

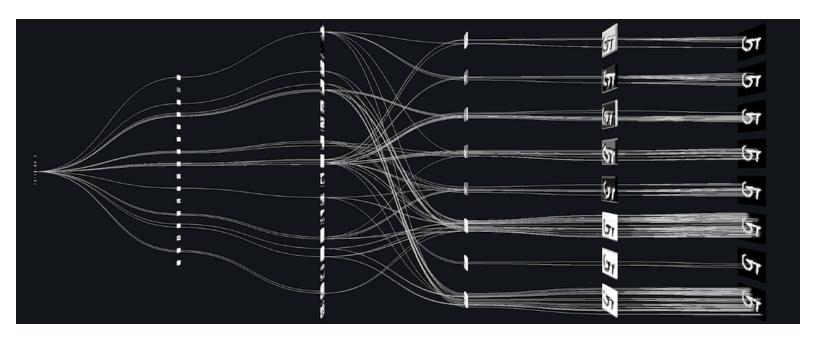
freetonik 8 сентября 2016 в 14:50

Что такое свёрточная нейронная сеть

Программирование *, Обработка изображений *, Машинное обучение *

Перевод

Автор оригинала: Adit Deshpande



Введение

Свёрточные нейронные сети (СНС). Звучит как странное сочетание биологии и математики с примесью информатики, но как бы оно не звучало, эти сети — одни из самых влиятельных инноваций в области компьютерного зрения. Впервые нейронные сети привлекли всеобщее внимание в 2012 году, когда Алекс Крижевски благодаря им выиграл конкурс ImageNet (грубо говоря, это ежегодная олимпиада по машинному зрению), снизив рекорд ошибок классификации с 26% до 15%, что тогда стало прорывом. Сегодня глубинное обучения лежит в основе услуг многих компаний: Facebook использует нейронные сети для алгоритмов автоматического проставления тегов, Google — для поиска среди фотографий пользователя, Amazon — для генерации рекомендаций товаров, Pinterest — для персонализации домашней страницы пользователя, a Instagram — для поисковой инфраструктуры.

Но классический, и, возможно, самый популярный вариант использования сетей это обработка изображений. Давайте посмотрим, как СНС используются для классификации изображений.

Задача

Задача классификации изображений — это приём начального изображения и вывод его класса (кошка, собака и т.д.) или группы вероятных классов, которая лучше всего характеризует изображение. Для людей это один из первых навыков, который они начинают осваивать с рождения.



What We See



What Computers See

Мы овладеваем им естественно, без усилий, став взрослыми. Даже не думая, мы можем быстро и легко распознать пространство, которое нас окружает, вместе с объектами. Когда мы видим изображение или просто смотрим на происходящее вокруг, чаще всего мы можем сразу охарактеризовать место действия, дать каждому объекту ярлык, и все это происходит бессознательно, незаметно для разума. Эти навыки быстрого распознавания шаблонов, обобщения уже полученных знаний, и адаптации к различной запечатлённой на фото обстановке не доступны нашим электронным приятелям.

Вводы и выводы

Когда компьютер видит изображение (принимает данные на вход), он видит массив пикселей. В зависимости от разрешения и размера изображения, например, размер массива может быть 32х32х3 (где 3 — это значения каналов RGB). Чтобы было понятней, давайте представим, у нас есть цветное изображение в формате JPG, и его размер 480х480. Соответствующий массив будет 480х480х3. Каждому из этих чисел присваивается значение от 0 до 255, которое описывает интенсивность пикселя в этой точке. Эти цифры, оставаясь бессмысленными для нас, когда мы определяем что на изображении, являются единственными вводными данными, доступными компьютеру. Идея в том, что вы даете компьютеру эту матрицу, а он выводит числа, которые описывают вероятность класса изображения (.80 для кошки, .15 для собаки, .05 для птицы и т.д.).

Чего мы хотим от компьютера

Теперь, когда мы определили задачу, ввод и вывод, давайте думать о том, как подбираться к решению. Мы хотим, чтобы компьютер мог различать все данные ему изображения и распознавать уникальные особенности, которые делают собаку собакой, а кошку кошкой. У нас и этот процесс происходит подсознательно. Когда мы смотрим на изображение собаки, мы можем отнести его к конкретному классу, если у изображения есть характерные особенности, которые можно идентифицировать, такие

изображений через поиск характеристик базового уровня, например границ и искривлений, а затем с помощью построения более абстрактных концепций через группы свёрточных слоев. Это общее описание того, что делают СНС. Теперь перейдём к специфике.

Биологические связи

В начале немного истории. Когда вы впервые услышали термин сверточные нейронные сети, возможно подумали о чем-то связанном с нейронауками или биологией, и отчасти были правы. В каком-то смысле. СНС — это действительно прототип зрительной коры мозга. Зрительная кора имеет небольшие участки клеток, которые чувствительны к конкретным областям поля зрения. Эту идею детально рассмотрели с помощью потрясающего эксперимента Хьюбел и Визель в 1962 году (видео), в котором показали, что отдельные мозговые нервные клетки реагировали (или активировались) только при визуальном восприятии границ определенной ориентации. Например, некоторые нейроны активировались, когда воспринимали вертикальные границы, а некоторые — горизонтальные или диагональные. Хьюбел и Визель выяснили, что все эти нейроны сосредоточены в виде стержневой архитектуры и вместе формируют визуальное восприятие. Эту идею специализированных компонентов внутри системы, которые решают конкретные задачи (как клетки зрительной коры, которые ищут специфические характеристики) и используют машины, и эта идея — основа СНС.

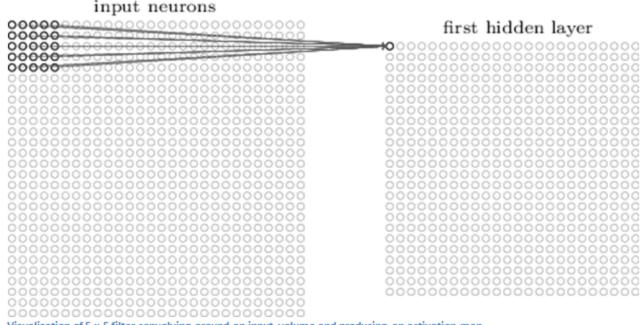
Структура

Вернёмся к специфике. Что конкретно делают СНС? Берётся изображение, пропускается через серию свёрточных, нелинейных слоев, слоев объединения и полносвязных слоёв, и генерируется вывод. Как мы уже говорили, выводом может быть класс или вероятность классов, которые лучше всего описывают изображение. Сложный момент — понимание того, что делает каждый из этих слоев. Так что давайте перейдем к самому важному.

Первый слой — математическая часть

Первый слой в СНС всегда свёрточный. Вы же помните, какой ввод у этого свёрточного слоя? Как уже говорилось ранее, вводное изображение — это матрица 32 х 32 х 3 с пиксельными значениями. Легче всего понять, что такое свёрточный слой, если представить его в виде фонарика, который светит на верхнюю левую часть изображения. Допустим свет, который излучает этот фонарик, покрывает площадь 5 х 5. А теперь давайте представим, что фонарик движется по всем областям вводного изображения. В терминах компьютерного обучения этот фонарик называется фильтром (иногда нейроном или ядром), а области, на которые он светит, называются рецептивным полем (полем восприятия). То есть наш фильтр — это матрица (такую матрицу ещё называют матрицей весов или матрицей параметров). Заметьте, что глубина у фильтра должна быть такой же, как и глубина вводного изображения (тогда есть гарантия математической верности), и размеры этого фильтра — 5 х 5 х 3. Теперь давайте за пример возьмем позицию, в которой находится фильтр. Пусть это будет левый верхний угол. Поскольку фильтр производит свёртку, то есть передвигается по вводному изображению, он умножает значения

фильтра на исходные значения пикселей изображения (поэлементное умножение). Все эти умножения суммируются (всего 75 умножений). И в итоге получается одно число. Помните, оно просто символизирует нахождение фильтра в верхнем левом углу изображения. Теперь повторим этот процесс в каждой позиции. (Следующий шаг — перемещение фильтра вправо на единицу, затем еще на единицу вправо и так далее). Каждая уникальная позиция введённого изображения производит число. После прохождения фильтра по всем позициям получается матрица 28 х 28 х 1, которую называют функцией активации или картой признаков. Матрица 28 х 28 получается потому, что есть 784 различных позиции, которые могут пройти через фильтр 5 х 5 изображения 32 х 32. Эти 784 числа преобразуются в матрицу 28 х 28.



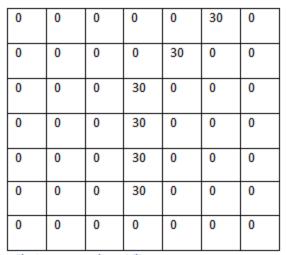
Visualization of 5 x 5 filter convolving around an input volume and producing an activation map

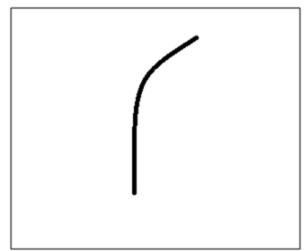
(Небольшая ремарка: некоторые изображения, в том числе то, что вы видите выше, взяты из потрясающей книги "Нейронные сети и глубинное обучение" Майкла Нильсена ("Neural Networks и Deep Learning", by Michael Nielsen). Настоятельно рекомендую).

Допустим, теперь мы используем два $5 \times 5 \times 3$ фильтра вместо одного. Тогда выходным значением будет $28 \times 28 \times 2$.

Первый слой

Давайте поговорим о том, что эта свертка на самом деле делает на высоком уровне. Каждый фильтр можно рассматривать как идентификатор свойства. Когда я говорю свойство, я имею в виду прямые границы, простые цвета и кривые. Подумайте о самых простых характеристиках, которые имеют все изображения в общем. Скажем, наш первый фильтр 7 х 7 х 3, и он будет детектором кривых. (Сейчас давайте игнорировать тот факт, что у фильтра глубина 3, и рассмотрим только верхний слой фильтра и изображения, для простоты). У фильтра пиксельная структура, в которой численные значения выше вдоль области, определяющей форму кривой (помните, фильтры, о которых мы говорим, это просто числа!).

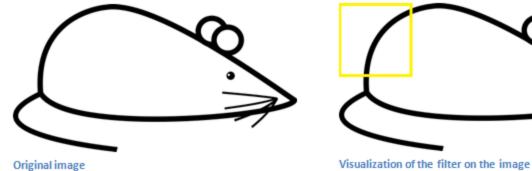


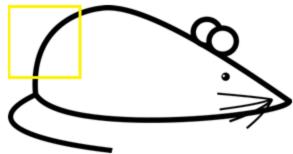


Pixel representation of filter

Visualization of a curve detector filter

Вернемся к математической визуализации. Когда у нас в левом верхнем углу вводного изображения есть фильтр, он производит умножение значений фильтра на значения пикселей этой области. Давайте рассмотрим пример изображения, которому мы хотим присвоить класс, и установим фильтр в верхнем левом углу.





Помните, всё что нам нужно, это умножить значения фильтра на исходные значения пикселей изображения.



Visualization of the receptive field

0	0	0	0	0	0	30
0	0	0	0	50	50	50
0	0	0	20	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0

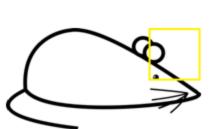
Pixel representation of the receptive field

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

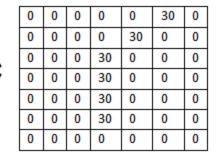
Pixel representation of filter

Multiplication and Summation = (50*30)+(50*30)+(50*30)+(20*30)+(50*30)=6600 (A large number!)

По сути, если на вводном изображении есть форма, в общих чертах похожая на кривую, которую представляет этот фильтр, и все умноженные значения суммируются, то результатом будет большое значение! Теперь давайте посмотрим, что произойдёт, когда мы переместим фильтр.



0	0	0	0	0	0	0
0	40	0	0	0	0	0
40	0	40	0	0	0	0
40	20	0	0	0	0	0
0	50	0	0	0	0	0
0	0	50	0	0	0	0
25	25	0	50	0	0	0



Visualization of the filter on the image

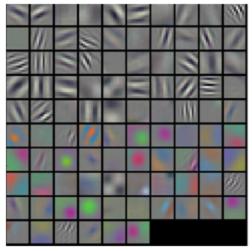
Pixel representation of receptive field

Pixel representation of filter

Multiplication and Summation = 0

Значение намного ниже! Это потому, что в новой области изображения нет ничего, что фильтр определения кривой мог засечь. Помните, что вывод этого свёрточного слоя — карта свойств. В самом простом случае, при наличии одного фильтра свертки (и если этот фильтр — детектор кривой), карта свойств покажет области, в которых больше вероятности наличия кривых. В этом примере в левом верхнем углу значение нашей 28 х 28 х 1 карты свойств будет 6600. Это высокое значение показывает, что, возможно, что-то похожее на кривую присутствует на изображении, и такая вероятность активировала фильтр. В правом верхнем углу значение у карты свойств будет 0, потому что на картинке не было ничего, что могло активировать фильтр (проще говоря, в этой области не было кривой). Помните, что это только для одного фильтра. Это фильтр, который обнаруживает линии с изгибом наружу. Могут быть другие фильтры для линий, изогнутых внутрь или просто прямых. Чем больше фильтров, тем больше глубина карты свойств, и тем больше информации мы имеем о вводной картинке.

Ремарка: Фильтр, о котором я рассказал в этом разделе, упрощён для упрощения математики свёртывания. На рисунке ниже видны примеры фактических визуализаций фильтров первого свёрточного слоя обученной сети. Но идея здесь та же. Фильтры на первом слое сворачиваются вокруг вводного изображения и "активируются" (или производят большие значения), когда специфическая черта, которую они ищут, есть во вводном изображении.



Visualizations of filters

(Заметка: изображения выше — из Стэнфордского курса 231N, который преподают Андрей Карпатый и Джастин Джонсон (Andrej Karpathy and Justin Johnson). Рекомендую тем, кто хочет лучше изучить СНС).

Идём глубже по сети

Сегодня в традиционной сверточной нейронной сетевой архитектуре существуют и другие слои, которые перемежаются со свёрточными. Я очень рекомендую тем, кто интересуется темой, почитать об этих слоях, чтобы понять их функциональность и эффекты. Классическая архитектура СНС будет выглядеть так:

Input -> Conv -> ReLU -> Conv -> ReLU -> Pool -> ReLU -> Conv -> ReLU -> Pool -> Fully Connected

Последний слой, хоть и находится в конце, один из важных — мы перейдём к нему позже. Давайте подытожим то, в чём мы уже разобрались. Мы говорили о том, что умеют определять фильтры первого свёрточного слоя. Они обнаруживают свойства базового уровня, такие как границы и кривые. Как можно себе представить, чтобы предположить какой тип объекта изображён на картинке, нам нужна сеть, способная распознавать свойства более высокого уровня, как например руки, лапы или уши. Так что давайте подумаем, как выглядит выходной результат сети после первого свёрточного слоя. Его размер 28 х 28 х 3 (при условии, что мы используем три фильтра 5 х 5 х 3). Когда картинка проходит через один свёрточный слой, выход первого слоя становится вводным значением 2-го слоя. Теперь это немного сложнее визуализировать. Когда мы говорили о первом слое, вводом были только данные исходного изображения. Но когда мы перешли ко 2-му слою, вводным значением для него стала одна или несколько карт свойств — результат обработки предыдущим слоем. Каждый набор вводных данных описывает позиции, где на исходном изображении встречаются определенные базовые признаки.

Теперь, когда вы применяете набор фильтров поверх этого (пропускаете картинку через второй свёрточный слой), на выходе будут активированы фильтры, которые представляют свойства более

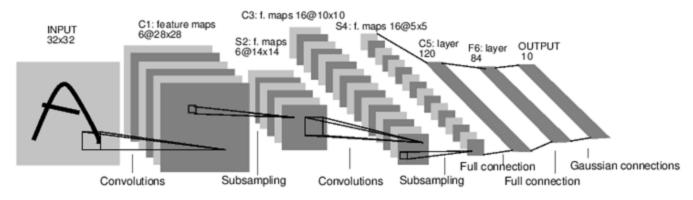
высокого уровня. Типами этих своиств могут оыть полукольца (комоинация прямои границы с изгиоом) или квадратов (сочетание нескольких прямых ребер). Чем больше свёрточных слоёв проходит изображение и чем дальше оно движется по сети, тем более сложные характеристики выводятся в картах активации. В конце сети могут быть фильтры, которые активируются при наличии рукописного текста на изображении, при наличии розовых объектов и т.д. Если вы хотите узнать больше о фильтрах в свёрточных сетях, Мэтт Зейлер и Роб Фергюс написали отличную научно-исследовательскую работу на эту тему. Ещё на ютубе есть видео Джейсона Йосински с отличным визуальным представлением этих процессов.

Еще один интересный момент. Когда вы двигаетесь вглубь сети, фильтры работают со все большим полем восприятия, а значит, они в состоянии обрабатывать информацию с большей площади первоначального изображения (простыми словами, они лучше приспособляются к обработке большей области пиксельного пространства).

Полносвязные слои

Теперь, когда мы можем обнаружить высокоуровневые свойства, самое крутое — это прикрепление полносвязного слоя в конце сети. Этот слой берёт вводные данные и выводит N-пространственный вектор, где N — число классов, из которых программа выбирает нужный. Например, если вы хотите программу по распознаванию цифр, у N будет значение 10, потому что цифр 10. Каждое число в этом N-пространственном векторе представляет собой вероятность конкретного класса. Например, если результирующий вектор для программы распознавания цифр это [0 0,1 0,1 0,75 0 0 0 0 0,05], значит существует 10% вероятность, что на изображении "1", 10% вероятность, что на изображение "2", 75% вероятность — "3", и 5% вероятность — "9" (конечно, есть и другие способы представить вывод).

Способ, с помощью которого работает полносвязный слой — это обращение к выходу предыдущего слоя (который, как мы помним, должен выводить высокоуровневые карты свойств) и определение свойств, которые больше связаны с определенным классом. Например, если программа предсказывает, что на каком-то изображении собака, у карт свойств, которые отражают высокоуровневые характеристики, такие как лапы или 4 ноги, должны быть высокие значения. Точно так же, если программа распознаёт, что на изображении птица — у неё будут высокие значения в картах свойств, представленных высокоуровневыми характеристиками вроде крыльев или клюва. Полносвязный слой смотрит на то, что функции высокого уровня сильно связаны с определенным классом и имеют определенные веса, так что, когда вы вычисляете произведения весов с предыдущим слоем, то получаете правильные вероятности для различных классов.



A Full Convolutional Neural Network (LeNet)

Обучение (или "Что заставляет эту штуку работать")

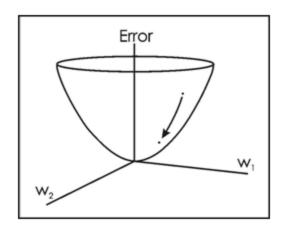
Это один из аспектов нейронных сетей, о котором я специально до сих пор не упоминал. Вероятно, это самая важная часть. Возможно, у вас появилось множество вопросов. Откуда фильтры первого свёрточного слоя знают, что нужно искать границы и кривые? Откуда полносвязный слой знает, что ищет карта свойств? Откуда фильтры каждого слоя знают, какие хранить значения? Способ, которым компьютер способен корректировать значения фильтра (или весов) — это обучающий процесс, который называют методом обратного распространения ошибки.

Перед тем, как перейти к объяснению этого метода, поговорим о том, что нейронной сети нужно для работы. Когда мы рождаемся, наши головы пусты. Мы не понимаем как распознать кошку, собаку или птицу. Ситуация с СНС похожа: до момента построения сети, веса или значения фильтра случайны. Фильтры не умеют искать границы и кривые. Фильтры верхних слоёв не умеют искать лапы и клювы. Когда мы становимся старше, родители и учителя показывают нам разные картинки и изображения и присваивают им соответствующие ярлыки. Та же идея показа картинки и присваивания ярлыка используется в обучающем процессе, который проходит СНС. Давайте представим, что у нас есть набор обучающих картинок, в котором тысячи изображений собак, кошек и птиц. У каждого изображения есть ярлык с названием животного.

— предсказание) в квадрате.

$$E_{total} = \sum \frac{1}{2} (target - output)^2$$

Примем это значение за переменную L. Как вы догадываетесь, потеря будет очень высокой для первых двух обучающих изображений. Теперь давайте подумаем об этом интуитивно. Мы хотим добиться того, чтобы спрогнозированный ярлык (вывод свёрточного слоя) был таким же, как ярлык обучающего изображения (это значит, что сеть сделала верное предположение). Чтобы такого добиться, нам нужно свести к минимуму количество потерь, которое у нас есть. Визуализируя это как задачу оптимизации из математического анализа, нам нужно выяснить, какие входы (веса, в нашем случае) самым непосредственным образом способствовали потерям (или ошибкам) сети.



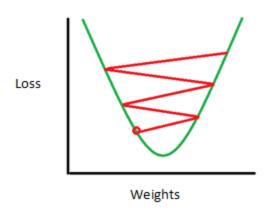
One way of visualizing this idea of minimizing the loss is to consider a 3-D graph where the weights of the neural net (there are obviously more than 2 weights, but let's go for simplicity) are the independent variables and the dependent variable is the loss. The task of minimizing the loss involves trying to adjust the weights so that the loss decreases. In visual terms, we want to get to the lowest point in our bowl shaped object. To do this, we have to take a derivative of the loss (visual terms: calculate the slope in every direction) with respect to the weights.

(Один из способов визуализировать идею минимизации потери — это трёхмерный график, где веса нейронной сети (очевидно их больше, чем 2, но тут пример упрощен) это независимые переменные, а зависимая переменная — это потеря. Задача минимизации потерь — отрегулировать веса так, чтобы снизить потерю. Визуально нам нужно приблизиться к самой нижней точке чашеподобного объекта. Чтобы добиться этого, нужно найти производную потери (в рамках нарисованного графика — рассчитать угловой коэффициент в каждом направлении) с учётом весов).

Это математический эквивалент **dL/dW**, где W — веса определенного слоя. Теперь нам нужно выполнить **обратное распространение** через сеть, которое определяет, какие веса оказали большее влияние на потери, и найти способы, как их настроить, чтобы уменьшить потери. После того, как мы вычислим производную, перейдём к последнему этапу — обновлению весов. Возьмём все фильтровые веса и обновим их так, чтобы они менялись в направлении градиента.

$$w = w_i - \eta \frac{dL}{dW}$$

w = Weight w_i = Initial Weight η = Learning Rate **Скорость обучения** — это параметр, который выбирается программистом. Высокая скорость обучения означает, что в обновлениях веса делались более крупные шаги, поэтому образцу может потребоваться меньше времени, чтобы набрать оптимальный набор весов. Но слишком высокая скорость обучения может привести к очень крупным и недостаточно точным скачкам, которые помешают достижению оптимальных показателей.



Consequence of a high learning rate where the jumps are too large and we are not able to minimize the loss.

Процесс прямого распространения, функцию потери, обратное распространение и обновление весов, обычно называют одним периодом дискретизации (или epoch — эпохой). Программа будет повторять этот процесс фиксированное количество периодов для каждого тренировочного изображения. После того, как обновление параметров завершится на последнем тренировочном образце, сеть в теории должна быть достаточно хорошо обучена и веса слоёв настроены правильно.

Тестирование

Наконец, чтобы увидеть, работает ли СНС, мы берём другой набор изображений и ярлыков и пропускаем изображения через сеть. Сравниваем выходы с реальностью и смотрим, работает ли наша сеть.

Как компании используют СНС

Данные, данные, данные. Компании, у которых тонны этого шестибуквенного магического добра, имеют закономерное преимущество перед остальными конкурентами. Чем больше тренировочных данных, которые можно скормить сети, тем больше можно создать обучающих итераций, больше обновлений весов и перед уходом в продакшн, получить лучше обученную сеть. Facebook (и Instagram) могут использовать все фотографии миллиарда пользователей, которые у них сегодня есть, Pinterest — информацию из 50 миллиардов пинов, Google — данные поиска, а Amazon — данные о миллионах продуктов, которые ежедневно покупаются. И теперь вы знаете какое волшебство они используют в своих целях.

Ремарка

Q

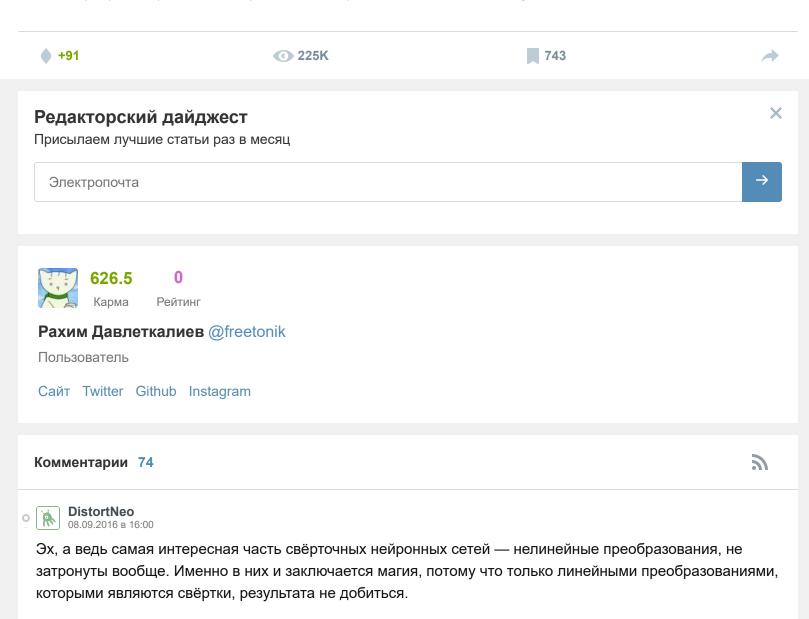
которые не обсуждались в этой статье, относятся нелинейные (nonlinear) слои и слои объединения (pooling), а также гиперпараметры сети, такие как размеры фильтров, шаги и отступы. Сетевая архитектура, пакетная нормализация, исчезающие градиенты, выпадение, методы инициализации, невыпуклая оптимизация, сдвиг, варианты функций потерь, расширение данных, методы регуляризации, машинные особенности, модификации обратного распространения и много других необсуждённых (пока;-) тем.

(Перевод Наталии Басс)

Обратите внимание на схему: Input -> Conv -> **ReLU** -> Conv ->...

Теги: глубинное обучение, машинное зрение, нейронные сети, мозг, зрение

Хабы: Программирование, Обработка изображений, Машинное обучение

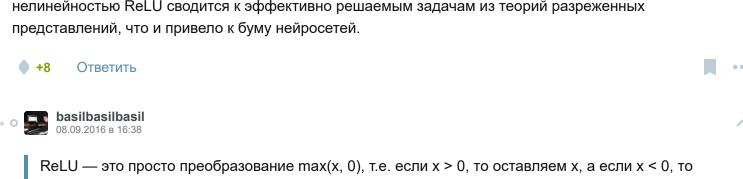


ReLU — это просто преобразование $\max(x, 0)$, т.е. если x > 0, то оставляем x, а если x < 0, то заменяем

на 0.

Почему в качестве нелинейного слоя в свёрточных нейронных сетях используется именно ReLU? Да чисто из практических соображений. Существует большое количество нелинейных преобразований, в т.ч. и более эффективных (размер сети получается меньше). Но именно с ReLU есть возможность обучения сравнительно большой сети за разумное время.

Дело в том, что обучение сети — это математическая задача минимизации ошибки, как это написано в статье. Но размерности задачи настолько огромны, что долгое время не существовало эффективных методов минимизации, и нейронные сети были похоронены. Однако, в последнее десятилетие была создана теория разреженных представлений, созданы математические методы приближённого решения задач L0 и L1 минимизации. А потом выяснилось, что задача обучения свёрточной нейронной сети с нелинейностью ReLU сводится к эффективно решаемым задачам из теорий разреженных представлений, что и привело к буму нейросетей.



заменяем на 0.

да, не раскрыта тема использования гиперболического тангенса.



НЛО прилетело и опубликовало эту надпись здесь

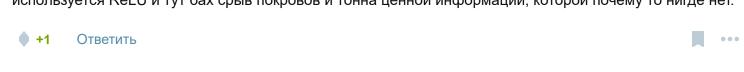


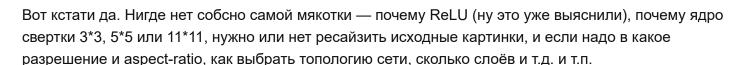
Seiten 08.09.2016 в 20:43

bertmsk

DistortNeo

А откуда у вас такие детальные познания? Что такого почитать если основы уже ясны понятны? При чем хотелось бы именно в духе того что вы написали. Ну типа причина того почему именно используется ReLU и тут бах срыв покровов и тонна ценной информации, которой почему то нигде нет.







почему ядро свертки 3*3, 5*5 или 11*11

09.09.2016 в 01:52

А это просто объяснить через аналогию с фонариком. Вырезаете из картинки кусочек определённого размера и пытаетесь понять, можно ли только по этому кусочку сделать обработку или классификацию. Если нет (кусочек слишком мал) — берите размер фильтра побольше. Если же вам кажется, что размер кусочка можно уменьшить, не потеряв в точности классификации — смело уменьшайте. Чем меньше лишних данных принимает алгоритм машинного обучения, тем легче он обучается и более точные результаты выдаёт.

нужно или нет ресайзить исходные картинки, и если надо в какое разрешение и aspect-ratio

Зависит от конкретной задачи. В общем случае да, нужно подавать на вход нормализованные данные. Вместо обучения нейросети на нахождение объектов разных размеров (отличие в разы) эффективнее обучать на нахождение объектов фиксированного размера и масштабировать саму картинку.

как выбрать топологию сети, сколько слоёв

Если честно, то:

- 1. Методом научного тыка: попробовать разные варианты и выбрать наиболее удачный.
- 2. Поизучать научную литературу по данной тематике и выбрать готовую успешную топологию из статьи, авторы которой потратили кучу времени и добились успешных результатов, применяя один из этих двух методов.



НЛО прилетело и опубликовало эту надпись здесь



Hy так работа обязывает во всём этом разбираться. Что почитать — к сожалению, ничего посоветовать не могу. В моём случае это научные статьи: читаете статью, смотрите ссылки, которые кажутся интересными, читаете их и так пока не надоест. Со временем в голове начинает складываться целостная картина.



Ответить

Жду появления статьи о том как научить нейронную сеть играть в крестики-нолики.

Эксперты — АУ!!!





Оптимальная игра в крестики-нолики неплохо реализуется в виде обычного алгоритма. Нейронные же сети следует применять там, где ничто другое не помогает, как, например, в задачах машинного зрения.



0

НЛО прилетело и опубликовало эту надпись здесь



Ответить

Может вам лучше сделать забивалку гвоздей, в которой в качестве ударного элемента использовался бы микроскоп?

НЛО прилетело и опубликовало эту надпись здесь



Ну как бы разные сетки в последние годы берут первые места в соревнованиях по распознаванию/ локализации объектов

https://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html



НЛО прилетело и опубликовало эту надпись здесь



Вобще-то самое прямое. И к науке и к бизнесу. Для решения задач классификации/ распознавания/еtc есть огромное количество подходов разной степени паршивости. Для того, чтобы эту самую степень паршивости определить требуется некоторый объективный критерий. Один из таких критериев — точность алгоритма на известных и публично доступных датасетах (MNIST, CIFAR, ImageNet, и т.д.).



НЛО прилетело и опубликовало эту надпись здесь

>Все «нейросетки», которые попадались мне, удается обмануть легко

Если вы работали с компьютерным зрением, то должны понимать, что там практически нет задач которые были бы решены полностью. Всегда есть некоторый предел. Новые методы этот предел отодвигают. Для многих задач, на текущий момент самые продвинутые методы основанны на Deep Learning.

Та же фигня с распознаванием голоса.

На телефоне точно есть Machine Learning. Именно с его помощью происходит поиск лиц в кадре. В те времена когда я интересовался Computer Vision лучше всего эту задачу решали каскады. Теперь, возможно это делают сетки.

Всякие распознавалки текста с большой вероятностью используют внутри собственно сверточные сети.

Prisma — это тоже сети.

Подозреваю что можно еще много примеров найти, но мне лень.

>И ещё разок. Котенка никто не учит, что такое собака.

А какая разница как кто учит котенка? «Биологические» нейронные сети и то что мы называем Deep Learning это довольно разные вещи.



НЛО прилетело и опубликовало эту надпись здесь



> Я и возражал. Deep Learning и «нейросети» не имеют никакого отношения ни к живым организмам, ни к ИИ.

Ну я этого нигде не утверждал. И это не мешает сетям показывать лучший результат во многих задачах.

Подозреваю, чтобы сделать «распознавание как у котенка» надо бы хорошо понимать как собственно оно у котенка работает. И даже после этого нет гарантии что получится нечто лучше существующих методов.

>Но опять, сделайте ловушку для тараканов, станете миллиардером Для начала — я не специалист по Deep Learning/Computer Vision. Ну и подозреваю, что себестоимость роботизированных систем для истребления тараканов и стоимость RnD просто не позволят выйти даже в плюс, не то чтобы заработать миллиарды.

Если что — я никого не минусую.



НЛО прилетело и опубликовало эту надпись здесь



Могу перестать спорить, это уже становится довольно скучно и однообразно, если честно =(

RnD дорогой не потому, что Deep learning такой плохой, а потому, что кейз «распознай таракана в темноте между кучей объектов» довольно сложен и не тривиален. Но это не самая большая проблема. Я даже думаю что некоторые успехи можно даже получить используя классические методы компьютерного зрения. Большая часть средств уйдет на сам инструмент уничтожения тараканов. Не окупится это все дело потому, что себестоимость железки будет большой, относительно классических средств. При этом физически уничтожать тараканов это не так эффективно как их травить. Т.е. мы получаем дорогое и не эффективное средство которое надо разработать, против вроде как хорошего и дешевого которое уже существует.

Не стоит сваливать на DL/CV недостатки бизнес-модели.



НЛО прилетело и опубликовало эту надпись здесь



RnD — https://en.wikipedia.org/wiki/Research and development.

DL/CV — Deep Learning / Computer Vision, мне просто надоело писать длинные названия

Мне кажется вам стоит несколько лучше разобраться в той сфере о которой вы пытаетесь делать выводы.

Можно начать с этого https://habrahabr.ru/post/274725/

На этом у меня все.



Ответить

Да, если что, на гиктаймс проскакивала статья про отпугивание котов при помощи DL. Как первое приближение можете взять локализацию для тараканоловушки оттуда





Ответить

Вы с животными видимо только по книжкам общались, да в зоопарке за решёткой видели?

Когда собака выкармливает котят или кошка щенков, система даёт сбой? В быту кошки с собаками порой неплохо ладят.

А котёнок любит спину дыбить даже на руку — этому его кто учил?

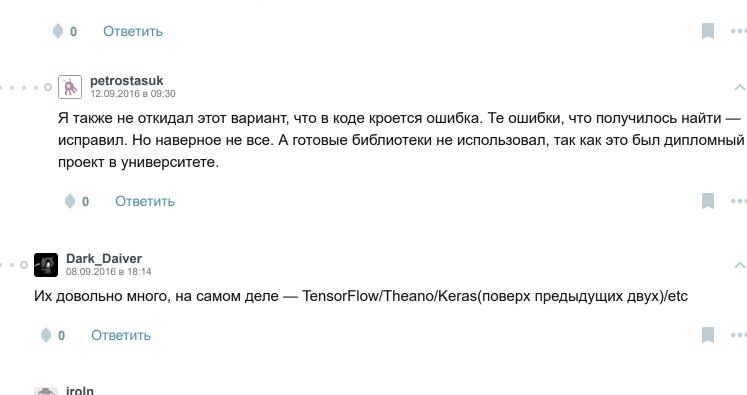
0

pelioslasuk 08.09.2016 в 16:22 Еще будучи студентом, использовал этот тип НС для решения задачи распознавания цифр из базы MNIST. При этом весь процесс распознавания запускал на видеокарте, используя технологию CUDA. Тогда получилось достичь точности распознавания около 97,8%, но в научных работах людям удавалось достичь точности около 99%. Подозреваю, что здесь большую роль играют начальные настройки НС, когда только начинаем процесс обучения. 0 Ответить **DistortNeo** 08.09.2016 в 16:34 Большую роль играет ещё и training data set. 0 Ответить samodum 09.09.2016 в 10:12 Он же написал, что training data set — это MNIST 0 Ответить **DistortNeo** 09.09.2016 в 10:19 MNIST — это просто набор изображений. Training data set они становятся только после соответствующей предобработки, учитывающей особенности алгоритма машинного обучения. 0 Ответить Alex ME 08.09.2016 в 17:40 Скажите, Вы реализовывали все сами или использовали какие-то готовые решения для нейронных сетей? Если да, то какие позволяют использовать видеокарту? 0 Ответить petrostasuk Писал все сам. Использовал язык С и SDK CUDA, тогда еще версия была примерно 2.0. Компилировал всё в Visual Studio. 0 Ответить supersonic_snail 09.09.2016 в 15:20 > Писал все сам.

Вполне возможно, что в этом и проблема, если вы не делали gradient checking. Очень просто

сделать какую-то мелкую ошибку в паре прямое-обратное распространение, типа индекса сдвинутого на 1, с которой все в принципе работает, но не так, как могло бы.

Из-за того, что сеть по сути функция, а обратное распространение считает градиент, его правильность можно очень просто проверить численно через определение градиента (df(x)/dx = (f(x+eps) - f(x-eps)) / (2*eps)). В данном случае f(x) — cost function, которые минимизируется, x — параметры сети. Результаты, полученные численно и обратным распространением, должны совпадать числа до 5-6. Если разница больше, то что-то не так.



какие-то готовые решения для нейронных сетей? Если да, то какие позволяют использовать видеокарту?

09.09.2016 в 17:37 🕏

Сейчас все современные библиотеки для нейронных сетей используют GPU или кластер из GPU. Это просто необходимость. На CPU в production-применении сети не обучают и ничего не классифицируют ими, потому что это очень-очень медленно. Берите любую библиотеку: caffe, tensorflow, cntk, theano

Ho надо учесть, что требуется современная мощная видеокарта (обычно NVIDIA с compute capability не меньше 3.0).



Большое спасибо! Я посмотрел разные библиотеки, пока остановился на TensorFlow, показалось самой интересной, читаю туториалы.





Интересно, но что по можете указать по поводу искусственной нейронной сети Хопфилда? Она обладает свойством памяти, фильтрации и востановления изображения.





Скорость обучения — это параметр, который выбирается программистом. Высокая скорость обучения означает, что в обновлениях веса делались более крупные шаги, поэтому образцу может потребоваться меньше времени, чтобы набрать оптимальный набор весов. Но слишком высокая скорость обучения может привести к очень крупным и недостаточно точным скачкам, которые помешают достижению оптимальных показателей.

а почему бы не ввести скорость обучения сети как один из выходных параметров, чтобы сеть сама оптимально задавала себе её?





Большая часть продвинутых алгоритмов оптимизации и так стараются адаптивно регулировать скорость обучения.





Это параметр процесса оптимизации, а не самой модели. Если Вы хотите оптимизировать оптимизацию :) то это другой процесс которому тоже нужны параметры. Но есть похожая рабочая идея — синтетические градиенты. Цель, правда, там другая.



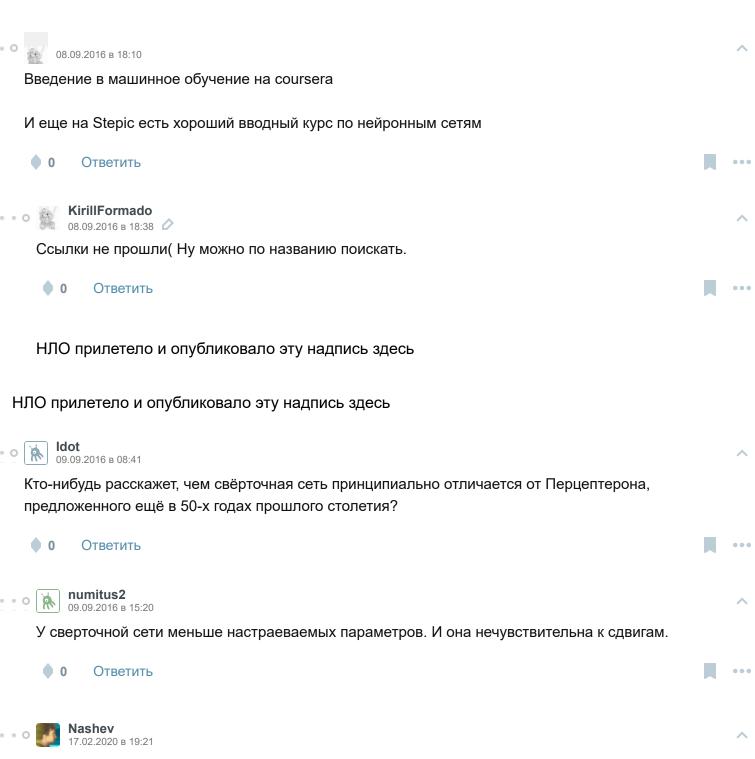


Подозреваю, что создать алгоритм, который бы оптимально подбирал параметры оптимизации — это на данном этапе развития более сложная задача, чем подобрать параметры вручную.





Сделали бы вы курс по машинному обучению, даже самый вводный. С радостью бы прошел...



Хороший вопрос, хоть и задан 4 года назад. Простите за некрокоммент.

Персептрон (однослойный) — это на каждом нейроне собранная сумма с входов с коэффициентами. Это, по сути, то самое, как тут описан свёрточный фильтр. И фильтров таких в свёрточной сети несколько, ровно как нейронов в персептроне.

Только персептрон применялся сразу весь, и сразу ко всем своим входам. А каждый фильтр в свёрточной сети применяется к большой куче разных фрагментов входного слоя.

То есть, получается, свёрточный слой в свёрточной сети — это фактически тот самый персептрон, но который просто последовательно коммутируют к разным фрагментам входного потока. И его выходы накапливают в качестве входов для следующего слоя нейронной сети.



Мне кажется, что через 10-20 лет будет придуман более простой механизм распознавания образов.



ldot

Ответить

0

Мы хотим, чтобы компьютер мог различать все данные ему изображения и распознавать уникальные особенности, которые делают собаку собакой, а кошку кошкой.

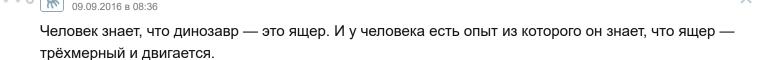
И мы думаем, что раз мы можем, значит и компьютер сможет. Забывая, что мы всегда видим мир в движении и когда оно застывает — это кажется неестественным. Думаю, это фундаментальная ошибка обучателей нейросетей.

Например, по картинкам сеть не узнает, что шерсть колышется на ходу.



Например, о динозаврах вы в детстве наверняка обучались по статичным картинкам, ничего не зная о моторике их шерсти/чешуи. Тем не менее, распознаете динозавров, я уверен, с высокой точностью. ;)

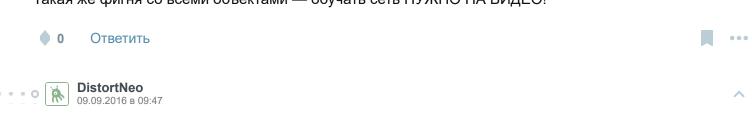




Поясню на знаменитом диване, который свёрточная сеть распознаёт как леопарда:

- для человека леопард это большая кошка, а кошка имеет четыре ноги (вход, выход, земля и питание), хвост, голову и так далее,
- для свёрточных сетей леопард это характерная текстура, и потому диван крашенный под леопарда, для свёрточной сети и есть леопард, потому что сеть не знает, что леопард должен иметь четыре ноги, хвост, голову и так далее.

Такая же фигня со всеми объектами — обучать сеть НУЖНО НА ВИДЕО!



Такова природа свёрточных сетей — они видят исключительно локально, небольшими фрагментами, без учёта контекста. Обычная свёрточная сеть способна определить текстуру, даже найти составные части (голова, конечность), но возможности абстрактного мышления у неё крайне

ограничены как концептуально, так и технически.

И видео не как набор изображений, а именно как связная последовательность, ей в данном случае не поможет никак.





Разрешите все же подвергнуть Ваш радикальный капс сомнению.

Проблема знания о четырех ногах и возможных действиях объекта типа «ящер» лежит не в плоскости видео/статика. Анализ видео теми же алгоритмами даст вам просто **гораздо** большую коллекцию картинок, которые будут чрезвычайно схожи, а после нормализации вообще одинаковы. В результате, обучение станет низко эффективным.

Допустим, Вы видите картинку, леопарда|дивана в плохом разрешении. Издалека. Вы распознаете по контексту (ковер на стене, часы, журнальный столик). Из этого контекста получается «комната», а в комнате леопарду делать нечего, следовательно это что-то, что может быть в комнате. О! Диван.

Надо понимать, что нахождение объектов на картинке (кадре) это еще не «компьютерное зрение». Это лишь один из слоев. Что, ни в коем случае, не уменьшает его важности.





У Путина дома живёт тигрёнок, так почему бы и леопарду не жить в комнате?

не в плоскости видео/статика. Анализ видео теми же алгоритмами даст вам просто гораздо большую коллекцию картинок, которые будут чрезвычайно схожи, а после нормализации вообще одинаковы. В результате, обучение станет низко эффективным.

Зрение должно выделять ДВИЖУЩИЕСЯ объекты. Например, нейросети лягушки вообще не видят, то что неподвижно.





А теперь представьте размер нейросети, которая способна решать подобные задачи на видео. И представьте, какой объём данных понадобится, чтобы её обучить.





Прежде всего нужно обучить нейросеть выделять движущиеся объекты отделяя от

неподвижных. А затем анализировать уже выделенные ооъекты. Что потреоует разделения нейронной сети на «отделы мозга».





Для выделения движущихся объектов давно созданы простые и быстрые алгоритмы— нейросеть здесь не нужна.

Сравните аналог: мозг человека имеет огромнейшую производительность, но математические операции все равно быстрее будет выполнять калькулятор, состоящий из сотни транзисторов.

И как вы собираетесь анализировать выделенные объекты? Объём данных (всевозможные видеофрагменты распознаваемых объектов), достаточных для обучения нейросети, будет настолько большой, а скорость обработки настолько медленной, что эффективнее будет использовать биологическую реализацию нейросети (живого человека).





Обучать нейросеть нужно именно на выделенных объектах. Иначе будет полнейшая хрень вроде дивана леопардового окраса.

Сам метод обучения на статических картинках — ущербный, потому что объект это не цветные пятна с характерным узором, а трёхмерный объект.





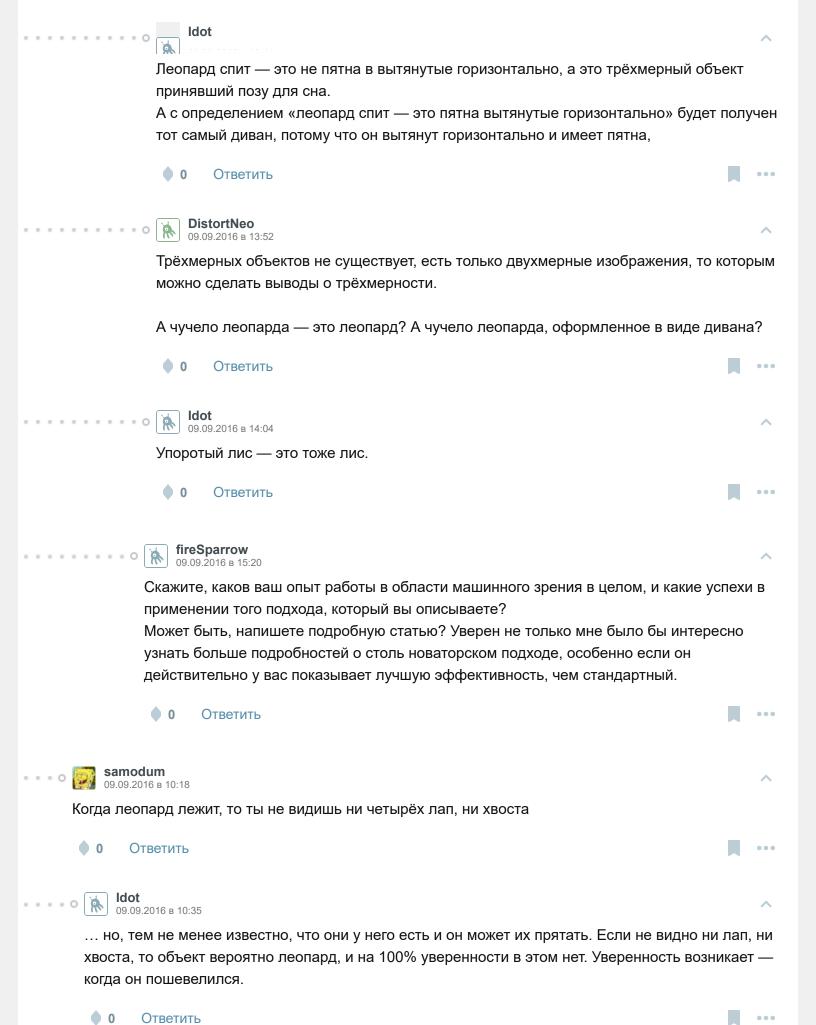
Обучать нейросеть нужно именно на выделенных объектах.

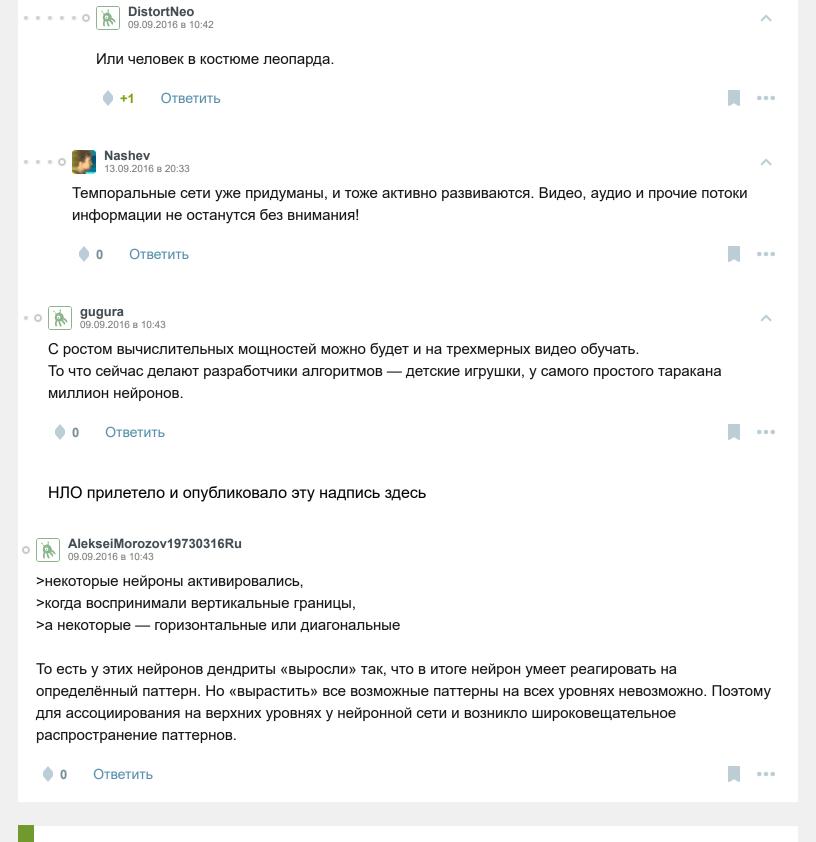
Обучайте. Я что, против? Я просто указал на то, что для забивания гвоздя достаточно использования молотка.

Сам метод обучения на статических картинках — ущербный, потому что объект это не цветные пятна с характерным узором, а трёхмерный объект.

Может и ущебный, но при современном уровне развития техники и математического аппарата только он и является доступным.

К тому же не забывайте про фотографии или про ситуацию, когда леопард спит.





Только полноправные пользователи могут оставлять комментарии. Войдите, пожалуйста.



Суперсила профессионалов.



ПОХОЖИЕ ПУБЛИКАЦИИ

2 июня 2020 в 12:10

Книга «Генеративное глубокое обучение. Творческий потенциал нейронных сетей»

+5

4.1K

31

0

2 ноября 2018 в 12:00

I see you: машинное обучение и искусственные нейронные сети в изучении зрения дрозофил

+17

◎ 6.7K

40

1 +1

8 октября 2018 в 14:38

Обучение и тестирование нейронных сетей на PyTorch с помощью Ignite

+36

② 25K

157

4 +4

минуточку внимания

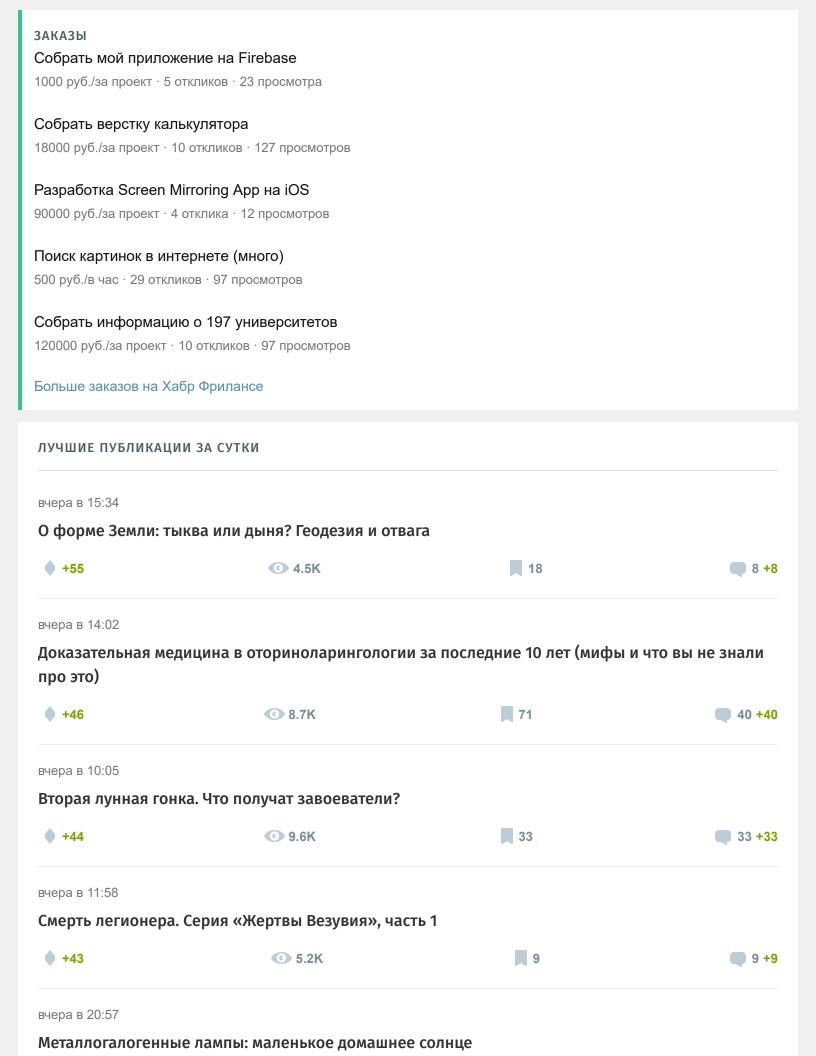
Разместить







Город, которого нет: OFDMA, MU-MIMO и 802.11ax

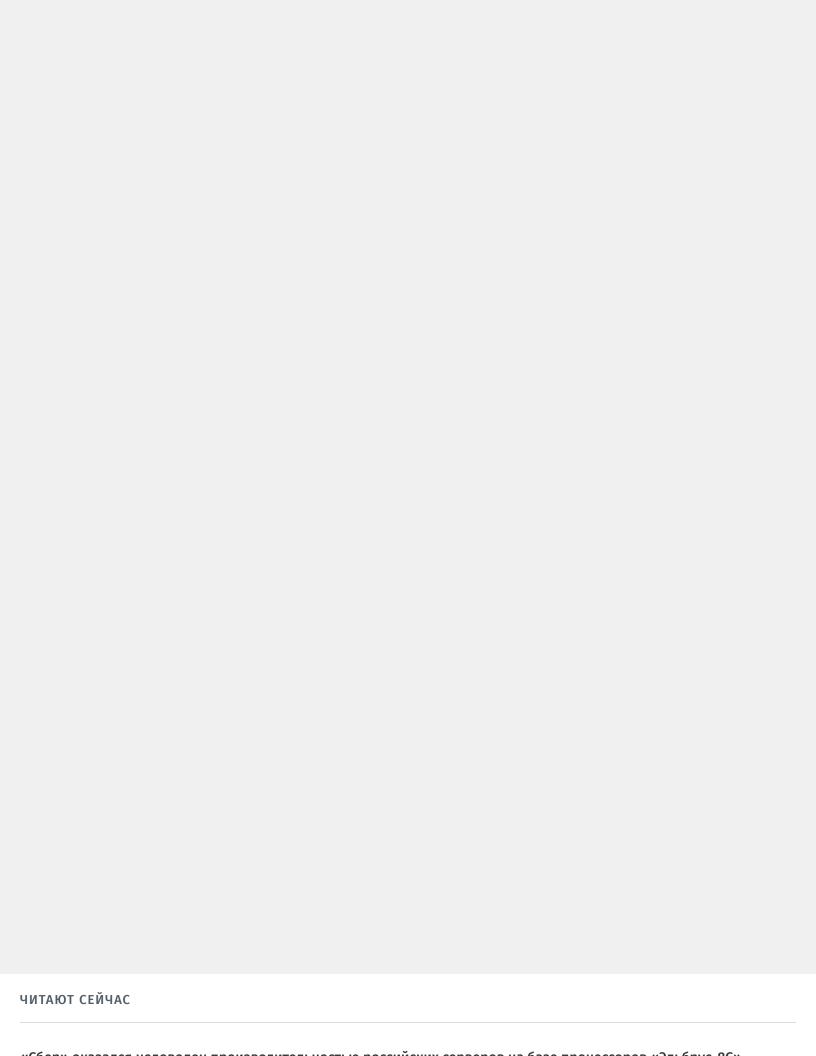


Реклама



Ваш аккаунт	Разделы	Информация	Услуги
Войти	Публикации	Устройство сайта	Реклама
Регистрация	Новости	Для авторов	Тарифы
	Хабы	Для компаний	Контент
	Компании	Документы	Семинары
	Авторы	Соглашение	Мегапроекты
	Песочница	Конфиденциальность	
	f w	✓	_

О сайте
Техническая поддержка
Вернуться на старую версию
© 2006–2021 «Habr»



Непереводимые слова: 7 русских лексем, которых не хватает в английском
○ 15K ○ 54 +54 ○ 15K ○ 54 +54 ○ 15K ○ 54 +54 ○ 15K ○ 15K
Авторы коллективного иска против Apple утверждают, что Apple Watch опасно носить
3.6K ■ 5 +5
Кризис видеокарт — что делать и кто виноват
② 5.7K ■ 32 +32
Госкорпорация Ростех показала беспилотник C-70 «Охотник» с плоским реактивным соплом
② 5.8K ■ 14 +14
Паззл DevSecOps: как из знаний, инструментов и процессов сложить слово «безопасность»
Турбо
РАБОТА
Data Scientist
144 вакансии
Все вакансии

«соер» оказался недоволен производительностью российских серверов на базе процессоров «эльорус-вс»