АЛГОРИТМЫ ДЕТЕКТИРОВАНИЯ ТЕКСТОВЫХ ОБЛАСТЕЙ НА ФОТОРЕАЛИСТИЧНЫХ ИЗОБРАЖЕНИЯХ СО СЛОЖНЫМ ФОНОМ

А.А. Друки, к.т.н., доцент,
М.В. Лазуков, студент гр. 8В7Б
Томский политехнический университет
E-mail: mvl16@tpu.ru

Введение

Задача детектирования текстовых областей на изображения привлекала исследователей ещё до появления первых сверточных нейронных сетей. С развитием цифровых технологий, портативных мобильных устройств, а также сети Интернет появляется все больше областей применения алгоритмов извлечения текстовой информации на изображениях. В связи с чем, в настоящее время детектирование текстовых областей является актуальной задачей компьютерного зрения, позволяющая получать важную информацию о семантике изображения.

Таким образом, целью работы является аналитический обзор современных алгоритмов детектирования текстовых областей, что является начальным этапом по созданию алгоритма распознавания текста на изображениях со сложным фоном.

Традиционные методы детектирования текстовых областей

Традиционные методы обнаружения текстовых областей можно разделить на два подхода:

- Методы, основанные на областях, зачастую используют технику скользящего окна для выбора областей-кандидатов [1], которые проверяются на наличие текста с помощью предобученного классификатора. Главными недостатками данного подхода является низкая точность на изогнутых текстовых областях или же областях, расположенных под углом.
- Методы, основанные на анализе компонентов связности, сначала извлекают компоненты кандидаты по схожим признакам [2], например, по таким признакам как: угловые точки, текстура текста, границы текста и т.п. После чего используются вручную сформированные правила или же предобученные классификаторы для фильтрации нетекстовых компонент. К недостаткам метода можно отнести низкое качество работы на изображениях со сложным фоном.

Современные подходы на основе глубоких нейронных сетей

Методы глубокого обучения позволяют сетям автоматически извлекать признаки из текста в результате обучения. Решения, основанные на глубоких сетях, зачастую являются более простыми и эффективными в сравнении с созданными вручную алгоритмами. Бурное развитие глубоких сетей привело к появлению подходов, успешно решающих задачу детекции текстовых областей, которые, можно классифицировать на 3 группы [3]:

1. Методы, основанные на регионах [4], обычно состоят из 2-х частей: классификатора и регрессора для областей-кандидатов. Данные методы имеют довольно высокую точность и полноту обнаружения текстовых областей. В то же время они опираются на сложные конструкции рамок и требуют больших вычислительных ресурсов.
2. Методы, основанные на сегментации [5] для классификации текстовых/не текстовых областей на уровне пикселей, стали основным направлением для обнаружения текста с различной ориентацией и произвольной формой. Однако, данная группа методов часто требует трудоемкой и сложной постобработки для решения сложных случаев.
3. Гибридные методы зачастую являются комбинацией вышеописанных методов с добавлением инновационных идей для решение специфичных задач.

Наборы данных

На данный момент существует большое количество размеченных наборов данных, содержащих как реальные изображения текстовых областей, так и сгенерированные изображения. Текст может быть написан на различных языках, иметь различный цвет, шрифт и ориентацию. В таблице 1 приведены сведения о некоторых распространенных наборах данных.

Таблица 1. Наиболее распространенные наборы данных

Название	Язык	Число изображений в выборке		
		Общее	Тренировочная	Тестовая
MSRA-TD500	Английский/ Китайский	500	300	200
ICDAR 2015	Английский	1500	1000	500
COCO-Text	Английский	63686	43686	20000
SynthText	Английский	858750	-	-
ICDAR 2019 MLT	Мультиязычный	20000	10000	10000
ICDAR 2019 LSVT	Английский/ Китайский	450000	430000	20000

Метрики для оценки качества работы

В качестве основных характеристики для оценки качества работы алгоритмов детекции текстовых областей используется точность (Precision, P), полнота (Recall, R) и общий оценочный индекс (F-measure, F), которые рассчитываются следующим образом:

$$\mathrm{R}ecall(G,D) = \frac{\sum_{i=1}^{|G|} BestMatch_G(G_i)}{|G|},$$

$$Precision(G,D) = \frac{\sum_{i=1}^{|D|} BestMatch_D(D_i)}{|D|},$$

$$f = 2\,\frac{Recall * Precision}{Recall + Precision},$$

где G – множество прямоугольников, достоверно описывающих расположение текста на изображении, D – множество предсказанных прямоугольников, полученных в результате работы алгоритма.

Функции *BestMatch* определены следующим образом:

$$BestMatch_G(G_i) = \max_{j=1\ldots|D|} \frac{2\,Area(G_i \cap D_j)}{Area(G_i) + Area(D_j)}$$

$$BestMatch_D(D_j) = \max_{i=1\ldots|G|} \frac{2\,Area(D_j \cap G_i)}{Area(D_j) + Area(G_i)}$$

Результаты работы различных архитектур глубоких нейронных сетей

Для сравнения были выбраны современные популярные архитектуры нейронных сетей для решения задачи детектирования текстовых областей. Сравнение проводилось на основе набора данных ICDAR 2015. Данные по точности были взяты из официальных репозиториев рассматриваемых сетей.

Таблица 2. Результат детекции произвольного четырехугольного текста на наборе ICDAR 2015

Название	P	R	F	FPS
EAST-PVANet	83,27	78,33	80,72	13,2
TextSnake	84,9	80,4	82,6	1,1
CRAFT	89,8	84,3	89,8	5,6
DB-ResNet18	84,8	77,5	81	55
DB-ResNet50	86,9	80,2	83,5	22

Заключение

В работе были рассмотрены подходы к решению задачи детектирования текста на изображениях, в том числе появившиеся в последнее время. Применение нейронных сетей при решении данной задачи позволило значительно увеличить скорость и точность обнаружения текста. Среди наиболее распространенных архитектур можно выделить DB-ResNet50 и DB-ResNet18 имеющие довольно хорошее качество при колоссальной скорости работы. В ходе дальнейшей работы планируется провести ряд экспериментов по модернизации данных архитектур в попытках улучшить качество детекции не потеряв при этом в производительности.

Список использованных источников

1. Fabrizio, J.; Marcotegui, B.; Cord, M. Text detection in street level images. Pattern Anal. Appl. 2013, c.519–533
2. Zhu, Y.; Yao, C.; Bai, X. Scene text detection and recognition: Recent advances and future trends. Front. Comput. Sci. 2015, 10, c. 19–36.
3. Cao, D.; Zhong, Y.; Wang, L.; He, Y.; Dang, J. Scene Text Detection in Natural Images: A Review, 2020, –27c.
4. Jaderberg, M.; Simonyan, K.; Vedaldi, A.; Zisserman, A. Reading Text in the Wild with Convolutional Neural Networks. Int. J. Comput. Vis. 2015, 116, c. 1–20.
5. Yao, C.; Bai, X.; Sang, N.; Zhou, X.; Zhou, S.; Cao, Z. Scene Text Detection via Holistic, Multi-Channel Prediction. arXiv 2016, arXiv:1606.09002