

РАЗРАБОТКА БИБЛИОТЕКИ НЕЙРОСЕТЕВОГО РАСПОЗНАВАНИЯ РУКОПИСНЫХ СИМВОЛОВ НА МАШИНОПЕЧАТАЕМЫХ БЛАНКАХ

*Н. Г. Авдеев, аспирант гр. А0-39
Томский политехнический университет
E-mail: ndb1@tpu.ru*

Введение

Контроль знаний, умений и навыков является важным звеном учебного процесса. От его организации во многом зависит результат обучения. В процессе контроля выявляются как достоинства, так и недостатки знаний и умений учащихся, что позволяет управлять учебным процессом, совершенствуя формы и методы обучения.

Массовые тестирования проводятся по регламентированным процедурам и требуют особого подхода к организации и обработке. К такому методу контроля знаний относится и Государственная итоговая аттестация (ГИА) – форма оценки степени и уровня освоения обучающимися основных образовательных программ, соответствующих требованиям федерального государственного образовательного стандарта. Для ознакомления учеников 9 и 11 классов с заданиями ГИА, а также с процедурой его проведения проводятся диагностические работы ОГЭ и ЕГЭ. Существующий формат проведения данного мероприятия ставит перед организаторами задачу анализа информации, представленной в виде изображений – произвести так называемое off-line распознавание уже написанного на бумаге текста. Задача обработки и распознавания изображений относится к разряду трудно формализуемых задач и является одной из наиболее важных на сегодняшний день.

Выбор технологии распознавания

Для поставленной задачи были выбраны сверточные нейронные сети. Выбор сверточных нейронных сетей обусловлен следующими факторами:

- Начиная с 2012 года сверточные сети занимают первые места на конкурсе ImageNet
- Сверточные нейронные сети устойчивы к аффинным преобразованиям
- Сверточные нейронные сети учитывают топологию изображений.

Подготовка обучающей выборки

Обучающая выборка была составлена из бланков репетиционного ЕГЭ. Бланки представляют собой монохромное изображение формата .tif, отсканированное с разрешением 300 dpi. Поля для вписывания букв на бланке имеют размер 64×65 пикселей. Для упрощения структуры нейронной сети разрешение изображений было уменьшено до размера 64×64 пикселя. Обучающая выборка была составлена из известных распознанных ответов. Вырезка букв осуществлялась по заранее известным координатам. Количество букв в каждом классе после обработки репетиционных бланков представлено на рисунке 1.

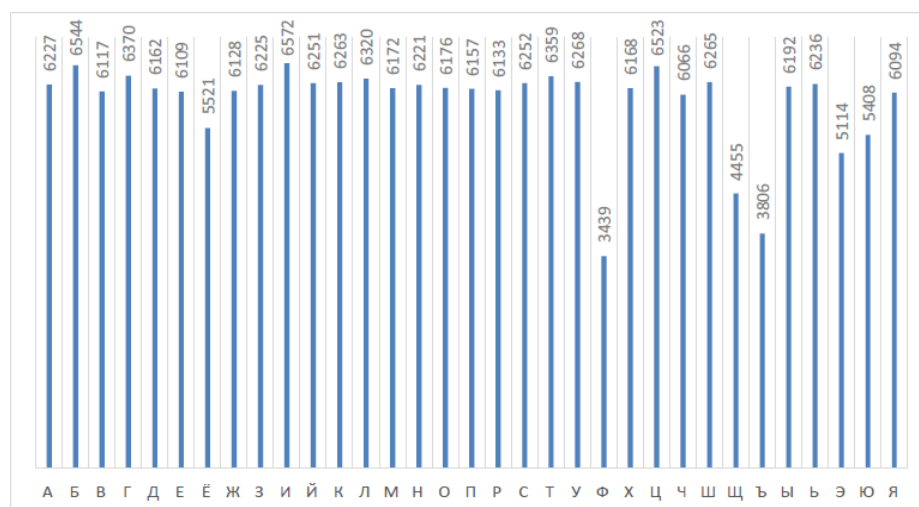


Рис. 1. Количество букв в обучающей выборке.

На рисунке 2 представлен пример изображений из обучающей выборки для букв класса «А».

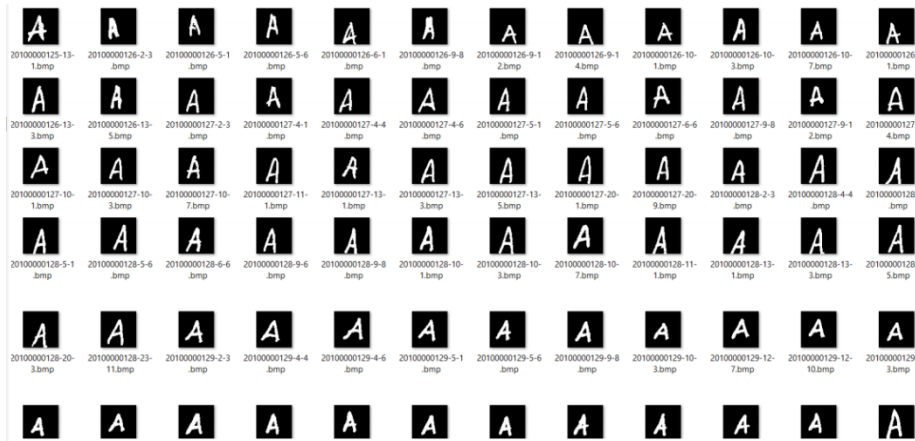


Рис. 3. Пример изображений из выборки для букв класса «А».

Изображения различных классов должны быть представлены в обучающей выборке примерно в одинаковых пропорциях. Если один из классов будет преобладать, то это может привести к «перекосу» в процессе обучения модели. Недостаток букв в классах «Ё», «Ф», «Ц», «Щ», «Ъ», «Э» и «Ю» был восполнен с помощью метода эластичных искажений.

Реализация метода распознавания на основе свёрточной нейронной сети

Для программы репетиционного экзамена было решено использовать две нейронные сети для распознавания следующих множеств символов:

- цифры, минус, запятая;
- символы русского алфавита.

Использование 2-х нейронных сетей необходимо для того, чтобы избежать неоднозначности в распознавании символов (цифра «0» и буква «О», цифра «3» и буква «з» и т. д.).

Были проведены эксперименты по подбору архитектуры нейронной сети. Количество слоев свертки варьировалось в пределах от 2 до 4, размер ядра свертки от 3x3 до 9x9, количество карт признаков на 1 слое – от 5 до 40. Количество карт признаков на каждом последующем слое удваивается. В подвыборочном слое использовалась операция MaxPooling с размером окна 2 и шагом 2. Для ускорения процесса обучение производилось на выборке, полученной следующим образом: из исходной выборки для каждого класса выдалось 100 случайных букв для обучения и 10 случайных букв для валидации результата. Было проведено 100 итераций обучения. Результирующее значение точности распознавания было получено путем 48 вычисления среднего арифметического из значений точности распознавания, полученных на каждой итерации.

На первом этапе эксперимента было протестировано 30 архитектур нейронных сетей. Количество слоев свертки варьировалось в пределах от 2 до 4, размер ядра свертки от 3x3 до 9x9, количество карт признаков на 1 слое оставалось равным 20. Архитектура выбиралась таким образом, чтобы не происходила потеря информации с изображения: если на одном из выходов сверточного слоя ширина или высота выходного изображения получалась нечетной, то при применении операции MaxPooling на подвыборочном слое крайние пиксели изображения не учитывались. Результаты эксперимента продемонстрированы на рисунке 3.

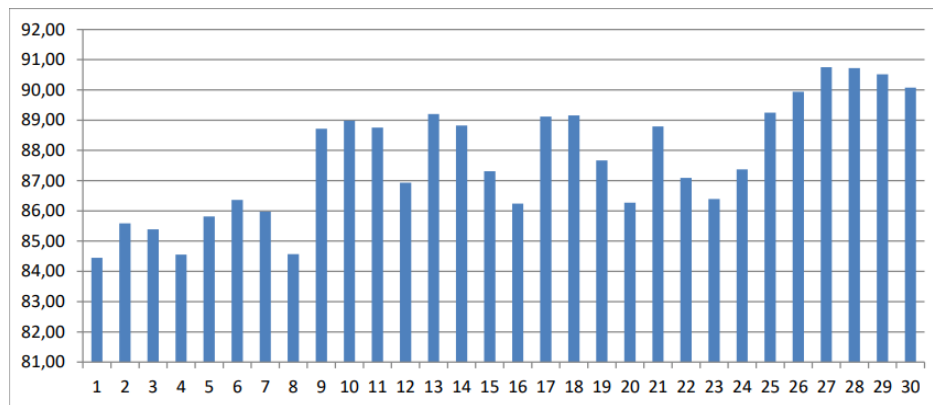


Рис. 4. Зависимость точности от архитектуры нейронной сети.

Исходя из результатов эксперимента, можно сделать вывод, что наибольшей точностью обладает нейронная сеть, которой соответствует 27-ой столбец диаграммы, изображенной на рисунке 4.24. Точность распознавания данной нейронной сети составляет 90.75%. Данная нейронная сеть имеет структуру, описанную в таблице 1.

Таблица 1. Архитектура выбранной сети для распознавания букв

Номер сверточного слоя	Размер ядра свертки	Количество карт признаков
1	5x5	20
2	3x3	40
3	7x7	80
4	3x3	160

Далее для данной архитектуры экспериментально было подобрано количество карт признаков. На рисунке 4 продемонстрирована зависимость точности распознавания от количества карт признаков на 1 слое нейронной сети (количество карт признаков на каждом последующем слое удваивается).

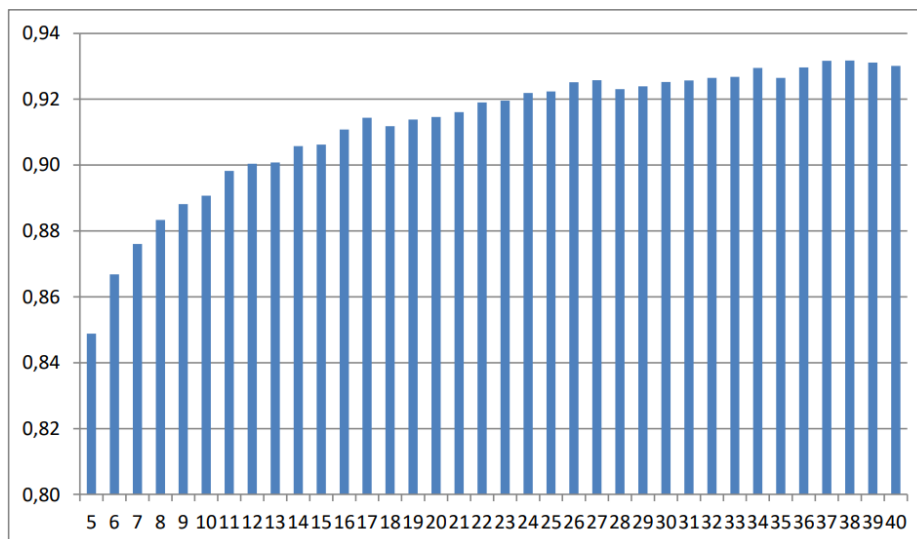


Рис. 5. Зависимость точности распознавания от количества фильтров на первом слое нейронной сети.

Исходя из полученных результатов, было решено взять количество фильтров на первом слое равным 25, так как дальнейшее увеличение значения количества фильтров не ведет к большому приросту точности, а ведет лишь к увеличению времени распознавания.

Аналогичный вышеописанному эксперимент был проведен для подбора архитектуры нейронной сети для распознавания цифр, минусов и запятых. Данная архитектура имеет структуру, описанную в таблице 2.

Таблица 2. Архитектура выбранной сети для распознавания цифр, минусов и запятых

Номер сверточного слоя	Размер ядра свертки	Количество карт признаков
1	5x5	25
2	3x3	50
3	3x3	100
4	5x5	200

Разработка библиотеки распознавания

Программная часть реализована в виде dll-библиотеки, написанной на языке программирования C#. Нейронная сеть сохранена в формате .onnx. Для запуска нейронной сети используется библиотека Microsoft ONNX Runtime. Диаграмма классов библиотеки продемонстрирована на рисунке 5.

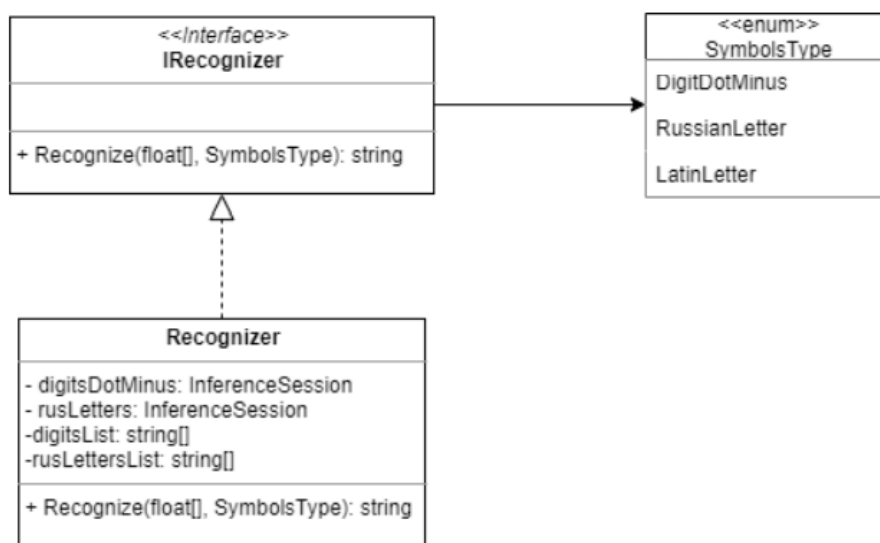


Рис. 6. Диаграмма классов разработанной библиотеки.

Был реализован интерфейс IRecognizer, который содержит метод Recognize. Создание данного интерфейса позволит использовать паттерн Dependency injection. У данного метода есть два аргумента: массив float[], которому соответствует изображение, преобразованное в одномерный массив. Значению 0 в данном массиве соответствует черный цвет на изображении, а значению 1 – белый. Вторым аргументом данного метода – перечисление SymbolsType, которое служит для определения нужного класса символов:

- цифры, минус, запятая;
- символы русского алфавита;
- символы латинского алфавита.

Шаблоны бланка устроены таким образом, что у каждого поля для ответа заранее predetermined класс вносимых в него символов. Класс Recognizer реализует интерфейс IRecognizer. У данного класса есть поля типа InferenceSession: digitsDotMinus и rusLetters. Данным полям соответствуют нейронные сети, которые выполняют распознавание. Кроме того в данном классе есть поля типа string[]: digitsList, rusLettersList. Данные поля нужны для сопоставления выходов нейронной сети символам.

Тестирование разработанной библиотеки

Было проведено 100 обучений нейронной сети для распознавания рукописных символов. Для каждого обучения выбирались лучшие значения точности распознавания валидационной выборки. Результирующее значение точности распознавания было получено путем усреднения полученных значений и составило 98.4 %.

Значение точности распознавания нейронной сети для распознавания цифр, минусов и запятых было получено аналогично вышеописанному способу и составило 99.7 %.

В силу того, что скорость работы верификатора зависит от процента верно распознанных ответов участника на бланке, имеет смысл вычислить данный показатель. Ответы вносятся участниками экзамена в поля для заполнения ответов. Пример распознаваемого бланка ответа продемонстрирован на рисунке 6.



Рис. 7. Пример распознаваемого бланка.

Ответы участника экзамена записаны в полях для символов, которые располагаются под надписью «результаты выполнения заданий с КРАТКИМ 53 ОТВЕТОМ», а также под надписью «Замена ошибочных ответов на задания с КРАТКИМ ОТВЕТОМ». Интерфейс программы распознавания, которая использует данную библиотеку, продемонстрирован на рисунке 7.



Рис. 8. Интерфейс программы для распознавания ответов.

Для сравнения результатов распознавания полученного программного продукта с программой Abbyy FormReader, которая используется в ЦОКО Томской области на данный момент, было распознано 300 бланков русского языка и 200 бланков базовой математики. Из бланков ответов русского языка были взяты только ответы, содержащие буквы. Результаты распознавания приведены в таблице 3.

Таблица 3. Результаты распознавания

Распознаваемый класс	Количество распознаваемых ответов	Количество верно распознанных ответов программой Abbyu FormReader	Количество верно распознанных ответов разработанной программой	Точность распознавания программы Abbyu FormReader	Точность распознавания разработанной программы
Цифры, минус, запятая	3722	2656	3178	71,4%	85,4%
Русские буквы	3534	2809	2244	79,5%	63,5%

Разработанная библиотека превзошла программу Abbyu FormReader по распознаванию ответов, содержащих цифры, минусы и запятые. Полученный результат обуславливается тем, что Abbyu FormReader часто ошибочно распознает запятые как символ «^».

Разработанная библиотека показала худший результат по распознаванию ответов, состоящих из русских букв. Это обуславливается тем, что программный продукт Abbyu FormReader лучше распознает схожие по написанию буквы, а также умеет распознавать строчные буквы, которые написаны с несоблюдением образца.

Список использованных источников

1. В.Г. Спицын, Интеллектуальные системы: учебное пособие / В.Г. Спицын, Ю.Р. Цой; Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2012. – 176 с.
2. CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more... [Электронный ресурс] – 2017. – Режим доступа: <https://medium.com/@sidereal/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5> (дата обращения: 01.02.2021).
3. SeNet – Winner Of ImageNet 2017 Classification Task (Squeeze-And-Excitation Networks) [Электронный ресурс] – 2017. – Режим доступа: <https://shaoanlu.wordpress.com/2017/08/17/senet-winner-of-imagenet-2017/> (дата обращения: 01.02.2021).
4. Affine transformation //Wikipedia [Электронный ресурс] – Режим доступа: https://en.wikipedia.org/wiki/Affine_transformation (дата обращения: 05.02.2021).
5. Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, «Gradient-based learning applied to document recognition» Proceedings of the IEEE, v. 86, pp. 2278- 2324, 1998.
6. Patrice Y. Simard Platt Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis / Patrice Y. Simard, Dave Steinkraus, John C. Platt – Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), 2003.
7. Dominic Masters Revisiting Small Batch Training for Deep Neural Networks / Dominic Masters, Carlo Luschi. [Электронный ресурс] Режим доступа: <https://arxiv.org/pdf/1804.07612.pdf> (дата обращения: 10.02.2021).
8. Ian H. Witten, Eibe Frank and Mark A. Hall Data Mining: Practical Machine Learning Tools and Techniques. — 3rd Edition. — Morgan Kaufmann, 2011. — 664 с.
9. Gentle Introduction to the Adam Optimization Algorithm for Deep Learning [Электронный ресурс] – 2017. Режим доступа: <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/> (дата обращения: 05.02.2021).
10. Калиновский И. А. Метод нейросетевого детектирования лиц в видеопотоке сверхвысокого разрешения: автореферат диссертации на соискание ученой степени кандидата технических наук : 05.13.11 / Калиновский Илья Андреевич. - Томск : [б. и.], 2016. URL: <http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000549544> (дата обращения 07.02.2021).