

Выделение и классификация объектов на изображениях с помощью сверточных нейронных сетей

Ключевые слова : Компьютерное зрение(КЗ), распознавание объектов, классификация, локализация(выделение), сверточные нейронные сети(CNN), машинное обучение(ML), глубокое обучение.

Аннотация

Компьютерное зрение существует уже более 50 лет, но в последнее время мы наблюдаем значительное возрождение интереса к тому, как машины "видят" и как компьютерное зрение может быть использовано для создания продуктов для потребителей и предприятий. Несколько примеров таких приложений - Amazon Go, Google Lens, автономные транспортные средства, распознавание лиц. Ключевым движущим фактором всего этого является компьютерное зрение. Проще говоря, компьютерное зрение[6]-это дисциплина в рамках широкой области искусственного интеллекта, которая учит машины видеть. И обнаружение объектов на изображениях, видео и в реальном времени было целью изучения многих исследователей.

В рамках данной статьи были анализированы и тестированы различные сверточные методы для классификации и локализации объектов на изображениях.

Введение

Область компьютерного зрения имеет дело с целью создания систем, которые могут обрабатывать изображения так же, как люди, или даже лучше. Возможность сортировки изображений в различные классы в зависимости от содержания называется классификацией изображений. Возможность идентифицировать несколько объектов на изображениях и определить их местоположение, поместив ограничивающую рамку вокруг объекта, называется локализацией объекта. Обнаружение объектов - это под часть компьютерного зрения, которое в настоящее время в значительной степени основано на машинном обучении. В течение последнего десятилетия в области машинного обучения доминировали так называемые глубокие нейронные сети, которые используют преимущества улучшения вычислительной мощности и доступности данных. Сверточные нейронные сети являются одним из наиболее эффективных и популярных методов локализации и идентификации объектов на изображениях. Основной целью данной статьи является анализ и тестирование классификации и обнаружения объектов с использованием сверточных нейронных сетей.

Кроме того, для достижения цели в данной работе поставлены следующие задачи:

1. Обзор и анализ существующих решений проблемы распознаваний объектов
2. Определить недостатки и преимущества методов
3. Тестирование сверточных методов

Обзор предметной области

Для определения требований к формируемому методу, необходимо рассмотреть существующие решения. Аналоги должны различаться по способу решения проблемы компьютерного зрения с использованием сверточных методов.

- Распознавание объектов на аэрофотоснимках с использованием CNN (с YOLO)

Обучение сверточных нейронных сетей на множестве аэрофотоснимков для эффективного и автоматизированного распознавания объектов. Сверточная нейронная сеть реализована в платформе "YOLO" (You Look Only Once). Объекты могут быть отслежены, обнаружены и классифицированы с помощью видеопотоков, подаваемых БПЛА (беспилотным летательным аппаратом) в режиме реального времени.

- Распознавание английского текста сверточной нейронной сетью

Оффлайн-распознавание текста, то есть, распознавание образов с входного изображения в текстовый формат. Разработка модели сверточной нейронной сети осуществлялась в Matlab с использованием MATLAB Library for Convolutional Neural Networks

- Сверточная нейронная сеть для решения задачи классификации

Сверточная нейронная сеть для решения задачи классификации текстов, ее структуру, метод обучения и преимущества перед полно связной нейронной сетью.

- Распознавание Лиц: Сверточный Нейросетевой Подход

Реализовано гибридное нейросетевое решение, которое выгодно отличается от других методов. Система объединяет локальную выборку изображений, самоорганизующуюся карту (СОК) нейронной сети и сверточной нейронной сети. СОК обеспечивает квантование образцов изображения в топологическое пространство, где входы, которые находятся рядом в исходном пространстве, также находятся рядом в выходном пространстве, тем самым обеспечивая уменьшение размерности и инвариантность к незначительным изменениям в образце изображения, а сверточная нейронная сеть обеспечивает частичную инвариантность к перемещению, вращению, масштабированию и деформации.

- Обнаружение пешеходов с помощью сверточных нейронных сетей

Новый метод обнаружения пешеходов, основанный на использовании классификатора сверточной нейронной сети (CNN). оценивается предложенный метод на сложной базе данных, содержащей пешеходов в городской среде без ограничений по позе, действию, фону и условиям освещения

Анализ аналогов проводился по следующим критериям:

Критерий 1

Решает ли проблему компьютерного зрения(кз)? (Классификация и локализация)
Нейронная сеть должна решать проблему классификации и локализации объектов.

Критерий 2

Наличие(использование) разных сверточных методов(R-CNN, Fast R-CNN, Faster R-CNN).

Критерий 3

Реализация на ЯП Python. Open-Source, больше реализаций и примеров.

Критерии	Аэрофото	Английский текст	Задачи классификации	Распознавание лиц	Пешеходы
Решение проблемы КЗ	+	+	+	+	+
Наличие разных методов	+	-	+	-	-
ЯП Python	-	-	+	-	-

Таблица 1 : Сравнение аналогов

Преимущество первого и третьего аналога перед другими является то, что они четко решают все проблемы и подходит по критериям. Так же недостатком первого и третьего аналога является их реализация не в ЯП Python и что они используют только не более двух сверточных методов. Третий аналог четко решает проблему КЗ но там кроме сверточного метода используется другие технологии одновременно, а так же статья в основном концентрируется на распознавании выражения, детали, и позу лица. Последний аналог обращает внимание на распознавание пешеходов в видео потоке, что конечно же, решает проблему КЗ, но в целом решает проблему другого направления.

Выбор метода решения

В результате обзора аналогов было выяснено, что разработано достаточное количество методов выполнения этой задачи. Так же было выявлено, что в аналогах не применяется разные сверточные методы распознавание объектов.

Следовательно, разрабатываемое решение должно представлять собой:

- Использование более чем двух сверточных методов.
- Анализ и сравнение этих методов
- Реализация на ЯП Python

Решение поставленных задач

Обзор и анализ существующих методов

Основная идея сверточной нейронной сети (CNN) исходит из рассмотрения того, как наш мозг работает с нейронами, связанными с другими нейронами, где один нейрон может активировать другой нейрон, [6]. Таким же образом строится CNN. Нейроны организованы в слои, связанные с другими слоями нейронов. Как эти нейроны связаны зависит от структуры каждого слоя и каждый слой имеет определенную функцию (т. е. сверточный слой, объединяющий слой, выравнивающий слой и т. д.). Способ объединения этих различных типов слоев для выполнения конкретной задачи называется нейронной сетью и для обработки изображений обычно используется множество сверточных слоев, поэтому и название: Сверточная Нейронная Сеть.

Детальный обзор слоев:

1. *Сверточный слой*: слои свертки характеризуются весами (значениями фильтра). Существует несколько сверток на слой с фиксированным размером, и каждое ядро применяется поверх все изображение с фиксированным шагом (stride). Первые слои свертки изучают низкоуровневые объекты такие как края, линии и углы. Следующие слои изучают более сложные представления (например, части и модельный). Чем глубже сеть, тем выше уровень изучаемых функций.
2. *Объединение слоев*: объединение слоев выполняет нелинейную понижающую дискретизацию. В этой категории, есть также несколько вариантов слоя, с максимальным объединением, являющимся самым популярным. Это в основном занимает фильтр (обычно размером 2x2) и шаг той же длины; затем применяет его к входному объему и выводит максимальное число в каждом субрегионе, вокруг которого свертывается фильтр. Таким образом количество параметров или весов уменьшено и переобучение проконтролирован.
3. *Активационные слои*: активационные функции имитируют поведение аксона нейрона, который запускает сигнал при предъявлении определенного стимула. Некоторые из наиболее распространенных активаций функции-это гиперболический тангенс[8], сигмоид[7] и выпрямленные линейные единицы.
4. *Выпадающие слои*: выпадающие слои " выпадают " случайный набор активаций путем установки их до нуля в прямом проходе. Это заставляет сеть быть избыточной; это означает, что сеть должна быть в состоянии обеспечить правильную классификацию или вывод для конкретного примера, даже если некоторые из активаций выпадают, что гарантирует, что сеть не становится слишком приспособлено к данным по тренировки и таким образом помогает разрешить проблему переобучение.
5. *Полностью подключенный слой* (ППС) отличается от вышеупомянутого слоев тем, что все выходы предыдущего слоя соединены со всеми входами слоями ППС. Эти слои могут быть математически представлены внутренними продуктами.

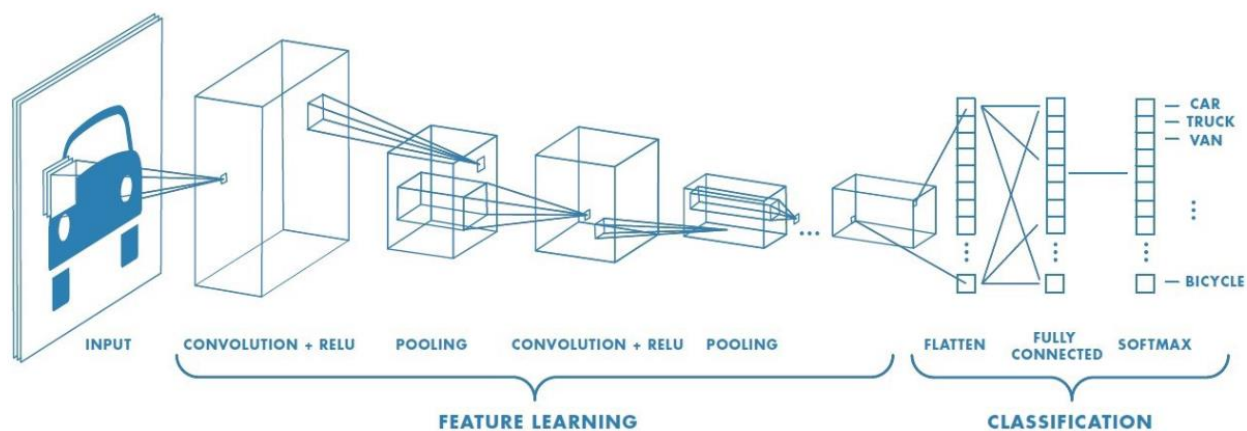


Рисунок 1 : Архитектура CNN

Анализ других методов:

Region Based Convolutional Neural Network (R-CNN) [2]

Сверточная нейронная сеть на основе региона. В R-CNN изображение сначала делится примерно на 2000 рекомендационных региона (предложений региона), а затем CNN применяется для каждого региона соответственно. Размер областей определяется, и правильная область вставляется в искусственную нейронную сеть. Самая большая проблема этого метода - время. Поскольку каждый регион на картинке применяется CNN отдельно, время обучения составляет примерно 84 часа, а время прогноза - примерно 47 сек.

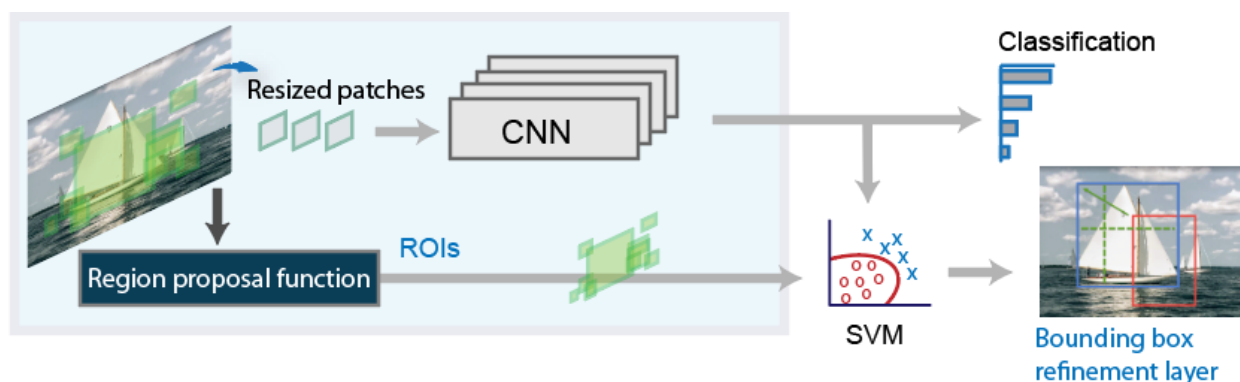


Рисунок 2 : Как работает R-CNN[2]

Fast on R-CNN.

Отличие этого метода от R-CNN заключается в том, что он сначала не разбивается на официальные рекомендации зон, а сначала применяет CNN, а затем распределяет его по рекомендациям зон на карте свойств. Кроме того, в части классификации метод использует классификацию softmax[2], которая является классификацией глубокого обучения, которая может быть реализована в искусственных слоях нейронной сети. Поскольку он использует CNN один раз, есть большой выигрыш во

времени. Время обучения составляет приблизительно 8,75 часа, а расчетное время - около 2,3 секунды.

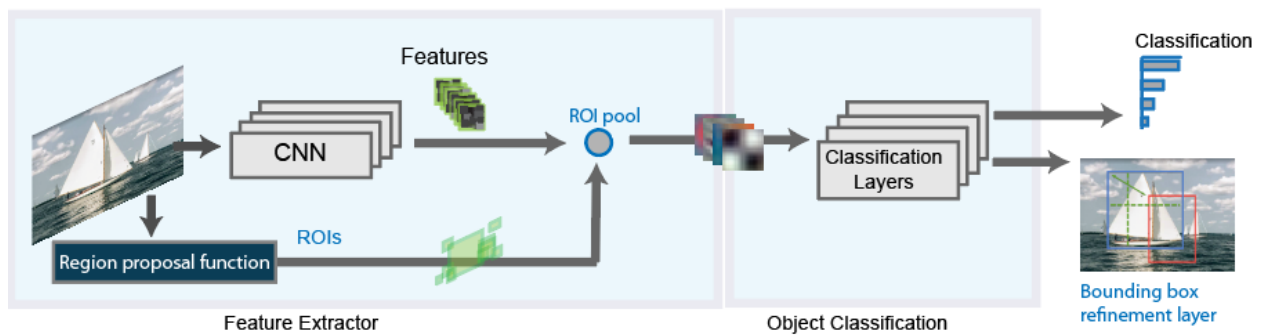


Рисунок 3: Как работает Fast R-CNN[2]

Faster on R-CNN.

В этом методе мы сначала применяем CNN к изображению, как в Fast R-CNN, и создаем карту объектов. С этого момента начинается изменение в соответствии с Faster R-CNN. Мы не участвуем в предложениях зон, создавая отдельную сеть предложений зон, но с выборочным поиском. Остальная часть выглядит практически так же, как, Fast R-CNN. Здесь, с учетом времени, которое мы получаем от регионального предложения, мы сокращаем время прогноза примерно до 0,3 секунды.

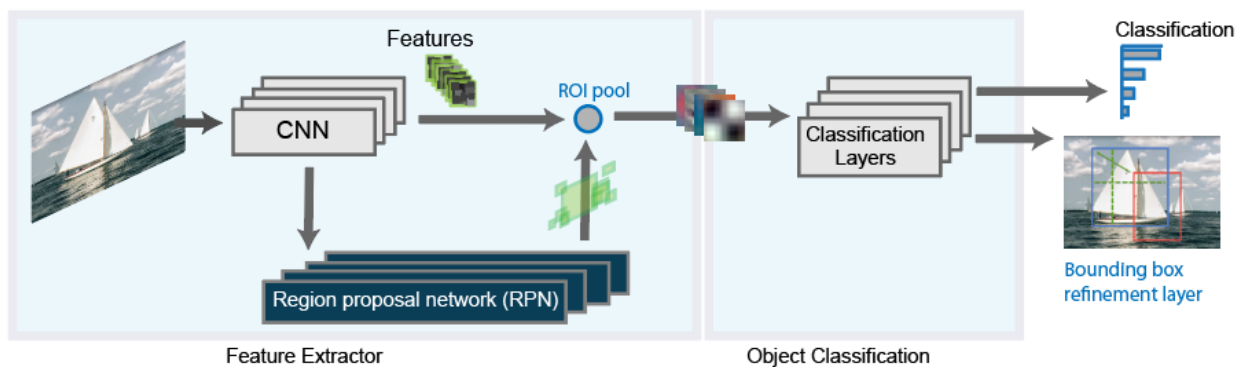


Рисунок 4: Как работает Faster R-CNN[2]

Резюме рассмотренных методов

Метод	Характеристики	Время прогнозирование	Пределы
CNN	Делит изображение на несколько областей и затем классифицирует каждую область на различные классы	–	Требуется много регионов для точного прогнозирования и, следовательно, высокого времени вычислений.
R-CNN	Использует выборочный поиск для создания регионов. Извлекает около 2000 областей из каждого изображения.	40-50 сек	Высокое время вычислений, поскольку каждый регион передается CNN отдельно, также он использует три разные модели для создания прогнозов.
Fast R-CNN	Каждое изображение передается только один раз в CNN, и карты объектов извлекаются. Выборочный поиск используется на этих картах для создания прогнозов. Объединяет все три модели, используемые в R-CNN вместе.	2-2,5 сек	Выборочный поиск выполняется медленно, и, следовательно, время вычисления все еще велико.
Faster R-CNN	Заменяет метод выборочного поиска сетью предложений по регионам, что значительно ускоряет работу алгоритма.	0,2 сек	Предложенная мысль требует времени, и поскольку существуют различные системы, работающие одна за другой, производительность систем зависит от того, как выполнялась предыдущая система

Таблица 2 : Сравнение методов

Заключение

Таким образом, в заключении можно сказать, что все рассмотренные методы уверенно решают проблему компьютерного зрения и с легкостью справляются с задачами обнаружения и локализации объектов на изображениях. При исследовании, так же было обнаружено что, Python лучше всего подходит для проектов машинного обучения и искусственного интеллекта, включая в себя простоту и согласованность, доступ к большим библиотекам и фреймворкам для искусственного интеллекта и машинного обучения (ML), гибкость, независимость от платформы и широкого общество.

Список литературы

1. Softmax
<https://ru.wikipedia.org/wiki/Softmax>
2. <https://www.mathworks.com/help/vision/ug/getting-started-with-r-cnn-fast-r-cnn-and-faster-r-cnn.html>
3. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
4. <https://www.analyticsvidhya.com/blog/2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1/>
5. Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016, <http://www.deeplearningbook.org>
6. <https://towardsdatascience.com/computer-vision-an-introduction-bbc81743a2f7>
7. Сигмоид
<https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%B3%D0%BC%D0%BE%D0%B8%D0%B4%D0%B0>
8. гиперболический тангенс
<https://ru.wikipedia.org/wiki/%D0%93%D0%B8%D0%BF%D0%B5%D1%80%D0%B1%D0%BE%D0%BB%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B5%D1%84%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D0%B8>
9. F.Li, J.Johnson and S.Young, Course notes from the course Neural Networks for Visual Recognition, Stanford University, 2017, <http://cs231n.github.io/>
10. Object Recognition in Aerial Images Using Convolutional Neural Networks Matija Radovic, Offei Adarkwa and Qiaosong Wang
11. Chatfield, K., Simonyan, K., Vedaldi, A., and Zisserman, A. Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531 (2014).