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.
Gaiduchenko.NE@gmail.com

В данной работе рассматривается задача детекции текста на фотографиях документов. Приведён сравнительный анализ современных нейросетевых архитектур (CTPN, EAST и др.) и непрерывно-морфологических методов (Auto Canny, Hough Lines, MSER и др) обучения без учителя. Модели протестированы на синтетически-сгенерированных и реальных выборках с различными функциями ошибки. Предложен алгоритм, основанный на использовании как нейросетевых, так и непрерывно-морфологических методов в зависимости от специфики задачи.

Ключевые слова: *нейронные сети, непрерывно-морфологические методы, распознование текста, распознование изображений обнаружение текста на изображении, детекция текста, морфологические методы, нейронные сети для обнаружения текста, обучение без учителя, анализ текстов.*

1 Введение

Особое место в списке активно исследуемых областей заслуживает распознавание текстов на изображениях. Обычно данная задача разбивается на две более конкретных: обнаружение ограничивающих прямоугольников для текста на изображении (детекция текста, text detection) и распознавание текста внутри ограничивающего прямоугольника (OCR, text recognition).

Решением задачи детекции текста, иногда называемой локализацией текста, является алгоритм, принимающий на вход изображение и возвращающий координаты мест, содержащих текст (как правило, координаты углов ограничивающих прямоугольников). Последние достижения в области глубокого обучения и распознавания объектов на изображениях обеспечили появление всё более точных и эффективных алгоритмов детекции текста, например CTPN [4], TextBoxes++ [2], SegLink [3], EAST [5] and PixelLink [1]. Большинство из этих state-of-the-art методов построены на использовании свёрточных нейронных сетей и используют два подхода:

Классификация "текст / не текст". Такие предсказания выносятся чаще всего в виде вероятности для каждого пиксела, что он принадлежит или не принадлежит ограничивающему прямоугольнику с текстом. Но они зачастую используются и в процессе регрессии (например в TextBoxes, SegLink, EAST, PixelLink).

Регрессия местонахождения. Местонахождение текстовых объектов предсказываются как отступ от их собственных ограничивающих прямоугольников (например, TextBoxes, SegLink, CTPN) или в виде абсолютных координат углов ограничивающих прямоугольников (EAST).

*Работа выполнена при финансовой поддержке РФФИ, проект №00-00-00000. Научный руководитель: Стрижов В. В. Задачу поставил: Жариков И. Н. Консультант: Местецкий Л М.

Существуют также методы детекции текста (например, SegLink), предсказывающие также связи между сегментами текста. После предсказаний проводится операция постобработки, в основном включающая в себя объединение сегментов (SegLink, CTPN) или подавление немаксимумов (TextBoxes, EAST) для формирования ограничивающих прямоугольников в качестве итогового результата.

2 Постановка задачи

Главной задачей данной работы является сравнительный анализ различных методов обнаружения текста, обученных на различных датасетах в рамках задачи детекции текста на документах. Входными данными для задачи служат алгоритмы, использующие нейросетевые методы глубокого обучения или непрерывно-морфологические методы. Выходными данными служит трёхмерная таблица "модели-выборки-метрики ошибки иллюстрирующая наглядно, какие модели лучше всего работают на каких выборках.

Каждая модель решает задачу обнаружения текста. Ставится задача определения координат ограничивающих прямоугольников для слов или сегментов текста на реальных изображениях и отсканированных документах. Эта задача является частью более общей проблемы детекции сегментов текста (строк, абзацев), которые впоследствие разбиваются на слова. В качестве исходных данных рассматриваются изображения, содержащие текст. Выходными данными служат координаты ограничивающих прямоугольников для текстовых сегментов. Данные прямоугольники (bounding boxes) можно визуализировать, нарисовав их на исходном изображении для наглядности.

3 Описание алгоритма

Происходит поиск и сбор данных для обучения. Для упрощения данного этапа работы используются синтетические выборки (например SynthText), позволяющие быстро получить большие выборки данных, экономя время и Интернет-трафик на скачивание больших объёмов данных.

Подготавливаются алгоритмы моделей детекции текста, происходит сопряжение форматов синтетических выборок с форматом данных на вход для каждой модели.

Алгоритмы детекции текста обучаются на синтетических выборках. Затем тестируются на выборках для тестирования - реальных или синтетических выборках с документами. Таким образом каждая модель получает оценку - величину ошибки на данной выборке.

Строится трёхмерная таблица сравнения результатов. По оси 0 откладываются использованные модели, по оси 1 откладываются выборки, на которых данные модели были обучены. По оси 2 - величина ошибки на тестовом датасете.

Производится анализ таблицы, результаты которого описаны в данной работе.

4 Анализ ошибки

Публике доподлинно известно, что точность и полнота не дают достаточно адекватные оценки ошибок модели. Поэтому, требуется использовать метрику, использующую в себе информацию о точности и полноте алгоритма в совокупности. F-мера представляет собой гармоническое среднее между precision и recall. Она стремится к нулю, если точность или полнота стремится к нулю.

$$F = 2\frac{Precision \cdot Recall}{Precision + Recall}$$

Данная формула придает одинаковый вес точности и полноте, поэтому F-мера будет падать одинаково при уменьшении и точности и полноты. Возможно рассчитать F-меру придав различный вес точности и полноте:

$$F = (\beta^2 + 1)\frac{Precision \cdot Recall}{\beta^2 Precision + Recall}$$

где β принимает значения в диапазоне $0 < \beta < 1$ если необходимо отдать приоритет точности, а при $\beta > 1$ приоритет отдается полноте. При $\beta = 1$ формула сводится к предыдуще и получается сбалансированная F-мера (также ее называют F1).

F-мера является хорошим кандидатом на формальную метрику оценки качества классификатора. Она сводит к одному числу две других основополагающих метрики: точность и полноту. Имея в своем распоряжении подобный механизм оценки вам будет гораздо проще принять решение о том являются ли изменения в алгоритме в лучшую сторону или нет.

Альтернативной метрикой оценки ошибки задачи детекции текста является Intersection over Union. Прогнозируемый ограничивающий прямоугольник изображается красным цветом, а ограничивающий прямоугольник с истинными (ground truth) координатами рисуется зеленым цветом. Функция вычисляется как отношение площади перекрытия этих рамок к площади их объединения:

$$IoU = \frac{AreaofOverlap}{AreaofUnion}$$

5 Результаты

Сравнение моделей на датасете Total-Text

Сравнение моделей на датасете Total-Text			
Модель	Precision	Recall	F-мера %
EAST VGG16	50.0	36.2	42.0
SegLink	30.3	23.8	26.7
TextBoxes	**62.1**	**45.5**	**52.5**

Сравнение моделей на датасете ICDAR 2015			
Модель	Precision	Recall	F-мера %
EAST	80.5	72.8	76.5
CTPN	74.2	51.6	60.9
TexbBoxes++	**87.2**	76.7	81.6
PixelLink 2s	85.5	**82.0**	**83.7**
SegLink	73.1	76.8	74.9

Литература

[1] Dan Deng, Haifeng Liu, Xuelong Li, and Deng Cai. Pixellink: Detecting scene text via instance segmentation.

[2] Minghui Liao, Baoguang Shi, and Xiang Bai. Textboxes++: A single-shot oriented scene text detector.

[3] Baoguang Shi, Xiang Bai, and Serge Belongie. Detecting oriented text in natural images by linking segments.

[4] Zhi Tian, Weilin Huang, Tong He, Pan He, and Yu Qiao. Detecting text in natural image with connectionist text proposal network.

[5] Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, and Jiajun Liang. East: An efficient and accurate scene text detector.