МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «МИСиС»**

Институт ИТКН

Кафедра инженерной кибернетики

Направление подготовки: 01.03.04 Прикладная математика

Квалификация (степень): бакалавр

Группа: БПМ-17-1

**ОТЧЕТ**

**ПО КУРСОВОЙ РАБОТЕ**

**«ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ»**

на тему: Свёрточная нейронная сеть для распознавания текста по изображению

**2020 – 2021 у.г.; VII семестр**

**Студентка:** Куликова Д. М.

**Преподаватель:** Кондыбаева А. Б.

**Москва 2020 г.**

**Содержание**

Список исполнителей 3

Введение 4

Теоретическая часть 5

Искусственные нейронные сети 5

Свёрточные нейронные сети 8

Содержательная постановка задачи 11

Реализация 12

Результаты 14

Список использованных источников 17

# **Список исполнителей**

Бердичевская А. Г.: сбор данных для обучения модели, проектирование архитектуры сети, определение путей улучшения модели.

Куликова Д. М.: препроцессинг данных, обучение модели, тестирование модели и анализ результатов.

# **Введение**

Существует огромное количество архивов с текстами, напечатанными на машинке, оцифровка подобных данных имеет важное значение для сохранение этих текстов на многие годы. Однако хранение текстовой информации просто в отсканированном виде не позволяет осуществлять быстрый поиск информации по документу, что существенно усложняет процесс поиска важных данных. Распознавание текста по изображению позволяет решить данную проблему и преобразовать старые документы к современному формату.

Данная задача могла бы быть решена нейронными сетями прямого хода, однако, так как картинки имеют огромное количество признаков, обучение таких моделей занимает слишком много времени. Наиболее подходящим средством для решения задачи распознавания текста по изображению являются свёрточные нейронные сети. Они заточены для работы с картинками, поэтому гораздо быстрее обучаются на подобных данных и дают хорошие результаты.

# **Теоретическая часть**

## **Искусственные нейронные сети**

Искусственные нейронные сети вдохновлены биологией, и они демонстрируют большое количество свойств, которые присущи мозгу. Например, они могут учиться на основе опыта, могут обобщать предыдущие прецеденты на новые случаи и выявлять существенные особенности из входных данных, которые содержат избыточную информацию.

Биологический нейрон состоит из тела и отростков. Отростки, по которым нейрон получает возбуждение, называются дендритами. Отросток, по которому нейрон передает возбуждение, называется аксоном, причем аксон у каждого нейрона один. Дендриты и аксон имеют довольно сложную ветвистую структуру. Место соединения аксона нейрона — источника возбуждения с дендритом называется синапсом. Основная функция нейрона состоит в передаче возбуждения из дендритов в аксон. Но сигналы, которые поступают из разных дендритов, могут влиять на сигнал в аксоне. Нейрон выдаст сигнал, если суммарное возбуждение превысит некоторое предельное значение, которое в общем случае меняется в некоторых границах. В противном случае на аксон сигнал выдан не будет: нейрон не ответит на возбуждение. Нейроны в искусственных нейронных сетях имитируют именно это свойство биологических нейронов.

Рассмотрим один искусственный нейрон. Множество входных сигналов, обозначенных x1, x2, ..., xn, поступают на искусственный нейрон. Эти входные сигналы соответствуют сигналам, которые приходят в синапсы биологического нейрона. Каждый сигнал множится на соответствующий вес w1, w2,..., wn, и поступает на суммирующий блок, обозначенный ∑. Каждый вес отвечает «силе» одной биологической синаптической связи. Суммирующий блок, который соответствует телу биологического элемента, алгебраически объединяет взвешенные входы, создавая выход NET:

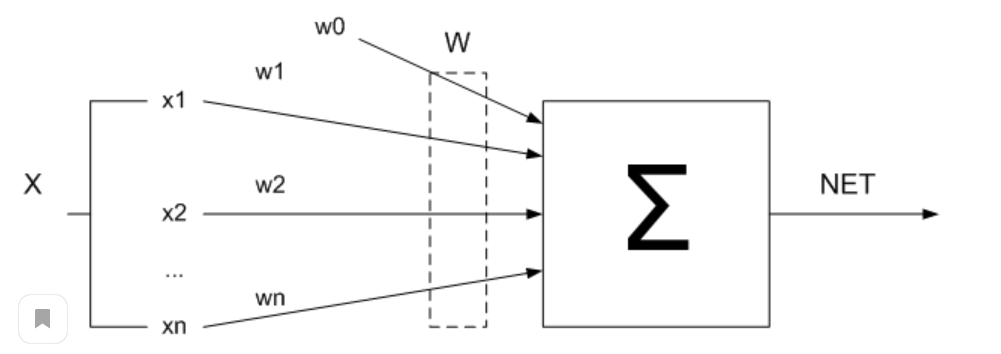
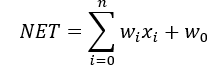


Рисунок 1 - искусственный нейрон

Данное описание можно представить следующей формулой



где

 — биас;

 — вес i- го нейрона;

 — выход i- го нейрона;

 — количество нейронов, которые входят в обрабатываемый нейрон

Сигнал w0, который имеет название биас, отображает функцию предельного значения, сдвига. Этот сигнал позволяет сдвинуть начало отсчета функции активации, которая в дальнейшем приводит к увеличению скорости обучения. Этот сигнал добавляется к каждому нейрону, он учится, как и все другие веса, а его особенность в том, что он подключается к сигналу +1, а не к выходу предыдущего нейрона.

Полученный сигнал NET как правило обрабатывается функцией активации и дает выходной нейронный сигнал:

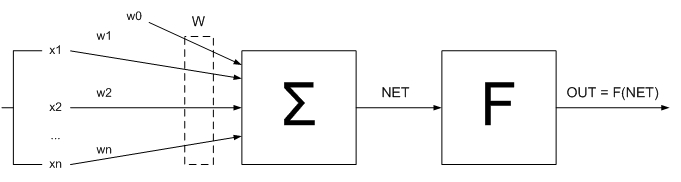
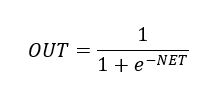


Рисунок 2 - искусственный нейрон с функцией активации

В качестве функции активации часто используются сигмоидальная функция. Эта функция математически выражается следующим образом:



Основное преимущество такой функции — то, что она имеет простую производную и дифференцируется по всей оси абсцисс. График функции имеет следующий вид:

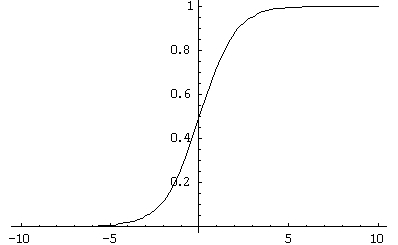


Рисунок 3 - график сигмоидной функции

Искусственная нейронная сеть представляет из себя совокупность искусственных нейронов, соединенных друг с другом определенным образом в зависимости от архитектуры сети. На вход каждого искусственного нейрона поступает некоторое множество сигналов, каждый из которых является выходом другого нейрона. Ниже представлена стандартная схема нейронной сети.

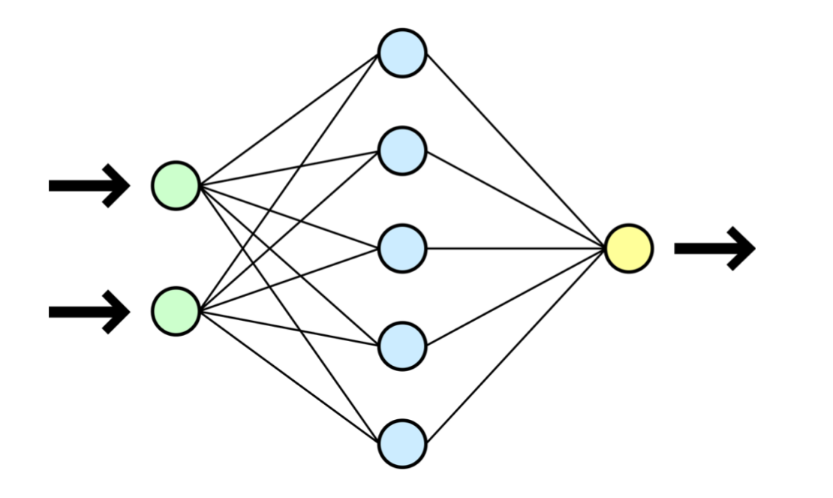


Рисунок 4 - схема нейронной сети

## **Свёрточные нейронные сети**

Свёрточная нейронная сеть — специальная архитектура искусственных нейронных сетей, нацеленная на эффективное распознавание образов. Данной архитектуре удаётся гораздо точнее распознавать объекты на изображениях, так как, в отличие от многослойного персептрона, учитывается двухмерная топология изображения. При этом свёрточные сети устойчивы к небольшим смещениям, изменениям масштаба и поворотам объектов на входных изображениях.

В отличие от сетей прямого распространения, которые работают с данными в виде векторов, свёрточные сети работают с изображениями в виде тензоров. На рисунке представлена архитектура одной из первых свёрточных сети.

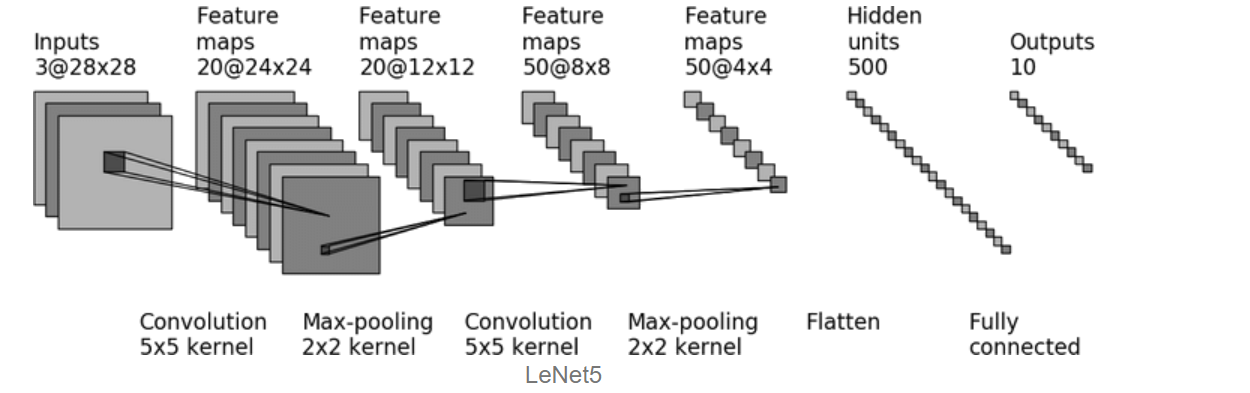


Рисунок 5 - схема свёрточной нейронной сети

Свёрточные нейронные сети состоят из базовых блоков,которые можно чередовать в произвольном порядке в зависимости от особенностей решаемой задачи. Основными блоками свёрточных нейронных сетей являются свёрточные слои, слои подвыборки (пулинга), слои активации и полносвязные слои.

**Слой свёртки**

Слой свёртки является самым главным слоем сети. Его основное назначение – выделить признаки на входном изображении и сформировать карту признаков. Для того, чтобы слой мог выделять признаки, в нём имеются так называемые фильтры (или ядра).

Для того, чтобы сформировать карту признаков из входного изображения, производится операция свёртки входного тензора с каждым из фильтров. Свёртка – это операция вычисления нового значения выбранного пикселя, учитывающая значения окружающих его пикселей. Алгоритм получения результата свёртки описан ниже.

Фильтр накладывается на левую верхнюю часть изображения и производится покомпонентное умножение значений фильтра и значений изображения, после чего фильтр перемещается дальше по изображению до тех пор, пока аналогичным образом не будут обработаны все его участки. Затем числа полученных матриц суммируются в единую матрицу — результат применения фильтра. После этого к каждому значению матрицы добавляется одинаковое число – значение смещения данного фильтра. Полученная матрица составляет один канал выходной карты признаков. После того, как будут получены каналы для каждого из фильтров, матрицы объединяются в единый тензор, благодаря чему на выходе снова получается изображение, с другим числом каналов и, возможно, другим размером.

**Слой пуллинга**

Данный слой позволяет уменьшить пространство признаков, сохраняя наиболее важную информацию. Существует несколько разных версий слоя пулинга, среди которых максимальный пулинг, средний пулинг и пулинг суммы. Наиболее часто используется слой макспулинга.

Слою подвыборки требуется всего один гиперпараметр — шаг пулинга, то есть число раз, в которое нужно сократить пространственные размерности. Наиболее часто используется слой макспулинга с уменьшением размера входного тензора в два раза. Некоторые библиотеки позволяют задавать раздельные параметры уменьшения по высоте и ширине, однако чаще всего эти параметры совпадают.

**Слой активации**

Данный слой представляет из себя некоторую функцию, которая применяется к каждому числу входного изображения. Наиболее часто используются такие функции активации, как ReLU, Sigmoid, Tanh, LeakyReLU. Обычно активационный слой ставится сразу после слоя свёртки, из-за чего некоторые библиотеки даже встраивают ReLU функцию прямо в свёрточный слой.

**Полносвязный слой**

Данный слой содержит матрицу весовых коэффициентов и вектор смещений и ничем не отличается от такого же слоя в обыкновенной полносвязной сети. Единственным гиперпараметром слоя является количество выходных значений. При этом результатом применения слоя является вектор или тензор, у которого матрицы в каждом канале имеют размер 1х1.

# **Содержательная постановка задачи**

Цель: Разработать программное обеспечение для решения задачи распознавания текста с картинки с применением свёрточной нейронной сети.

Для достижения поставленной цели требуется выполнить следующие задачи:

1. Выбрать подходящий набор данных для обучения модели распознавания текста;
2. Выполнить предварительную подготовку набора данных (разбить набор на обучающую и тестовую выборки);
3. Выполнить предварительную обработку входных данных (выделить на картинке отдельные символы);
4. Выбрать архитектуру свёрточной нейронной сети;
5. Обучить модель на обучающей выборке данных;
6. Провести тестирование модели.

# **Реализация**

Решение было реализовано на языке Python с использованием библиотек openCV, imghdr, numpy, tensorflow, keras, idx2numpy.

Для распознавания символов была использована сверточная нейронная сеть, собранная на основе линейной модели Sequential из библиотеки keras.

Использованная в работе сеть состоит из двух свёрточных слоёв, одного слоя максимального пуллинга и двух полносвязных слоёв.

Первый свёрточный слой содержит 32 фильтра размером 3\*3 и использует функцию активации ReLU. Второй свёрточный слой подержит 64 фильтра размера 3\*3 и также использует функцию активации ReLU. Затем следует слой макспуллинга, уменьшающий пространство признаков в 2 раза. После слоя макспуллинга применяется dropout для предотвращения переобучения модели. Затем применяется слой Flattern для выравнивания входа, после чего следует полносвязный слой, содержащий 512 нейронов и использующий функцию активации ReLU. После него снова применяется dropout и следует последний слой нейронной сети - ещё один полносвязный слой, состоящий столько нейронов, сколько различных символов представлено в обучающей выборке, то есть 62 (арабские цифры, буквы латинского алфавита и т.д.). Последний слой использует функцию активации softmax.

Нейронная сеть была обучена на датасете EMNIST ByClass. Датасет состоит из 814255 картинок рукописных символов размера 28х28 пикселей, разбитых на 62 класса (A, …, Z, 0, …, 9 и пр.). На рисунке 1 представлен пример рукописных символов из рассматриваемого датасета. В связи с большим объёмом датасета и долгим временем обучения модели была использована только 1/3 всех данных.



Рисунок 6 - Пример рукописных символов из датасета EMNIST

Модель обучалась на протяжении 30 эпох. На рисунке 7 представлен результат обучения на нескольких последних эпохах.

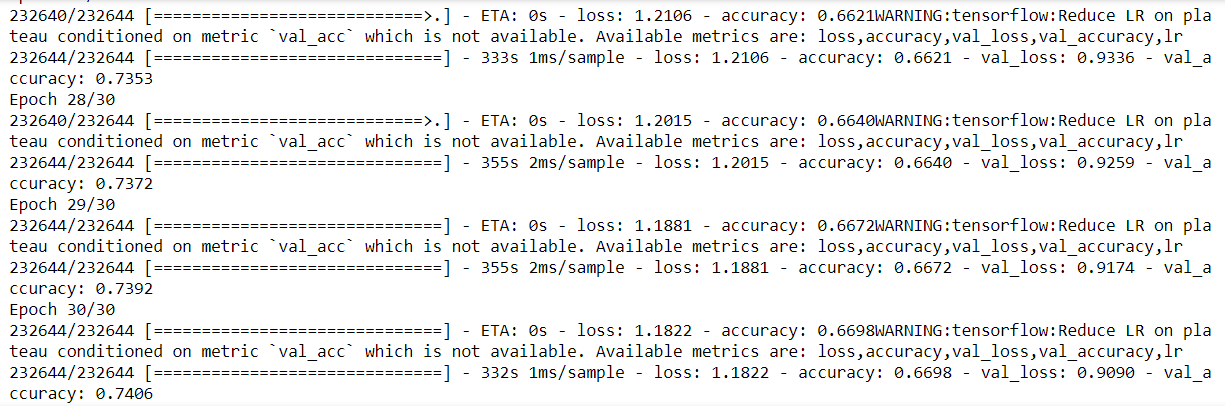


Рисунок 7 - Обучение модели

Перед передачей данных модели необходимо провести препроцессинг в результате которого текст будет разбит на отдельные символы. Для этого была использована библиотека openCV. Сначала входное изображение преобразуется к чёрно-белому, далее символы немного расширяются, находятся контуры символов и строятся ограничивающие прямоугольники для каждого символа по его контуру. На последнем этапе полученные прямоугольники преобразуются до квадрата 28х28 пикселей. Пример препроцессинга представлен на рисунке 8.

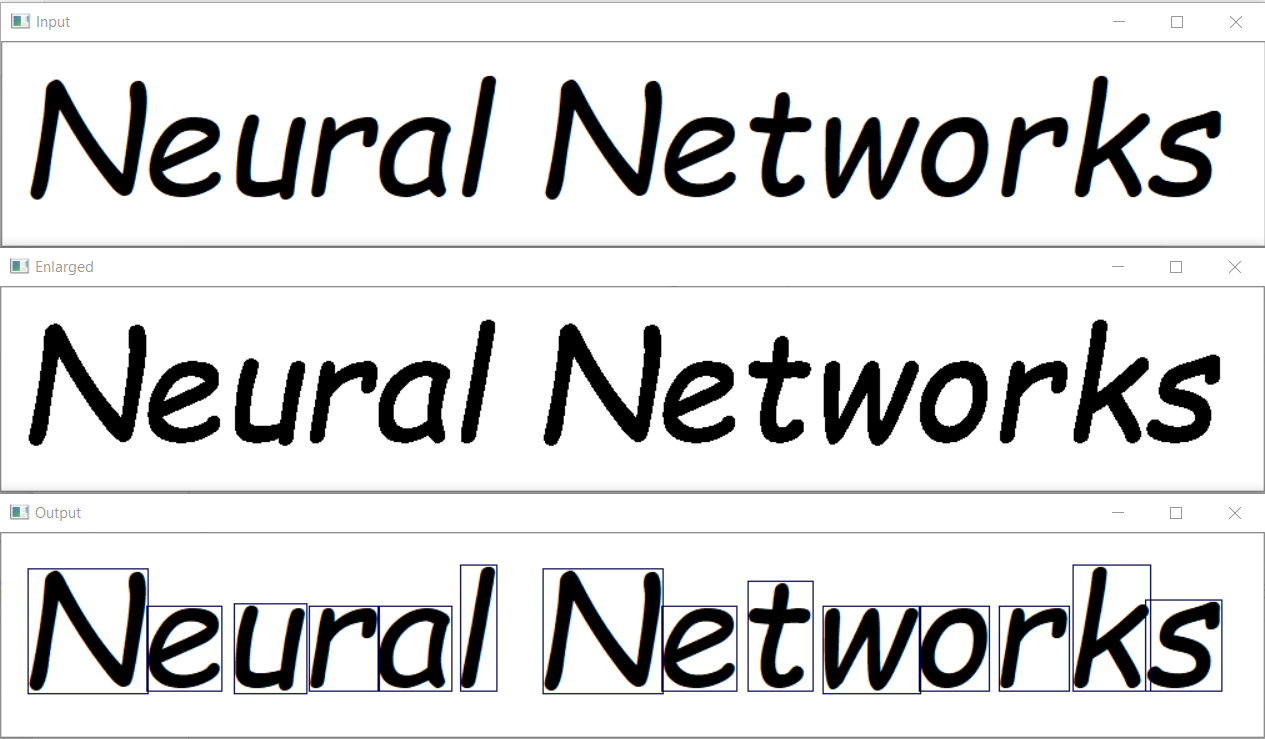


Рисунок 8 - Препроцессинг данных

# **Результаты**

Результатом работы модели для словосочетания “Neural Networks”, написанного шрифтом Comic Sans, с рисунка 8 является строка “NeUrU1 NetW0rk5”. На изображении видно, что буквы “l”, “o” и “s” были ошибочно распознаны как цифры “1”, “0” и “5”, соответственно, буква “a” как “U”. Некоторые прописные буквы были распознаны как заглавные, это вызвано тем, что написание прописных и заглавных букв может быть похоже, при этом процесс распознавание для каждой буквы происходит отдельно, вне слова.

Название датасета, написанное шрифтом Calibri и представленное на рисунке 9, было распознано как “EMN15T BYC1A55” “EMN1ST Byc1a55”. В данном тесте также некоторые буквы были ошибочно распознаны как похожие по написанию цифры, однако, большая часть букв была распознана верно (не считая ошибочного регистра).



Рисунок 9 – Тестовая картинка 2

Тестовое слово “coursework” (рис. 10, шрифт Times New Roman) было распознано моделью как “CUUF5eWUFK”. В данном тесте модель допустила довольно много ошибок, распознав лишь половину букв верно.

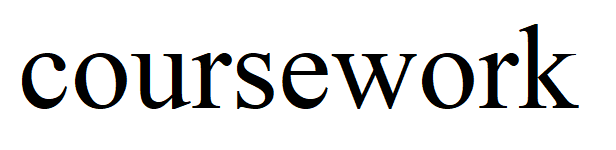


Рисунок 10 – Тестовая картинка 3

Полученные результаты могут быть улучшены путём дообучения модели на большем наборе данных. Также можно заметить тенденцию, что чем меньше шрифт похож на рукописный, тем хуже результаты распознавания. Это связано с тем, что датасет, на котором обучалась модель, состоит из рукописных символов. Таким образом, для улучшения результатов можно создать свой датасет, содержащий напечатанные символы с помощью различных шрифтов и провести аугментацию данных.

**Заключение**

В процессе работы были изучены свёрточные нейронные сети, а также был выбран датасет, подходящий для поставленной задачи. Был разработан алгоритм для препроцессинга данных, позволяющий разбивать текст, представленный на картинке, на отдельные символы. На ⅓ данных из выбранного датасета была обучена модель, представляющая собой свёрточную нейронную сеть, в течение 30 эпох. Весь программный код был написан на языке Python 3.8 с использованием библиотек openCV и TensorFlow.

Было проведено тестирование полученной модели и анализ результатов. Также были определены пути улучшения уровня распознавания, такие как увеличение объёма данных для обучения модули или создание нового датасета, содержащего печатные символы.

# **Список использованных источников**

1. Ф. Уоссермен. Нейрокомпьютерная техника: теория и практика. Перевод на русский язык Ю. А. Зуев, В. А. Точенов. Мир, 1992;

2. И. В. Заенцев. Нейронные сети: основные модели. Учебное пособие к курсу “Нейронные сети”. Воронеж, 1999.