

К. Бостанбеков¹, Д.Б. Нурсейтов¹, М.Ж. Сакыпбекова^{2*},
Ш.Н. Шекербаева³, И. Шайя⁴

¹ТОО «ҚазМұнайГаз Инжиниринг», г. Астана, Казахстан

²Казахский Национальный университет имени Аль-Фараби, г. Алматы, Казахстан

³Алматинский технологический университет, г. Алматы, Казахстан

⁴Стамбульский технический университет (ITU), г. Стамбул, Турция

*e-mail: sakypbekova.meruyert@gmail.com

ТЕМАТИЧЕСКИЕ ИССЛЕДОВАНИЯ ПО СБОРУ И ОБРАБОТКЕ ДАННЫХ ДЛЯ ОБНАРУЖЕНИЯ ЛЮДЕЙ В ГОРОДСКОМ ТРАНСПОРТЕ

Аннотация

В условиях роста городской мобильности задача точного подсчёта пассажиров становится особенно важной для планирования транспортной инфраструктуры. Целью настоящей статьи — разработка системы автоматического обнаружения и подсчёта людей в городском транспорте с использованием современных методов компьютерного зрения. В работе описаны этапы сбора и аннотирования данных, обучение модели и оценка её эффективности. В рамках исследования были поставлены задачи по сбору и аннотированию видеоданных, обучению модели и оценке её эффективности. Был создан специализированный датасет (4 047 изображений, 8 918 объектов), и модель обучена до достижения значения F1-показателя 0.90. Проведена серия экспериментов с различными алгоритмами трекинга. Результаты подтвердили высокую точность системы и её пригодность для применения в режиме реального времени. Разработанное решение может быть использовано для мониторинга пассажиропотока, оптимизации маршрутов и повышения эффективности управления городским транспортом.

Ключевые слова: подсчёт людей, интеллектуальный транспорт, компьютерное зрение, общественный транспорт, YOLO, датасет.

К. Бостанбеков¹, Д.Б. Нурсейтов¹, М.Ж. Сакыпбекова², Ш.Н. Шекербаева³, И. Шайя⁴

¹«ҚазМұнайГаз Инжиниринг» ЖШС, Астана қ., Қазақстан

²Әл-Фараби атындағы Қазақ Ұлттық университеті, Алматы қ., Қазақстан

³Алматы технологиялық университеті, Алматы қ., Қазақстан

⁴Стамбұл техникалық университеті (ITU), Стамбул қ., Түркия

ҚАЛАЛЫҚ ҚОҒАМДЫҚ КӨЛІКТЕ АДАМДАРДЫ АНЫҚТАУҒА АРНАЛҒАН ДЕРЕКТЕРДІ ЖИНАУ ЖӘНЕ ӨңДЕУ БОЙЫНША ТАҚЫРЫПТЫҚ ЗЕРТТЕУЛЕРІ

Аңдатпа

Қалалық мобильділіктің артуы жағдайында жолаушыларды дәл есептеу міндеті көлік инфрақұрылымын жоспарлау үшін аса маңызды болып отыр. Бұл мақаланың мақсаты — заманауи компьютерлік көру әдістерін пайдалана отырып, қалалық қоғамдық көлікте адамдарды автоматты түрде анықтау және санау жүйесін әзірлеу. Жұмыста деректерді жинау мен аннотациялау кезеңдері, үлгіні оқыту және оның тиімділігін бағалау сипатталған. Арнайы деректер жиынтығы құрылды (4 047 сурет, 8 918 объект), және модель F1 көрсеткіші 0.90 жеткенше оқытылды. Әртүрлі трекинг алгоритмдерімен бірқатар эксперименттер жүргізілді. Нәтижелер жүйенің жоғары дәлдігі мен нақты уақыт режимінде қолдануға жарамдылығын растады. Әзірленген шешім жолаушылар ағынын бақылауға, маршруттарды оңтайландыруға және қалалық көлікті басқару тиімділігін арттыруға пайдалануға болады.

Түйін сөздер: адамдарды санау, интеллектуалды көлік, компьютерлік көру, қоғамдық көлік, YOLO, деректер жиынтығы.

K. Bostanbekov¹, D.B. Nurseitov¹, M.Zh. Sakypbekova², Sh.N. Shekerbayeva³, I.Shayea⁴

¹“KazMunayGas Engineering” LLP, Astana, Kazakhstan

²Al-Farabi Kazakh National University, Almaty, Kazakhstan

³Almaty Technological University, Almaty, Kazakhstan

⁴ Istanbul Technical University (ITU), Istanbul, Turkey

CASE STUDIES OF COLLECTING AND PROCESSING DATA ON DETECTING PEOPLE IN URBAN TRANSPORT

Abstract

In the context of increasing urban mobility, the task of accurate passenger counting becomes critically important for transportation infrastructure planning. The aim of this article is to develop a system for automatic detection and counting of people in urban transport using modern computer vision methods. The paper describes the stages of data collection and annotation, model training, and evaluation of its effectiveness. The aim of this study is to develop a system for automatic detection and counting of people in urban transport using modern computer vision methods. The paper describes the stages of data collection and annotation, model training, and evaluation of its effectiveness. A specialized dataset was created (4,047 images, 8,918 objects), and the model was trained to achieve an F1-score of 0.90. A series of experiments with different tracking algorithms was conducted. The results confirmed the system's high accuracy and real-time applicability. The developed solution can be used for monitoring passenger flow, optimizing routes, and improving the efficiency of urban transport management.

Keywords: people counting, intelligent transportation, computer vision, public transport YOLO, dataset.

Введение

В условиях стремительного расширения урбанизированных территорий и возрастающего спроса на эффективные системы общественного транспорта, задача точной оценки пассажиропотока приобретает ключевое значение в процессе транспортного планирования и управления. Достоверная информация о количестве пассажиров, пользующихся автобусами, трамваями и метрополитеном в различные периоды суток, позволяет транспортным операторам оптимизировать графики движения, рационально распределять ресурсы и повышать качество обслуживания населения. Однако традиционные методы сбора таких данных – включая ручной подсчёт, инфракрасные датчики и напольные коврики, чувствительные к давлению – демонстрируют ограниченную эффективность вследствие проблем масштабируемости, чувствительности к внешним условиям и отсутствия гибкости в адаптации к изменяющейся среде в режиме реального времени. В последние годы значительный прогресс в области компьютерного зрения и технологий глубокого обучения открыл новые перспективы для автоматизированного подсчёта пассажиров в условиях общественного транспорта. Современные методы основаны на использовании RGB-камер, сенсоров глубины и мультисенсорных систем, обеспечивающих высокоточное обнаружение, отслеживание и количественную оценку пассажиропотока. Ряд научных исследований подтвердил эффективность свёрточных нейронных сетей (CNN), а также моделей обнаружения объектов, таких как YOLO, RetinaNet и Faster R-CNN, в решении подобных задач. В ряде недавних исследований рассматривается применение технологий глубокого обучения и обработки изображений для решения задачи подсчёта пассажиров в общественном транспорте. Так, в работе Ren, Fang и Djahel был предложен оптимизированный метод подсчёта людей в реальном времени – YOLO-PC, отличающийся от стандартной реализации YOLO использованием сетки размером 9×9. Это позволяет повысить точность обнаружения за счёт увеличения числа ограничивающих рамок и коэффициентов уверенности. Ключевым нововведением является стратегия выбора границ, обеспечивающая выборочный подсчёт людей, проходящих через заранее определённые зоны, с эффективной фильтрацией нерелевантных объектов. Метод достигает точности 64% mAP при скорости обработки более 40 кадров в секунду. Однако его применение требует точного позиционирования камер и демонстрирует высокую чувствительность к перекрытиям, что создаёт дополнительные сложности при эксплуатации в условиях плотного пассажиропотока [1].

В статье Sun и соавторов представлен метод подсчёта людей в общественном транспорте в реальном времени с применением RGB-D камер и алгоритмов компьютерного зрения, основанных на глубине. Ключевым вкладом работы стало создание открытого набора данных PCDS, включающего более 4500 видеозаписей, что позволило восполнить дефицит эталонных данных для оценки подобных методов. Предложенный алгоритм, основанный на 3D-реконструкции и проекции на опорную плоскость, обеспечивает точность до 92% при скорости 45 FPS, но характеризуется зависимостью от глубинных сенсоров и чувствительностью к углу обзора камеры [2]. В работе Kusuma, T.A.A.H., Usman, K., Saidah рассматривается применение модели YOLOv4 для подсчёта людей в режиме реального времени в условиях общественного транспорта, таких как поезда и самолёты. В исследовании используется метод трансферного обучения: модель обучается на пользовательском датасете, в основном сформированном из Open Image Dataset (управляемого Google), дополненном небольшим количеством вручную собранных реальных тестовых данных. Наилучшее достигнутое значение средней точности (mAP) составило 72,68%, что подтверждает эффективность модели в распознавании пассажиров в различных условиях. Несмотря на использование компьютерного зрения без необходимости в дополнительном оборудовании, сохраняются такие проблемы, как сложность обнаружения пассажиров в масках или головных уборах, а также перекрытия в переполненных пространствах, что указывает на необходимость дальнейшего улучшения модели и увеличения разнообразия обучающих данных [3]. В исследовании Baumann, D., Sommer, M., Schrempf, Y., Sax, E. представлен двухступенчатый подход к подсчёту пассажиров с применением методов глубокого обучения. На первом этапе применяется модель RetinaNet с архитектурой ResNet50, обученная на датасете PCDS, что позволило достичь высокой точности распознавания (mAP – 97,34%). Далее результаты детекции передаются в рекуррентную нейронную сеть, осуществляющую оценку количества вошедших и вышедших пассажиров. Несмотря на значительные вычислительные требования, предложенный метод показывает высокую эффективность и способен заменить традиционные сенсорные решения [4]. Обзорная статья Radovan A. и соавторов представляет всесторонний анализ различных методов подсчёта пассажиров в общественном транспорте [5]. В работе рассматриваются как традиционные методы обработки изображений, так и современные подходы на основе глубокого обучения, с подробным анализом их преимуществ и ограничений в условиях реального применения. Отдельное внимание уделяется интеграции мультисенсорных систем – включая инфракрасные датчики, Wi-Fi и RFID – с целью преодоления таких проблем, как перекрытия, переменные условия освещения и загромождённость сцены. Кроме того, авторы анализируют влияние нормативных требований, таких как GDPR, на внедрение подобных технологий, и дают практические рекомендации по использованию надёжных датасетов (например, PAMELA) и стратегическому размещению камер для повышения точности обнаружения в динамичной среде общественного транспорта.

В целом, рассмотренные исследования демонстрируют быстрое развитие и разнообразие подходов к подсчёту пассажиров в транспорте. Несмотря на ряд преимуществ – включая работу в реальном времени, высокую точность обнаружения, использование качественных датасетов и мультисенсорную интеграцию – сохраняются такие проблемы, как зависимость от сенсоров, точность позиционирования камер и соответствие нормативным требованиям. В связи с этим необходимы дальнейшие исследования и технологические усовершенствования для полноценной реализации данных систем в реальных условиях.

Методология исследования

Сбор датасета для обнаружения людей в городском транспорте. Сбор датасета для обнаружения людей в городском транспорте – это ключевой этап при разработке систем компьютерного зрения, таких как детекция пассажиров. Вот поэтапная научно обоснованная инструкция, как можно собрать и подготовить такой датасет. Один из наиболее надёжных способов сбора данных для обнаружения людей в городском транспорте – это организация

реальной видеосъёмки в целевых зонах. Камеры рекомендуется размещать на входах в автобус, внутри салона, у дверей метро и в других местах с интенсивным пассажиропотоком. Для получения репрезентативного набора данных важно фиксировать видео в различные периоды суток и при разных условиях – как при дневном, так и при вечернем освещении, в часы пик и в периоды низкой загрузки, при различных погодных условиях. При этом необходимо строго учитывать юридические и этические нормы, включая соблюдение требований по защите персональных данных (например, в соответствии с GDPR). Для исключения возможности идентификации личности использовались методы автоматического маскирования и размывания лиц, а также маскирования уникальных признаков. Дополнительно на этапе формирования датасета проводился специальный отбор: в конечный набор включались только те кадры, где лица пассажиров не были чётко различимы. Таким образом, конфиденциальность персональных данных была полностью обеспечена без необходимости дополнительных мер анонимизации.

Предлагаемая система автоматизированного подсчёта пассажиров включает в себя четыре основных модуля. На этапе сбора данных видеозаписи фиксируются изнутри салона автобуса с использованием как стационарных CCTV-камер, так и мобильных устройств (например, смартфонов), установленных в положении, обеспечивающем максимальный охват зоны присутствия пассажиров. На этапе предобработки каждый видеокادر масштабируется до требуемого разрешения с помощью библиотеки OpenCV. При этом дополнительная нормализация данных не применяется, что позволяет сохранить вычислительную эффективность при обработке в реальном времени. Для обнаружения объектов используется модель YOLOv8 (You Only Look Once, версия 8) [6] с предварительно обученными весами. Детектор осуществляет идентификацию каждого человека на изображении, формируя ограничивающие рамки и присваивая объектам соответствующую метку класса "person".

Финальный модуль системы отвечает за подсчёт количества обнаруженных пассажиров, визуализацию результатов в пользовательском интерфейсе в реальном времени, а также генерацию звукового сигнала при превышении заданного порога вместимости транспортного средства. Для обучения модели как показано на рисунке 1, было снято 20 видеозаписей с установленной в салоне автобуса камеры, в транспортных средствах, обслуживающих маршруты городского общественного транспорта в Астане.

На рисунке 1 представлен коллаж, составленный из аннотированных видеокладов, предназначенных для решения задачи детекции людей в условиях городского общественного транспорта. Каждый кадр содержит ограничивающие рамки (bounding boxes) с меткой «Человек», обозначающей объекты класса *person*, распознанные в сцене. Разметка выполнена с использованием инструмента CVAT, что подтверждается унифицированным стилем визуализации и структурой аннотаций [7]. Синие прямоугольники демонстрируют точное позиционирование объектов даже в условиях пониженной освещённости и частичного перекрытия. На изображениях также отображаются метаданные — идентификатор камеры, дата и время съёмки, что обеспечивает возможность синхронизации данных и последующего контекстного анализа. Представленный коллаж иллюстрирует этап ручной аннотации, необходимый для формирования обучающего датасета и последующей валидации моделей компьютерного зрения.

Как показано на рисунке 2, объекты класса «Человек» были размечены с помощью полуавтоматической масочной сегментации в инструменте CVAT, что позволяет точно выделить контуры каждого объекта даже в условиях частичного перекрытия и ограниченной видимости. Этот пример иллюстрирует процесс семантической сегментации пассажиров в городской среде, что является ключевым этапом при формировании датасетов для обучения моделей компьютерного зрения (например, Mask R-CNN и др.). Аннотация таких масок используется для задач распознавания и подсчёта людей в условиях общественного транспорта.

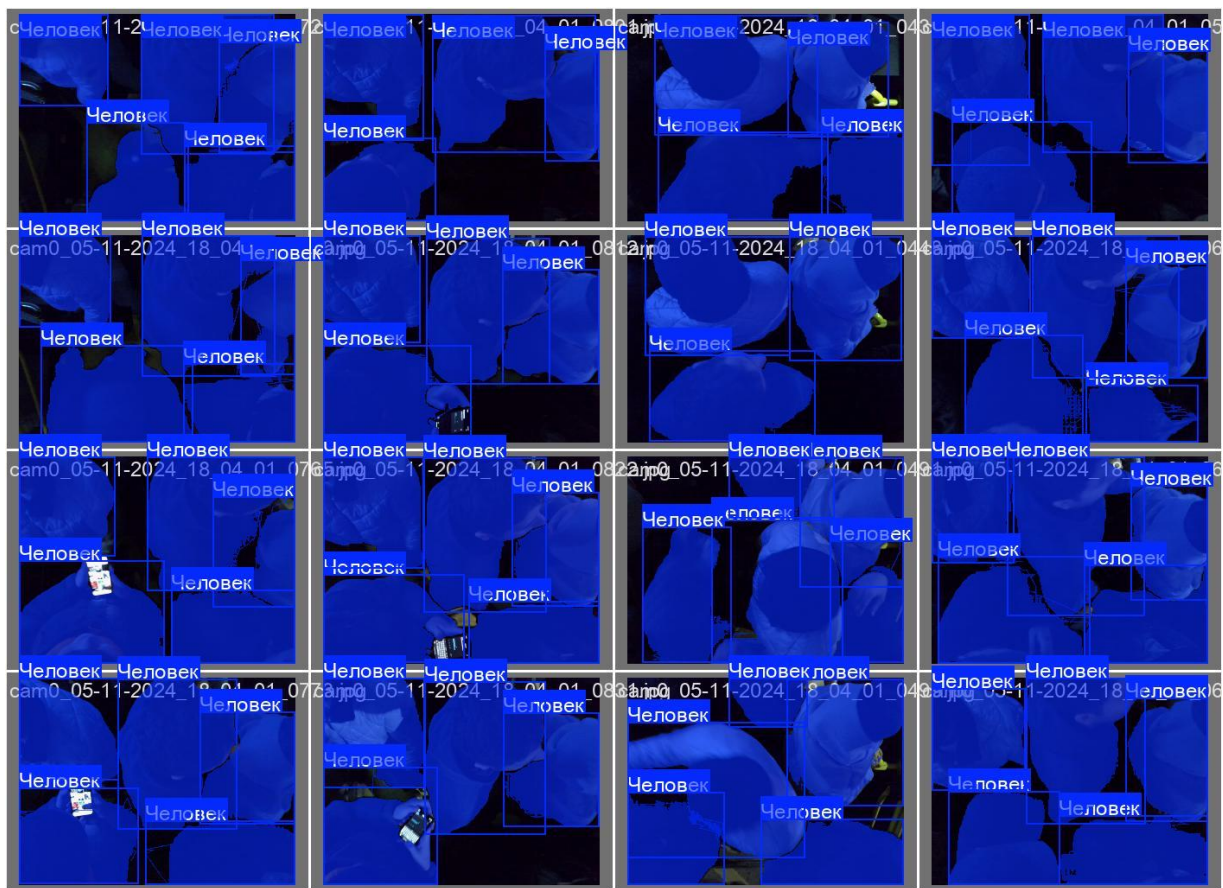


Рисунок 1. Примеры аннотированных видеокадров с детекцией объектов класса «Человек», размеченных в системе CVAT

Для экспериментов использовалась RGB-D камера с лазерной технологией LiDAR Intel RealSense L515, высокой точностью измерения глубины в диапазоне 0.25–9 м и интеграцией IMU. Калибровка выполнялась по методу Zhang [8] с дополнительной температурной компенсацией. Для синхронизации потоков RGB и depth использовался аппаратный триггер с точностью ± 1 мс.

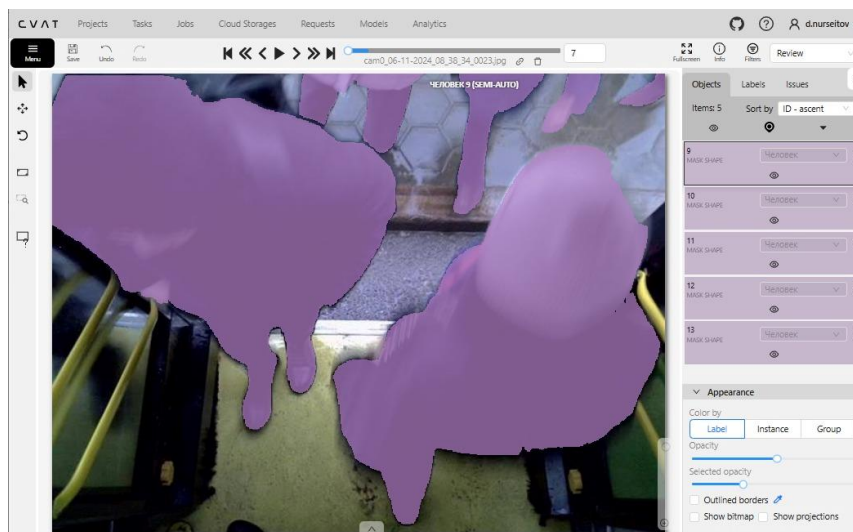


Рисунок 2. Пример аннотирования объектов класса «Человек» с использованием масочной сегментации в CVAT

Модель YOLO (You Only Look Once) стала ключевым решением в области обнаружения объектов в реальном времени и широко применяется в таких сферах, как робототехника, беспилотный транспорт и системы видеонаблюдения. Исследователи Terven, J., Córdova-Esparza, D.-M. и Romero-González, J.-A. провели всесторонний обзор эволюции архитектур YOLO — от оригинальной версии до современных модификаций, включая YOLOv8, YOLO-NAS и модели с использованием трансформеров [9]. В работе рассматриваются применяемые метрики оценки качества, методы постобработки, а также ключевые изменения в архитектуре нейросетей и стратегии обучения, характерные для каждой итерации модели. В завершение авторы обобщают полученные на протяжении развития YOLO уроки и обозначают перспективные направления дальнейших исследований, направленных на совершенствование систем детекции объектов в реальном времени.

Модель YOLOv8, разработанная компанией Ultralytics, представляет собой современную и высокоэффективную архитектуру для решения задач компьютерного зрения, включая детекцию объектов, сегментацию и классификацию. В отличие от предыдущих версий, YOLOv8 реализует anchor-free подход, позволяющий напрямую предсказывать координаты ограничивающих рамок без необходимости использования заранее заданных anchor-box шаблонов. Это упрощает обучение модели и повышает её универсальность в различных сценариях. Архитектура YOLOv8 реализована в нескольких вариантах масштабирования (n, s, m, l, x), что позволяет пользователю выбирать оптимальный баланс между точностью и производительностью. Модель также поддерживает экспорт в форматы ONNX, TensorRT, CoreML и другие, что облегчает интеграцию в прикладные системы. Таким образом, YOLOv8 сочетает в себе модульную архитектуру, высокую скорость обработки и широкую функциональность, что делает её одним из наиболее перспективных решений в области автоматизированной обработки визуальных данных.

Процедура аннотирования – это этап подготовки датасета, на котором каждому объекту на изображениях или видеокадрах присваиваются метки (аннотации), соответствующие его классу и координатам. В задачах компьютерного зрения аннотирование необходимо для обучения моделей, таких как YOLOv8, так как оно позволяет алгоритму «научиться» распознавать и локализовать объекты. Аннотирование видеоданных в рамках данного исследования осуществлялось следующим образом:

1. Автоматическая разметка:

- С последующей ручной проверкой.
- Средняя точность детекции: 72.3% (IoU=0.5)

2. Форматы хранения:

- COCO – это обширный датасет для задач компьютерного зрения, включая 2D-детекцию объектов, сегментацию и captioning [10]. В контексте COCO и вообще компьютерного зрения captioning – это задача автоматической генерации текстового описания изображения.

- KITTI – это специализированный набор данных для разработки и оценки алгоритмов компьютерного зрения в системах автономного вождения, включающий синхронизированные данные с камер, лидаров и GPS/IMU [11].

- Собственный JSON-формат для хранения дополнительных метаданных.

3. Статистика датасета:

- Общий объем: 2.5 часов видео.
- Количество кадров: 27 тыс.
- Размеченных объектов: 2.5 тыс.

Все собранные экспериментальные данные структурируются по следующей иерархии:

/datasets/

- MaixSense_A075V/
- StereoVR_USB/
- RealSense_L515/

Для каждого датасета хранятся:

- Цветовые видео и изображения (.mp4, .png)
- Глубинные карты (.png, .bag)
- Аннотации и метки (.json, .csv)
- Логи экспериментов и конфигурации оборудования

Для обучения и тестирования модели была сформирована структура датасета, включающая в себя 20 папок с видеокадрами, полученными с камер, установленных в салоне городского автобуса. В общей сложности было извлечено 4047 изображений и размечено 8918 объектов класса «человек». Из них 16 папок (3293 изображения, 7322 объекта) были использованы для обучения, а 4 папки (754 изображения, 1596 объектов) — для тестирования.

В таблице 1 представлена соотношение между обучающей и тестовой выборками составило 81% и 19% по количеству изображений соответственно, а по числу аннотированных объектов — 82% и 18%. Такое распределение обеспечивает репрезентативность тестовой выборки при сохранении достаточного объема обучающих данных для корректной настройки модели.

Таблица 1. Распределение изображений и аннотированных объектов в обучающей и тестовой выборках

Категория	Папки	Изображения	Объекты
<i>Train</i>	16	3293(81%)	7322(82%)
<i>Test</i>	4	754(19%)	1596(18%)
<i>Всего</i>	20	4047	8918

Алгоритм подсчета людей реализован в виде класса, который отслеживает перемещение объектов (например, людей) относительно заданной виртуальной линии, представляющей границу входа или выхода. При инициализации задаются координаты этой линии, а также создаются внутренние структуры для хранения истории движения объектов, их последних появлений в кадре и счетчиков пересечений. Основная идея заключается в том, чтобы для каждого обнаруженного объекта сохранять его положение относительно линии, вычисляя знак векторного произведения между направлением линии и вектором от начала линии до центра объекта. Эти значения со временем накапливаются, формируя историю движения объекта. Затем через заданное количество кадров проводится анализ накопленной траектории каждого объекта. Для сглаживания данных используется полиномиальная аппроксимация третьего порядка, позволяющая оценить направление движения объекта. Если аппроксимированная траектория пересекает ноль, это может свидетельствовать о пересечении границы. При этом определяется начальное положение объекта относительно линии, если количество пересечений нечетное, считается, что объект пересек границу. В зависимости от того, с какой стороны он начал движение, алгоритм фиксирует либо вход, либо выход.

Итогом работы алгоритма является счетчик входов и выходов, ID последнего пересекшего границу объекта, а также направление его движения. Алгоритм устойчив к шуму за счет сглаживания траектории, но его эффективность зависит от качества трекинга и корректности ID объектов.

Результаты исследования

Для обучения модели сегментации были аннотированы 4047 изображений, в которых были 8918 объектов. Для теста отобраны 754 изображений, где 1596 объектов. Остальные 3293 изображения использовались для обучения, где присутствовало 7322 объекта. Обучение проводилось на 200 эпохах, показанного на рисунке 3. По графику наблюдаем, что после 150 эпох модель начинает переобучаться. Матрица ошибок показывает точность модели 92%, показанного на рисунке 4.

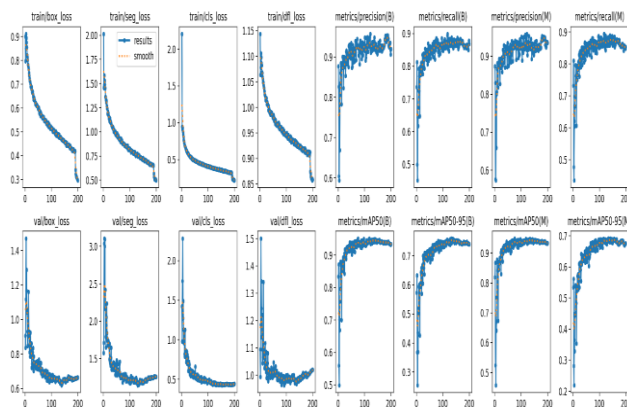


Рисунок 3. Функция потерь

Максимальное значение F1-достигает 0.90, что свидетельствует о высокой сбалансированности между точностью (precision) и полнотой (recall) при заданном пороге в 0.553.

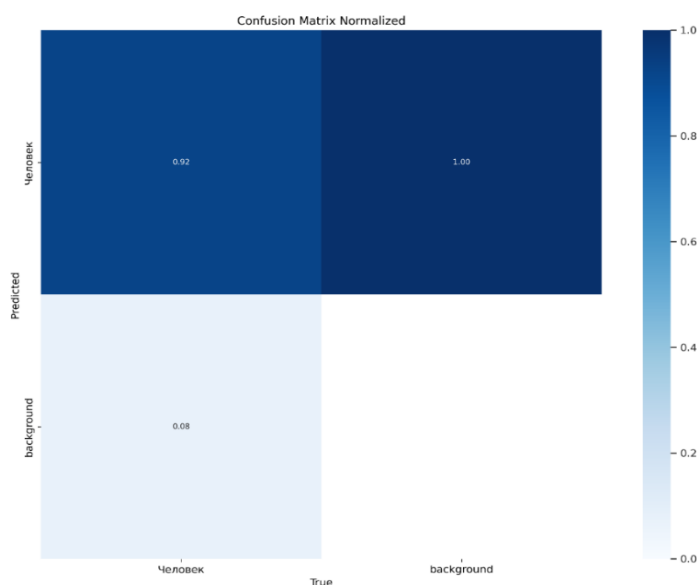


Рисунок 4. Матрица ошибок

Такой результат указывает на высокую эффективность модели в задаче обнаружения объектов класса «человек» в условиях видеонаблюдения. Достигнутое значение F1-показателя является признаком хорошей обобщающей способности модели и её пригодности для практического применения в реальном времени, которая представлена на рисунке 5.

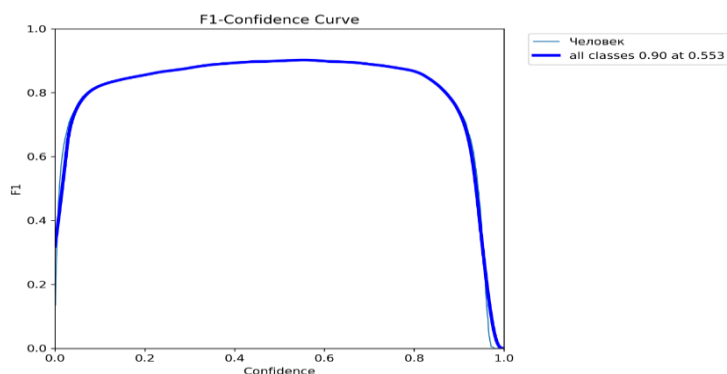


Рисунок 5. Показатель F-I

В таблице 2 представлена серия экспериментов по сравнению алгоритмов трекинга.

Таблица 2. Сравнительный анализ алгоритмов отслеживания объектов

Метод	MOTA	IDF1	Глубина	Задержка	Потребление памяти
<i>SORT [6]</i>	0.58	0.70	Нет	10 мс	500 МБ
<i>DeepSORT [4]</i>	0.65	0.78	Частично	45 мс	1.2 ГБ
<i>FairMOT [7]</i>	0.67	0.80	Нет	35 мс	1.5 ГБ

Результаты показали, что наибольшую точность отслеживания обеспечил алгоритм FairMOT, достигая значений MOTA = 0.67 и IDF1 = 0.80. Тем не менее, с учётом требований к производительности в условиях ограниченных ресурсов предпочтительным оказался DeepSORT, показавший сбалансированные результаты при умеренных вычислительных затратах. Использование глубинной информации (depth data) в рамках DeepSORT способствовало улучшению устойчивости трекинга при частичных перекрытиях объектов.

Эти результаты подчёркивают целесообразность выбора трекеров с учётом специфики целевого применения – в частности, при построении встраиваемых систем мониторинга пассажиропотока на базе компактных вычислительных платформ.

Для верификации эффективности разработанной модели была проведена серия экспериментов на заранее записанных видеоданных, не использованных в обучающем процессе.

Основными критериями оценки выступали качество детекции объектов (метрики F1-score, precision, recall) и производительность модели в кадрах в секунду (FPS) при различных аппаратных конфигурациях, которого представлены на рисунке 6.

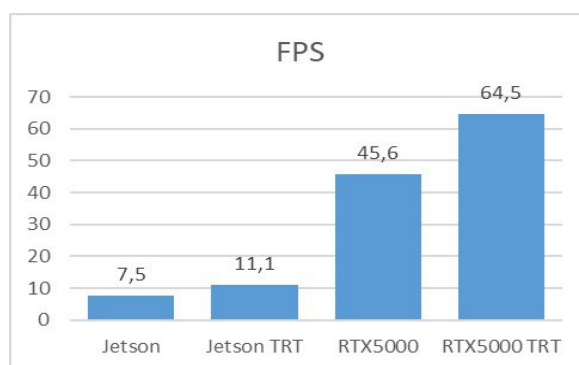


Рисунок 6. Производительность модели

Одновременно была проведена оценка точности на отложенной тестовой выборке. Как показывают полученные метрики, модель достигает значения F1-score = 0.90 при оптимальном пороге уверенности, что подтверждает её высокую надёжность и пригодность к практическому применению в условиях городской транспортной системы.

Дискуссия

Представленные результаты демонстрируют эффективность применения глубинных камер и модели YOLOv8 для автоматизированного подсчёта пассажиров в общественном транспорте. Интеграция depth-данных с алгоритмами трекинга, в частности модифицированным DeepSORT, позволила повысить устойчивость системы к перекрытиям и динамике сцены. Несмотря на более высокую точность FairMOT, DeepSORT показал лучший баланс между скоростью и точностью, что делает его оптимальным для использования на ограниченных по ресурсам платформах. Однако выявлены и слабые стороны подхода: при

плотных скоплениях пассажиров возникают ошибки в ID-трекинге, что приводит к двойному или пропущенному подсчёту. Кроме того, перекрытия и быстрые перемещения пассажиров могут снижать надёжность системы. В дальнейшем планируется исследование алгоритмов коррекции идентификаторов, а также интеграция методов на основе трансформеров для повышения устойчивости.

Формирование собственного датасета обеспечило достаточный объём и качество аннотированных данных, позволив обучить и протестировать систему в приближённых к реальным условиям. Однако точность метода подсчёта зависит от стабильности ID, присваиваемых трекером, что остаётся зоной для оптимизации. В перспективе предполагается использование трансформеров и адаптивных трекинг-методов для повышения надёжности системы в сложных условиях эксплуатации.

Заключение

В ходе проведённого исследования была разработана система автоматизированного обнаружения и подсчёта пассажиров в общественном транспорте с применением современных методов компьютерного зрения. Выбор подхода на основе модели YOLO Pose обусловлен необходимостью обеспечения высокой точности и скорости обработки в условиях плотной сцены и ограниченных вычислительных ресурсов. Использование глубинной информации (depth-данных) позволило повысить устойчивость системы к частичным перекрытиям и сложным ракурсам.

В процессе работы:

1. Разработана трекинг-система, интегрирующая RGB- и глубинные данные, что позволило улучшить точность идентификации и отслеживания пассажиров.
2. Сформирован и аннотирован специализированный датасет, содержащий более 27 000 кадров и 2 500 объектов, адаптированный под задачи обучения нейросетей.
3. Выполнена оптимизация модели, обеспечившая работоспособность в режиме реального времени на различных аппаратных конфигурациях, включая встраиваемые платформы.

Созданная система может служить надёжным инструментом для мониторинга пассажиропотока и анализа загрузки маршрутов, что имеет прикладное значение для повышения эффективности управления городским транспортом и планирования инфраструктуры. Помимо общественного транспорта, подобные системы могут найти применение в торговых центрах, спортивных объектах, аэропортах и других местах массового скопления людей.

Благодарность

Данное исследование выполнено при финансовой поддержке Комитета науки МНВО РК в рамках программно-целевого финансирования BR24992852 «Разработка интеллектуальных моделей и методов цифровой экосистемы Smart City для устойчивого развития города и повышения качества жизни горожан»

Список использованных источников

- [1] Ren, P., Fang, W., & Djahel, S. A novel YOLO-based real-time people counting approach // *International Smart Cities Conference (ISC2)*. – 2017. – P. 1–2. – <https://doi.org/10.1109/ISC2.2017.8090864>
- [2] Sun, Y., Wu, M., & Li, H. Using GPS trajectories to adaptively plan bus lanes // *Applied Sciences*. – 2021. – Vol. 11(3). – P. 1035. – <https://doi.org/10.3390/app11031035>
- [3] Kusuma, T.A.A.H., Usman, K., & Saidah, S. People counting for public transportations using You Only Look Once method // *Jurnal Teknik Informatika (JUTIF)*. – 2021. – <https://api.semanticscholar.org/CorpusID:234045399>

- [4] Baumann, D., Sommer, M., Schrempf, Y., & Sax, E. Use of deep learning methods for people counting in public transport. // 2022 International Conference on Connected Vehicle and Expo (ICCVE). – 2022. – P. 1–6. – <https://doi.org/10.1109/ICCVE52871.2022.9742924>
- [5] Radovan, A., Mršić, L., Đambić, G., & Mihaljević, B. A review of passenger counting in public transport concepts with solution proposal based on image processing and machine learning // Engineering– 2024. – Vol. 5. – P. 3284–3315. – <https://doi.org/10.3390/eng5040172>
- [6] Zhang, Z. A flexible new technique for camera calibration // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2000. – Vol. 22(11). – P. 1330–1334. – <https://doi.org/10.1109/34.888718>
- [7] CVAT Documentation. Getting Started Overview [Электронный ресурс]. – 2025. https://docs.cvat.ai/docs/getting_started/overview/ (дата обращения: 27.06.2025)
- [8] Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS // Machine Learning and Knowledge Extraction. – 2023. – Vol. 5. – P. 1680–1716. – <https://doi.org/10.3390/make5040083>
- [9] Jocher, G., et al. YOLOv8: State-of-the-Art Object Detection Model // Ultralytics. [Online]. Available. – 2023. <https://github.com/ultralytics/ultralytics>
- [10] Lin, T.-Y., et al. Microsoft COCO: Common Objects in Context // European Conference on Computer Vision (ECCV). – 2014. – P. 740–755. – https://doi.org/10.1007/978-3-319-10602-1_48
- [11] Geiger, A., Lenz, P., & Urtasun, R. Are we ready for autonomous driving? The KITTI vision benchmark suite. // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2012. – P. 3354–3361. – <https://doi.org/10.1109/CVPR.2012.6248074>

References

- [1] Ren, P., Fang, W., & Djahel, S. A novel YOLO-based real-time people counting approach // International Smart Cities Conference (ISC2). – 2017. – P. 1–2. – <https://doi.org/10.1109/ISC2.2017.8090864>
- [2] Sun, Y., Wu, M., & Li, H. Using GPS trajectories to adaptively plan bus lanes // Applied Sciences. – 2021. – Vol. 11(3). – P. 1035. – <https://doi.org/10.3390/app11031035>
- [3] Kusuma, T.A.A.H., Usman, K., & Saidah, S. People counting for public transportations using You Only Look Once method // Jurnal Teknik Informatika (JUTIF). – 2021. – <https://api.semanticscholar.org/CorpusID:234045399>
- [4] Baumann, D., Sommer, M., Schrempf, Y., & Sax, E. Use of deep learning methods for people counting in public transport. // 2022 International Conference on Connected Vehicle and Expo (ICCVE). – 2022. – P. 1–6. – <https://doi.org/10.1109/ICCVE52871.2022.9742924>
- [5] Radovan, A., Mršić, L., Đambić, G., & Mihaljević, B. A review of passenger counting in public transport concepts with solution proposal based on image processing and machine learning // Engineering– 2024. – Vol. 5. – P. 3284–3315. – <https://doi.org/10.3390/eng5040172>
- [6] Zhang, Z. A flexible new technique for camera calibration // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2000. – Vol. 22(11). – P. 1330–1334. – <https://doi.org/10.1109/34.888718>
- [7] CVAT Documentation. Getting Started Overview [Electronic resource]. – 2025. https://docs.cvat.ai/docs/getting_started/overview/ (дата обращения: 27.06.2025)
- [8] Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS // Machine Learning and Knowledge Extraction. – 2023. – Vol. 5. – P. 1680–1716. – <https://doi.org/10.3390/make5040083>
- [9] Jocher, G., et al. YOLOv8: State-of-the-Art Object Detection Model // Ultralytics. [Online]. Available. – 2023. <https://github.com/ultralytics/ultralytics>
- [10] Lin, T.-Y., et al. Microsoft COCO: Common Objects in Context // European Conference on Computer Vision (ECCV). – 2014. – P. 740–755. – https://doi.org/10.1007/978-3-319-10602-1_48
- [11] Geiger, A., Lenz, P., & Urtasun, R. Are we ready for autonomous driving? The KITTI vision benchmark suite. // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2012. – P. 3354–3361. – <https://doi.org/10.1109/CVPR.2012.6248074>