Титов П.С., Чупеев А.Д., Шеремет А.А. Создание приложения для распознавания и перевода текста с изображений с использованием компьютерного зрения и обработки естественного языка // Международный журнал информационных технологий и энергоэффективности.– 2025. – Т. 10 № 3(53) с. 149–153

Международный журнал информационных технологий и энергоэффективности

Сайт журнала:
http://www.openaccessscience.ru/index.php/ijcse/

УДК 004.8

# СОЗДАНИЕ ПРИЛОЖЕНИЯ ДЛЯ РАСПОЗНАВАНИЯ И ПЕРЕВОДА ТЕКСТА С ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ КОМПЬЮТЕРНОГО ЗРЕНИЯ И ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

**[1]Титов П.С., [2]Чупеев А.Д., Шеремет А.А.**
*ФГБОУ ВО "ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ", Тюмень, Россия (625000, Тюменская область, город Тюмень, ул. Володарского, д. 38), e-mail: [1]tpptgs@bk.ru, [2]cenz1217@gmail.com*

**В данном работе рассматривается процесс разработки приложения для распознавания и перевода текста с изображений с одного языка на другой. Основной целью работы является создание программного средства, использующее алгоритмы компьютерного зрения для точного извлечения текстовой информации из визуальных данных и применяющее технологии машинного перевода для автоматического преобразования текста на целевой язык.**
**Структура приложения включает несколько ключевых модулей: модуль оптического распознавания символов (OCR), система машинного перевода, а также пользовательский интерфейс для удобного доступа к функционалу приложения.**
**Для реализации проекта используются следующие библиотеки и фреймворки: OpenCV для задач компьютерного зрения и TensorFlow для разработки и обучения нейронных сетей, что обеспечивает высокую точность и производительность системы.**

Ключевые слова: Оптическое распознавание символов, компьютерное зрение, машинное обучение, перевод текста, обработка изображений.

# CREATING AN APPLICATION FOR RECOGNIZING AND TRANSLATING TEXT FROM IMAGES USING COMPUTER VISION AND NATURAL LANGUAGE PROCESSING

**[1]Titov P.S., [2]Chupeev A.D., Sheremet A.A.**
*TYUMEN INDUSTRIAL UNIVERSITY, Tyumen, Russia (625000, Tyumen Region, Tyumen, Volodarskogo St., 38), e-mail: [1]tpptgs@bk.ru, [2]cenz1217@gmail.com*

**This paper examines the process of developing an application for recognizing and translating text from images from one language to another. The main goal of the work is to create a software tool that uses computer vision algorithms to accurately extract text information from visual data and uses machine translation technologies to automatically convert text into the target language.**
**The application structure includes several key modules: an optical character recognition (OCR) module, a machine translation system, and a user interface for easy access to the application's functionality.**
**The following libraries and frameworks are used to implement the project: OpenCV for computer vision tasks and TensorFlow for the development and training of neural networks, which ensures high accuracy and system performance.**

Keywords: Optical character recognition, computer vision, machine learning, text translation, image processing.

## Введение

Современные технологии компьютерного зрения и обработки естественного языка (NLP) значительно расширили возможности автоматического распознавания и перевода текста. Однако, несмотря на значительные достижения в этих областях, интеграция оптического распознавания символов (OCR) и машинного перевода для создания универсальных приложений, способных работать с изображениями и видео в реальном времени, остается сложной задачей. Основная цель данного проекта заключается в разработке приложения, способного распознавать текст с изображений и переводить его с одного языка на другой. Выделенный такие задачи проекта**:**

1. Изучить и проанализировать текущие достижения в области компьютерного зрения и NLP, связанные с распознаванием и переводом текста.
2. Разработать модуль оптического распознавания символов (OCR), способный извлекать текст из изображений и видео с высокой точностью.
3. Интегрировать систему машинного перевода, обеспечивающую точный и контекстно-зависимый перевод текста на целевой язык.
4. Провести тестирование и оценку эффективности приложения в различных сценариях использования.

В процессе работы мы использовали такие технологии: Tesseract (для OCR) [1, 6], EasyOCR [2], Google Translate [3], так как они показали высокую эффективность в своих областях. А также модели BERT [4] и GPT [5].

**Теоретическая часть**

Для распознавания символов на изображении используется алгоритм сегментации текста на уровне слов. Изображение разбивается на части, соответствующие отдельным словам, после чего внутри каждого слова осуществляется дальнейшая сегментация на буквы. Каждая выделенная буква передаётся на классификацию в качестве отдельного изображения, результаты которой сохраняются в массив символов. По завершении, массив символов преобразуется в строку, которая затем передаётся на этап перевода.

Для сегментации слов на изображении был применён метод *morphologyEx* из библиотеки OpenCV с использованием ядра свёртки размером 8x8 пикселей, что обеспечило эффективное выделение текста на уровне слов. Функция *findContours* из библиотеки OpenCV была использована для обнаружения границ каждого слова, что позволило сформировать массив изображений, содержащих отдельные слова. При помощи встроенной функции *sorted* из Python контуры были отсортированы сначала по вертикали (сверху вниз), что позволило разбить текст на строки, затем по горизонтали (слева направо), что сохранило порядок слов в строке.

Для изображений отдельных слов вновь применялись морфологические преобразования, которые описаны в предыдущем абзаце, но с меньшим ядром свёртки размером 8x1 пикселей. Это позволило объединить точки над буквами i и j с основными частями букв, сохраняя при этом разделение между символами.

Пример алгоритма приведен на Рисунке 1

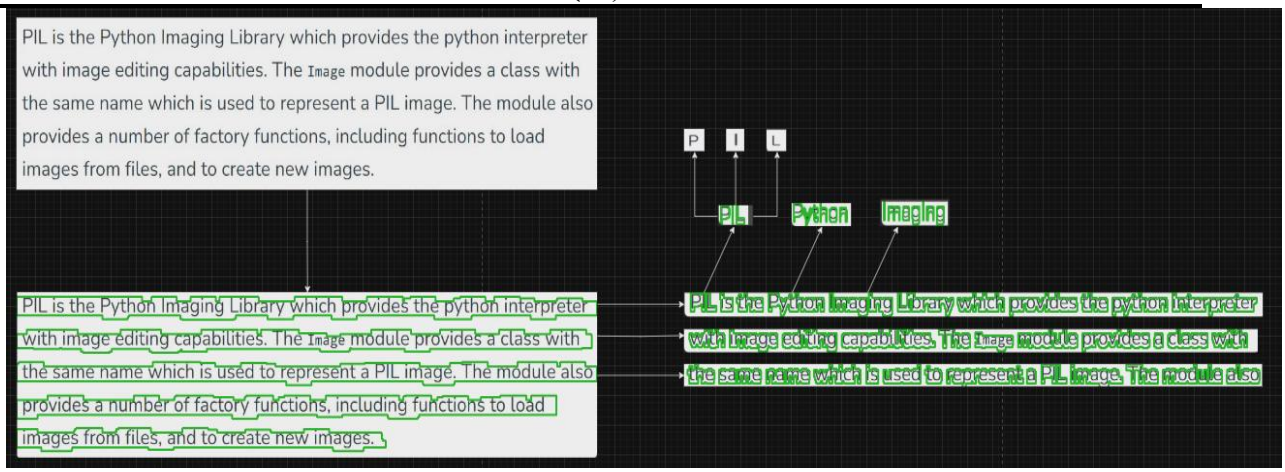


Рисунок 1 - Пошаговое преобразование

**Практическая часть**

Для решения задачи «распознавания букв на изображении», нужно было подготовить данные. Был собран набор данных, включающий изображения текста в различных стилях и шрифтах. Далее буквы различных шрифтов и стилей были выделены в изображения 28x28 пикселей. Далее для расширения выборки и улучшения качества обучение модели была проведена аугментация данных, а именно повороты, сдвиги, изменения яркости изображений. Также для повышения точности обучения модели на каждом изображении были размечены точные границ букв.Следующим этапы было предобработка изображений алгоритмом, описанным в теоретической частей, выделение, сегментация слов и букв на изображении для дальнейшей обработки. Методы предобработки изображений были реализованы с использованием библиотеки OpenCV.

Для решения задачи распознавания букв на изображении, была разработана и обучена модель сверточной нейронной сети (CNN) на основе архитектуры Xception, которая состояла из 4 сверточных слоев и 2 полносвязных слоев. В качестве функции активации использовалась ReLu, а также функция регуляризации для предотвращения переобучения. Обучение с оптимизационным алгоритмом Adam проходило на ранее размеченном наборе данных. Для оценки точности использовался валидационный набор данных, который не участвовал в процессе обучения. Точность на валидационной выборке 70% и 95% на обучающей. Для обучения и создания модели применялся фреймворк TensorFlow.

После успешного завершения обучения модель была использована для распознавания текста на изображении. Модель предсказывала буквы, которые сохранялись в массив символов. Массив преобразовывался в строку и передавался на этап перевода. Постобработка результатов с использованием Google Translate и ChatGPT позволила автоматически исправлять и корректировать распознанные символы, что значительно повысило качество итогового текста.

**Результаты**

Результаты работы программы представляются в виде 3 изображений: исходное изображение с текстом (Рисунок 2), распознанный текст и перевод уже распознанного текста.

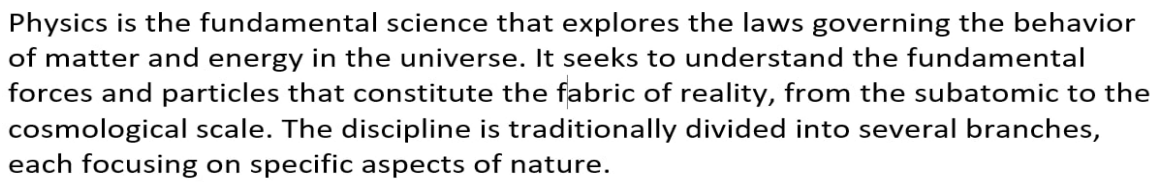


*Рисунок 2 - Исходный текст*

Распознанный нашей моделью текст: Physics is the fundamental science that explores the laws governing the behavior of matter and energy in the universe. It seeks to understand the fundamental forces and particles that constitute the fabric of reality, from the subatomic to the cosmological scale. The discipline is traditionally divided into several branches, each focusing on specific aspects of nature.

Распознанный Tesseract текст: Physics is the fundamental science that explores the laws governing the behavior of matter and energy in the universe. It seeks to understand the fundamental forces and particles that constitute the fabric of reality, from the subatomic to the cosmological scale. The discipline is traditionally divided into several branches, each focusing on specific aspects of nature.

Текст, распознанный нашей моделей и переведенный ChatGPT: Физика — это основная наука, исследующая законы, управляющие материей и энергией во Вселенной. Её цель понять фундаментальные силы и частицы, из которых состоит ткань реальности, охватывая масштабы от субатомного уровня до космологических явлений. Эта дисциплина традиционно делится на несколько областей, каждая из которых сосредоточена на изучении определённых аспектов природы.

Текст, распознанный Tesseract и переведенный GoogleTranslate: Физика - является фундаментальной наукой, изучающей законы, которые определяют поведение материи и энергии во Вселенной. Её задача состоит в исследовании фундаментальных сил и частиц, формирующих структуру реальности, начиная от субатомного уровня и заканчивая космологическими масштабами. Традиционно физика делится на несколько направлений, каждое из которых посвящено изучению конкретных аспектов природы.

**Заключение**

В результате работы разработано приложение, распознающее текст с изображений и переводящее его в реальном времени. Приложение показало удовлетворительные результаты, но выявило области для улучшений.

1. Расширение датасета**:** увеличение тренировочного набора данных повысит устойчивость модели к различным шрифтам, улучшая точность распознавания.

2. Фильтрация некорректных символов**:** для устранения ошибок распознавания символов, отсутствующих в обучающем датасете, возможно создание алгоритмов фильтрации и применение методов обработки естественного языка. Это улучшит качество перевода и итоговые результаты.

Дополнительно можно внедрить интерактивную систему для ручной коррекции ошибок пользователем, что повысит точность работы и улучшит процесс.

**Список литературы**

1. Смит Р. Обзор движка Tesseract OCR // Труды Девятой международной конференции по анализу и распознаванию документов. – IEEE, 2007. – С. 629–633.
2. Джайдед Дж. EasyOCR: Простое оптическое распознавание текста с использованием OpenCV и глубокого обучения [Электронный ресурс]. – Репозиторий GitHub, 2020. – URL: https://github.com/JaidedAI/EasyOCR (дата обращения: 30.03.2024).
3. Ву Й., Шустер М., Чен З., Ле К.В., Норуози М., Махере В. Нейронная система машинного перевода от Google: преодоление разрыва между человеком и машинным переводом // arXiv preprint arXiv:1609.08144, 2016. – 23 с.
4. Девлин Дж., Чанг М.-В., Ли К., Тоутаван К. BERT: Предобучение глубоких двунаправленных трансформеров для понимания языка // Труды Конференции 2019 года Североамериканского отделения Ассоциации по вычислительной лингвистике: Технологии обработки естественного языка. – 2019. – С. 4171–4186.
5. Браун Т., Манн Б., Райдер Н., Суббия М., Каплан Дж., Дхаривал П. Языковые модели как обучающиеся с ограниченным количеством примеров // arXiv preprint arXiv:2005.14165, 2020. – 61 с.
6. Tesseract OCR. Tesseract documentation [Электронный ресурс]. – Режим доступа: https://tesseract-ocr.github.io/tessdoc/ (дата обращения: 05.04.2024).

**References**

1. Smith R. An overview of the Tesseract OCR engine // Proceedings of the Ninth International Conference on Document Analysis and Recognition. – IEEE, 2007. – pp. 629–633. Smith R.
2. Jaided J. EasyOCR: Easy-to-use Optical Character Recognition with OpenCV and deep learning [Electronic resource]. – GitHub Repository, 2020. – URL: https://github.com/JaidedAI/EasyOCR (accessed: 30.03.2024).
3. Wu Y., Schuster M., Chen Z., Le Q.V., Norouzi M., Macherey W. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation // arXiv preprint arXiv:1609.08144, 2016. – p.23/
4. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. – 2019. – pp. 4171–4186.
5. Brown T., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P. Language Models are Few-Shot Learners // arXiv preprint arXiv:2005.14165, 2020. – p.61.
6. Tesseract OCR. Tesseract documentation [Electronic resource]. – URL: https://tesseract-ocr.github.io/tessdoc/ (accessed: 05.04.2024).