

ИССЛЕДОВАНИЕ СОВРЕМЕННЫХ СИСТЕМ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ СИМВОЛОВ

А.С. Филиппов¹

¹ Самарский университет, г. Самара
e-mail: alexfil1402@gmail.com

Аннотация

В настоящей работе приведены результаты сравнительного анализа open-source систем оптического распознавания символов (OCR): EasyOCR и Tesseract. Рассмотрены методы цифровой обработки изображений и детектирования документа на кадрах видео. Экспериментально доказано превосходство EasyOCR в распознавании символов в различных сценариях работы.

Ключевые слова: OCR, цифровая обработка изображений, матрица преобразования перспективы, CER, WER

Введение

В наши дни возрастает потребность в использовании решений с открытым исходным кодом, в том числе и для задач оптического распознавания символов. Цель работы – исследование open-source OCR-систем, включая анализ архитектур и оценку точности распознавания текста в различных сценариях (печатный и рукописный текст, обработка видеопотока).

1. Обзор OCR-систем

Проведя анализ научных публикаций и обзор современных исследований, в качестве основных кандидатов были выделены EasyOCR и Tesseract. Данные OCR-системы занимают лидирующие позиции по соотношению точности распознавания, скорости работы и удобству прототипирования. Вокруг них сформировалось активное сообщество и имеется большая база наработок и исследований.

Tesseract – это библиотека с открытым исходным кодом, поддерживаемая Google [1]. В версиях 4.x+ основу составляет рекуррентная нейронная сеть LSTM, обеспечивающая высокий уровень точности и устойчивости к искажениям на входных изображениях [1,2]. Система поддерживает более 123 языков и форматы PNG, JPG, TIFF, PDF [3]. Однако, Tesseract не имеет нативной поддержки работы с видеопотоком в качестве входных данных.

В отличие от Tesseract, EasyOCR является более гибким решением, построенным на PyTorch и распространяемым под лицензией Apache 2.0 [4]. Архитектура модульная: детектор CRAFT локализует текст, CRNN-модель распознаёт символы [5]. Финальный этап декодирования выполняется жадным алгоритмом (Greedy Decoder). В качестве преимуществ EasyOCR можно выделить поддержку гибридного ускорения на CPU/GPU и возможность работы с видеопотоком.

2. Методы предобработки изображений и детектирования документа

Разработанная система предобработки включает детектирование документа в видеопотоке, коррекцию перспективы, выделение области интереса (ROI) и бинаризацию. Для снижения вычислительной нагрузки из видео извлекается каждый десятый кадр. Далее, кадры переводятся в градации серого и бинаризуются для выделения внешних контуров. Путем фильтрации по площади и по соотношению сторон близкому к $1/\sqrt{2}$ (формат A4) позволяет найти оптимальный для распознавания контур документа. На рисунке 1 представлен пример детектирования.

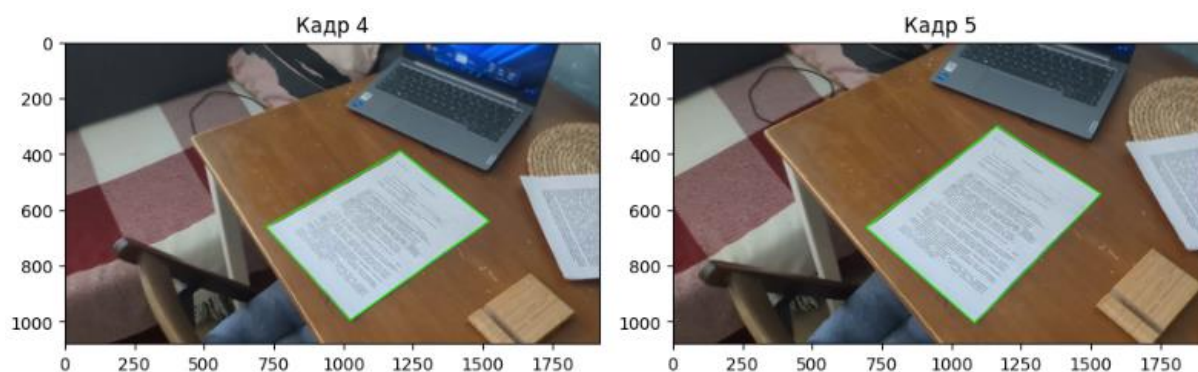


Рисунок 1 – Примеры кадров с детектированным документом

Для коррекции перспективы используется функция `getPerspectiveTransform` опенсорс библиотеки OpenCV [6], вычисляющая матрицу преобразования по координатам углов найденного контура. Затем применяется `warpPerspective`, выравнивающий изображение в проекции, как если бы его сняли сверху перпендикулярно плоскости, на которой он расположен, что позволяет корректно выделить ROI без геометрических искажений.

Для улучшения контраста в области интереса применяются методы CLAHE и маски нерезкости. CLAHE выравнивает освещённость локально, подавляя шумы, а маска нерезкости повышает чёткость границ текста за счёт вычитания из оригинального изображения его размытой копии. Финальная сегментация символов выполняется бинаризацией методом Оцу, который находит порог t , максимизирующий межклассовую дисперсию [8]:

$$\sigma_w^2(t) = w_0(t)\sigma_0^2(t) + w_1(t)\sigma_1^2(t), \quad (1)$$

где w_0 и w_1 – вероятности классов; t – порог бинаризации; σ_0^2 и σ_1^2 – дисперсии классов. Полученное бинарное изображение передаётся в OCR-систему, что повышает точность распознавания. Результат работы обоих методов представлен на рисунке 2.

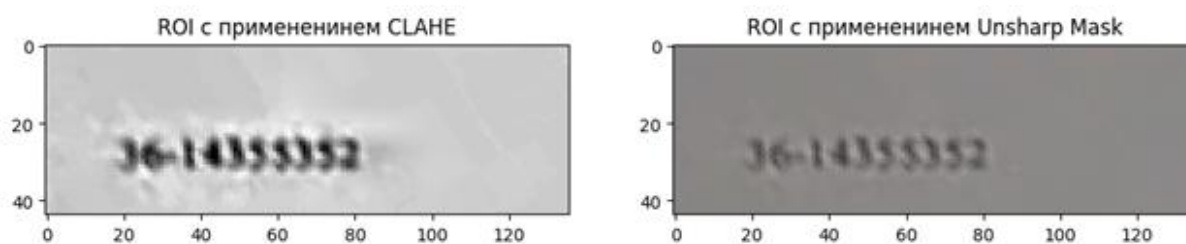


Рисунок 2 – Изображение после применения CLAHE и маски нерезкости

3. Экспериментальная оценка точности

Тестирование опенсорс систем проводилось в различных сценариях: на отсканированных документах, на кадрах из видеопотока и на синтетических данных. Для минимизации ошибок в обеих системах был задан ограниченный словарь символов (английский алфавит, цифры). Все изображения проходили описанный ранее этап предобработки.

На отсканированных документах обе системы корректно распознали печатный текст. При работе с рукописным текстом Tesseract допустил ошибки классификации, тогда как EasyOCR показала корректный результат после корректировки параметра порога достоверности (с 0,7 до 0,5). Аналогичная тенденция сохранилась при тестировании на кадрах из видео: после коррекции перспективы EasyOCR распознала рукописный текст без ошибок, а Tesseract допустил ошибки в схожих между собой символах.

Для получения численной оценки использовались синтетические датасеты (по 1000 изображений) с шумами и искажениями различного вида. Результаты по метрикам CER и WER представлены в таблице 1.

Таблица 1 – Результаты тестирования на синтетических данных

Датасет	Метрика	Tesseract	EasyOCR
Сложный (5 - 12 символов)	CER, %	29,1	8,4
	WER, %	58,6	45,8
Простой (12 цифр)	CER, %	15,8	2,3
	WER, %	18,1	11,7

EasyOCR продемонстрировала существенное преимущество: на сложном датасете метрика CER составила 8,4% против 29,1% у Tesseract, на простом – 2,3% против 15,8%. Разработанные методы предобработки позволили повысить точность распознавания для обеих систем.

Заключение

По результатам исследования рассмотренных OCR-систем установлено, что обе системы демонстрируют высокую точность при распознавании печатного текста на качественных изображениях. Однако, в сценариях с рукописным и зашумлённым текстом EasyOCR показывает значительное преимущество: на синтетических датасетах метрика CER составила 8,4% против 29,1% у Tesseract, а WER – 45,8% против 58,6%.

При работе с видеопотоком EasyOCR также распознала текст корректно после настройки порога достоверности, тогда как Tesseract допустил ошибки классификации символов. Разработанные методы предобработки (детектирование документа, коррекция перспективы, CLAHE, бинаризация по Оцу) позволили повысить точность распознавания для обеих систем.

Список использованных источников

1. Tesseract Open Source OCR Engine // GitHub: [сайт]. – [Б. м.], 2023. – URL: <https://github.com/tesseract-ocr/tesseract> (дата обращения: 29.03.2026)].

2. Модели LSTM и GRU // Машинное и глубокое обучение: [сайт]. – [Б. м.], 2016. – URL: https://deeplearning.ru/docs/Neural_networks/Recurrent-neural-nets/LSTM-GRU (дата обращения: 19.03.2026).
3. Павлов, П.М. Комплексная оценка OCR-систем в задачах распознавания математических выражений на русском и английском языках / П.М. Павлов, В.А. Бекмурзаев // Вестник науки. – 2025. – Т. 1, № 5 (86). – С. 633–648. – URL: <https://cyberleninka.ru/article/n/kompleksnaya-otsenka-ocr-sistem-v-zadachah-raspoznavaniya-matematicheskikh-vyrazheniy-na-russkom-i-angliyskom-yazykah> (дата обращения: 10.03.2026).
4. EasyOCR: [репозиторий] // GitHub. – 2025. – URL: <https://github.com/JaidedAI/EasyOCR> (дата обращения: 11.01.2026).
5. Изучение нейросетевого подхода к решению OCR на примере задачи распознавания арабского текста // Habr.com: [сайт]. – 2023. – URL: <https://habr.com/ru/articles/682270/> (дата обращения: 18.02.2026).
6. OpenCV modules // OpenCV: [сайт]. – [Б. м.], 2023. – URL: <https://docs.opencv.org/4.11.0/index.html> (дата обращения: 23.01.2026).
7. Разработка алгоритма обнаружения дефектов в стеклянных изоляторах на основе компьютерного зрения с использованием нейросетевого подхода / А.В. Коржов, В.А. Сурин, М.А. Ческидова [и др.] // Вестник Южно-Уральского государственного университета. Серия: Математика. Механика. Физика. – 2024. – Т. 16, № 4. – С. 35–42.
8. Доронин, Ю. Д. Методы обработки изображений и использование компьютерного зрения в OCR / Ю. Д. Доронин // StudNet. – 2021. – Т. 4, № 5. – С. 3–20. – URL: <https://cyberleninka.ru/article/n/metody-obrabotki-izobrazheniy-i-ispolzovanie-kompyuternogo-zreniya-v-ocr> (дата обращения: 25.03.2026).

RESEARCH OF MODERN SYSTEMS OPTICAL CHARACTER RECOGNITION

A.S. Filippov¹

¹*Samara National Research University, Samara
e-mail: alexfil1402@gmail.com*

Abstract

This paper presents the results of a comparative analysis of open-source optical character recognition (OCR) systems: EasyOCR and Tesseract. The methods of digital image processing and document detection on video frames are considered. EasyOCR's superiority in character recognition in various operating scenarios has been experimentally proven.

Keywords: OCR, digital image processing, perspective transformation matrix, CER, WER.