

УДК 004.85

## СОВМЕСТНАЯ ДЕТЕКЦИЯ И РАСПОЗНАВАНИЕ СИМВОЛОВ НА ИЗОБРАЖЕНИЯХ С ПОМОЩЬЮ ГЛУБОКОГО ОБУЧЕНИЯ

Буковкина Т.А. (Университет ИТМО, Санкт-Петербург)

Научный руководитель – к.т.н. Сметанников И.Б.

(Университет ИТМО, Санкт-Петербург)

В данной работе исследуется одна из задач компьютерного зрения – задача распознавания символов на изображениях. Разработан метод для решения данной задачи с помощью глубокого обучения, проведено сравнение с передовыми решениями в данной области.

**Введение.** В настоящее время распознавание того, что изображено на фотографии, считается одной из самых сложных и важных задач компьютерного зрения. Оптическое распознавание символов (OCR – от англ. optical character recognition) – это преобразование изображения текста – напечатанного или рукописного – в текстовые данные для последующего использования и хранения на электронных устройствах. Сфер жизни, в которых может применяться распознавание текста, немало: заполнение информации текстовых полей при помощи камеры или фотографий, оцифровка старых документов или книг для хранения в электронном виде, перевод текста с помощью камеры мобильного устройства и многое другое. Задача OCR состоит из двух последовательных этапов: обнаружение области текста на изображении и распознавание того, что в этой области написано.

Задачу распознавания текста решают с использованием глубокого обучения. За последние несколько лет предложено не мало методов для разрешения данной задачи, но универсального решения с высокой точностью так и не найдено.

Целью данной работы является исследование и разработка алгоритма распознавания текста на изображениях естественной сцены.

**Основная часть.** Один из путей решения в разработке методов распознавания текста – это модификация какого-либо существующего алгоритма для получения лучшего результата. Поэтому в качестве базового метода был выбран алгоритм для обнаружения текста на изображении AdvancedEAST. Основное преимущество данного метода – это возможность точно определять вершины начала и конца текстовой области. Все пиксели формируют рамку текстового поля, а затем только граничные пиксели используются для прогнозирования координат возвращаемой вершины.

Оригинальная архитектура обучена на наборе данных Tianchi ICPR MTWI 2018. Изображения в нем представляют из себя синтетические фотографии с описанием продуктов преимущественно на китайском языке. Наиболее интересная область – распознавание текста в естественных сценах, чему соответствуют изображения из наборов данных ICDAR.

В рамках данной работы был изменен процесс подготовки изображений для нейронной сети таким образом, чтобы отвечало требованиям выбранных наборов данных. Также было проведено обучение данной архитектуры на тренировочных наборах данных ICDAR 2015 и ICDAR 2017 MLT, в состав которых входит 10000 изображений, и проведен анализ полученных результатов. По оценкам стало понятно, что также необходимо модифицировать процесс постобработки, так как модель показывала неточные выводы для отдельных случаев. В частности, длинные фразы перестали соединяться в одну, а разбивались по словам; для изогнутых текстов координаты граничных вершин изменялись так, чтобы вся площадь текста находилась внутри рамки.

После настройки метода детекции текста была добавлена часть для распознавания текста в найденных областях. Для этого использовался измененный существующий метод. Полученная нейронная сеть обучалась сквозным способом и в результате была протестирована на тестовых изображениях ICDAR.

**Выводы.** В рамках данной работы был модифицирован существующий метод обнаружения текста на изображениях и добавлен алгоритм распознавания текста для найденных областей, метод был протестирован на наборах данных ICDAR и показал неплохие результаты. Полученный метод можно внедрять в другие проекты, в которых задача распознавания текста окажется полезной.