УДК 004.932

РАСПОЗНАВАНИЕ ТЕКСТА НА ИЗОБРАЖЕНИЯХ СЛОЖНЫХ ГРАФИЧЕСКИХ СЦЕН

К. т. н. А. А. Николенко, Тьен Т. К. Нгуен, П. В. Гордиенко

Одесский национальный политехнический университет
Украина, г. Одесса
anatolyn@ukr.net, ktien85@ukr.net

Рассмотрена задача разработки системы распознавания текста на изображениях сложных графических сцен. При решении задач разработаны и реализованы алгоритмы для локализации текстовых областей на изображении, сегментации символов, распознавания символов.

Ключевые слова: распознавание символов, сверточная нейронная сеть, сложная графическая сцена, вейвлет-преобразование.

Задача распознавания текста на изображениях сложных графических сцен подразумевает локализацию, сегментацию и распознавание текста в ситуации, когда изображения кроме текста содержат другие объекты или их фрагменты (например, деревья, машины, здания и пр.) [1]. Существующие на данный момент системы распознавания текста обычно ориентированы на типовые условия, в которых однотонный текст расположен на однотонном контрастном фоне. Такие изображения могут содержать нетекстовую информацию, например рисунки, графики. Однако эта информация легко локализуется и отделяется от текстовых областей, что облегчает дальнейшее распознавание самого текста. В то же время распознавание текста на сложных графических сценах осложняется тем, что текст на таких изображениях часто не отделен от прочей информации явно, а является частью этой информации. Невозможно заранее предугадать, в какой области изображения расположен текст, какое он имеет искажение. Поэтому распознавание текста на изображениях сложных графических сцен на сегодняшний день является трудно решаемой задачей.

Целью данной работы является разработка и реализация системы распознавания текста на фотографиях и видеокадрах сложных графических сцен.

Система распознавания текста (СРТ) предполагает наличие на входе изображения с текстом (в формате данных графического файла). Решение задач распознавания текста на изображении включает в себя следующие подзадачи — локализация текстовых области (ТО) на изображении, сегментация символов, присутствующих в ТО, распознавание символов.

На начальных этапах обработки в СРТ выполняется процедура локализации ТО на изображении с целью отделения их от фона. Точность выполнения этой процедуры определяет эффективность работы СРТ в целом. В данной работе локализация осуществляется по двухэтапной схеме [2]. На первом этапе с использованием градиентных методов или вейвлет-преобразований [2, 3] на основе анализа перепадов интенсивности в локальных областях изображения осуществляется отбор областей изображений, в которых может быть текстовая информация. Как правило, на первом этапе могут быть обнаружены и локализованы собственно ТО на однородном фоне, а также составные текстовые области (СТО) с наложенными другими объектами. На втором этапе используется классификатор на основе сверточной нейронной сети с многомасштабным представлением изображения на основе дискретного вейвлет-преобразования для того, чтобы оценить степень принадлежности к тексту каждого пиксела отобранных СТО [2, 3]. Для обучения сверточной нейронной сети была создана обучающая выборка из 676 изображений (36×64 пиксела) и контрольная выборка из 560 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 точность классификации изображений обучающей выборки составила 99%, а кон-

трольной выборки — 77%. Средняя скорость обучения сети составила 0,31 с на эпоху.

Сегментация изображений ТО, полученной после локализации, проводилась в три этапа:

— выделение строк — исходная ТО разделялась на полосы-строки нужной ширины;

— сегментация слов — в изображении текстовой строки выделялись изображения слов;

— сегментация символов — в изображении слова определялись границы символов.

Для улучшения сегментации использовались также морфологические операции, позволяющие уменьшить шум в ТО и количество объектов, которые не являются символами.

Алгоритм распознавания символов также реализован на основе сверточной нейронной сети [4]. Входное изображение — серое изображение с белым фоном размера 36×36 выходной результат — соответствующий символ. Для обучения сети была создана обучающая выборка из 1452 изображений (36×36 пикселов) с белым фоном и контрольная выборка из 350 изображений разных размеров со сложным фоном. Выбранные изображения включали текст с различными размерами, типами шрифтов (Times New Roman, Cambria, Calibri, Arial, Adobe Garamond Pro и т. д.), видами шрифтов (полужирный, курсив, обычный, полужирный курсив) и цветом шрифтов. Для обучения сети был выбран алгоритм обратного распространения ошибки. После обучения сети точность распознавания обучающей выборки составила 97%, а контрольной выборки — 93%. Средняя скорость обучения сети составила 0,31 с. Результат работы системы распознавания представлен в таблице.

Результат работы СРТ на изображениях сложных графических сцен

Локализация ТО	Сегментация ТО	Распознавание символов
		SHOW YOUR STRIPES s

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

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: 27.01.2014).

2. Николенко А. А., Нгуен Тьен Т. К. Обнаружение текстовых областей в видеопоследовательностях // 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 сверточной нейронной сети // Вестник НТУ «ХПИ».— 2013.— № 19 (992).— С. 121—127.

4. Changan Han. Neural Network Based Off-line Handwritten Text Recognition System.— 2011.— 75 p. http://digitalcommons.fiu.edu/cgi/viewcontent.cgi?article=1436&context=etd

---

A. A. Nikolenko, Tien T. K. Nguyen, P. V. Gordienko

Text recognition in images of complex graphic scenes.

The task of developing text recognition system in the images of complex graphic scenes was considered. Algorithms for text localization, segmentation and recognition were created and implemented when solving problems.

Keywords: *character recognition, convolution neural network, image processing, complex graphic scene, text regions, wavelet transformation.*