

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ/SYSTEM ANALYSIS,
MANAGEMENT AND PROCESSING OF INFORMATIONDOI: <https://doi.org/10.60797/IRJ.2025.157.38>ИССЛЕДОВАНИЕ ВОЗМОЖНОСТИ СОЗДАНИЯ СРЕДНЕБЮДЖЕТНЫХ ОЧКОВ ДОПОЛНЕННОЙ
РЕАЛЬНОСТИ И АНАЛИЗ ПОЛУЧИВШЕГОСЯ ИЗДЕЛИЯ

Научная статья

Сурин Н.В.^{1,*}, Лунегов И.В.²¹ ORCID : 0000-0001-7501-7964;² ORCID : 0000-0002-8204-3084;^{1,2} Пермский государственный университет, Пермь, Российская Федерация

* Корреспондирующий автор (n.surin[at]psu.ru)

Аннотация

Статья посвящена разработке AR-очков, в которых используются две ключевые нейронные модели: YOLO для обнаружения объектов в реальном времени и GPT для генерации текстовых описаний на основе визуальных данных.

Основная цель статьи — рассказать о способе создания бюджетного AR-устройства, способного не только распознавать объекты, но и предоставлять пользователю детальное описание сцены, что позволит сделать его удобным инструментом для профессионального использования в различных сферах, таких как логистика, медицина, образование и т. д.

Были изучены существующие решения в области AR-устройств, их достоинства и недостатки. Используя анализ существующих AR-продуктов, были разработаны различные компоненты системы. Эксперимент же стал ключевым методом для тестирования гипотез и оптимизации системы.

Основные результаты эксперимента представляют из себя оценку устройства группой лиц в широком диапазоне возрастов.

Ключевые слова: искусственный интеллект, дополненная реальность, 3д печать, нейросеть, большие данные.

STUDY OF THE POSSIBILITY TO CREATE MID-BUDGET AUGMENTED REALITY GLASSES AND ANALYSIS
OF THE RESULTING PRODUCT

Research article

Surin N.V.^{1,*}, Lunegov I.V.²¹ ORCID : 0000-0001-7501-7964;² ORCID : 0000-0002-8204-3084;^{1,2} Perm State University, Perm, Russian Federation

* Corresponding author (n.surin[at]psu.ru)

Abstract

The article is devoted to the development of AR glasses that utilise two key neural models: YOLO for real-time object detection and GPT for generating textual descriptions from visual data.

The main objective of the article is to describe a way to create a budget AR device that can not only recognise objects, but also provide the user with a detailed description of the scene, which will make it a convenient tool for professional use in various fields, such as logistics, medicine, education, etc.

Existing AR device solutions, their advantages and disadvantages were studied. Using analyses of existing AR products, various components of the system were developed. The experiment, on the other hand, was the key method for testing hypotheses and optimising the system.

The main results of the experiment are an evaluation of the device by a group of individuals across a wide range of ages.

Keywords: artificial intelligence, augmented reality, 3d printing, neural network, big data.

Введение

Современные технологии дополненной реальности (AR) уже давно перешагнули порог научной фантастики и стали частью нашей повседневной жизни. Это крайне мощные технологии, которые могут значительно изменить наш опыт восприятия окружающей среды [1, С. 4]. Однако, несмотря на стремительное развитие большинства коммерческих AR-решений, таких как Apple Vision Pro или Microsoft HoloLens, они по-прежнему остаются недоступными для широкого круга пользователей. Высокая стоимость компонентов, сложная архитектура и зависимость от мощного аппаратного обеспечения делают такие устройства уделом специализированных корпораций или энтузиастов с большим бюджетом. В связи с этим возникает актуальная потребность в разработке альтернативных решений, которые смогут обеспечить мощный функционал AR-анализа, сохраняя при этом доступность.

Целью исследования, представленного в этой статье, является создание прототипа AR-очков, который будет способен выполнять распознавание и анализ объектов и людей в реальном времени, причем устройство должно быть максимально похоже на обычные солнцезащитные очки, а его стоимость не должна превышать цену среднебюджетного смартфона.

Самый сложный компонент системы — модуль визуализации информации, который самостоятельно реализовать почти невозможно и единственным на сегодняшний день бюджетным решением, внешне максимально схожим с

солнцезащитными очками, являются носимые мониторы XREAL. Такое устройство максимально схоже с обычными очками, что очень важно, так как это необходимо для создания элегантного прототипа, который не отпугнет пользователей, в отличие от систем, подобных Apple Vision Pro.

Отличительной особенностью проекта является использование локальных алгоритмов обработки данных, что исключает зависимость от облачных сервисов и повышает уровень приватности. Кроме того, в работе предложено нововведение — разделение задач по приоритетам: модель YOLO используется для критичных ко времени операций, а GPT модель — для сложных анализов, не требующих мгновенной реакции.

Практическая значимость исследования заключается в демонстрации возможности создания AR-устройств с ограниченным бюджетом без ущерба для ключевых функций. Такой подход открывает путь к персонализированным AR-приложениям в образовании, медицине, безопасности и других областях, где стоимость и адаптируемость системы играют ключевую роль. Таким образом, данное исследование нацелено на решение актуальной проблемы доступности AR-технологий и задаёт новый стандарт для разработки бюджетных решений, способных конкурировать с промышленными аналогами.

Методы и принципы исследования

Разработка самодельных AR-очков с ограниченным бюджетом требует тщательного подбора компонентов, которые бы обеспечивали высокую производительность, энергоэффективность и компактность. Поэтому разработку необходимо начать с анализа ключевых элементов аппаратной части проекта, включая выбор USB-камер, сравнение одноплатного компьютера Raspberry Pi и видеокарты RTX 2080 Ti, конструирование 3D-корпуса, а также затронуть использование носимых мониторов XREAL Air 2.

Одной из нетривиальных задач при создании AR-очков является выбор USB-камер, которые должны обеспечивать качественный захват видео в реальном времени и при этом быть компактными. С последним, а именно компактностью, есть огромная проблема, так как такие камеры могут рассматриваться как шпионские и их приобретение может повлечь ненужное внимание со стороны правоохранительных органов. Поэтому затронем тему видеокамер поверхностно.

Основными критериями при подборе камер являются:

- разрешение и частота кадров;
- наличие автофокуса и стабилизации изображения;
- работа в условиях слабой освещённости; компактность и вес.

Стоит отметить, что использование нескольких камер открывает дополнительные возможности. Например, стереокамеры могут улучшить точность распознавания объектов за счёт трёхмерного анализа пространства. Другим вариантом использования нескольких камер может быть одновременный захват изображения с разных углов или даже использование удаленных камер, что открывает огромный простор для расширения применения устройства. Однако, такой подход требует более мощного аппаратного обеспечения и тщательной синхронизации данных.

Для тестирования и начальной реализации были выбраны бюджетные USB-камеры, как выяснится в дальнейшем это будет оправданно, так как размер кадра при программном вычислении будет обрезаться примерно до квадрата со стороной 600 пикселей из-за компромисса между точностью и быстродействием.

Другим сложным вопросом при разработке AR-очков становится выбор вычислительного устройства. На рассмотрение представлены два доступных варианта: одноплатный компьютер Raspberry Pi 4/5 и видеокарта Nvidia (конкретика видеокарты не имеет особого значения, важен факт превосходства в вычислениях).

Raspberry Pi 4/5 — это компактный микрокомпьютер оснащён четырёхъядерным процессором и оперативной памятью до 8 ГБ. Его главное преимущество — низкая стоимость и простота интеграции. Однако для задач, требующих высокой вычислительной мощности, таких как обучение моделей и обработка больших данных в реальном времени, его производительность может оказаться недостаточной.

Nvidia RTX 2080 Ti — это видеокарта предоставляет колоссальную вычислительную мощность и приемлемую энергоэффективность, что делает её идеальной для задач машинного обучения и обработки видео. Однако большой размер делают её не подходящей для портативных устройств. Возможно использование удаленных вычислений, но это несет необходимость разработки линии удаленной связи.

В итоге самым оптимальным решением было выбрано использование Raspberry Pi 4/5 для начальных этапов разработки и тестирования, а для более сложных задач, таких как обучение и запуск сложных моделей, использование RTX 2080 Ti.

Ещё одной из важных задач аппаратной части проекта является создание функционального и компактного корпуса для размещения всех необходимых компонентов. Важно не только обеспечить надёжное крепление элементов, но и сохранить минимальный вес и удобство использования. Идеально под эту задачу подходит 3D-печать. Аддитивное производство, или 3D-печать, — процесс создания трехмерных объектов практически любой геометрической формы на основе их цифровых моделей [2, С. 4]. Основные требования к корпусу включают: надёжность, компактность и эргономику.

Для разработки 3D-модели корпуса можно воспользоваться программой Blender, которая предоставляет широкие возможности для проектирования сложных художественных конструкций и достаточно проста в освоении, однако не совсем подходит для инженерного моделирования, но в данном случае это не особо важно, так корпус не имеет подвижных частей и ювелирная точность, следовательно, не нужна. После разработки модели в Blender она должна быть экспортирована в формате поддерживаемом 3д-слайсером (например, STL), после чего необходимо создать в 3D-слайсере G-code и отправить на 3D-печать (например, на 3D-принтер Elegoo Neptune 3 Pro).

В итоге был создан корпус, включающий отсеки для микрокомпьютера, аккумуляторов и USB-камер. Конструкция корпуса предусматривает удобное подключение проводов и быстрый доступ к компонентам для их обслуживания или замены.

Что касается самих очков, то их конструкция не требовала использования 3D-печати, так как в качестве основы были выбраны уже готовые носимые мониторы XREAL Air 2. Использование 3D-печати для создания корпуса микрокомпьютера, камер и аккумуляторов позволило получить компактное и надёжное решение, соответствующее всем требованиям проекта. В то же время, выбор готовых AR-очков упростил процесс разработки, обеспечив высокое качество и эстетичный дизайн. Такой комбинированный подход демонстрирует, как можно эффективно сочетать ручное проектирование и готовые компоненты для создания сложных устройств.

Один из ключевых плюсов XREAL Air 2 — это хорошие дисплеи (1920x1080 пикселей на каждый глаз), которые обеспечивают чёткое и детализированное изображение. Это особенно важно для AR-приложений, где точность отображения информации играет критическую роль. Устройство выполнено в форме, которая визуально почти не отличается от обычных солнцезащитных очков. Это делает их удобными для повседневного использования и снижает нагрузку на пользователя.

Однако, наряду с преимуществами, XREAL Air 2 имеют и некоторые недостатки. Основным недостатком является небольшой угол обзора. Угол обзора составляет всего 46 градусов, что существенно ограничивает поле зрения пользователя. Это может вызывать дискомфорт при длительном использовании, особенно в приложениях, требующих широкого охвата окружающего пространства.

Устройство поддерживает подключение через USB-C, но в данном проекте гораздо проще воспользоваться HDMI-переходником для совместимости с микрокомпьютером Raspberry Pi 4/5. Это решение добавило дополнительный элемент в систему. Чтобы этого избежать, можно использовать вместо Raspberry Pi 4/5 официально поддерживаемые смартфоны Samsung S22 или Samsung S24, однако, это значительно повысит стоимость проекта.

Внешний вид собранного устройства представлен на рис. 1.

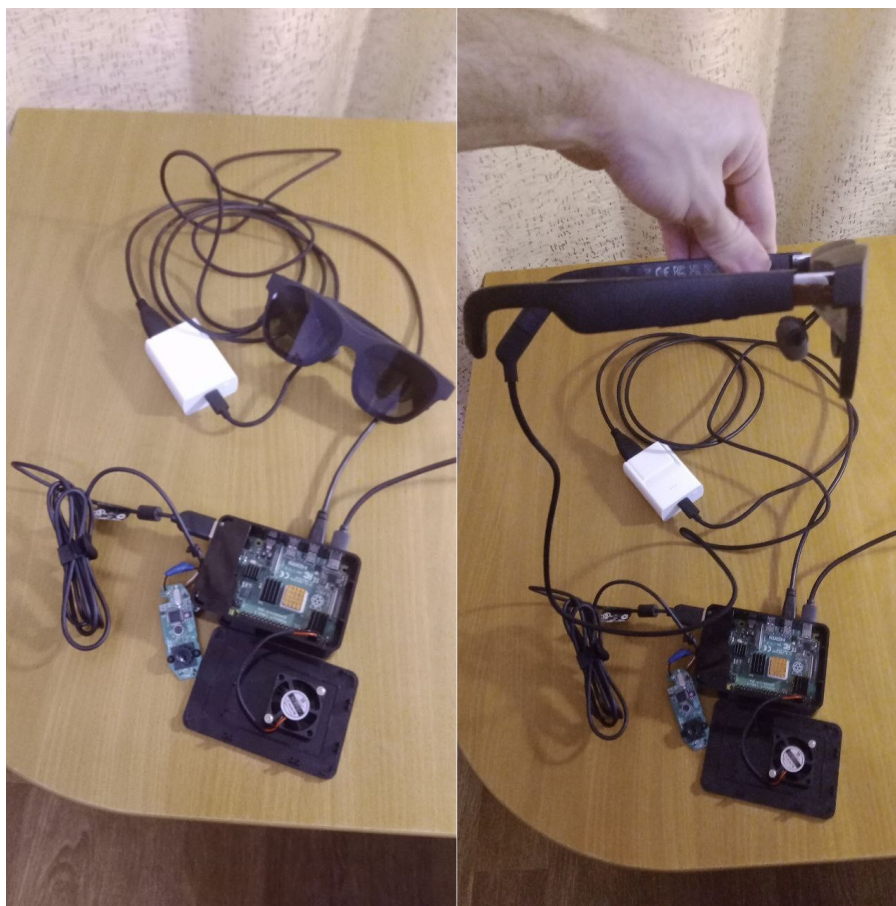


Рисунок 1 - Собранное устройство
DOI: <https://doi.org/10.60797/IRJ.2025.157.38.1>

Создание программного обеспечения для AR-очков является сложнейшим этапом проекта. Задача сводится к разработке конвейера обработки данных, который бы эффективно работал в реальном времени, и состоял из следующих стадий: захват изображения, анализ данных, вывод результата.

Самая трудозатратная стадия конвейера — обработка данных. Чтобы её решить можно воспользоваться методами искусственного интеллекта. Искусственный интеллект — это полностью или частично автономная самоорганизующаяся и самоорганизующаяся компьютерно-аппаратно программная виртуальная (virtual) или киберфизическая (cyber-physical), в том числе биокрибернетическая (bio-cybernetic), система (юнит), неживая в биологическом смысле этого понятия, с соответствующим математическим обеспечением, наделённая / обладающая программно-синтезированными (эмулированными) способностями и возможностями [3, С. 69]. В число вышеупомянутых возможностей и способностей входят антропоморфно-разумные и когнитивные процессы, такие как восприятие,

понимание, анализ, обучение и другие. Простыми словами, ИИ призван программными средствами имитировать естественный интеллект [4, С. 5]. Искусственный интеллект — это достаточно широкая отрасль, которая в свою очередь охватывает и машинное и глубокое обучение. Машинное обучение является областью искусственного интеллекта, которая характеризуется настройкой алгоритма на основе больших данных. Обычно этот алгоритм представляет собой нейронную сеть. Нейросети обычно описываются как совокупность связанных единиц, называемых искусственными нейронами, организованными слоями. Настройка нейросетей возможна при помощи глубокого обучения [5, С. 57]. Самыми популярными на сегодняшний день нейросетевыми моделями для обработки изображений являются YOLO (конкретно будет использована самая новая и легкая YOLO 11n) и GPT (конкретно будет использована самая новая и легкая LLAVA 7b). Под лёгкой версией подразумевается самый компактный вариант модели, то есть с минимальным количеством характеристик; это означает, что для обработки данных требуется меньше всего вычислительных ресурсов, а значит, процесс выполняется быстрее и требует меньше ресурсов, что особенно важно для автономных портативных устройств. В основе работы LLAVA лежит нейросетевая архитектура под названием “Transformer” [6]. В основе же работы YOLO лежит CNN-нейросеть. По сравнению с GPT её устройство гораздо проще, что делает её более легковесной и нетребовательной к ресурсам вычислительных машин.

Для реализации программного обеспечения были выбраны три ключевые библиотеки: Ultralytics, OpenCV и Unsloth. Их выбор обусловлен функциональностью, производительностью и доступностью, что делает их идеальными инструментами для задач компьютерного зрения и машинного обучения.

Ultralytics: Эта библиотека используется для работы с моделями YOLO, которые обеспечивают быстрое и точное обнаружение объектов. YOLO сочетает в себе высокую скорость обработки данных и точность распознавания.

Unsloth: Эта библиотека используется для работы с моделями машинного обучения, такими как GPT, которые выполняют сложный анализ изображений, а именно генерацию текстовых описаний изображений. Unsloth оптимизирует процесс обучения и запуска модели, что особенно важно для ресурсоёмких задач.

OpenCV стала основой для захвата и обработки видео, а также для рендеринга AR-меток. Её простота использования и широкая функциональность, включая работу с камерами, фильтрацию изображений и наложение графики, делает её незаменимой для задач компьютерного зрения.

После захвата изображение передаётся на анализ в зависимости от задачи. YOLO подходит для задач, требующих быстрого обнаружения объектов. Она анализирует изображение, идентифицирует объекты и возвращает их координаты, классы и уровень уверенности. Это критически важно для приложений, где требуется мгновенная реакция, таких как навигация или взаимодействие с окружающей средой. GPT, в свою очередь, больше подходит для сложных задач, таких как анализ сцен или генерация текстовых описаний.

После обработки данных результат интегрируется в видеопоток с использованием библиотеки OpenCV. Это включает: наложение AR-меток (например, рамок вокруг объектов) и вывод текстовых описаний, если требуется. Процесс вывода изображения также оптимизирован для минимизации задержки, что обеспечивает синхронность между действиями пользователя и отображаемой информацией. Одним из простых методов оптимизации вывода является использование только латинского алфавита, потому что именно он оптимизирован в библиотеке OpenCV.

Комплексная схема системы представлена на рис. 2.

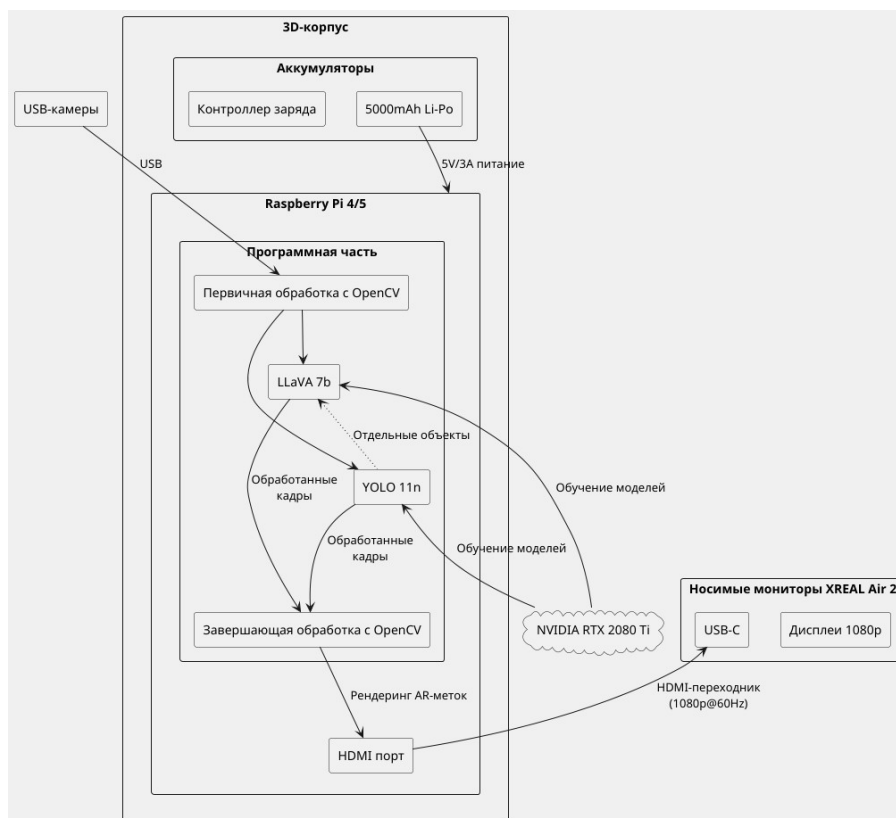


Рисунок 2 - Схема системы
DOI: <https://doi.org/10.60797/IRJ.2025.157.38.2>

Подготовка данных является самым объемным этапом разработки AR-очков, так как качество и объем данных напрямую влияют на эффективность моделей машинного обучения. Можно с уверенностью сказать, что самое сложное в машинном обучении — это данные [7, С. 65]. В данном проекте правильнее всего воспользоваться комбинированным подходом к сбору и обработке данных, включающем получение изображений из общедоступных источников, синтетическую генерацию данных и применение автоматизированной разметки с использованием интеллектуального помощника.

Набор данных, который содержит изображения всех классов распознаваемых объектов, сгруппированных в соответствующие разделы, носит название датасет [8, С. 10]. Для первоначального этапа обучения моделей можно загрузить датасеты из общедоступных источников. Эти датасеты содержат миллионы изображений с объектами различных классов, а также их аннотации, что делает их идеальными для задач обнаружения и классификации. Однако, несмотря на их богатство, общедоступные датасеты не всегда покрывают специфические задачи проекта. Например, для обнаружения объектов в условиях низкой освещенности или для анализа людей в нестандартных позах требуются дополнительные данные.

Для восполнения пробелов в данных логично использовать синтетическую генерацию изображений, например, с помощью Stable Diffusion XL. Эта программа позволяет генерировать изображения на основе текстовых описаний, а также дополнять уже существующие изображения новыми элементами. Но с синтетическими данными всегда есть опасность того, что они не будут репрезентативны реальному миру, где будет функционировать модель [9, С. 8].

Далее данные необходимо разметить, то есть подготовить текстовые описания изображений в специальном формате. Для ускорения процесса разметки данных лучше всего разработать собственный серверный помощник на базе библиотеки FastAPI, интегрированный с Label Studio ML Backend. Этот инструмент автоматически размечает изображения на основе уже обученных моделей, таких как YOLO. Основные преимущества серверного помощника: интеграция с Label Studio, что позволяет пользователям редактировать и подтверждать разметку через удобный интерфейс; автоматическая начальная разметка снижает нагрузку на человека. Однако, не нужно забывать о необходимости проверки автоматически сгенерированных разметок.

Для обучения модели LLaVA через библиотеку Unsloth требуется создать свой набор данных. Этот процесс включает в себя не только получение данных, но и создание структурированных Parquet-файлов, поддерживаемых библиотекой Unsloth. Parquet файл идеально подходит для работы с большими объемами данных благодаря своей эффективности и совместимости с большинством инструментов обработки. Каждый созданный для обучения файл Parquet содержит две ключевые колонки: изображение, преобразованное в байтовый формат для удобства хранения, и текстовое описание, соответствующее этому изображению.

Процесс тестирования и оптимизации системы самодельных AR-очков является завершающим этапом, позволяющим не только выявить скрытые недостатки архитектуры, но и трансформировать сырой прототип в устойчивое к реальным условиям устройство. На начальном этапе тестирования следует сосредоточиться на функциональной проверке всех компонентов: от корректности захвата видеопотока с камер до синхронизации AR-

меток с движениями пользователя. Для этого необходимо разработать серию сценариев, имитирующих типичные варианты использования — от прогулки по квартире до анализа сложных визуальных композиций, для чего проще всего выводить различные изображения или даже видео на дополнительный экран и подбирать такой угол камеры, чтобы она фокусировалась на изображении с монитора. Каждый сценарий необходимо сопровождать детальным протоколом, фиксирующим такие параметры, как задержка обработки кадра, процент ложных срабатываний модели, субъективную оценку удобства интерфейса и так далее.

Тестирование позволяет оценить эффективность модели YOLO 11n и LLaVA 7b в реальных условиях. Необходимо тестировать модели на изображениях различной сложности — от простых сцен с одним объектом до сложных композиций с множеством взаимодействующих элементов.

Основные результаты

YOLO 11n демонстрирует высокую скорость обработки, что критически важно для AR-приложений, где задержка недопустима. Однако, несмотря на её эффективность в обнаружении объектов, модель ограничена в плане понимания контекста.

С другой стороны, LLaVA 7b, которая, в отличие от YOLO 11n, способна анализировать изображения на более глубоком уровне и генерировать текстовые описания, но значительно уступает YOLO 11n в плане производительности, а именно, YOLO 11n выполняет работу по обработке одного кадра за несколько миллисекунд, в то время как LLaVA 7b за несколько секунд. Однако, модель LLaVA 7b показывает впечатляющие результаты в понимании контекста и описании сцен. Например, она может не только определить объекты на изображении, но и объяснить их взаимодействие, что делает её идеальной для задач, требующих детального анализа. Тестирование LLaVA 7b проводится на тех же тестовых экземплярах, что и YOLO 11n, чтобы обеспечить объективность сравнения.

Помимо тестирования отдельных моделей, была проведена оценка их совместной работы в рамках AR-очков. В этом сценарии YOLO 11n используется для быстрого обнаружения объектов, а LLaVA 7b — для их детального анализа и генерации текстовых описаний. Такой подход показывает себя эффективным, но имеет в себе ряд проблем, связанных с синхронизацией данных и нагрузкой на вычислительные ресурсы. Например, в некоторых случаях задержка между обнаружением объекта и получением его описания оказывается слишком большой, что снижает удобство использования системы.

Снимок работы устройства показан на рис. 3.



Рисунок 3 - Пример работы прототипа изделия
DOI: <https://doi.org/10.60797/IRJ.2025.157.38.3>

Для объективности результатов необходимо количественно оценить работу моделей. Изначально в качестве наилучшей оценки искомой величины можно взять среднее арифметическое всех полученных результатов [10, С. 9]. Для сочетания сравнения YOLO и GPT необходимо стандартизировать оценку, и если с YOLO все понятно, то оценка GPT почти всегда не объективна, чтобы минимизировать субъективность можно воспользоваться дискретной оценкой, например порогом в 0,5. Это означает, что, если модель делает предсказание с уверенностью менее 50%, оно считается ошибочным, либо в случае GPT — результат приемлем или нет.

Сравнение моделей YOLO 11n и LLaVA 7b представлено в табл. 1.

Таблица 1 - Сравнение общих ошибок YOLO и GPT

DOI: <https://doi.org/10.60797/IRJ.2025.157.38.4>

Этап тестирования	Процент ошибки YOLO 11n	Процент ошибки LLaVA 7b
Простые сцены (один объект)	20	1
Сложные сцены (композиция из нескольких объектов)	25	5
Сложные сцены (множество отдельных объектов)	25	N/A (LLaVA не поддерживает)
Высокие шумы	45	10
Описание взаимодействий	N/A (YOLO не поддерживает)	5

Как можно понять из табл. 1, LLaVA 7b выигрывает во всем, кроме выделения отдельных частей изображения, так эта нейросеть не предназначена для этого, хотя в теории это возможно реализовать. Причем, если учесть, что это самая слабая версия модели LLaVA, то даже страшно предположить на что способна самая большая версия, однако для её запуска необходим настоящий суперкомпьютер, который никак не вписывается в рамки поставленной цели.

Также, обе нейросети были дообучены на распознавание особенностей человека. Детальное сравнение представлено в табл. 2.

Таблица 2 - Сравнение частных ошибок YOLO и GPT

DOI: <https://doi.org/10.60797/IRJ.2025.157.38.5>

Этап тестирования	Пояснение визуальных признаков	Процент ошибки YOLO 11n	Процент ошибки LLaVA 7b
Сильная усталость	Прищуренные глаза, зевание, сутулость	15	5
Радость	Улыбка	5	1
Гнев	Хмурость	5	1
Оружие в руках	Нож, пистолет	20	5
Пол	Мужчина, женщина	5	1
Признаки заболеваний	Чихание, бледность, потливость	10	5
Признаки обмана	Неестественные выражения, прикосновения к лицу	20	5

Для оценки практической применимости AR-очков в реальных сценариях было проведено тестирование с участием 20 добровольцев в возрасте от 7 до 65 лет. Участники, разделённые на три возрастные группы (7–18, 18–40, 40–65), выполняли задачи, имитирующие повседневные и профессиональные сценарии: поиск объектов в помещении, анализ текстовых инструкций в режиме реального времени, взаимодействие с AR-метками в динамической среде. Особое внимание уделялось субъективному восприятию удобства интерфейса, скорости отклика системы и точности распознавания. Результаты опроса участников представлены в табл. 3.

Таблица 3 - Оценка устройства группами лиц

DOI: <https://doi.org/10.60797/IRJ.2025.157.38.6>

Этап тестирования	Оценка группы 7–18 лет	Оценка группы 18–40 лет	Оценка группы 40–65 лет
Компактность (оценка 1-5)	3	4	3
Четкость изображения (оценка 1-5)	4	5	3
Полезность AR-информации (оценка 1-5)	4	5	2
Точность информации (%)	80	70	50
Время отклика YOLO	1	1	1

Этап тестирования	Оценка группы 7–18 лет	Оценка группы 18–40 лет	Оценка группы 40–65 лет
(сек/запрос)			
Время отклика LLAVA (сек/запрос)	20	15	15
Работа при слабом освещении (оценка 1-5)	2	3	2
Автономность (оценка 1-5)	5	5	5
Приятность материалов (оценка 1-5)	3	4	4
Вес (оценка 1-5)	4	5	3

Заключение

Проведённое исследование демонстрирует, что создание функциональных очков дополненной реальности на базе доступных компонентов — задача не только достижимая, но и открывающая новые горизонты для персонализированных решений в области компьютерного зрения. Разработанный прототип, сочетающий носимые мониторы XREAL, алгоритмы YOLO и мультимодальные модели GPT, подтверждает возможность балансировки между стоимостью и производительностью, бросая вызов коммерческим аналогам с их закрытыми экосистемами и завышенными ценниками. Важнейшим достижением становится доказательство того, что даже на платформе Raspberry Pi 4/5, чья вычислительная мощность сопоставима со слабым смартфоном, можно реализовать интерактивный AR-опыт с минимальной задержкой, что ранее считалось прерогативой специализированного оборудования.

Тестирование AR-очков в группах с возрастным диапазоном от 7 до 65 лет выявило как сильные стороны устройства, так и направления для дальнейшей оптимизации. Наибольшую эффективность система продемонстрировала среди пользователей 18–40 лет: высокая оценка компактности (4/5), четкости изображения (5/5) и полезности AR-информации (5/5) подтверждает соответствие устройства запросам активной аудитории. Однако для младшей (7–18 лет) и старшей (40–65 лет) групп выявлены существенные ограничения.

Перспективы проекта видятся в нескольких направлениях. Во-первых, интеграция нейросетей следующего поколения, способных работать в несколько раз эффективнее текущих моделей, однако небольшой командой крайне непросто создать новый тип нейросети, следственно следует ждать новых open-source решений. Во-вторых, переход на более хорошие дисплеи может устранить проблему неприятных эффектов при длительной носке, но пока что такие устройства слишком дорогие. И наконец, использование более мощного микрокомпьютера (например, NVIDIA JETSON) так же повысит качество AR-опыта, и также значительно повысит стоимость устройства.

В глобальном контексте эта работа бросает вызов индустрии, доказывая, что инновации не обязаны быть дорогими или эксклюзивными. Самодельные AR-очки, собранные за стоимость среднебюджетного смартфона, уже сегодня способны трансформировать, например, образование, помогая студентам визуализировать сложные концепции. Остаётся надеяться, что данный проект станет катализатором для новых исследований, где доступность и открытость будут не исключением, а нормой, ведь будущее дополненной реальности должно принадлежать не только корпорациям, но и каждому, кто готов взять паяльник, написать пару строк кода и увидеть невидимое.

Конфликт интересов

Не указан.

Рецензия

Рудой Е.М., ООО «ГК «Иннотех», Москва Российская Федерация
DOI: <https://doi.org/10.60797/IRJ.2025.157.38.7>

Conflict of Interest

None declared.

Review

Rudoi E.M., Innotech Group LLC, Moscow Russian Federation
DOI: <https://doi.org/10.60797/IRJ.2025.157.38.7>

Список литературы / References

1. Смолин А.А. Системы виртуальной, дополненной и смешанной реальности / А.А. Смолин, Д.Д. Жданов, А.В. Меженин [и др.]. — Санкт-Петербург : Университет ИТМО, 2018. — 59 с.
2. Шкуро А.Е. Технологии и материалы 3D-печати / А.Е. Шкуро, П.С. Кривоногов. — Екатеринбург : Уральский государственный лесотехнический университет, 2017. — 98 с.
3. Морхат П.М. Право и искусственный интеллект / П.М. Морхат. — Москва : Юнити-Дана, 2018. — 544 с.
4. Ватян А.С. Системы искусственного интеллекта / А.С. Ватян, Н.Ф. Гусарова, Н.В. Добренко. — Санкт-Петербург : Университет ИТМО, 2022. — 186 с.

5. Вакуленко С.А. Практический курс по нейронным сетям / С.А. Вакуленко, А.А. Жихарева. — Санкт-Петербург : Университет ИТМО, 2018. — 71 с.
6. Vaswani A. Attention is All you Need / A. Vaswani, N. Shazeer, N. Parmar [et al.] // Neural Information Processing Systems. — 2017. — № 30.
7. Домингос П. Верховный алгоритм. Как машинное обучение изменит наш мир / П. Домингос. — Москва : Манн, 2016. — 366 с.
8. Ярышев С.Н. Технологии глубокого обучения и нейронных сетей в задачах видеоанализа / С.Н. Ярышев, В.А. Рыжова. — Санкт-Петербург : Университет ИТМО, 2022. — 82 с.
9. Кугаевских А.В. Классические методы машинного обучения / А.В. Кугаевских, Д.И. Муромцев, О.В. Кирсанова. — Санкт-Петербург : Университет ИТМО, 2022. — 53 с.
10. Воронова Н.С. Анализ и представление результатов эксперимента / Н.С. Воронова, С.Г. Бежанов, С.А. Воронов [и др.]. — Москва : НИЯУ МИФИ, 2015. — 120 с.

Список литературы на английском языке / References in English

1. Smolin A.A. Sistemi virtualnoi, dopolnennoi i smeshannoi realnosti [Virtual, augmented, and mixed reality systems] / A.A. Smolin, D.D. Zhdanov, A.V. Mezhenin [et al.]. — Saint Petersburg : Universitet ITMO, 2018. — 59 p. [in Russian]
2. Shkuro A.E. Tekhnologii i materialy 3D-pechati [3D printing technologies and materials] / A.E. Shkuro, P.S. Krivonogov. — Yekaterinburg : Ural State Forest Engineering University, 2017. — 98 p. [in Russian]
3. Morkhat P.M. Pravo i iskusstvennii intellekt [Law and artificial intelligence] / P.M. Morkhat. — Moscow : Yuniti-Dana, 2018. — 544 p. [in Russian]
4. Vatan A.S. Sistemi iskusstvennogo intellekta [Artificial intelligence systems] / A.S. Vatan, N.F. Gusarova, N.V. Dobrenko. — Saint Petersburg : Universitet ITMO, 2022. — 186 p. [in Russian]
5. Vakulenko S.A. Prakticheskii kurs po neironnim setyam [Practical course on neural networks] / S.A. Vakulenko, A.A. Zhikhareva. — Saint Petersburg : Universitet ITMO, 2018. — 71 p. [in Russian]
6. Vaswani A. Attention is All you Need / A. Vaswani, N. Shazeer, N. Parmar [et al.] // Neural Information Processing Systems. — 2017. — № 30.
7. Domingos P. Verkhovnii algoritm. Kak mashinnoe obuchenie izmenit nash mir [The supreme algorithm. How machine learning will change our world] / P. Domingos. — Moscow : Mann, 2016. — 366 p. [in Russian]
8. Yarishev S.N. Tekhnologii glubokogo obucheniya i neironnikh setei v zadachakh videoanaliza [Deep learning and neural network technologies in video analysis tasks] / S.N. Yarishev, V.A. Rizhova. — Saint Petersburg : Universitet ITMO, 2022. — 82 p. [in Russian]
9. Kugaevskikh A.V. Klassicheskie metodi mashinnogo obucheniya [Classical machine learning methods] / A.V. Kugaevskikh, D.I. Muromtsev, O.V. Kirsanova. — Saint Petersburg : Universitet ITMO, 2022. — 53 p. [in Russian]
10. Voronova N.S. Analiz i predstavlenie rezultatov eksperimenta [Analysis and presentation of experimental results] / N.S. Voronova, S.G. Bezhanov, S.A. Voronov [et al.]. — Moscow : NIYaU MIFI, 2015. — 120 p. [in Russian]