

УДК 004.932

МЕТОДИКА ЛОКАЛИЗАЦИИ ТЕКСТА НА ИЗОБРАЖЕНИЯХ СЛОЖНЫХ ГРАФИЧЕСКИХ СЦЕН

У Чжисян

к.т.н., доцент каф. ИС Николенко А.А.

Одесский Национальный Политехнический Университет, УКРАИНА

АННОТАЦИЯ. Рассмотрена разработка и реализация методики локализации текстовых областей на изображениях со сложным фоном на фотографиях и видеокдрах сложных графических сцен. Для решения задачи локализации предлагается сформировать пирамиду изображений с использованием дискретного двумерного вейвлет-преобразования и классификатор на основе свёрточной нейронной сети.

Введение. Одной из базовых составляющих информационной технологии распознавания текста является процедура локализации текстовых областей (ТО) на изображениях, применение которой позволяет сократить объем обрабатываемой информации и определяет оперативность и качество информационных систем переработки графической информации. Однако, на практике качество локализации ТО на изображении существенно снижается в зависимости от сложности и неоднородности фона. Поэтому разработка системы и методики локализации текстовых областей на изображениях со сложным фоном является трудно решаемой задачей, актуальной на сегодняшний день.

Цель работы – разработка и реализация методики локализации текстовых областей на изображениях со сложным фоном на фотографиях и видеокдрах сложных графических сцен, используемой в различных системах обработки графической информации и распознавания текста.

Основная часть работы. Для решения задачи локализации ТО в научно-технической литературе предложено множество методов на основе корреляции, контурной и текстурной сегментации, нейронных сетей, преобразования Фурье, вейвлет-преобразования. Проведенный анализ показал, что методы локализации, приведенные в большинстве из опубликованных работ, основаны при поиске текстов на анализе обособленных пикселей изображения без учета их пространственной взаимосвязи [1]. Это не обеспечивает хорошие результаты локализации ТО на реальных изображениях, при наличии слитного текста (номера домов, рекламные щиты, вывески магазинов, автомобильные номера и т.п.). Очевидно, что для эффективной работы систем локализации метод локализации ТО на изображении должен учитывать особенности представления данных, а именно – строгую двумерную структуру изображения, в которой существует пространственная зависимость между значениями соседних пикселей.

Одним из наиболее перспективных для решения задачи локализации текстовых областей на изображениях со сложным и неоднородным фоном является подход с учетом пространственного представления символов.

В работах [2-5] предложено для локализации ТО на изображении использовать вейвлет-преобразование и в качестве классификатора свёрточную нейронную сеть (СНС), которая по сравнению с классическими нейронными сетями обладает следующими преимуществами: возможностью учета пространственной структуры, снижением сложности архитектуры и обучения, устойчивостью к искажениям символов.

В данной работе локализация осуществляется по двухэтапной схеме [2,6].

На первом этапе с использованием градиентных методов или вейвлет-преобразований на основе анализа перепадов интенсивности в локальных областях изображения осуществляется отбор областей изображений, в которых может быть текстовая информация. Как правило, на первом этапе обнаружены и локализованы собственно ТО на однородном фоне, а также

составные текстовые области с наложенными другими объектами. На втором этапе используется классификатор на основе свёрточной нейронной сети с многомасштабным представлением изображения на основе дискретного вейвлет-преобразования для того, чтобы оценить степень принадлежности к тексту каждого пикселя отобранных составных текстовых областей.

Таким образом предложена следующая методика локализации ТО на изображениях со сложным фоном, использующая многомасштабное представление изображения на основе двумерного дискретного вейвлет-преобразования (вейвлет-пирамида), классификатор, построенный на основе свёрточной нейронной сети для отбора текстовых областей-кандидатов:

Шаг 1. Из исходного цветного изображения (цветовое пространство RGB) выделяется прямоугольный фрагмент.

Шаг 2. Формируется пирамида изображений для выделенного фрагмента. Для этого осуществляется преобразование значений интенсивности фрагмента изображения к диапазону $[-1, 1]$ и выполняется двумерное дискретное вейвлет-преобразование с разложением изображения на два уровня по каждой составляющей каналов цвета R, G, B .

Шаг 3. Полученный набор пирамид изображений поступает на предварительно обученную свёрточную нейронную сеть, которая классифицирует данный участок изображения как содержащий текст (текстовую область) или не содержащий текст — проводится выбор текстовых областей-кандидатов.

Шаг 4. Повторяются шаги алгоритма 1 – 3 (для всего изображения).

Шаг 5. Отобранные области подвергаются верификации по форме и размерам. Формируется список координат текстовых областей, содержащих символы.

Для обучения свёрточной нейронной сети была создана обучающая выборка из 650 изображений (36×64 пикселя) и контрольная выборка из 500 изображений. Изображения выборок включали текст с различными размерами, типами и цветом шрифтов, многострочный текст. Кроме символов текста они содержали и другие объекты (фрагменты домов, деревьев и т.д.). Также были добавлены изображения, содержащие только часть символов текста и без текста. После обучения сети были правильно классифицированы 98 % изображений обучающей выборки, а контрольной выборки – 78 %. средняя скорость обучения сети составила 0,35 с на эпоху.

Выводы. В данной работе рассмотрена методика локализации текстовых областей на изображениях с применением дискретного вейвлет-преобразования и классификатора на основе свёрточной нейронной сети. Экспериментальная проверка предложенных решений подтвердила их способность локализовать текст на изображениях в условиях сложных графических сцен при наличии множества нетекстовых объектов (людей, фрагментов домов и т.п.). Свёрточная нейронная сеть правильно классифицировала 98 % изображений обучающей выборки и 78 % контрольной выборки.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Андрианов А.И. Локализация текста на изображениях сложных графических сцен // Современные проблемы науки и образования. – 2013. – № 3; URL: www.science-education.ru/109-9311 (дата обращения: 30.04.2018)
2. Николенко, А. Обнаружение текстовых областей в видео-последовательностях [Текст] / А.А. Николенко, Тьен Т.К. Нгуен // Искусственный интеллект. — 2012. — № 4. — С. 227 — 234.
3. Николенко А.А., Бабилунга О.Ю., Нгуен Тьен Т.К. Локализация текстовых областей на изображениях с использованием свёрточной нейронной сети — ISSN 2079-0031 Вестник НТУ "ХПИ". — 2013. — № 19 (992). — С. 121 — 127.
4. Антошук С. Г. Метод локализации текста на изображении с обучением на основе вероятностной модели символов [Текст] / С. Г. Антошук, А. А. Николенко, Тьен Т. К. Нгуен // Вимірювальна та обчислювальна техніка в технологічних процесах. – Хмельницьк :– 2014. – ХНУ. – № 3. – С. 79 – 84.
5. Антошук С. Г. Информационная технология локализации текстовых областей на сложном фоне [Текст] / С. Г. Антошук, А. А. Николенко, Тьен Т. К. Нгуен // Електротехн. та комп'ют. системи. – К. : Техніка.– 2014. – № 16 (92). – С. 114 – 121.
6. Delakis, M. Text detection with convolutional neural networks [Text] / M. Delakis, Cr. Garcia // International Conference on Computer Vision Theory and Applications. – 2008. – P. 290-294.