**Государственное бюджетное общеобразовательное учреждение**

**города Москвы «Школа №1347»**

**Разработка системы автоматического анализа фото, видео с распознаванием объектов, текста и речи**

Выполнил:

Учащийся 10 «Б» класса ГБОУ Школа № 1347

Смирнов Михаил Павлович

Руководитель работы:

Учитель информатики ГБОУ Школа № 1347

Кузнецов Илья Николаевич

**Москва, 2024**

Оглавление

[Введение 3](#_Toc65352580)

[Цель, постановка задачи 5](#_Toc65352581)

[Методика выполнения 6](#_Toc65352582)

[Практическая и теоретическая значимость 15](#_Toc65352583)

[Результаты и выводы 16](#_Toc65352584)

[Перспективы развития 17](#_Toc65352585)

[Список литературы 18](#_Toc65352586)

# Введение

Современные технологии анализа видео, фото или аудио записи позволяют автоматизировать многие процессы, включая контроль безопасности, анализ поведения, автоматическое создание субтитров и машинный перевод речи. Важно создавать доступные и удобные решения, способные обрабатывать видео и аудио потоки и извлекать полезную информацию.

# Обоснование выбора темы

Распознавание объектов, текста и речи является ключевой задачей компьютерного зрения и искусственного интеллекта. Создание приложения, объединяющего эти функции, поможет автоматизировать обработку видео в различных сферах: от мониторинга городской среды до анализа образовательного контента.

Существующие системы анализа видео зачастую узкоспециализированы: одни решения ориентированы только на детекцию объектов, другие – на распознавание речи, третьи – на обработку текста. Комплексное решение, объединяющее эти технологии, будет полезным инструментом для анализа видеоданных.

# Цель, постановка задачи

**Цель**

Разработать программный инструмент для анализа видео, включающий распознавание объектов, текста и речи.

**Задачи**

1. Реализовать систему детекции объектов с использованием нейросетевой модели YOLOv8.
2. Внедрить алгоритмы распознавания текста на изображениях и кадрах видео (OCR).
3. Добавить функционал обработки аудиодорожки видео и конвертации речи в текст.
4. Разработать графический интерфейс пользователя (GUI) на PyQt5.
5. Реализовать возможность сохранения полученных результатов.
6. Провести тестирование системы на различных видеоданных.

# Методика выполнения работы

## **Программное обеспечение и инструменты**

1. **Язык программирования:** Python 3.12
2. **Библиотеки:**
   * OpenCV – обработка изображений и видео
   * Ultralytics YOLO – детекция объектов
   * EasyOCR – распознавание текста
   * Deep Translator – машинный перевод
   * SpeechRecognition – распознавание речи
   * Pydub – обработка аудиофайлов
   * Avconv – обработка видео/аудио
3. **Фреймворк для GUI:** PyQt5
4. **Среда разработки:** PyCharm
5. **Операционные системы:** Windows

# Аппаратное обеспечение

1. Компьютер с GPU (желательно для ускорения работы YOLO).
2. Видеокамера (для тестирования в реальном времени).
3. Микрофон (для тестирования распознавания речи).

# Методы работы

* Предварительная обработка кадров и применение YOLO для детекции объектов.
* Оптическое распознавание текста (OCR) на отдельных кадрах.
* Конвертация аудиодорожки в текст с помощью SpeechRecognition.
* Внедрение механизмов перевода текста.
* Тестирование системы на различных видеофайлах.

# Место и сроки выполнения работы

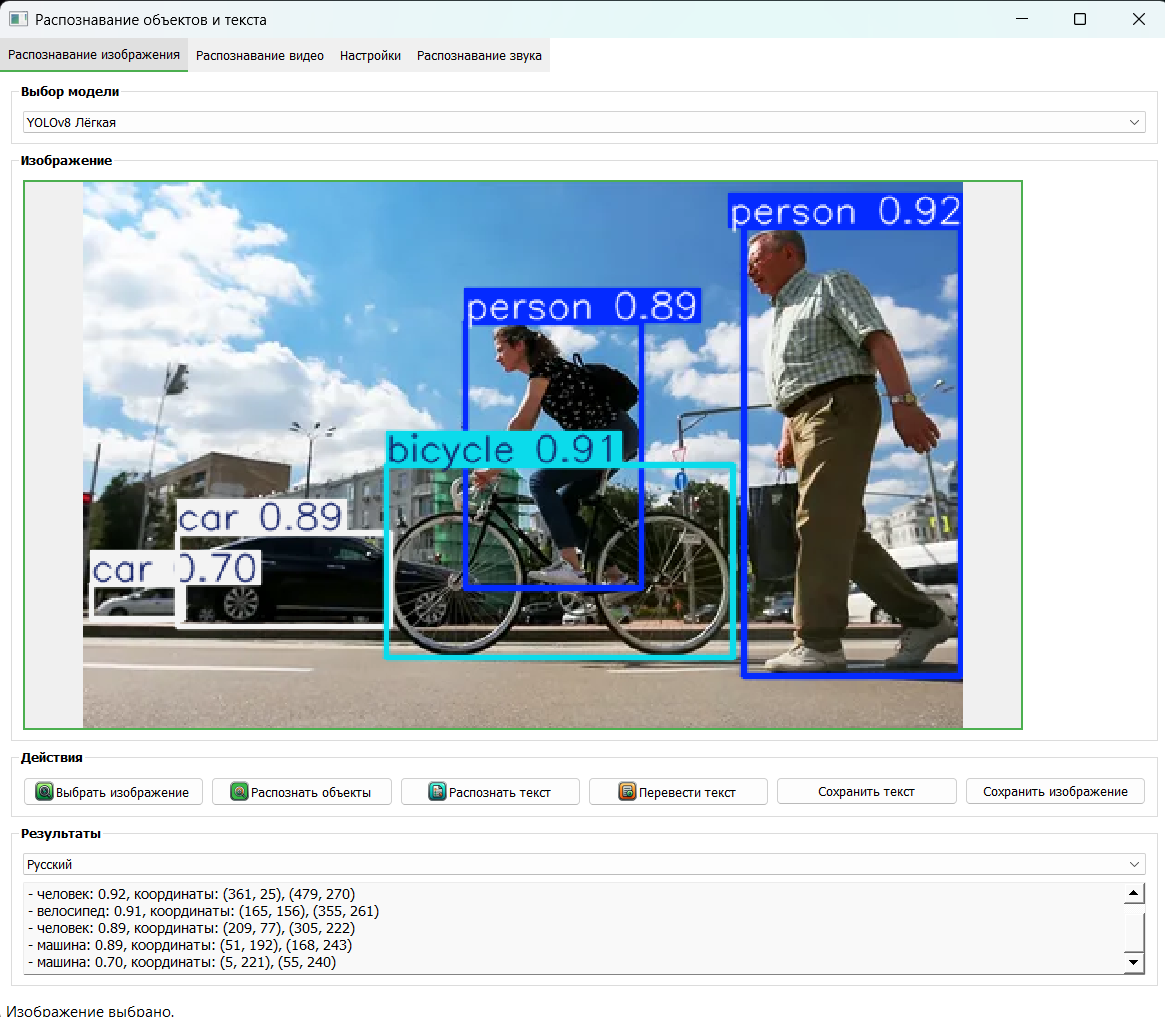
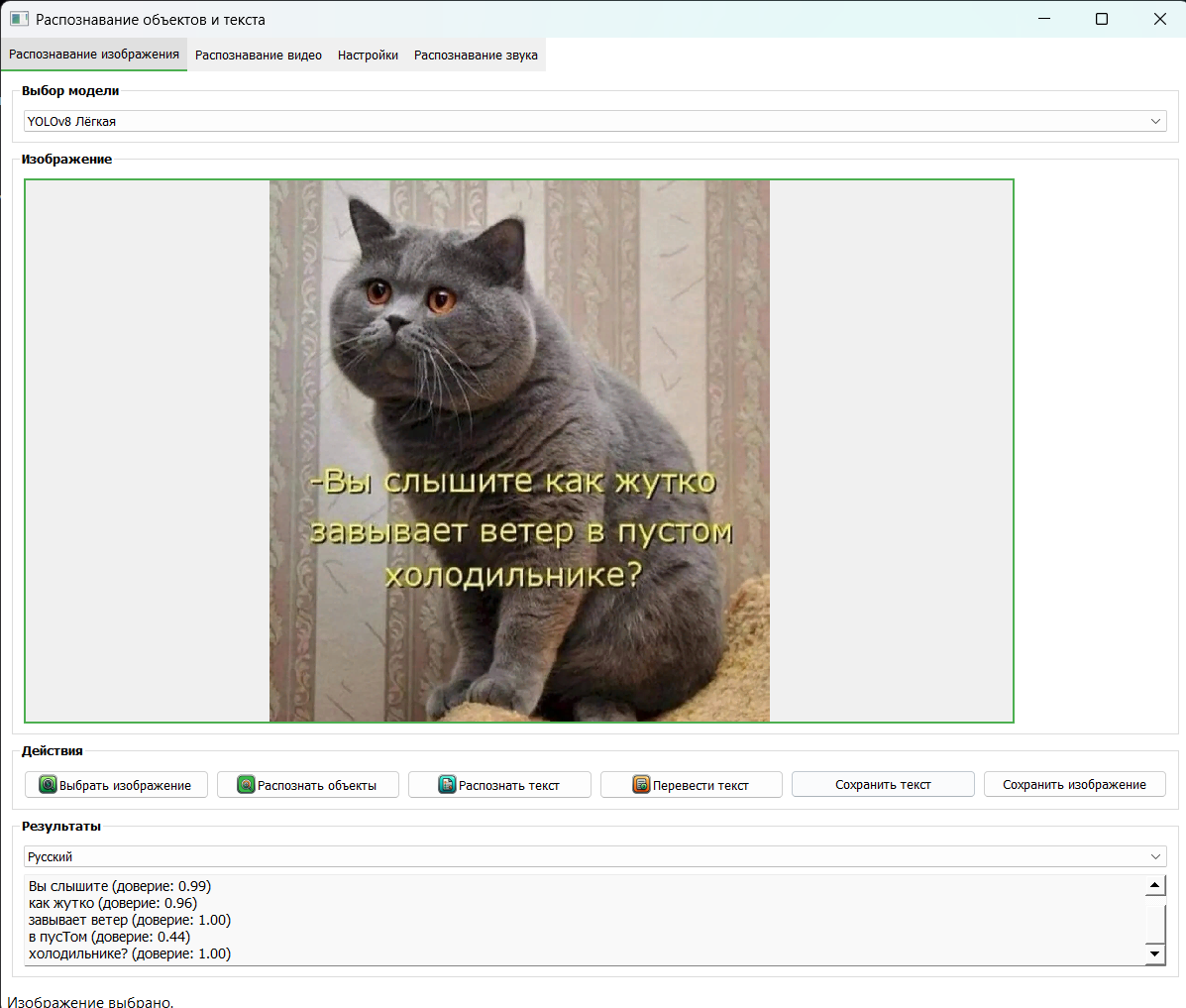
Разработка проводилась в ИТ полигоне ГБОУ Школа 1347 в течении 3 месяцев.

# Результаты работы

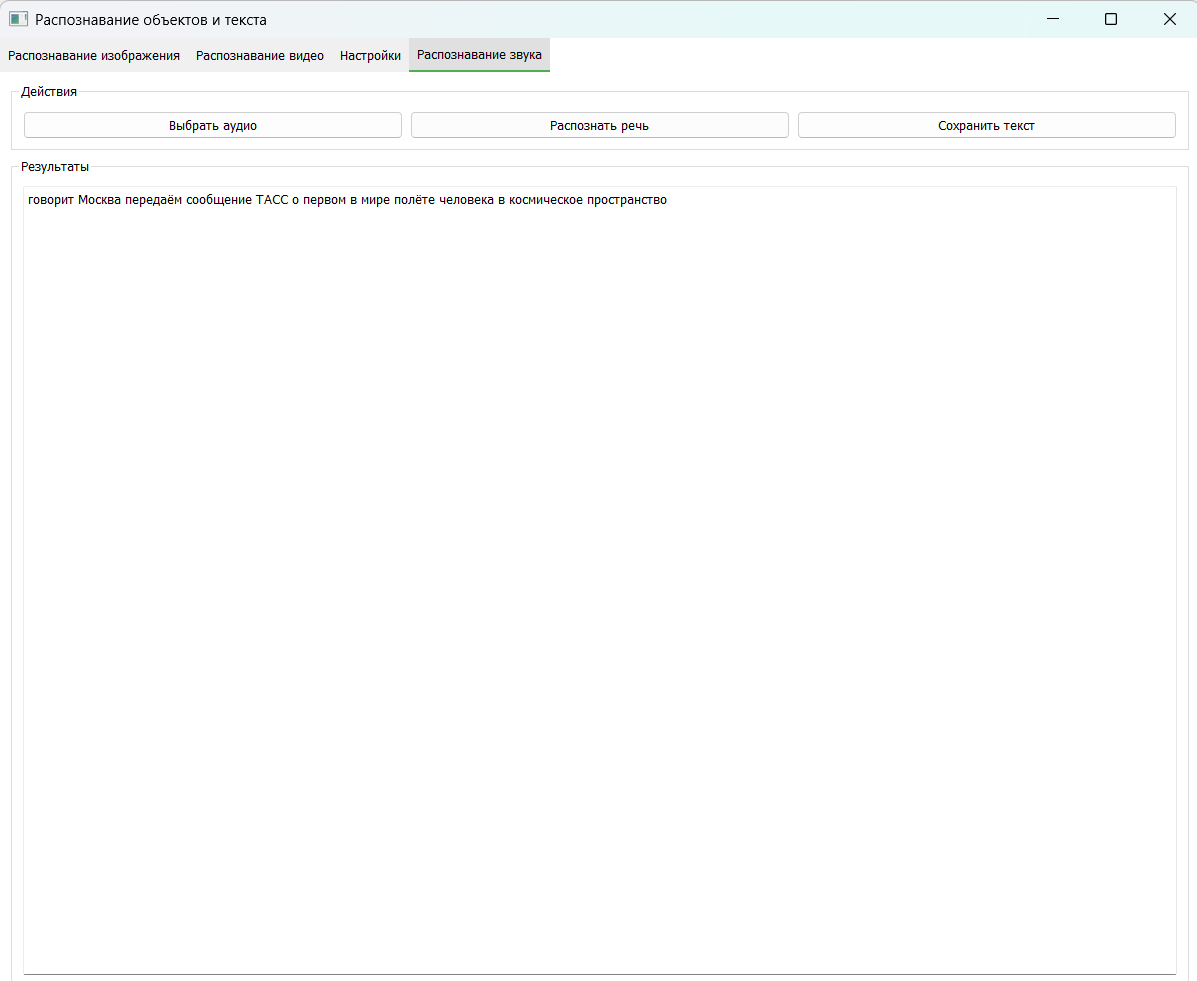
**Функциональные возможности программы**

* Графический интерфейс позволяет загружать и анализировать видео файлы.
* Детекция объектов в кадрах видео с возможностью выбора модели YOLOv8.
* Автоматическое распознавание текста и перевод.
* Анализ аудиофайлов с извлечением текста речи.
* Поддержка различных форматов видео и аудио.
* Возможность сохранения обработанных данных.

**Примеры работы**

Распознавание объектов (рис.1):  
**Распознавание текста (рис.2):  
**

Распознавание речи (рис.3):

https://disk.yandex.ru/d/T3CPckxaxyywvg  
**

**Выводы и перспективы дальнейшей работы**

**Выводы**

1. Разработано программное обеспечение, которое объединяет анализ видео, текста и речи.
2. Система успешно выполняет распознавание объектов, текстов и аудио.
3. Интеграция с PyQt5 обеспечивает удобный пользовательский интерфейс.
4. Использование YOLOv8 позволяет достичь высокой точности детекции объектов.

**Перспективы дальнейшей работы**

* Добавление поддержки онлайн-анализа видеопотоков (например, с камер наблюдения).
* Оптимизация модели YOLO для работы на мобильных устройствах.
* Улучшение алгоритмов распознавания речи и добавление поддержки нескольких языков.

**Список используемой литературы**

1. Redmon J., Farhadi A. "YOLOv3: An Incremental Improvement", 2018.
2. Ultralytics. "YOLOv8 Documentation". https://docs.ultralytics.com
3. Smith R. "An Overview of the Tesseract OCR Engine", 2007.
4. Python SpeechRecognition Library. <https://pypi.org/project/SpeechRecognition/>
5. PyQt5 Documentation. https://www.riverbankcomputing.com/software/pyqt/intro