

УДК 004.383.8.032.26

Синтез искусственной нейронной сети для распознавания литературных источников по фотографиям

Щаников С.А., Борданов И.А., Данилин С.Н.

В данной статье описана информационная система распознавания книг по фотографиям, которая может быть использована в составе средств автоматизированного учета и инвентаризации книг в библиотеках и фондохранилищах. В основе разрабатываемой системы лежит применение искусственной нейронной сети (ИНС), обученной распознаванию печатных символов русского алфавита. Повышение точности распознавания достигается за счет коррекции ошибок в выходных данных книг с применением алгоритма вычисления расстояния Дамерау-Левенштейна и регулярных выражений.

Ключевые слова: искусственные нейронные сети, распознавание образов, моделирование, инженерное проектирование, точность.

Введение

В настоящее время задачи обработки и распознавания изображений очень популярны [1-3]. С каждым годом в мире появляется все больше аппаратных и программных средств фото- и видео фиксации (мобильные телефоны, камеры видеонаблюдения, спутниковые системы и т. д.), которые внедряются во все сферы человеческой деятельности. Данным технологиям можно найти применение и в библиотечной сфере, например, автоматизировав процесс распознавания книг путем создания информационных систем их инвентаризации и учета с применением мобильных устройств. Это позволит сократить время, затрачиваемое на проведение данных операций, снизить количество ошибок и оптимизировать требования к профессиональным навыкам работников.

Существующие в настоящее время системы основаны на распознавании штрих-кодов, в которых зашифрован ISBN (International Standard Book Number) книги. Применение таких систем требует наличия специализированного оборудования, такого как сканеры штрих-кодов. Это влияет на дополнительные затраты для его приобретения, а так же на удобство проведения инвентаризации и учета в случае стационарных сканеров. Стоит отметить, что в последнее время появляются и мобильные системы распознавания штрих-кодов [4].

Кроме того, не каждая старая или изданная не официально книга имеет ISBN. Для таких

книг нужно или придумывать собственную систему кодирования и последующего представления в виде штрих или QR кодов или искать другие варианты. Одним из подобных вариантов может являться разработка системы идентификации книги путем распознавания фотографии её внутренних или внешних элементов. Это такие элементы как титульный лист, контртитул, авантитул, фронтиспис, форзац, переплет, обложка, корешок и так далее.

Для систем распознавания наиболее удобным было бы использование внешних элементов книги. Для этих целей можно использовать, например, обложку, однако на обложке книги очень часто размещается множество бесполезной информации и иллюстраций. Минимально необходимая информация размещена на корешке книги. Так же корешок книги является более доступным, так как при размещении книг на полках и стеллажах он является основной внешней видимой частью книги. Такой подход будет очень полезен при создании мобильных систем учета и инвентаризации библиотечного фонда.

Таким образом, целью данной работы является создание автоматизированной системы распознавания книг по фотографиям.

Метод

В основе разрабатываемой системы лежит алгоритм, состоящий из следующих шагов:



1. Фиксация изображения с фотокамеры мобильного устройства.

2. Поиск корешка книги на изображении и геометрическая трансформация (поворот на определенный угол до горизонтальной ориентации).

3. Предварительная обработка, целью которой является сегментация и бинаризация области корешка книги на изображении.

4. Сегментация символов и пробелов на изображении корешка книги.

5. Распознавание букв.

6. Коррекция ошибок распознавания.

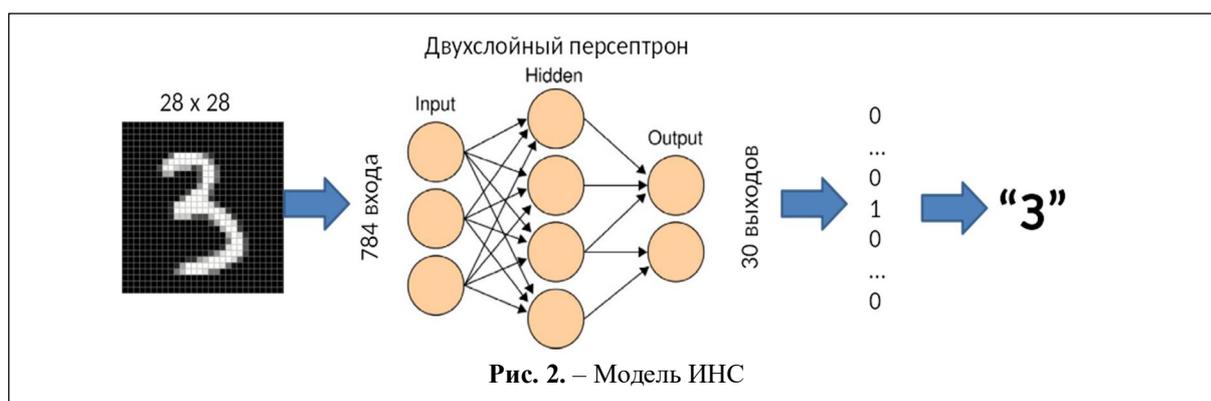
7. Добавление/редактирование данных об экземпляре книги в базе данных библиотечного фонда.

Для предварительной обработки изображения используется библиотека `opencv` и язык программирования Python. В процессе предобработки на изображении выделяется корешок книги, определяется угол его наклона и изображение поворачивается до горизонтального состояния, обрезается и бинаризуется. В основе алгоритма сегментации корешка книги и букв на корешке лежит функция библиотеки `OpenCV` для выделения контуров `findContours`. Для каждого контура символа так же фиксируются координаты X и Y,

чтобы можно было понять, какой строке принадлежит символ (по координате Y) и где пробел между словами (по координате X) (рис. 1).

Как показано в ряде работ Галушкина А.И., в частности в [5] для решения различных задач распознавания и преобразования изображений по всем основным критериям наиболее подходят ИНС, так как базис задач соответствует базису вычислительной системы. ИНС обладают потенциальными преимуществами по точности, отказоустойчивости, быстродействию, надёжности, энергозатратам, перед информационными системами с последовательными технологиями обработки информации [6,7].

Для распознавания букв создана двухслойная ИНС прямого распространения (рис. 2). Она имеет 784 входа, 100 нейронов в скрытом слое и 30 выходных нейронов. На вход ИНС подается изображение буквы размером 28x28 пикселей, представленное в виде вектора размером 784 элемента. Каждый элемент вектора принимает значение от 0 до 255 в оттенках серого до нормализации и от 0 до 1 после нормализации. На выходном нейроне ИНС, соответствующем распознаваемой букве должно быть «1», а на остальных «0». Для создания ИНС используется модуль `NeuroLab`.



Математическая модель ИНС следующая

$$\hat{y} = f \left(\sum_{j=1}^{N_{\text{hidden}}} w_{\text{out}j} \cdot f \left(\sum_{i=1}^{N_{\text{input}}} w_{ij} \cdot x + b_j \right) + b_{\text{out}} \right), \quad (1)$$

где \hat{y} — выходная информация ИНС; j — номер нейрона скрытого слоя ($N_{\text{hidden}} = 2$); i — номер нейрона входного слоя ($N_{\text{input}} = 2$); w_{out} — массив весовых коэффициентов синапсов выходного нейрона, w — весовые коэффициенты синапсов нейронов скрытого слоя, x — входная информация ИНС; b , b_{out} — пороговые смещения скрытого и выходного слоя.

Функция активации нейронов всех слоев — гиперболический тангенс. Для обучения ИНС создана обучающая выборка, состоящая из 30 букв русского алфавита (за исключением букв «Ё», «Й», «Ъ»), написанных разными шрифтами (10 наиболее распространенных шрифтов). Обучение персептрона проведем методом обратного распространения ошибки на основе квазиньютоновского алгоритма (данный метод реализован в модуле NeuroLab). Точность в процессе обучения оценим с помощью функции среднего квадрата ошибок MSE (mean squared error)

$$\text{MSE} = \frac{1}{H} \cdot \sum_{h=1}^H (y_h - \hat{y}_h)^2, \quad (2)$$

где H — общее количество элементов обучающей, валидационной или тестовой выборки, y_h — значение функции для h -го значения выборки, \hat{y}_h — значение выходной информации ИНС для h -го значения выборки.

Точность распознавания оценим по формуле

$$P_{\text{ош}} = \frac{E}{H} \cdot 100\%, \quad (3)$$

где E — количество неверно распознанных букв.

Таким образом каждая буква после сегментации масштабируется до размера 28x28 пикселей и представляется в виде вектора размером 784. Данный вектор подается на вход ИНС и на выходе определяется нейрон с максимальным значением. Так в цикле каждая буква подается на вход ИНС и в результате работы алгоритма получается слово.

Для коррекции ошибок в распознанном слове вычисляется расстояние Дамерау-Левенштейна. Это мера разницы двух строк символов, определяемая как минимальное количество операций вставки, удаления, замены и транспозиции (перестановки двух соседних символов), необходимых для перевода одной строки в другую. Название книги после коррекции ищется в базе данных (БД) с помощью регулярных выражений. Применение данных средств (расстояние Дамерау-Левенштейна и регулярные выражения) повышает вероятность найти правильную запись о книге в БД.

Результаты и обсуждение

Проведем исследование разработанного алгоритма. Для этого создадим несколько фотографий книги, расположив её корешком к камере мобильного телефона в хорошо освещаемом помещении на светлом однородном фоне. Сохраним полученные изображения в папку с исходным кодом программы и запустим её.

Программа предварительно выделила контуры на изображении. Таких контуров всегда получается много, однако в программе есть условия, которые позволяют отфильтровать контуры с диагональю меньше, чем 75% от диагонали снимка и с отношением сторон менее чем 1:3. После выделения контура алгоритм определил координаты прямоугольника и затем по теореме Пифагора рассчитал угол, на который нужно повернуть изображение, чтобы корешок книги стал изображен на фото строго горизонтально.

Следующим этапом стала обрезка изображения по имеющимся координатам. Данные координаты получены после поворота на заданный угол из исходных. После обрезки к изображению применена пороговая обработка. Она позволила отсечь фон от букв, для того чтобы далее для каждой буквы можно было найти контур. Результат данной операции представлен на рис. 3.



Далее на бинаризованном изображении ко-решка находим буквы, сегментируем их (рис. 4), масштабируем и подаем в ИНС.

Перейдем к обучению ИНС. Как видно из рис. 5 ИНС обучалась до заданного целевого значения (красная линия), которое было задано на уровне 0.005, по критерию среднего квадрата ошибок.

Обучение прошло за 364 эпохи и продолжалось около 20 минут. Это достаточно быстрый результат, и он обуславливается маленькой обучающей выборкой и размером изображения и ИНС. Для улучшения результата распознавания нужно увеличивать обучающую выборку и добавлять больше слоев и нейронов в ИНС.

Для распознавания на вход ИНС нужно подать изображение буквы 28x28 пикселей преобразованное в вектор размером 784. На 30 выходах ИНС появляются определенные значения, которые в лучшем случае должны быть распределены таким образом, что на одном из 30 нейронов ожидается цифра 1, а на других 29 – 0. Однако на выходе ИНС всегда есть погрешность и значения далеки от ожидаемых.

Результат тестирования ИНС для каждого символа и каждого выходного нейрона представлен на рис. 6. Как видно из данного рисунка на выходах ИНС, значения группируются в районе 0,7-0,8 и 0,1-0,2. Однако для некоторых случаев данные значения сильно отличаются, как например показано на рис. 7.

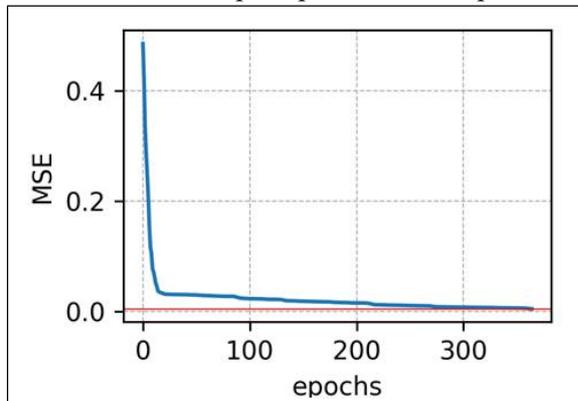


Рис. 5. – Кривая обучения ИНС

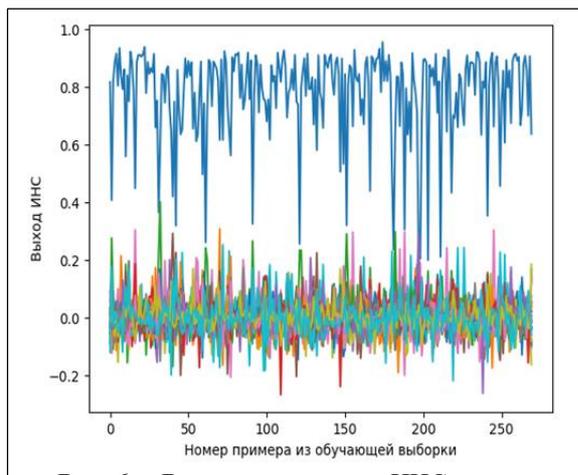


Рис. 6. – Выходные сигналы ИНС со всех нейронов. Выходные сигналы сгруппированы для всех букв. Синяя кривая - все значения выходных сигналов для распознаваемых букв остальные 29 кривых - все значения выходных сигналов на 29 оставшихся выходных нейронах

Из рис. 6 и 7 видно, что экспериментальным образом можно подобрать некоторый порог распознавания, при котором можно считать, что нейронная сеть распознала ту или иную букву. Для этого в цикле будем изменять значение порога от 0 до 1 и считать вероятность распознавания. Результаты представим на рис. 8.

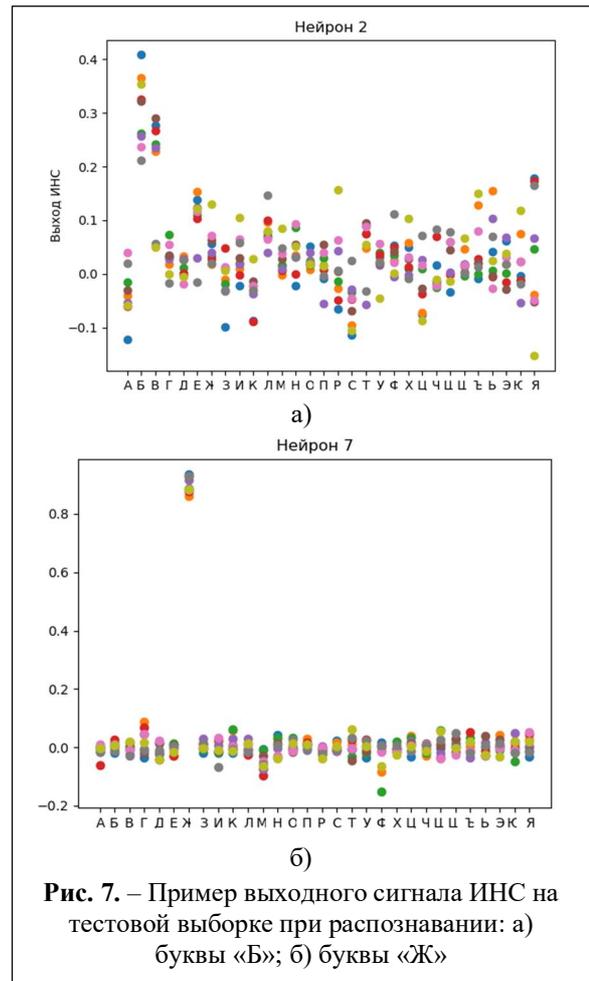


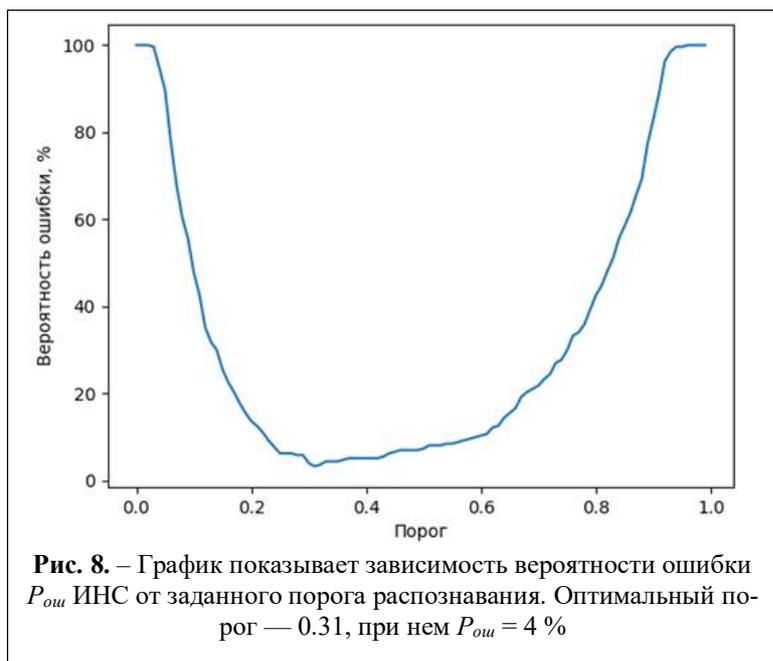
Рис. 7. – Пример выходного сигнала ИНС на тестовой выборке при распознавании: а) буквы «Б»; б) буквы «Ж»

Из рис. 8 видно, что при значениях порога 0 и 1 вероятность ошибки равна 100 %, однако при приближении значения порога к 0,31 вероятность ошибки снижается до 4%.

Заключение

В результате проведенной работы создана автоматизированная система распознавания книг по фотографиям корешка. В её основе лежит двухслойная ИНС прямого распространения, которая имеет 784 входа, 100 нейронов в скрытом слое и 30 выходов. Работа системы протестирована на реальных фотографиях книг. Созданная система может быть использована в составе средств автоматизированного учета и инвентаризации книг в библиотеках и фондохранилищах. К нерешенным задачам и путям дальнейшего развития можно отнести:

- увеличение обучающей выборки для улучшения качества распознавания отдельных символов;



- увеличение количества выходных нейронов сети для обеспечения возможности распознавания большего количества символов (знаки препинания и английские символы);

- улучшение алгоритмов предварительной обработки для обеспечения возможности распознавания названий книг, написанных в несколько строчек;

- модернизация алгоритма коррекции ошибок на основании имеющихся в базе данных названий книг.

Литература

1. Захаров А.А., Булаев А.В., Жизняков А.Л. Разработка методов визуального внимания на основе структурного представления признаков

Работа выполнена при поддержке гранта РФФИ №19-07-01215.

Поступила 11 сентября 2019 г.

This article describes an information system for recognizing books by photos, which can be used as part of automated accounting and inventory of books in libraries and storage facilities. The system under development is based on the use of an artificial neural network (ins) trained to recognize printed characters of the Russian alphabet. Improved recognition accuracy is achieved by correcting errors in the output data of books using the Damerau-Lowenstein distance calculation algorithm and regular expressions.

Key words: artificial neural networks, pattern recognition, modeling, engineering design, accuracy.

Щаников Сергей Андреевич – кандидат технических наук, декан факультета информационных технологий Муромского института (филиала) федерального государственного образовательного учреждения высшего образования "Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых".

E-mail: seach@inbox.ru

изображений // Оптико-электронные приборы и устройства в системах распознавания образов и обработки изображений. распознавание – 2019 сборник материалов XV Международной научно-технической конференции. 2019. С. 86-88.

2. Варламов А.Д. Машинное обучение анализу изображений на предмет спама // Алгоритмы, методы и системы обработки данных. 2014. №4(29). С. 31-36.

3. Жиганов С.Н., Сеннов Д.В. Комплексирование разночастотных изображений одного и того же участка земной поверхности // DSPA: Вопросы применения цифровой обработки сигналов. 2017. С. 149-151.

4. Астафьев А.В., Орлов А.А. Алгоритм локализации изображений символьных и штрих-кодовых мар-

кировок промышленных изделий для контроля движения продукции // Устойчивость и процессы управления. Материалы III международной конференции. 2015. С.559-560.

5. Галушкин, А. И. Нейронные сети: основы теории. - М.: Горячая линия-Телеком, 2013. - 496 с.

6. Данилин С.Н., Щаников С.А., Борданов И.А., Зуев А.Д., Пантюхин Д.В., Пантелеев С.В. Состояние исследований в области инженерного проектирования и производства нейрокомпьютеров // Алгоритмы, методы и системы обработки данных. 2019. №1(39). С.14-15.

7. Щаников С.А., Никитаев А.С., Борданов И.А. Перспективные технологии автоматизации производственных процессов при переходе к цифровой индустрии с использованием нейрокомпьютеров на базе мемристоров // Методы и устройства передачи и обработки информации. 2018. № 20 (20). С. 70-79.

Борданов Илья Алексеевич - инженер кафедры информационных систем Муромского института (филиала) федерального государственного образовательного учреждения высшего образования "Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых".

E-mail: bordanov2011@yandex.ru

Данилин Сергей Николаевич - кандидат технических наук, доцент программной инженерии Муромского института (филиала) федерального государственного образовательного учреждения высшего образования "Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых".

E-mail: dsn-55@mail.ru.

Адрес: 602264, г. Муром, ул. Орловская, д. 23.