|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ **Информатика и системы управления**

КАФЕДРА **Компьютерные системы и сети (ИУ6)**

НАПРАВЛЕНИЕ ПОДГОТОВКИ **09.04.01 Информатика и вычислительная техника**

МАГИСТЕРСКАЯ ПРОГРАММА **09.04.01/12 Интеллектуальный анализ больших**

**данных в системах поддержки принятия решений.**

**Отчет**

**по лабораторной работе № 1-4**

**Название ЛР 1:** исследование возможностей использования генеративных помощников и их применения в платформах промышленной аналитики

**Название ЛР 2:** сбор и предобработка данных для модели машинного обучения

**Название ЛР 3:** построение модели машинного обучения на базе платформ промышленной аналитики

**Название ЛР 4:** доработка модели и оценка ее качества

**Дисциплина:** платформы промышленной аналитики

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент | ИУ6-43М |  |  | В.Н. Зыкин |
|  | (Группа) |  | (Подпись, дата) | (И.О. Фамилия) |
|  |  |  |  |  |
| Преподаватель |  |  |  | М.А.Скворцова |
|  |  |  | (Подпись, дата) | (И.О. Фамилия) |

Москва, 2024

**Лабораторная работа 1**

**Цель:** ознакомиться с возможностями работы чатов/генеративных помощников на базе LLM для применения в исследовательской деятельности студентов и решения профессиональных задач.

Для исследования были выбраны следующие генеративные помощники: СhatGPT-4o, DeepSeek, GigaChat и YandexGPT.

**Задание 1:** необходимо задать вопросы по вашей тематике магистерской диссертации и проанализировать ответы чат-ботов для актуальности и корректности применения данных ответов в ВКРМ (должно быть сформулировано и проверено не менее 5 гипотез по каждому генеративному помощнику).

Были выдвинуты следующие гипотезы:

*Гипотеза 1.* Для оптимизации нейронных сетей под мобильные устройства можно использовать следующие методы: квантование, прунинг, кластеризация, использование специальных фреймворков, использование облегчённых архитектур, сжатие, дистилляция знаний.

Вопрос: какие есть способы оптимизации нейронных сетей под мобильные устройства?

*Гипотеза 2.* Квантование нейронных сетей может быть постобучающим, с учетом обучения, динамическим и гибридным.

Вопрос: какие есть основные способы квантования нейронных сетей?

*Гипотеза 3*. Tensorflow предоставляет больше всего инструментов для оптимизации нейронных сетей под мобильные устройства.

Вопрос: Какой фреймворк предоставляет больше всего инструментов для оптимизации и инференса нейронных сетей на мобильных устройствах?

*Гипотеза 4.* Для обучения моделей семейства YOLO проще всего использовать библиотеку Ultralytics.

Вопрос: Какой фреймворк лучше использовать для обучения модели YOLOv8?

*Гипотеза 5.* Модели сеймества YOLO чаще всего используются для задач детекции.

Вопрос: для каких задач чаще всего используются модели семейства YOLO?

ChatGPT

Ответ ИИ на гипотезу 1.

Оптимизация нейронных сетей под мобильные устройства — важная задача, так как такие устройства ограничены по ресурсам (CPU, GPU, память, батарея).

Чтобы нейросеть работала на мобильном, её обычно упрощают. Самый популярный способ — квантизация: заменяют 32-битные числа на 8-битные. Это сильно ускоряет работу и уменьшает размер модели. Ещё модели «обрезают» — убирают неважные веса (прунинг). Иногда берут большую модель и обучают маленькую повторять её поведение — это называется дистилляция знаний.

Также есть специальные лёгкие архитектуры вроде MobileNet или EfficientNet-Lite — они изначально сделаны для телефонов. Чтобы всё это работало быстрее, используют фреймворки вроде TensorFlow Lite, Core ML или ONNX, которые умеют превращать модель в формат, удобный для смартфона. Иногда модель сжимают ещё больше с помощью матричных разложений или деления весов на общие группы. И, наконец, всё это хорошо работает только при наличии аппаратного ускорения — нейросети используют GPU, NPU или DSP, которые есть в современных телефонах.Начало формы

Анализ: **ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 2.

Квантование нейронных сетей — это техника оптимизации, при которой значения весов и активаций сжимаются с помощью уменьшения разрядности (битности) представления чисел. Это позволяет сократить размер модели и ускорить её выполнение, особенно на устройствах с ограниченными ресурсами (мобильные устройства, edge-устройства и т.п.).

Вот основные способы квантования:

1. Пост-тренировочное квантование (Post-Training Quantization, PTQ)

Квантование применяется к уже обученной модели без дообучения.

Квантование весов – веса переводятся из 32-битных float в 8-битные int (INT8).

Квантование весов и активаций – и веса, и выходы слоёв преобразуются в INT8.

Преимущества:

Простота: не нужно заново обучать модель.

Быстрое применение.

Недостатки:

Возможна потеря точности (особенно у чувствительных моделей).

2. Квантование с дообучением (Quantization-Aware Training, QAT)

Модель обучается или дообучается с учётом будущего квантования.

Во время обучения симулируется эффект квантования.

Позволяет сохранить точность близкую к оригинальной модели.

Используется, когда PTQ даёт слишком большую деградацию качества.

3. Динамическое квантование (Dynamic Quantization)

Квантуются веса до INT8, но активации квантуются на лету во время инференса.

Применимо, например, к трансформерам (BERT, GPT) и RNN.

Хороший компромисс между скоростью и точностью.

Используется в PyTorch, например: torch.quantization.quantize\_dynamic.

4. Гибридное квантование (Hybrid Quantization)

Комбинируются разные типы квантования: часть слоёв квантована, часть остаётся в float32.

Пример: веса в INT8, активации – в float16 или float32.

Используется для балансировки между производительностью и точностью.

Анализ: **ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 3.

На сегодняшний день наиболее полный и мощный фреймворк для оптимизации и инференса нейронных сетей на мобильных устройствах — это:

TensorFlow Lite (TFLite)

Почему именно он:

Поддержка широкого набора оптимизаций:

Квантование (Post-Training, QAT)

Pruning, Clustering

Delegate'ы (GPU, NNAPI, Hexagon, CoreML)

Поддержка Android и iOS

Совместим с моделями TensorFlow / Keras

Простой экспорт и конвертация

Есть TensorFlow Model Optimization Toolkit

Инструменты профилирования и анализа производительности

Лёгкий запуск на устройствах (C++, Java, Swift, Python)Начало формы

Анализ: **ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 4.

Проще всего обучить модель семейства YOLO (например, YOLOv5, YOLOv8) с использованием готовых тулов, где уже всё "упаковано" — загрузка данных, аугментации, обучение, инференс и т.д.

Анализ: **ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 5. Начало формы

Модели семейства YOLO чаще всего используются для задач обнаружения объектов (object detection).

Основные задачи:

* Обнаружение объектов на изображениях и видео (люди, машины, животные и т.д.)
* Системы видеонаблюдения и безопасности
* Аналитика в ритейле (подсчёт людей, товаров на полках)
* Робототехника и автономные системы
* Медицинская визуализация (обнаружение аномалий)
* Анализ изображений в агро- и индустрии (напр., распознавание дефектов)

Анализ: **ответ соответствует гипотезе**.

YandexGpt

Ответ ИИ на гипотезу 1.

Для оптимизации нейронных сетей под мобильные устройства можно использовать следующие методы:

1. Квантование весов и активаций:
   * Уменьшение точности вычислений (например, использование 16-битных чисел с плавающей точкой вместо 32-битных) позволяет сократить объём занимаемой памяти и ускорить вычисления.
2. Сжатие модели:
   * Применение алгоритмов сжатия, таких как Huffman coding или LZW, для уменьшения размера модели без значительной потери точности.
   * Использование методов низкорангового разложения матриц (например, Singular Value Decomposition, SVD) для сжатия весовых матриц.
3. Архитектура модели:
   * Выбор более простых архитектур, таких как MobileNet, SqueezeNet или ShuffleNet, которые специально разработаны для работы на мобильных устройствах.
   * Использование глубинных separable свёрток, которые значительно уменьшают количество параметров и вычислений по сравнению с обычными свёртками.
4. Оптимизация вычислений:
   * Применение библиотек и фреймворков, оптимизированных для мобильных устройств, таких как TensorFlow Lite, PyTorch Mobile или Core ML.
   * Использование аппаратных ускорителей, таких как GPU или DSP, для ускорения вычислений.

Анализ: **ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 2.

Основные методы квантования нейронных сетей:

1. Квантование с фиксированной точкой (Fixed-point quantization) — преобразование весов и активаций из чисел с плавающей точкой в числа с фиксированной точкой для уменьшения размера модели и ускорения вычислений.
2. Квантование с обучением (Quantization-Aware Training, QAT) — учёт ограничений, связанных с квантованием, в процессе обучения модели для минимизации потерь точности.
3. Посттренировочное квантование (Post-Training Quantization, PTQ) — квантование модели после завершения обучения без дополнительного обучения.
4. Динамическое квантование (Dynamic Quantization) — адаптация уровня квантования в зависимости от текущих входных данных для оптимизации баланса между точностью и эффективностью.
5. Симметричное и асимметричное квантование:
   * Симметричное квантование использует одинаковый масштабный коэффициент для всех весов или активаций, упрощая вычисления.
   * Асимметричное квантование позволяет использовать разные масштабные коэффициенты для разных частей модели, что может улучшить точность.
6. Квантование с перекалибровкой (Quantization with Recalibration) — дополнительный этап перекалибровки модели после квантования для корректировки ошибок и улучшения точности.
7. Квантование с учётом ошибок (Error-Aware Quantization) — учёт ошибок, возникающих из-за уменьшения точности представления данных при квантовании, для минимизации их влияния на общую точность модели.

Анализ: **ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 3.

Одним из наиболее популярных фреймворков, предоставляющих широкий набор инструментов для оптимизации и инференса нейронных сетей на мобильных устройствах, является TensorFlow Lite. Он разработан специально для развёртывания моделей машинного обучения на мобильных и встраиваемых устройствах и предлагает различные инструменты и техники для оптимизации моделей, включая квантование, обрезку и сжатие моделей.

Анализ: **ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 4:

Для обучения модели семейства YOLO (You Only Look Once) можно следовать следующим шагам:

1. Выбор версии YOLO.
2. Подготовка данных.
3. Настройка конфигурации.
4. Использование готовых решений.
5. Запуск обучения.
6. Оценка и тестирование.
7. Доработка и оптимизация.
8. Использование предобученных моделей

Следуя этим шагам, вы сможете обучить модель YOLO для задачи обнаружения объектов. Важно помнить, что процесс обучения может потребовать значительных вычислительных ресурсов и времени, особенно при работе с большими наборами данных и сложными моделями.

Анализ: **ответ не соответствует гипотезе.**

Ответ ИИ на гипотезу 5.

Модели семейства YOLO чаще всего используются для задачи обнаружения объектов (Object Detection) — определения местоположения и классов различных объектов на изображениях или видео.

Анализ: **ответ соответствует гипотезе**.

GigaChat

Ответ ИИ на гипотезу 1.

Оптимизация нейронной сети для мобильных устройств — важная задача, особенно учитывая ограниченность ресурсов таких устройств, таких как память, вычислительная мощность и энергопотребление. Вот несколько ключевых способов сделать это:

1. Quantization

Quantization — это процесс уменьшения битности весовых коэффициентов модели (например, с 32 до 8 бит). Это значительно уменьшает размер модели и ускоряет вычисления, поскольку меньшие значения требуют меньше памяти и быстрее обрабатываются.

Существует несколько подходов:Post-training quantization: Применяется после тренировки модели, когда веса преобразуются в низкобитный формат.Quantization-aware training: Модель тренируется сразу с учетом будущей квантизации, что позволяет избежать ухудшения точности.

2. Pruning (Обрезка)

Pruning удаляет избыточные параметры модели, такие как неважные веса или целые слои. Существуют разные методы обрезки:Magnitude-based pruning: Удаляются веса с наименьшими значениями.Structured pruning: Обрезаются целые фильтры или каналы.

После обрезки часто применяется ретренировка для восстановления потерянной точности.

3. Knowledge Distillation (Дистилляция знаний)

Knowledge distillation используется для переноса знаний от большой модели (учителя) к маленькой модели (ученику). Ученик учится предсказывать выход учителя, а также исходные метки, что помогает сохранить точность даже при уменьшении размера модели.

4. Использование легких архитектур

Выбор архитектуры с минимальными требованиями к ресурсам — один из самых простых способов оптимизации. Например:MobileNet: Семейство моделей, разработанных специально для мобильных устройств. Они используют depthwise separable convolutions, что снижает количество операций.EfficientNet: Оптимизированная архитектура, основанная на MobileNet, которая сочетает глубину, ширину и разрешение входного изображения для достижения высокой эффективности.ShuffleNet: Использует pointwise group convolutions и shuffle operations для снижения вычислительной сложности.

5. Компрессия модели

Помимо квантизации и обрезки, существуют другие методы компрессии:Huffman coding: Сжатие весов модели с использованием кодирования Хаффмана.Sparse matrix representation: Представление весов в виде разреженных матриц, где сохраняются только ненулевые элементы.

6. Hardware acceleration

Использование аппаратных ускорителей, таких как GPU или специализированные ASIC/TPU чипы, может значительно ускорить выполнение модели на мобильном устройстве. Некоторые библиотеки, такие как TensorFlow Lite, поддерживают оптимизацию под конкретную архитектуру процессора.

7. Разделение вычислений между устройством и облаком

В некоторых случаях часть вычислений можно перенести на облачные серверы, оставляя на устройстве только обработку данных низкого уровня. Это снижает нагрузку на мобильное устройство, хотя требует постоянного подключения к интернету.

8. Минимизация ввода-вывода

Для ускорения работы модели важно минимизировать операции чтения-записи в память. Например, использование буферов и кеширование промежуточных результатов может существенно улучшить производительность.

9. Удаление ненужных слоев

Иногда можно удалить некоторые части модели, которые не критичны для конечного результата. Например, можно убрать избыточные полносвязные слои или уменьшить количество фильтров в свёрточных слоях.

Заключение

Выбор конкретного метода зависит от требований к производительности, размеру модели и допустимому уровню потери точности. Часто комбинирование нескольких методов даёт наилучший результат.

Анализ: **ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 2.

Основные типы квантования нейронных сетей: Пост-тренировочное квантование — применяется после тренировки модели, веса и активации преобразуются в низкоразрядный формат (например, 8 бит). Быстрота внедрения, сохранение точности. Квантование с учётом обучения — интегрировано в процесс тренировки, учитывается влияние квантования на точность. Более точное, но требует большего времени на тренировку.

Типы квантования: Uniform (равномерное) Non-uniform (неравномерное) Fixed-point (с фиксированной точкой) Dynamic range (динамический диапазон)

Примеры библиотек: TensorFlow Lite, PyTorch, ONNX Runtime.

Анализ: **ответ не соответствует гипотезе.**

Ответ ИИ на гипотезу 3.

Краткий ответ: TensorFlow Lite. Этот фреймворк предлагает широкий набор инструментов для оптимизации и инференса нейронных сетей на мобильных устройствах, включая квантование, сжатие моделей и поддержку различных аппаратных ускорений.

Анализ: **ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 4.

Для обучения модели семейства YOLOv8 рекомендуется использовать YOLOv8 (или YOLOv8s) от компании Roboflow. Это специализированный фреймворк, основанный на PyTorch, который разработан именно для работы с моделями YOLO последней версии. Он включает в себя все необходимые инструменты для подготовки данных, обучения и оценки моделей, а также имеет удобную документацию и сообщество поддержки.

Преимущества YOLOv8:Простая установка и настройка.Поддержка различных версий YOLO (включая v8).Гибкая система конфигурирования и управления обучением.Возможность интеграции с другими инструментами машинного обучения.

Таким образом, YOLOv8 является оптимальным выбором для обучения моделей данного семейства

Анализ: **ответ не соответствует гипотезе.**

Ответ ИИ на гипотезу 5.

Модели семейства YOLO (You Only Look Once) чаще всего используются для решения задач объектной детекции в режиме реального времени. Эти модели предназначены для обнаружения и классификации объектов на изображениях или видео, определяя их местоположение с помощью bounding box'ов (ограничивающих рамок).

Основные области применения YOLO включают:Автоматическое вождение: Детекция автомобилей, пешеходов, дорожных знаков и других объектов на дороге.Безопасность и видеонаблюдение: Определение людей, транспортных средств и подозрительных действий в системах видеонаблюдения.Робототехника: Ориентация роботов в пространстве и взаимодействие с окружающими объектами.Медицина: Анализ медицинских изображений, таких как рентгенограммы или МРТ-снимки, для выявления патологий.Розничная торговля: Подсчет товаров на полках, контроль за запасами и автоматизация кассовых процессов.Производство: Контроль качества продукции, обнаружение дефектов и автоматизация производственных линий.

Благодаря своей быстроте и точности, модели YOLO широко применяются там, где требуется оперативная обработка больших объемов визуальной информации.

Анализ: **ответ соответствует гипотезе**.

DeepSeek

Ответ ИИ на гипотезу 1.

Оптимизация нейронных сетей для мобильных устройств является важной задачей, так как мобильные устройства имеют ограниченные вычислительные ресурсы, память и энергопотребление. Вот основные способы оптимизации нейронных сетей для таких устройств:

### 1. Сжатие модели

* Квантование (Quantization)**:**
* Обрезка (Pruning)**:**
* Дистилляция (Knowledge Distillation)**:**

### 2. Использование легковесных архитектур

### 3. Оптимизация вычислений

* Использование специализированных библиотек
* Использование аппаратного ускорения
* Оптимизация операций

Анализ: **ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 2.

Квантование (quantization) — это процесс уменьшения точности представления чисел в нейронной сети, что позволяет сократить размер модели, ускорить её выполнение и снизить энергопотребление. Это особенно важно для мобильных устройств и embedded-систем, где ресурсы ограничены. Основные способы квантования нейронных сетей включают:

1. Посттренировочное квантование (Post-Training Quantization, PTQ)

Описание:

Модель квантуется после завершения обучения. Веса и активации преобразуются из 32-битных чисел с плавающей запятой (float32) в 8-битные целые числа (int8) или другие форматы.

Преимущества:

Простота реализации, не требует повторного обучения модели.

Недостатки:

Может привести к потере точности, особенно если модель не была оптимизирована для квантования.

Типы:

Только веса: Квантуются только веса, а вычисления выполняются в float32.

Веса и активации: Квантуются как веса, так и активации, что позволяет ускорить выполнение операций.

Примеры:

TensorFlow Lite, PyTorch, ONNX Runtime поддерживают посттренировочное квантование.

2. Квантование с учетом обучения (Quantization Aware Training, QAT)

Описание:

Модель обучается с учетом будущего квантования. В процессе обучения имитируется квантование, что позволяет модели адаптироваться к потере точности.

Преимущества:

Сохраняет высокую точность после квантования.

Недостатки:

Требует дополнительного времени на обучение.

Примеры:

TensorFlow (tf.quantization), PyTorch (torch.quantization).

3. Динамическое квантование (Dynamic Quantization)

Описание:

Веса квантуются заранее, а активации квантуются динамически во время выполнения. Это позволяет сохранить гибкость для разных входных данных.

Преимущества:

Уменьшает размер модели и ускоряет выполнение.

Недостатки:

Менее эффективно, чем полное квантование весов и активаций.

Примеры:

PyTorch поддерживает динамическое квантование.

4. Статическое квантование (Static Quantization)

Описание:

Веса и активации квантуются заранее, на этапе подготовки модели. Для активаций используются заранее вычисленные диапазоны значений.

Преимущества:

Максимальное ускорение и уменьшение размера модели.

Недостатки:

Требует калибровки для определения диапазонов активаций.

Примеры:

TensorFlow Lite, PyTorch.

5. Гибридное квантование (Hybrid Quantization)

Описание:

Часть модели квантуется (например, сверточные слои), а другая часть остается в float32 (например, полносвязные слои).

Преимущества:

Позволяет сохранить точность в критически важных частях модели.

Недостатки:

Меньшая оптимизация по сравнению с полным квантованием.

Примеры:

TensorFlow Lite.

**Анализ: ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 3.

Если говорить о одном фреймворке, который предоставляет наибольшее количество инструментов для оптимизации и инференса нейронных сетей на Android, то это TensorFlow Lite. Он является наиболее полным и зрелым решением для мобильных устройств, особенно для Android, благодаря своей интеграции с экосистемой TensorFlow и поддержке множества аппаратных ускорителей.

**Анализ: ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 4.

Для обучения модели YOLOv8 лучше всего использовать фреймворк Ultralytics, так как YOLOv8 был разработан именно этой командой и максимально оптимизирован для работы с их инструментами. Ultralytics предоставляет простой и мощный API, который значительно упрощает процесс обучения, валидации и развертывания моделей YOLO.

**Анализ: ответ соответствует гипотезе.**

Ответ ИИ на гипотезу 5.

Модели семейства YOLO (You Only Look Once) чаще всего используются для задачи обнаружения объектов (Object Detection). Это одна из ключевых задач компьютерного зрения, которая заключается в локализации и классификации объектов на изображении или видео. YOLO особенно популярен благодаря своей скорости и точности, что делает его идеальным выбором для реального времени.

**Анализ: ответ соответствует гипотезе.**

**Задание 2:** сформировать аннотацию/введение для своей магистерской диссертации с помощью генеративных помощников. Определить какой из них справляется с этой задачей лучше.

Попытка ChatGPT

Выпускная квалификационная работа посвящена разработке программной системы визуального осмотра транспортного средства, предназначенной для упрощения, автоматизации и повышения достоверности процесса фиксации состояния автомобилей. Актуальность темы обусловлена растущими требованиями к цифровизации процессов в автомобильной, страховой и логистической сферах, где критически важным становится оперативный и объективный осмотр транспортных средств с возможностью последующего анализа и документирования.

Основная цель работы — создание инструмента, который позволяет пользователю с помощью мобильного устройства проводить осмотры автомобилей, фиксируя их текущее визуальное состояние. В процессе осмотра осуществляется фото- и видеосъёмка, на основании которой система извлекает ключевую информацию о транспортном средстве. Полученные материалы и данные автоматически сохраняются и используются для формирования отчётов, доступных как пользователям, так и специалистам, принимающим решения на их основе.

Разрабатываемая система ориентирована на практическое применение в таких сферах, как страхование (например, при оформлении страховых случаев или расчёте стоимости полиса), контроль состояния автопарков, приёмка и возврат автомобилей в сервисах аренды и каршеринга, а также в автосервисах и при предпродажной подготовке. Использование современных технологий обработки изображений и автоматического распознавания объектов позволяет сократить время осмотра, снизить влияние человеческого фактора и повысить объективность оценки.

Таким образом, разработанная программная система не только упрощает проведение осмотров, но и создаёт надёжную цифровую основу для хранения, анализа и документирования результатов, что делает её эффективным инструментом в современных бизнес-процессах, связанных с эксплуатацией автотранспорта.

Попытка YandexGpt

Дипломная работа посвящена разработке программной системы для визуального осмотра транспортных средств. Система представляет собой комплексное решение, включающее Android-приложение для проведения осмотров, серверную часть с базой данных и хранилищем MinIO для хранения фото- и видеоматериалов, а также сервис для формирования отчётов в формате PDF.

Основная цель работы — создание эффективного инструмента для визуального контроля состояния автомобилей, который позволит автоматизировать процесс сбора и анализа данных. В Android-приложении пользователь может проводить осмотры, делая снимки и видеозаписи. Во время съёмки задействуются нейронные сети, которые анализируют полученные данные и извлекают важную информацию об автомобиле.

Разработанная система предназначена для использования в автомобильной отрасли, включая сервисы технического обслуживания, страховые компании и другие организации, которым требуется быстрый и точный визуальный контроль транспортных средств. Она позволяет упростить процесс осмотра, сократить время на анализ данных и повысить точность результатов. Это способствует улучшению качества обслуживания клиентов и оптимизации рабочих процессов.

Попытка GigaChat

Выпускная квалификационная работа посвящена созданию программной системы для визуального осмотра транспортных средств. Система разработана для повышения эффективности и автоматизации процесса осмотра автомобилей, обеспечивая удобный интерфейс для пользователей. Основой системы является мобильное приложение на платформе Android, позволяющее проводить осмотр транспортных средств путём фотографирования и видеосъёмки. Во время съемки задействованы нейронные сети, обеспечивающие получение важных данных об автомобиле в реальном времени. Серверная часть системы включает базу данных и хранилище Minio для безопасного хранения фотографий и видеозаписей осмотров. Дополнительно предусмотрен сервис формирования отчётов в формате PDF, содержащих результаты проведённых осмотров. Система направлена на улучшение точности осмотров, сокращение временных затрат и повышение удобства работы пользователей.

**Попытка DeepSeek**

Дипломная работа посвящена разработке программной системы визуального осмотра транспортных средств, предназначенной для автоматизации и упрощения процесса диагностики автомобилей. Система представляет собой комплексное решение, состоящее из мобильного приложения на платформе Android, серверной части и облачного хранилища данных. Основная цель работы — создание инструмента, который позволяет пользователям проводить осмотры транспортных средств с использованием современных технологий компьютерного зрения и нейронных сетей, а также автоматически формировать отчёты о результатах осмотра.

Мобильное приложение является ключевым компонентом системы, предоставляя пользователю интуитивно понятный интерфейс для создания и проведения осмотров. В процессе съёмки транспортного средства задействуются нейронные сети, которые анализируют изображения и видео в реальном времени, извлекая данные о состоянии автомобиля, такие как наличие повреждений, дефектов или других особенностей. Это позволяет значительно ускорить процесс осмотра и повысить его точность за счёт автоматизированного анализа.

Серверная часть системы обеспечивает хранение и обработку данных. Для хранения информации о проведённых осмотрах используется база данных, а для хранения медиафайлов (фотографий и видео) — облачное хранилище MinIO. Это позволяет эффективно управлять большими объёмами данных и обеспечивает их доступность для дальнейшего анализа. Дополнительно реализован сервис формирования PDF-отчётов, который автоматически генерирует документы с результатами осмотра, что упрощает документирование и передачу информации заинтересованным сторонам.

Разработанная система имеет широкий спектр применения. Она может быть использована в автосервисах, страховых компаниях, а также при продаже подержанных автомобилей для объективной оценки их состояния. Внедрение такой системы позволяет минимизировать человеческий фактор, снизить время проведения осмотров и повысить качество диагностики.

В рамках дипломной работы были решены задачи проектирования архитектуры системы, разработки мобильного приложения, интеграции нейронных сетей для анализа изображений, настройки серверной части и облачного хранилища, а также создания сервиса формирования отчётов. Результатом работы стала полнофункциональная система, готовая к использованию в реальных условиях.

Данная работа вносит вклад в развитие технологий автоматизации процессов диагностики транспортных средств и демонстрирует возможности применения нейронных сетей и облачных технологий для решения практических задач в области автомобильной индустрии.

**Вывод к заданию 2:** лучше всех справились СhatGPT и DeepSeek, так как сформировали наиболее объёмный и содержательный текст.

**Задание 3:** описать достоинства и недостатки применения данного инструмента для профессиональной деятельности.

Использование генеративных помощников очень сильно упрощает поиск большого объёма информации. Зачастую они формируют достаточно структурированный и ёмкий ответ. Кроме того, они помогают в решение достаточно типовых проблем, что также сильно сокращает время. Однако если задачи касаются достаточно узких областей, то ИИ могут формировать странные и далёкие от реальности ответы, и приходится очень много времени тратить на написание понятного ИИ промта, что может быть дольше, чем самому решить проблему. В целом генеративные ИИ в несколько раз упрощают и ускоряют работу, однако к его ответы лучше дополнительно анализировать. Кроме того, некоторые модели страдают одной проблемой: склонность оформлять текст в виде огромного количества вложенных списков, что затрудняет использование их ответов в отчётах и работах. Для решения этой проблемы иногда приходится напрямую просить их оформить текст в как можно более сплошном виде.

**Задание 4:** приложить несколько запросов и ответов по использованию генеративных помощников по использованию в конструкторской/технологической части ВКРМ, при формировании списка литературы.

**Запрос**: я собираюсь использовать Kotlin при разработке приложения. Кратко одним абзацем опиши данный инструмент и почему он был выбран.

**ChatGPT**

Kotlin — это современный, статически типизированный язык программирования, разработанный JetBrains и полностью совместимый с Java, что делает его отличным выбором для разработки Android-приложений. Он предлагает более лаконичный, безопасный и выразительный синтаксис по сравнению с Java, снижает количество шаблонного кода и уменьшает вероятность ошибок, связанных с null-значениями благодаря встроенной системе null-безопасности. Kotlin активно поддерживается Google и стал официальным языком разработки под Android, что делает его актуальным, удобным и перспективным инструментом для мобильной разработки.

**YandexGpt**

Kotlin — это современный язык программирования, который работает на платформе Java и хорошо подходит для разработки приложений под Android. Он выбран за свою лаконичность, безопасность типов, поддержку функционального программирования и возможность полной интеграции с существующим Java-кодом, что упрощает процесс разработки и делает код более читаемым и поддерживаемым.

**GigaChat**

Kotlin — это современный статически типизированный язык программирования от JetBrains, который особенно популярен среди Android-разработчиков благодаря своей лаконичности, безопасности и совместимости с Java. Он позволяет писать чистый и понятный код, снижает вероятность ошибок типа NullPointerException благодаря системе типов, поддерживает многоплатформенную разработку и легко интегрируется с существующим Java-кодом. Выбор Kotlin обусловлен его производительностью, поддержкой от Google как основного языка для разработки под Android, а также активным сообществом разработчиков и богатой экосистемой библиотек.

**DeepSeek**

Kotlin — современный, статически типизированный язык программирования, работающий на JVM и полностью совместимый с Java, что делает его идеальным для Android-разработки и кроссплатформенных проектов. Его ключевые преимущества включают лаконичный синтаксис, сокращающий объем кода, встроенную защиту от ошибок (например, null-безопасность), поддержку функционального программирования и расширенные возможности для многопоточности. Kotlin был выбран благодаря официальной поддержке Google для Android, высокой производительности, активному сообществу, а также возможности интеграции с существующим Java-кодом, что ускоряет разработку и упрощает поддержку приложений.

**Запрос**: я использую следующие средства, инструменты, архитектуры разработки клиентской части: Model-View-ViewModel, Kotlin, Jetpack Compose, CameraX, Room, TensorFlow Lite, Gradle, Android Studio. Составь мне по ним список литературы по ГОСТ (для каждого по одному источнику).

**ChatGPT**

1. Госсман, Дж. Model-View-ViewModel (MVVM) [Электронный ресурс] // Википедия. URL: https://ru.wikipedia.org/wiki/Model-View-ViewModel (дата обращения: 02.04.2025).
2. Исакова, С., Жемеров, Д. Kotlin в действии. М.: Питер, 2018. 360 с.
3. Фацио, М. Kotlin and Android Development featuring Jetpack: Build Better, Safer Android Apps. Packt Publishing, 2020. 350 p.
4. Darwin, I. Android. Сборник рецептов: задачи и решения для разработчиков приложений. 2-е изд. СПб.: Питер, 2018. 608 с.
5. Глушенко, С. А., Долженко, А. И. Разработка мобильных приложений: учебное пособие. Ростов-на-Дону: РГЭУ (РИНХ), 2018. 221 с. URL: https://rsue.ru/sveden/files/09.03.03\_Razrabotka\_mobilynyx\_prilogheniy.pdf (дата обращения: 02.04.2025).
6. Kose, U., Deperlioglu, O. Deep Learning for Biomedical Applications. Springer, 2021. 189 p.
7. Loughran, S., Hatcher, E. Ant in Action. 2nd ed. Manning Publications, 2007. 600 p.
8. Черников, В. Разработка мобильных приложений на C# для iOS и Android. М.: ДМК Пресс, 2017. 350 с.

**YandexGpt**

1. Гулаков, С. Д. Android-разработка на Kotlin с нуля до PRO / С. Д. Гулаков. — Москва : Диалектика, 2020. — 464 с.
2. Бертт, Э. Программирование на Kotlin / Э. Бертт. — Москва : Вильямс, 2017. — 512 с.
3. Google. Compose for Android [Электронный ресурс] / Google. — Режим доступа: https://developer.android.com/jetpack/compose (дата обращения: [указать дату]).
4. Google. CameraX API [Электронный ресурс] / Google. — Режим доступа: https://developer.android.com/training/camerax (дата обращения: [указать дату]).
5. Fabián, G. Android Architecture Patterns with Kotlin: Implementing Clean Architecture, MVVM, and Repository Pattern / G. Fabián. — Birmingham : Packt Publishing, 2021. — 356 p.
6. Abhishek, N. TensorFlow 2.x YOLO for Object Detection: Build a real-time object detection system with TensorFlow Lite / N. Abhishek. — Birmingham : Packt Publishing, 2020. — 282 p.
7. Baeldung. Gradle Tutorial [Электронный ресурс] / Baeldung. — Режим доступа: https://www.baeldung.com/gradle (дата обращения: [указать дату]).
8. Google. Android Studio [Электронный ресурс] / Google. — Режим доступа: https://developer.android.com/studio (дата обращения: [указать дату]).

**GigaChat**

1. Model-View-ViewModel (MVVM)Блинов, В. Н. Разработка мобильных приложений на основе паттерна MVVM / В. Н. Блинов // Программирование и информационные технологии. — 2019. — № 4. — С. 20–25.
2. KotlinИванов, А. А. Основы программирования на языке Kotlin / А. А. Иванов. — М.: ДМК Пресс, 2018. — 272 с.
3. Jetpack ComposeAndroid Developers. Guide to Jetpack Compose [Электронный ресурс]. — Режим доступа: https://developer.android.com/jetpack/compose/documentation. — Дата обращения: 10.11.2023.
4. CameraXGoogle Developers. Introducing CameraX, a new Jetpack support library for Android camera development [Электронный ресурс]. — Режим доступа: https://developers.googleblog.com/2019/05/introducing-camerax-jetpack-support.html. — Дата обращения: 15.09.2023.
5. RoomPapinutto, F. Mastering Room: A deep dive into Android's powerful database library / F. Papinutto. — Packt Publishing, 2020. — 360 p.
6. TensorFlow LiteTensorFlow Team. An Introduction to TensorFlow Lite: On-device Machine Learning Framework [Электронный ресурс]. — Режим доступа: https://www.tensorflow.org/lite/guide/get\_started. — Дата обращения: 01.10.2023.
7. GradleHäyrynen, K. Gradle Beyond the Basics: Customizing Next-Generation Builds / K. Häyrynen. — O'Reilly Media, Inc., 2016. — 250 p.
8. Android StudioBurnette, E. Hello, Android: Introducing Google's Mobile Development Platform / E. Burnette. — Pragmatic Bookshelf, 2017. — 320 p.

**DeepSeek**

1. **Model-View-ViewModel (MVVM)**  
   Android Developers. ViewModel overview | App architecture [Электронный ресурс]. – URL: https://developer.android.com/topic/libraries/architecture/viewmodel (дата обращения: 03.04.2025).
2. **Kotlin**  
   JetBrains. Kotlin Language Documentation [Электронный ресурс]. – URL: https://kotlinlang.org/docs/home.html (дата обращения: 03.04.2025).
3. **Jetpack Compose**  
   Smith J. Jetpack Compose: A Complete Guide for Kotlin Android Developers. – 1st ed. – 2024. – 400 p. – ISBN-10: B0D6L39Y7V.
4. **CameraX**  
   Delasign. How to setup a camera in an Android app using Jetpack Compose [Электронный ресурс]. – 2023. – URL: https://www.delasign.com/blog/android-studio-jetpack-compose-kotlin-camerax-composable/ (дата обращения: 03.04.2025).
5. **Room**  
   Gures C. Basic Implementation of Room Database With Repository and ViewModel | Android Jetpack // Medium. – 2020. – URL: https://medium.com/swlh/basic-implementation-of-room-database-with-repository-and-viewmodel-android-jetpack-8945b364d322 (дата обращения: 03.04.2025).
6. **TensorFlow Lite**  
   TensorFlow Developers. TensorFlow Lite Documentation [Электронный ресурс]. – URL: https://www.tensorflow.org/lite (дата обращения: 03.04.2025).
7. **Gradle**  
   Gradle Inc. Gradle User Manual [Электронный ресурс]. – URL: https://docs.gradle.org/current/userguide/userguide.html (дата обращения: 03.04.2025).
8. **Android Studio**  
   Google Developers. Android Studio User Guide [Электронный ресурс]. – URL: https://developer.android.com/studio/intro (дата обращения: 03.04.2025)

**Вывод:** были изучены и исследованы на темах ВКРМ чатыChatGPT, GigaChat, YandexGpt и DeepSeek.