OBJECT DETECTION ДЛЯ АВТОМАТИЗАЦИИ ОБРАБОТКИ ДОКУМЕНТОВ

Козак А. В.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Теслюк В. Н. – кандидат физ.-мат. наук, доцент

В данной работе были исследованы и проанализированы различные методы распознавания объектов в условиях высокоструктурированных данных (на примере изображений документов). Также было проведено исследование потенциальных проблем, которые могут препятствовать качественной обработке и извлечению интересующей нас информации, а также их решения.

Компьютерное зрение – это область, которая в последние годы становится все более и более популярной (особенно начиная с развития CNN). И в последнее время центральное место заняли задачи с извлечением полезной информации из документов (изображений документов). Еще одной неотъемлемой частью компьютерного зрения является обнаружение (нахождение) объектов (Object Detection), которая играет важную роль для создания алгоритмов автоматической обработки документов. Обнаружение объектов помогает в оценке скелета объекта, обнаружении структуры документа, наблюдении и т. д. Разница между алгоритмами обнаружения объектов и алгоритмами классификации заключается в том, что в алгоритмах обнаружения мы пытаемся локализовать интересующий объект на изображении. Кроме того, мы можем находить не обязательно только один объект, а находить множество различных объектов. В задаче классификации мы можем только ответить на вопрос: есть ли на заданном изображение интересующий нас объект или нет, что накладывает ограничения на результат работы алгоритма.

Основная причина, по которой мы не можем использовать в решении этой задачи, построение стандартной сверточной сети с последующим полносвязанным слоем, заключается в том, что длина выходного слоя является переменной, а не постоянной, это связано с тем, что число вхождений

интересующих объектов не является фиксированным. Наивный подход к решению этой проблемы состоял бы в том, чтобы взять различные области интереса из изображения и использовать CNN для классификации присутствия объекта в этой области. Проблема такого подхода заключается в том, что интересующие объекты могут иметь различное пространственное расположение внутри изображения и различное соотношение сторон. Следовательно, вам придется выбрать огромное количество областей, а это может быть вычислительно взорвано. Для этого и были разработаны алгоритмы, такие как R-CNN, YOLO и другие, чтобы находить объекты быстро и эффективно.

Стандартная задача обнаружение объектов (т.е. локализация объекта и определение его категории) определяется следующим образом. По изображению целью обнаружения (нахождения, локализации) объекта состоит в том, чтобы определить, существуют ли экземпляры объектов из предопределенных категорий на изображении или нет и, если есть, вернуть пространственное расположение и размер каждого из найденных экземпляров. Больший акцент в задаче локализации делается на обнаружении широкого диапазона естественных категорий, в отличие от обнаружения специфических категорий объектов, где более узкая предварительно определенная категория нашего интереса может присутствовать (фактически одна категория содержит объект другой категории). Хотя тысячи объектов представляют визуальный мир, в котором мы живем, в настоящее время исследовательское сообщество в первую очередь заинтересовано в локализации высокоструктурированных объектов (например, автомобили, лица, велосипеды и самолеты) и сборных объектов (например, людей, коров и лошадей), нежели неструктурированные (например, небо, трава и облака).

В идеале задача обнаружения объектов заключается в разработке универсального алгоритма, который позволяет достичь двух конкурирующих целей: качество / точность и высокая эффективность. Высококачественное обнаружение объектов должно точно определять местоположение и распознавать объекты на изображениях (или видеокадрах), так чтобы можно было находить объекты, различных классов несмотря на большое разнообразие объектов (то есть иметь высокую степень различимости), и локализовать и узнавать экземпляры объектов одной и той же категории, учитывая внутриклассовые вариации (то есть обладать высокой надежностью). Высокая эффективность требует выполнения задачи обнаружения в режиме реального времени с приемлемыми требованиями к потребляемым ресурсам.

Задача автоматической обработки документов главным образом включает в себя нахождение областей интереса. А уже потом в извлечение информации (перевод в цифровой, обрабатываем и анализируемом) из областей интереса. Задача нахождения решается как правило задачей локализации объектов на изображение, а для извлечения информации (как правило представленной в буквенно-цифирном или текстовом виде) используются методы OCR, которые также используют алгоритмы распознавания объектов. Стоит отметить, что изображения документов (и, вообще говоря, документы в целом) являются высокоструктурированными данными. Таким образом, к таким данным необходимо применять соответствующие алгоритмы обработки-детекции. Последние время исследования показывают, что с задачей распознавания на высокоструктурированных изображениях справляются лучше всего алгоритмы локализации объектов семейства R-CNN, а именно R-CNN, Fast R-CNN, Faster R-CNN.

Алгоритм R-CNN (сокращено от Region-based Convolutional Neural Networks) был предложен в 2014 году Ross Girshick. Предложенный им алгоритм состоит из трех основных этапов. На первом этапе генерируются потенциальные классонезависимые области. Эти области представляют собой потенциальные распознаные объекты, то есть потенциальные предсказания детектора или области, содержащие необходимые объекты. Второй этап представляет собой большую глубокую сверточную нейронную сеть, которая генерирует фиксированной длины вектор признаков для каждой потенциальной области. И, наконец, третий этап – множество линейных SVM (для каждого класса свой классификатор), которые позволяют определить, что за объект находится в потенциальной области. Если внимательнее изучить этапы обучения R-CNN, мы можем легко обнаружить, что обучение модели R-CNN является дорогостоящим и медленным, так как все три этапа являются независимыми и требуют трудоемкой работы.

Чтобы ускорить R-CNN, Girshick усовершенствовал процедуру обучения, объединив три независимых модели в одну совместно обучающую структуру и ускорив общие результаты вычисления предсказаний, назвав полученный алгоритм Fast R-CNN. Вместо того, чтобы извлекать векторы признаков с помощью CNN независимо для каждого потенциальной области, полученной с помощью выборочного поиска, эта модель объединяет их в один прямой проход CNN по всему изображению, а уже потенциальные области совместно используют полученную после прохода матрицу признаков. Затем та же матрица признаков разветвляется для использования каждой областью для применения классификатора объекта и регрессора ограничивающего прямоугольника. Однако, выборочный поиск очень ресурсоемкий алгоритм.

Интуитивное ускоренное решение (т.е. улучшение Fast R-CNN) заключается в интеграции алгоритма генерации потенциальных регионов в саму модель CNN. Faster R-CNN делает именно это:

создает единую унифицированную модель, состоящую из RPN (сети предложения региона) и Fast R-CNN с общими сверточными слоями признаков. Фактически Faster R-CNN состоит из двух частей: RPN и Fast R-CNN. Фактически данный алгоритм использует недавно популярную терминологию нейронных сетей с механизмами «внимания», модуль RPN сообщает модулю Fast R-CNN, где стоит искать объект.

В своей работе я проанализировал работу каждого из алгоритмов семейства R-CNN на предмет качественного обнаружения интересующих областей в документах, то есть с областями данных необходимых для извлечения. Полученные результаты показали, что наиболее эффективный и качественный алгоритм для работы с высокоструктурированными данными является алгоритм Faster R-CNN. Это объясняется тем, что только в этой архитектуре используется обучаемый модуль по извлечению потенциальный областей, который потом используются для детектирования. Алгоритмы R-CNN и Fast R-CNN же используют выборочный поиск, который плохо показывает себя в работе со структурированными изображениями, так как является детерминированным алгоритмом, основанным на необучаемом алгоритме попиксельной иерархической сегментации.

Полученные результаты и реализованные методы были использованы в разработке ПО автоматической обработки и оценивания контрольных и / или проверочных работ студентов и / или учащихся в средних и / или высших учебных заведениях и исследовании применение методов машинного обучения в системе образования.

Список использованных источников:

1. Rich feature hierarchies for accurate object detection and semantic segmentation [Электронный ресурс]. — Режим доступа: https://arxiv.org/pdf/1311.2524.pdf

2. Fast R-CNN [Электронный ресурс]. — Режим доступа: https://arxiv.org/pdf/1504.08083.pdf

3. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [Электронный ресурс]. — Режим доступа: https://arxiv.org/pdf/1506.01497.pdf