

Московский государственный университет имени М. В. Ломоносова

Факультет вычислительной математики и кибернетики

Кафедра математических методов прогнозирования



## КУРСОВАЯ РАБОТА

**«Верификация автора рукописных документов на основе  
сиамской сети»**

**«Handwritten documents author verification based on the  
Siamese network»**

Выполнила:

студентка 5 курса 117 группы

*Пронина Наталия Михайловна*

Научный руководитель:

д.т.н., профессор

*Местецкий Леонид Моисеевич*

Москва, 2024

# Содержание

<b>1 Введение</b>	<b>3</b>
<b>2 Постановка задачи</b>	<b>3</b>
<b>3 Основные понятия</b>	<b>5</b>
3.1 Сиамская нейронная сеть . . . . .	5
3.2 Triplet loss function . . . . .	6
<b>4 Метод решения</b>	<b>6</b>
<b>5 Эксперименты</b>	<b>7</b>
5.1 Baseline модель . . . . .	7
5.2 Бинаризация изображений . . . . .	9
5.3 Сиамская сеть . . . . .	11
5.4 Результат . . . . .	12
5.5 Датасет IAM . . . . .	13
<b>6 Заключение</b>	<b>15</b>

## **Аннотация**

В данной работе представлен метод верификации определённого автора в корпусе документов на основании малого количества образцов, также предложен алгоритм предобработки данных.

Предлагается использование сиамской нейронной сети для сравнения и анализа уникальных характеристик почерка, стиля письма. Такой метод обучения позволяет получать эффективные эмбеддинги изображений, на основе которых можно произвести качественную классификацию автора.

Предложенный подход позволил выделить возможные работы Жуковского среди рукописных текстов неизвестных авторов. Подход также был проверен на полностью размеченном датасете IAM и показал высокую точность классификации.

# 1 Введение

Глубокие нейронные сети уже давно демонстрируют высокие результаты на задачах классификации изображений. Ярчайший пример - в 2015 году нейросеть ResNet [1] с 152 слоями показала меньшее количество ошибок на конкурсе ImageNet по сравнению с мануальной разметкой. Однако чем больше параметров у модели, тем выше риск переобучения и тем выше потребность в большем количестве данных. Эта проблема встаёт наиболее остро в задаче верификации автора.

Актуальность задачи верификации обусловлена постепенной оцифровкой многих архивов рукописных текстов 17-18 века. Возможность выделять небольшое подмножество подозрительных элементов поможет ускорить профессиональную работу филолога по атрибуции текста - проверке подлинности и установлению автора.

## 2 Постановка задачи

Дана небольшая коллекция почерков Жуковского, его *автографов* (рис. 1), из разных периодов жизни — раннего, зрелого и позднего, разной степени аккуратности — идеальный, обычный, неаккуратный. Всего - 22 изображения.

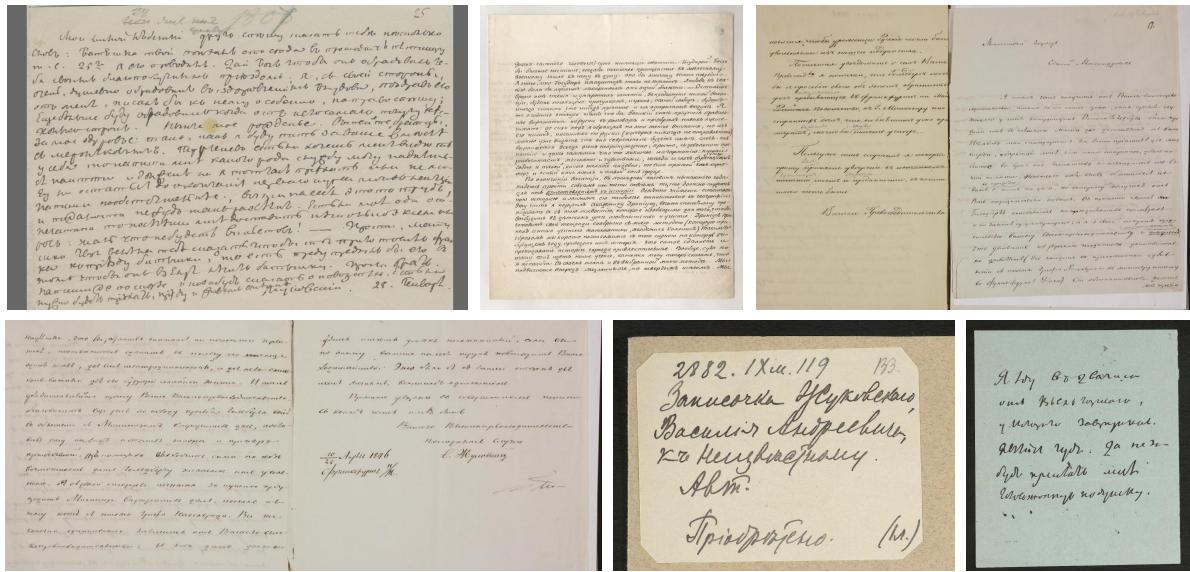


Рис. 1: Автографы Жуковского

Так же дан набор снимков рукописных документов неизвестных авторов, *конволют* (рис. 2), всего - 224 изображения. Требуется среди конволют выделить документы «подозрительные» на авторство Жуковского.

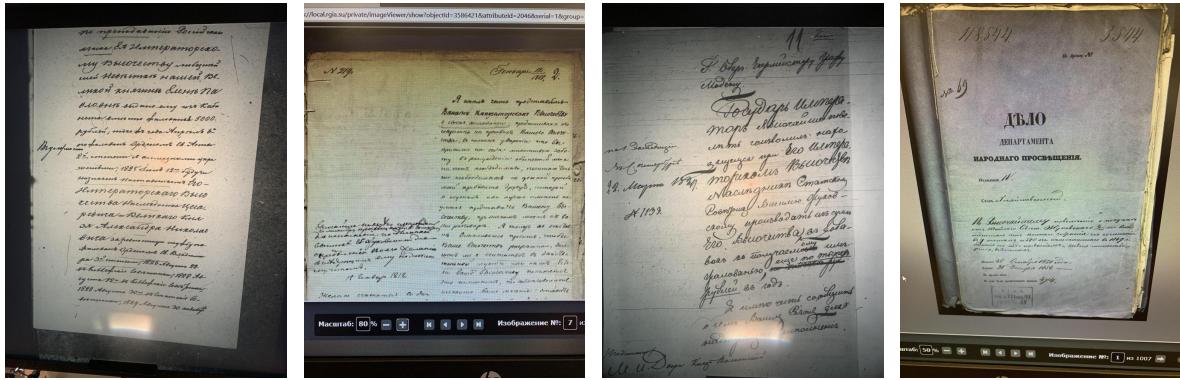


Рис. 2: Конволюты

Формальная постановка задачи: требуется разработать алгоритм, который будет принимать на вход изображение из конволют и выдавать вероятность того, что этот документ является автографом Жуковского. Далее по пороговому значению вероятности пользователь сможет отобрать топ документов для ручной экспертной проверки.

Особенности предоставляемых данных:

1. Коллекция писем Жуковского слишком мала, всего 22 изображения. Этого не хватит для эффективного обучения глубоких свёрточных сетей классическим методом.
2. Автографы и конволюты имеют разный фон
3. Почерки имеют разный масштаб
4. Конволюты оцифрованы намного хуже, чем автографы
5. Конволюты имеют искажение в виде муарового узора, так как сняты с экрана, возможны блики

### 3 Основные понятия

#### 3.1 Сиамская нейронная сеть

Решить главную проблему данных, их недостаток, предполагается с помощью алгоритма однократного обучения — *сиамской нейронной сети*.

Сиамская сеть является типом нейронной сети для глубокого обучения, которая использует две или больше идентичных подсети с одинаковой архитектурой. Также они используют одни параметры для обучения.

Сиамские сети особенно полезны в случае классификации с большим количеством классов и с небольшим числом объектов каждого класса. В таких случаях недостаточно примеров каждого класса, чтобы обучить глубокую свёрточную нейронную сеть. К тому же при добавлении новых классов пришлось бы менять архитектуру сети и переобучать. Вместо этого сиамская сеть учится на задачу бинарной классификации пар объектов: принадлежат ли объекты одному классу или нет. Это делает их особенно полезными в задачах, где требуется гибкость и эффективность при ограниченном наборе данных.

Данный вид сетей позволяет получить вектора признаков, эмбеддинги, двух объектов, отражающие их семантическое сходство или различие. Примеры приложений для сиамских сетей: распознавание лиц [2], верификация подписи [3].

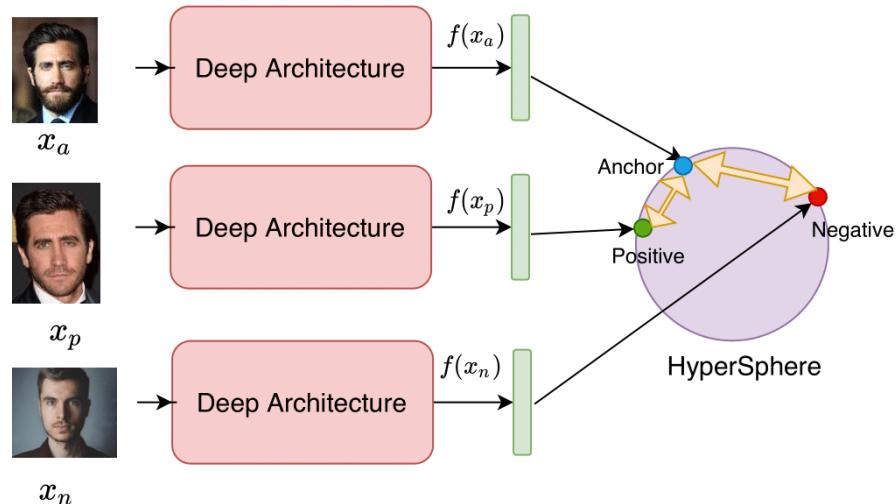


Рис. 3: Обучение сиамской сети с помощью triplet loss (источник: Shivam Chandhok)

Метки класса получаются после обучения на эмбеддингах простого классификатора, в данной работе будет использоваться двухслойная полносвязная нейронная сеть. Похожий метод классификации символов описан в [4].

### 3.2 Triplet loss function

Две наиболее популярные функции потерь для обучения сиамских нейронных сетей — contrastive loss function [5] и triplet loss function [6]. Они решают разные проблемы: если известна мера сходства двух объектов, тогда уместен contrastive loss; если известно положительное /отрицательное отношение (например, принадлежность к одному классу), тогда уместен triplet loss. К условиям нашей постановки задачи близок второй вариант.

Triplet loss function использует объект рассматриваемого класса (anchor), с которым будет проводиться сравнение, а также два других объекта: один принадлежащий к тому же классу (positive), и один не принадлежащий к этому классу (negative):

$$L(A, P, N) = \max\{d(A, P) - d(A, N) + margin, 0\}, \quad d(x, y) = \|x - y\|_p \quad (1)$$

Из формулы (1) видно, что функция стремится приблизить похожие объекты и увеличить расстояние между разными объектами. *margin* — заранее задаваемый параметр, показывающий, за какую разницу расстояний следует штрафовать.

## 4 Метод решения

Предполагается, что среди конволютов не более 2 % автографов Жуковского. Поэтому возьмём в качестве *отрицательного* класса все конволюты, а в качестве *положительного* — автографы Жуковского.

Общий метод получения вероятности принадлежности к положительному классу:

1. Обучим модель на задачу бинарной классификации (функция потерь — кросс энтропия).
2. Получим предсказания модели для отрицательного класса.

3. Применим функцию softmax, получим число из вероятностного симплекса, оценивающее степень принадлежности к положительному классу (далее для простоты будем называть его *вероятностью*)
4. Выделим топ объектов отрицательного класса с наибольшей вероятностью. Далее будем называть такие объекты *подозрительными* на авторство Жуковского.

## 5 Эксперименты

### 5.1 Baseline модель

В качестве базового метода обучим ResNet18 на автографах и конволютах на задачу бинарной классификации. Возьмём веса, предобученные на ImageNet, заменим последний линейный слой размерности 1000 на линейный слой размерности 2, так как мы решаем задачу бинарной классификации. Будем обучать только последние два слоя. Таким образом получим 2050 обучаемых параметров.

Чтобы увеличить количество положительных и отрицательных объектов и сбалансировать классы была проведена аугментация с помощью случайного вырезания фрагмента 300 x 300 пикселей. Количество положительных экземпляров увеличилось в 40 раз, количество отрицательных — в 4 раза. Итого: 880 положительных, 896 отрицательных.

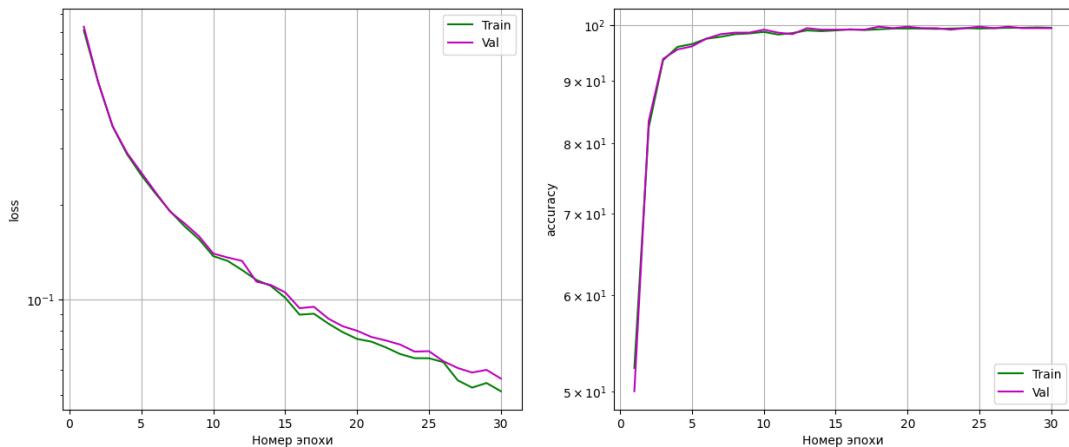


Рис. 4: Loss и accuracy

По графику 4 видно, что даже с малым количеством обучаемых параметров модель довольно быстро обучается и достигает 99.7% точности на валидационной выборке. Однако если посмотреть на топ объектов негативного класса с наибольшей вероятностью (рис. 5), то увидим картинки с печатным текстом, что точно не может быть автографом Жуковского.



Рис. 5: «Подозрительные» конволюты

Если посмотреть на положительные объекты с наименьшей вероятностью (рис. 6), то становится понятно, почему так происходит: модель сильно переобучается на фон. Картинки с тёмным фоном модель относит скорее к отрицательному классу, со светлым — к положительному. Поэтому несмотря на печатный текст модель отнесла 5 скорее к положительному классу.

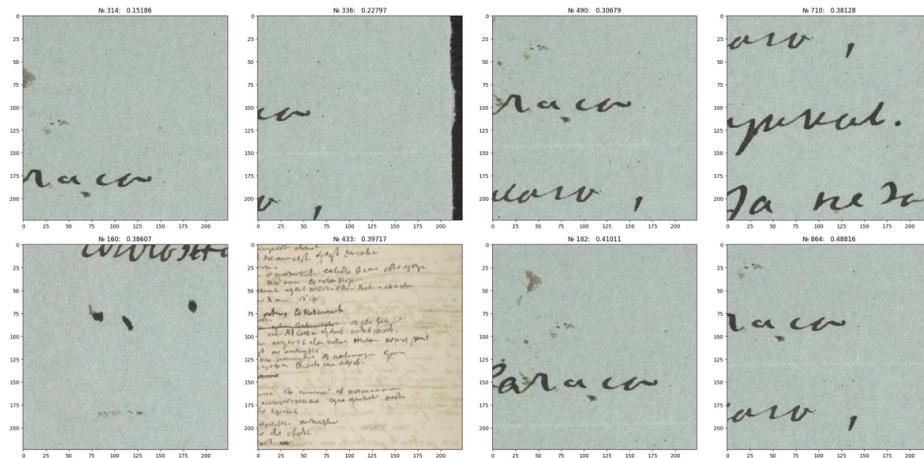


Рис. 6: Автографы Жуковского с наименьшей вероятностью

## 5.2 Бинаризация изображений

Чтобы предотвратить переобучение на фон, бинаризуем изображения с помощью трансформера DocEnTr [7]. На рисунке 7 приведён пример бинаризации автографов и конволютов. Для предотвращения попадания «пустых» изображений в выборки аугментация производилась таким образом, чтобы доля белых пикселей была не более 95%.

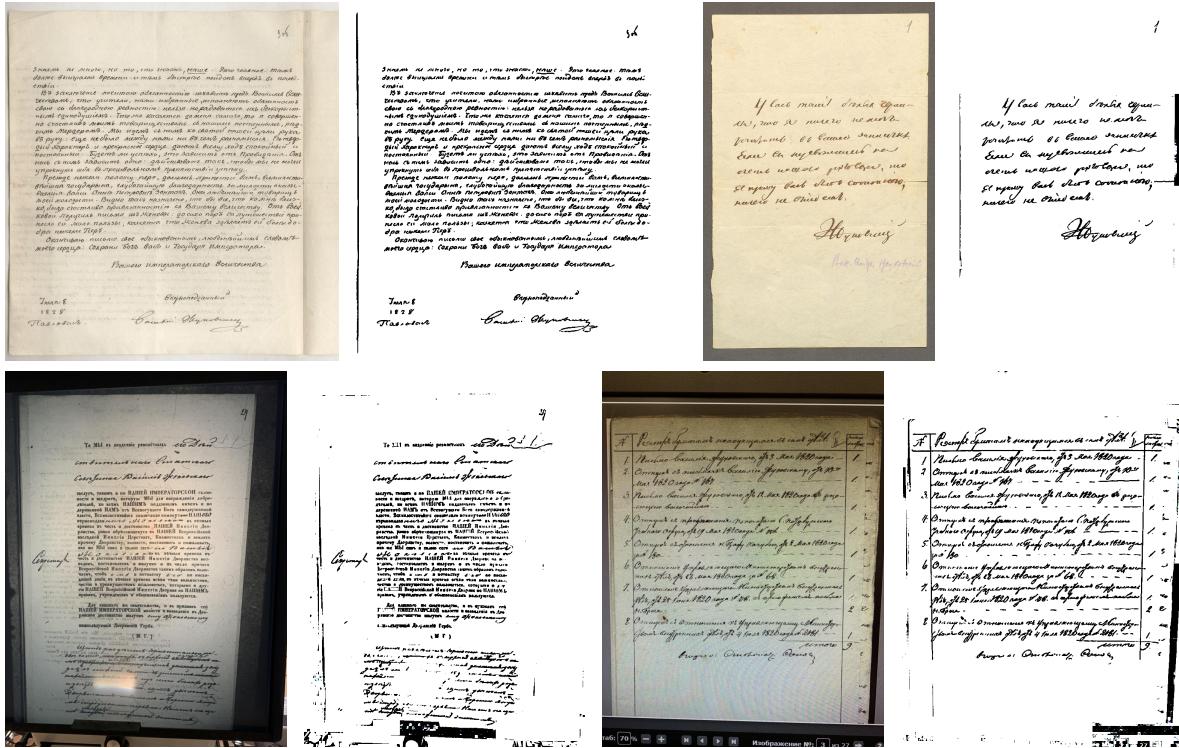


Рис. 7: Бинаризация с помощью DocEnTr

Модели стало заметно сложнее классифицировать объекты: качество снизилось до 87% на валидации, выросло переобучение (рис. 8). По матрице ошибок (рис. 9) видно, что модель меньше ошибается на автографах Жуковского, чем на конволютах, но распределения вероятности по положительным и отрицательным выборкам получились очень «гладкими»: среди конволютов нет ярко выраженных выбросов, которые мы с уверенностью могли бы верифицировать как автограф Жуковского.

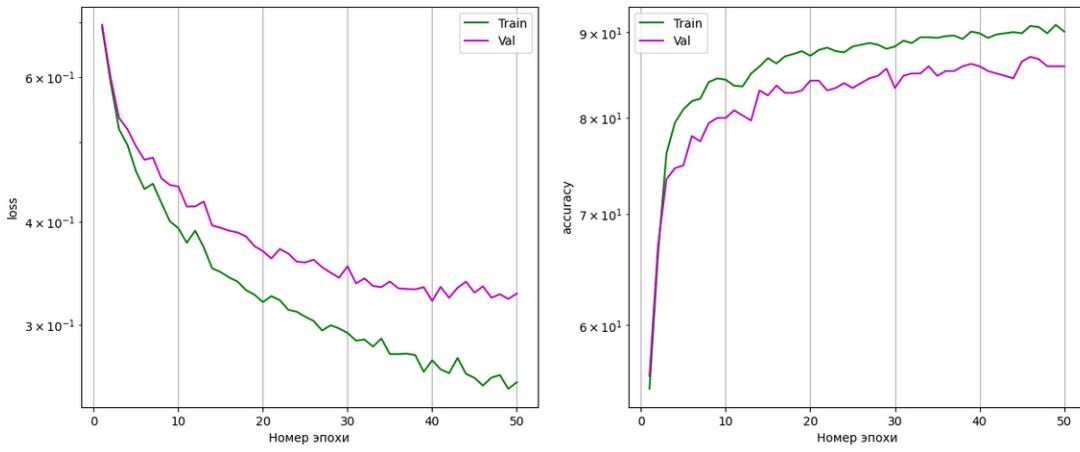


Рис. 8: Loss и accuracy

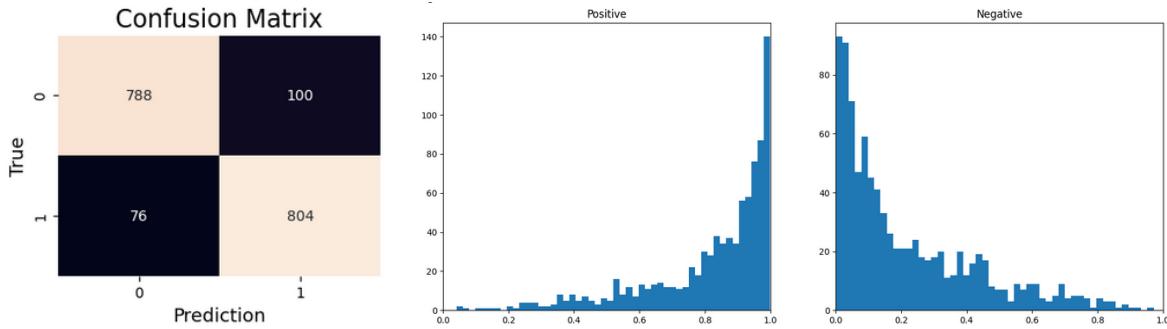


Рис. 9: Матрица ошибок и распределение вероятностей

Посмотрим на «подозрительные» конволюты (рис. 10), среди них снова попадается печатный текст, значит, модели не удаётся выделить эффективные признаки рукописного текста.

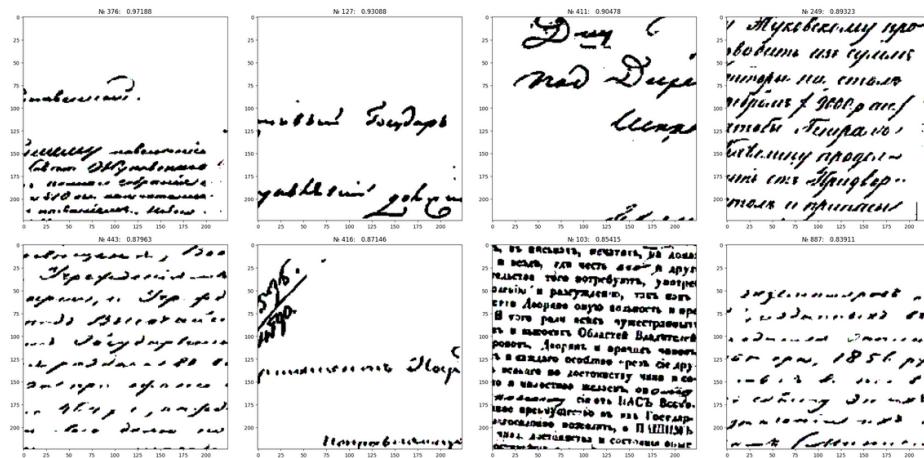


Рис. 10: «Подозрительные» конволюты

### 5.3 Сиамская сеть

Применим идею обучения сиамской сети на бинарных изображениях. Возьмём предобученный ResNet18. Последний слой имеет размерность 1000 на выходе, поэтому сиамская сеть будет выучивать 1000-размерный эмбеддинг изображения. Будем обучать последние два слоя, 513000 обучаемых параметров.

Построим датасет из 10 000 троек изображений, *tripлетов*: в качестве anchor и positive будем брать случайные автографы Жуковского, в качестве negative — случайную конволюту.

На графиках 11 приведены значения triplet loss и точность на обучении. Точность считается как доля триплетов, для которых евклидово расстояние между эмбеддингами anchor и positive меньше, чем между anchor и negative. Удалось достичь точности в 99.7% на валидации.

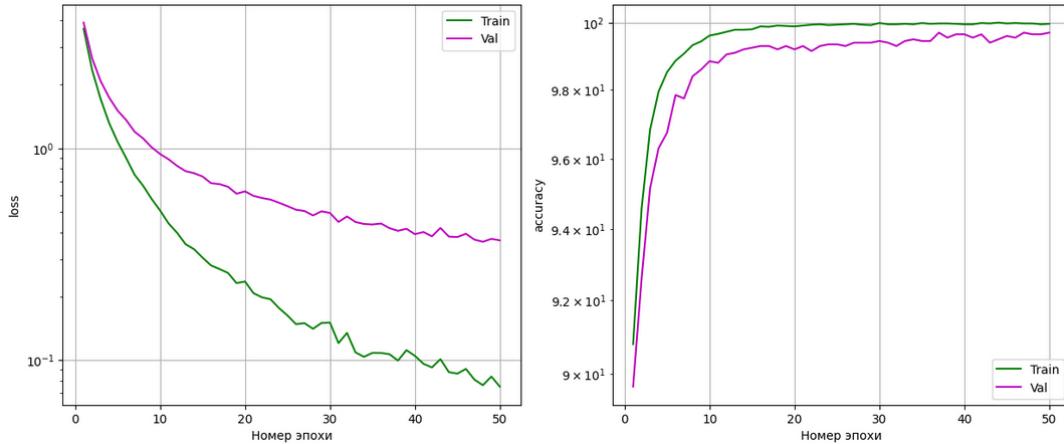


Рис. 11: Triplet loss и accuracy

На полученных эмбеддингах обучим классификатор: двуслойную полносвязную нейронную сеть с 513 538 параметрами. Качество на валидации получилось 97.1%, что на 10% выше, чем у предыдущего подхода.

По матрице ошибок (рис. 12) видно, что модель на автографах Жуковского практически не ошибается, на конволютах — 1% ошибок. Распределения вероятности по положительным и отрицательным выборкам получились сильно прижаты к краям, среди конволютов есть ярко выраженные выбросы.

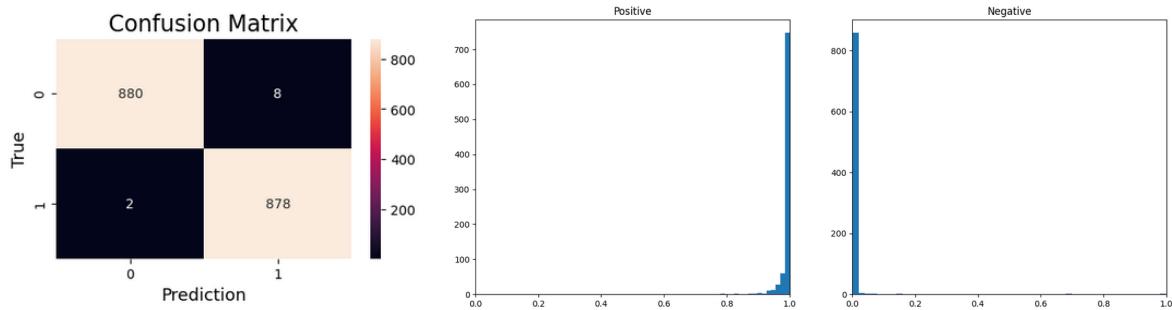


Рис. 12: Матрица ошибок и распределение вероятностей

«Подозрительные» конволюты так же получились хорошего качества.

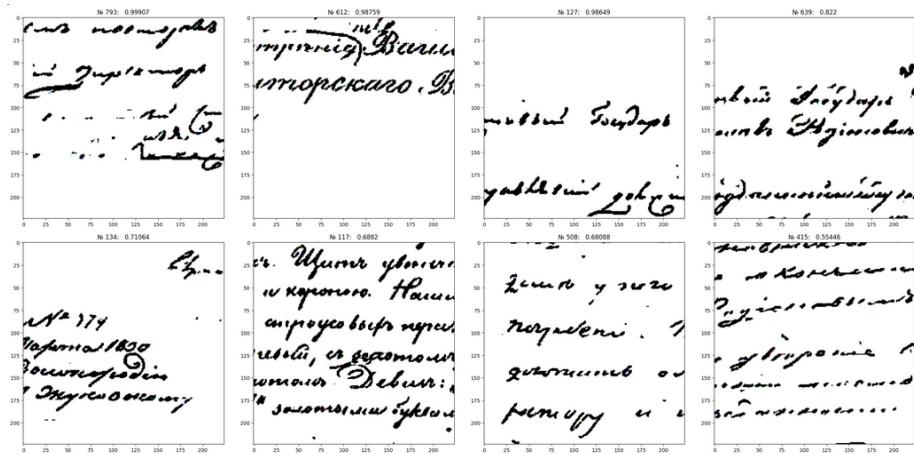


Рис. 13: «Подозрительные» конволюты

Также был проведён аналогичный эксперимент для эмбеддингов размерности 128. Качество получилось незначительно хуже: 99.1% - точность сиамской сети, 96.3% - точность классификации эмбеддингов, но как «подозрительные» модель классифицировала те же конволюты.

## 5.4 Результат

Для получения итоговой вероятности для каждой конволюты были проведены следующие операции: вырезались 5 случайных фрагментов 300 x 300 пикселей, для каждого фрагмента модель предсказывала вероятность, итоговой вероятностью конволюты считается максимальная вероятность среди 5 фрагментов.



Рис. 14: Топ конволют с максимальной вероятностью

## 5.5 Датасет IAM

Так как к конволютам не прилагается точная разметка авторов, мы не можем объективно оценить качество результата.



Рис. 15: IAM

Протестируем предложенный подход на аналогичной задаче классификации почерков на корпусе IAM [8]. Он содержит 1539 картинок с 657 авторами. Эксперимент будет проводиться на укороченном IAM, так как для многих авторов приведён только один пример. Отберём только тех авторов, у которых не менее пяти объектов.

Укороченный IAM содержит 397 картинок с 39 авторами. Также обрежем верхнюю и нижнюю часть изображения с печатным текстом и подписью автора. Примеры изображений разных авторов приведены на рисунке 15.

Сиамская сеть с размерностью эмбеддинга 1000 обучалась на наборе из 1000 триплетов и достигла точности 100% на валидации (рис. 16). Точность двухслойного классификатора на 39 классов — 98.75% (рис. 17).

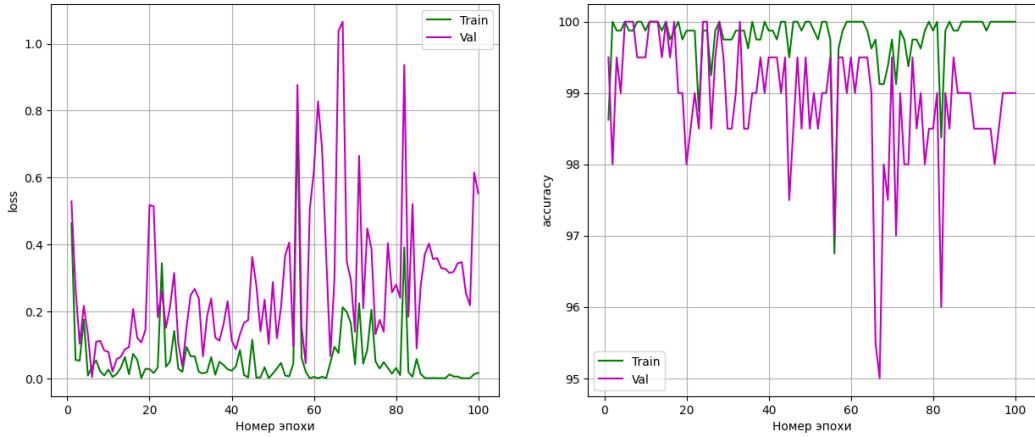


Рис. 16: Triplet loss и accuracy

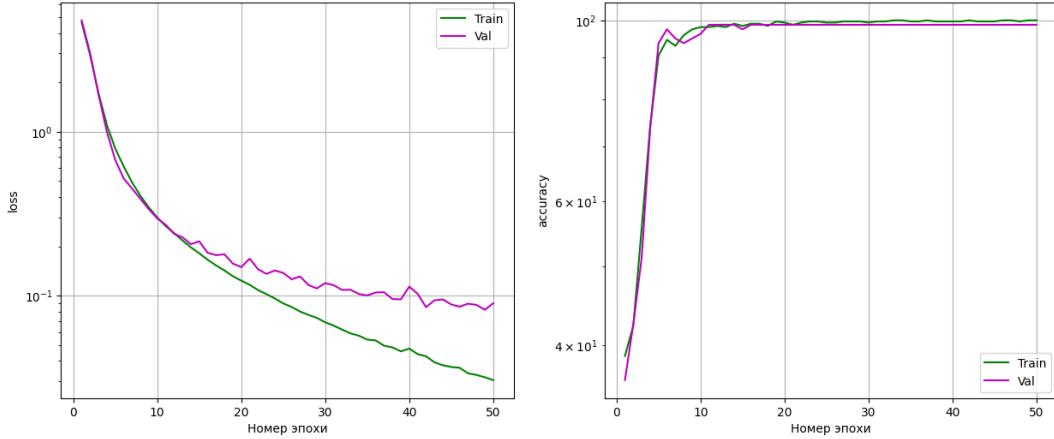


Рис. 17: Loss и accuracy

Такая высокая точность по сравнению с результатом на конволютах объясняется лучшим качеством датасета IAM: в нем нет проблемы разного фона, масштаба написания и расположения текста.

## **6 Заключение**

В данной работе был предложен подход к верификации определённого автора в корпусе документов на основании малого количества образцов. Были поставлены эксперименты, подтверждающие его эффективность, и проведено сравнение с классическим подходом к обучению нейронной сети.

Предложенный метод может быть улучшен с помощью перехода от классификации страницы к классификации строк рукописного текста, это поможет устранить проблему разного масштаба почерка.

## Список литературы

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [2] Enoch Solomon, Abraham Woubie, and Eyael Solomon Emiru. Deep learning based face recognition method using siamese network, 2024.
- [3] Jane Bromley, James Bentz, Leon Bottou, Isabelle Guyon, Yann Lecun, Cliff Moore, Eduard Sackinger, and Rookpak Shah. Signature verification using a "siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7:25, 08 1993.
- [4] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2. Lille, 2015.
- [5] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE, 2006.
- [6] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2015.
- [7] Mohamed Ali Souibgui, Sanket Biswas, Sana Khamekhem Jemni, Yousri Kessentini, Alicia Fornés, Josep Lladós, and Umapada Pal. Docentr: An end-to-end document image enhancement transformer, 2022.
- [8] H. Bunke U. Marti. The iam-database: An english sentence database for off-line handwriting recognition. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 5, pages 39 – 46. IEEE, 2002.