

УДК 004.93:656.25

UDC 004.93:656.25

5.2.2. Математические, статистические и инструментальные методы в экономике

5.2.2. Mathematical, statistical and instrumental methods in economics

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ МОНИТОРИНГА БЕЗОПАСНОСТИ ОБЪЕКТОВ ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА НА БАЗЕ YOLOv11

INTELLIGENT COMPUTER VISION SYSTEM FOR MONITORING THE SAFETY OF RAILWAY TRANSPORT FACILITIES BASED ON YOLOv11

Павлов Дмитрий Алексеевич
к.ф.-м.н., доцент
РИНЦ SPIN-код, 8822-5089
dp.logic@gmail.com

Pavlov Dmitriy Alexeevich
Cand.Phys.-Math.Sci., associate professor
RSCI SPIN-code, 8822-5089
dp.logic@gmail.com

Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный аграрный университет имени И.Т. Трублина», г. Краснодар, Россия

Federal State Budgetary Educational Institution of Higher Education «Kuban State Agrarian University named after I.T. Trubilin», Krasnodar, Russia

Жуков Максим Максимович
zhukov3004@yandex.ru
Студент факультета прикладной информатики
Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный аграрный университет имени И.Т. Трублина», г. Краснодар, Россия

Zhukov Maxim Maksimovich
zhukov3004@yandex.ru
Student, Faculty of Applied Computer Science
Federal State Budgetary Educational Institution of Higher Education «Kuban State Agrarian University named after I.T. Trubilin», Krasnodar, Russia

Критическая инфраструктура железнодорожного транспорта (ОАО «РЖД») ежегодно сталкивается с сотнями инцидентов, связанных с несанкционированным доступом на ж/д пути, падением пассажиров и посторонними предметами. Традиционные системы видеонаблюдения с ручным мониторингом характеризуются высокой нагрузкой на оператора и значительным уровнем ложных срабатываний. В статье представлены результаты исследования и разработки интеллектуальной системы компьютерного зрения для автоматического обнаружения угроз в реальном времени. На основе сравнительного анализ архитектур сверточных нейронных сетей выбрана одностадийная модель YOLOv11, обеспечивающая наилучший баланс точности ($mAP@0.5 = 0.90$) и производительности (до 128 FPS). Предложена трехуровневая распределенная архитектура, гарантирующая сквозную задержку обработки 0,1–0,4 с. Разработан и реализован двухэтапный метод детекции падения на железнодорожные пути, сочетающий быстрый кинематический детектор и верификацию по скелетной модели, что позволило достичь полноты детекции 0.95 при низком уровне ложных тревог (<1 события/час/камеру). Оценка экономической эффективности показала, что годовой предотвращенный ущерб составляет 478 млн руб., чистый дисконтированный доход (NPV)

Critical rail infrastructure (Russian Railways) faces hundreds of incidents annually related to unauthorized access to tracks, passenger falls, and foreign objects. Traditional video surveillance systems with manual monitoring are characterized by a high operator workload and a significant rate of false alarms. This article presents the results of the research and development of an intelligent computer vision system for automatic threat detection in real time. Based on a comparative analysis of convolutional neural network architectures, the single-stage YOLOv11 model was selected, providing the best balance of accuracy ($mAP@0.5 = 0.90$) and performance (up to 128 FPS). A three-tier distributed architecture is proposed, guaranteeing an end-to-end processing latency of 0.1–0.4 s. A two-stage method for detecting falls on railway tracks was developed and implemented. It combines a fast kinematic detector and skeletal model verification, achieving a detection completeness of 0.95 with a low false alarm rate (<1 event/hour/camera). A cost-effectiveness assessment showed that the annual prevented damage amounted to 478 million rubles, the net present value (NPV) was 1,615 million rubles, and the payback period was 12.4 months. The developed prototype is ready for pilot implementation at Russian Railways facilities

— 1 615 млн руб., срок окупаемости — 12,4 месяца. Разработанный прототип готов к пилотному внедрению на объектах ОАО «РЖД»

Ключевые слова: КОМПЬЮТЕРНОЕ ЗРЕНИЕ, ДЕТЕКЦИЯ ПАДЕНИЯ, ЖЕЛЕЗНОДОРОЖНЫЙ ТРАНСПОРТ, СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ, БЕЗОПАСНОСТЬ КРИТИЧЕСКОЙ ИНФРАСТРУКТУРЫ, YOLO, BYTETRACK

Keywords: COMPUTER VISION, FALL DETECTION, RAIL TRANSPORT, CONVOLUTIONARY NEURAL NETWORKS, CRITICAL INFRASTRUCTURE SECURITY, YOLO, BYTETRACK

<http://dx.doi.org/10.21515/1990-4665-220-030>

Введение

Обеспечение безопасности на объектах железнодорожного транспорта является стратегической задачей, от решения которой зависят жизнь и здоровье пассажиров, а также бесперебойность грузовых перевозок. Ежегодно на сети ОАО «РЖД» протяженностью более 85 тыс. км регистрируются тысячи инцидентов, связанных с несанкционированным доступом граждан на пути (до 250 наездов на людей в год), падением пассажиров на рельсы (около 180 случаев), хищением, вандализмом и задержкой поездов из-за посторонних предметов. Суммарный прямой экономический ущерб от этих событий превышает 780 млн руб. в год.

Традиционные системы видеонаблюдения, основанные на ручном мониторинге десятков видеопотоков, демонстрируют принципиальные ограничения: концентрация внимания оператора необратимо снижается уже через 20-30 минут работы, а классические детекторы движения (вычитание фона, оптический поток) дают неприемлемо высокий уровень ложных срабатываний при изменении освещенности, осадках и движении теней от подвижного состава [1].

Современные достижения в области глубокого обучения и компьютерного зрения открывают новые возможности для автоматизации обнаружения угроз. Архитектуры сверточных нейронных сетей (CNN), в особенности одностадийные детекторы типа YOLO (You Only Look Once), способны выполнять детекцию объектов в реальном времени с точностью,

<http://ej.kubagro.ru/2026/06/pdf/30.pdf>

сопоставимой с двухстадийными аналогами [2, 3]. Однако существующие коммерческие решения либо не адаптированы к специфике железнодорожных сцен (высокая динамика, специфическая перспектива, работа в полосе отчуждения), либо не обеспечивают требуемого баланса между скоростью, точностью и стоимостью развертывания [1].

Цель настоящего исследования – разработка и исследование системы компьютерного зрения на базе архитектуры YOLOv11, обеспечивающей автоматическое обнаружение нарушителей и опасного поведения пассажиров на объектах железнодорожной инфраструктуры ОАО «РЖД» в реальном времени.

Материалы и методы исследования

Работа выполнена в рамках системного подхода к проектированию информационных систем критического назначения. Используются методы системного анализа, теории машинного обучения (сверточные нейронные сети, трансферное обучение), алгоритмы компьютерного зрения (детекция объектов, трекинг), методы статистической оценки качества моделей, а также методы экономического анализа (методы чистой приведенной стоимости (NPV), дисконтированного срока окупаемости (DPP) и внутренней нормы доходности (IRR)) [4].

Разработана трехуровневая распределенная архитектура (рис. 1).

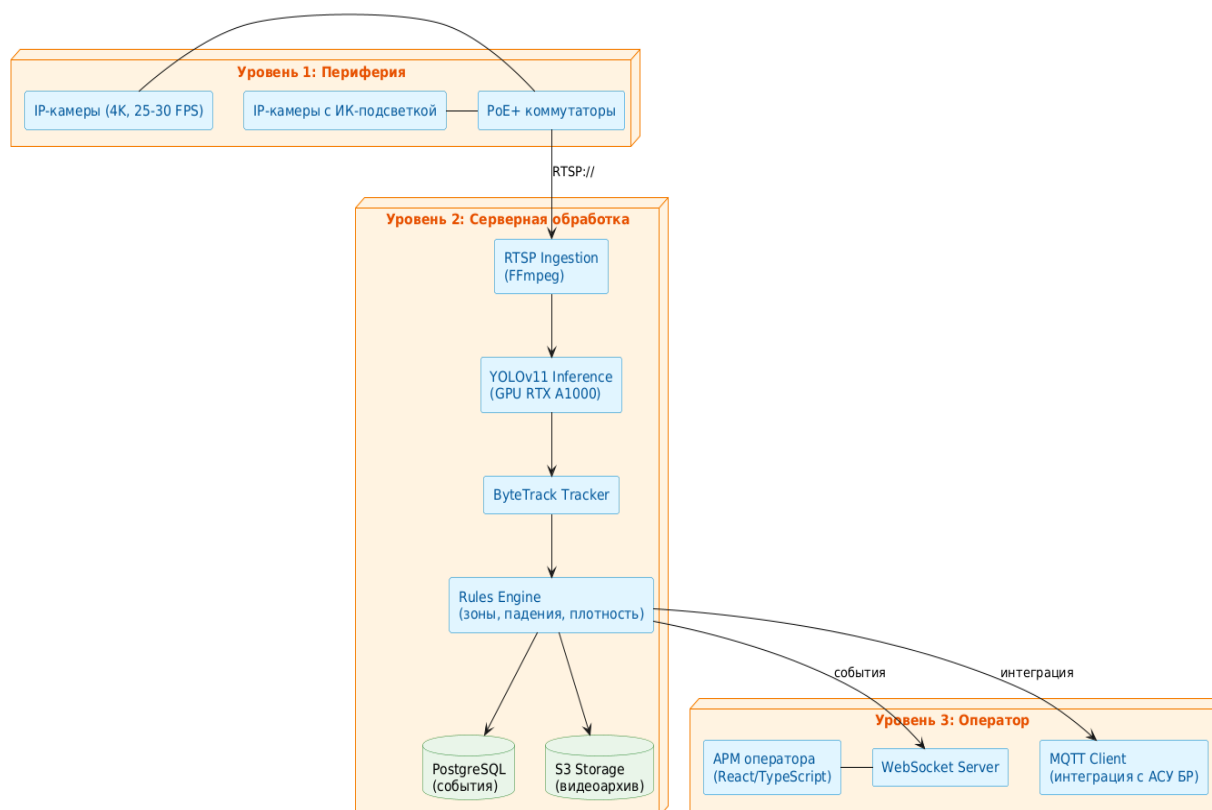


Рисунок 1 - Схема развертывания системы компьютерного зрения

Конвейер обработки включает приём RTSP-потоков (Real Time Streaming Protocol или протокол потока реального времени), декодирование, предобработку (масштабирование до 640×640 , нормализация), вывод результатов работы обученной модели (инференс) YOLOv11, постобработку NMS (Non-Maximum Suppression) [5], трекинг ByteTrack (современный алгоритм множественного отслеживания объектов) [6], анализ сценариев безопасности (зоны, падения, линия безопасности, плотность) и генерацию событий.

Для выбора архитектур сравнение (см. табл. 1) выполнено по метрикам $mAP@0.5:0.95$ на COCO (Common Objects in Context - масштабный набор данных, используемый для обнаружения объектов) и производительности на GPU RTX A1000.

Таблица 1 – Сравнение архитектур детекторов объектов

Архитектура	mAP@0.5:0.95, %	FPS (RTX A1000)	Детекция мелких объектов	Встроенный трекинг
Faster R- CNN	42.0	12	Средняя	Нет
SSD	31.2	46	Низкая	Нет
RetinaNet	39.1	18	Хорошая	Нет
YOLOv11	44.9	128	Отличная	Да (ByteTrack)

Модель YOLOv11 выбрана как базовая архитектура благодаря балансу скорости и точности, встроенному трекингу и возможности экспорта в оптимизированные форматы (TensorRT, ONNX) [3, 5].

Для отслеживания объектов во времени использован алгоритм ByteTrack, сохраняющий низкоуверенные детекции, что критически важно при частичном перекрытии людей в толпе. Для детекции падения разработан двухэтапный комбинированный метод (рис. 2).

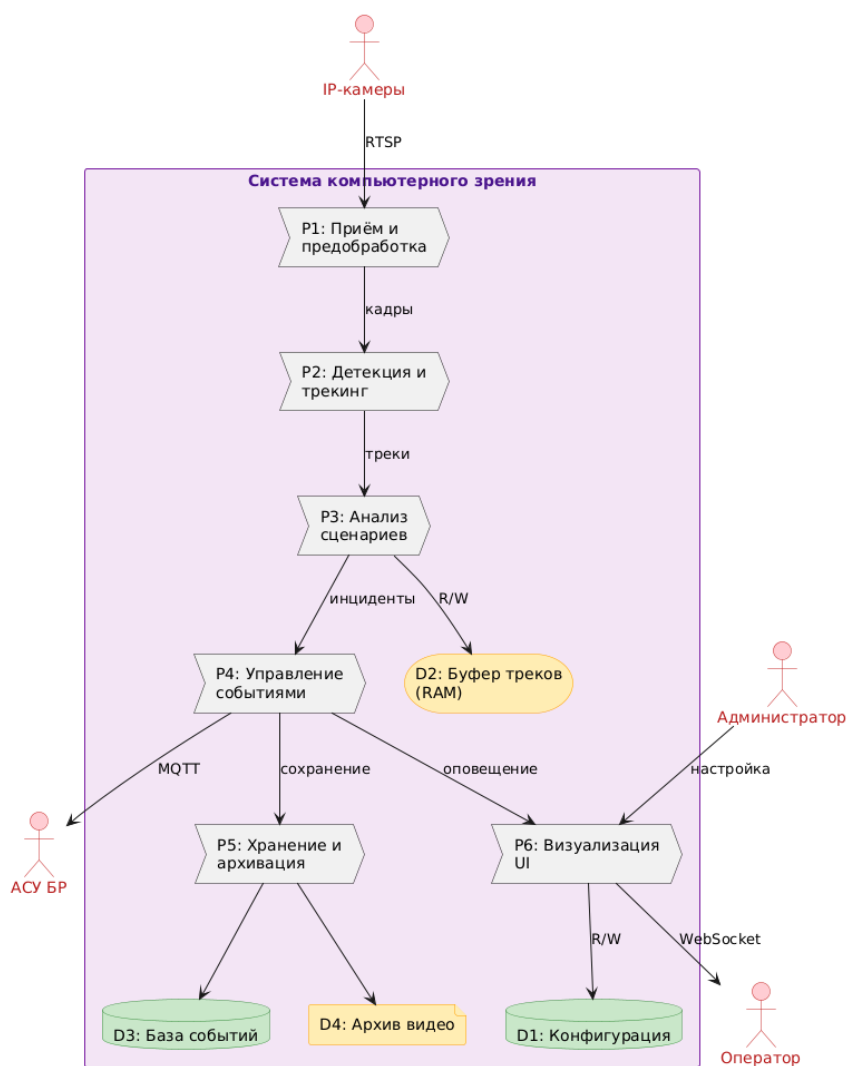


Рисунок 2 - Диаграмма потоков данных системы компьютерного зрения

На рисунке 3 представлена схема двухэтапного детектора падения.

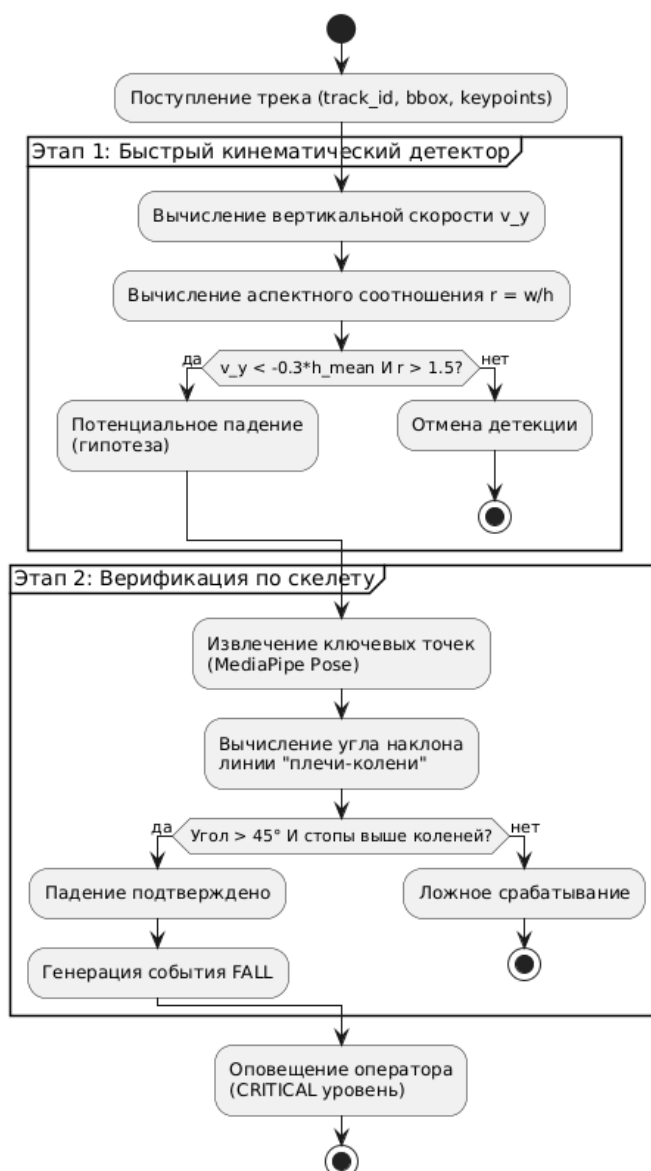


Рисунок 3 – Схема двухэтапного детектора падения

Этап 1 реализует быстрый кинематический детектор, который работает на каждом кадре. В результате чего анализирует вертикальную скорость

$$v_y(t) = y_{bottom}(t) - y_{bottom}(t - 1)$$

где $v_y(t)$ - вертикальная скорость человека в момент времени t (пиксель/кадр). Положительное значение означает движение вниз (например, приседание), отрицательное — движение вверх (характерно для падения, когда тело резко опускается); $y_{bottom}(t)$ - Y-координата нижней границы ограничивающего прямоугольника (bounding box) человека в

текущем кадре t . Обычно отсчитывается от верхнего левого угла изображения (0 – верх, максимальное значение — низ); $y_{bottom}(t-1)$ - Y-координата нижней границы ограничивающего прямоугольника человека в предыдущем кадре ($t-1$).

А также на этапе 1 рассчитывается аспектное соотношение

$$r(t) = w(t)/h(t)$$

где $r(t)$ - отношение ширины к высоте, ограничивающего прямоугольника человека в текущем кадре t . Характеризует, насколько объект «вытянут» по горизонтали или вертикали, $w(t)$ - ширина ограничивающего прямоугольника человека в текущем кадре t , $h(t)$ - высота ограничивающего прямоугольника человека в текущем кадре t .

Падение характеризуется резким отрицательным скачком v_y и увеличением r . Пороговые значения: $v_y < -0.3 \cdot h_{mean}$ и $r > 1.5$

где h_{mean} - средняя высота ограничивающего прямоугольника человека, вычисленная по последним N кадрам (например, за 1 секунду), которая позволяет нормализовать порог вертикальной скорости относительно реального размера человека в кадре.

Пороговый коэффициент для вертикальной скорости в условии $v_y < -0.3 \cdot h_{mean}$ означает, что человек опустился более чем на 30% своей средней высоты за один кадр – резкое вертикальное движение вниз.

Пороговое значение аспектного соотношения. Когда $r > 1.5$, ширина ограничивающего прямоугольника превышает высоту в 1.5 раза, что соответствует горизонтальному (лежащему) положению тела. Для стоящего человека характерно $r < 0.8$ (высота больше ширины).

Этап 2 запускает верификатор на основе скелета и анализирует угол наклона линии «плечи–колени» и взаимное расположение ключевых точек.

Для реализации системы сформирован специализированный датасет, включающий 45 часов видеозаписей с реальных объектов ОАО «РЖД» (платформы, перегоны, стрелочные переводы) при различных условиях освещенности и погоды, синтезированные сцены падений и ночных инцидентов (аугментация: поворот $\pm 15^\circ$, масштабирование 0.8–1.2, изменение яркости $\pm 30\%$, гауссов шум $\sigma=0.02$, размытие $\sigma=1.0$) и публичные наборы COCO и CrowdHuman для предобучения.

Разметка выполнена в формате YOLO с использованием CVAT [7]. CVAT широко используется в проектах машинного обучения и компьютерного зрения (Computer Vision). Его основная задача — это разметка или, как ее еще называют, аннотация данных. Итоговый объем: 1500 изображений, 24 500 аннотаций, классы: «человек», «группа людей», «животное», «посторонний предмет».

Обучение YOLOv11m выполнено на GPU NVIDIA RTX A1000 (48 ГБ VRAM) с параметрами: эпохи = 150, $\text{imgsz} = 640$, $\text{batch} = 16$, $\text{patience} = 20$, оптимизатор = AdamW, $\text{learning rate} = 0.001$. Применено трансферное обучение от предобученной на COCO модели. Используются аугментации: мозаика, микширование, горизонтальное отражение.

Расчет выполнен для типового железнодорожного узла с 32 камерами. Горизонт планирования — 5 лет, норма дисконта — 15%. Используются методы чистой приведенной стоимости (NPV), дисконтированного срока окупаемости (DPP) и внутренней нормы доходности (IRR).

Результаты и обсуждение

Обученная модель YOLOv11 достигла метрик, представленных в табл. 2. Средневзвешенная точность ($\text{Precision} = 0,87$) и полнота ($\text{Recall} = 0,88$) удовлетворяют требованиям технического задания ($\text{Precision} \geq 0.85$, $\text{Recall} \geq 0.88$). Наилучшие результаты достигнуты для класса «человек» ($\text{Precision} 0,89$, $\text{Recall} 0,92$, $\text{mAP}@0.5} 0,94$), что критически важно для

основной задачи. Более низкие показатели для класса «посторонний предмет» ($mAP@0.5 = 0,77$) обусловлены высокой вариативностью формы и текстуры предметов, что требует дальнейшего сбора данных.

Таблица 2 – Итоговые метрики обученной модели YOLOv11

Класс	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Человек (person)	0.89	0.92	0.94	0.67
Группа людей (crowd)	0.84	0.79	0.85	0.52
Животное (animal)	0.91	0.85	0.89	0.58
Посторонний предмет	0.76	0.71	0.77	0.43
Средневзвешенная	0.87	0.88	0.90	0.61

Для снижения задержки прямого прохода обученной модели (инференса) выполнено квантование весов до INT8 и экспорт в TensorRT. Результаты представлены в табл. 3. Квантование в INT8 позволило сократить время инференса с 18.2 мс до 6.8 мс (в 2.7 раза) при незначительном снижении $mAP@0.5$ (с 0.90 до 0.89).

Таблица 3 – Сравнение производительности форматов модели

Формат	Размер, МБ	Инференс, мс/кадр	mAP@0.5
PyTorch (.pt)	49	18.2	0.90
ONNX	49	15.8	0.90
TensorRT (FP16)	24	9.4	0.90
TensorRT (INT8)	12	6.8	0.89

Анализ временных характеристик конвейера (табл. 4) показал, что общая сквозная задержка от момента возникновения инцидента до появления сигнала на АРМ оператора составляет 102–378 мс, что значительно ниже требуемых 2 с и позволяет квалифицировать систему как работающую в реальном времени.

Таблица 4 – Бюджет задержки конвейера обработки

Этап	Типичная задержка, мс	Максимальная задержка, мс
Передача RTSP	50–100	200
Декодирование кадра	5–10	15
YOLOv11 инференс (640×640)	12–18	25
NMS + ByteTrack	3–5	8
Анализ сценариев (4 параллельных)	10–15	20
Генерация события → очередь	2–5	10
WebSocket → АРМ оператора	20–50	100
Итого	102–203	378

Разработанный двухэтапный метод детекции падения протестирован на тестовой выборке из 120 видеозаписей реальных и смоделированных падений (50 сценариев). Результаты: полнота (Recall) - 0.95, точность (Precision) = 0.93, частота ложных тревог - 0.8 события/час/камеру. Сравнение с существующими подходами: метод на основе только кинематических параметров дает полноту 0.88 при частоте ложных тревог 2.1 события/час; метод на основе только анализа скелета требует значительных вычислительных ресурсов (≈ 45 мс/кадр). Предложенный комбинированный подход обеспечивает оптимальный компромисс.

Прогнозируемое снижение основных типов инцидентов после внедрения системы представлено в табл. 5. Прогноз снижения инцидентов выполнен экспертным методом на основе анализа технических характеристик систем компьютерного зрения, представленных в научной литературе [8, 9]. Согласно данным [8], современные модели детекции достигают точности 95.68% при обнаружении человека в реальных сценах. Исследование [9] демонстрирует 95% точности обнаружения вторжений с нулевым уровнем ложных срабатываний при пороге уверенности 0.65, а

также снижение затрат на развертывание на 75%. Достижимый уровень точности и эффективности позволяет прогнозировать существенное (до 60-70%) сокращение целевых типов инцидентов при пилотном внедрении разработанной системы. Годовой предотвращенный ущерб составляет 478,1 млн руб. Экономия от автоматизации видеоконтроля (снижение нагрузки на 3 операторов в смену на 50%) – 10,8 млн руб./год. Годовые эксплуатационные затраты (ОРЕХ) – 4,64 млн руб. Годовой чистый экономический эффект – 484,26 млн руб.

Таблица 5 – Прогноз снижения инцидентов и предотвращенный ущерб

Тип инцидента	Базовый ущерб, млн руб./год	Снижение, %	Предотвращ. ущерб, млн руб./год
Наезд на людей на путях	375	60	225,0
Травмирование на платформах	144	70	100,8
Хищение грузов и имущества	60	50	30,0
Повреждение инфраструктуры	90	55	49,5
Задержки поездов	112	65	72,8
ИТОГО	781	—	478,1

Расчет эффективности инвестиций (табл. 6) показал: чистый дисконтированный доход (NPV) за 5 лет – 1 615,45 млн руб., дисконтированный срок окупаемости (DPP) – 12,4 месяца, внутренняя норма доходности (IRR) – 182%. Все показатели свидетельствуют о высокой инвестиционной привлекательности проекта.

Таблица 6 – Денежные потоки и расчет NPV (тыс. руб.)

Показатель	Год 0	Год 1	Год 2	Год 3	Год 4	Год 5
Капитальные затраты (CAPEX)	-6 770	0	0	0	0	0
Эксплуатационные затраты (OPEX)	0	-4 643	-4 877	-5 121	-5 377	-5 646
Предотвращенный ущерб	0	478 100	478 100	478 100	478 100	478 100
Экономия автоматизации	0	10 800	10 800	10 800	10 800	10 800
Чистый денежный поток (CF)	-6 770	484 257	484 023	483 779	483 523	483 254
Коэффициент дисконтирования	1	0,87	0,756	0,658	0,572	0,497
Дисконтированный денежный поток	-6 770	421 304	365 923	318 194	276 573	240 227
Накопленный (DCF)	-6 770	414 534	780 457	1 098 651	1 375 224	1 615 451

Разработанный прототип включает: бэкенд-сервис на Python (FastAPI, асинхронная обработка, RabbitMQ, PostgreSQL), обученную модель YOLOv11 (экспорт в TensorRT), веб-интерфейс оператора на React/TypeScript. Обеспечена интеграция с АСУ БР ОАО «РЖД» через MQTT. Система контейнеризована с использованием Docker и может быть развернута как на одном сервере (до 16 камер), так и в кластере Kubernetes (до 128 камер).

Заключение

В результате проведенного исследования решена актуальная научно-практическая задача – разработана и исследована интеллектуальная система компьютерного зрения для мониторинга безопасности объектов железнодорожного транспорта.

Основные научные результаты:

1. Выполнен сравнительный анализ архитектур CNN для задач детекции на железнодорожных объектах, обоснован выбор YOLOv11 как обеспечивающей наилучший баланс точности ($mAP@0.5 = 0.90$) и производительности (128 FPS на RTX A1000).

2. Предложена трехуровневая распределенная архитектура системы (Edge — Compute — Operator), гарантирующая сквозную задержку обработки 0,1–0,4 с и горизонтальное масштабирование до 128 камер.

3. Разработан и реализован двухэтапный метод детекции падения на железнодорожные пути, сочетающий быстрый кинематический детектор и верификацию по скелетной модели, что позволило достичь полноты 0.95 при частоте ложных тревог менее 1 события/час/камеру.

4. Экспериментально подтверждено, что квантование модели до INT8 и экспорт в TensorRT сокращают время инференса в 2.7 раза (с 18.2 до 6.8 мс) при незначительной потере точности ($mAP@0.5$: 0.90→0.89).

Экономическая эффективность подтверждена расчетами: годовой предотвращенный ущерб – 478 млн руб., NPV – 1 615 млн руб., DPP – 12,4 месяца, IRR – 182%;

Социальный эффект от внедрения – снижение травматизма и гибели людей на железнодорожной инфраструктуре.

Литература

1. Chen P., Zhang Z., Huang Y. et al. Railway obstacle intrusion warning mechanism integrating YOLO-based detection and risk assessment // Journal of Industrial Information Integration. – 2024. – Vol. 38. – P. 100571.

2. Jocher G., Chaurasia A., Qiu J. YOLO by Ultralytics. Version 11.0. – 2024. – URL: <https://github.com/ultralytics/ultralytics>

3. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // CVPR. – 2016. – P. 779-788.

4. Медведева Е.В., Перевощикова А.А. Исследование нейросетевых алгоритмов распознавания объектов железнодорожной инфраструктуры на видеокдрах // Компьютерная оптика. – 2025. – Т. 49, № 3. – С. 443-452.

5. Zhang Y., Sun P., Jiang Y. et al. ByteTrack: Multi-Object Tracking by Associating Every Detection Box // ECCV. – 2022. – P. 1-18.

6. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. – М.: ДМК Пресс, 2018. – 652 с.
7. <https://www.cvat.ai/>
8. Chang C.C., Huang K.H., Lau T.K., Huang C.F., Wang C.H. Using deep learning model integration to build a smart railway traffic safety monitoring system // Scientific Reports. – 2025. – Vol. 15. – Article 4224. – DOI: 10.1038/s41598-025-88830-7.
9. NETRA: Smart Railway Obstruction Detection System using IoT and Computer Vision / A. Kumar, R. Singh, P. Sharma, V. Patel // arXiv. – 2026. – arXiv:2605.08246.
10. Федеральный закон от 09.02.2007 № 16-ФЗ «О транспортной безопасности».
11. ОАО «РЖД». Стандарт организации. Системы видеонаблюдения на объектах железнодорожной инфраструктуры. СТО РЖД 1.12.002-2020.

Reference

- 1) 1. Chen P., Zhang Z., Huang Y. et al. Railway obstacle intrusion warning mechanism integrating YOLO-based detection and risk assessment // Journal of Industrial Information Integration. – 2024. – Vol. 38. – P. 100571.
2. Jocher G., Chaurasia A., Qiu J. YOLO by Ultralytics. Version 11.0. – 2024. – URL: <https://github.com/ultralytics/ultralytics>
3. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // CVPR. – 2016. – P. 779-788.
4. Medvedeva E.V., Perevoshnikova A.A. Issledovanie nejrosetevy`x algoritmov raspoznavaniya ob`ektov zheleznodorozhnoj infrastruktury` na videokadrax // Komp`yuternaya optika. – 2025. – T. 49, № 3. – S. 443-452.
5. Zhang Y., Sun P., Jiang Y. et al. ByteTrack: Multi-Object Tracking by Associating Every Detection Box // ECCV. – 2022. – P. 1-18.
6. Gudfellou Ya., Bendzhio I., Kurvill` A. Glubokoe obuchenie. – М.: ДМК Press, 2018. – 652 с.
7. <https://www.cvat.ai/>
8. Chang C.C., Huang K.H., Lau T.K., Huang C.F., Wang C.H. Using deep learning model integration to build a smart railway traffic safety monitoring system // Scientific Reports. – 2025. – Vol. 15. – Article 4224. – DOI: 10.1038/s41598-025-88830-7.
9. NETRA: Smart Railway Obstruction Detection System using IoT and Computer Vision / A. Kumar, R. Singh, P. Sharma, V. Patel // arXiv. – 2026. – arXiv:2605.08246.
10. Federal`ny`j zakon ot 09.02.2007 № 16-FZ «O transportnoj bezopasnosti».
11. ОАО «RZhD». Standart organizacii. Sistemy` videonablyudeniya na ob`ektax zheleznodorozhnoj infrastruktury`. STO RZhD 1.12.002-2020.