

Earthquake-resistant construction. 1979, Issue 5, pp. 1-3.

2. Kilimnik L.Sh., Soldatova L.L., Lyakhina L.I. Analysis of the operation of buildings with a sliding belt using a multi-mass calculation model. Structural mechanics and calculation of structures. 1986, No. 6, pp. 96-103.

3. Recommendations for the design of buildings with a seismically insulating sliding belt and dynamic vibration dampers. - М.: TSNIISK them. Kucherenko, NIIOSP them. Gersanova, 1984, 55p.

4. Polyakov V.S., Kilimnik L.Sh., Cherkashin A.V. Modern methods of seismic protection. - М.: Stroyizdat, 1989, -320 p., Pp. 249-250.

5. Davydova G.V., Ermoshin A.A., Uzdin A.M., Rummyantsev A.Yu. Assessment of the movements of buildings with a seismically insulated sliding belt.

Earthquake-resistant construction. Safety of facilities. No. 3, 2007, pp. 34-35.

6. Patent No. 90, of the Kyrgyz Republic / Seismic isolating support // Mamatov Zh.Y., Toktonasarov Zh.M., Matozimov BS, Andashev A. Zh.

7. Patent No. 91, of the Kyrgyz Republic / Seismic isolating support // Mamatov Zh.Y., Toktonasarov Zh.M., Matozimov BS, Andashev A. Zh.

8. Patent No. 92, of the Kyrgyz Republic / Earthquake-resistant house // Mamatov Zh.Y., Toktonasarov Zh.M., Matozimov BS, Andashev A. Zh.

9. Abdybaliev M.K. The results of the seismic reaction of buildings on sliding supports. Republican scientific-practical conference. The influence of regional climatic factors on the organizational, technical and economic features of construction and the Kyrgyz SSR. Abstracts of reports. 1989.

РЕШЕНИЕ ЗАДАЧИ РАСПОЗНАВАНИЯ СИМВОЛОВ НА ИЗОБРАЖЕНИЯХ ТОВАРНЫХ ЦЕННИКОВ

Марков Виталий Владиславович

*студент, Челябинский государственный университет,
Россия, г. Челябинск*

АБСТРАКТ

В работе сравниваются модель STAR-net и технология tesseract-ocr. Датасет представляет собой вырезанные слова с магазинных ценников, размеченных в автоматическом режиме с помощью сервиса Yandex OCR. Использовались такие метрики как доля правильных ответов и расстояние левенштейна. Наилучший результат показала модель STAR-net. Была проведена аналитика ошибок и даны рекомендации по улучшению качества.

Ключевые слова: OCR, задача распознавания символов, tesseract-ocr, STAR-net, расстояние левенштейна.

Введение

Одной из классических задач компьютерного зрения является задача распознавания символов на изображениях. В первую очередь, это связано с большим количеством различных приложений таких решений: начиная от разбора изображений документов, заканчивая капчей в интернете. Одним из таких приложений является чтение информации с изображений магазинных ценников. Такая модель может быть частью системы по сбору информации с ценников, позволяя автоматизировать процесс заполнения информации к изображению. Основными трудностями в решении данной задачи является разнообразие шрифтов, фонов, размеров символов, различные сокращения слов, редкие символы. Иногда, для правильного распознавания символов, необходимо знать некоторую семантику.

Обзор

Использовать подходы глубокого обучения в задаче распознавания символов начали относительно недавно. В рамках данной задачи существует несколько датасетов, на которых проверяется качество моделей. В качестве примеров таких датасетов можно привести следующие: ICDAR 2013[8], ICDAR 2015[9], Street View Images[10] и т.д.. Данные в таких датасетах представляют собой изображения и разметках к ним, где разметка это координаты слово и само слово.

В задаче распознавания символов на изображении существует несколько подходов. Одним из типов подходов заключается в тренировки end-to-end модели, которая решает не только задачу распознавания символов, но и нахождение их. В качестве примера можно привести модель FOTS[5], архитектура которой состоит из двух пайплайнов. Оба пайплайна основаны на сети ResNet[3], при этом первый пайплайн занимается решением задачи нахождения текста, а второй пайплайн, основываясь также на признаках, полученных из первого пайплайна, решает задачу распознавания символов. В пайплайне распознавания символов используется рекуррентная нейронная сеть с функционалом качества CTC[2]. Такой подход, как правило, требует больше данных, дольше и труднее обучается, но позволяет получить качество на каждой подзадаче выше, в отличие если бы модели обучались по отдельности.

В качестве второго типа подходов можно выделить подходы, при которых используется энкодер в виде сверточной нейронной сети и декодера в виде рекуррентной нейронной сети. При этом, как правило, в качестве функционала качества используют CTC. Данные для таких подходов представляют собой уже вырезанные слова, а разметка здесь это текст, который содержится на изображении. Одной из моделей

такого подхода является STAR-net[4]. Архитектура сети состоит из 4-х этапов: обработка изображения TPS[11], энкодер в виде сверточной нейронной сети, декодер в виде BiLSTM и функционал качества CTC. Преимуществом таких подходов заключается в том, что нет необходимости в разметке каждого символа, получение которой может быть достаточно трудозатратно.

Также существуют открытые решения с уже обученными моделями. В качестве примера такого решения можно привести tesseract ocr[7]. Tesseract-ocr способен распознавать более 100 различных языков, а также работать с разными видами текста. Также tesseract можно дообучить на своих данных. При этом стоит заметить, что tesseract хоть и является открытым решением, но из-за того, что tesseract предоставляет распознавание большого количества языков, может работать с низким качеством из-за специфики конкретного датасета.

Помимо открытых решений, также есть облачные решения. Одно из таких решений

предоставляет Google Cloud Platform. Как правило, такие решения натренированы на большом объеме данных, из-за чего они способны распознавать большое количество различных языков и символов. Однако, облачные технологии накладывают некоторые ограничения на возможности системы. Так, распознавание не будет работать в режиме реального времени, а если количество изображений огромно, то это может дорого стоить.

В рамках данной работы решено было сравнить модель STAR-net и tesseract: tesseract выступает в роли базового решения, STAR-net как достаточно легкая модель распознавания символов.

Данные

Датасет представляет собой около 12000 изображений ценников, которые были предварительно размечены с помощью сервиса Yandex OCR. Затем, размеченные слова вырезались. Таким образом получилось собрать 141049 изображений слов. Пример изображения можно увидеть на рисунке 1.



Рисунок 1. Пример изображения из тестовой выборки.

В дальнейшем был проведен первичный анализ данных, в ходе которого выяснилось, что в данных присутствует неравномерность частот в словах. Редкие слова решено было не удалять, так как они состояли из цифр, то есть не было повторяющихся паттернов, которые могли бы привести к переобучению. С другой стороны, количество самых частотных слов было уменьшено, чтобы уменьшить переобучение на отдельные слова. Также, были исключены слова, в которых присутствовали редкие символы, чтобы сократить алфавит. Количество слов, в которых были редкие символы, было меньше 1% от всего датасета. После обработки получилось 108838 изображений, из них 88584 изображений - обучающая выборка, а 20254 изображений - тестовая. А размер алфавита составил 90 символов, включая строчные символы русского и английского алфавитов. Средняя высота изображения составляет 33 пикселя, а средняя ширина 101 пиксель.

Эксперимент

В эксперименте решено было сравнить две модели:

tesseract-ocr (<https://tesseract-ocr.github.io/tessdoc/>)

STARnet (<https://github.com/clovaai/deep-text-recognition-benchmark>).

В качестве метрик были выбраны следующие: доля правильно распознанных слов и расстояние Левенштейна. Обучение модели STAR-net проходило с базовыми аргументами, но был заменен алфавит на алфавит, полученный из первичного анализа данных. Результаты tesseract-ocr на тестовом датасете были получены с параметрами `--lang=rus+eng` и `--oem=1`, а параметр `--psm` перебирался среди следующего множества {10, 13}. Такие значения параметра `--psm` были выбраны, так как они лучше всего подходили к датасету. Значения метрик представлены в таблице 1.

Таблица 1.

Значения метрик на тестовом датасете.

| | tesseract-ocr (psm 11) | tesseract-ocr (psm 13) | STAR-net |
|----------------------|------------------------|------------------------|----------|
| accuracy | 0.3627 | 0.3650 | 0.8085 |
| levenshtein distance | 2.9988 | 2.9769 | 0.5215 |

Заключение

В данной работе был проведен обзор существующих подходов к распознаванию текста на изображениях. В результате был поставлен эксперимент по сравнению двух моделей на вырезанных словах с товарных ценников: STAR-net и tesseract-ocr. По полученным метрикам видно, что наилучший результат показала модель STAR-net. Анализ ошибок показал, что если модель и ошибается, то делает чаще всего это в одном символе, что может говорить о высоком результате распознавания. Также были замечены изображения, на которых разметка не соответствовала действительности. При этом tesseract-ocr часто путал схожие символы, а также не находил часть текста. В качестве улучшения можно предложить попробовать другие параметры обучения модели STAR-net. Для tesseract-ocr же можно попробовать препроцессинг данных, а также дообучить его на датасете.

Список литературы

1. Baek, Jeonghun, et al. "What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis." 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, doi:10.1109/iccv.2019.00481.
2. Graves, Alex, et al. "Connectionist Temporal Classification." Proceedings of the 23rd International Conference on Machine Learning - ICML '06, 2006, doi:10.1145/1143844.1143891.
3. He, Kaiming, et al. "Deep Residual Learning for

Image Recognition." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, doi:10.1109/cvpr.2016.90.

4. Liu, Wei, et al. "STAR-Net: A SpaTial Attention Residue Network for Scene Text Recognition." Proceedings of the British Machine Vision Conference 2016, 2016, doi:10.5244/c.30.43.

5. Liu, Xuebo, et al. "FOTS: Fast Oriented Text Spotting with a Unified Network." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, doi:10.1109/cvpr.2018.00595.

6. Stanford cs class cs231n: Convolutional neural networks for visual recognition. <http://cs231n.github.io/neural-networks-case-study/>.

7. Tesseract OCR, <https://tesseract-ocr.github.io/tessdoc/>

8. Karatzas, Dimosthenis, et al. "ICDAR 2013 Robust Reading Competition." 2013 12th International Conference on Document Analysis and Recognition, 2013, doi:10.1109/icdar.2013.221.

9. Karatzas, Dimosthenis, et al. "ICDAR 2015 Competition on Robust Reading." 2015 13th International Conference on Document Analysis and Recognition (ICDAR), 2015, doi:10.1109/icdar.2015.7333942.

10. Wang, Kai, et al. "End-to-End Scene Text Recognition." 2011 International Conference on Computer Vision, 2011, doi:10.1109/iccv.2011.6126402.

11. Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In NIPS, pages 2017–2025, 2015.

ПРОВЕДЕНИЕ СРАВНИТЕЛЬНОГО АНАЛИЗА ATTENTION OCR И TESSERACT В ЗАДАЧЕ РАСПОЗНАВАНИЯ СИМВОЛОВ НА ИЗОБРАЖЕНИЯХ ПРЕЙСКУРАНТОВ.

Марков Андрей Владиславович

студент,

Челябинский государственный университет,

Россия, г. Челябинск

АБСТРАКТ

Решения классической задачи распознавания символов является высоко востребованной на практике. В рамках данной работы будет рассматриваться задача распознавания символов с изображений прайс-листов табачной продукции. Для разметки изображений использовался сервис Yandex OCR. Сравнилась модель Attention OCR и технология Tesseract по качеству распознавания изображений слов, вырезанных с прайс-листов. Attention OCR показала более качественное распознавание символов по сравнению с Tesseract.

Ключевые слова: Tesseract, Attention OCR, прайс-листы.

Введение

Задача распознавания символом является одной из базовых задач компьютерного зрения. Отличительной особенностью данной задачи состоит в разнообразии данных. Текст может быть представлен различными символами, языками, иметь разный шрифт, фон, размер, а также ориентацию в пространстве. Актуальность же данной задачи состоит в большом практическом значении в областях, где решение данной задачи позволяет автоматизировать процесс сбора информации с изображений. В рамках данной работы будет рассматриваться задача

расознавания символов с изображений сигаретных прайс-листов. Решение задачи распознавания символом на этих данных полезна для автоматического сбора различной информации, в том числе цен, представленных на прайс-листах табачной продукции. Данная информация может быть полезна для дальнейшего исследования полноценного end-to-end решения для сбора информации с изображений прайс-листов.

Обзор

Одно из первых открытых решений задачи распознавания символов является технология Tesseract[5] (<https://tesseract-ocr.github.io/tessdoc/>).