

УДК 004.89

ИССЛЕДОВАНИЕ И РАЗРАБОТКА СИСТЕМЫ РАСПОЗНАВАНИЯ ТЕКСТА НА ИЗОБРАЖЕНИИ

Денисенко А.А.*ЧП Денисенко, Ирпень, e-mail: alexey.denisenko.work@gmail.com*

В работе рассмотрена технология создания системы распознавания символов (с применением сверточных нейронных сетей) на изображении, а именно символов номерного знака автомобиля. На сегодня существует множество подходов к решению этой проблемы, причем большинство из них неэффективны для изображений, символы которых расположены на сложном фоне и подвержены шумовым, аффинным и проекционным искажениям. Предлагаемая в статье методика состоит из следующих этапов: предварительная обработка изображения, сегментация текста и распознавание, которое осуществляется сверточными нейронными сетями. В ходе исследования был проведен ряд экспериментов, а именно: эксперимент по определению наиболее подходящего метода бинаризации изображений, эксперимент по определению оптимальной топологии нейронной сети, для задачи распознавания символов на изображении. Методика, представленная в данной статье применительно к задаче распознавания автомобильных номеров, в результате проведенных экспериментов демонстрирует высокую надежность и точность – около 81 %, в том числе в условиях слабого освещения, именно поэтому разработанный метод распознавания может быть рекомендован для использования в коммерческих целях. Также в статье предложены возможные варианты улучшения данной методики.

Ключевые слова: машинное обучение, глубокое обучение, сверточные нейронные сети, автоматическое распознавание номеров, обработка изображений, бинаризация, сегментация

RESEARCH AND DEVELOPMENT OF TEXT RECOGNITION SYSTEM

Denisenko A.A.*PE Denysenko, Irpin, e-mail: alexey.denisenko.work@gmail.com*

The paper discusses the technology of creating character recognition (using convolutional neural networks) systems on the image. These days, there are many approaches to solving this problem, and most of them are ineffective for images whose symbols are located on a complex background and are vulnerable to noise, affine and projection distortions. The proposed technique consists of the following stages: image pre-processing, text segmentation, and recognition by convolutional neural networks. During research was conducted a series of experiments, namely: experiment to select the most suitable method of binarization of digital images, experiment to select the most efficient convolutional neural network topology form text recognition problem. As a result of the experiments performed, this technique as applied to the recognition of car numbers demonstrates high reliability and accuracy, including in low light conditions, therefore, the developed recognition method can be recommended for commercial use. As an additional field of experiments was suggested a bunch of approaches of how to improve this technique.

Keywords: machine learning, deep learning, convolutional neural networks, automatic number-plate recognition, image processing, binarization, segmentation

Современные методы распознавания символов текста позволяют решать ряд научных, а также прикладных задач, таких как восстановление документов, публикация текста на веб-странице, оцифровка книг, автоматизация систем учета в бизнесе, определение номера банковской карты. Поскольку ряд характеристик текстовых данных имеет свойство меняться (информация может быть нанесена на изображения вручную или с использованием разных шрифтов; символы могут содержать цифровые дефекты или отображаться на изображениях частично; сами изображения могут иметь сложную фоновую структуру), методы, положенные в основу программных систем, должны обеспечивать высокую точность и быстродействие, при этом оставаясь эффективными в естественных условиях.

В связи с этим особую актуальность приобретает разработка систем распознавания символов с большой нагрузкой, которые

ориентированы на распознавание коротких текстов, не имеющих строгого стандарта, например американских автомобильных номеров. Разработка программной системы сопряжена с рядом проблем:

1. Освещение: из-за воздействия окружающей среды (свет фар, дождь и т.д.) освещение входного изображения меняется.

2. Сложный фон: фон номерных пластин может содержать рисунки со сложными объектами, которые трудно отделимы от символов, находящихся на переднем плане.

3. Расположение региона (штата): расположение идентификатора в номерных знаках США варьируется от штата к штату. Это затрудняет обобщение методов, лежащих в основе системы распознавания, и требует больших вычислений.

4. Наличие контуров, теней, нежелательных символов и т.д.

Этап предобработки изображения, содержащего автомобильный номер, включа-

ет в себя коррекцию изображения (удаление шумов с фона номерной пластины, устранение неравномерного распределения яркости и эффектов потери фокуса) и устранение избыточной информации. Этап предварительной обработки является не менее важным, чем все последующие, – от его успеха зависит качество сегментации изображения. Предлагаемый в работе метод использует анизотропную диффузию и эквализацию гистограммы изображения.

Поскольку изображения могут иметь ряд многочисленных особенностей из-за специфики окружающей среды, рассмотрение одного метода бинаризации неэффективно. Для лучшей сегментации символов предлагается провести бинаризацию гибридным методом [1]: к входному изображению применить пять методов бинаризации, среди которых выбрать лучший в зависимости от качества получаемого результата.

Материалы и методы исследования

Глобальный метод выбирает пороговое значение для классификации пикселя изображения – фон или передний план [2]. Пороговое значение основано на требуемом проценте фоновых пикселей и рассчитывается для части изображения, содержащей необходимую текстовую информацию.

Метод Саувола [3] относится к методам локальной адаптивной бинаризации – он вычисляет индивидуальный порог бинаризации $T(x, y)$ для каждого пикселя (x, y) :

$$T(x, y) = m(x, y) \left[1 + k \left(\frac{\sigma(x, y)}{R-1} \right) \right], \quad (1)$$

где $m(x, y)$ – среднее значение, $\sigma(x, y)$ – среднеквадратическое отклонение в точке (x, y) , R – максимальное отклонение ($R = 128$ для оттенков серого), k – смещение, которое принимает положительные значения в диапазоне $[0,2; 0,5]$.

Алгоритм Оцу [2] позволяет минимизировать среднюю ошибку сегментации, возникающую при принятии решения о принадлежности пикселя фону или объекту изображения:

1. Вычислить сумму $P_i(k)$ для $k = 0, 1, 2, \dots, L-1$ по формуле

$$P(k) = \sum_{i=0}^k p_i, \quad (2)$$

где p_i – компоненты нормализованной гистограммы для $i = 0, 1, 2, \dots, L-1$, L – максимальное значение шкалы оттенков серого.

2. Вычислить средние значения $m(k)$ для $k = 0, 1, 2, \dots, L-1$ по формуле

$$m(k) = \sum_{i=0}^k i \times p_i. \quad (3)$$

3. Вычислить общую яркость m_G по формуле

$$m_G = \sum_{i=0}^{L-1} i \times p_i. \quad (4)$$

4. Вычислить межклассовую дисперсию $\sigma_b^2(k)$ для $k = 0, 1, 2, \dots, L-1$ по формуле

$$\sigma_b^2 = \frac{[m_G \times P_1(k) - m(k)]^2}{(P_1(k)[1 - P_1(k)])}. \quad (5)$$

5. Найти порог Оцу как значение, для которого $\sigma_b^2(k)$ максимальна.

Часто используемым в задачах компьютерного зрения является алгоритм Кэнни обнаружения границ [2]:

1. Применить к входному изображению фильтр Гаусса для удаления шума.

2. Найти градиенты яркости, применяя матрицы свертки к каждому пикселю изображения:

$$G_x(i, j) = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix},$$

$$G_y(i, j) = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix},$$

где (i, j) – координаты пикселя в исходном изображении.

3. Вычислить значение градиента G и угол направления вектора градиента θ , используя соответствующие формулы:

$$G(i, j) = \sqrt{G_x^2(i, j) + G_y^2(i, j)}, \quad (6)$$

$$\theta(i, j) = \tan^{-1} \left(\frac{G_y(i, j)}{G_x(i, j)} \right). \quad (7)$$

4. Отметить как границы только локальные максимумы.

5. Определить итоговые границы путем удаления всех «слабых» границ.

Полученное в результате работы алгоритма Кэнни изображение делится на области, которые используются для вычисления локального порога. Пороговое значение для каждого блока находится при помощи шка-

лы оттенков серого исходного изображения и соседних для всех граничных пикселей, затем полученные бинарные блоки объединяются для создания целого двоичного изображения.

Медианный стекинг представляет собой технологию наложения изображений с целью уменьшения шума, при которой значение яркости каждого пикселя вычисляется путем нахождения медианного значения его яркости из набора изображений. В настоящей работе для медианного стекинга предлагается использовать бинаризованные изображения, полученные после применения вышеприведенных алгоритмов.

На рис. 2 представлены изображения (1–5), которые были получены в результате применения пяти выбранных методов бинаризации к некоторому входному изображению (рис. 1). Изображение номерной пластины, бинаризованное при помощи обнаружения границ, содержит меньшее количество так называемых артефактов – отношение пикселей переднего края к фоновым пикселям является максимальным. Таким образом, для данного входного изображения бинаризация на основе обнаружения границ выбирается как лучший среди имеющихся методов бинаризации.



Рис. 1. Входное изображение номерной пластины

Предположив, что идентификатор региона расположен в верхней части номерного знака, легко осуществить поиск и извлечение названия штата и государственного номера при помощи метода горизонтальной проекции. Сегментация с использованием проекционных профилей [4] использует тот факт, что в бинарном изображении пиксели переднего плана принимают одни и те же значения, противоположные фоновым пикселям. Локальные минимумы горизонтального профиля проекции соответствуют фоновым пикселям. Горизонтальный профиль $H(y)$ находит сумму яркостей всех пикселей изображения $I[x, y]$ размерности $M \times N$, где x , y – координаты пикселя:

$$H(y) = \sum_x I[x, y]. \quad (8)$$

Вертикальный профиль проекции $V(x)$ способен справиться с поворотами сим-

волов [4]. Он помогает в поиске координат предполагаемых частей номерного знака:

$$V(x) = \sum_y I[x, y]. \quad (9)$$



Рис. 2. Бинаризованные различными методами номерные пластины

Для того, чтобы уменьшить использование памяти и выполнить суммирование с минимальными временными затратами, перед подсчетом горизонтальных и вертикальных проекционных значений входное изображение при помощи морфологических операций преобразуется в более компактную версию под названием «скелет» [6]. Скелет государственного номера номерной пластины из рис. 1 сохраняет структуру всех его объектов, но удаляет лишние пиксели, как показано на рис. 3. Извлеченное название штата («Texas»), которое находится в верхней части номерной пластины входного изображения, проиллюстрировано на рис. 4.



Рис. 3. Скелет части входного изображения с государственным номером



Рис. 4. Извлеченное из входного изображения название штата («Texas»)

На рис. 5 изображен извлеченный при помощи метода горизонтальной проекции гистограммы государственный номер. Результаты сегментации символов приводятся на рис. 6.

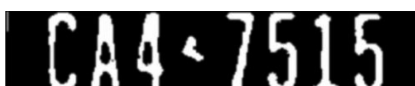


Рис. 5. Извлеченный из входного изображения государственный номер



Рис. 6. Сегментированные символы государственного номера

Предлагаемая в данной работе сверточная нейронная сеть используется в качестве экстрактора признаков и классификатора для распознавания символов английского алфавита и чисел. Регистрационный номер и идентификатор региона транспортного средства распознаются отдельными моделями сверточной сети. Первая архитектура обучается на бинаризованных изображениях и состоит из трех слоев свертки, пулинга и нормализации, а также двух полносвязных слоев. Во второй архитектуре количество слоев свертки увеличено до шести, а входные изображения находятся в оттенках серого.

Результаты исследования и их обсуждение

Оценка сегментации проводилась на базе данных из 394 изображений. Результаты сравнивались вручную путем оценки соответствия регистрационного номера и извлеченной информации. 372 изображения были сегментированы удовлетворительно, сегментация 22 изображений не удалась. Тем не менее некоторые из изображений с неудачной сегментацией были сегментированы частично. Это дает точность сегментации регистрационного номера 94,4%. Идентификатор региона был извлечен корректно для большинства изображений.

Первая модель сверточной нейронной сети при обучении с использованием би-

нарных изображений продемонстрировала точность распознавания 88,97% для идентификатора региона – 331 распознанное изображение из 372 – и 15,3% для государственного номера – 57 распознанных изображений из 372. Общая точность всех этапов составила 13% с учетом изображений, сегментированных некорректно.

Точность распознавания государственного номера для изображений в оттенках серого составила 91,12% – 339 верно распознанных изображений из 372. Несмотря на то, что эта модель сети обладает способностью различать «0» и «O», они считаются принадлежащими к одному и тому же классу из-за геометрических свойств символов. Точность распознавания идентификатора региона – 88,97%, или 331 распознанное изображение из 372. Общая точность для всех этапов составила 81,1% – 302 из 394 изображений сегментированы успешно и распознаны корректно.

Конкурс-тест систем распознавания автомобильных номеров, проведенный журналом «ProSystem CCTV» [7], показал результаты, приведенные в таблице ниже. Сравнение предложенной системы с ближайшими аналогами позволяет сделать вывод о возможности ее использования в коммерческих целях при следующих улучшениях:

1. Получение данных в режиме реального времени для их обработки и распознавания. К данным, полученным в режиме реального времени, относятся, например, изображения, собранные камерами видеонаблюдения с автостоянок, платных автодорог и т.д.

2. Использование локализации государственного номера.

3. Автоматическое выполнение этапов сегментации и локализации идентификатора региона.

Общая точность распознавания номерных знаков в сравнении с коммерческими системами-аналогами

Система распознавания автомобильных номеров	Общая точность распознавания
«Авто-Инспектор»	94,19%
«Auto-Trassin»	83,07%
«CVS-Авто»	82,09%
«АвтоУраган-ВСМ»	90,45%
Предложенная система	81,1%

Заключение

В работе исследовано использование методов глубокого обучения в области распоз-

навания символов текста на изображении. Предложена технология сегментации и распознавания текста применительно к задаче распознавания автомобильных номеров. Важнейшими аспектами построенной системы распознавания являются внедрение гибридного метода бинаризации, позволившего улучшить качество сегментации, и использование сверточных нейронных сетей для выделения признаков и распознавания номерных знаков.

Улучшение результатов может быть достигнуто за счет включения в предлагаемую методику этапа обнаружения идентификатора региона [8], а также реализации этапа сегментации номерного знака с использованием методов глубокого обучения [9]. Реализация всех этапов предложенной системы с использованием глубокого обучения может потребовать больших временных затрат и является сложной задачей из-за ограниченной доступности наборов данных. В качестве возможного решения этой проблемы может выступать использование современных графических процессоров.

Список литературы

1. Yang Y., Yan H. An adaptive logical method for binarization of degraded document images, *Pattern Recognition*. 2010. vol. 33. P. 787–807.
2. Gonzalez R. and Woods R. *Digital Image Processing*, 4th edition. Pearson Education, 2018. 1168 с.
3. Sauvola J., Pietikainen M. Adaptive document image binarization. *Pattern Recognition*. 2000. vol. 33. no. 2. P. 225–236.
4. Kim K.K., Kim K., Kim J., Kim H.J. Learning-based approach for license plate recognition. *Proceedings of the IEEE Signal Processing Society Workshop*. 2000. vol. 2. P. 614–623.
5. Du S., Ibrahim M., Shehata M. and Badawy W. Automatic license plate recognition (alpr): A state-of-the-art review. *IEEE Trans. Circuits and Syst. for Video Technol.* 2013. vol. 23. no. 2. P. 311–325.
6. Kresch R. and Malah D. Skeleton-based morphological coding of binary images. *IEEE Transactions on Image Processing*. 1998. vol. 7, no. 10. P. 1387–1394.
7. Распознавание автомобильных номеров: «Авто-Инспектор» – Лидер тестирования. [Электронный ресурс]. URL: https://www.adron-perm.ru/test_leader/ (дата обращения: 06.05.2020).
8. Sermanet P., Eigen D., Zhang X., Mathieu M., Fergus R., LeCun Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. 2013. arXiv preprint arXiv:1312.6229.
9. Badrinarayanan V., Kendall A. and Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017 vol. 39. no. 12. P. 2481–2495. DOI: 10.1109/TPAMI.2016.2644615.