Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**Пермский национальный исследовательский**

**политехнический университет**

Факультет прикладной математики и механики

Кафедра «Вычислительная математика, механика и биомеханика»

Направление 09.04.02 «Информационные системы и технологии»

Профиль программы магистратуры «Информационные технологии и системная инженерия»

**КУРСОВАЯ РАБОТА**

по дисциплине **«Нейронные сети»**

на тему

**«Классификация образов на изображениях»**

Выполнил студент группы ИТСИ-20-1м

Лоргов Евгений Сергеевич

(Фамилия, Имя, Отчество полностью)

Проверил:

канд. физ.-мат. наук, доц.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (Клюев А.В.)

(подпись) (ФИО)

Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2021 г.

**Пермь 2021**

Содержание

1. Введение 3

2. Общий подход к задаче 5

3. Сверточная нейронная сеть 6

4. Модель нейронной сети для текущей задачи 9

5. Обучение нейронной сети 10

6. Результаты 11

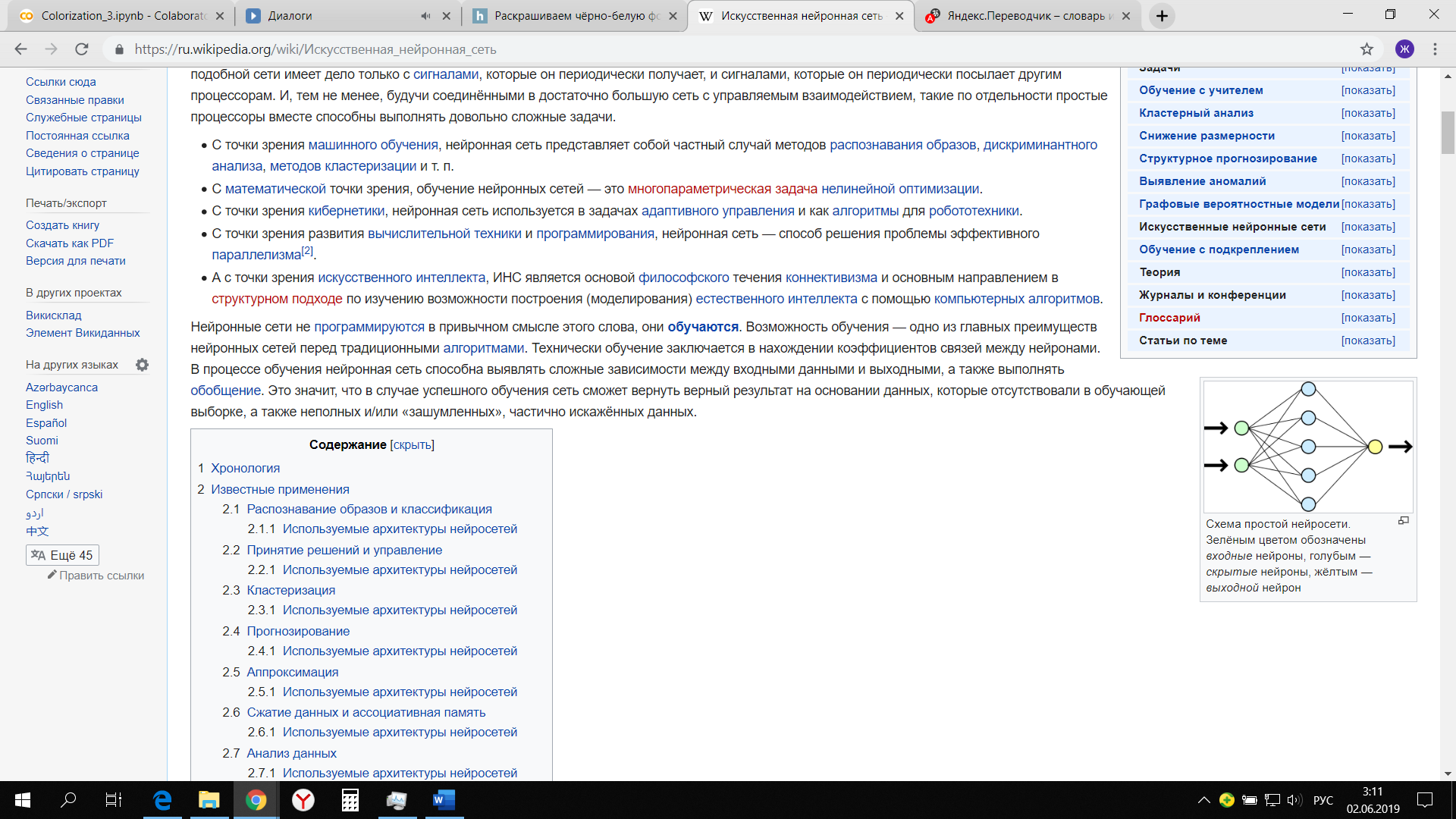
7. Заключение 12

Введение

Задача распознавания объектов на изображении является довольно актуальной в наше время. Задача применяется в таких сферах как маркетинг (отслеживание упоминания брендов в социальных сетях), видеонаблюдение (распознавание образов на камерах городского видеонаблюдения) и т. д.

Данная задача была решена при помощи нейронной сети. Нейронная сеть представляет собой систему соединенных и взаимодействующих простых узлов – нейронов. Каждый такой узел имеет дело только с сигналами либо от других узлов, либо от него самого к следующим нейронам.

Нейронная сеть состоит из входного, выходного и скрытого (если такой имеется) слоев. На входном слое подаются начальные данные, затем они обрабатываются на скрытом слое, и, наконец, попадая на выходной слой, преобразуются в необходимую информацию. С математической точки зрения каждый такой слой представляет из себя *n* – мерный вектор, состоящий из действительных чисел. Например, нейронная сеть, созданная для решения задачи классификации объектов на изображении, принимает на вход изображение (если оно черно – белое, то его размерность равна 2, если оно цветное, то 3), а на выход – вектор из *N* чисел – объектов изображений. В таком векторе содержится информация о вероятности принадлежности к конкретному объекту.



**Рис. 1** *Простая модель нейронной сети, которая содержит входной (зеленый), скрытый (голубой) и выходной (желтый) слои.*

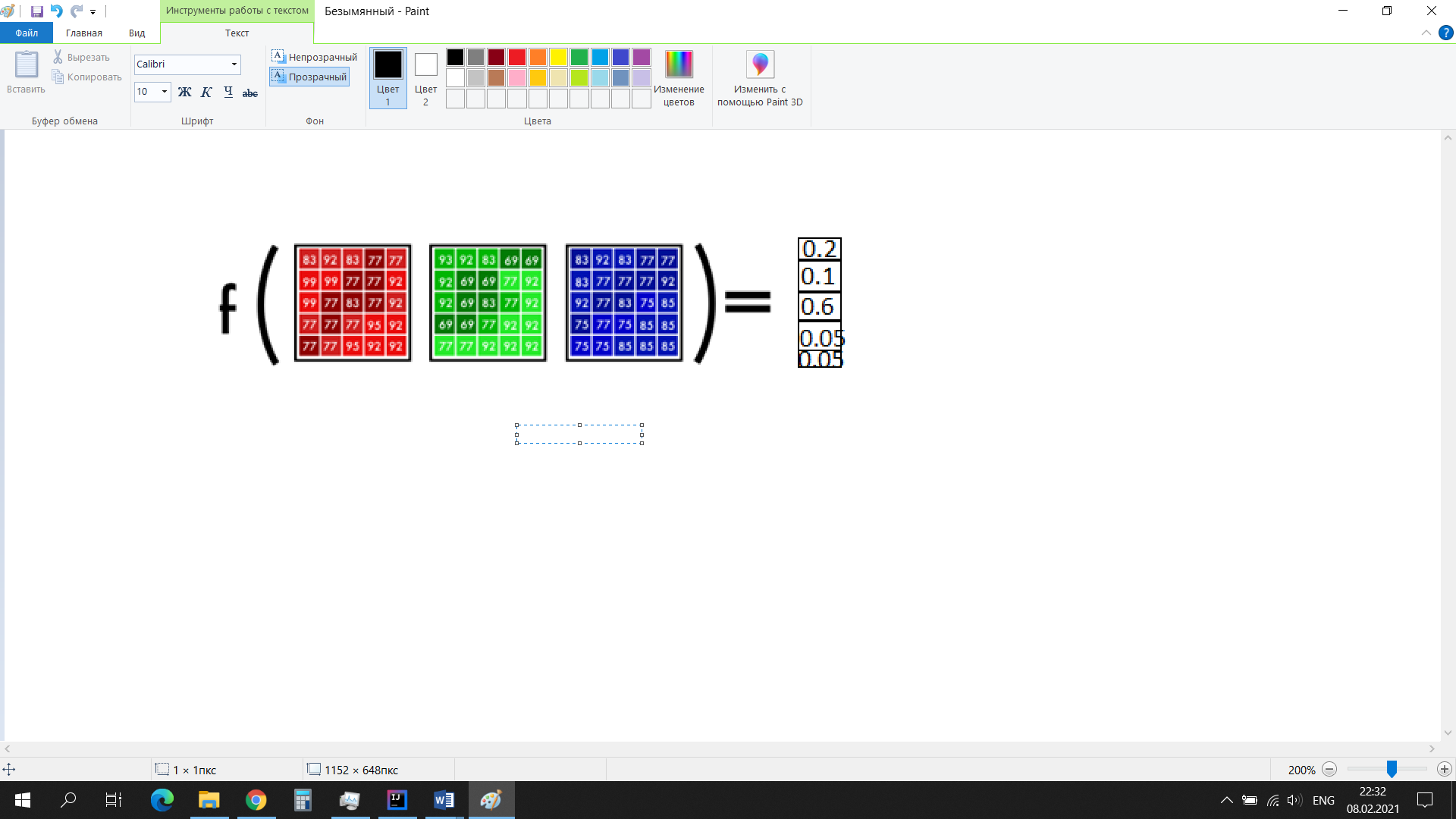
Нейронная сеть не может обходиться без специальных коэффициентов – весов, служащих для соединения нейронов. Веса показывают важность сигнала от одного нейрона к другому. Чаще всего веса представляют из себя числа от 0 до 1. От значения весов напрямую зависит полученный результат, поэтому их необходимо корректировать путем обучения нейронной сети. Суть обучения – минимизация ошибки, выдаваемой сетью. Существуют разные способы обучения, но конкретно для задачи распознавания объектов на изображении представлено обучение с учителем. Данный тип обучения работает следующим образом:

1. Нейронная сеть выдает результаты;
2. Учитель дает ей результаты, являющиеся правильными;
3. Сеть сравнивает их и при помощи специальной функции потери считает ошибку между ними;
4. При помощи специальных методов оптимизации ошибка поддается процессу минимизации, и веса корректируются в правильное направление.

Данный процесс происходит до тех пор, пока веса не будут полностью (или почти полностью) скорректированы.

1. Общий подход к задаче

Главная цель в поставленной задаче – взять изображение и получить из него число, говорящее о распознанном объекте. Можно поставить задачу как изучение такого отображения *f*, для которого , где – вектор с вероятностями распознавания образов, N – количество образов, которые может распознать нейросеть, а - изображение в оттенках RGB. Разрешение входного изображения – фиксированное и одинаковое. Значения входного изображения варьируются от 0 до 255. Значения выходов от 0 до 1.



**Рис. 1.1** *Схематичное представление задачи, представленное выше в пространстве RGB.*

В качестве объекта, который распознала нейронная сеть на изображении, выбирается тот, у которого самое большое значение в выходном векторе вероятностей.

1. Сверточная нейронная сеть

Сверточные нейронные сети — это широкий класс архитектур, основная идея которых состоит в том, чтобы использовать одни и те же части нейронной сети для работы с разными маленькими, локальными участками входов. Такой тип сети выдает наиболее лучшие результаты в области работы с изображениями.  Сверточные нейронные сети объединяют три архитектурных идеи, для обеспечения инвариантности к изменению масштаба, повороту сдвигу и пространственным искажениям:

1. Локальные рецепторные поля, обеспечивающие локальную двумерную связность нейронов;
2. Общие синаптические коэффициенты, дающие возможность детектировать некоторые черты в любом месте изображения и уменьшает общее число весовых коэффициентов;
3. Иерархическая организация с пространственными подвыборками.

В моделях сверточной нейронной сети можно столкнуться со следующими операциями:

1. Сверточный слой – основной блок сверточной нейронной сети. Слой свёртки включает в себя для каждого канала свой фильтр (один или несколько), ядро свёртки которого обрабатывает предыдущий слой по фрагментам. Сверточный слой может иметь свой шаг, определяющий, в окрестностях каких пикселей будет производиться операция свертки. Таким образом, шаг свертки позволяет получить либо такое же по размерам изображение, либо уменьшенное.
2. Слой активации. Скалярный результат каждой свёртки попадает на функцию активации, которая представляет собой некую нелинейную функцию. Обычно для сверточных нейронных сетей применяют функцию гиперболического тангенса , а также сигмоида . Но наиболее широкое применение получила функция ReLU, которая выглядит следующим образом:

, (2.1)

1. Частичное отключение весовых коэффициентов. В нейронный сетях существует специальная операция, позволяющая отключать часть весов с заданной вероятностью. Она необходима, чтобы избежать проблемы переобучения сети – явления, когда построенная модель хорошо объясняет примеры из обучающей выборки, но относительно плохо работает на примерах, не участвовавших в обучении (на примерах из тестовой выборки).
2. Слой субдискретизации. Слой пулинга (иначе подвыборки, субдискретизации) – нелинейное уплотнение карты признаков, при этом группа пикселей (обычно размера 2×2) уплотняется до одного пикселя, проходя нелинейное преобразование. Наиболее употребительна при этом функция максимума. Преобразования затрагивают непересекающиеся прямоугольники или квадраты, каждый из которых ужимается в один пиксель, при этом выбирается пиксель, имеющий максимальное значение. Операция пулинга позволяет существенно уменьшить пространственный объём изображения. Пулинг интерпретируется так: если на предыдущей операции свёртки уже были выявлены некоторые признаки, то для дальнейшей обработки настолько подробное изображение уже не нужно, и оно уплотняется до менее подробного.
3. Полносвязный слой. После нескольких прохождений свёртки изображения и уплотнения с помощью пулинга система перестраивается от конкретной сетки пикселей к более абстрактным картам признаков. В конце концов остаётся большой набор каналов, хранящих небольшое число данных (даже один параметр), которые интерпретируются как самые абстрактные понятия, выявленные из исходного изображения. Эти данные объединяются и передаются на обычную полносвязную нейронную сеть, которая тоже может состоять из нескольких слоёв. При этом полносвязные слои уже утрачивают пространственную структуру пикселей и обладают сравнительно небольшой размерностью (по отношению к количеству пикселей исходного изображения).
4. Выравнивание (Flatten). Используется при переходе от сверточного к полносвязному слою нейронной сети. Массив, полученный в результате свертки, имеющий размерность 3, разворачивается в обычный вектор.
5. Модель нейронной сети для текущей задачи

В данной работе нейронная сеть обучается на основе датасета MNIST (распознавание цифр), в котором изображения имеют размер 28\*28\*1 (черно-белое изображение), а количество образов равно 10. Для решения данной задачи была составлена следующая модель нейросети:

1. Входящий слой. Размер 28\*28\*1;
2. Свертка. Размер ядра 3. Количество карт признаков 15. Stride = 1. Activation = ReLu. Dropout = 0.0;
3. MaxPooling. Stride = 2;
4. Свертка. Размер ядра 3. Количество карт признаков 30. Stride = 1. Activation = ReLu. Dropout = 0.0;
5. Полносвязный слой. Кол-во нейронов 10. Activation = SoftMax. Dropout = 0.1;
6. Обучение нейронной сети

Для обучения нейронной сети в данной задаче была использована квадратичная функция ошибки:

, (4.1)

где L – размер подвыборки, M – количество образов, – полученный ответ нейросети, – ожидаемый ответ нейросети.

Далее функция потерь (4.1) минимизируется оптимизационным методом, корректируя веса нейросети в нужное направление. В данном случае был применен метод обратного распространения ошибки. В данном методе корректировка весов идет с конца модели сети. Для этого сначала считается ошибка на слое:

, (4.2)

где – значение производной функции активации по ее аргументу для нейрона i, – ошибка на следующем слое.

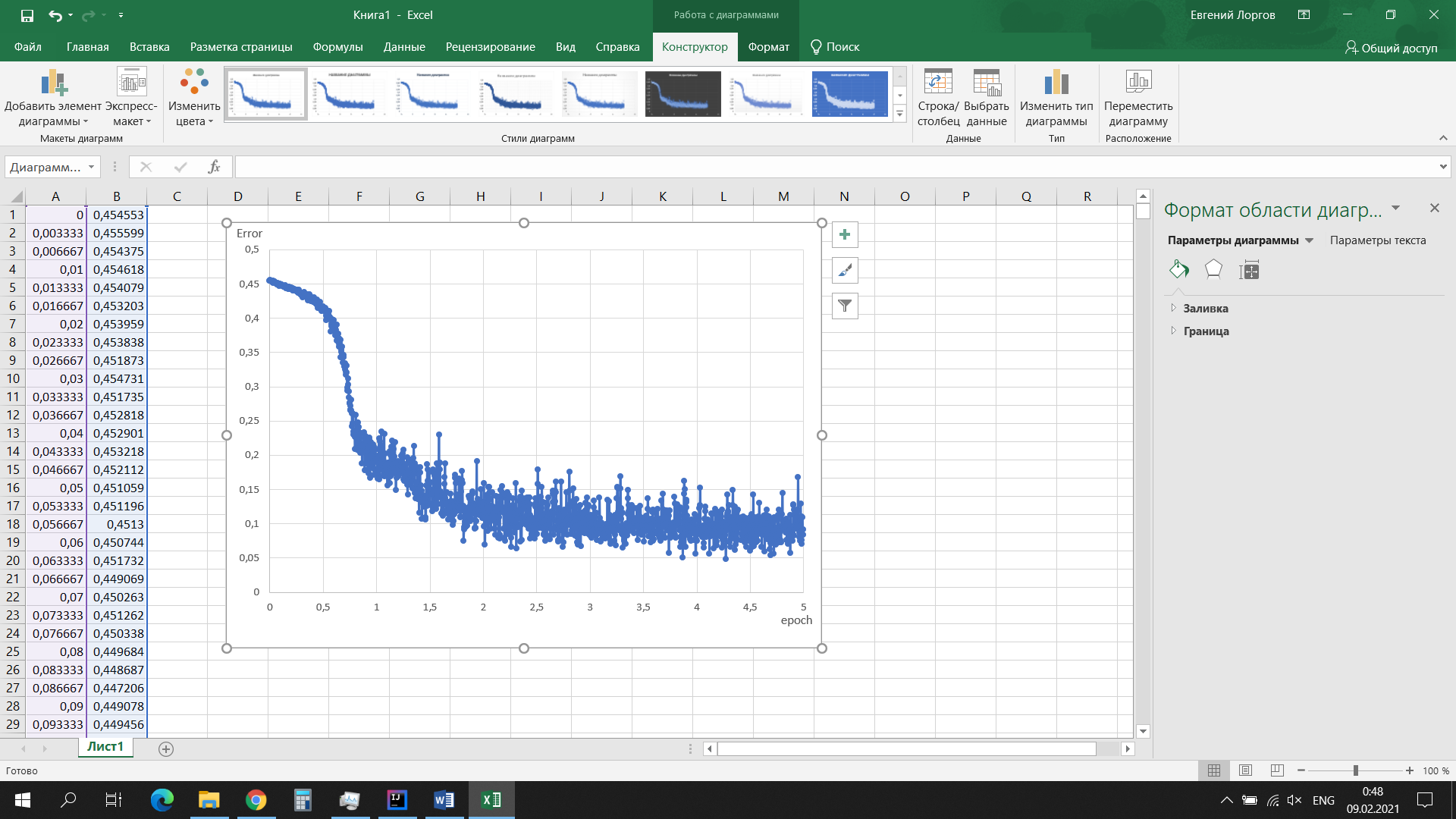
После подсчета ошибки на слое считается новое значение весов на нем:

, (4.3)

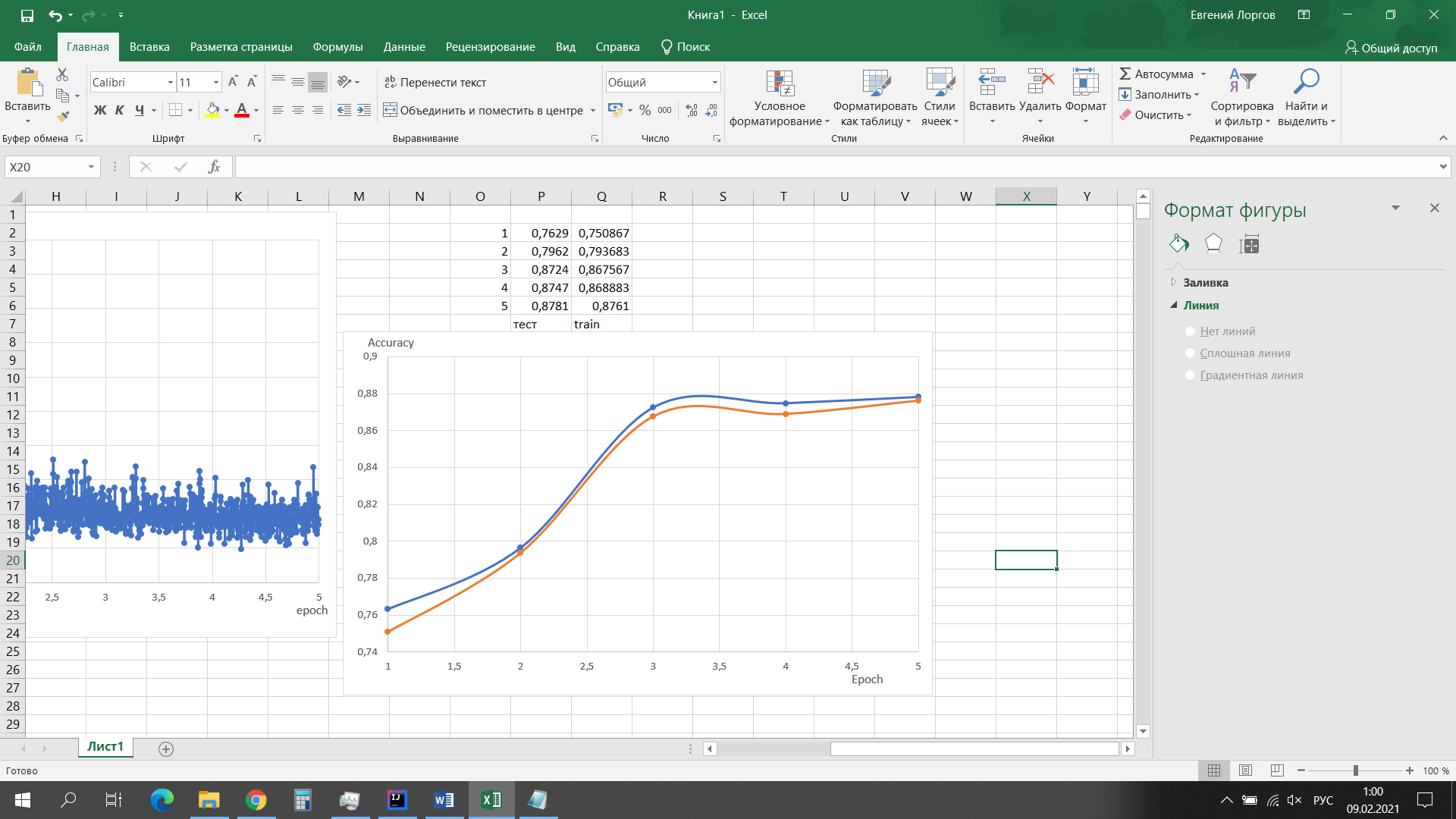
где alpha – коэффициент скорости обучения. На первом слое веса корректируются на основе входного изображения (вместо будет соответствующее значение пикселя).

1. Результаты

Нейронная сеть была реализована при помощи языка программирования Java. В датасете находится около 60000 картинок для обучения и 10000 картинок для тестирования. Обучение производилось по мини выборкам размером по 200. Количество эпох равно 5. Коэффициент скорости обучения равен 0,0001. Перед обучением весам были приписаны случайные значения от -0,002 до 0,05.



**Рис. 5.1** *Зависимость значения функции потери от эпохи обучающего множества.*



**Рис. 5.2** *Зависимость точности от эпохи. Синий график – тестовое множество, оранжевый – обучающее.*

Заключение

В данной работе была решена задача классификации объектов на изображении при помощи сверточной модели нейронной сети. В результате в процессе обучения была получена относительно высокая точность (почти 90%) распознавания объекта и маленькое значение функции ошибки.