

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.93'12

Тодрик
Антоний Иванович

Использование нейронных сетей для распознавания текста на
изображениях

АВТОРЕФЕРАТ

на соискание степени магистра информатики и вычислительной техники
по специальности 1-40 81 01 «Информатика и технологии разработки
программного обеспечения»

Научный руководитель
Теслюк Владимир Николаевич
доцент, кандидат физико-
математических наук

Минск 2018

ВВЕДЕНИЕ

В настоящее время компьютеры проникли во все сферы жизнедеятельности человека. Они используются в производстве, медицине, для работы, учебы и развлечений. Часто возникают задачи, когда требуется определить наличие каких-либо признаков, параметров, выявить закономерности в наборе данных, сделать какие-либо прогнозы, например, определить курс акций на завтра или перевести записанную речь в текст. Такие задачи в большинстве своем являются нетривиальными, для их решения нет единого подхода, который подойдет для всех случаев.

Одними из видов упомянутых задач являются распознавание графических образов и, его частный случай, распознавание текста. От компьютера требуется по изображению определить улыбается человек или нет, определить его возраст, пол, расу, определить изображена ли на картинке собака или кот, найти и считать с картинки текст, чтобы представить его в строковом виде, удобном для обработки. Некоторую сложность в решение проблемы вносят такие факторы, как шум на изображении, различные вариации расположения искомого объекта.

Для распознавания можно применить следующие методы: перебор видов объекта с разным масштабом, под разными углами, с различными смещениями; определение и исследование контура объекта и его характеристик (форма, углы и т.д.); использование технологий машинного обучения (потребуется большое количество примеров распознавания с корректным результатом).

Ранее методы машинного обучения были намного менее доступны в следствие необходимости большого количества вычислений и низких вычислительных мощностей, сегодня же современные процессоры и видеокарты позволяют реализовывать такие алгоритмы на домашнем компьютере.

В данной работе рассматривается использование нейронных сетей для распознавания текста на изображениях, в частности, на автомобильных номерах.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задача исследования

Целью диссертационной работы является исследование возможности применения искусственных нейронных сетей для распознавания текста на изображениях, в частности для распознавания автомобильных регистрационных номеров.

Для реализации поставленной цели необходимо выполнить промежуточные задачи исследования:

- провести сравнительный анализ возможных решений;
- описать возможные виды используемых нейронных сетей;
- подготовить данные для обучения сетей;
- сравнить различные конфигурации сетей, сравнить изменение параметров топологии на точность распознавания.

Объектом исследования является возможность распознать текст на изображениях. Предметом исследования является применение для этого нейронных сетей.

Связь работы с приоритетными направлениями научных исследований

В современном мире человек постоянно сталкивается с задачей распознавания (классификации) образов. Растет необходимость производить классификацию в автоматическом режиме с достаточной точностью без участия человека. Т.к. в подавляющем большинстве случаев классификация является трудно формализуемой задачей, традиционные алгоритмы слабо применимы. Но с этой задачей хорошо справляются искусственные нейронные сети, аппарат которых активно исследуется и развивается в последние годы. В работе рассматривается конкретный случай применения нейронных сетей для распознавания текста на автомобильных номерах.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя, В. Н. Теслюка, заключается в формулировке целей и задач исследования.

Структура и объем диссертации

Пояснительная записка по диссертационной работе включает в себя оглавление, общую характеристику работы, введение, основную часть, состоящую из 4 глав, заключения, списка использованной литературы.

Первая глава содержит анализ предметной области, сравнение возможных путей решения поставленной задачи. Также глава содержит краткий анализ наиболее важных работ за последнее время.

Вторая глава содержит теоретическое описание искусственных нейронных сетей, их концепции, видов. Также глава содержит описание алгоритмов выбора области изображения с искомым объектом. В главу включен обзор используемых программных средств и технологий, использованных при разработке, таких как язык программирования Python, библиотека для обработки изображений OpenCV, набор инструментов Microsoft Cognitive Toolkit.

В третьей главе описана используемая модель системы, используемые подходы для обнаружения области с автомобильным номером и выделения областей с символами. Также в главе описана подготовка обучающих данных к использованию, дано сравнение нескольких конфигураций многослойных перцептронов и сверточных сетей, описан используемый подход для визуализации промежуточных слоев сверточной сети.

В заключении подводятся итоги проделанной работы, делаются выводы и описывается дальнейший план развития проекта.

Общий объем работы составляет 51 с., 24 рис., 3 табл., 29 источников.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В рассматриваемой задаче классификации(распознавания) символов на автомобильных номерах классы соответствуют 22 буквам и цифрам латинского алфавита (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, E, I, K, M, N, O, P, S, T, X). Показателем работы сети будем считать значение ошибки распознавания, которое равно отношению количества неверно распознанных примеров к верно распознанным, и значение ошибки n -лучших (пример правильно классифицируется, если он присутствует среди n -лучших предположений классификатора).

Для обучения сети использовался подход *train-test*, т.е. сеть обучается на одном наборе данных (2200 изображений), а итоговая ошибка сети вычисляется на другом (660 изображений), не пересекающимся с первым.

Рассмотрим использованную процедуру обучения сетей. Обучение проводилось итерациями (эпохами). На каждой итерации количество примеров, которые могут быть обработаны, было ограничено сверху. В рамках одной итерации примеры обрабатывались пакетами по 32 примера. Пакетная обработка примеров позволяет значительно снизить время обучения за счет возможности параллельно обрабатывать элементы одного пакета. Статистика для пакетов различных размеров представлена в таблицах 1, 2, 3.

Таблица 1 – Статистика для пакета размером 1

Итерация	Потери	Метрика, %	Затраченное время, сек	Скорость, примеров/сек
1	1,261526	34,41	8,246	266,8
2	1,026506	25,05	6,745	326,2
3	0,970758	22,00	6,748	326,0
4	1,141645	23,91	6,830	322,1
5	0,951448	20,77	6,834	321,9

Таблица 2 – Статистика для пакета размером 32

Итерация	Потери	Метрика, %	Затраченное время, сек	Скорость, примеров/сек
1	1,003463	26,45	1,853	1187,3
2	0,510983	14,73	0,417	5275,8
3	0,398598	11,09	0,418	5263,2
4	0,380279	10,95	0,417	5275,8
5	0,352885	10,09	0,418	5263,2

Таблица 3 – Статистика для пакета размером 64

Итерация	Потери	Метрика, %	Затраченное время, сек	Скорость, примеров/сек
1	1,311655	32,91	1,730	1271,7
2	0,646902	16,77	0,308	7142,9
3	0,482456	12,73	0,301	7309,0
4	0,443054	12,50	0,305	7213,1
5	0,413858	11,32	0,302	7284,8

Как видим, при увеличении размера пакета растет не только скорость его обработки, но и суммарная ошибка классификации. Поэтому пакет размером 32 примера является разумным компромиссом между скоростью и точностью.

После обработки каждого пакета вычисляются суммарная разность полученных и желаемых результатов и суммарное значение функции ошибки, которое затем используется для обновления синапсов нейронов. После завершения каждой итерации выводилась собранная статистика, полученная модель сохраняется в файл.

Рассмотрим три персептрона, которые содержат входной слой с 784 нейронами, один скрытый с 600 нейронами, используют логистическую функцию активации. Первый для обновления параметров будет использовать стохастический градиентный спуск со скоростью 0,1, второй – градиентный спуск с моментом, позволяющий рассчитывать изменения весов с учетом предыдущих изменений, третий – алгоритм adadelata, позволяющий регулировать скорость для каждого параметра индивидуально, с базовой скоростью 0,1. График обучения изображен на рисунке 1.

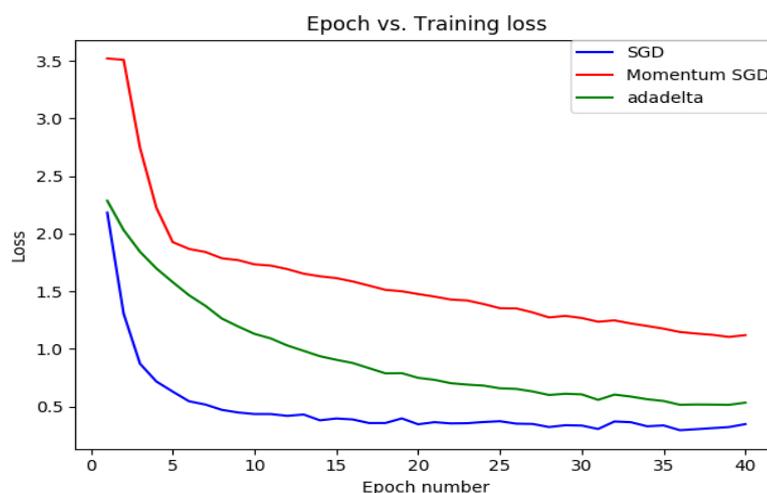


Рисунок 1 – Использование SGD, momentum SGD, adadelata для изменения параметров персептрона

Практика показывает, что модели, использующие адаптивное изменение скорости обучения, более склонны к переобучению (т.е. сеть "заучивает" данные), что ведет к росту ошибки на незнакомых примерах.

Сравним следующие 3 модели персептронов: все используют входной слой с 784 нейронами, используют стохастический градиентный спуск и сигмоиду, как функцию активации, но первый содержит три скрытых слоя с 600, 450 и 200 нейронами, второй – 2 скрытых слоя с 600 и 400 нейронами, третий – один скрытый слой с 600 нейронами.

Первая конфигурация имеет ошибку 15,15%, ошибку 5-лучших – 4,69%, вторая – ошибку 10,15% и ошибку 5-лучших 5,0%, третья – ошибку 9,69% и ошибку 5-лучших 4,54%. График обучения изображен на рисунке 2.

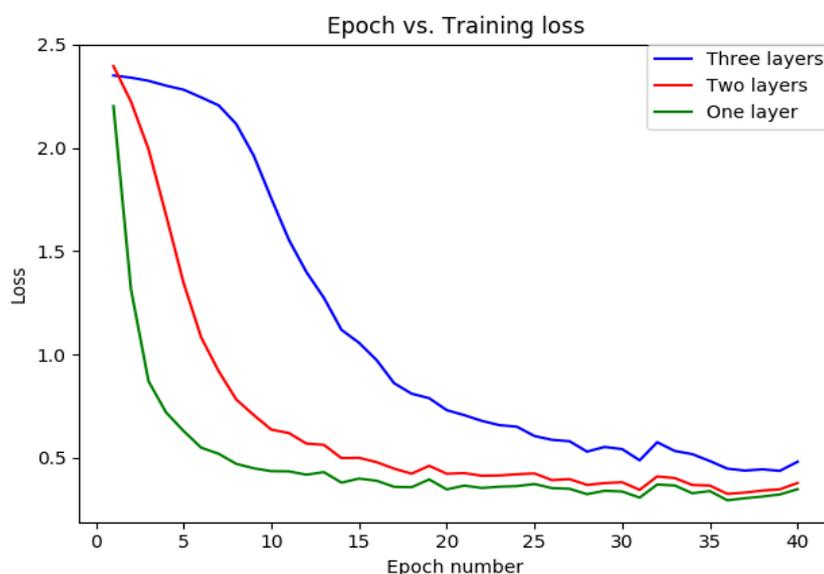


Рисунок 2 – Использование персептронов с разным количеством слоев

Несмотря на то, что ошибка 5-лучших для трехслойного персептрона имеет более низкое значение, чем для двухслойного, можно сделать вывод, что для конкретной задачи распознавания на автомобильных номерах, увеличение количества скрытых слоев не даст значительного прироста точности.

Рассмотрим влияние функции активации на значение ошибки. Пусть имеется три персептрона. Все используют входной слой с 784 нейронами, стохастический градиентный спуск и один скрытый слой с 600 нейронами, но первый используют в качестве функции активации сигмоиду, второй – функцию гиперболического тангенса. Третий использует функцию ReLU.

Первая модель имеет ошибку 9,29% и ошибку 5-лучших 4,54%, вторая – ошибку 7,27% и ошибку 5-лучших 2,87%, третья – ошибку 3,33% и ошибку 5-лучших 1,21%. График обучения изображен на рисунке 3.

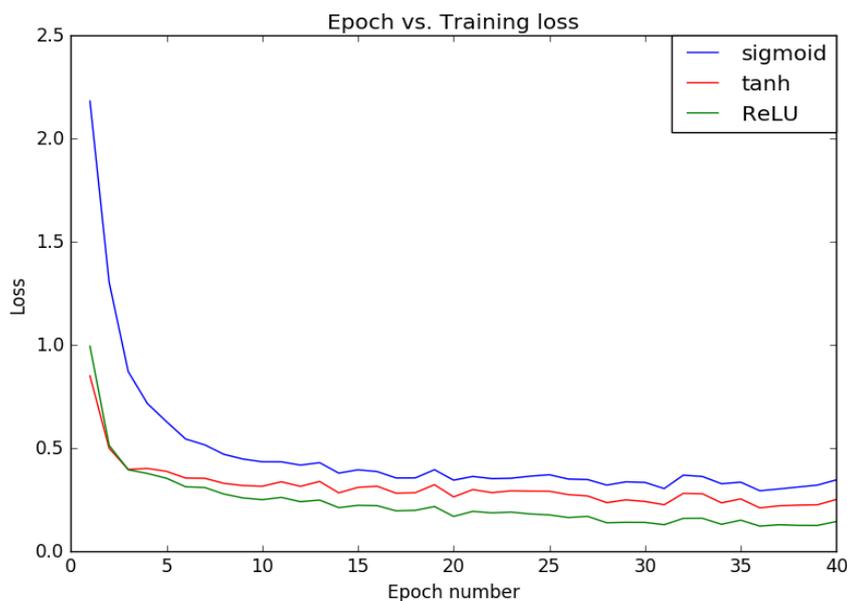


Рисунок 3 – Использование различных видов функции активации

Функция ReLU показывает более высокие результаты, т.к. она более способствует сходимости градиента ошибки и требует меньше вычислительных ресурсов.

В описании сверточных сетей будем записывать их структуру используя следующие обозначения:

- Conv((5, 5), 32) – сверточный слой, в котором используются 32 фильтра размером 5x5;
- MaxPooling((3, 3), (2, 2)) – слой субдискретизации с фильтром размером 3x3 и шагом 2 по горизонтали и 2 по вертикали;
- Dense(96) – полносвязный слой, состоящий из 96 нейронов;
- Dropout(0,5) – дропаут слой, в котором выходной сигнал нейрона будет сброшен в нуль с вероятностью 50%.

Рассмотрим влияние количества сверточных слоев на точность классификации. Пусть есть три сети. Первая имеет топологию Conv((5, 5), 32), MaxPooling((3, 3), (2, 2)), Conv((3, 3), 48), MaxPooling((3, 3), (2, 2)), Dense(96), Dropout(0,5). Вторая – Conv((5, 5), 32), MaxPooling((3, 3), (2, 2)), Dense(96), Dropout(0,5). Третья – Conv((5, 5), 32), MaxPooling((3, 3), (2, 2)), Conv((3, 3), 48), MaxPooling((3, 3), (2, 2)), Conv((3, 3), 64), MaxPooling((3, 3), (2, 2)), Dense(96), Dropout(0,5). График обучения изображен на рисунке 4.

Первая сеть имеет ошибку 1,67%, вторая – 1,52%, третья – 1,36%. Как видно из графика сети быстро сходятся к одному уровню ошибку уже на 10 итерации. Можно сделать вывод, что в конкретной задаче увеличение количества сверточных слоев не даст значительного прироста точности.

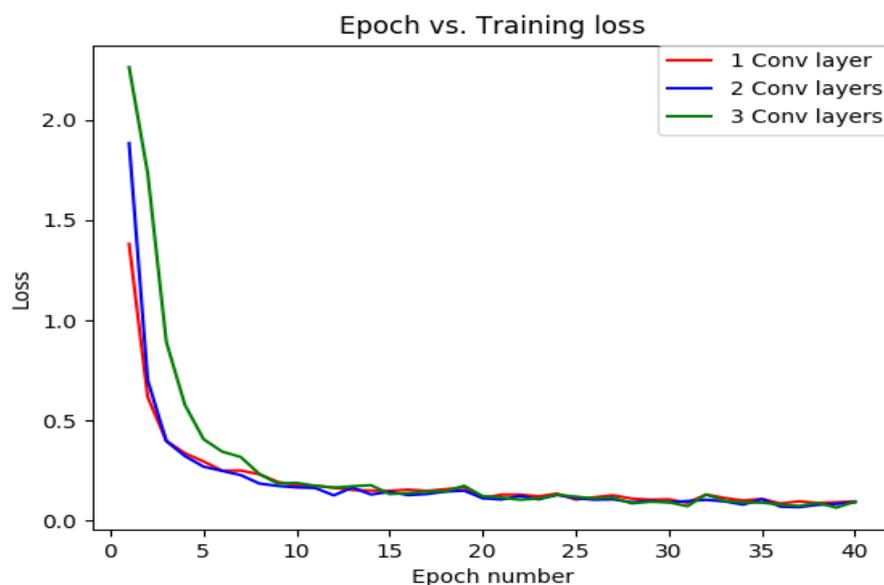


Рисунок 4 – Использование разного количества сверточных слоев

На рисунке 5 приведен график изменения значения функции потерь при использовании Dropout слоев: Dropout(0,0), Dropout(0,3), Dropout(0,6). Первая конфигурация имеет ошибку классификации 1,52%, вторая – ошибку 1,36%, третья – ошибку 1,06%. Несмотря на то, что самое низкое значение функции потерь наблюдается при использовании Dropout(0,0), наиболее точной является модель с Dropout(0,6), что говорит о лучшей способности к обобщению незнакомых данных.

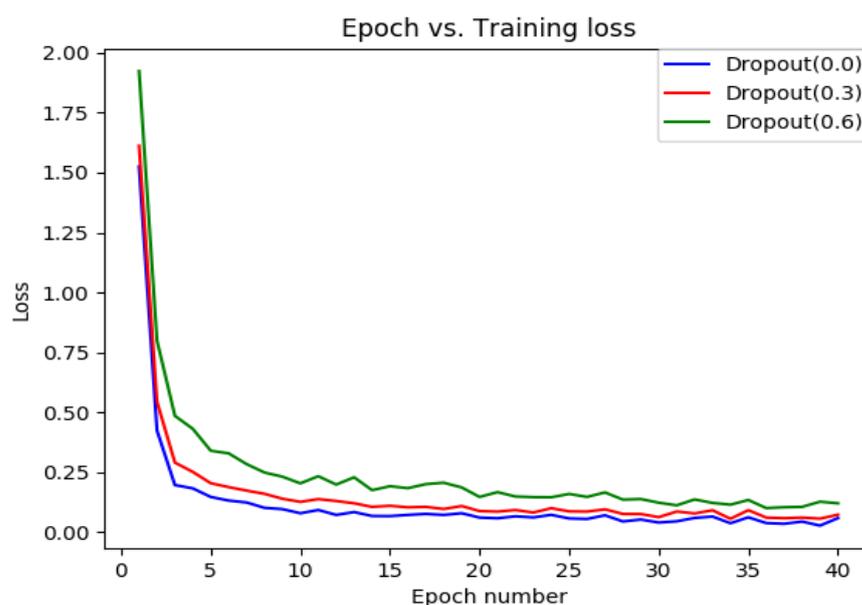


Рисунок 5 – Использование разных Dropout слоев

ЗАКЛЮЧЕНИЕ

Данная работа рассматривает использование искусственных нейронных сетей для распознавания текста на изображениях на примере автомобильных номеров.

В работе были рассмотрены искусственные нейронные сети, их виды, используемые для распознавания текста, были сравнены различные виды функций активации нейронов, методы изменения весов-параметров и их влияние на точность сети. Также было рассмотрено влияние количества промежуточных слоев на ошибку классификации, был приведен способ визуализации полученных карт признаков в сверточных слоях, были рассмотрены алгоритмы и методы выделения областей изображения с искомым объектом.

В качестве продолжения исследования можно рассмотреть влияние сверхглубоких сетей на качество классификации, реализовать распознавание разных видов номеров, распознавание рукописного текста с добавлением мультязычности, реализовать обнаружение искомых объектов на основе нейронных сетей.

Полученные результаты можно использовать для распознавания номеров на изображениях, полученных с дорожных скоростных камер, для автоматических систем парковки, для мобильных приложений-сканеров.