МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ**

**Н. Г. ЧЕРНЫШЕВСКОГО»**

Факультет компьютерных наук и информационных технологий

УТВЕРЖДАЮ

Зав. кафедрой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ФИО)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись, дата)

**ОТЧЕТ ПО ПРАКТИКЕ**

студента 3 курса, 341 группы, факультета КНиИТ

**Протасова Ивана Леонидовича**

вид практики: учебная ("Научно-исследовательская работа")

кафедра: кафедра информатики и программирования

(наименование кафедры)

курс: 3 .

(номер курса)

семестр: 5 .

(номер семестра)

продолжительность: с 02.09.24 по 27.12.24

(кол-во недель) (дата начала) (дата окончания)

Руководитель практики:

зав. кафедрой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Огнева М. В. .

(должность) (подпись, дата) (ФИО)

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 3](#_Toc185503250)

[1. Обзор существующих решений задач помощи слабовидящим людям с использованием компьютерного зрения 4](#_Toc185503251)

[2. Обзор и выбор технологий для реализации приложения 5](#_Toc185503252)

[2.1. Языки программирования 5](#_Toc185503253)

[2.2. Библиотеки для работы с нейронными сетями 6](#_Toc185503254)

[2.3. Инструменты для мобильной разработки 7](#_Toc185503255)

[2.4. Инструменты для тестирования 8](#_Toc185503256)

[3. Описание предметной области 9](#_Toc185503257)

[3.1. Компьютерное зрение 9](#_Toc185503258)

[3.1.1. Задача распознавания объектов 9](#_Toc185503259)

[3.1.2. Задача распознавания текста 10](#_Toc185503260)

[3.2. Свёрточные нейронные сети 12](#_Toc185503261)

[3.3. Архитектура свёрточной нейронной сети 13](#_Toc185503262)

[3.4. Обучение свёрточной нейронной сети 14](#_Toc185503263)

[3.6. Работа сети после завершения обучения 16](#_Toc185503264)

[ЗАКЛЮЧЕНИЕ 17](#_Toc185503265)

[СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ 18](#_Toc185503266)

# ВВЕДЕНИЕ

**Актуальность** данной темы заключается в необходимости создания решений для облегчения доступа слабовидящих людей к общественному транспорту. Существующие методы, такие как звуковые оповещения и дисплеи высокой контрастности, не всегда эффективны. Использование методов компьютерного зрения позволяет предложить современное и персонализированное решение. Несмотря на успехи в данной области, включая алгоритмы OCR (optical character recognition) и технологии обработки изображений, специализированных решений для распознавания номеров автобусов в сложных условиях пока недостаточно. Важным аспектом данной работы является создание собственного набора данных и разработка специализированной нейронной сети, что позволит учесть специфику задачи и улучшить точность и надежность распознавания.

**Цель научно-исследовательской работы:** изучить методы компьютерного зрения и глубокого обучения, провести анализ существующих решений в области помощи слабовидящим людям, а также определить инструменты и подходы для разработки мобильного приложения с интеграцией нейронной сети.

**Задачи научно-исследовательской работы:**

1. Изучить теоретические основы компьютерного зрения, свёрточных нейронных сетей и глубокого обучения.
2. Провести анализ существующих решений для распознавания текстов и номеров, их применимости в условиях городской среды.
3. Проанализировать возможности существующих инструментов для разработки нейросетей и их интеграции в мобильные приложения.
4. Сформулировать требования к архитектуре системы для распознавания номеров автобусов

# ****1. Обзор существующих решений задач помощи слабовидящим людям с использованием компьютерного зрения****

На сегодняшний день существует множество решений, направленных на распознавание текста и объектов с помощью компьютерного зрения. Однако большинство из них ориентированы на другие задачи или имеют ограничения, которые делают их недостаточно эффективными для распознавания номеров автобусов в условиях городской среды.

**Системы автоматического распознавания номеров (ANPR)**  
широко используются для распознавания автомобильных номеров. Они основаны на алгоритмах компьютерного зрения, таких как YOLO или SSD для детекции объекта и OCR для извлечения текста. Эти системы демонстрируют высокую точность в контролируемых условиях (фиксированные камеры, стабильное освещение). Однако их применение ограничено, поскольку они не предназначены для мобильных устройств, работы в сложных условиях освещения или распознавания номеров на движущемся транспорте.  
 **Seeing AI** от Microsoft способен распознавать текст и объекты, предоставляя голосовую обратную связь. Тем не менее, это приложение являются универсальным и не адаптировано для задач, связанных с идентификацией номеров автобусов. Это приводит к снижению эффективности и удобства использования в подобных сценариях.

**Инфраструктурные решения в транспорте некоторых** городов, такие как внедрение систем озвучивания маршрутов или дисплеев высокой контрастности на автобусах, требуют значительных финансовых вложений и модернизации инфраструктуры, что делает их недоступными для большинства регионов. Более того, их эффективность ограничена, так как они не учитывают индивидуальные потребности слабовидящих пользователей.

**Таким образом**, несмотря на наличие аналогов, ни одно из существующих решений не удовлетворяет всем требованиям задачи распознавания номеров автобусов для слабовидящих пользователей.

# ****2.**** Обзор и выбор технологий для реализации приложения

## 2.1. Языки программирования

**Python** — это высокоуровневый язык программирования, который широко используется в области машинного обучения и искусственного интеллекта. Он имеет богатую экосистему библиотек для работы с данными, таких как NumPy, Pandas, PyTorch, TensorFlow, а также поддерживает удобные фреймворки для разработки и обучения нейронных сетей.

* **Почему выбран**: Python был выбран для разработки нейронной сети в связи с его высокой продуктивностью в области искусственного интеллекта и машинного обучения. Он обладает простой и понятной синтаксисической структурой, что позволяет быстро прототипировать и разрабатывать модели, а также активно поддерживается научным сообществом.
* **Альтернативы**: C++ может быть более эффективным по времени выполнения и имеет большой контроль над низкоуровневыми операциями, что важно для разработки высокоскоростных систем, однако не имеет таких удобных и мощных библиотек для машинного обучения, что делает процесс разработки более трудозатратным. [1]

**Kotlin** — это современный язык программирования, который используется для создания приложений под Android. Kotlin полностью совместим с Java, но обладает более современным синтаксисом и функциональными возможностями.

* **Почему выбран**: Kotlin выбран для создания Android-приложения, поскольку это официальный язык разработки для Android, поддерживаемый Google. Он предоставляет безопасную работу с нулевыми значениями, удобную работу с функциональными методами и лаконичный синтаксис, что делает разработку более быстрой и менее подверженной ошибкам.
* **Альтернативы**: **Java** является еще одним популярным языком для Android-разработки, однако Kotlin предлагает более простое и выразительное решение для современных приложений. [2]

## 2.2. Библиотеки для работы с нейронными сетями

**PyTorch** — это популярная библиотека для работы с нейронными сетями, которая позволяет легко строить, обучать и развертывать модели машинного обучения. Она поддерживает динамическое вычисление графов, что делает её удобной для быстрого прототипирования.

* **Почему выбран**: PyTorch был выбран за свою гибкость, удобство в обучении нейронных сетей и активную поддержку мобильных платформ через **PyTorch Mobile**. PyTorch также имеет развитую документацию и сообщество, что значительно облегчает процесс разработки и внедрения.
* **Альтернативы**: **TensorFlow** — конкурент PyTorch, предоставляющий аналогичные возможности, однако для мобильных приложений PyTorch Mobile обеспечивает более простую интеграцию и поддержку. [3]

**TorchScript** — это способ сериализации моделей PyTorch, который позволяет использовать их на мобильных устройствах. С помощью TorchScript модель PyTorch конвертируется в формат, который можно легко загрузить и выполнить на Android.

* **Почему выбран**: TorchScript выбран для конвертации модели в формат, совместимый с Android-приложением. Это позволяет использовать ту же модель, которая была обучена в Python, в мобильном приложении, обеспечивая высокий уровень производительности и совместимости.
* **Альтернативы**: **ONNX** (Open Neural Network Exchange) также является форматом для сериализации моделей и может быть использован для работы с TensorFlow, PyTorch и другими библиотеками. Однако для использования PyTorch в Android приложении TorchScript является более естественным выбором.

## 2.3. Инструменты для мобильной разработки

**Android Studio** — это официальная среда разработки для Android, которая предоставляет весь необходимый инструментарий для создания приложений, включая редактор кода, эмуляторы, инструменты для отладки и анализа производительности.

* **Почему выбран**: Android Studio это официальная среда разработки для Android, которая поддерживает все необходимые инструменты для интеграции с PyTorch и OpenCV.
* **Альтернативы**: **IntelliJ IDEA** также поддерживает Android-разработку, но Android Studio предоставляет более тесную интеграцию с платформой Android.

**Gradle** — это инструмент для автоматизации сборки, который используется для управления зависимостями, настройки и сборки проекта. Он широко используется в экосистеме Android.

* **Почему выбран**: Gradle это стандартный инструмент для управления зависимостями и сборки проектов на Android, и он идеально подходит для интеграции библиотек, таких как OpenCV, PyTorch и CameraX.
* **Альтернативы**: **Maven** — еще один инструмент для управления зависимостями, однако Gradle предоставляет больше возможностей для гибкой настройки сборки в Android-приложениях.

**CameraX** — это инструменты для работы с камерами на Android. CameraX упрощает работу с камерой, обеспечивая простоту использования и поддержку различных типов устройств.

* **Почему выбран**: CameraX был выбран для работы с камерой, так как он упрощает захват изображений, что особенно важно для приложения, использующего компьютерное зрение для распознавания номеров автобусов. [4]
* **Альтернативы**: **Camera2 API** предоставляет более низкоуровневый доступ к камере, что дает больше возможностей, но требует более сложной настройки. CameraX является более высокоуровневым API, которое подходит для большинства случаев использования.

**OpenCV (for android)** — это библиотека для обработки изображений и видео в реальном времени. В контексте мобильных приложений, она предоставляет удобный набор инструментов для работы с изображениями и видео.

* **Почему выбран**: OpenCV был выбран для обработки изображений с камеры, так как эта библиотека предоставляет широкий функционал для манипуляции изображениями, а также для предварительной обработки данных перед подачей их в нейронную сеть.
* **Альтернативы**: **TensorFlow Lite** также предоставляет инструменты для обработки изображений, но OpenCV предлагает более широкие возможности и гибкость в области обработки изображений.

## 2.4. Инструменты для тестирования

**Android Emulator** — это инструмент для тестирования Android-приложений на различных устройствах без необходимости использовать физическое устройство.

* **Почему выбран**: Android Emulator был выбран для тестирования приложения, так как он позволяет имитировать работу различных Android-устройств, что важно для проверки совместимости и функциональности приложения.
* **Альтернативы**: **Физическое устройство** — может быть использовано для тестирования, однако Android Emulator предоставляет более быстрый и удобный способ тестирования на разных устройствах.

# ****3. Описание предметной области****

## 3.1. Компьютерное зрение

Компьютерное зрение — это область искусственного интеллекта, направленная на создание алгоритмов и систем, которые могут интерпретировать и анализировать содержимое цифровых изображений и видео. Основная задача состоит в том, чтобы компьютер «понимал» изображения или видеопотоки так же, как это делает человек. С точки зрения техники, компьютерное зрение включает в себя обработку данных (цифровых изображений) и применение алгоритмов для выполнения задач, таких как распознавание, сегментация, отслеживание и анализ объектов.

Ключевая особенность компьютерного зрения — способность машин анализировать визуальные данные в реальном времени и адаптироваться к динамически изменяющимся условиям. Это качество имеет решающее значение для задач, связанных с городской средой, где изображения часто содержат большое количество объектов и сложные условия освещения.

### 3.1.1. Задача распознавания объектов

Распознавание объектов (Object Detection) — одна из центральных задач в компьютерном зрении. Её цель состоит в определении местоположения объектов определённых классов на изображении. Задача включает два этапа: детекция объектов и классификация объектов. [5]

**Детекция объектов** — это процесс нахождения и выделения области, где находится объект интереса, на изображении или видео. Она включает не только идентификацию наличия объекта, но и определение его местоположения. Основные этапы детекции:

1. Локализация объекта**:** определение координат прямоугольной рамки, в которой находится объект.
2. Анализ контекста**:** использование дополнительных признаков для подтверждения наличия объекта.

**Классификация объектов** — это процесс присвоения объекта к одному из предопределённых классов. Основные этапы детекции **классификации:**

1. **Подготовка данных:**
   * Удаление шумов и улучшение качества изображения (например, выравнивание освещения).
   * Преобразование выделенной области в формат, пригодный для подачи в нейронную сеть.
2. **Извлечение признаков:**
   * Использование алгоритмов глубокого обучения, чтобы извлечь особенности объекта, такие как контуры и текстуры.
3. **Присвоение класса:**
   * На основе извлечённых признаков объект классифицируется как определённый номер или символ.

В задаче распознавания номеров автобусов детекция помогает выделить область табло автобуса, после чего происходит дальнейший анализ для считывания номера. Это особенно важно в условиях городской среды, где объекты могут частично перекрываться или находиться в движении.

### 3.1.2. Задача распознавания текста

Распознавание текста (Optical Character Recognition, OCR) направлено на извлечение текстовой информации из изображений. Основная сложность OCR состоит в том, что текст может быть написан разными шрифтами, находиться под углом или быть частично затенённым. Задача включает несколько этапов. [6]

1. Предобработка изображения:
   * Градации серого: преобразование изображения в оттенки серого для уменьшения размерности данных.
   * Бинаризация: преобразование изображения в двоичный формат (черное-белое) для акцентирования текста.
   * Удаление шумов: фильтрация мелких дефектов, вызванных внешними условиями (например, дождем или пылью).
   * Выравнивание: коррекция перспективных искажений текста из-за углов съемки.
2. Детекция текста:
   * На данном этапе выделяются области, содержащие текст
3. Распознавание текста (OCR):
   * После выделения текстовой области используется Optical Character Recognition - технология преобразования текстовой информации в цифровой, машиночитаемый текст.
4. Постобработка результатов
   * Исправление ошибок: использование словарей или статистических моделей для устранения ошибок распознавания.
   * Устранение ложных срабатываний: фильтрация неверных областей, например, случайных шумов, распознанных как текст.

В контексте задачи распознавания номеров автобусов OCR используется для извлечения числовой информации с табло автобуса, что требует адаптации алгоритмов к условиям городской среды, например, к низкому качеству изображения или яркому солнечному свету.

## 3.2. Свёрточные нейронные сети

Свёрточные нейронные сети (Convolutional Neural Networks, CNN) — это класс глубоких нейронных сетей, которые особенно эффективны в задачах обработки изображений. CNN автоматически извлекают признаки из изображений, такие как контуры, текстуры и формы, что делает их идеальными для задач компьютерного зрения. [7]

CNN состоят из нескольких основных компонентов:

* **Свертка** — основной строительный блок CNN. Она заключается в применении фильтров к входным данным для выделения различных признаков, таких как края, текстуры и формы. Фильтр скользит по изображению и вычисляет свертку, создавая карту признаков. Этот процесс позволяет выделить важные детали изображения, которые затем используются для дальнейшей обработки.
* После свертки применяется **функция активации**, которая добовляет нелинейность в модель. Это помогает ей лучше справляться с сложными задачами, так как нелинейные функции позволяют нейронной сети моделировать более сложные зависимости.
* **Пулинг** (субдискретизация) уменьшает размер карты признаков, сохраняя наиболее значимые признаки. Это помогает уменьшить количество параметров и вычислительных затрат, сохраняя при этом важные характеристики изображения.

Архитектуры CNN предоставляют гибкость в настройке модели для работы на устройствах с ограниченными вычислительными ресурсами, например, на смартфонах. В задаче распознавания номеров автобусов CNN используются для детекции табло и распознавания текста. [8]

## 3.3. Архитектура свёрточной нейронной сети

Перед подачей в сеть изображения проходят предобработку: изменение размера (например, до 128x128 пикселей), нормализация значений пикселей и преобразование формата данных в тензор. **Входной слой** принимает данный тензор (то есть массив чисел) для дальнейшей обработки нейронной сетью.

**Сверточные слои** выполняют свертку с различными фильтрами для выделения признаков. Каждый сверточный слой создает несколько карт признаков, которые затем передаются на следующий слой. Эти слои являются основными компонентами CNN и выполняют основную работу по выделению признаков из изображения. Каждый сверточный слой может использовать несколько фильтров, каждый из которых выделяет разные признаки. Для распознавания текста на автобусах свёрточные слои помогают выделять ключевые элементы, такие как формы цифр. Глубокие слои могут изучать более сложные абстракции, например, стиль шрифта или расположение текста.

**Слои активации** применяют функцию активации (например, ReLU) к картам признаков, добавляя нелинейность в модель. Это позволяет модели лучше справляться с сложными задачами, так как нелинейные функции позволяют нейронной сети моделировать более сложные зависимости.

**Пулинговые слои** уменьшают размер карт признаков, сохраняя важные признаки и уменьшая вычислительные затраты. Это помогает модели быть более эффективной и быстрой, особенно при обработке больших изображений.

**Полносвязные слои** соединяют все нейроны предыдущего слоя с каждым нейроном текущего слоя. Они используются для окончательной классификации или регрессии. Эти слои являются завершающими компонентами CNN и выполняют основную работу по принятию решений на основе выделенных признаков. В задаче распознавания номера автобуса выход полносвязных слоев может быть интерпретирован как вероятности принадлежности изображения к различным номерам.

**Выходной слой** генерирует окончательный результат, для задачи распознавания текста это может быть один из заранее заданных классов (например, цифры от 0 до 9 и буквы).

## 3.4. Обучение свёрточной нейронной сети

**Датасет**: для обучения модели необходимо собрать и разметить набор данных. В данному случае это изображения автобусов с чётко видимыми номерами. Данные проходят предобработку: кадрирование для выделения области с номером, изменение размера, нормализация. Для улучшения качества обучения применяется аугментация данных — добавление шумов, повороты, изменение яркости и контраста, чтобы сеть могла адаптироваться к реальным условиям.

**Гиперпараметры сети**:

* **Размер свёрток** определяет область изображения, которую анализирует фильтр за один проход. В выбранной архитектуре используется размер 3x3. Это значение позволяет эффективно выделять локальные признаки, такие как линии, границы и мелкие текстовые элементы, не увеличивая избыточность параметров.
* **Количество фильтров** в каждом слое определяет, сколько различных признаков будет извлечено на каждом уровне. В начальных слоях используется 32 фильтра для захвата базовых признаков (например, текстур и краёв). В средних слоях это число увеличивается до 64, чтобы извлечь более сложные паттерны. В глубоких слоях 128 фильтров используются для формирования высокоуровневых признаков, таких как формы букв или цифр.
* **Шаг свёртки** определяет, насколько фильтр смещается по изображению при обработке. Выбранный шаг равен 1, что позволяет сохранить высокую пространственную детализацию изображения.
* **Размер batch** определяет количество примеров, обрабатываемых моделью за один проход. В данном случае выбрано значение 32, которое позволяет эффективно использовать ресурсы графического процессора, обеспечивая стабильность вычислений и плавную сходимость модели.
* ReLU (Rectified Linear Unit) используется как **функция активации**. Её простота и эффективность обусловлены тем, что она устраняет проблему затухания градиента и ускоряет обучение модели. ReLU позволяет сохранять значимую информацию для последующих слоёв, оставляя только положительные значения.
* Алгоритм Adam (Adaptive Moment Estimation) выбран в качестве **оптимизатора**. Он автоматически корректирует шаг градиента для каждого параметра, что делает его подходящим для задач с шумными данными.
* **Функция потерь:** Cross-Entropy Loss используется для измерения расхождения между предсказанными вероятностями и истинными метками. Эта функция оптимальна для задач классификации, таких как определение номера автобуса. Она минимизирует ошибку, способствуя повышению точности модели.
* **Пуллинг:** MaxPooling (размер 2x2, шаг 2) используется для уменьшения размерности карты признаков, сохраняя наиболее важные из них. Этот подход снижает вычислительные затраты и уменьшает риск переобучения, сохраняя при этом ключевые характеристики текста.

Процесс обучения заключается в подаче размеченных изображений на вход сети, вычислении функции потерь, обновлении весов с помощью градиентного спуска и проверки результатов на валидационной выборке. Обучение продолжается до достижения оптимального значения точности. [9, 10]

## 3.6. Работа сети после завершения обучения

После обучения модель экспортируется в формат, подходящий для мобильного приложения (TorchScript). Это позволяет запускать модель на устройстве без серверной обработки. Интеграция осуществляется с использованием PyTorch Mobile.

Изображения с камеры смартфона поступают в приложение, где проходят предобработку, аналогичную той, что использовалась при обучении. Это обеспечивает соответствие формата данных и корректность работы сети.

Сеть предсказывает номер автобуса на изображении, преобразует результат в текстовую или голосовую форму. Для удобства пользователей, особенно слабовидящих, информация отображается крупным шрифтом или озвучивается через аудиоинтерфейс.

# ЗАКЛЮЧЕНИЕ

В рамках данной научно-исследовательской работы изучены методы компьютерного зрения, проведен анализ существующих решений для распознавания текстов и помощи слабовидящим людям. Рассмотрены особенности работы свёрточных нейронных сетей, инструменты разработки мобильных приложений и методы интеграции нейронных сетей в мобильные устройства.

Проведенное исследование позволяет сформировать теоретическую базу для дальнейшей разработки приложения, которое будет способно эффективно распознавать номера автобусов и предоставлять информацию слабовидящим пользователям.

# СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. **C++ vs Python for Machine Learning: Which Language Should You Choose** [Электронный ресурс] : статья / LinkedIn. - URL: <https://www.linkedin.com/pulse/c-vs-python-machine-learning-which-language-should-you> (дата обращения: 14.12.2024).
2. **Kotlin vs Java** [Электронный ресурс] : статья / GeeksforGeeks. - URL: <https://www.geeksforgeeks.org/kotlin-vs-java> (дата обращения: 14.12.2024).
3. **PyTorch vs TensorFlow vs Keras** [Электронный ресурс]: статья / DataCamp. - URL: <https://www.datacamp.com/tutorial/pytorch-vs-tensorflow-vs-keras> (дата обращения: 14.12.2024).
4. **Getting Started with CameraX** [Электронный ресурс] : документация / Android Developers. - URL: <https://developer.android.com/codelabs/camerax-getting-started?hl=en#2> (дата обращения: 14.12.2024).
5. **Object Recognition** [Электронный ресурс] : статья / ScienceDirect. - URL: <https://www.sciencedirect.com/topics/computer-science/object-recognition> (дата обращения: 18.12.2024).
6. **Optical Character Recognition** [Электронный ресурс] : статья / IBM. - URL: <https://www.ibm.com/think/topics/optical-character-recognition> (дата обращения: 18.12.2024).
7. **Howard, Gugger. Deep Learning for Coders with Fastai and PyTorch: AI Applications Without a PhD [Электронный ресурс] : книга** / Amazon. - URL: <https://www.amazon.com/Deep-Learning-Coders-fastai-PyTorch/dp/1492045527>
8. **Свёрточные нейронные сети (CNN): что это и как они работают** [Электронный ресурс] : статья // Sky.pro. - URL: <https://sky.pro/wiki/python/svertochnye-nejronnye-seti-cnn-chto-eto-i-kak-oni-rabotayut/> (дата обращения: 18.12.2024).
9. **Zhang, Liptom. Dive into Deep Learning** [Электронный ресурс] : книга / Dive Into Deep Learning. - URL: <https://d2l.ai/> (дата обращения: 18.12.2024).
10. **Stevens, Antiga, Viehmann. Deep Learning with PyTorch** [Электронный ресурс] : книга/ ISIP. - URL: <https://isip.piconepress.com/courses/temple/ece_4822/resources/books/Deep-Learning-with-PyTorch.pdf> (дата обращения: 18.12.2024).