

УДК 004.932

А.А. НИКОЛЕНКО, канд. техн. наук, доц., ОНПУ, Одесса,
О.Ю. БАБИЛУНГА, канд. техн. наук, доц., ОНПУ, Одесса,
ТЬЕН Т.К. НГУЕН, магистр, ОНПУ, Одесса

ЛОКАЛИЗАЦИЯ ТЕКСТОВЫХ ОБЛАСТЕЙ НА ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ

Предложен метод локализации текстовых областей на изображениях со сложным фоном, использующий сверточную нейронную сеть и многомасштабное представление изображения на основе дискретного вейвлет-преобразования. Проведено тестирование нейронной сети. Ил.: 2. Библиогр.: 13 назв.

Ключевые слова: метод локализации текстовых областей, сверточная нейронная сеть, дискретное вейвлет-преобразование.

Постановка проблемы. Текстовые области (ТО) на изображении содержат информацию, которая используется в интеллектуальных системах обработки и распознавания изображений (ИСОРИ) при решении задач автоматического создания аннотированного описания изображений, интерпретации контекстного содержания изображения, распознавания символов текста на изображении и т.п.

На начальных этапах обработки в ИСОРИ выполняется процедура локализации ТО на изображении с целью отделения их от фона. Точность выполнения этой процедуры определяет эффективность работы ИСОРИ в целом. 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 [1, 2]. Однако, окончательного решения задачи автоматической локализации ТО на реальных изображениях со сложным фоном, который может содержать фрагменты других объектов, например, деревьев, машин, зданий и пр., на сегодняшний день – нет. Представляется перспективной разработка метода локализации ТО на изображениях со сложным фоном на базе нейросетевых технологий с использованием многомасштабного разложения изображения по вейвлет-базису, поскольку данная технология позволяет учесть все многообразие представления ТО на изображениях.

Анализ литературы. Из научно-технических источников известен

ряд методов обработки изображений на основе вейвлет-преобразований, применяемых для решения задачи локализации областей на изображении [2 – 7]. Все они базируются на построчной или постолбцовой обработке изображений с использованием дискретного или непрерывного вейвлет-преобразований, которые позволяют определить границы объекта в анализируемой строке (или столбце) изображения. Для текстовых областей разработаны методики автоматического выбора масштаба непрерывного вейвлет-преобразования, обеспечивающие локализацию ТО при различных размерах шрифтов символов [4, 5, 7]. При этом практически не учитываются пространственные соотношения между элементами объекта, т.е. особенности его формы. В [1] для учета формы объектов применены корреляционные методы, однако их реализация в ИСОРИ затрудняется тем, что объект может иметь различные размеры и расположение на изображении.

Очевидно, что для эффективной работы ИСОРИ 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 [8]. В данной работе рассмотрены сверточные сети – один из подвидов нейронных сетей, который позволяет учесть особенности представления данных на изображениях [9 – 11].

Цель статьи – разработка метода локализации текстовых областей на изображениях на основе сверточной нейронной сети с многомасштабным представлением изображения для обнаружения текстовых областей на изображениях со сложным фоном.

Метод локализации текстовых областей. Предлагаемый метод локализации ТО на изображении, использует сверточную нейронную сеть и многомасштабное представление изображения на основе дискретного вейвлет-преобразования и реализуется следующим алгоритмом.

Шаг 1. Из исходного цветного изображения (формат *RGB*) выделяется фрагмент, соответствующий прямоугольному окну с размерами 36×64 пикселя.

Шаг 2. Формируется пирамида изображений для выделенного фрагмента. Для этого осуществляется преобразование значений интенсивности фрагмента изображения к диапазону $[-1, 1]$ и выполняется 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 сверточную нейронную сеть, которая классифицирует данный участок изображения, как содержащий текст

(текстовую область) или не содержащий текст.

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

В результате применения предложенного метода локализации ТО ко всему изображению, определяются координаты прямоугольных областей, содержащих символы текста.

Архитектура сети. Топология свёрточной сети представлена на рис. 1.

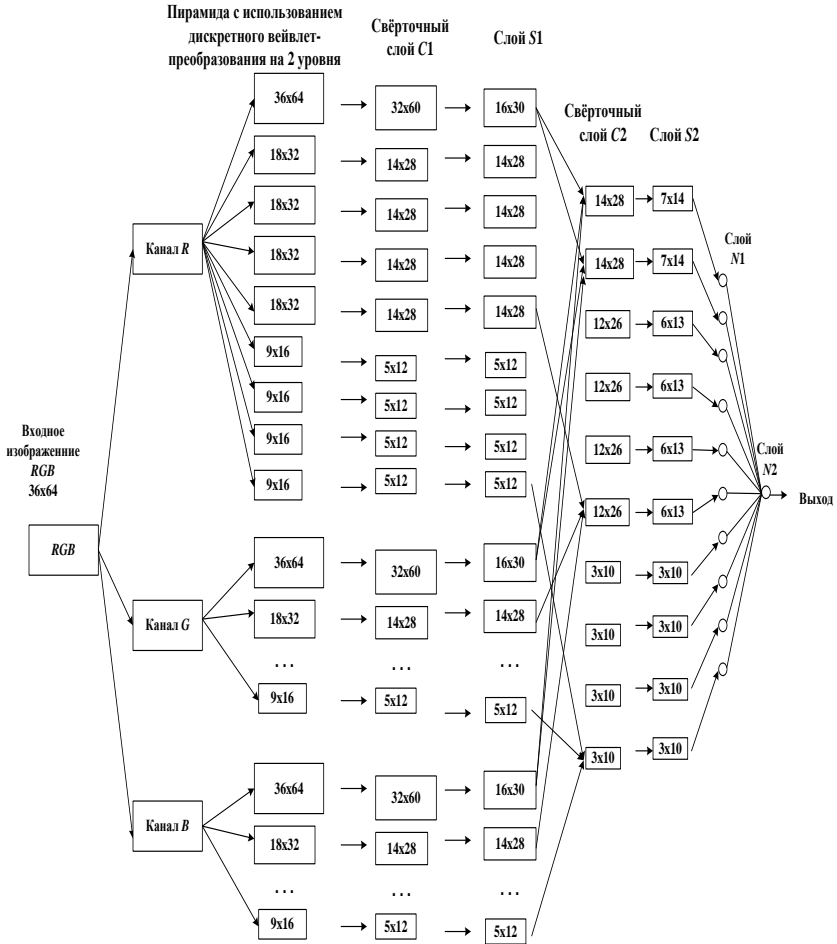


Рис. 1. Схема предложенной сетевой топологии

Исходные данные для работы сети формируются из цветного изображения (формат *RGB*) размером 36×64 пикселя. Сеть должна определить, содержит данный фрагмент изображения текст или нет.

Для обнаружения текста из входного изображения с помощью дискретного вейвлет-преобразования [2] с разложением на два уровня (используется вейвлет Хаара) формируются еще два набора по 4 изображения с размерами 18×32 и 9×16 для каждой составляющей каналов цвета *R*, *G*, *B*. В результате исходное изображение представляется одновременно на нескольких уровнях разрешения.

Свёрточная нейронная сеть состоит из двух сверточных слоев *C1* и *C2* (для их создания используются рецептивные матрицы 5×5 и 3×3 пикселя соответственно), двух подвыборочных слоев *S1* и *S2* (размер матрицы 2×2) и двух полносвязных слоев *N1* (10 нейронов) и *N2* (один нейрон). Слой *C1* имеет 27 свёрточных плоскостей (карт), которые группируются согласно каналу изображения. Каждая карта получается в результате обработки пирамиды входных изображений с помощью свертки с маской 5×5 . Особенностью подвыборочного слоя *S1* является то, что он уменьшает в 2 раза только три карты слоя *C1* размером 32×60 пикселей, остальные остаются неизменными. Слой *C2* содержит 10 свёрточных карт, каждая из которых получена с помощью свертки с маской 3×3 некоторых карт слоя *S1*. В подвыборочном слое *S2* карты слоя *C2* размером 3×10 дублируются, остальные уменьшаются в 2 раза. Полносвязные слои *N1* и *N2* содержат нейроны с функцией активации в виде гиперболического тангенса. Каждый нейрон в слое *N1* соединяется только с одной картой слоя *S2*.

Особенности обучения сети. Для обучения сети был выбран алгоритм обратного распространения ошибки. В результате обучения сеть формировала отклики со значениями -1 для фрагмента изображения, не содержащего текст, и $+1$ – для фрагмента изображения с текстом.

Значения исходных синаптических весов для всех сверточных слоёв сети генерировались в соответствии с законом равномерного распределения с нулевым математическим ожиданием и дисперсией, равной обратной величине квадратного корня из количества синаптических связей нейрона [12]. Алгоритм использует методику, позволяющую быстро вычислять вектор частных производных (градиент) сложной функции многих переменных, если структура этой функции известна. В качестве такой функции в алгоритме рассматривается функция ошибки сети и учитывается, что структура функции ошибки сети определяется архитектурой нейронной сети. В процессе обучения достигается настройка межнейронных связей для сверточных и

полносвязных слоев, коэффициенты связей для подвыборочных слоев остаются неизменными и равными 0,25.

Для обучения сети была создана обучающая выборка из 676 изображений (36×64 пикселя) 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.

Тестирование работы нейронной сети при локализации ТО. В ходе компьютерного эксперимента оценивались скорость обучения сети, точность классификации фрагментов изображений из обучающей и контрольной выборок, которые содержали текстовые области.

Оценка времени выполнения классификации ТО выполнялась на персональном компьютере со следующей конфигурацией: процессор – Intel (R) Core (TM) 2 Duo CPU T9400 @ 2.53GHz, RAM – 4GB, тип системы – 32 bit, операционная система – Windows 7.

Для оценки точности классификации изображений, содержащих текст, использовалась мера точности [13], которая показывает, сколько процентов областей, распознанных на изображении как текстовые, составляют действительно текстовые области.

После обучения сети точность классификации изображений обучающей выборки составила 99,3 %, а контрольной выборки – 77,7 %. средняя скорость обучения сети составила 0,31 с на эпоху, среднее время классификации текстовых областей на изображении 36×64 пикселя составило 0,17 с. Пример работы обученной сети на реальном изображении представлен на рис. 2.



Рис. 2. Пример локализации ТО на изображении с применением предложенного метода: *а* – исходное изображение; *б* – результат локализации текстовой области

Выводы. Предложенный метод локализации текстовых областей на изображениях на основе сверточной нейронной сети с многомасштабным представлением изображения позволяет обнаруживать текстовые области на изображениях со сложным фоном в ИСОРИ различного назначения.

Список литературы. 1. Шапиро Л. Компьютерное зрение / Л. Шапиро, Дж. Стокман. – М.: БИНОМ. Лаб. знаний, 2006. – 752 с. 2. Малла С. Вейвлеты в обработке сигналов / С. Малла. – М.: Мир, 2005. – 671 с. 3. Добеши И. Десять лекций по вейвлетам / И. Добеши. – М.: Регулярная и хаотическая динамика, 2001. – 464 с. 4. Антощук С.Г. Анализ базисных функций вейвлет-преобразования при мультимасштабном контурном представлении изображений / С.Г. Антощук, А.А. Николенко, Е.В. Ткаченко // Электромашинобуд. та електрообладнання. – 2009. – Вип. 72. – С. 15-19. 5. Использование вейвлет-преобразования при локализации последовательностей символов / С.Г. Антощук, А.А. Николенко, Е.В. Ткаченко, О.Ю. Бабилунга // Искусственный интеллект. – 2009. – № 4. – С. 23-29. 6. Адаптивна локалізація символних написів на зображеннях методом вейвлет-аналізу / Антощук С.Г., Бабилунга О.Ю., Николенко А.О., Ткаченко О.В. // Вісник ЖІТІ. – 2008. – № 4. – С. 125-130. 7. Антощук С.Г. Локализация квазипериодических последовательностей на изображениях / С.Г. Антощук, А.А. Николенко, Тьен Т.К. Нгуен // Компьютерные науки и технологии. – 2009. – № 2.– С. 35-40. 8. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – М.: Вильямс, 2006 – 1104 с. 9. LeCun Y. Scaling learning algorithms towards AI / Y. LeCun, Y. Bengio. – MIT Press, 2007. 10. LeCun Y. Gradient-based learning applied to document recognition / Y. LeCun, L. Bottou, Y. Bengio, P. Haffner // Proceedings of the IEEE. – 1998. – 86 (11). – P. 2278-2324. 11. Delakis M. Text detection with convolutional neural networks. International Conference on Computer Vision Theory and Applications / M. Delakis, Cr. Garcia. – VISAPP. – 2008 – P. 290-294. 12. LeCun Y. Efficient BackProp in Neural Networks: Tricks of the trade / Y. LeCun, L. Bottou, G. Orr, K. Muller. – Springer, 1998. – 44 p. 13. Вежневцев В. Оценка качества работы классификаторов / В. Вежневцев // Компьютерная графика и мультимедиа. – 2006. – № 4. – С. 106.

Поступила в редакцию 25.06.2013

Статью представил д-р техн. наук, проф. ОНПУ Крисилов В.А.

УДК 004.932

Локалізація текстових областей на зображеннях з використанням згорткової нейронної мережі / Николенко А.О., Бабилунга О.Ю., Нгуен Т.К. Тьен // Вісник НТУ "ХПИ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПИ". – 2013. – № 19 (992). – С. 121 – 127.

Запропоновано метод локалізації текстових областей на зображеннях з складним фоном, що використовує згорткову нейронну мережу та багатомасштабне представлення зображення на основі дискретного вейвлет-перетворення. Проведено тестування нейронної мережі. Іл.: 2. Бібліогр.: 13 назв.

Ключові слова: метод локалізації текстових областей, згорткова нейронна мережа, дискретне вейвлет-перетворення.

UDC 004.932

Localization of the text area on the images using a convolution neural network / Nikolenko A.A., Babilunga O.Ju., Nguyen T.K. Tien // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2013. – № 19 (992). – P. 121 – 127.

Method of localization of the text area on the images with a difficult background, using a convolution neural network and multi-scale representation of the image based on discrete wavelet transform was proposed. Testing of neuron network is conducted. Figs.: 2. Refs.: 13 titles.

Keywords: method of localization of the text area, convolution neural network, discrete wavelet transform.