

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Южно-Уральский государственный университет (национальный исследовательский университет)»
Высшая школа электроники и компьютерных наук
Кафедра системного программирования

Разработка веб-сервиса для распознавания рукописного текста с использованием нейросетевых технологий

Научный руководитель:
ст. преподаватель кафедры СП
Я.А. Краева

Автор работы:
студент группы КЭ-401
А.В. Послыхалин

Челябинск, 2024

Актуальность

- Цифровизация рукописных медицинских карт пациентов
- Распознавание и обработка банками рукописных чеков и платежных документов
- Приведение рукописных записей конспектов студентами в удобный цифровой вид

Цель и задачи исследования

Цель работы: разработать веб-сервис для распознавания рукописного текста с использованием нейросетевых технологий

Задачи:

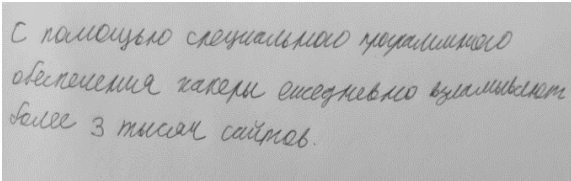
1. Провести анализ предметной области
2. Спроектировать архитектуру сервиса для распознавания рукописного текста с фотографии
3. Спроектировать и реализовать нейросетевую модель для распознавания рукописного текста
4. Разработать веб-сервис, предоставляющий возможность распознать текст с фотографии
5. Провести вычислительные эксперименты по исследованию эффективности нейросети
6. Выполнить тестирование веб-сервиса

Обзор аналогов

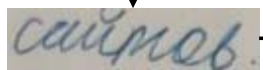
Критерий	rehand.ru	
Использование нейронных сетей для распознавания	+	+
Возможность распознавания текста на русском языке	+	-
Возможность распознавания с группы изображений и формирования единого текстового документа	-	+
Бесплатность платформы	-	-

Распознавание рукописного текста с изображения

Изображение
рукописного
текста



Сегментация



Предобработка

Распознавание 

Постобработка

СЛОВО
«сайтов»

конкатенация

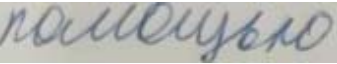
⋮

⋮

⋮

⋮

+



Предобработка

Распознавание 

Постобработка

СЛОВО
«ПОМОЩЬЮ»

Изображения
слов

 – алгоритмы компьютерного зрения

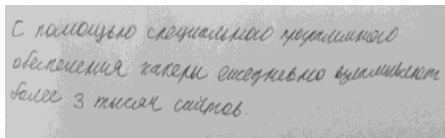
 – нейронные сети

Распознанный
текст

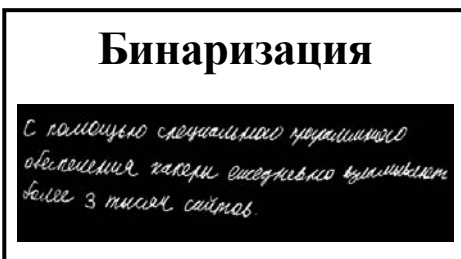
С помощью специального программного обеспечения хакеры ежедневно взламывают более 3 тысяч сайтов.

Сегментация изображения текста

Изображение
рукописного текста



Бинаризация

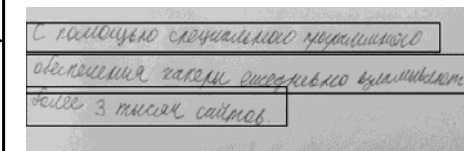


Сегментация текста на строки

Дилатация



Поиск контуров строк

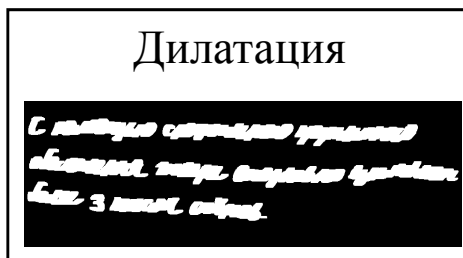


Изображения
слов

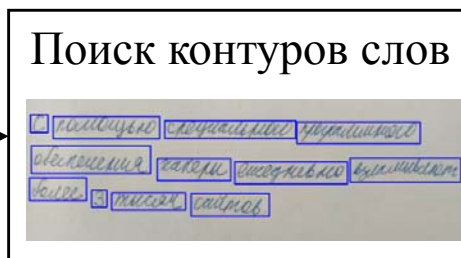


Сегментация строк на слова

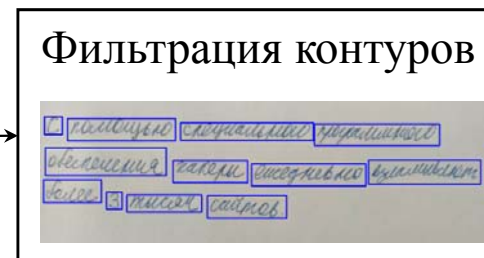
Дилатация



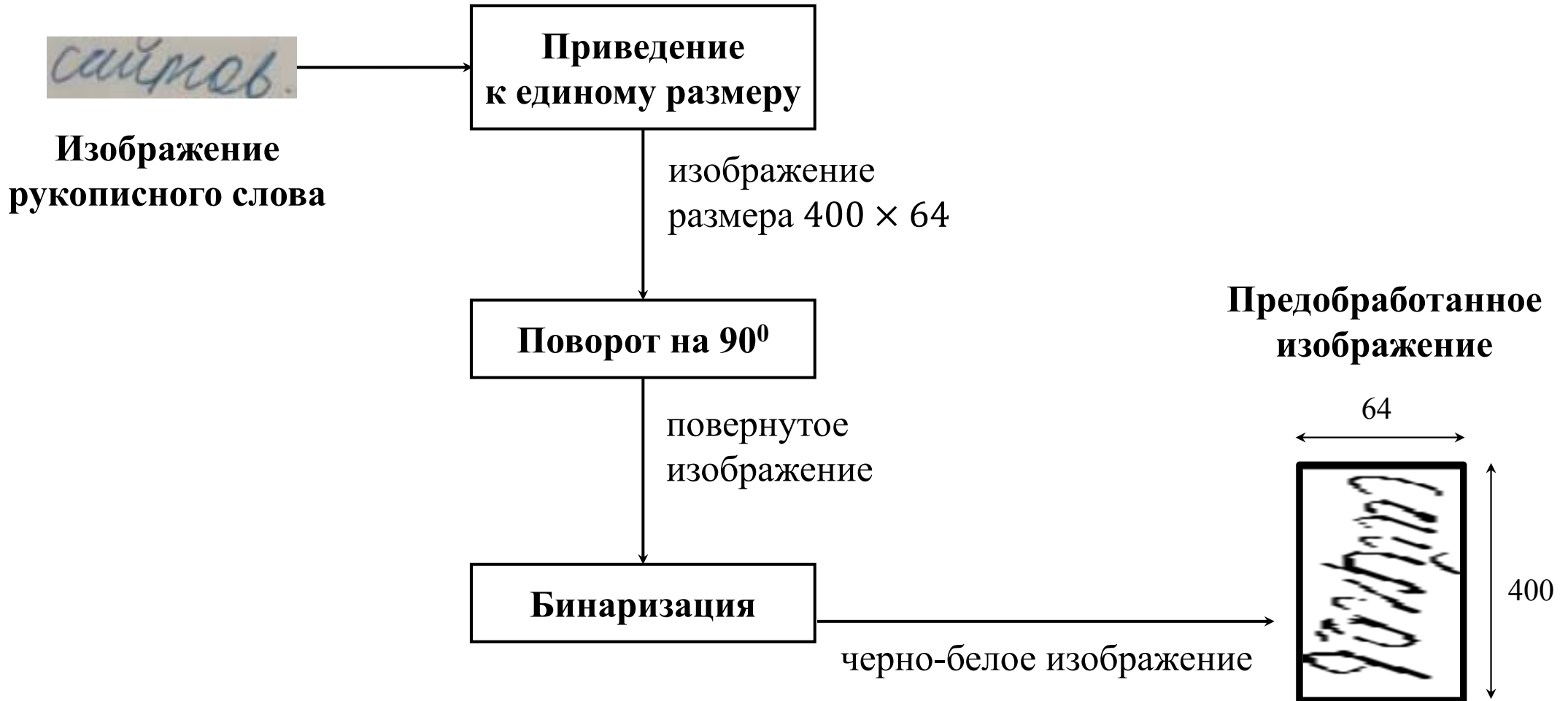
Поиск контуров слов



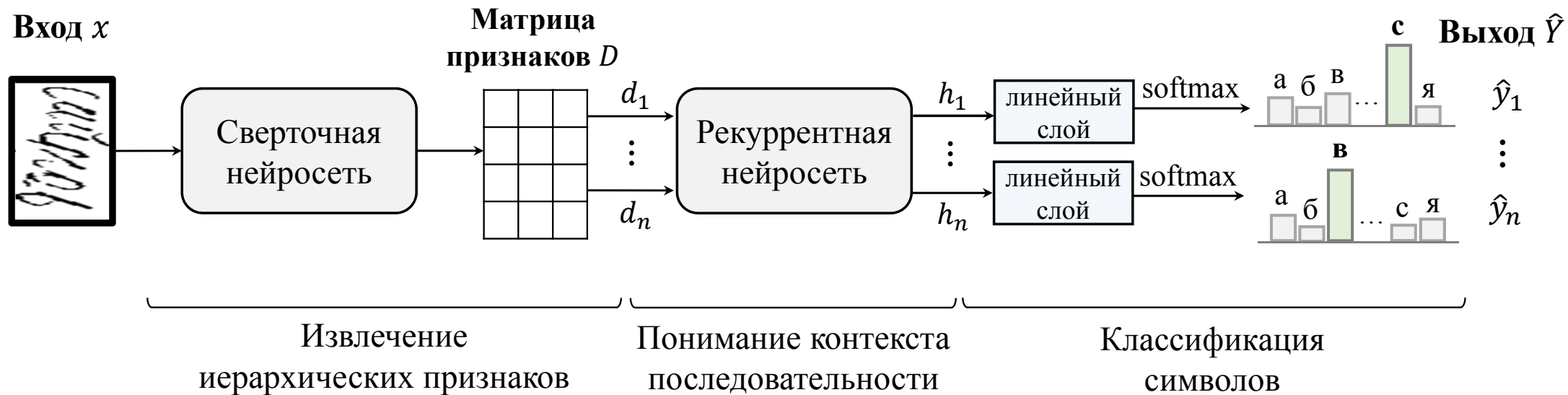
Фильтрация контуров



Предобработка изображения слова

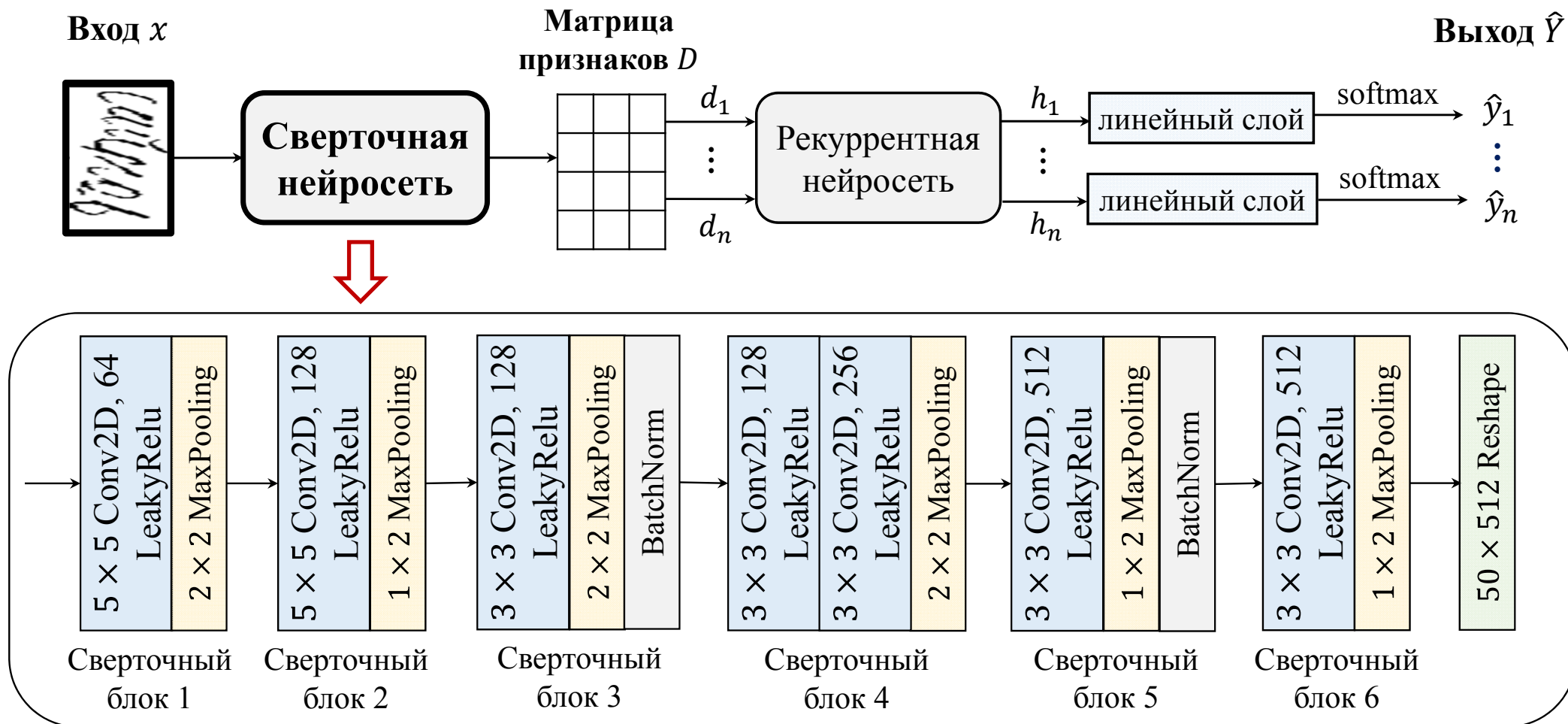


Распознавание: архитектура нейронной сети

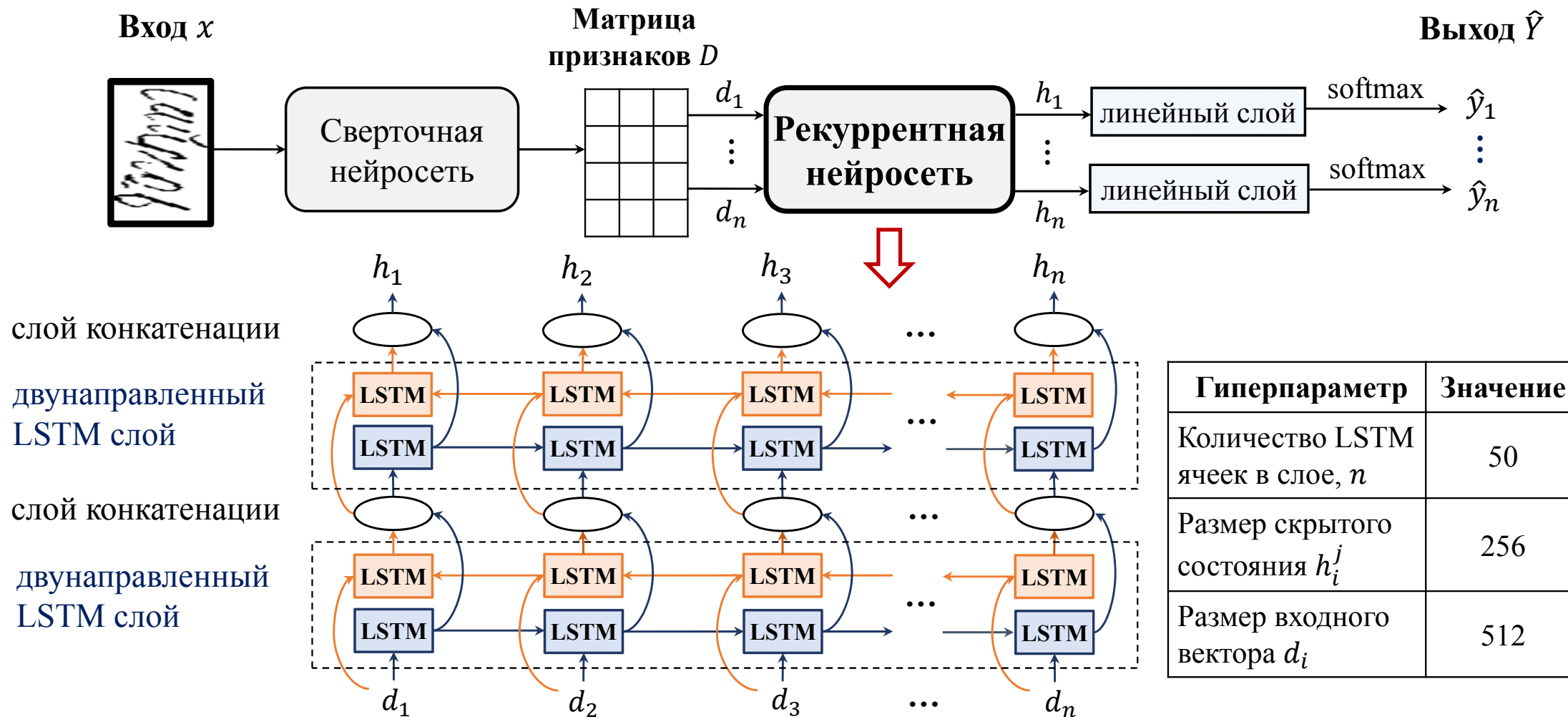


- **Вход:** цифровое представление изображения слова $x \in \mathbb{R}^{400 \times 64}$
- **Выход:** классы символов $\hat{Y} = \{\hat{y}_i\}_{i=1}^n$, $\hat{y}_i \in \mathcal{A}$ из алфавита $\mathcal{A} = \{a, б, \dots, я, -\}$ для входного изображения x
- **Функция потерь:** Connectionist Temporal Classification (CTC)

Распознавание: сверточная часть нейросети



Распознавание: рекуррентная часть нейросети



Архитектура системы распознавания текста



Средства разработки

Клиентская часть:

- Язык программирования: JavaScript ES5
- Верстка: HTML 5 и CSS 3

Серверная часть:

- Язык программирования: Python 3.10
- Библиотеки:
 - создание и обучение нейронных сетей: Tensorflow 2.5.0, Keras 2.12.0
 - обработка и анализ данных: Numpy 1.23.5, Pandas 2.0.0, Scikit-learn 1.2.2
 - предобработка изображений: OpenCV 4.8.0
- Фреймворк: Flask 2.3.2

Репозиторий: <https://github.com/AleksandrPoslihaln/HandWrittenTextRecognition>

Графический интерфейс

Вид сайта *до* загрузки изображения

Распознавание рукописного текста с изображения

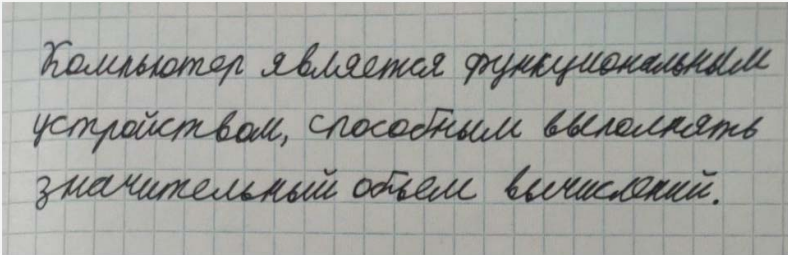
Выберите файл

Распознанный текст:

Вид сайта *после* загрузки изображения

Распознавание рукописного текста с изображения

Выберите файл



Распознанный текст:

компьютер является функциональным устройством, способным выполнять значительный объем вычислений

Вычислительные эксперименты: настройки

- **Набор данных:** Cyrillic Handwriting Dataset*

- 72 286 размеченных изображений для тренировочной и валидационной выборок
- 1 544 для тестовой выборки

- **Гиперпараметры обучения**

Гиперпараметр	Значение
Скорость обучения	0.001
Оптимизатор	Nadam
Максимальное количество эпох	75
Размер подвыборки	64

- **Метрика точности:** CER (Char Error Rate)

$$CER = \frac{I + D + S}{N},$$

- где S – число замен,
 D – число удалений,
 I – число необходимых вставок,
 N – число символов в исходной строке.

* Набор данных изображений рукописных слов и словосочетаний на русском языке. URL: <https://www.kaggle.com/datasets/constantinwerner/cyrillic-handwriting-dataset>

Вычислительные эксперименты: точность распознавания

Итоговое значение метрики
точности распознавания

$$CER = 0.21$$



Правильно распознанных
символов составило 79 %

Входные изображения		Результат распознавания
Изображение	Расшифровка	
	Осталось	осталось
	Оптическое	ОлТиическое
	г. Ульяновск	г. кльеновси
	паспорт	поспор
	отправление	отправление

Тестирование веб-сервиса

№	Описание теста	Действие	Ожидаемый результат	Результат
1	Отображение элементов интерфейса веб-сервиса	Пользователь заходит на сайт	Отображаются все элементы интерфейса веб-сервиса	Пройден
2	Отображение загруженного изображения и распознанного текста после выбора файла с корректным расширением и нажатия кнопки «Загрузить»	Пользователь выбирает файл с корректным расширением, нажимает на кнопку «Загрузить»	В области с результатами отображается загруженное изображение и весь распознанный с него текст	Пройден
3	Отображение сообщения об ошибке после выбора файла с некорректным расширением и нажатие кнопки «Загрузить»	Пользователь выбирает файл с некорректным расширением нажимает на кнопку «Загрузить»	Выводится уведомление о неправильно выбранном файле, вывод в области результатов отсутствует	Пройден
4	Отображение сообщения об ошибке после нажатия кнопки «Загрузить», которому не предшествовал выбор файла	Пользователь нажимает на кнопку «Загрузить», не выбирая файл перед этим	Выводится уведомление о необходимости выбрать файл, вывод в области результатов отсутствует	Пройден

Основные результаты

1. Проведен анализ предметной области
2. Спроектирована архитектура веб-сервиса для распознавания рукописного текста на фотографии
3. Спроектирована и реализована нейросетевая модель для распознавания рукописного текста
4. Разработан веб-сервис, предоставляющий возможность распознавать рукописный текст с фотографии
5. Проведены вычислительные эксперименты по исследованию эффективности нейронной сети
6. Выполнено тестирование веб-сервиса

Диаграмма вариантов использования

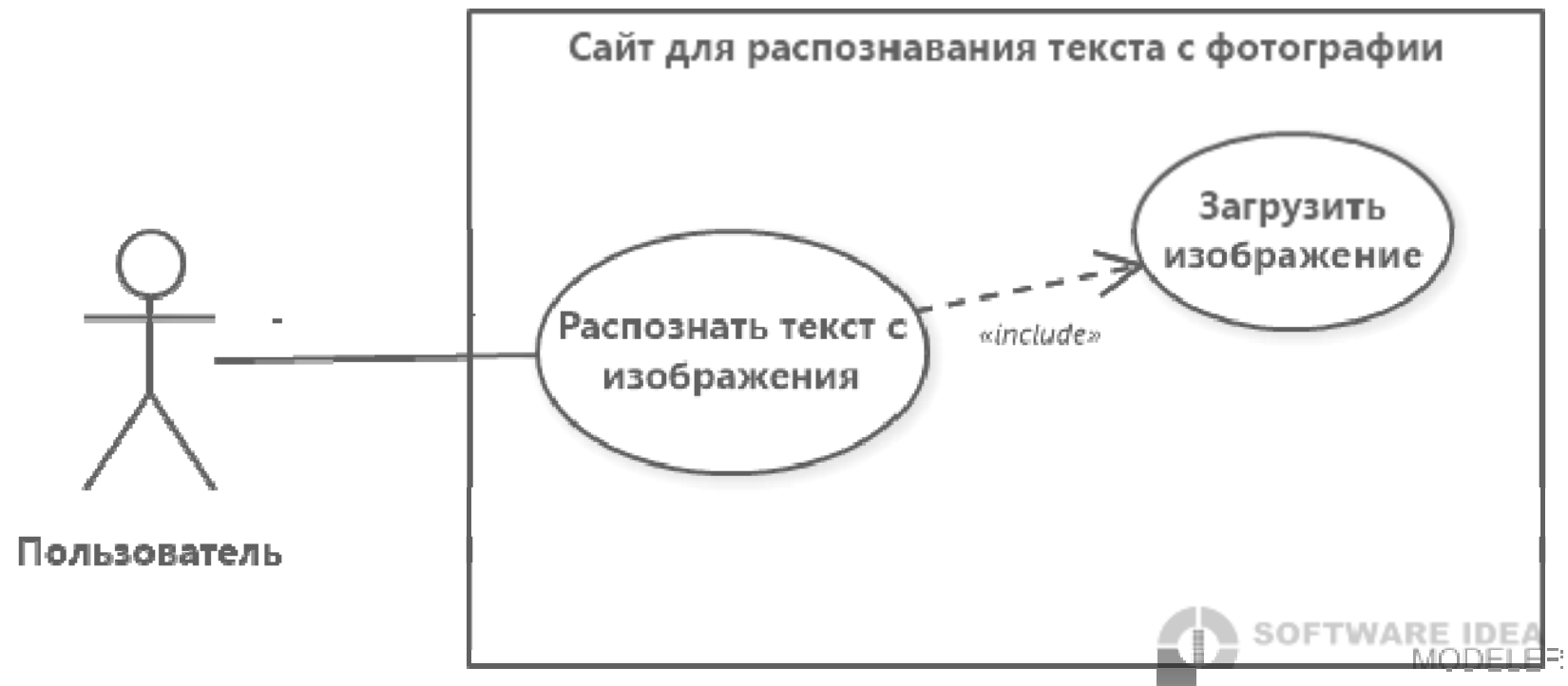
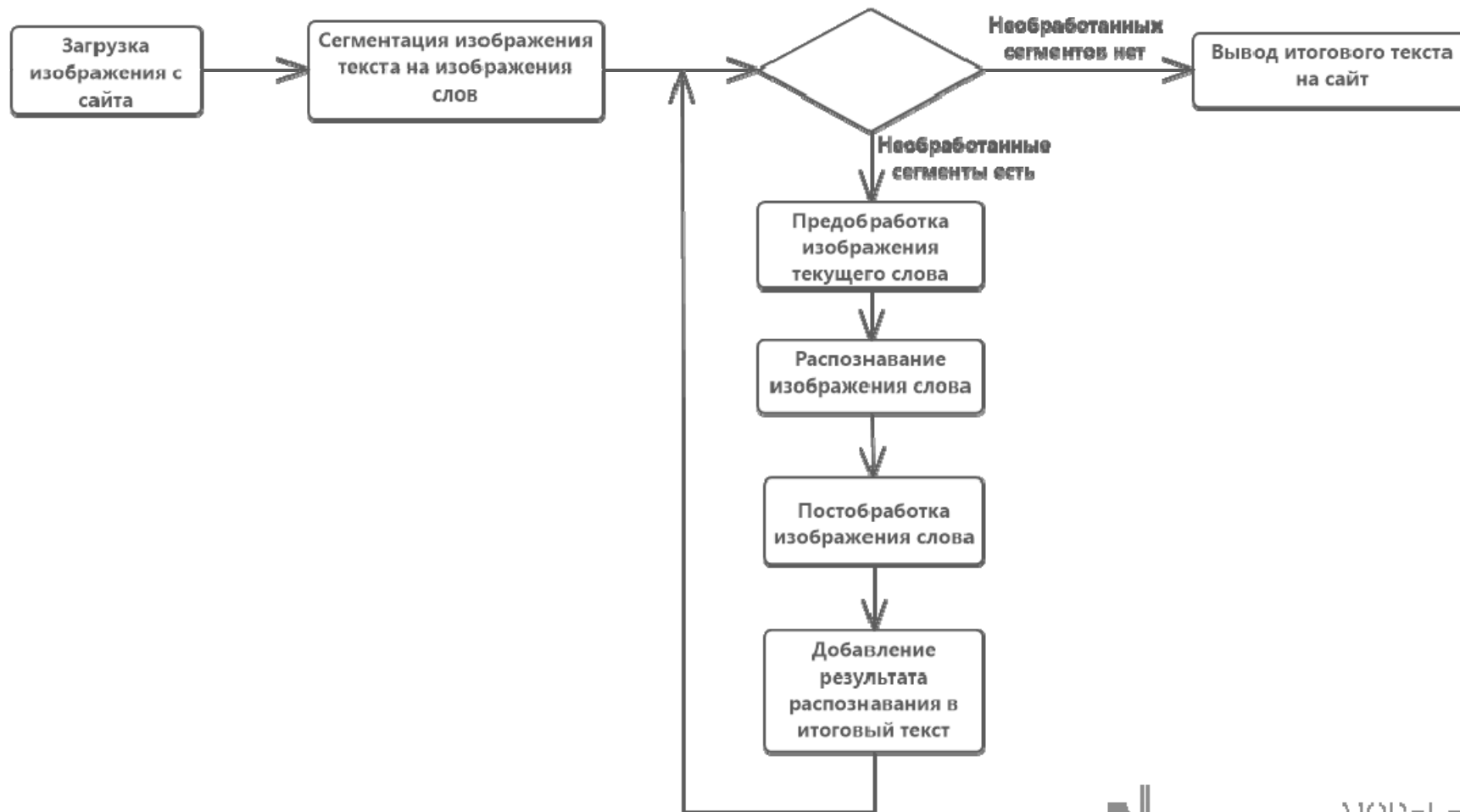


Диаграмма деятельности создания и обучения нейросети



Схема работы приложения



Тестирование компонента сегментации

- **Набор данных:** 10 собственноручно написанных текстов, содержащих 180 слов
- **Метрики точности:** Precision (точность) и Recall (полнота)

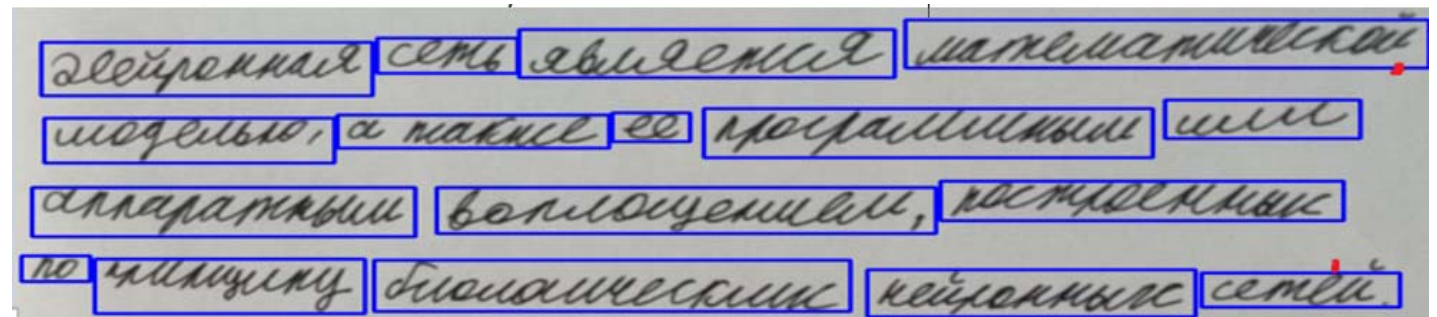
$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN},$$

где TP , FP , FN – количество истинно-положительных, ложно-положительных и ложно-отрицательных сегментированных слов

- **Результаты:**

$Precision = 95.6\%$

$Recall = 98.3\%$



Пример результатов сегментации текста