный зрительный путь.



Дорсальный зрительный путь начинается в первичной зрительной зоне, в нашем темечке и продолжается вверх, в то время как вентральный путь начинается на нашем затылке и заканчивается примерно за ушами. Все важное распознавание образов, которое у нас происходит, все смысловое, то что мы осознаём, проходит именно там же, за ушами.

Все области, которые используются в нейронных сетях для распознавания образов, пришли к нам именно из вентрального зрительного пути, где каждая маленькая зона отвечает за свою строго определенную функцию.

Изображение попадает к нам из сетчатки глаза, проходит череду зрительных зон и заканчивается в височной зоне.

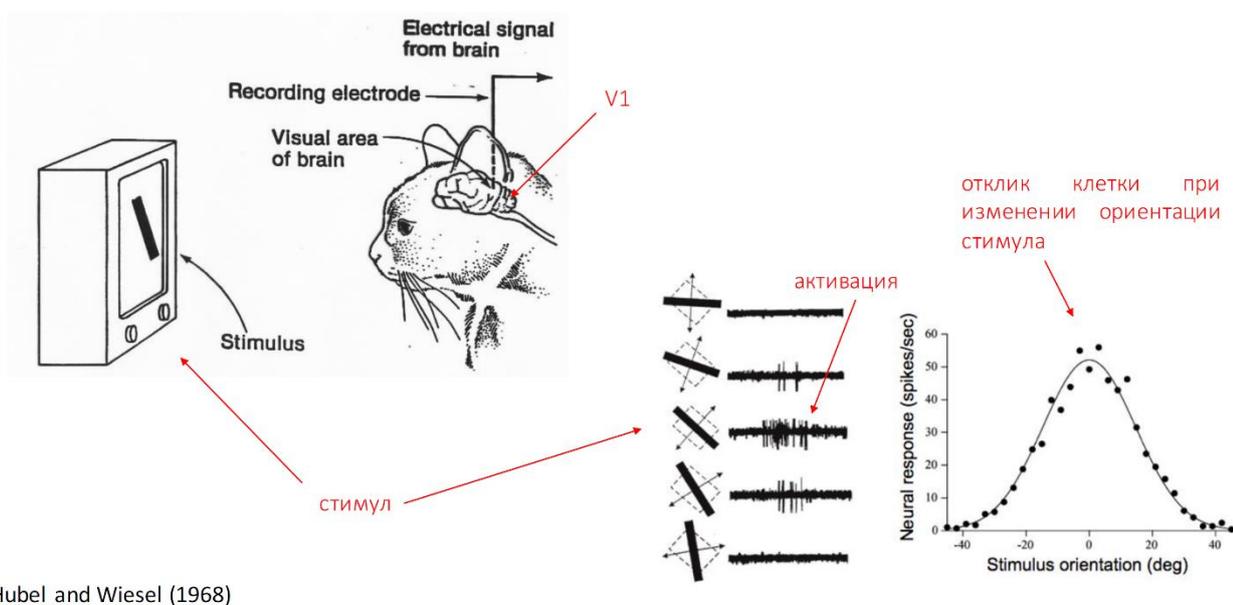
В далекие 60-е годы прошлого века, когда только начиналось изучение зрительных зон мозга, первые эксперименты проводились на животных.

Исследовали мозг с помощью электродов, вживлённых в различные зрительные зоны.

Первая зрительная зона была исследована Дэвидом Хьюбелем и Торстенем Визелем в 1962 году. Они проводили эксперименты на кошках. Кошкам показывались различные движущиеся объекты. На что реагировали клетки мозга, то и было тем стимулом, которое распознавало животное. Даже сейчас многие эксперименты проводятся подобными способами. Но тем не менее, это самый эффективный способ узнать, что делает каждая мельчайшая клеточка в нашем мозгу.

Таким же способом были открыты еще многие важные свойства зрительных зон, которые мы используем в глубинном обучении (англ. Deep learning) сейчас. Одно из важнейших свойств — это увеличение рецептивных полей наших клеток по мере продвижения от первичных зрительных зон к височным долям, то есть более поздним зрительным зонам. Рецептивное поле — это та часть изображения, которую обрабатывает каждая клеточка нашего мозга. У каждой клетки своё рецептивное поле. Это же свойство сохраняется и в нейронных сетях.

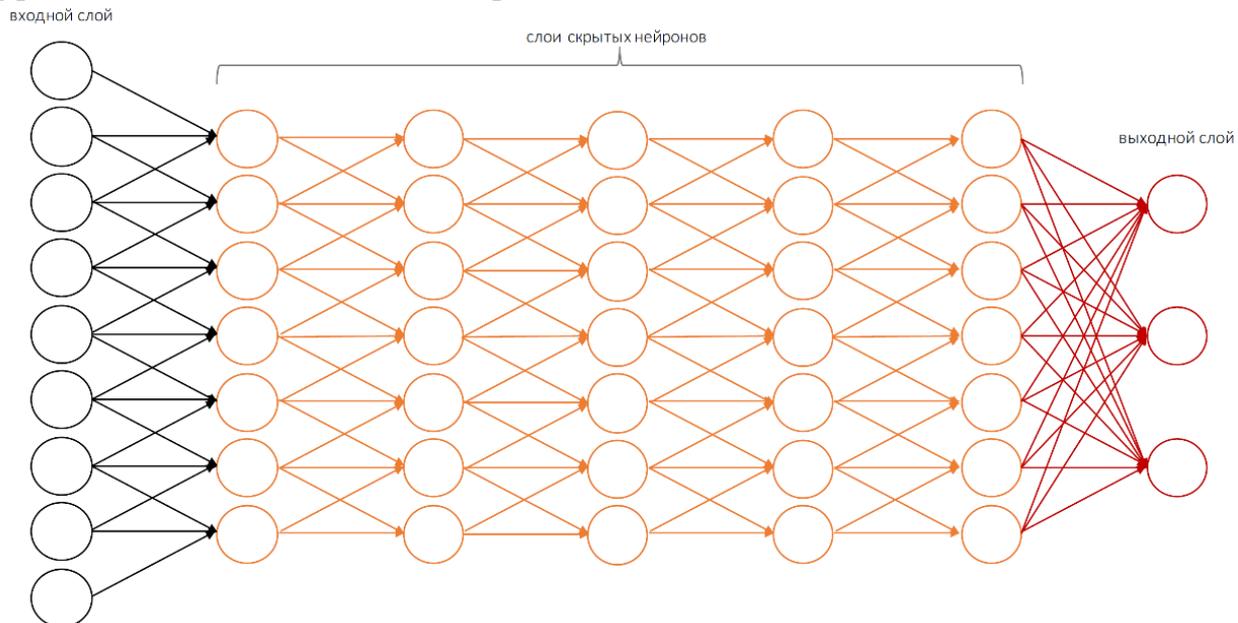
Также с возрастанием рецептивных полей увеличиваются сложные стимулы, которые обычно распознают нейронные сети.



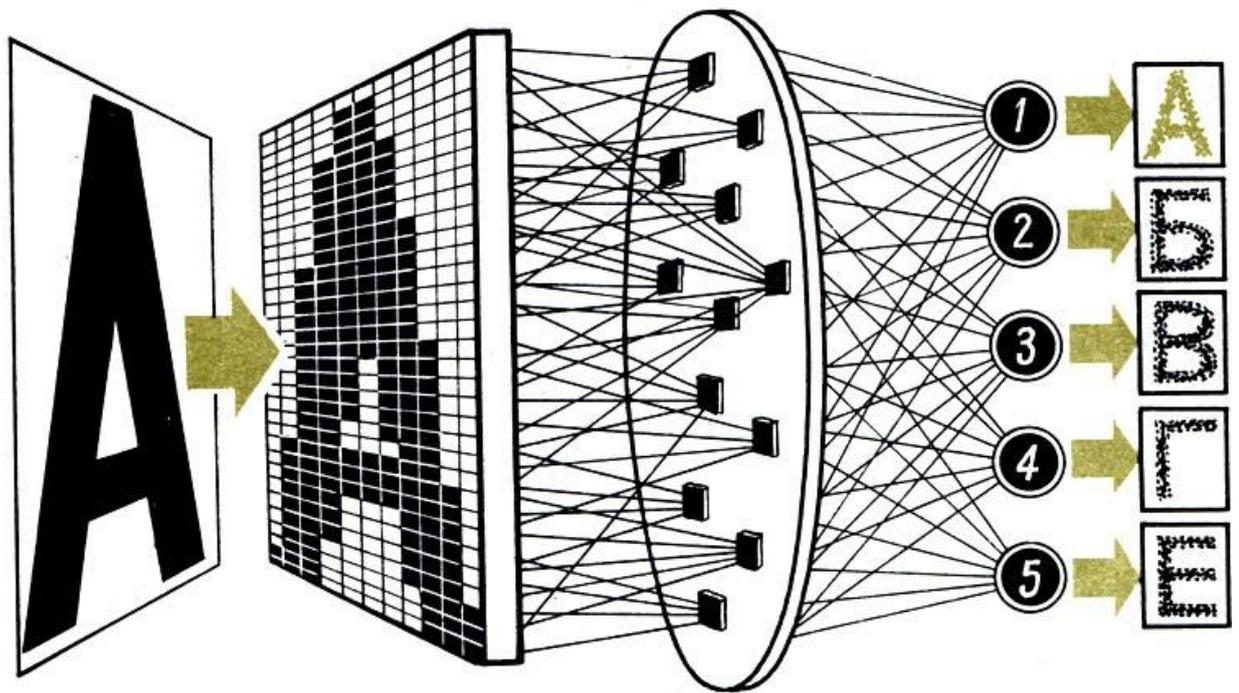
Компьютерное зрение

Но сначала немного о простейшем перцептроне. Он также образован по образу и подобию нашего мозга. Простейший элемент напоминающий клетку мозга — нейрон. Имеет входные элементы, которые по умолчанию располагаются слева направо, изредка снизу-вверх. Слева это входные части нейрона, справа выходные части нейрона.

Простейший перцептрон способен выполнять только самые простые операции. Для того, чтобы выполнять более сложные вычисления, нам нужна структура с большим количеством скрытых слоёв.



В случае компьютерного зрения нам нужно еще больше скрытых слоёв. И только тогда система будет осмысленно распознавать то, что она видит.



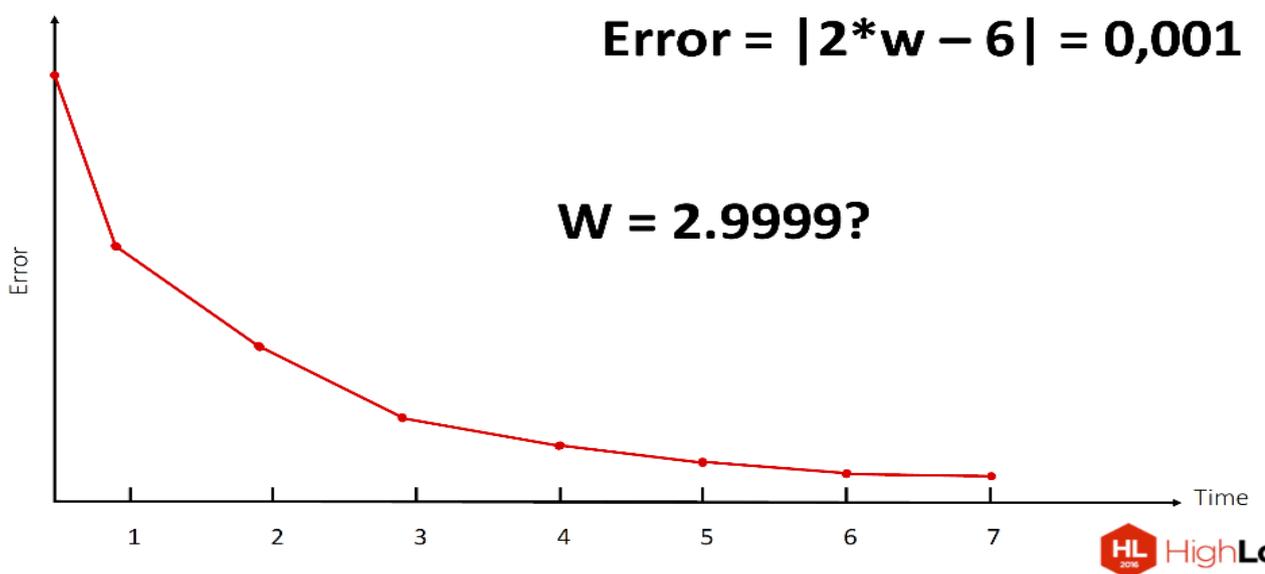
Итак, что происходит при распознавании изображения, опишем на примере лиц.

Для нас посмотреть на картинку и сказать, что на ней изображено именно лицо статуи, достаточно просто. Однако до 2010 года для компьютерного зрения это было невероятно сложной задачей. Нам нужно это было сделать каким-то геометрическим способом, описать объект, описать взаимосвязи объекта, как могут эти части относиться к друг другу, потом найти это изображение на объекте, сравнить их и получить, что мы распознали плохо.

Сейчас это происходит не так. Мы разбиваем наше изображение либо на пиксели, либо на некие патчи: 2×2 , 3×3 , 5×5 , 11×11 пикселей — как удобно создателям системы, в которой они служат входным слоем в нейронную сеть.

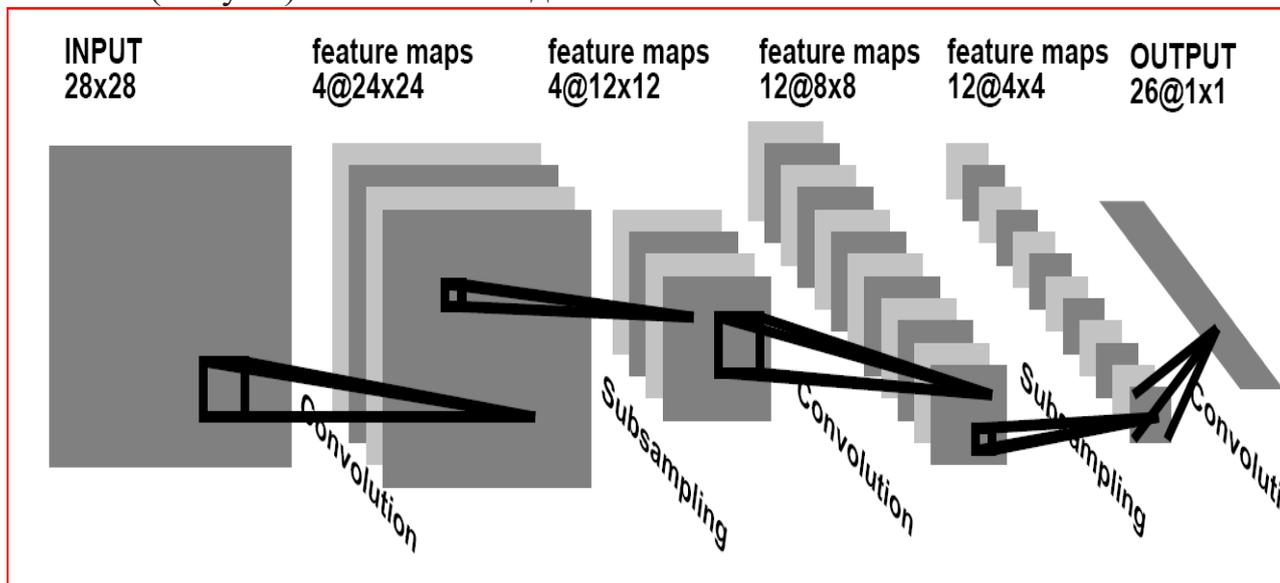
Сигналы с этих входных слоёв передаются от слоя к слою с помощью синапсов, каждый из слоёв имеет свои определенные коэффициенты. Итак, мы передаём от слоя к слою, от слоя к слою, пока мы не получим, что мы распознали лицо.

Условно все эти части можно разделить на три класса, мы их обозначим X , W и Y , где X — это наше входное изображение, Y — это набор лейблов, и нам нужно получить наши веса. Как мы вычислим W ? При наличии нашего X и Y это, кажется, просто. Однако то, что обозначено звездочкой, очень сложная нелинейная операция, которая, к сожалению, не имеет обратной. Даже имея 2 заданных компоненты уравнения, очень сложно ее вычислить. Поэтому нам нужно постепенно, методом проб и ошибок, подбором веса W сделать так, чтобы ошибка максимально уменьшилась, желательно, чтобы стала равной нулю. Этот процесс происходит итеративно, мы постоянно уменьшаем, пока не находим то значение веса W , которое нас достаточно устроит.



Сверточная нейронная сеть (англ. convolutional neural networks)

Идея сверточных нейронных сетей заключается в чередовании сверточных слоев (C-layers), субдискретизирующих слоев (S-layers) и наличии полносвязных (F-layers) слоев на выходе.



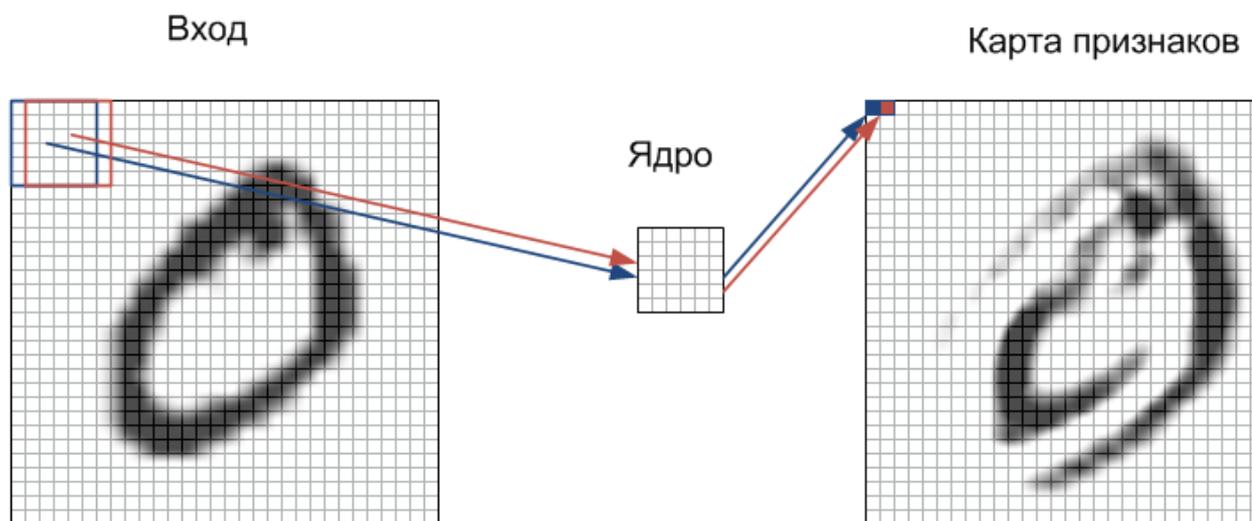
Такая архитектура включает в себе 3 основных парадигмы:

1. *Локальное восприятие.*
2. *Разделяемые веса.*
3. *Субдискретизация.*

Локальное восприятие подразумевает, что на вход одного нейрона подается не все изображение (или выходы предыдущего слоя), а лишь некоторая его область. Такой подход позволил сохранять топологию изображения от слоя к слою.

Концепция *разделяемых весов* предполагает, что для большого количества связей используется очень небольшой набор весов. Т.е. если у нас имеется на входе изображение размерами 32×32 пикселя, то каждый из нейронов следующего слоя примет на вход только небольшой участок этого изображения размером, к примеру, 5×5 , причем каждый из фрагментов будет обработан одним и тем же набором. Важно понимать, что самих наборов весов может быть много, но каждый из них будет применен ко **всему** изображению. Такие наборы часто называют ядрами (kernels). Нетрудно посчитать, что даже для 10 ядер размером 5×5 для входного изображения размерами 32×32 количество связей окажется равным примерно 256000 (сравниваем с 10 млн.), а количество настраиваемых параметров **всего 250!**

А как же это скажется на качестве распознавания? Как ни странно, в лучшую сторону. Дело в том, что такое искусственно введенное ограничение на веса улучшает обобщающие свойства сети (generalization), что в итоге позитивно сказывается на способности сети находить инварианты в изображении и реагировать главным образом на них, не обращая внимания на прочий шум. Можно посмотреть на этот подход немного с другой стороны. Те, кто занимался классикой распознавания изображений и знает как это работает на практике (например в военной технике) знают, что большинство таких систем строятся на основе двумерных фильтров. Фильтр представляет собой матрицу коэффициентов, обычно заданную вручную. Эта матрица применяется к изображению с помощью математической операции, называемой сверткой. Суть этой операции в том, что каждый фрагмент изображения умножается на матрицу (ядро) свертки поэлементно и результат суммируется и записывается в аналогичную позицию выходного изображения. Основное свойство таких фильтров заключается в том, что значение их выхода тем больше чем больше фрагмент изображения похож на сам фильтр. Таким образом, изображение, свернутое с неким ядром, даст нам другое изображение, каждый пиксел которого будет означать степень похожести фрагмента изображения на фильтр. Иными словами, это будет карта признаков.



Каждый фрагмент изображения поэлементно умножается на небольшую матрицу весов (ядро), результат суммируется. Эта сумма является пикселем выходного изображения, которое называется картой признаков. Следует сказать, что в идеале не разные фрагменты проходят последовательно через ядро, а параллельно все изображение проходит через идентичные ядра. Кроме того, количество ядер (наборов весов) определяется разработчиком и зависит от того какое количество признаков необходимо выделить. Еще одна особенность сверточного слоя в том, что он немного уменьшает изображение за счет краевых эффектов.

Суть *субдискретизации* и S-слоев заключается в уменьшении пространственной размерности изображения. Т.е. входное изображение грубо (усреднением) уменьшается в заданное количество раз. Чаще всего в 2 раза, хотя может быть и не равномерное изменение, например, 2 по вертикали и 3 по горизонтали. Субдискретизация нужна для обеспечения инвариантности к масштабу.

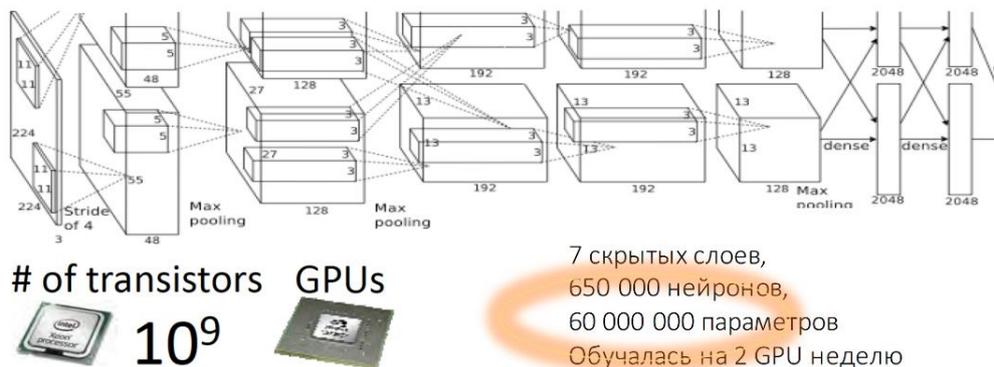
Чередование слоев позволяет составлять карты признаков из карт признаков, что на практике означает способность распознавания сложных иерархий признаков.

Обычно после прохождения нескольких слоев карта признаков вырождается в вектор или даже скаляр, но таких карт признаков становится сотни. В таком виде они подаются на один-два слоя полносвязной сети. Выходной слой такой сети может иметь различные функции активации. В простейшем случае это может быть тангенциальная функция, также успешно используются радиальные базисные функции.

Современные базы данных(изображений).

Перед вами первая сеть, которая победила на международном соревновании ImageNet в 2012 году. Это так называемый AlexNet.

2012
Krizhevsky
et al.

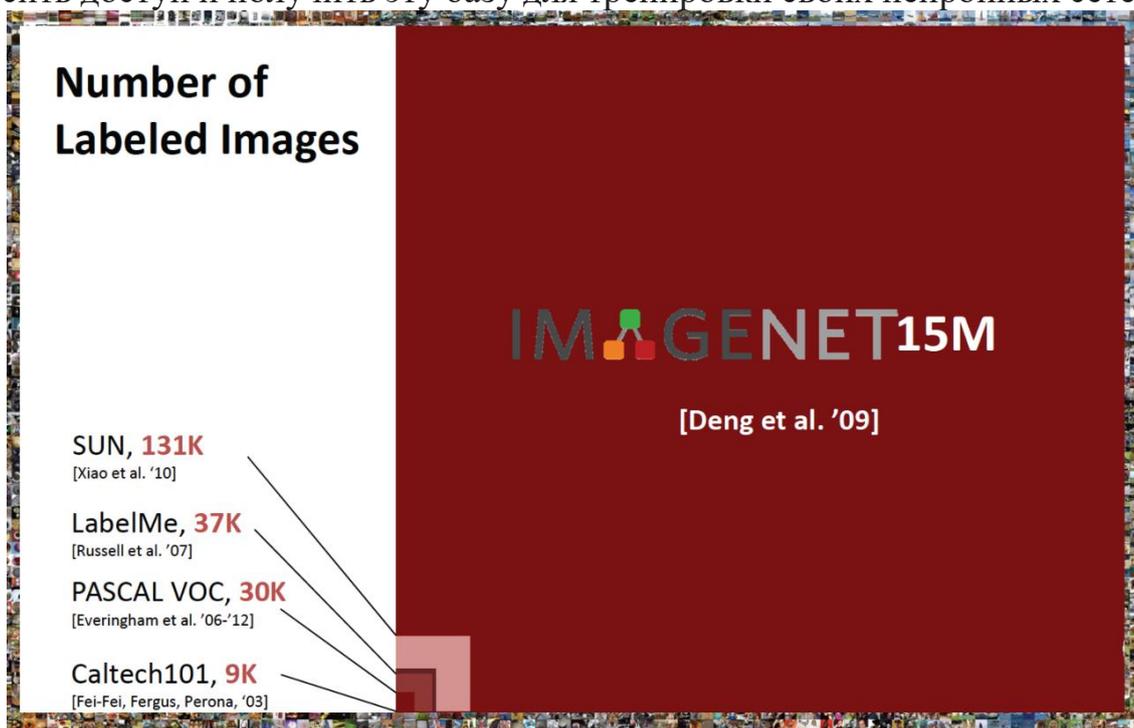


Несмотря на то, что эта сеть достаточно мелкая (в ней всего 7 скрытых слоёв), она содержит 650 тысяч нейронов с 60 миллионами параметров. Для того, чтобы итеративно научиться находить нужные веса, нам нужно очень много примеров.

Нейронная сеть учится на примере картинки и лейбла. Как нас в детстве учат «это кошка, а это собака», так же нейронные сети обучаются на большом количестве картинок. Но дело в том, что до 2010 не существовало достаточно большого объема данных, который способен был бы научить такое количество параметров распознавать изображения.

Самые большие базы данных, которые существовали до этого времени: PASCAL VOC, в который было всего 20 категорий объектов, и Caltech 101, который был разработан в California Institute of Technology. Тем же, кто не сумел найти свои объекты ни в одной из этих баз данных, приходилось стоять свои базы данных.

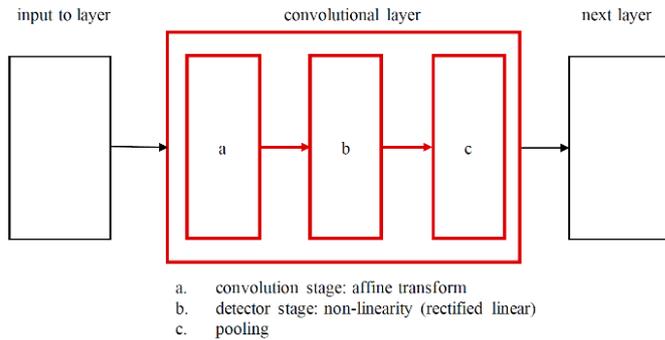
Однако, в 2010 году появилась база ImageNet, в которой было 15 миллионов изображений, разделённые на 22 тысячи категорий. Это решило проблему обучения нейронных сетей. Сейчас все желающие, у кого есть какой-либо академический адрес, могут спокойно зайти на сайт базы, запросить доступ и получить эту базу для тренировки своих нейронных сетей.



Архитектура глубокой нейронной сети

Условно ее можно разделить на 2 части: те, которые учатся, и те, которые не учатся.

Архитектура

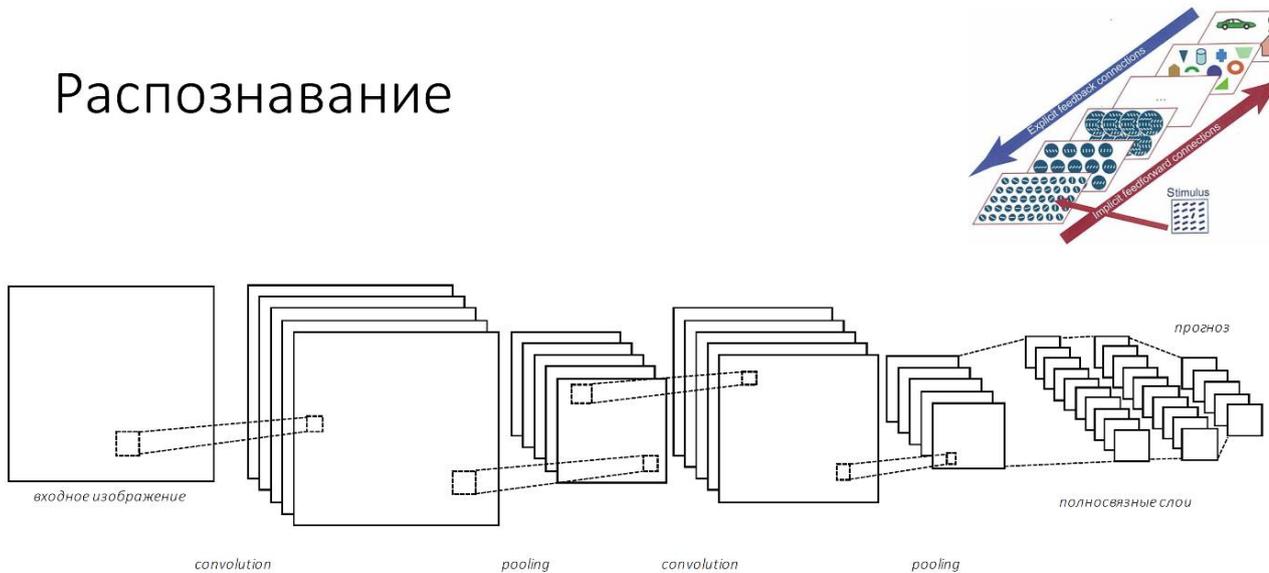


(Goodfellow, 2016)

Чёрным обозначены те части, которые не учатся, все остальные слои способны обучаться. Существует множество определений того, что находится внутри каждого сверточного слоя. Одно из принятых обозначений — один слой с тремя компонентами разделяют на convolution stage(свертка), detector stage(определение) и pooling stage(объединение).



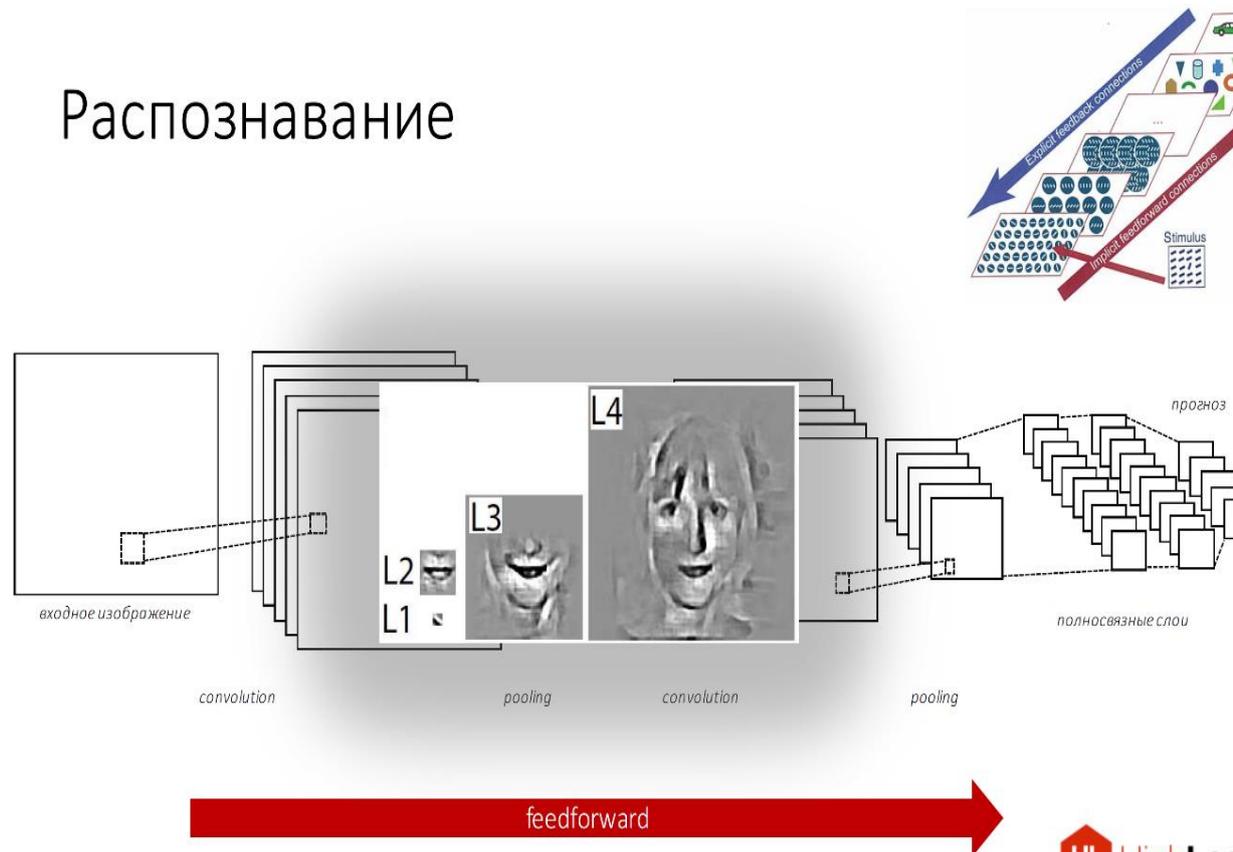
Распознавание



Итак, входное изображение попадает в сеть слоёв, которые можно назвать фильтрами разного размера и разной сложности элементов, которые они распознают. Эти фильтры составляют некий свой индекс или набор признаков, который потом попадает в классификатор. Обычно это либо SVM, либо MLP — многослойный перцептрон, кому что удобно.

По образу и подобию с биологической нейронной сетью объекты распознаются разной сложности.

Распознавание

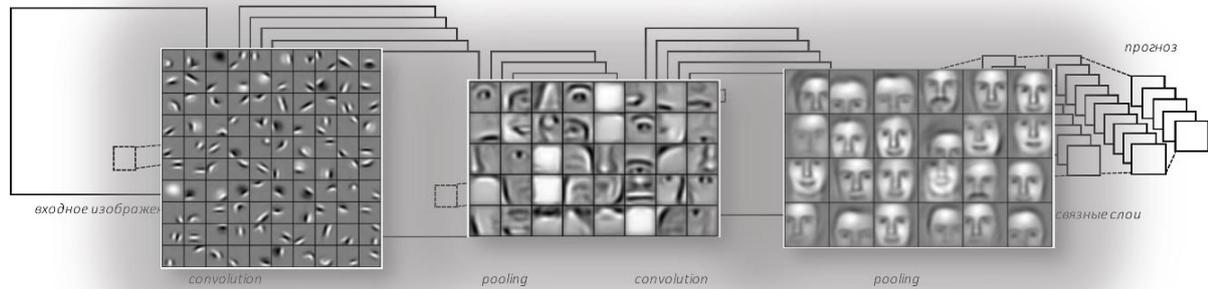
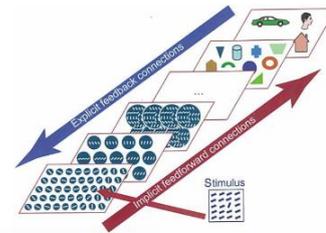


M.Zieler, CVSS 2015

HL High Load 2016

Если рассмотреть на примере распознавания лиц, то у нас рецептивное поле первого слоя будет маленьким, потом чуть побольше, побольше, и так до тех пор, пока наконец мы не сможем распознавать уже лицо целиком.

Распознавание



H.Lee et al., 2011



С точки зрения того, что находится у нас внутри фильтров, сначала будут наклонные палочки плюс немного цвета, затем части лиц, а потом уже целиком лица будут распознаваться каждой клеточкой слоя.

Этапы распознавания картинок в общих чертах выглядят примерно так: Предварительная фильтрация изображения — здесь распознаватель пытается найти наиболее выделяющиеся в изображении черты, позволяющие выявить общие закономерности его строения, цветовых характеристик и др. Сравнить этот процесс можно с написанием портрета, при котором художник старается отыскать черты лица человека, придающие ему индивидуальность.

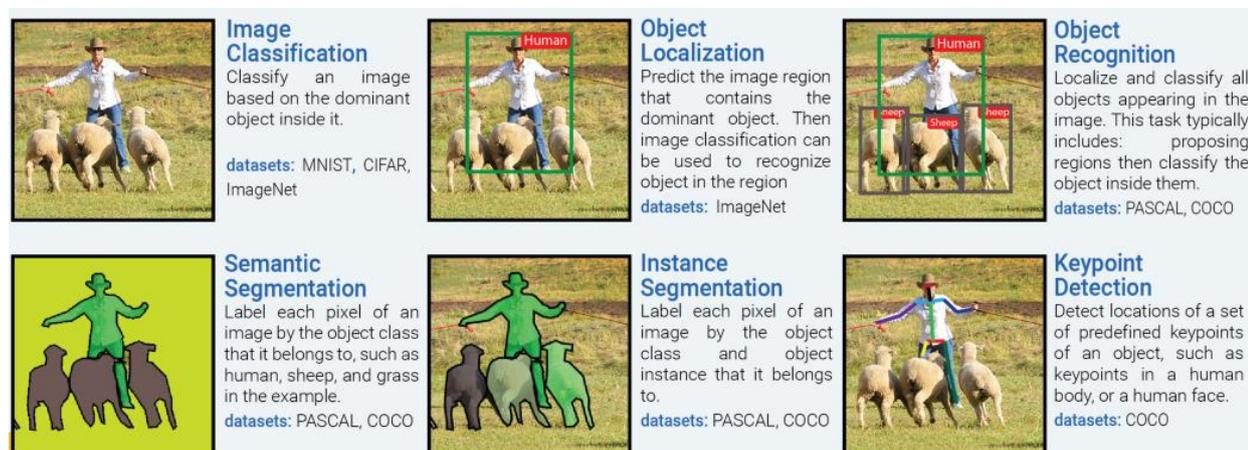
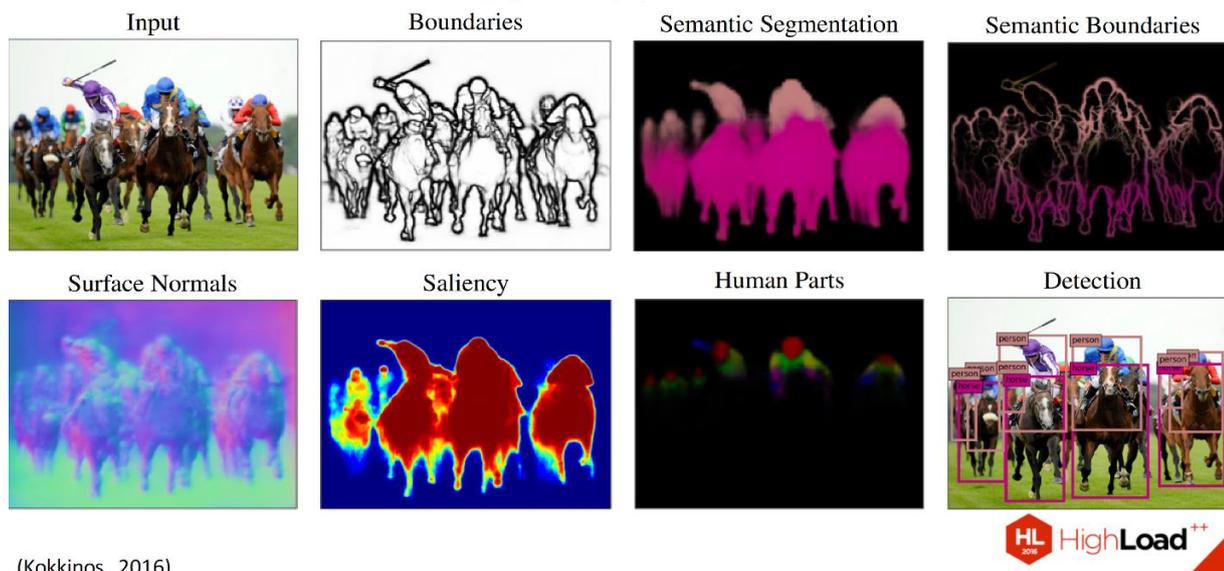


Бинаризация по порогу — один из простейших, на сегодняшний день, методов фильтрации изображений. Работает он по принципу перевода цветного изображения в бинарное, то есть, содержащее два цвета — белый и черный. В зависимости от порога яркости пикселя ИИ причисляет его к соответствующему значению (0 — белый, 1 — черный).

Классические задачи сверточных нейронных сетей

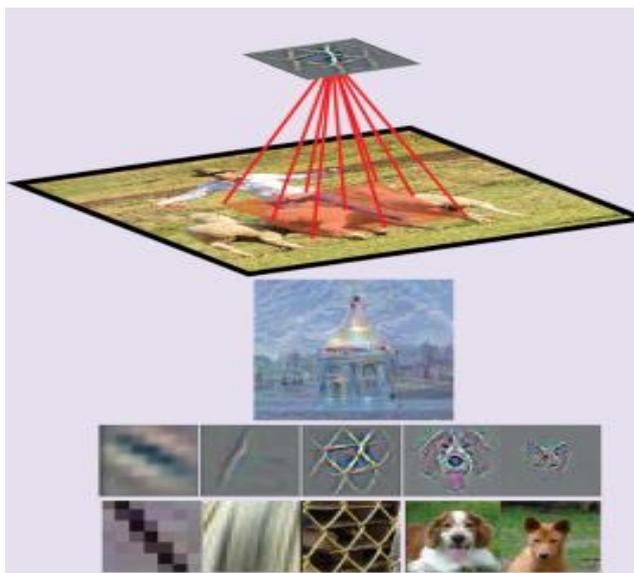
Их на самом деле не так много, они относятся к трём классам. Среди них — такие задачи, как идентификация объекта, семантическая сегментация, распознавание лиц, распознавание частей тела человека, семантическое определение границ, выделение объектов внимания на изображении и выделение нормалей к поверхности. Их условно можно разделить на 3 уровня: от самых низкоуровневых задач до самых высокоуровневых задач.

Классические задачи для CNN



На примере этого изображения рассмотрим, что делает каждая из задач.

- Определение границ — это самая низкоуровневая задача, для которой уже классически применяются сверточные нейронные сети.
- Определение вектора к нормали позволяет нам реконструировать трёхмерное изображение из двухмерного.
- Saliency, определение объектов внимания — это то, на что обратил бы внимание человек при рассмотрении этой картинки.
- Семантическая сегментация позволяет разделить объекты на классы по их структуре, ничего не зная об этих объектах, то есть еще до их распознавания.
- Семантическое выделение границ — это выделение границ, разбитых на классы.
- Выделение частей тела человека.
- И самая высокоуровневая задача — распознавание самих объектов, которое мы сейчас рассмотрим на примере распознавания лиц.



Распознавание лиц

На первом этапе — используется face detector (определитель лиц) по изображению для того, чтобы найти лицо. Далее мы нормализуем, центрируем лицо и запускаем его на обработку в нейронную сеть. После чего получаем набор или вектор признаков однозначно описывающий черты этого лица.

Поиск по базе данных



Система распознавания 3D Guard

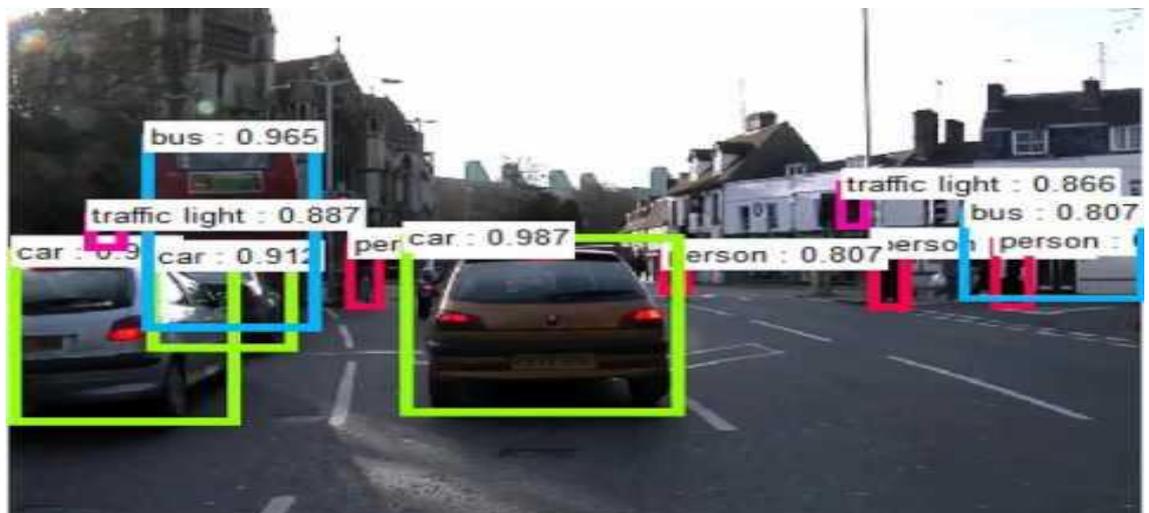
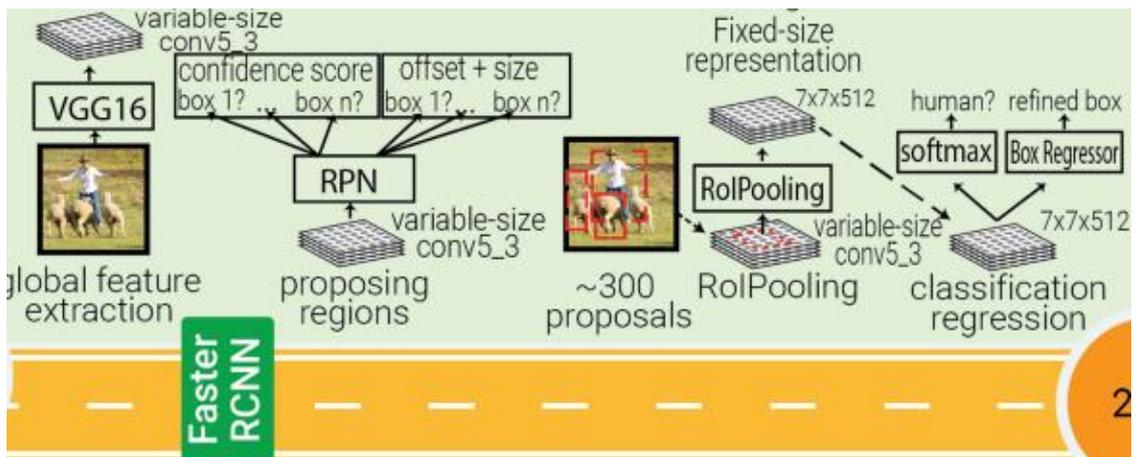
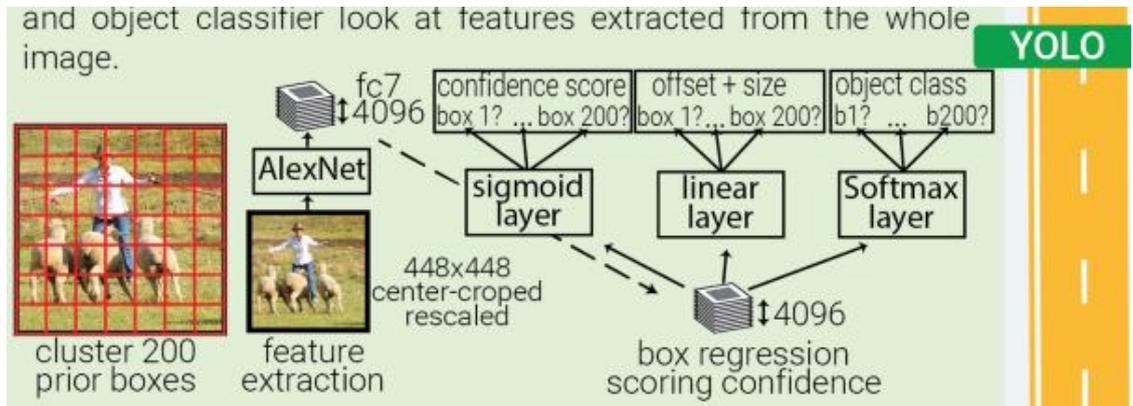


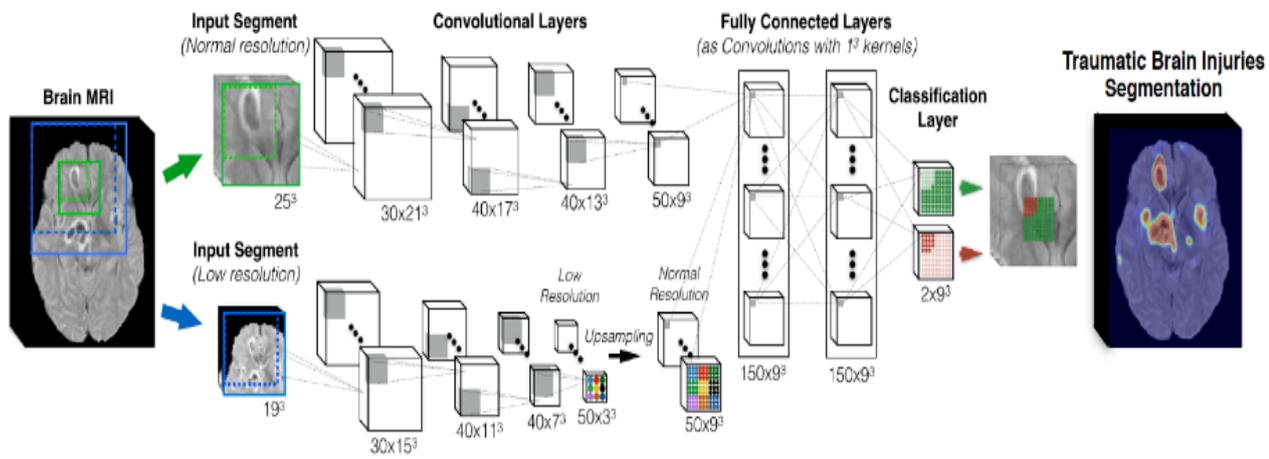
Рис.2.7. Принцип работы системы распознавания

Затем можно этот вектор признаков сравнивается со всеми векторами признаков, которые хранятся у нас в базе данных, и получить отсылку на конкретного человека, на его имя, на его профиль — всё, что может храниться в базе данных.

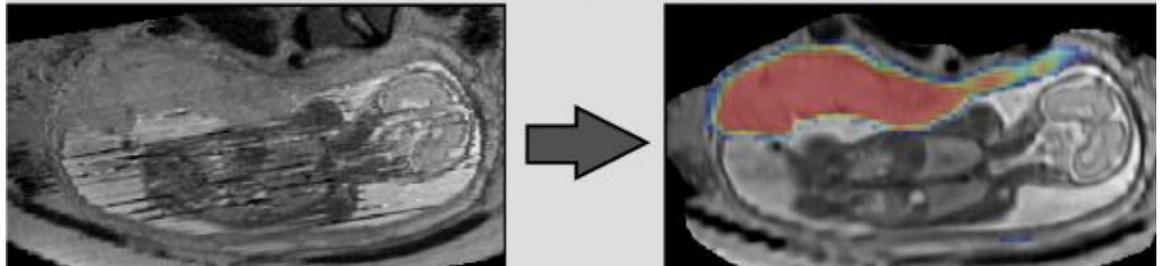
Новейшие технологии

На данный момент Faster-RCNN, R-FCN, Mask-RCNN, YOLO, SSD – наиболее перспективные вариации на архитектуры сети распознавания изображений. Они уже используются в самых разнообразных задачах: от поиска котов на картинке до определения мини-объектов на спутниковых снимках и использования в медицине.

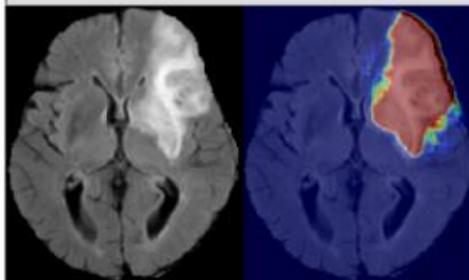




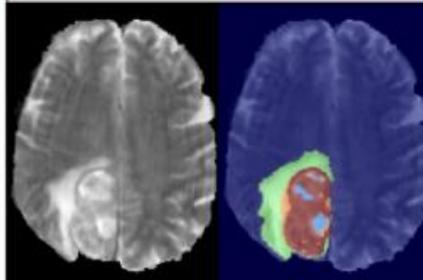
Placenta Segmentation from Motion Corrupted Data (Alansary et al, MICCAI 2016)



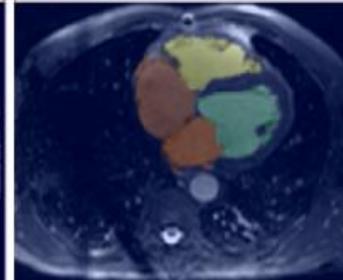
Ischemic Stroke Lesions



Brain Tumors



Cardiac



Заключение

Даны определения нейросетям и их видам, дан обзор различных нейросетевых методов распознавания изображений. Рассмотрены достоинства и недостатки этих методов при распознавании объектов. Указаны проблемы при распознавании объектов. Выделены перспективные направления в распознавании трёхмерных объектов. Отмечены возможности применения нейросетевых методов для задачи распознавания человека по изображению лица.

Список использованной литературы и источников.

1. Yann LeCun, J. S. Denker, S. Solla, R. E. Howard and L. D. Jackel: Optimal Brain Damage, in Touretzky, David (Eds), Advances in Neural Information Processing Systems 2 (NIPS*89), Morgan Kaufman, Denver, CO, 1990
2. Y. LeCun and Y. Bengio: Convolutional Networks for Images, Speech, and Time-Series, in Arbib, M. A. (Eds), The Handbook of Brain Theory and Neural Networks, MIT Press, 1995
3. Y. LeCun, L. Bottou, G. Orr and K. Muller: Efficient BackProp, in Orr, G. and Muller K. (Eds), Neural Networks: Tricks of the trade, Springer, 1998
4. Ranzato Marc'Aurelio, Christopher Poultney, Sumit Chopra and Yann LeCun: Efficient Learning of Sparse Representations with an Energy-Based Model, in J. Platt et al. (Eds), Advances in Neural Information Processing Systems (NIPS 2006), MIT Press, 2006
5. Наталия Ефремова, лекция с конференции Highload++, 2017
6. Adam Geitgey, medium.com, Machine Learning is Fun! Part 3, Deep Learning and Convolutional Neural Networks

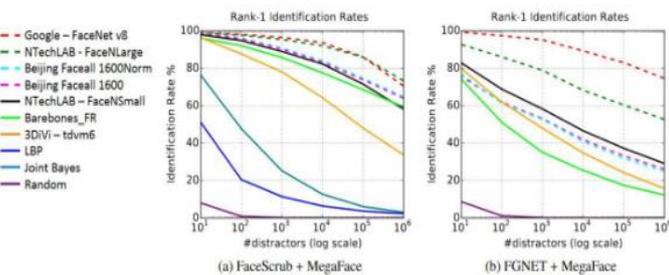
Нейронные сети

Инвестирование в ИИ



Распознавание изображений с помощью нейронных сетей

В 2016 году было проведено соревнование среди алгоритмов распознавания лиц. Организаторы начали с существующих помеченных наборов изображений людей - одного набора, состоящего из знаменитостей разных ракурсов, другого - из обычных людей с разным возрастом. Они добавили шум к этому сигналу в виде «дистракторов». Количество дистракторов варьировалось от 10 до миллиона. Победителями стали Google - Facenet и российская команда NTechLAB



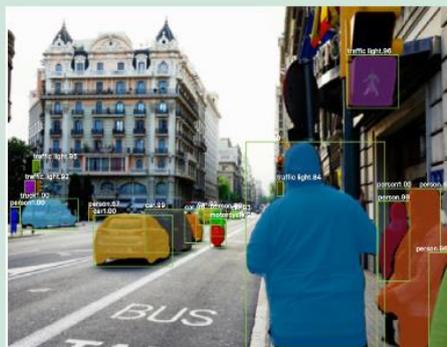
Страны и технологии

Развитие нейронных сетей как в распознавании изображений, так и в других сферах невозможно без должного уровня компьютеризации страны. Несмотря на то, что российские разработчики всегда на лидирующих позициях, общий уровень компьютеризации и технологического развития России крайне мал.



Современные архитектуры сети распознавания изображений

Сфера информационных технологий стремительно развивается, методы совершенствуются ежедневно. В настоящее время самыми перспективными архитектурами распознавания изображений являются RPN и Faster RCNN, обладающие быстрым и точным поиском гипотез (Гипотеза - это любой объект, отличный от фона).

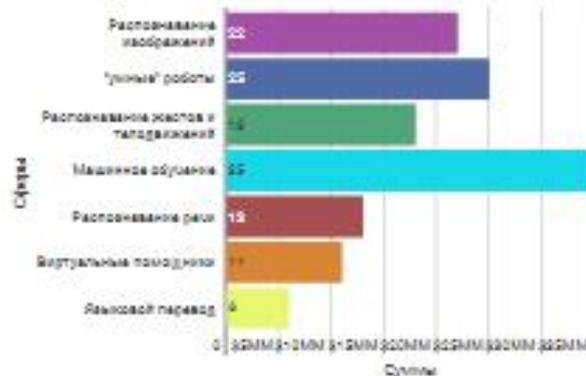


Заключение

Нейронные сети - крайне интересное направление в сфере информационных технологий. В настоящее время их понемногу внедряют в различные сферы жизни, а их незаменимость - лишь вопрос времени. Вскоре, с их помощью, люди разных национальностей смогут понимать друг друга, не зная языка, врачи смогут определять наличие опухолей у пациента, а геологи и работники нефтегазовой сферы определять нужные месторождения, экономя огромные деньги.

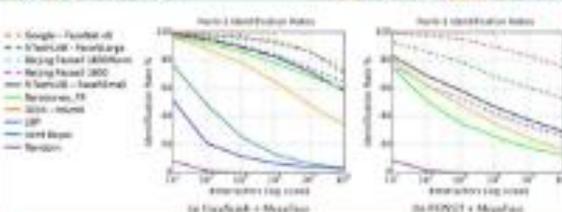
Нейронные сети

Инвестирование в ИИ



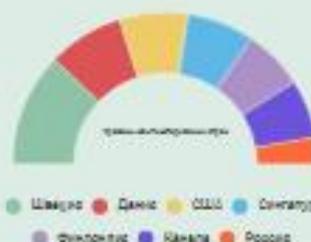
Распознавание изображений с помощью нейронных сетей

В 2015 году была проведена крупнейшая международная соревнование по распознаванию изображений. Организаторы выбрали в качестве задачи пометить тысячи изображений людей - указать пол, возраст, цвет волос и т.д. Победителем оказался DeepMind, команда которого использовала архитектуру AlexNet.



Страны и технологии

Развитие нейронных сетей как в распознавании изображений, так и в других сферах невозможно без должного уровня компьютеризации страны. Несмотря на то, что российская разработка всегда на лидирующей позиции, общий уровень компьютеризации и технологического развития России крайне низок.



Современные архитектуры сети распознавания изображений

Сфера информационных технологий стремительно развивается, методы совершенствуются. В настоящее время самыми перспективными архитектурами распознавания изображений являются RNN и GAN (Generative Adversarial Network), обладающие высокой и точной работой. Например, GAN - это модель, которая способна генерировать изображения, которые очень похожи на реальные.



Заключение

Нейронные сети - крайне интересное направление в сфере информационных технологий. В настоящее время оно активно внедряется в различные сферы жизни, а его популярность - лишь вопрос времени. Вскоре, с их помощью, люди разных национальностей смогут понимать друг друга, машины смогут определять наличие опухолей у пациента, а геологи и работники нефтегазовой сферы определять нужные месторождения, машина сможет читать документы.