

### **Аннотация**

Задача идентификации авторов для рукописных текстов является актуальной задачей в области компьютерного зрения, заключающейся в определении количества авторов набора рукописных документов и их кластеризации по писателям. Данная работа посвящена решению данной задачи путем использования различных архитектур сверточных нейронных сетей, а также различных методов кластеризации. В результате работы было протестировано множество моделей и методов, и предложены улучшения полученных результатов.

# Содержание

<b>1</b>	<b>Введение</b>	<b>4</b>
1.1	Актуальность . . . . .	4
1.2	Постановка задачи . . . . .	4
<b>2</b>	<b>Обзор существующих методов</b>	<b>5</b>
<b>3</b>	<b>Разработанные решения</b>	<b>6</b>
3.1	Препроцессинг и дальнейшее агрегирование . . . . .	6
3.1.1	Нарезание текста на равные фрагменты . . . . .	6
3.1.2	Corner-detectors . . . . .	6
3.1.3	Агрегирование средним и VLAD . . . . .	6
3.2	Обучение энкодера . . . . .	6
3.2.1	Auto-encoder . . . . .	7
3.2.2	Сиамская нейронная сеть . . . . .	7
3.2.3	Обучение на задаче классификации . . . . .	7
3.3	Кластеризация . . . . .	7
3.3.1	Уменьшение размерности . . . . .	7
3.3.2	Определение количество кластеров . . . . .	7
3.3.3	Алгоритм кластеризации . . . . .	7
<b>4</b>	<b>Результаты проведения экспериментов</b>	<b>7</b>
<b>5</b>	<b>Заключение</b>	<b>7</b>

# **1 Введение**

## **1.1 Актуальность**

Идентификация авторов рукописных текстов является актуальной задачей в области компьютерного зрения. Среди сфер использования данной технологии можно выделить анализ исторических документов, обработку рукописных текстов в судебной практике, кластеризацию огромного количества рукописных текстов в образовательной сфере, а также разметку датасета по писателям в автоматическом режиме, что может помочь улучшить качество работы генеративных нейронных сетей, обученных на рукописных текстах.

## **1.2 Постановка задачи**

Работы по данной теме выделяют онлайн и оффлайн методы распознавания авторов. Онлайн метод подразумевает обработку рукописного текста, который представлен в виде временных фрагментов штрихов, из которых извлекается уникальная информация о писателе. В свою очередь, оффлайн метод проводит анализ изображения уже написанного рукописного текста, из которого извлекаются признаки, по которым выявляется автор.

Задачу идентификации авторов можно решать в постановке как задачи классификации, так и задачи кластеризации. В случае задачи классификации каждый автор представляется из себя отдельным классом, который модель предсказывает, имея на вход рукописный текст. В случае задачи кластеризации, не зная заранее множество авторов и их количество, рукописные фрагменты разбиваются на кластеры, каждый из которых предположительно написал один человек. Стоит отметить, что если задача решена в постановке кластеризации, то она решена в постановке классификации, так как в случае успешной кластеризации, можно сопоставить полученные кластеры уже известным классам.

Данная дипломная работа будет решать задачу идентификации авторов в формулировке оффлайн кластеризации. Имея на входе документы с рукописным текстом, нужно определить количество писателей и кластеризовать тексты по авторам. Документы могут из себя представлять как полноценные тексты на бумаге, так и отдельно написанные от руки слова или предложения. Обученной модели на стадии inference могут подаваться тексты писателей, которых она не видела во время обучения.

## 2 Обзор существующих методов

Исследования в области идентификации авторов рукописных текстов проводились в течении многих лет, и улучшали постепенно результаты, предлагая различные методы и идеи извлечения и обработки признаков рукописного текста.

Для выявления признаков из полученного изображения современные научные работы в основном делают выбор на сверточных нейронных сетях. Используются различные архитектуры, включая ResNet-18 (источник), ResNet-50 (источник), VGG (источник). Данные модели показали хорошие результаты в классификационной постановке задачи, где их применяли в качестве энкодеров.

Научные работы предлагают различные варианты обучения энкодера. Например, некоторые исследования обучают энкодер в паре с полносвязной нейронной сетью, используя функцию потерь CrossEntropy (источник). Также есть исследования в области применения сиамской архитектуры обучения энкодера на данной задаче (источник).

Существует также несколько способов извлечения фрагментов из рукописного текста для дальнейшего извлечения признаков. Один из самых простых способов заключается в нарезании рукописного текста на слова или просто на фрагменты определенной ширины (источник). В некоторых работах из рукописного текста извлекаются самые информативные элементы почерка, которые обнаруживаются различными алгоритмами обнаружения углов (corner-detectors), например, HARRIS и FAST. После прохождения через сверточную нейронную сеть, полученные эмбединги потом агрегируются различными способами. Например находится среднее арифметическое векторов или используется алгоритм агрегации VLAD.

В целях значительного увеличения датасета и, в последствии, улучшения качества обучения, существует идея синтетической генерации датасета рукописных текстов, используя шрифты, похожие на рукописный текст, и применяя аугментацию (источник).

Также, в исследованиях распознавания лиц применяется техника обучения Metric Learning, которая помогает в той области получить более репрезентативные эмбединги. Так, используя функцию потерь ArcFace удалось достичь значительного улучшения результата в задачи классификации фотографий лиц людей (источник). Не исключено, что данный метод хорошо себя может показать и на рукописных текстах.

### **3 Разработанные решения**

Исходя из вышеописанных работ, можно составить общую архитектуру решения поставленной задачи. Рукописные тексты сначала проходят через стадию предобработки, во время которой улучшается качество самого рукописного текста, а также происходит его разбивка на фрагменты, либо путем нарезания на слова/части одинаковой ширины, либо путем применения алгоритма нахождения углов для получения максимально репрезентативных элементов почерка. Далее, эти фрагменты поступают в энкодер, который представляет из себя сверточную нейронную сеть, в результате чего получаются эмбединги. Далее, эти эмбединги при необходимости агрегируются в глобальный эмбединг фрагмента текста, если ранее был применен corner-detector. Наконец, применяется алгоритм уменьшения размерности эмбедингов для улучшения качества кластеризации и применяется сам алгоритм кластеризации.

#### **3.1 Препроцессинг и дальнейшее агрегирование**

На вход энкодеру не подается целое изображение документа рукописного текста, так как в нем может содержаться лишняя информация, и энкодеру может быть сложно извлечь репрезентативные признаки из него. Вместо этого рукописный текст поддается предобработке.

##### **3.1.1 Нарезание текста на равные фрагменты**

##### **3.1.2 Corner-detectors**

##### **3.1.3 Агрегирование средним и VLAD**

#### **3.2 Обучение энкодера**

Вышеописанная архитектура предполагает, что энкодер уже был обучен выдавать репрезентативные эмбединги. Этого можно добиться несколькими способами, которые будут описаны далее в данной главе.

### **3.2.1 Auto-encoder**

### **3.2.2 Сиамская нейронная сеть**

### **3.2.3 Обучение на задаче классификации**

## **3.3 Кластеризация**

Существует множество алгоритмов кластеризации, некоторые из которых принимают на вход уже известное количество кластеров, которое в нашем случае является неизвестным.

### **3.3.1 Уменьшение размерности**

### **3.3.2 Определение количество кластеров**

### **3.3.3 Алгоритм кластеризации**

## **4 Результаты проведения экспериментов**

## **5 Заключение**