Программа и алгоритм сегментации и распознавания рукопечатных символов с помощью сверточных нейронных сетей

Е.С. Попова, В.Г. Спицын, Ю.А. Болотова
esp9@tpu.ru|spvg@tpu.ru|jbolotova@tpu.ru
Национальный Исследовательский Томский Политехнический Университет, Томск, Россия

Статья посвящена разработке алгоритма сегментации и распознавания рукопечатных символов на изображениях. Излагается обобщенный алгоритм работы системы распознавания текстов. Приводится описание методов сегментации текстовых документов. Предлагается применение нейросетевого подхода, основанного на архитектуре сверточных нейронных сетей, для решения задачи распознавания рукопечатных символов. Результаты численных экспериментов по распознаванию бланков ЕГЭ на основе предложенного подхода показали точность распознавания 94.1%, что превосходит точность распознавания 82%, полученную с помощью системы ABBYY FineReader.

Ключевые слова: *компьютерное зрение, распознавание образов, сверточные нейронные сети, сегментация текстовых изображений.*

Program and algorithm for segmentation and recognition of hand-printed characters using convolutional neural networks

E.S. Popova, V.G. Spitsyn, Yu.A. Bolotova
esp9@tpu.ru|spvg@tpu.ru|jbolotova@tpu.ru
National Research Tomsk Polytechnic University, Tomsk, Russia

The article is devoted to the development of the algorithm for segmentation and recognition of hand-printed symbols on images. A generalized algorithm for the operation of the text recognition system is presented. The paper describes the methods of segmentation of text documents. The application of a neural network approach based on the architecture of convolutional neural networks is proposed to solve the problem of recognizing hand-printed characters. Results of numerical experiments on the recognition of USE forms on the basis of the proposed approach showed a recognition accuracy of 94.1%, which exceeds the recognition accuracy of 82% obtained with the help of ABBYY FineReader.

Keywords: *computer vision, pattern recognition, convolutional neural networks, segmentation of text images.*

1. Введение

На сегодняшний день существует множество направлений науки и техники, которые в значительной степени ориентированы на развитие систем, анализирующих информацию, представленную в виде изображений. Задача обработки и распознавания изображений относится к разряду трудно формализуемых задач, и является одной из наиболее важных на сегодняшний день.

Задача сегментации и распознавания текстовых областей, содержащих рукопечатные символы, является актуальной, так как в мире существует большое количество документов, нуждающихся в оцифровке.

Рукописные цифры или буквы, очевидны для человека, но для компьютеров идентификация таких символов — сложная задача, для решения которой часто применяются сверточные нейронные сети, являющиеся наиболее подходящими для анализа изображений.

Свёрточная нейронная сеть впервые была представлена в 1998 году французским исследователем Яном Лекуном, как развитие модели неокогнитрона, предназначенного для эффективного распознавания изображений.

Процесс распознавание текстовых областей на изображениях условно можно разделить на два этапа: этап сегментации изображения на строки, слова и символы, и этап распознавания символов.

Целью данной работы является разработка алгоритма сегментации и распознавания бланков ЕГЭ, которые на сегодняшний день могут обрабатываться как вручную, так и с помощью специализированного ПО. Однако, используемое программное обеспечение не всегда дает достаточную точность распознавания, необходимую для полной автоматизации процесса.

В данной работе проводится сравнение различных архитектур нейронных сетей и алгоритмов обучения, основанных на алгоритме градиентного спуска для решения поставленной задачи. Так же приводится описание алгоритма для сегментации текстового изображения, основанного на комбинации метода связанных компонент и порогового метода анализа гистограмм.

Далее в статье будут рассмотрены методы и алгоритмы, используемые для решения поставленной задачи.

2. Сегментация текстового изображения

2.1 Бинаризация изображения

Прежде чем произвести бинаризацию изображения, его необходимо перевести в градации серого, яркость пикселя вычисляется по формуле [1-3]:

$$I = R * 0.299 + G * 0.587 + B * 0.114,$$

где R, G, B — красный, зеленый и синий канал соответственно.

Для последующей бинаризации изображения яркость каждого пикселя $I(x,y)$ сравнивается с некоторым пороговым значением P. Если яркость пикселя больше порогового значения, то цвет пикселя принимается равным 1, иначе 0.

Глобальный порог бинаризации вычисляется по формуле [8, 9]:

$$P = \frac{I_{max}+I_{min}}{2},$$

где I_{max} – максимальное значение яркости изображения, I_{min} – минимальное значение яркости изображения.

The publication of the conference proceedings was supported by the RFBR, grant No. 18-07-20045\18

2.2 Сегментация изображения на строки и слова

Для сегментации фрагмента текста на строки и слова используется метод гистограмм.

Метод предполагает построение гистограммы для черных пикселей бинарного изображения. По оси *Y* располагается шкала распределения пикселей по количеству, а на оси *X* размещены номера строк, в случае сегментирования изображения на строки.

Алгоритм основан на предположении, что количество черных пикселей в межстрочных интервалах существенно меньше, чем в текстовых строках. Основываясь на этом предположении, определим каким должно быть наименьшее количество черных пикселей в строке, чтобы отнести ее к текстовой строке. Рассчитаем значение по формуле:

$$I = 0.1 * Str_{max},$$

где Str_{max} – максимальное число черных пикселей в строке изображения. Следовательно, используя найденный порог разделим все изображение на строки.

Работа алгоритма сегментации строк заключается в последовательном просмотре массива, содержащего количество черных пикселей для каждой строки и сравнение их с минимальным количеством N, затем выявлении множества пар индексов (s^1_i, s^2_i) строк, соответствующих границам печатных строк.

Аналогично осуществляется сегментация строк на слова, только в данном случае условием выделения пробелов является последовательность из *K* белых пикселей в строке изображении.

2.3 Метод связанных компонент

Для сегментации слов на символы в работе используется двухпроходной метод связанных компонент (МСК) [3,4]. Под выделением связных компонент понимают присвоение уникальной метки каждому объекту изображения. При следующем анализе данные метки служат в качестве идентификаторов при обращении к объектам.

Для описания алгоритма введем некоторые понятия. Обозначим через *I* матрицу изображения. Если $I(i, j) = 0$, то пиксель является фоновым. Если $I(i, j) = 1$, то пиксель принадлежит объекту. Через *L* обозначим двумерную матрицу (скан-маску), которая используется для хранения информации о метках и имеющую размеры, равные размерам изображения.

Первый проход по изображению осуществляется из верхнего левого угла, слева на право и сверху вниз. На внешнем цикле – перебор строк, на внутреннем – перебор столбцов строки, анализируются только соседи сверху и слева.

Каждый раз при обнаружении черного пикселя его соседи, принадлежащие скан-маске, исследуются для определения метки, которая будет присвоена рассматриваемому пикселю. Если в скан-маске содержатся только фоновые пиксели, то рассматриваемый пиксель получает новую промежуточную метку, если скан-маска содержит только один пиксель интереса, то рассматриваемый пиксель получает метку соседа. Если скан-маска содержит несколько точек интереса, то их промежуточные метки являются эквивалентными, среди них выбирается метка с наименьшим значением и пикселю присваивается значение выбранной метки-представителя.

После окончания первого прохода каждый объектный пиксель получает временную метку, на втором проходе значение метки уточняется.

Второй проход осуществляется в противоположенном направлении, снизу-вверх и справа налево, на втором проходе исследуются все соседние пиксели.

Следовательно, на втором проходе осуществляется поиск связей между маркированными пикселями различных категорий. Если связь найдена, то все «старшие» метки заменяются на «младшие».

3. Распознавание символов

3.1. Обоснования выбора сверточных нейронных сетей

Свёрточные нейронные сети, в отличие от других нейросетевых архитектур, обеспечивают частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям. Они объединяют три архитектурные идеи, для обеспечения инвариантности к изменению масштаба, повороту сдвигу и пространственным искажениям [6]:

- локальные рецепторные поля (обеспечивают локальную двумерную связность нейронов);
- общие синаптические коэффициенты (обеспечивают детектирование некоторых черт в любом месте изображения и уменьшают общее число весовых коэффициентов);
- иерархическая организация с пространственными подвыборками.

3.2. Обучающая выборка

В качестве наборов данных для обучения и тестирования будет использоваться обучающая выборка, содержащая 146324 изображений рукописных символов, включающая цифры и буквы русского алфавита.

Изображения в наборе данных имеют разрешение 32×32 пикселей и хранятся в формате оттенков серого, следовательно, каждое значение пикселя лежит в диапазоне от 0 (представляет белый цвет) до 255 (представляет черный цвет).

Для ускорения работы сети необходимо инвертировать значения пикселей следующим образом 0 – белый цвет, 255 – черный цвет, по следующей формуле:

$$I = 255 - x_i.$$

Для ускорения сходимости обучения сети значения входных пикселей изображений нормализуются по формуле:

$$y_i = \frac{x_i}{255},$$

где x_i– значение *i*-го пикселя изображения из базы, y_i – значение, подаваемое на вход сети.

3.3. Алгоритм обучения

В данной работе используется наиболее распространенный алгоритм обучения нейронных сетей, основанный на методе градиентного спуска (метод обратного распространения ошибки) и его модификации: SGD, Adam, AdaGrad, AdaDelta. [7, 10-13].

Алгоритм обратного распространения ошибки имеет несколько режимов обучения [7]. В данной работе используется подход с Mini-batch, как компромисс между последовательным и пакетным режимами, в этом случае корректировка синаптических весов сети происходит после небольшого количества обучающих образцов.

3.4. Функция активации

В данной работе используются функции активации ReLU и для выходного слоя Softmax.

ReLU является выпрямленной линейной функцией и на данным момент считается гораздо более простым и эффективным с точки зрения вычислительной сложности вариантом передаточной функции [5]:

$$f(S) = \max(0, S) = \begin{cases} 0, \text{при } S < 0 \\ S, \text{при } S \geq 0 \end{cases}.$$

Работа опубликована при финансовой поддержке РФФИ, грант 18-07-20045\18

На сегодняшний день существует семейство различных модификаций функции ReLU, решающих проблемы надёжности этой передаточной функции при прохождении через нейрон больших градиентов: Leaky ReLU, Parametric ReLU, Randomized ReLU.

Softmax функция активации разработана, чтобы превратить любой вектор с реальными значениями в вектор распределения вероятностей и определяется для *i*-ого нейрона следующим образом [7]:

$$z_i = \frac{\exp(y_i)}{\sum_{j=1}^{n} \exp(y_j)},$$

где z_i – искомое значение выхода *i*-ого нейрона, y_i – исходное значение выхода *i*-ого нейрона.

3.5. Функция потерь

Так как в качестве функции активации в выходном слое будет использоваться функция активации softmax, которая преобразует любой входной вектор в вектор вероятностей, то для сравнения двух вероятностных распределений необходимо выбрать корректную меру. В качестве такой меры будет использоваться перекрестная энтропия [3, 7]:

$$\mathrm{C} = -\sum_{j=1}^{n} t_j \log(y_j),$$

где t_j – требуемый выход для текущего обучающего примера, y_j – реальный выход нейронной сети.

3.6. Выбор библиотеки машинного обучения

Для подбора оптимальных параметров разрабатываемой сети, был произведен обзор существующих библиотек машинного обучения. На основе выбранной библиотеки спроектирована нейронная сеть с различными параметрами, для выявления наиболее подходящей конфигурации сверточной сети.

В качестве библиотеки машинного обучения для проектирования и тестирования сети была выбрана открытая нейросетевая библиотека Keras, которая позволяет на более высоком уровне работать с нейросетями. В качестве базовой библиотеки для вычислений Keras может использовать Theano и Tenzorfow, в данной работе использовалась библиотека машинного обучения Theano.

3.7. Обучение сети

Для решения поставленной задачи была спроектирована сверточная нейронная сеть с одним сверточным и подвыборочным слоем, размер матрицы свертки 5×5, размер окна подвыборки 2×2. Конфигурация сети представлена в таблице 1.

Таблица 1. Конфигурация сверточной нейронной сети

Тип слоя	Функция активации	Кол-во настраиваемых параметров
Слой свертки, Кол-во карт признаков: 32	ReLU	832
Слой подвыборки Кол-во карт признаков: 32	–	–
Полносвязный слой, кол-во нейронов:128	ReLU	802944
Полносвязный слой, кол-во нейронов:32	Softmax	4128
Общее кол-во настраиваемых параметров		807 904

В таблицах 2 и 3 приведен анализ влияния количества эпох обучения и размера минивыборки на точность распознавания сети для различных модификаций метода градиентного спуска.

Таблица 2.Влияние количества эпох на точность распознавания.

Размер мини выборки = 100				
Кол-во эпох	Алгоритмы оптимизации			
	SGD	Adam	AdaGrad	AdaDelta
5	81.03%	94.53%	96.14%	93.09%
10	84.41%	95.18%	96.30%	94.21%
20	88.42%	95.98%	96.46%	95.82%

Таблица 3. Влияние размера минивыборки на точность распознавания.

Количество эпох = 10				
Размер мини выборки	Алгоритмы оптимизации			
	SGD	Adam	AdaGrad	AdaDelta
50	89.23%	95.98%	96.46%	95.18%
100	84.41%	95.18%	96.30%	94.21%
200	83.92%	94.53%	96.46%	94.21%

Так же было проанализировано влияние количества карт признаков на точность распознавания, результат представлен в таблице 4.

Таблица 4. Влияние количества карт признаков на точность распознавания

Количество эпох = 10, размер минивыборки = 100				
Количество карт признаков выборки	Алгоритмы оптимизации			
	SGD	Adam	AdaGrad	AdaDelta
32	84.41%	95.18%	96.30%	94.21%
64	85.21%	95.34%	96.78%	95.98%

Исходя из полученных результатов были выбраны оптимальные параметры обучения и конфигурации сети:

- Алгоритм оптимизации: AdaGrad
- Количество эпох обучения: 10
- Размер минивыборки: 200
- Количество карт признаков: 32

4. Результаты тестирования

Для оценивания результатов классификации была рассчитана доля верно классифицированных объектов к общему количеству объектов. На данный момент лучшая точность распознавания на тестовой выборке, которой удалось добиться при реализации выбранной модели составляет 94.1%. Значение точности вычислялось по формуле:

$$R = \frac{n}{N} = \frac{623}{662} = 0.941\ ,$$

где R – точность распознавания по всему набору тестовой выборки, n – количество правильно распознанных символов из тестовой выборки, N – количество элементов в тестовой выборке.

На рисунке 1 представлены некоторые примеры неверного распознавания символов различных категорий.

The publication of the conference proceedings was supported by the RFBR, grant No. 18-07-20045\18

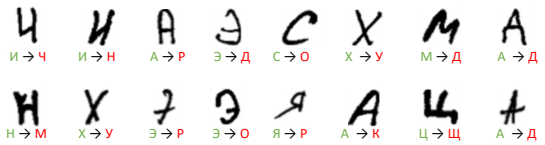

Рис. 1. Примеры не верно распознанных символов

Были проведены исследования по распознаванию бланков ЕГЭ с помощью системы ABBYY FineReader, которая показала точность распознавания 82%. В связи с этим, можно сделать вывод о том, что разрабатываемая система является перспективной, а дальнейшее ее улучшение – востребованным.

5. Заключение

В ходе работы опытным путем была выявлена наиболее подходящая архитектура сверточной нейронной сети для решения поставленной задачи, исследованы алгоритмы оптимизации на основе метода градиентного спуска и виды функции активации нейронов. Все тесты проводились с использованием базы данных рукописных символов русского алфавита.

Были исследованы алгоритмы сегментации текстовых изображений, на основе пороговых алгоритмов и метода связанных компонент.

6. Благодарности

Работа выполнена в рамках Программы повышения конкурентоспособности ТПУ при финансовой поддержке РФФИ в рамках научного проекта № 18-08-00977 А.

7. Литература

[1] Грузман И.С., Киричук В.С., Косых В.П., Перетягин Г.И., Спектор А.А. Цифровая обработка изображений в информационных системах: Учеб. пособие. – Новосибирск.: Изд-во НГТУ, 2003. – 352 с.

[2] Гонсалес Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М.: Техносфера, 2005. – 1072 с.

[3] Маркелов А.А. Алгоритмы и программная система классификации полутоновых изображений на основе нейронных сетей: Диссертация на соискание ученой степени кандидата технических наук / А. А. Маркелов. – Томск, 2007.

[4] Поршнев С.В., Левашкина А.О., Универсальная классификация алгоритмов сегментации изображений // Журнал научных публикаций аспирантов и докторантов [Электронный ресурс] — Электронный научный журнал – 2006. –Режим доступа: http://jurnal.org/articles/2008/inf23.html

[5] Спицын В.Г., Интеллектуальные системы: учебное пособие / В.Г. Спицын, Ю.Р. Цой; Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2012. – 176 с.

[6] Солдатова О.П., Гаршин А.А. Применение сверточной нейронной сети для распознавания рукописных цифр. Компьютерная оптика. – 2010. – Том 34,№2. – с. 252-260 – ISSN 0134-2452.

[7] Хайкин С. Нейронные сети: полный курс. М.: Вильямс, 2006. - 1104 с.

[8] Шапиро Л. Компьютерное зрение / Л. Шапиро, Дж. Стакан, 206. – 53 с.

[9] Введение в цифровую обработку изображений: лекция 3. [Электронный ресурс] – 2011. – Режим доступа: http://cvbeginner.blogspot.ru/2011/09/3.html

[10] LeCun, Y. Efficient BackProp in Neural Networks: Tricks of the trade / Y.LeCun, L. Bottou, G. Orr, K. Muller – Springer, 1998.

[11] LeCun, Y. Scaling learning algorithms towards AI / Y.LeCun, Y. Bengio – MIT Press, 2007.

[12] Mike O'Neill. Neural Network for Recognition of Handwritten Digits. [Электронный ресурс] Точка доступа: https://www.codeproject.com/Articles/16650/Neural-Network-for-Recognition-of-Handwritten-Digi

Об авторах

Спицын Владимир Григорьевич, д.т.н., профессор инженерной школы информационных технологий и робототехники Томского политехнического университета. E-mail spvg@tpu.ru.

Болотова Юлия Александровна, к.т.н., доцент инженерной школы информационных технологий и робототехники Томского политехнического университета. E-mail: jbolotova@tpu.ru.

Попова Екатерина Сергеевна, аспирант инженерной школы информационных технологий и робототехники Томского политехнического университета. E-mail: esp9@tpu.ru.

Работа опубликована при финансовой поддержке РФФИ, грант 18-07-20045\18