УДК 004.932

Тьен Т.К. Нгуен

ОБНАРУЖЕНИЕ И РАСПОЗНАВАНИЕ ТЕКСТОВ НА ИЗОБРАЖЕНИЯХ СЛОЖНЫХ ГРАФИЧЕСКИХ СЦЕН С ПОМОЩЬЮ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Аннотация. *Рассмотрена задача разработки интеллектуальной системы распознавания текста на фотографиях и видеокадрах сложных графических сцен. При решении задачи реализованы методы для обнаружения и локализации текстовых областей, распознавания символов с помощью свёрточных нейронных сетей.*

Ключевые слова: *распознавание символов, свёрточная нейронная сеть, обработка изображений, сложная графическая сцена, локализация текстовых областей*

Tien T. K. Nguyen

TEXT DETECTION AND CHARACTER RECOGNITION IN IMAGES OF COMPLEX GRAPHIC SCENES USING SUPER PRECISIONOUS NEURAL NETWORKS

Abstract. *The problem of developing the intellectual system of text recognition in photographs and video of complex graphic scenes was considered. Methods for the text detection and text localization, character recognition using convolutional neural networks were implemented to solve the problem.*

Keywords: *character recognition, convolution neural network, image processing, complex graphic scene, text detection*

Тієн Т. К. Нгуєн

ВИЯВЛЕННЯ І РОЗПІЗНАВАННЯ ТЕКСТІВ НА ЗОБРАЖЕННЯХ СКЛАДНИХ ГРАФІЧНИХ СЦЕН ЗА ДОПОМОГОЮ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Анотація. *Розглянуто задачу розробки інтелектуальної системи розпізнавання тексту на фотографіях і відеокадрах складних графічних сцен. При вирішенні завдання реалізовані методи для виявлення і локалізації текстових областей , розпізнавання символів за допомогою згорткових нейронних мереж.*

Ключові слова: *розпізнавання символів, згорткова нейронна мережа , обробка зображень , складна графічна сцена, локалізація текстових областей*

Введение. В тех случаях, когда изображения, кроме текста, содержат другие объекты или их фрагменты (например, деревья, машины, здания и пр.), задача распознавания текста на изображениях рассматривается как задача анализа сложных графических сцен и предполагает выполнение процедур: локализации, сегментации и распознавания текста [1]. Существующие на данный момент системы распознавания текста, как правило, ориентированы на работу в условиях дополнительных ограничений, например – однотонный текст расположен на однотонном контрастном фоне. Такие изображения могут содержать нетекстовую информацию (рисунки, графики и др.), 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 явно не отделен от прочей информации (фона), а является частью этой информации. В таком случае невозможно заранее предугадать, в какой области изображения расположен текст и, какое он имеет искажение. Поэтому распознавание текста на изображениях сложных графических сцен является актуальной задачей, требующей проведения дополнительных исследований с привлечением современных средств интеллектуального анализа данных

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

На вход интеллектуальной системы распознавания текста (ИСРТ) подается изображение с текстом в формате данных графического файла. Задача распознавания текста на изображении решается поэтапно: локализация текстовых областей (ТО) на изображении; сегментация символов, присутствующих в ТО, и распознавание символов.

Локализации ТО. Процедура локализации выполняется с целью отделения ТО от фона изображения. В данной работе процедура локализации выполняется по двухэтапной схеме [2]. На первом этапе (эвристический этап) с использованием градиентных методов на основе анализа перепадов интенсивности в локальных областях цветного изображения (цветовое пространство RGB) осуществляется отбор областей изображений, в которых может быть текстовая информация. Как правило, на первом этапе бывают 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 свёрточной нейронной сети с многомасштабным представлением изображения на основе дискретного вейвлет-преобразования для того, чтобы оценить вероятность принадлежности к тексту пикселей отобранных областей СТО [3]. Для обучения свёрточной нейронной сети с многомасштабным представлением изображения была создана обучающая выборка из 676 цветных изображений (цветовое пространство RGB, 36x64 пикселя). Контрольная выборка из 560 изображений – была создана для проверки работы сети. 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 в выборки были добавлены изображения, содержащие только часть символов текста и без текстовых областей.

После обучения сети точность классификации изображений обучающей выборки составила 99,3 %, а контрольной выборки – 77,7 %.

Сегментация символов в ТО. Процедура сегментации символов в ТО [4] проводится в три этапа:

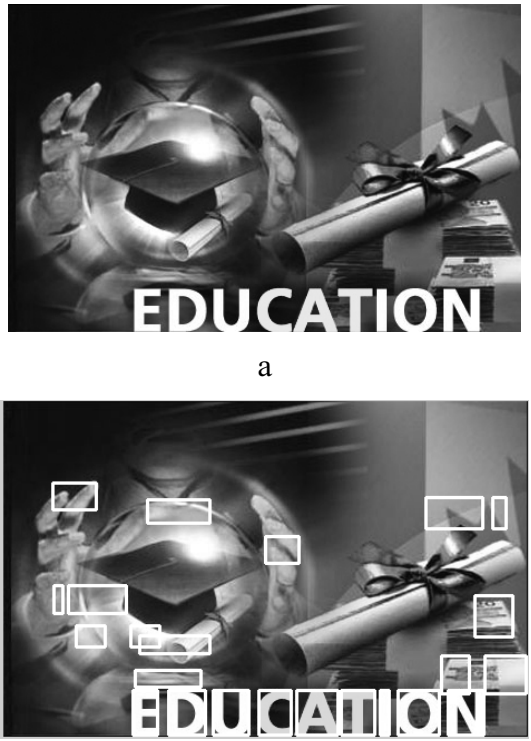
– выделение строк текста методом горизонтального, вертикального проектирования со значением порога, определяемого по усреднению минимальных экстремумов суммы интенсивности каждой строки.

– сегментация слов. В изображении текстовой строки выделяются изображения слов и определяются их координаты;

– сегментация символов. В изображении слова выделяются области, соответствующие отдельным символам.

Для улучшения качества результата сегментации применяются морфологические операции [5], которые позволяют удалить в области символа шум и другие объекты, которые не связаны с символом или не имеют границ.

Пример локализации ТО и сегментации символов в ТО приведен на рис. 1.



а

б

Рис. 1. Пример локализации ТО на изображениях сложных графических сцен:
а) исходное изображение; б) полученное изображение после локализации ТО и сегментации символов в ТО

Распознавание символов. В данной работе предлагается алгоритм распознавания символов, который построен на основе свёрточной нейронной сети (СНС) [6 – 9].

Входное изображение – полутоновое изображение с белым фоном размера 36×36, интенсивность пикселей лежит в диапазоне [-1;1], и выходной результат – соответствующий символ или объявление ошибки (т.е. изображение не содержит символ).

Топология свёрточной нейронной сети предложена на рис. 2.

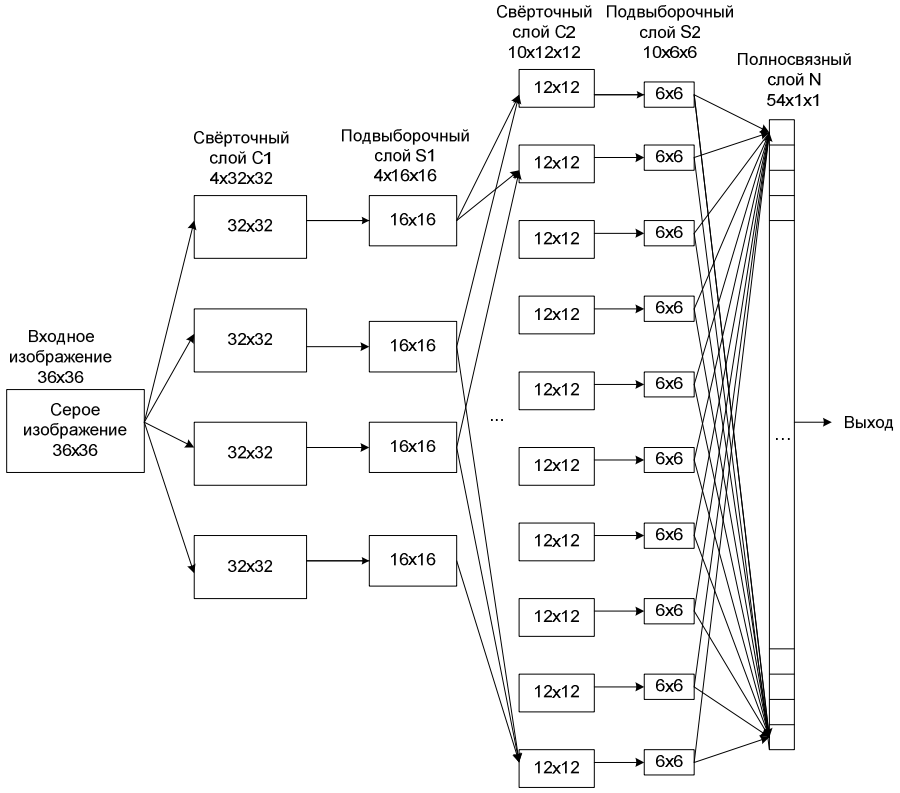


Рис. 2. Схема предложенной сетевой топологии

Сеть состоит из двух свёрточных слоев $C1$ и $C2$ (для их создания используются рецептивные матрицы 5×5 пикселей соответственно), двух подвыборочных слоев $S1$ и $S2$ (размер рецептивной матрицы 2×2), и одного полносвязного слоя N (54 нейрона).

Слой $C1$ содержит 4 карты размером 32×32. Каждая карта получается в результате обработки входных изображений с помощью свертки с маской 5×5. В качестве функции активации нейронов слоя $C1$ используется гиперболический тангенс. Значение элемента карты вычисляется по формуле

$$X_i^{h,l} = f\left(\sum_{k=1}^{n_l-1}\sum_{j=-\infty}^{\infty} X_{i-j}^{k,l} W_{i-j,i}^{h,k,l} + B_i^{h,l}\right), \quad (1)$$

где $X_i^{h,l}$ – значение элемента i в карте признаков h слоя l, f – функции активации (гиперболический тангенс), n_l – количество карт признаков в слое l, $W_{i-j,i}^{h,k,l}$ – синоптический вес связи между элементом i в карте признаков h слоя l и элементом $i-j$ карты k слоя $l-1$, $B_i^{h,l}$ – значение сдвига (смещения) для элемента i в карте признаков h слоя l.

Выход слоя $C1$ подключен к подвыборочному слою $S1$ для увеличения помехоустойчивости сети к входным деформациям. Подвыборочный слой состоит из 4 карт признаков размером 16×16. Каждый из элементов карт размером 16×16 этого слоя соединен с областью 2×2 в соответствующей карте признаков предыдущего слоя. Для элементов подвыборочного слоя эти области не перекрываются, следовательно, карты признаков этого слоя содержат в 2 раза меньше строк и столбцов, чем в предыдущем слое.

Слой $C2$ содержит 10 свёрточных карт, каждая из которых получена с помощью свертки с маской 5×5 некоторых карт слоя $S1$. Связь карт, используемая при проектировании данной сети, приведена в табл. 1. Каждый нейрон слоя C2 в качестве функции активации использует гиперболический тангенс.

1. Параметры соединения слоя $S1$ и слоя $C2$

№	1	2	3	4	5	6	7	8	9	10
1	x	x	x				x	x	x	
2	x			x	x		x	x		x
3		x		x		x	x		x	x
4			x		x	x		x	x	x

В подвыборочном слое *S*2 карты слоя *C*2 размером 6x6 дублируются, остальные уменьшаются в 2 раза. Наконец, полносвязанный слой *N* содержит 54 стандартных сигмоидальных нейрона. Каждый нейрон в слое *N* соединяется только с одной картой слоя подвыборки *S*2.

Сеть, представленная на рис.2, содержит 88848 связей, но только 890 из них требуют настройки весов.

Особенности обучения сети. Для обучения сети был выбран алгоритм обратного распространения ошибки [10–11]. Обучающая выборка содержит много изображений 43-х английских и русских отличающихся по начертанию букв (a, .., z, A, .., Z), десятичных цифр (0, .., 9), и нетекстовых объектов. Изображениям, содержащим буквы или цифры, соответствуют целевые выходные маски со значением от 1 до 53, а нетекстовым изображениям – со значением 54. В результате обучения сеть формирует цифровой массив, содержащий значения выходов 54 нейронов слоя *N* в диапазоне [-1; 1]. В массиве только одно значение равно 1, а остальные -1. Позиция значения «1» в массиве соответствует распознанному символу.

Обучающая выборка была создана из 1697 изображений (36×36 пикселей) с белым фоном и интенсивностью пикселей в диапазоне [-1; 1]. Контрольная выборка была создана из 568 различных изображений. 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. Остальные выб-ранные изображения включали нетекстовые объекты, которые на втором этапе процедуры локализации [2, 3] смешивались с текстовыми.

Значения исходных синаптических весов для всех сверточных слоёв сети генерировались в соответствии с законом равномерного распределения с нулевым математическим ожиданием и дисперсией, обратной квадратному корню из количества синаптических связей нейрона [11]. Алгоритм обратного распространения ошибки использует методику, позволяющую быстро вычислять вектор частных производных (градиент) сложной функции многих переменных, если структура этой функции известна.

В этом алгоритме происходит распространение ошибки от выходов СНС к входам, то есть в направлении обратном распространению сигналов в обычном режиме работы. Согласно методу наименьших квадратов минимизируемой целевой функцией ошибки сети является величина

$$E(w) = \frac{1}{2}\sum_{j-p}(y_{j-p}^{(N)} - d_{j-p})^2 ,$$

где $y_{j-p}^{(N)}$ – реальное выходное состояние нейрона j выходного слоя N нейронной сети при подаче на ее входы p-го образа; d_{j-p} – желаемое (идеальное) выходное состояние этого нейрона. Суммирование происходит по всем нейронам выходного слоя и по всем обрабатываемым сетью образам.

Минимизация ведется методом градиентного спуска, что означает подстройку весовых коэффициентов следующим образом:

$$\Delta w_{ij}^{(n)} = -\eta \cdot \frac{\partial E}{\partial w_{ij}},$$

где η – коэффициент скорости обучения, $0<\eta<1$.

В процессе обучения достигается настройка межнейронных связей для сверточных и полносвязных слоев, коэффициенты связей для подвыборочных слоев остаются неизменными и равными 0,25.

После обучения сети точность распознавания обучающей выборки составила 96,88 %, а контрольной выборки – 93 %.

В качестве иллюстрации работы системы можно рассмотреть результаты, полученные при обработке изображения, представленного на рис.1. Предлагаемая система обнаружила на этапе локализации все ТО, соответствующие символам, а также ряд областей, не содержащих символы (рис.1). На втором этапе распознавания из полученных ТО система отсеяла нетекстовые области (за исключением одной) и распознала текст на исходном изображении в виде слова EUUCATION (вместо правильного EDUCATION) и отдельного сивола q, который соответствовал ошибочно локализованной нетекстовой области.

Следует отметить, что такие средства распознавания текста как FineReader 11 и CuneiForm V12 на исходном изображении в автоматическом режиме не обнаружили текстовых областей и никакого текста не распознали.

Выводы. В данной работе рассмотрена интеллектуальная система распознавания текста на фотографиях и видеокадрах сложных графических сцен, методы и алгоритмы локализации и сегментации ТО, распознавания символов. Реализация системы включает две основные части, каждая из которых выполнена на основе отдельной сверточной нейронной сети. Экспериментальная проверка предложенных решений подтвердила их способность обнаруживать и распознавать текст на изображениях в условиях сложных графических сцен при наличии множества нетекстовых объектов (людей, фрагментов домов и т.п.). Дальнейшее повышение качества функционирования интеллектуальной системы распознавания текста может быть достигнуто с использованием лингвистической коррекции распознанного текста.

Список использованной литературы

1. Андрианов А. И. Локализация текста на изображениях сложных графических сцен / А. И. Андрианов // Современные проблемы науки и образования. – 2013. – № 3. URL: www.science-education.ru/109-9311 (дата обращения: 27.01.2014).

2. Николенко А. А. Обнаружение текстовых областей в видеопоследовательностях. / А. А. Николенко, Нгуен Тьен Т.К. – Искусственный интеллект. – 2012. – № 4. – С. 227 – 234.

3. Николенко А.А. Локализация текстовых областей на изображениях с использованием сверточной нейронной сети / А. А. Николенко, О. Ю. Бабилунга, Нгуен Тьен Т.К. Вестник НТУ «ХПИ». – 2013. – № 19 (992). – С. 121 – 127. ISSN 2079-0031

4. Danial Md Nor, Rosli Omar , M. Zarar M. Jenu, and Jean-Marc Ogier. Image Segmentation and Text Extraction: Application to the Extraction of Textual Information in Scene Image, (2011), *International Seminar on Application of Science Mathematics 2011* (*ISASM 2011*). URL:http://eprints.uthm.edu.my/2380/1/995.pdf

5. Frank Y. Shin. Image Processing and Pattern Recognition: Fundamentals and Techniques, (2010), *Hoboken*, New Jersey, *Wiley-IEEE Press*.

6. Han Changan. Neural Network Based Off-line Handwritten Text Recognition System, (2011), *FIU Electronic Theses and Dissertations,* 363 p. URL:http://digitalcommons.fiu.edu/cgi/viewcontent.cgi?article=1436&context=etd.

7. Simadr P., Steinkraus D., and Platt J. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis, (2003), *International Conference on Document Analysis and Recognition (ICDAR)*, *IEEE Computer Society,* Los Alamitos, pp. 958 – 962.

8. Mirovski P., LeCun Y., Madkhavan D., and Kujneskii R. Comparing SVM and Convolutional Networks for Epileptic Seizure Prediction from Intracranial EEG, (2008), *Proceeding. Machine Learning and Signal Processing (MLSP'08)*, IEEE, 2008. URL:http://yann.lecun.com/exdb/publis/pdf/mirowski-mlsp-08.pdf (accessed 01.03.2014).

9. Ebrahimpour R., Esmkhani A., and Faridi S. Farsi Handwritten Digit Recognition Based on Mixture of RBF Experts, (2010), *IEICE Electronics Express*, Vol. 7, No. 14, pp. 1014 – 1019.

10. Khashman, A. A Modified Backpropagation Learning Algorithm With Added Emotional Coefficients, (2008), *IEEE Transactions on Neural Networks*, Vol. 19, No. 11, November 2008.

11. .LeCun Y., Bottou L., Orr G., and Muller K. Efficient BackProp, (1998). URL:http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf (accessed 01.03.2014).

Получено 02.03.2014

References

1. Andrianov A.I. Lokalizacija teksta na izobrazhenijah slozhnyh graficheskih scen [Text Localization in Images of Complex Graphic Scenes], (2013), *Sovremennye Problemy Nauki i Obrazovanija*, Vol. 3 (In Russian). URL: www.science-education.ru/109-9311 (accessed: 27.01.2014).

2. Nikolenko A.A., and Nguen Tien T.K. Obnaruzhenie tekstovyh oblastej v video-posledovatel'nostjah [Text Regions Detection in Video Frames], (2012), *Iskusstvennyj Intellect Publ.,* Vol. 4, – pp. 227 – 234 (In Russian).

3. Nikolenko A.A., Babilunga O.Ju., and Nguen Tien T.K. Lokalizacija tekstovyh oblastej na izobrazhenijah s ispol'zovaniem svertochnoj nejronnoj seti [Localization of the Text Area on the Images Using a Convolution Neural Network], (2013), *Vestnik* N*acional'nogo Politehnicheskogo universiteta "HPI"Publ.*, Vol. 19 (992), pp. 121 – 127(In Russian).

4. Danial Md Nor, Rosli Omar, M. Zarar M. Jenu, and Jean-Marc Ogier. Image Segmentation and Text Extraction: Application to the Extraction of Textual Information in Scene Image, (2011), *International Seminar on Application of Science Mathematics 2011 (ISASM 2011)* (In English). URL:http://eprints.uthm.edu.my/2380/1/995.pdf

5. Frank Y. Shin, (2010), Image Processing and Pattern Recognition: Fundamentals and Techniques) *Hoboken,* New Jersey, *Wiley-IEEE Press* (In English)

6. Changan Han. Neural Network Based Off-line Handwritten Text Recognition System, (2011), *FIU Electronic Theses and Dissertations,* 363 p. (In English). URL:http://digitalcommons.fiu.edu/cgi/viewcontent.cgi?article=1436&context=etd

7. Simadr P., Steinkraus D., and Platt J. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis, (2003), *International Conference on Document Analysis and Recognition (ICDAR)*, *IEEE Computer Society,* Los Alamitos, pp. 958 – 962 (In English).

8. Mirovski P., LeCun Y., Madkhavan D., and Kujneskii R. Comparing SVM and Convolutional Networks for Epileptic Seizure Prediction from Intracranial EEG, (2008). *Proceeding. Machine Learning and Signal Processing (MLSP'08)*, IEEE, 2008 (In English) URL:http://yann.lecun.com/exdb/publis/pdf/mirowski-mlsp-08.pdf (accessed 01.03.2014)

9. Ebrahimpour R., Esmkhani A., and Faridi S. Farsi Handwritten Digit Recognition Based on Mixture of RBF Experts, (2010), *IEICE Electronics Express*, Vol. 7, No. 14, p. 1014 –1019 (In English).

10. Khashman A. A Modified Backpropagation Learning Algorithm With Added Emotional Coefficients, (2008), *IEEE Transactions on Neural Networks*, Vol. 19, No. 11, November 2008 (In English).

11. LeCun Y, Bottou L., Orr G., and K. Muller. Efficient BackProp, (1998) (In English). URL:http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf (accessed 01.03.2014).

Нгуен Тьен Тхи Кхань,
аспирантка каф. информационных систем
Одесского нац. политехн. ун-та,
тел. 705-8356.
E-mail:
ktien85@ukr.net