

Сравнение нейросетевых и непрерывно-морфологических методов в задаче детекции текста (Text Detection)

Гайдученко Н.Е., Труш Н.А., Торлак А.В., **Миронова Л.Р.**, **Акимов К.М.**, Гончар Д.А.

October 20, 2018

Анализ ошибки

Наша задача - минимизация функционала потерь. Одной из метрик оценки ошибки задачи классификации является Intersection over Union. Назовем множество рамок с текстом, которые мы ожидаем получить на выходе алгоритма ожидаемым. После работы алгоритма мы получили другое множество рамок, заданных координатами: $x1_i, x2_i, y1_i, y2_i$, по которым мы можем вычислить площади: ожидаемых - $S1_i$, полученных - $S2_i$ и площади их пересечений. Функция вычисляется как отношение площади перекрытия рамок ожидаемой и полученной к площади их объединения:

$$IoU = \sum((S1_i \cap S2_i) / (S1_i \cup S2_i)). \quad (1)$$

Две другие метрики - precision и recall. Мы знаем правильные ответы для задачи классификации изображений на текст (положительный класс) и не текст (отрицательный класс). Есть алгоритм, который определяет, является ли изображение текстом, то есть относит его или к положительному, или к отрицательному классу. Тогда определим следующие величины:

tp- количество изображений, правильно отнесенных к тексту;

tn- количество правильно не отнесенных к тексту;

fp- количество неправильно отнесенных к тексту;

fn- количество неправильно не отнесенных к тексту.

$$P(\text{precision}) = \frac{tp}{tp + fp}, \quad R(\text{recall}) = \frac{tp}{tp + fn} \quad (2)$$

Мера точности характеризует, сколько полученных от классификатора положительных ответов являются правильными. Чем больше точность, тем меньше число ложных попаданий. Мера полноты - способность классификатора «угадывать» как можно большее число положительных ответов из ожидаемых.

Precision и Recall дают характеристику классификатора с разных сторон, и увеличение одной из них приводит к уменьшению другой. Поэтому удобно использовать метрику F1, их среднее гармоническое:

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \quad (3)$$

Выбор модели

По сравнению с другими моделями, Text Detect оказалась самой результативной. Однако, требует наибольшее количество времени на вычисление среди вышеописанных алгоритмов. В среднем 1,5 мин на изображение. Объясняется это длительным поиском экстремальных регионов и с последующей их обработкой.



Заключение

В рамках данной статьи были получены следующие результаты: Рассмотрена область машинного обучения по извлечению признаков. Реализованы несколько базовых алгоритмов для обнаружения текста на изображении, проведен их сравнительный анализ и поставлен ряд экспериментов. Было произведено сравнение работы нескольких моделей: непрерывно морфологических, а также нейросетевых. В ходе эксперимента была замечена зависимость успешности каждого метода от используемого набора данных. Дальнейшие действия направлены на совершенствование архитектуры модели и исследование ее поведения на других данных.

Список литературы

- [1] Е. О. Захаров, Л. М. Местецкий. *Сегментация текстовых блоков в изображениях рукописных документов*. 2016.
- [2] Q. Ye, D. Doermann. *Text Detection and Recognition in Imagery: A Survey*. IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 37, Issue: 7, July 1 2015).
- [3] D. Aldavert, M. Rusiñol. *Manuscript Text Line Detection and Segmentation Using Second-Order Derivatives*. 2018 13th IAPR International Workshop on Document Analysis Systems (DAS).
- [4] Zh. Liu, Y. Li, X. Qi, Y. Yang, M. Nian, H. Zhang, R. Xiamixiding. *Method for unconstrained text detection in natural scene image*. IET Computer Vision (Volume: 11, Issue: 7, 10 2017).
- [5] T. Wang, David J. Wu, A. Coates, A. Y. Ng. *End-to-End Text Recognition with Convolutional Neural Networks*.
- [6] M. Liao, B. Shi, X. Bai. *A Single-Shot Oriented Scene Text Detector*. arXiv:1801.02765v3 [cs.CV] 27 Apr 2018.
- [7] Zh. Tian, W. Huang, T. He, P. He Y. Qiao. *Detecting Text in Natural Image with Connectionist Text Proposal Network*. arXiv:1609.03605v1 [cs.CV] 12 Sep 2016.