

УДК 004.93`1

МЕТОДЫ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ РАСПОЗНАНИЯ ТЕКСТА ИЗ ИЗОБРАЖЕНИЙ

С.Н. МАЛАШОНОК

(Представлено: канд. техн. наук, доц. И.Б. БУРАЧЕНОК)

В данной статье описывается использование машинного зрения для распознавания текста, рассматриваются основные проблемы и подходы.

Многие компании имеют тысячи документов для обработки, анализа, преобразования и выполнения различных операций. Примерами таких документов являются квитанции, счета-фактуры, формы, выписки, контракты и многие другие фрагменты неструктурированных данных, и необходимо иметь возможность быстро распознать информацию, хранящуюся в неструктурированных данных, подобные этим.

Последние достижения в области компьютерного зрения позволяют добиваться больших успехов в облегчении процесса анализа и понимания документов.

Задача извлечения текста из изображений документов называется оптическим распознаванием символов (OCR – Optical Character Recognition) и является предметом многих исследований. OCR является одной из самых ранних задач компьютерного зрения, поскольку в некоторых аспектах не требует глубокого обучения. Когда документы четко размечены и имеют определенную структуру (например, деловое письмо), существующие инструменты для OCR могут работать достаточно хорошо. Популярным инструментом для распознавания текста с открытым исходным кодом является проект Tesseract, который изначально разрабатывался Hewlett-Packard, но в последние годы находится под поддержкой Google. В последнее время провайдеры облачных услуг внедряют возможности обнаружения текста наряду с их другими предложениями, что делает методы компьютерного зрения широко доступными. Среди них GoogleVision, AWS Textract, Azure OCR и Dropbox. [1]

Однако есть много случаев использования, которые можно назвать нетрадиционным OCR, когда существующие общие решения не подходят. К таким случаям можно отнести обнаружение текста, который может находиться на дорожных знаках, номерах домов, рекламных объявлениях и т.д. Изображения могут включать в себя разделение фона, различные ориентации и длины текста, разнообразие шрифтов, отвлекающие объекты.

Другой случай, который создает аналогичные проблемы, связан с извлечением текста из изображений сложных документов. В отличие от документов с определенным макетом (таких как страница из книги, столбец из газеты), многие типы документов относительно неструктурированы по макету и имеют разбросанные по всему тексту элементы (такие как квитанции, формы, и счета). Эти изображения характеризуются сложным расположением текстовых элементов, разбросанных по всему документу и окруженных множеством «отвлекающих» объектов. В этих изображениях основная проблема заключается в правильном сегментировании объектов в изображении для идентификации нужных текстовых блоков.

На основании этих двух случаев можно выделить следующие критерии задач OCR:

- плотность текста – на печатной или письменной странице плотность текста будет велика в сравнении с текстом на изображении улицы;
- структура текста – текст на странице может располагаться в соответствии с каким-либо макетом или может располагаться в хаотичном порядке и иметь различные углы поворота;
- шрифты – печатные шрифты проще, так как они более структурированы по сравнению с рукописными символами;
- артефакты – изображения улиц имеет большее количество шумов, чем изображения страниц книг [2].

Основными этапами распознавания текста являются обнаружение текста на изображении и распознавание обнаруженного текста с помощью одного из следующих подходов: классические методы компьютерного зрения, специализированное глубокое обучение или стандартный подход глубокого обучения.

Классический методы компьютерного зрения включают в себя:

- применение различных фильтров для выделения символов из фона;
- распознавание контура каждого отдельного символа;
- применение классификации контуров символов для идентификации символов.

Обнаружение контура является довольно сложной задачей для обобщения, поэтому требует много ручной тонкой настройки. Классификация контура может быть реализована либо с помощью сопоставления с образцом, либо с помощью машинного обучения.

К специализированному глубокому обучению относится сверточная рекуррентная нейронная сеть (CRNN – Convolutional Recurrent Neural Network). Эта архитектура нейронной сети объединяет функции извлечения, моделирование последовательности и транскрипцию в единую структуру. Эта модель не нуждается в сегментации символов. Сверточная нейронная сеть извлекает элементы из входного изображения (области обнаружения текста). Глубокая двунаправленная рекуррентная нейронная сеть предсказывает последовательность меток с некоторым соотношением между символами. Уровень транскрипции преобразует кадр, созданный рекуррентной нейронной сетью, в последовательность меток. Существует два способа транскрипции: безлексиконный и основанный на лексиконах. В подходе, основанном на лексиконе, будет предсказана наиболее вероятная последовательность меток [3].

Таким образом, стандартные подходы глубокого обучения, такие как SSD, YOLO и Mask RCNN, используются для обнаружения символов и слов. Однако они часто не достигают желаемой точности, и поэтому необходимы специальные подходы.

Вышеизложенные данные поспособствуют разработке мобильного приложения для учета и планирования расходов посредством сканирования чеков. Актуальность данного приложения заключается в введении статистики по различным видам и категориям товаров с целью контроля и оптимизации финансовых затрат.

ЛИТЕРАТУРА

1. Learning to Read: Computer Vision Methods for Extracting Text from Images [Электронный ресурс]. – Режим доступа: <https://medium.com/capital-one-tech/learning-to-read-computer-vision-methods-for-extracting-text-from-images-2ffcdae11594>. – Дата доступа: 24.09.2019.
2. A gentle introduction to OCR [Электронный ресурс]. – Режим доступа: <https://towardsdatascience.com/a-gentle-introduction-to-ocr-ee1469a201aa>. – Дата обращения: 24.09.2019.
3. Deep Learning Based OCR for Text in the Wild [«ktrnhjyysq htcehc/ – Ht;bv ljcnegf^ https^|nanonets/com|blog|deep-learning-ocr|#text-detection/ – Lfnf ljcnegf^ 24/09/2019/.