**Аннотация проекта: Система распознавания объектов, текста и аудио в видео**

**1. Описание проекта**

Проект представляет собой десктопное приложение для обработки видеофайлов, включающее следующие функции:

* **Распознавание объектов** в видео с использованием нейросетевой модели YOLOv8.
* **Распознавание текста** на отдельных кадрах видео и изображениях с помощью OCR (EasyOCR).
* **Распознавание речи** в аудиодорожке видео с последующей конвертацией в текст.
* **Перевод распознанного текста** на различные языки с использованием Google Translator.
* **Обнаружение и добавление аудиодорожки** в обработанное видео.
* **Сохранение результатов** работы (изображения, текст, аудиофайлы).

**2. Основные функции**

**Обработка изображений**

* Загрузка изображений (форматы: .png, .jpg, .jpeg, .bmp).
* Распознавание объектов с использованием YOLOv8.
* Распознавание текста на изображениях.
* Перевод распознанного текста.
* Сохранение изображения с аннотациями.

**Обработка видео**

* Загрузка видеофайлов (форматы: .mp4, .avi, .mov).
* Распознавание объектов на каждом кадре видео с YOLOv8.
* Остановка видео для распознавания текста на определённом кадре.
* Конвертация аудиодорожки в текст с помощью speech\_recognition.
* Добавление аудиодорожки обратно в обработанное видео с использованием avconv.
* Сохранение обработанного видео с аннотациями.

**Обработка аудио**

* Извлечение аудиодорожки из видео (.mp3, .wav).
* Конвертация аудиофайла в текст.
* Сохранение полученного текста.

**Настройки**

* Выбор модели YOLOv8 для детекции (n, s, m, l, x).
* Изменение порога уверенности для детекции.
* Переключение светлой и тёмной темы интерфейса.
* Возможность выбора языка перевода текста.

**3. Используемые технологии**

**Библиотеки Python**

* **PyQt5** – создание графического интерфейса.
* **OpenCV (cv2)** – обработка изображений и видео.
* **Ultralytics YOLO** – нейросетевая модель для детекции объектов.
* **EasyOCR** – распознавание текста на изображениях.
* **Deep Translator** – перевод текста на другие языки.
* **SpeechRecognition** – конвертация речи в текст.
* **Pydub** – обработка аудиофайлов.
* **Avconv** – обработка и конвертация видео/аудио.

**4. Архитектура проекта**

* **Графический интерфейс (PyQt5)** – вкладки для работы с изображениями, видео и настройками.
* **Обработчики видео** – детекция объектов и распознавание текста.
* **Обработчики аудио** – извлечение и конвертация аудиодорожки.
* **Система сохранения данных** – сохранение обработанных изображений, видео и текстов.

**5. Требования к установке**

* Python 3.8+
* Установленные зависимости:
* pip install PyQt5 opencv-python-headless ultralytics easyocr deep-translator speechrecognition pydub
* **Установленный avconv** (например, через libav-tools для Linux):
* sudo apt install libav-tools

Для Windows рекомендуется скачать avconv.exe и добавить его в PATH.

**6. Заключение**

Этот проект предоставляет мощный инструмент для анализа видео, объединяя технологии компьютерного зрения, обработки текста и речи. Его можно использовать в различных областях, таких как автоматическое создание субтитров, мониторинг объектов, анализ видеоданных и машинный перевод.