Международный университет информационных технологий

УДК: 004.85:004.932

На правах рукописи

ОЛЖАЕВ ОЛЖАС МҰРАТҰЛЫ

Разработка системы обнаружения повреждений дорог с использованием методов глубокого обучения на основе видеоданных

8D06105 – Наука о данных

Диссертация на соискание степени
доктора философии (PhD)

Отечественный научный консультант:
PhD, профессор-исследователь,
Омаров Батырхан Султанович

Зарубежный научный консультант:
PhD, PhD professor Azizah Suliman,
университет Asia Metropolitan University,
Subang Jaya, Selangor, Malaysia,

Республика Казахстан
Алматы 2026

СОДЕРЖАНИЕ

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ДТП	– Дорожно-транспортное происшествие
ОЭСР	– Организация экономического сотрудничества и развития
API	– Application Programming Interface (Программный интерфейс приложения)
BCE	– Binary Cross-Entropy (Бинарная кросс-энтропия)
BSH	– Boundary-Aware Segmentation Head (Сегментационная ветвь с учетом границ)
CIoU	– Complete Intersection over Union (Полное пересечение по объединению)
CNN	– Convolutional Neural Network (Сверточная нейронная сеть)
CRM	– Classification Refinement Module (Модуль уточнения классификации)
CV	– Computer Vision (Компьютерное зрение)
DDH	– Decoupled Detection Head (Модуль раздельного обнаружения)
DIP	– Digital Image Processing (Цифровая обработка изображений)
DL	– Deep Learning (Глубокое обучение)
FLOPs	– Floating Point Operations per Second (Количество операций с плавающей запятой в секунду)
FN	– False Negative (Ложноотрицательное срабатывание)
FP	– False Positive (Ложноположительное срабатывание)
FPS	– Frames Per Second (Кадры в секунду)
IoU	– Intersection over Union (Пересечение по объединению)
ITS	– Intelligent Transport Systems (Интеллектуальные транспортные системы)
mAP	– mean Average Precision (Средняя средняя точность)
mIoU	– mean Intersection over Union (Среднее попиксельное пересечение по объединению)
NMS	– Non-Maximum Suppression (Подавление немаксимумов)
PCI	– Pavement Condition Index (Индекс состояния покрытия)
RDD	– Road Damage Dataset (Глобальный набор данных повреждений дорог)
RNN	– Recurrent Neural Network (Рекуррентная нейронная сеть)
RoI	– Region of Interest (Область интереса)
SVM	– Support Vector Machine (Метод опорных векторов)
TCR	– Transformer Context Refinement (Контекстное уточнение на основе трансформера)
TP	– True Positive (Истинно-положительное срабатывание)
ViT	– Vision Transformer (Визуальный трансформер)
YOLO	– You Only Look Once (Семейство одностадийных детекторов объектов)

ВВЕДЕНИЕ

Актуальность

Обеспечение эксплуатационной надежности дорожной инфраструктуры следует рассматривать как критический фактор, определяющий не только темпы экономического развития, но и уровень общественной безопасности. Процесс ухудшения дорожного полотна, как правило, инициируется формированием сетки микротрещин, которые при отсутствии своевременного вмешательства неизбежно перерастают в глубокие выбоины. Подобные дефекты создают прямую угрозу для участников движения. Согласно глобальным отчетам Всемирной организации здравоохранения (ВОЗ), ежегодная смертность в результате инцидентов на дорогах достигает 1,19 миллиона человек [1]. При этом диапазон несмертельного травматизма, нередко влекущего за собой стойкую инвалидность, охватывает до 50 миллионов пострадавших. Важно учитывать, что состояние дорожного покрытия детерминирует не менее 10% всех зарегистрированных аварий. В этой связи переход к мониторингу дорожных сетей трансформируется из узкотехнической задачи в приоритетную стратегию в области здравоохранения.

Прямые затраты на ликвидацию последствий ДТП на прямую влияют на финансовый аспект, обусловленный эксплуатацией изношенных дорожных сетей. Глобальная исследовательская практика [2] показывает, что совокупный ущерб от дорожно-транспортного травматизма и деградации покрытия может достигать 3% от валового внутреннего продукта (ВВП) государства. Прямые расходы на медицину, долгосрочное снижение производительности труда, а также форсированный износ инженерных сооружений и транспортных средств формируют этот показатель. И более того, ежегодно возникают миллиардные издержки на внеплановое сервисное обслуживание из-за эксплуатации автопарка на дефектном полотне, что ведет к преждевременному исчерпанию технического ресурса машин. Учитывая эти условия, когда дорожные ведомства стремятся к оптимизации операционных расходов, переход к стратегиям проактивного мониторинга становится безальтернативным инструментом поддержания активов.

На сегодняшний день популярным методом аудита дорожного хозяйства остается визуальное обследование, делегированное персоналу эксплуатационных служб. Однако потенциальная глубина такой экспертной оценки нивелируется ограниченностью технических ресурсов. В действительности ручной контроль представляет собой крайне инертный процесс, сопряженный с высокими операционными издержками и, что критично для точности данных, существенным риском субъективных искажений. Физические ограничения человеческого фактора делают невозможным обеспечение адекватного пространственного охвата, необходимого для непрерывного мониторинга дорожных сетей в режиме реального времени [3]. На фоне стремительной урбанизации и усложнения топологии транспортных узлов традиционная визуальная инспекция

окончательно утратила свою эффективность, и сдерживает развитие цифровизации технического обслуживания дорог технического обслуживания дорог.

Стремительная эволюция компьютерного зрения и алгоритмов глубокого обучения открыла принципиально новые возможности для создания автономных систем мониторинга дорожной инфраструктуры. Прорывные результаты, достигнутые в задачах классификации, объектной детекции и семантической сегментации, сосредоточили исследовательский интерес к автоматизации поиска дорожных дефектов. В рамках данного технологического вектора доминирующим подходом стало использование глубоких сверточных нейронных сетей (CNN) и современных детекторов. Такие архитектуры позволяют экстрагировать высокоуровневые визуальные признаки из снимков высокого разрешения, обеспечивая прецизионную идентификацию повреждений в условиях сложного визуального контекста.

Сфера автоматизированной детекции дорожных дефектов с применением компьютерного зрения на сегодняшний день считается одной из актуальных задач, и видим экспоненциальный рост интереса в данной сфере. Сделав анализ профильной литературы, наглядно можем видеть этот всплеск научной активности: если в 2015 году числилось лишь около 50 релевантных публикаций, то к 2025 году их количество превысило отметку в 900 работ [4]. Подобная динамика обусловлена тремя факторами: повсеместным внедрением сенсоров высокого разрешения, формированием масштабных репрезентативных датасетов и доступностью высокопроизводительных вычислительных мощностей, необходимых для обучения многослойных нейронных сетей.

Если проанализировать текущие научные работы, можно заметить, что основной упор в них делается на обработку одиночных снимков с регистраторов или смартфонов. Но практика показывает: для реального мониторинга этого мало. Нам нужно видеть состояние полотна в динамике, используя видеопоток. Именно работа с видео добавляет в систему «временную координату», которая помогает точнее распознавать дефекты и отсеивать случайные помехи. Когда камера на автомобиле пишет видео без перерывов прямо во время движения, это позволяет охватить огромные участки дорожной сети. В итоге получается инструмент для диагностики повреждений почти в реальном времени, что критически важно для оперативного обслуживания магистралей [5].

Несмотря на достигнутые успехи, сегмент автоматизированной детекции дорожных дефектов средствами дистанционного зондирования по-прежнему сталкивается с рядом фундаментальных барьеров. Эффективность существующих систем критически зависит от вариативности внешней среды: динамического диапазона освещенности, наличия глубоких окклюзий (теней), сезонных метаморфоз ландшафта и нестабильности метеоусловий [6]. Так же дополнительно надо учитывать сложную специфику дистантной съемки под острыми углами, что неизбежно ведет к перспективным искажениям объектов. Более того, морфология самих дефектов – их текстура, нелинейные формы и

широкий диапазон геометрических размеров – крайне затрудняет процесс их прецизионной дифференциации стандартными методами. Данные вызовы диктуют необходимость перехода к более сложным, иерархическим архитектурам глубокого обучения. В частности, требуется внедрение специализированных модулей экстракции признаков, обладающих высокой робастностью к визуальному шуму и способных выделять релевантные паттерны повреждений в условиях низкой контрастности сцены.

Целью диссертационного исследования является разработка интеллектуальной системы автоматического обнаружения и анализа повреждений дорожного покрытия в режиме реального времени на основе методов компьютерного зрения и глубокого обучения, обеспечивающей повышение точности выявления дефектов дорожной инфраструктуры и возможность их последующего картографирования.

Были поставлены следующие **задачи** исследования:

- провести сбор видеоданных с использованием камер высокого разрешения и других средств записи;
- выполнить ручную разметку данных, выделяя разнообразные типы повреждений дорожного покрытия, такие как трещины, выбоины и деформации;
- создать высококачественный размеченный датасет для обучения и тестирования моделей глубокого обучения;
- провести анализ визуальных признаков и характеристик изображений дорожного покрытия, влияющих на качество данных и устойчивость моделей глубокого обучения;
- разработать и обучить модель глубокого обучения для классификации и сегментации повреждений дорог;
- провести тестирование разработанных моделей на собранном наборе данных, с целью достижения высокой точности и устойчивости к различным условиям съемки;
- интегрировать обученную модель в приложение для автоматического мониторинга состояния дорог, обеспечив возможность построения интерактивных карт повреждений;

Объект исследования являются автомобильные дороги и визуальные характеристики повреждений дорожного покрытия, получаемые на основе анализа видеоданных.

Предметом исследования являются методы и алгоритмы компьютерного зрения, глубокого обучения и многозадачных нейронных сетей для обнаружения, классификации и сегментации повреждений дорожного покрытия в режиме реального времени.

Методологическую основу работы составляют методы цифровой обработки изображений, алгоритмы компьютерного зрения и современные методы глубокого машинного обучения. В работе используются сверточные нейронные сети (CNN), трансформерные механизмы внимания, методы многозадачного обучения и анализа видеоданных для построения

интеллектуальной системы обнаружения дорожных повреждений. Также применяются методы предварительной обработки данных, аугментации изображений, оптимизации нейронных сетей и экспериментальной оценки качества моделей с использованием метрик детекции и сегментации.

Научные положения, выносимые на защиту:

- разработанная многозадачная архитектура нейронной сети (TCR-RoadNet) для одновременного обнаружения, классификации и сегментации повреждений дорожного покрытия в режиме реального времени.
- разработанный модуль контекстного уточнения признаков на основе кросс-масштабного трансформерного внимания, повышающий точность локализации сложных и фрагментированных дефектов.
- предложенный метод формирования инвариантных визуальных признаков, обеспечивающий устойчивость распознавания дефектов к изменениям освещенности, фоновому шуму и погодным условиям.

Основные результаты исследования:

- разработана интеллектуальная система автоматического обнаружения повреждений дорожного покрытия на основе анализа видеоданных, обеспечивающая непрерывный мониторинг дорожной инфраструктуры в режиме реального времени с использованием методов компьютерного зрения и глубокого обучения;
- предложена многозадачная архитектура нейронной сети TCR-RoadNet, предназначенная для одновременного обнаружения, классификации и сегментации дорожных дефектов. Архитектура включает многомасштабный сверточный блок извлечения признаков, Transformer Context Refinement (TCR) модуль и специализированные ветви обработки, что позволило повысить точность локализации и устойчивость распознавания сложных повреждений;
- разработан и реализован модуль контекстного уточнения признаков на основе трансформерного механизма внимания, обеспечивающий эффективный анализ пространственных зависимостей между объектами различного масштаба. Использование данного модуля позволило улучшить качество сегментации и уменьшить количество ложноположительных срабатываний при наличии теней, бликов и неоднородной текстуры асфальта;
- выполнен сбор, предварительная обработка и разметка собственного набора видеоданных дорожного покрытия, содержащего различные типы повреждений, включая продольные и поперечные трещины, сетчатые разрушения и выбоины, зафиксированные в реальных условиях эксплуатации транспортной инфраструктуры;
- проведена экспериментальная оценка эффективности разработанной модели. Результаты вычислительных экспериментов показали устойчивую сходимость нейронной сети в процессе обучения, высокие показатели точности обнаружения и сегментации дефектов, а также возможность работы системы в режиме реального времени со скоростью до 57 кадров в секунду;
- выполнен сравнительный анализ разработанной архитектуры с существующими моделями глубокого обучения для задач дорожного

мониторинга, который показал повышение точности обнаружения дефектов и улучшение качества сегментации при сохранении высокой производительности системы;

- разработано программное обеспечение и веб-интерфейс интеллектуального мониторинга дорожной инфраструктуры, позволяющие автоматически визуализировать обнаруженные повреждения, выполнять их привязку к географическим координатам и отображать результаты анализа на интерактивной карте для дальнейшего использования дорожными службами и коммунальными организациями.

Научная новизна исследования заключается в разработке многозадачной нейросетевой модели TCR-RoadNet, которая одновременно находит, распознает и сегментирует дорожные дефекты в реальном времени. Ключевым элементом новизны является внедрение трансформерного механизма внимания, обеспечивающего анализ глобальных пространственных зависимостей дорожной сцены. Это позволило создать уникальный метод адаптации, благодаря которому система обеспечивает устойчивое распознавание повреждений в любую погоду и при любом освещении, игнорируя тени, лужи и визуальный шум.

Научный вклад:

- разработан и размечен уникальный набор видеоданных (датасет) изображений дорожного покрытия, содержащий различные типы повреждений, зафиксированных в реальных условиях эксплуатации транспортной инфраструктуры.
- определены и алгоритмически учтены дополнительные визуальные признаки, такие как влияние сезонных условий и изменчивость динамического освещения, что позволило существенно повысить устойчивость и робастность модели к сложным условиям съемки.
- разработана комплексная система автоматизированного мониторинга дорожной инфраструктуры, способная обрабатывать непрерывный видеопоток в режиме жесткого реального времени и предоставлять аналитическую отчетность через интегрированный веб-интерфейс.

Теоретическая значимость данной работы связана с применением методов компьютерного зрения и глубокого обучения для решения задачи обеспечения непрерывного мониторинга дорожного анализа. В отличие от большинства актуальных подходов, обрабатывающих статические изображения, данная работа увеличивает научную и методологическую базу для анализа пространственно-временных характеристик видеопоследовательностей. Разработанные методы способствуют алгоритмической поддержке интеллектуальных систем для разработки надежных систем машинного зрения в сложных условиях.

Практическая значимость исследования выражается в создании готового к внедрению программного обеспечения и специализированной информационной системы, которая автоматизировала бы поиск дефектов на дорогах и превратила бы дорожные инспекции, которая автоматизирует

процесс обнаружения дефектов дорожного покрытия и способствует переходу от ручных инспекций к интеллектуальным системам мониторинга дорожной инфраструктуры. Результатом исследования стала изученная и обоснованная модель нейронной сети для классификации и локализации дефектов дорог (выбоин, трещин) на основе видеомониторинга с оборудования в режиме реального времени, а также специализированное веб-приложение, реализованное в виде интерактивной панели управления. Интеграция обученной модели нейронной сети с данными веб-интерфейса позволяет автоматически привязывать обнаруженные дефекты дорог и отображать их на карте. Полученные результаты и разработанные программные средства могут быть использованы в оперативной реагировании дорог, для объективного определения приоритетного порядка ремонта дорог и для предотвращения расходования бюджета на техническое обслуживание специализированными коммунальными службами, дорожно-строительными компаниями и министерствами транспорта и дорожной инфраструктуры.

Достоверность полученных результатов обеспечивается использованием современных методов компьютерного зрения и глубокого обучения, применением сверточных нейронных сетей (CNN) и трансформерных механизмов внимания, а также проведением вычислительных экспериментов на размеченном наборе видеоданных дорожного покрытия, собранного в реальных условиях эксплуатации. Достоверность результатов подтверждается сравнительным анализом с существующими архитектурами и алгоритмами обнаружения дорожных дефектов, использованием общепринятых метрик оценки качества (mAP, IoU, Precision, Recall, F1-score), устойчивой сходимостью модели в процессе обучения и воспроизводимостью полученных результатов в различных условиях освещенности, погодных факторов и фонового шума.

Апробация диссертационной работы

По теме диссертации опубликовано 4 публикации [7], [8], [9], [10], в том числе 2 публикации в рейтинговых научных изданиях [7], [8], индексируемых в базе Scopus; 2 публикации в материалах международных конференций [9], [10]; получено 1 авторское свидетельство.

1. Olzhayev, O. M., Kulambayev, B. O., Sakenkyzy, N., & Belisbek, M. (2026). A Real-Time Multi-Scale Feature Pyramid YOLO Architecture for Accurate and Deployment-Efficient Road Damage Detection. *International Journal of Advanced Computer Science & Applications*, 17(3). https://doi.org/10.14569/IJACSA.2026.0170350

2. Kulambayev, B. O., Olzhayev, O. M., Altayeva, A. B., & Zhunisbekova, Z. (2025). A Multi-Scale ROI-Aligned Deep Learning Framework for Automated Road Damage Detection and Severity Assessment. *International Journal of Advanced Computer Science & Applications*, *16*(12). https://doi.org/10.14569/IJACSA.2025.01612107

3. Olzhayev, O., Kulambayev, B., & Omarov, B. (2025). Real-Time Pixel-Wise Segmentation of Road Surface Damage Using a 2D U-Net Architecture.

Procedia Computer Science, *269*, 131-139. https://doi.org/10.1016/j.procs.2025.08.266

4. Kulambayev, B., & Olzhayev, O. (2025). A Mask R-CNN Algorithm for Automated Segmentation of Asphalt Road Cracks. *Procedia Computer Science*, *269*, 39-48. https://doi.org/10.1016/j.procs.2025.08.257

5. Олжаев О. Свидетельство на право охраны программы для ЭВМ № 69666 Республики Казахстан. Программное обеспечение TCR-RoadNet от 7.04.2026

Во всех перечисленных публикациях соискателю принадлежит ведущая роль и основные результаты исследований, анализы, модели, программы созданы автором, выводы сделаны на основе результатов, полученных от работы и исследования соискателя.

Связь с государственными программами

Диссертационное исследование выполнено в рамках грантового финансирования (ИРН AP23487192) на тему «Разработка системы обнаружения повреждений дорожного покрытия в режиме реального времени с использованием компьютерного зрения и искусственного интеллекта».

Исследование соответствует стратегическим приоритетам развития Республики Казахстан и вносит вклад в реализацию Концепции развития искусственного интеллекта в Республике Казахстан на 2024–2029 годы, утвержденной Постановлением Правительства Республики Казахстан №592 от 24 июля 2024 года, в части разработки и внедрения интеллектуальных систем компьютерного зрения, анализа видеоданных и автоматизированного мониторинга транспортной инфраструктуры. Разработанная многозадачная архитектура нейронной сети TCR-RoadNet и система обнаружения повреждений дорожного покрытия способствуют развитию отечественных технологий искусственного интеллекта, интеллектуальных транспортных систем и цифровизации дорожной инфраструктуры Республики Казахстан.

Основное содержание диссертации

Диссертационная работа фокусируется на проблеме нахождения дефектов по задачам обнаружения, классификации и сегментации повреждений дорожного покрытия на основе анализа видеоданных, где используется передовые методы глубокого обучения.

В первом разделе обосновывается актуальность проблемы деградации дорожной инфраструктуры и необходимость перехода от ресурсоемких ручных инспекций к интеллектуальным системам мониторинга. Формулируются цель и задачи исследования, а также определяются научная новизна и практическая значимость разработанной многозадачной нейросетевой архитектуры и программно-аналитического комплекса.

Во втором разделе проводится масштабный обзор литературы, включающий систематизацию типов дорожных дефектов и анализ эволюции методов их обнаружения. Рассматривается исторический переход от алгоритмов классического компьютерного зрения к современным архитектурам глубокого обучения (сверточные сети, детекторы YOLO, визуальные трансформеры). Выявляются фундаментальные пробелы в

существующих исследованиях, такие как сложность обработки непрерывных видеопотоков из-за временной несогласованности, нехватка репрезентативных видео-датасетов и проблема адаптации тяжелых моделей для периферийных вычислений, что алгоритмически обосновывает направление текущей работы.

В третьем разделе подробно описываются материалы и методы исследования. Представляется процесс сбора, предварительной обработки и прецизионного аннотирования уникального набора видеоданных, зафиксированных в реальных условиях эксплуатации. Главное внимание уделяется математическому и структурному проектированию инновационной многозадачной нейросетевой архитектуры TCR-RoadNet. Детально описывается работа многомасштабной сверточной магистрали, модуля контекстного уточнения на основе кросс-масштабного трансформерного внимания (TCR), а также трех специализированных ветвей вывода: модуля раздельного обнаружения, блока уточнения классификации и сегментационной ветви с учетом границ. Дополнительно излагаются стратегия оптимизации и многозадачные функции потерь (комбинация CIoU, Focal Loss и BCE+Dice)

В четвертом разделе представлены результаты комплексной вычислительной и экспериментальной оценки разработанной системы. Проводится анализ динамики сходимости модели в процессе обучения и количественная оценка семантической точности с использованием стандартизированных метрик mAP и mIoU. Выполняется детальное аблиационное исследование, эмпирически доказывающее архитектурную целесообразность каждого внедренного вычислительного модуля. Сравнительный анализ с передовыми мировыми аналогами (State-of-the-Art) подтверждает, что предложенная модель TCR-RoadNet обеспечивает оптимальный компромисс между высокой точностью детализированного распознавания и способностью работать в режиме жесткого реального времени со скоростью 57 кадров в секунду.

Пятый раздел посвящен подведению общих итогов диссертационного исследования. Подтверждается успешное достижение поставленной цели, обобщаются главные научные результаты и формулируются выводы о готовности разработанного аппаратно-программного комплекса к практическому внедрению в задачи автоматизированного аудита транспортных сетей.

Общий объем диссертационной работы составляет 104 страниц, включая 28 иллюстраций, 8 таблиц и список литературы из 154 использованных источников.

1 АНАЛИЗ МЕТОДОВ И СРЕДСТВ АВТОМАТИЗИРОВАННОГО МОНИТОРИНГА ДОРОЖНОГО ПОКРЫТИЯ

В этой главе проводится критический разбор того, как сегодня развиваются технологии автоматического мониторинга дорог. Задача – не просто перечислить существующие методы распознавания ям и трещин, но и найти те сегменты, которые мешают внедрению таких систем в реальную эксплуатацию. Именно эти пробелы и определили вектор исследования.

Логика изложения материала выстроена от постановки фундаментальной проблемы к анализу передовых методов ее решения. Начинается с того, почему следить за дорогами так важно для экономики, и разберем, как именно выглядят основные дефекты на видео. Затем проследим технологическую эволюцию: как на смену простому анализу пикселей и ручному подбору фильтров пришли глубокие нейросети (Deep Learning), совершившие настоящий прорыв в инспекции дорог. Детально сравним классические сверточные сети (CNN) и новые гибридные модели.

Особый акцент в обзоре сделан на обработке видео в движении, так как это основа диссертации. Также проанализированы доступные мировые датасеты и с стало понятно, где их данных не хватает. В финале главы формулируется главные трудности: чувствительность нейросетей к визуальному шуму, проблемы при работе в реальном времени и временную рассинхронизацию кадров. Решение этих вопросов и стало основой научной новизны нашей работы.

1.1 Важность мониторинга дорожной сети и обнаружения повреждений

Дорожная инфраструктура представляет собой фундаментальную систему современных глобальных и региональных экономических моделей, обеспечивая беспрецедентный территориальный охват, гибкость маршрутизации и связность транспортных сетей по сравнению с любыми другими видами транспорта [11, 12]. В работе [13, 14] показывает, как функционирование автомобильных дорог имеет прямую и измеримую корреляцию с макроэкономическими показателями государств. Согласно масштабным исследованиям, проведенным Всемирным банком, Организацией экономического сотрудничества и развития (ОЭСР) и Всемирной дорожной ассоциацией, целевые инвестиции в дорожную инфраструктуру приносят значительную и долгосрочную экономическую отдачу, особенно в странах, чья логистика критически зависит от автомобильных грузоперевозок [11, 15]. Более того, такие авторитетные институты, как Всемирный банк, исторически используют плотность дорог с твердым покрытием, находящихся в хорошем эксплуатационном состоянии, в качестве одного из главных индикаторов «экономической силы» и глобальной конкурентоспособности страны на мировой арене [16]. В контексте экономического роста, качественная дорожная сеть рассматривается не просто как физический актив, но как катализатор. Он способен трансформировать натуральное сельское хозяйство

в коммерческое, стимулировать создание новых рабочих мест в производственном секторе, способствовать переходу работников из неформальной экономики в формальную и снижать стоимость жизни за счет удешевления логистики скоропортящихся продуктов питания [17, 18]. Эмпирические данные из развивающихся стран показывают, что эластичность благосостояния домохозяйств по отношению к качеству дорог составляет около 0,09, а соотношение выгод и затрат для инвестиций в техническое обслуживание дорог достигает 1,8 [19].

С течением времени под воздействием факторов, как описано в работе [16], циклических транспортных нагрузок, климатических факторов, процессов старения битумных вяжущих материалов и неадекватных строительных практик дорожное покрытие неизбежно деградирует. Своевременная оценка и классификация состояния дорог является одной из главных и наиболее ресурсоемких задач в обеспечении безопасности и устойчивости транспортных систем [20, 21]. Ухудшение состояния покрытия негативно сказывается на факторах, такие как комфорте вождения, пропускной способности магистралей, износе транспортных средств и, что наиболее критично, на безопасности дорожного движения [22]. Глобальная статистика дорожно-транспортных происшествий свидетельствует о том, что некачественное дорожное покрытие выступает сопутствующим фактором примерно в 16% всех аварий с тяжелыми последствиями [23]. Выбоины, глубокие продольные трещины и просадки основания могут вызывать внезапную потерю управления, серьезные повреждения подвески транспортных средств и приводить к авариям с фатальным исходом [23, 24]. Эта проблема особенно остро проявляется в развивающихся странах с высокой плотностью трафика и ограниченными бюджетами на ремонт: например, статистические данные по Индии показывают, что инциденты, связанные с выбоинами, унесли более 19 000 жизней в период с 2013 по 2023 год, что составляет около 10% всех смертей в ДТП в стране, вызванных открытыми люками или ямами на дорогах [25].

Помимо прямых угроз безопасности жизни и здоровью граждан, деградация дорожного полотна влечет за собой серьезные экономические последствия [26, 27]. Современные системы управления дорожным покрытием (Pavement Management Systems, PMS) опираются на фундаментальный принцип превентивного технического обслуживания: своевременное выявление и недорогое устранение мелких поверхностных дефектов предотвращает их перерастание в крупные структурные разрушения, требующие капитального ремонта всего дорожного полотна с заменой основания [28]. Отсрочка технического обслуживания приводит к экспоненциальному росту затрат на восстановление инфраструктуры [29]. Американское общество инженеров-строителей (ASCE) в своих отчетах предупреждает, что стремительное ухудшение состояния инфраструктуры может привести к экономическим потерям, исчисляемым триллионами долларов в течение ближайшего десятилетия [30]. Более того, плохое состояние дорог напрямую увеличивает эксплуатационные расходы

транспортных средств (Vehicle Operating Costs, VOC), повышает расход моторного топлива и, как следствие, объемы выбросов парниковых газов, усугубляя экологические проблемы в густонаселенных урбанизированных регионах [31]. Увеличение шероховатости покрытия и наличие дефектов нарушают плавность движения, что не только изнашивает автомобили, но и снижает общую логистическую эффективность государства [20].

Классический инструментарий для оценки состояния дорожного полотна, который и сегодня остается основным в мировой муниципальной практике, базируется преимущественно на ручном визуальном аудите [20]. Технология таких обследований предполагает физическое присутствие инженеров или ремонтных бригад непосредственно на объекте. Используя измерительные колеса и рулетки, специалисты вручную заполняют отчетную документацию, фиксируя параметры каждого обнаруженного дефекта [32]. При всей своей детальности, данный метод обладает критической уязвимостью. Он не только требует колоссальных трудозатрат и времени, но и неизбежно страдает от фактора субъективности: разные инспекторы могут по-разному оценивать степень деградации покрытия. Кроме того, работа на открытых участках дорог сопряжена с прямым риском для жизни персонала из-за близости плотных транспортных потоков [20]. В масштабах целой страны регулярный ручной мониторинг превращается в логистический тупик – бюджетных и человеческих ресурсов попросту не хватает для поддержания актуальности баз данных [33]. Существует и другая крайность – использование высокотехнологичных мобильных лабораторий. Такие системы, укомплектованные лазерными профилометрами, георадарами (GPR) и прецизионной оптикой, позволяют добиться миллиметровой точности измерений. Однако цена таких комплексов и стоимость их обслуживания настолько велики, что даже богатые дорожные ведомства заказывают сканирование сети в лучшем случае раз в несколько лет. Это создает опасный временной разрыв: между редкими проверками дефекты успевают развиться до критического состояния, оставаясь незамеченными для системы управления [33].

Учитывая все перечисленные барьеры, автоматизация процессов детекции дефектов с помощью компьютерного зрения и нейросетей становится не просто альтернативой, а стратегической необходимостью. Развертывание интеллектуальных транспортных систем (ITS), базирующихся на алгоритмах глубокого обучения, дает возможность организовать непрерывный и, что крайне важно, объективный мониторинг дорожной среды в режиме near real-time [34]. Использование доступного оборудования – от обычных видеорегистраторов и камер смартфонов до беспилотных летательных аппаратов (БПЛА) – радикально удешевляет сбор визуальной информации [34]. В результате формируются динамичные массивы больших данных. Это позволяет государственным ведомствам более эффективно распоряжаться бюджетами на ремонт, минимизировать простои из-за перекрытия магистралей и реально повысить безопасность всех участников движения [20]. Подобная технологическая трансформация – это фундамент

для реализации концепции «Умного города» (Smart City). В такой среде дорожная инфраструктура фактически переходит к модели самодиагностики и предиктивного (предупреждающего) обслуживания [35].

1.2 Типы повреждений дорожного покрытия и их характеристики

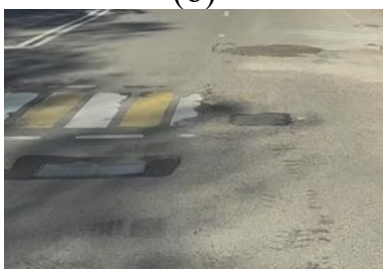
Для разработки надежных, математически обоснованных и легко интерпретируемых автоматизированных систем обнаружения дефектов на основе алгоритмов глубокого обучения необходимо концептуальное понимание физической природы, причин возникновения и визуальных характеристик различных типов повреждений дорожной одежды [36, 37]. В мировой инженерной и строительной практике классификация дефектов строго регламентирована признанными международными и национальными стандартами [38, 39]. Одним из наиболее авторитетных, детализированных и широко используемых руководств в этой области является стандарт Американского общества по испытаниям и материалам ASTM D6433 (Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys) [40]. Данный стандарт, изначально разработанный Инженерным корпусом армии США, определяет 38 различных видов дефектов: 18 специфичных для асфальтобетонных (гибких) покрытий (AC - Asphalt Concrete) и 18 для жестких портландцементных бетонных покрытий (PCC - Portland Cement Concrete) [41]. Стандарт требует классификации каждого обнаруженного дефекта по трем уровням серьезности (низкий, средний, высокий) и оценки плотности его распространения на контрольном участке для последующего математического расчета Индекса состояния покрытия (Pavement Condition Index, PCI) – универсальной метрики от 0 до 100, описывающей структурную целостность и эксплуатационную пригодность дороги [40]. Параллельно с ASTM используется Руководство по идентификации дефектов в рамках Программы долгосрочной эффективности дорожных покрытий (LTPP Distress Identification Manual), которое также предоставляет исчерпывающий словарь для описания разрушений [42].

На рисунке 1 представлены ключевые категории повреждений, классифицированные для задач автоматического распознавания. (а) – продольные трещины, возникающие в результате усталости покрытия или смещения основания; (б) – поперечные трещины, обусловленные термической усадкой материала; (в) – сетчатые (аллигаторные) трещины, свидетельствующие о глубоком структурном разрушении; (г) – выбоины, представляющие наибольшую опасность для безопасности движения; (д, е) – разрушенные знаки на дорогах. Различия в текстуре, форме и контрастности данных объектов формируют основу для обучения многомасштабных признаков в глубоких нейронных сетях.


(а)

(б)

(в)

(г)

(д)

(е)

Рисунок 1 – Общая архитектура комплексной системы обнаружения повреждений дорог

С точки зрения прикладного компьютерного зрения и процесса разметки данных (аннотации) для обучения сверточных нейросетей (CNN), всё многообразие дефектов асфальта обычно группируется в несколько крупных классов. Такое объединение диктуется визуальным сходством – геометрией и текстурой повреждений, которые преобладают на дорогах по всему миру. Причина подобного упрощения кроется в самой логике работы нейросетевых моделей: алгоритмы распознавания образов ищут характерные пиксельные паттерны на поверхности, а не анализируют скрытые физико-механические процессы, протекающие внутри дорожной одежды [43, 44]. Таким образом, для эффективного обучения детекторов целесообразно использовать макроклассификацию, основанную на морфологических признаках, которые представлены в Таблице 1:

Таблица 1 – Основные группы дефектов

Группа дефектов	**Тип повреждения**	**Визуальные характеристики и физические причины возникновения**
Трещины (Cracking)	Продольные трещины (Longitudinal Cracking)	Трещины, проходящие параллельно осевой линии дороги или направлению укладки асфальта. Визуально представляют собой длинные линейные разрывы. Основные причины: усталость покрытия, отраженные трещины от нижних слоев, плохая конструкция продольных стыков при укладке соседних полос, температурные градиенты или смещение подстилающего основания [45]. В системах компьютерного зрения (например, в наборе данных RDD) они часто обозначаются классом D00 и могут дополнительно подразделяться на трещины в колее (связанные с нагрузкой) и вне колеи (связанные с материалом) [46].
	Поперечные трещины (Transverse Cracking)	Трещины, ориентированные перпендикулярно осевой линии дороги. Возникают преимущественно из-за термической усталости и усадки асфальтового

Продолжение таблицы 1

		вяжущего при экстремально низких температурах, когда температурные напряжения превышают прочность материала на разрыв [45]. В датасетах маркируются как класс D10 [47].
	Усталостные/Сетчатые трещины (Alligator / Fatigue Cracking)	Серия многоугольных, пересекающихся трещин, формирующих паттерн, визуально напоминающий кожу аллигатора или проволочную сетку. Это один из наиболее серьезных структурных дефектов, возникающий в зонах колеи. Причина: усталостное разрушение слоя асфальтобетона под воздействием повторяющихся многократных транспортных нагрузок, усугубляемое недостаточной толщиной покрытия или слабым водонасыщенным основанием [45]. Обозначается классом D20 [47].
	Блочные трещины (Block Cracking)	Формируют крупные, прямоугольные или квадратные блоки на поверхности асфальта. В отличие от усталостных трещин, блочное растрескивание не связано напрямую с нагрузками от трафика. Оно возникает из-за усадки асфальта под воздействием суточных температурных циклов и старения (окисления) битумного вяжущего, что приводит к потере эластичности на значительных площадях [45].
Поверхностные деформации	Колееобразование (Rutting)	Линейные, продольные впадины или углубления вдоль траектории движения колес транспортных средств. С точки зрения визуального обнаружения, колеи крайне сложно зафиксировать с помощью обычных 2D-монокулярных камер, поскольку они требуют анализа карты глубины, 3D-профилирования лазером или использования стереозрения [42].
	Просадки и сдвиги (Depressions, Shoving)	Локализованные заниженные участки поверхности (впадины), способные удерживать воду (Depressions), или пластические сдвиговые деформации, образующие волны и бугры (Shoving). Сдвиги часто возникают в зонах интенсивного торможения или ускорения (на перекрестках) из-за нестабильности асфальтовой смеси [46].
Поверхностные дефекты	Выкрашивание (Raveling)	Прогрессирующий процесс потери мелкого, а затем и крупного каменного заполнителя с поверхности покрытия. Происходит из-за разрушения связи между асфальтовым вяжущим и заполнителем под воздействием влаги, солнца и трафика, делая поверхность шероховатой и пористой [42].
Локальные разрушения	Выбоины (Potholes)	Чашеобразные углубления и проломы в покрытии различных размеров с неровными, зазубренными краями [48]. Процесс формирования выбоин носит каскадный характер: вода проникает через трещины, разрушает связь в подстилающих слоях грунта, а затем интенсивный трафик (особенно в сочетании с циклами замерзания-оттаивания) выбивает куски ничем не поддерживаемого асфальта [48]. Выбоины (класс D40 в RDD) представляют наибольшую непосредственную опасность, способную спровоцировать ДТП [25].

Визуальный облик перечисленных дефектов на цифровых кадрах отличается крайней нестабильностью. Эффективность их распознавания

напрямую зависит от условий освещенности: резкое прямое солнце, глубокие сумерки или пасмурная погода кардинально меняют контрастность сцены. Серьезную проблему создают артефакты, такие как резкие падающие тени от зданий и деревьев, которые нейросеть может принять за глубокие трещины. Погодные факторы, такие как лужи или влажный блеск асфальта после дождя, создают блики, искажающие истинную текстуру полотна. Не стоит забывать и о вариативности самого фона: цвет асфальта меняется от выцветшего светло-серого до насыщенного черного, что сбивает настройки чувствительности алгоритмов [48]. Ситуацию осложняют многочисленные визуальные артефакты: дорожная разметка, масляные пятна, решетки ливневой канализации, палая листва и мусор. Все эти объекты по своим морфологическим признакам могут напоминать повреждения, что провоцирует рост ложных срабатываний (False Positives) [49]. В совокупности эти факторы формируют сложную вычислительную среду, где от алгоритмов требуется не просто фиксация объектов, а высоконадежная экстракция признаков, способная отделять полезный сигнал от визуального шума.

1.3 Традиционные методы обнаружения в компьютерном зрении

До появления мощных графических процессоров (GPU) и триумфа глубокого обучения, автоматика в дорожной диагностике работала иначе. Почти все изыскания конца 90-х и 2000-х годов строились на классической цифровой обработке сигналов (DIP) и базовых алгоритмах компьютерного зрения [50]. В основе лежал простой физический принцип: считалось, что любой дефект на снимке всегда темнее окружающего фона и выделяется резким перепадом яркости [51]. Исходя из этой гипотезы, инженеры собирали жесткие цепочки обработки (pipelines). Типичный алгоритм того времени выглядел как строгая последовательность: сначала шла фильтрация шумов и выравнивание гистограммы, затем выделялись границы объектов, а после – вручную подбирались статистические признаки для итоговой классификации [52]. Этот подход требовал филигранной настройки под каждый конкретный снимок, что и было его главным ограничением.

Традиционные подходы можно систематизировать в несколько ключевых категорий [35, 53]

Методы на основе пороговой обработки (Thresholding и Segmentation): Пороговая обработка самый старый и простой способ выделить дефекты на фото [52, 54]. Весь расчет здесь строится на элементарной логике: берется конкретное значение яркости (порог), и всё, что темнее него, считается трещиной, а всё, что светлее – фоном [52]. Самым известным инструментом здесь стал метод Оцу (Otsu’s algorithm). Его математика направлена на автоматический поиск такого порога, который бы максимально четко разделял пиксели на два класса через дисперсию [49]. Но на реальной дороге этот метод часто пасовал. Проблема в том, что асфальт освещен неравномерно, и один «золотой» порог для всего кадра просто не работает [55, 56]. Пытаясь спасти ситуацию, исследователи перешли к

адаптивным методам. Здесь порог уже не был статичным – он пересчитывался динамически для каждого небольшого участка изображения [35]. Однако даже такая гибкость не помогла победить визуальный шум. Резкие тени, масляные пятна или следы от протекторов по своей яркости почти не отличаются от реальных трещин. В итоге алгоритмы либо «склеивали» дефект с фоном, либо заваливали систему мусорными объектами, принимая любую тень за провал в покрытии [57].

Методы обнаружения краев (Edge Detection): Так как любая трещина – это прежде всего линейный объект с резким скачком яркости, именно детекторы границ стали фундаментом для ранних систем поиска дефектов [51]. В ход пошли классические дифференциальные фильтры первого порядка: операторы Собеля, Прюитта и Робертса. Через свертку кадра с матрицами 3x3 они вычисляли градиент интенсивности по осям, буквально «высвечивая» контуры объектов на снимке [58]. Параллельно развивались методы второго порядка – например, лапласиан, который искал точки перехода через ноль во второй производной [59]. Настоящим стандартом в индустрии стал многошаговый детектор Канни. Его ценят за надежность: алгоритм сначала гасит шум размытием по Гауссу, затем вычисляет направление градиента, убирает лишние пиксели (non-maximum suppression) и, наконец, связывает разрозненные точки в цельные линии через двойной порог [51]. Но здесь инженеры столкнулись с парадоксом. Несмотря на высокую точность, детектор Канни оказался слишком чувствителен к самой фактуре асфальта. Грубое покрытие с крупными камнями воспринималось системой как бесконечный набор мелких краев. Это порождало такой объем визуального «мусора», что даже последующая морфологическая очистка – эрозия и дилатация – не помогала отделить реальную трещину от естественной текстуры дороги [58].

Методы на основе фильтрации, преобразований и анализа текстур: Когда стало понятно, что простые дифференциальные фильтры не справляются с шумом, инженеры перешли к частотно-пространственному анализу. Фильтры Габора, работа которых напоминает механизмы зрительной коры человека, и вейвлет-преобразования (в том числе алгоритм A trous) показали отличные результаты в поиске направленных линий. С их помощью удавалось «вытаскивать» тонкие продольные и поперечные трещины, одновременно отсекая хаотичный фоновый шум асфальтового покрытия [49]. Для работы с более сложной геометрией, например выбоинами или участками выкрашивания, потребовались иные инструменты. Здесь на первый план вышел статистический анализ текстур. Использование матриц совпадений (GLCM) и локальных бинарных шаблонов (LBP) позволило вычислять такие параметры, как энтропия, контрастность и однородность кадра. Это дало возможность математически отличить реальную глубокую выбоину от плоского, но темного масляного пятна [60]. Если же стояла задача найти идеально прямые линии, исследователи чаще всего использовали преобразование Хафа в комбинации с направляемыми фильтрами и проекционными интегралами [61].

Классическое машинное обучение (Traditional Machine Learning): Венцом эволюции традиционного компьютерного зрения в этой области стало использование конвейеров, где извлеченные вручную (handcrafted) признаки (такие как векторы HOG – Histograms of Oriented Gradients, дескрипторы LBP, статистические моменты из GLCM или ответы фильтров Габора) подавались на вход алгоритмам классического машинного обучения (Machine Learning, ML) [60, 62].

Алгоритмы классификации, такие как Метод опорных векторов (Support Vector Machine, SVM), Случайный лес (Random Forest, RF), Деревья решений (Decision Trees), K-ближайших соседей (k-NN) и ансамблевые методы градиентного бустинга (например, LightGBM), обучались разделять фрагменты изображений на классы («дефект» или «фон»), а также распознавать паттерны повреждений [61, 63]. Исследования демонстрировали впечатляющие результаты работы этих методов, но преимущественно в строго контролируемых условиях [64, 65]:

- в одном из исследований гибридный подход, объединяющий извлечение признаков с помощью классификатора Random Forest (RF), позволил достичь выдающейся валидационной точности в 99,97% при распознавании трещин [61];
- в другой работе применение алгоритма Light Gradient Boosting Machine (LightGBM) в сочетании с фильтрами Гаусса и проекционными интегралами продемонстрировало высокую производительность с точностью более 0,96 (96%) при классификации шести различных паттернов трещин [66];
- метод SVM в комбинации с алгоритмом Оцу и LBP-дескрипторами широко применялся для базовой сегментации, показывая надежные результаты при условии однородного освещения [58].

Дальше представлены этапы работы классических алгоритмов компьютерного зрения при локализации трещин. На рисунке видно применение оператора Собеля, что позволяет подчеркнуть пространственные градиенты яркости, однако не обеспечивает четкой сегментации объекта.

Рисунок 2 – Применение оператора Собеля

В то же время на рисунке 3 видно, как детектор границ Канни формирует более тонкие контуры, но проявляет избыточную чувствительность к

естественной макротекстуре асфальтобетона и неоднородному освещению. Сформированный таким образом высокий уровень визуального «шума» (ложноположительных срабатываний) обуславливает низкую робастность традиционных методов в реальных условиях эксплуатации и необходимость использования иерархических признаков глубокого обучения.

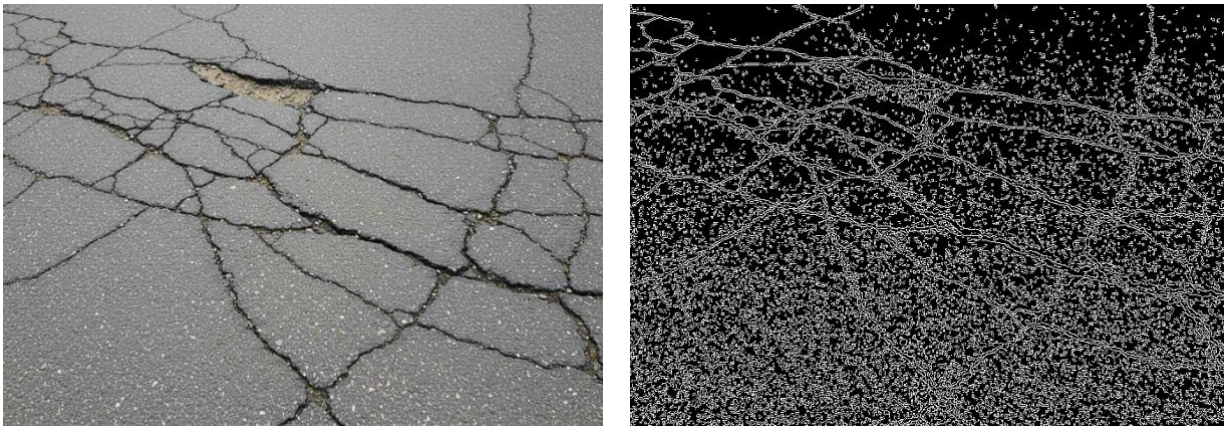


Рисунок 3 – Применение оператора Канни

Ограничения традиционных методов: Несмотря на вычислительную легкость, интерпретируемость и отсутствие необходимости в огромных массивах размеченных данных (в отличие от современных нейросетей), парадигма традиционного компьютерного зрения обладает рядом критических, системных недостатков. Главной проблемой является их полная зависимость от ручного конструирования признаков (handcrafted features). Исследователь должен был заранее математически описать, как именно выглядит дефект [13]. Это делает традиционные алгоритмы крайне негибкими (brittle) и неспособными к обобщению в сложных, неконтролируемых условиях реального мира (wild environments). Изменения интенсивности солнечного света, длинные тени, блики от луж, сложная геометрия дорожной разметки, листья и мусор легко «обманывают» фильтры Собеля и Канни, генерируя неприемлемо высокий уровень ложных срабатываний [58]. Любая смена типа дорожного покрытия (например, переход от темного свежего асфальта к выцветшему светлому или бетонному покрытию) требует ручной перекалибровки порогов и весов фильтров, что делает традиционные методы абсолютно непригодными для внедрения в глобальные, масштабируемые системы автоматического дорожного мониторинга [67].

1.4 Появление глубокого обучения в инспекции инфраструктуры

Стремительное развитие концепций глубокого обучения (Deep Learning, DL) и, в частности, архитектур глубоких сверточных нейронных сетей (Deep Convolutional Neural Networks, DCNNs) в середине 2010-х годов совершило подлинную парадигмальную революцию в области компьютерного зрения, изменив все устоявшиеся подходы к мониторингу и оценке состояния гражданской инфраструктуры [35, 68, 69].

Ниже, рисунок 4 демонстрирует трансформацию исследовательской парадигмы в области мониторинга дорог. (2011–2012) – доминирование классических методов image-processing и геометрической реконструкции (CrackTree); (2016–2017) – появление специализированных наборов данных (CrackForest, GAPs) для машинного обучения; (2018–2021) – переход к пиксельной сегментации на базе Deep Learning (DeepCrack) и массовому сбору данных (RDD); (2022–2024) – внедрение механизмов внимания (CrackFormer), адаптации домена (DA-RDD) и переход к анализу видеопоследовательностей с попиксельными масками.

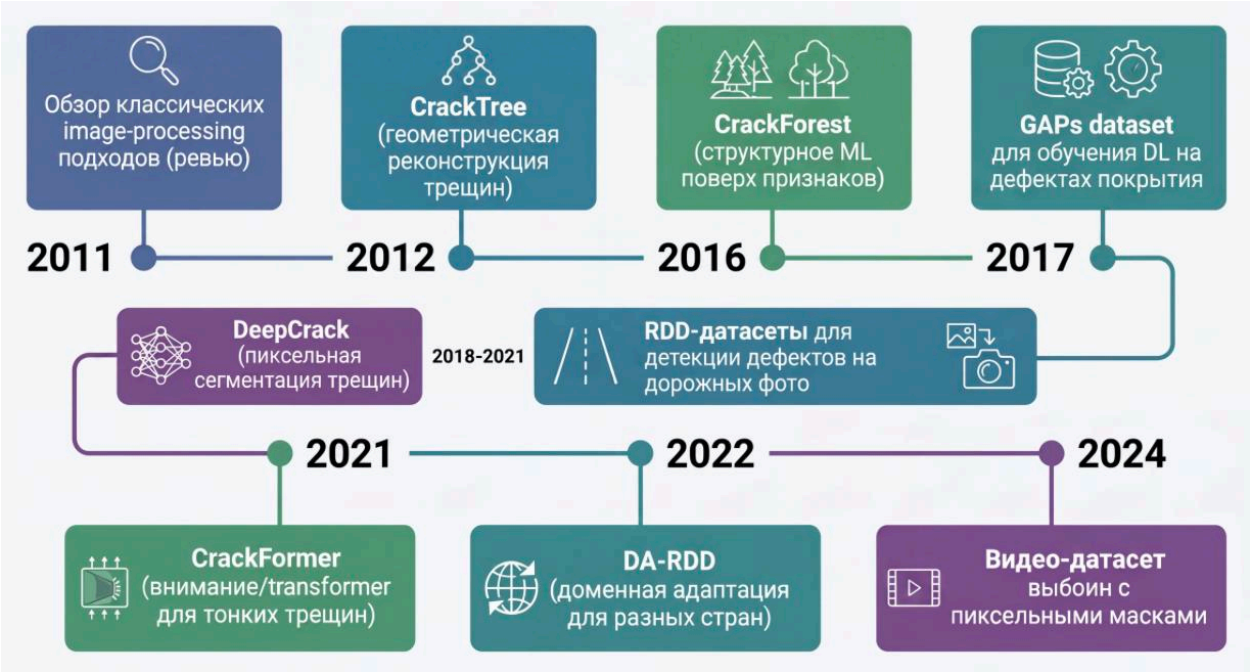


Рисунок 4 – Хронологическая эволюция методов и наборов данных для анализа дорожных повреждений

Фундаментальный сдвиг заключался в переходе от парадигмы «ручного конструирования признаков» (feature engineering), доминировавшей в классическом машинном обучении, к концепции «автоматического обучения признакам» (feature learning / representation learning) [13]. В отличие от алгоритмов Канни или фильтров Габора, где математические правила поиска краев задавались инженером априори, глубокие сверточные нейронные сети способны автоматически, в процессе обратного распространения ошибки (back-propagation), извлекать оптимальные, иерархические и высокоабстрактные визуальные признаки непосредственно из необработанных матриц пиксельных данных [70]. Нижние слои CNN выучивают простые фильтры (подобные детекторам краев Габора), средние слои комбинируют их в текстуры, а глубокие слои формируют сложные семантические представления, позволяющие отличать настоящую разветвленную усталостную трещину от сложной тени дерева или следа шины [66, 71]. Это наделило системы беспрецедентной устойчивостью (робастностью) к фоновым шумам, радикальным изменениям освещенности и

неоднородности макротекстуры асфальта, полностью превзойдя возможности SVM и RF [52, 72].

На рисунке 5 представлено сравнение двух методологических подходов к анализу дорожной инфраструктуры. В традиционном машинном обучении (сверху) ключевым этапом является ручное проектирование признаков, что ограничивает гибкость системы в неконтролируемых условиях. В парадигме глубокого обучения (снизу) реализован принцип сквозного обучения (end-to-end), при котором нейронная сеть автоматически формирует иерархическую систему визуальных дескрипторов. Это позволяет эффективно дифференцировать истинные дефекты от сложных визуальных помех, таких как тени, блики и неоднородность текстуры асфальта.

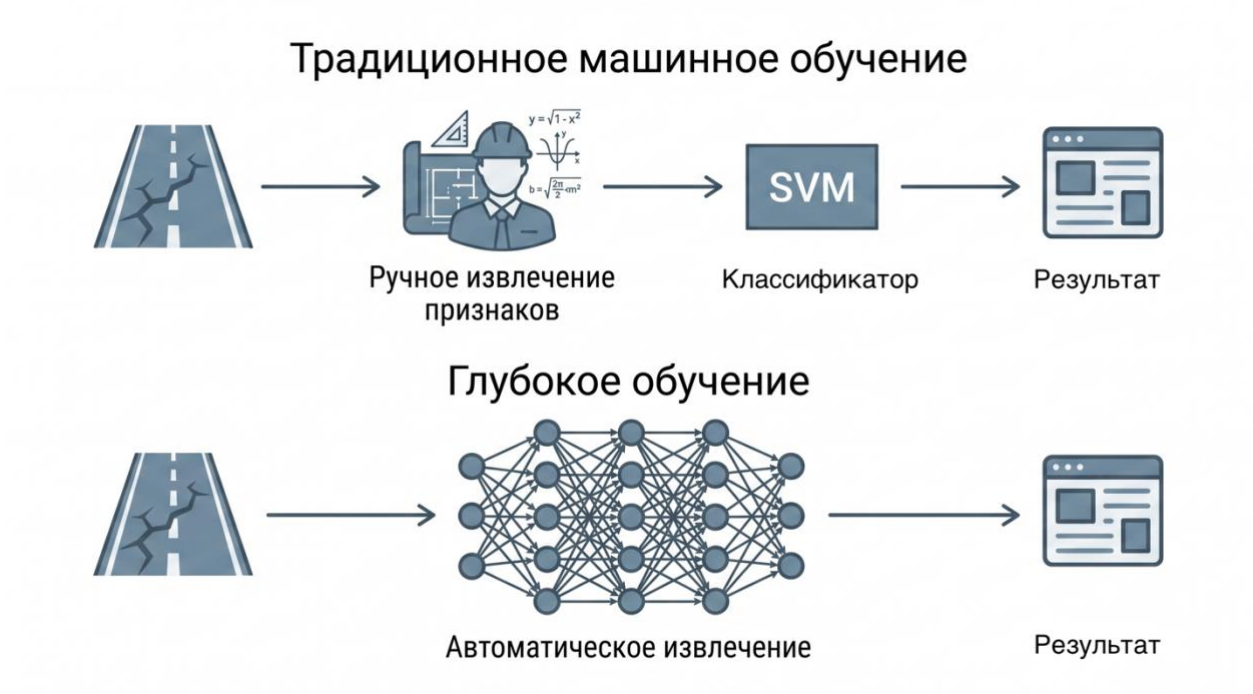


Рисунок 5 – Различие между традиционным машинным обучением и глубоким обучением в задачах инспекции дорог

Интеграция технологий глубокого обучения была катализирована и неразрывно связана с тремя факторами: экспоненциальным ростом вычислительных мощностей (в частности, разработкой специализированных графических ускорителей GPU), созданием оптимизированных фреймворков (TensorFlow, PyTorch) и, что не менее важно, демократизацией и удешевлением сенсорного оборудования [73]. Ранее высокоточные автоматизированные системы оценки дорожного покрытия (Automated Pavement Distress Systems) являлись монополией правительственных агентств, так как они полагались исключительно на громоздкие и баснословно дорогие специализированные автомобили-лаборатории. Такие комплексы были нашпигованы 3D-лазерными сканерами (LiDAR), георадарами (GPR) для подповерхностного зондирования, сложными инерциальными измерительными модулями (IMU) и линейными инфракрасными камерами [50]. Безусловно, такие системы обеспечивали превосходную, миллиметровую

точность реконструкции 3D-профиля дороги. Однако их астрономическая стоимость и сложность эксплуатации привели к тому, что мониторинг мог проводиться не чаще одного-двух раз в год, оставляя огромные «слепые зоны» во времени, за которое дефекты успевали развиться [33].

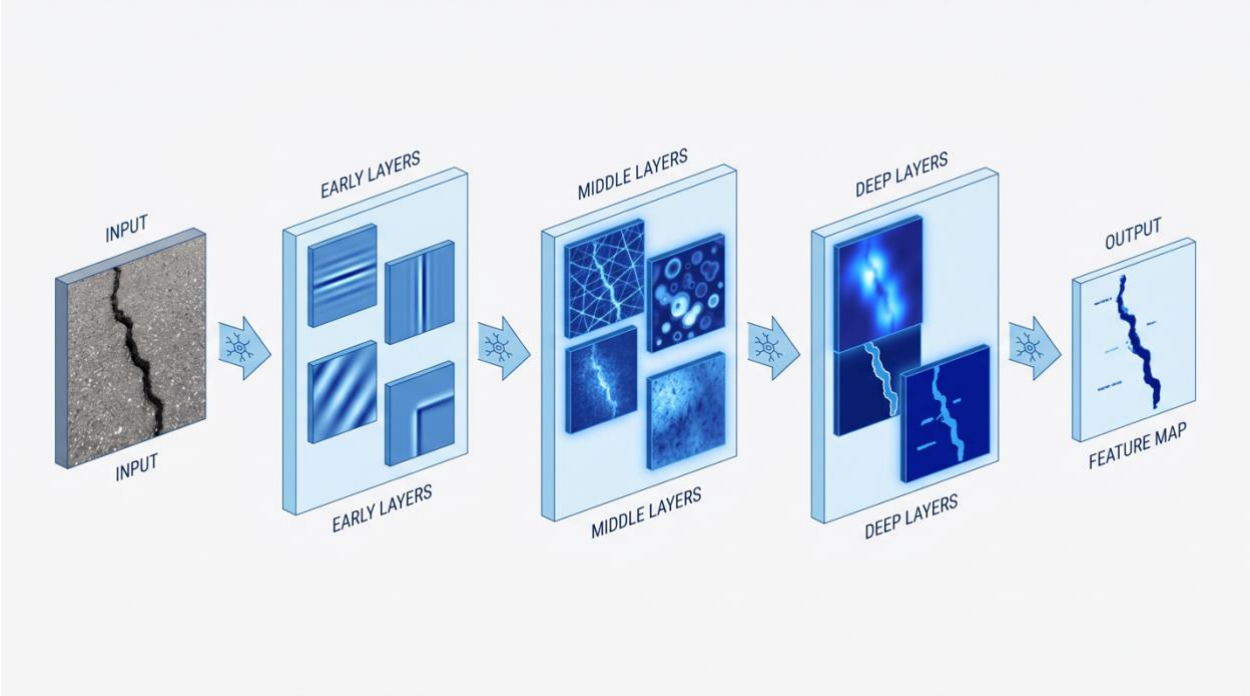


Рисунок 6 – Структура извлечения признаков в глубоких сверточных нейронных сетях

Глубокие сверточные сети (CNN) доказали свою способность извлекать высококачественную, семантически богатую информацию о дефектах из простых, недорогих 2D-изображений и видеоданных в стандартном оптическом (RGB) диапазоне [32]. Процесс, как видно на рисунке 6, начинается с анализа микроструктур на начальных слоях (выделение кромок и перепадов яркости). На промежуточных уровнях сеть формирует представления о макротекстуре асфальта и геометрии трещин. Глубокие слои генерируют высокоуровневые семантические карты, позволяющие системе надежно дифференцировать повреждение покрытия от визуально схожих артефактов, таких как тени или следы шин.Это открыло дорогу для массового использования повсеместно доступных (ubiquitous) устройств сбора данных. В научный и практический оборот вошли смартфоны с GPS, закрепленные на приборных панелях обычных автомобилей (dashcams), носимые экшен-камеры (GoPro) и коммерческие видеорегистраторы [74]. Использование смартфонов и видеорегистраторов в сочетании с алгоритмами DL сформировало мощную концепцию краудсенсинга (crowdsensing), позволяющую муниципальным службам привлекать обычных водителей, общественный транспорт или мусоровозы для непрерывного, фонового и практически бесплатного сбора данных о состоянии дорожной сети на уровне всего мегаполиса [34]. Параллельно произошел взрывной рост применения беспилотных летательных аппаратов (БПЛА / UAV). Оснащенные RGB-

камерами высокого разрешения, дроны обеспечили недорогой доступ к инспекции труднодоступных объектов (мостов, эстакад), позволили получать снимки в зенитном ракурсе без перспективных искажений и предоставили возможность оперативного сканирования гигантских территорий (road topology mapping), что стало критически важным инструментом для оценки инфраструктурного ущерба сразу после масштабных стихийных бедствий (землетрясений, наводнений) [75, 76].

1.5 Нейросетевые архитектуры для обнаружения дорожных дефектов

В настоящий момент академическое и индустриальное сообщество применяет колоссальный спектр архитектур глубокого обучения для инспекции гражданской инфраструктуры [77, 78]. Их можно классифицировать в зависимости от типа решаемой задачи компьютерного зрения: от простой бинарной классификации патчей до точного попиксельного выделения дефектов и анализа видеопотоков [79, 80]. Рисунок 7 ниже демонстрирует эволюцию подходов к цифровой инспекции инфраструктуры: (а) – классификация изображений, позволяющая определить только наличие дефекта в конкретном блоке (патче); (б) – обнаружение объектов, обеспечивающее локализацию повреждений с помощью ограничивающих рамок для оценки их количества и типа; (в) – семантическая сегментация, необходимая для прецизионного выделения общей площади разрушения на уровне пикселей; (г) – экземплярная сегментация, позволяющая разделять и индивидуально анализировать каждый дефект в отдельности.

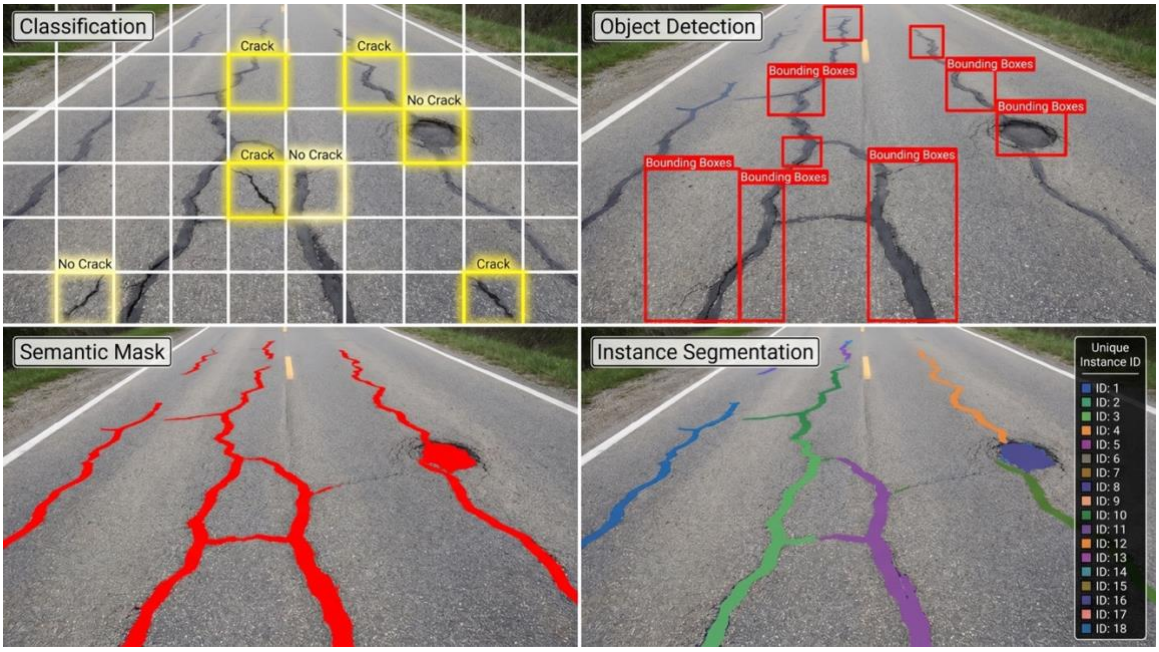


Рисунок 7 – Основные типы задач компьютерного зрения при анализе повреждений дорожного полотна

Ниже в таблице 2 приведена систематизация основных архитектур, применяемых в современных исследованиях [81, 82]

Таблица 2 – Основные архитектуры глубокого обучения для обнаружения повреждений дорожного покрытия

Категория задачи	**Основные архитектуры и базовые сети (Backbones)**	**Принципы работы, преимущества и ограничения**
Классификация изображений (Image Classification)	AlexNet, VGG (VGG-16, VGG-19), ResNet (ResNet-18, 34, 50, 101), Inception (InceptionV4), MobileNet, SqueezeNet, ShuffleNet. Специфические сети: CrackNet (I, II, V).	Исторически первый подход. Изображение нарезается на мелкие патчи, и сеть предсказывает наличие или отсутствие трещины в каждом патче [83]. Использование остаточных связей (residual connections) в **ResNet** решило проблему затухания градиента, позволив обучать сверхглубокие сети (до 101 слоя) для выявления микротрещин на сложном фоне [84]. Для развертывания на мобильных процессорах используются **MobileNet**, **SqueezeNet** и **ShuffleNet**, которые применяют разделяемые по глубине свертки (depthwise separable convolutions) для радикального снижения вычислительных затрат [85]. Отдельного упоминания заслуживает семейство **CrackNet**, разработанное специально для дорог. В отличие от классических CNN, CrackNet принципиально не содержит слоев объединения (pooling layers). Это позволяет сохранять исходное пространственное разрешение карт признаков до самого конца сети, предотвращая потерю тонких деталей и микротрещин, хотя и требует больше памяти [86, 87]. Ограничение классификаторов – они не дают точных координат и размеров дефекта [88].
Обнаружение объектов: Двухстадийные (Two-stage Object Detectors)	R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN.	Алгоритмы работают в два этапа: сначала специализированная сеть (Region Proposal Network, RPN) генерирует сотни регионов-кандидатов на изображении, где предположительно находится дефект, а затем классификатор проверяет каждый регион и уточняет координаты ограничивающей рамки (bounding box) [89]. Такие модели обеспечивают эталонную, высочайшую точность локализации, но обладают огромным количеством параметров. Их главный недостаток – низкая скорость вывода (inference), что делает их малопригодными для обработки видеопотоков с камеры движущегося автомобиля в реальном времени (без применения мощных серверных GPU) [32, 90].
Обнаружение объектов: Одностадийные (One-stage Object Detectors)	SSD (Single Shot MultiBox Detector), RetinaNet, EfficientDet, Семейство YOLO (You Only Look Once): YOLOv3, v4, v5, v7, v8, v9, v11. Специфические модификации: YOLOv8-PD, RepGD-YOLOv8W, YOLO-LWNet.	Эти алгоритмы предсказывают классы и координаты рамок за один проход нейронной сети по изображению в виде регрессионной задачи [91]. Это обеспечивает превосходный баланс между высокой точностью и скоростью (свыше 30–60 FPS), что критично для видео-аналитики [31, 92]. Семейство **YOLO** стало абсолютным отраслевым стандартом де-факто [91, 93]. Современные итерации вносят революционные изменения. Например, **YOLOv7** внедряет структуру E-ELAN (Extended Efficient Layer Aggregation Network), которая улучшает способность сети к обучению непрерывных признаков без увеличения длины градиентного пути [94]. Инновационный **YOLOv8** отличается “anchor-free” подходом (полным отказом от якорных рамок, требующих ручной настройки) и развязанной головной частью (decoupled head), где задачи определения класса и регрессии координат рамок выполняются физически параллельными ветвями, что резко повышает точность

Продолжение таблицы 2

		локализации мелких дефектов [95, 96]. Исследователи активно модифицируют YOLO для дорожных задач: внедряют модули внимания (CBAM, ECA) для фокусировки на трещинах [97], заменяют стандартные сверточные слои на легкие GhostConv для снижения энергопотребления (модели YOLO-LWNet, YOLOv8-PD) [98], интегрируют архитектуру RepViTBlock (модель RepGD-YOLOv8W) [99], а также применяют функции потерь Wise-IoU (WIOU) и GioU для более точной сходимости ограничивающих рамок при сложных перекрытиях дефектов [97, 100].
Семантическая сегментация (Semantic Segmentation)	FCN (Fully Convolutional Network), U-Net, DeepCrack, SegNet, RHA-Net, PavementNet, CrackUNet.	В отличие от ограничивающих рамок (Object Detection), сегментация классифицирует каждый отдельный пиксель изображения, создавая точную бинарную маску дефекта. Это критически важно для количественной инспекции: вычисления точной площади выбоины, физической длины и ширины трещины для расчета индекса PCI [83, 101]. Архитектура **U-Net** (с энкодером и декодером, соединенными skip-connections) является золотым стандартом, так как она великолепно восстанавливает пространственные детали потерянных на этапах свертки границ дефекта [49]. Модели, подобные **DeepCrack**, используют иерархическое слияние признаков с разных слоев свертки для достижения прецизионной сегментации сложных сетчатых трещин [95]. Недостаток – сегментация требует самых колоссальных вычислительных мощностей и невероятно трудоемкой ручной разметки обучающих данных на уровне пикселей [88, 102].
Vision Transformers (ViT) и Механизмы Внимания	Vision Transformer (ViT), Swin Transformer, PoFormer, DeIT, CrackFormer, RoadFormer.	Ключевым ограничением всех сверточных сетей (CNN) является физический размер локального рецептивного поля: ядро свертки видит лишь малую окрестность пикселей (например, 3x3), что затрудняет понимание глобального макро-контекста сцены [103]. В последние 2-3 года в инспекцию дорог ворвались трансформеры (Vision Transformers – ViT) [20]. Используя механизмы самовнимания (self-attention), модели ViT (и их гибриды с CNN, такие как **PoFormer** или иерархический **Swin Transformer**) способны улавливать долгосрочные зависимости между бесконечно удаленными участками одного изображения. Это концептуально важно для идентификации длинных, но прерывистых продольных трещин, или для понимания того, что массив разрозненных трещин на самом деле является единой структурой блочного растрескивания [20, 104].

1.6 Подходы к обнаружению повреждений дорожного покрытия на основе видео

Хотя глубокие нейронные сети (YOLO, ViT), работающие со статическими фотографиями (frame-wise detection), демонстрируют в лабораторных условиях выдающиеся показатели mAP (mean Average Precision), их прямое, наивное применение к непрерывным видеопотокам сталкивается с серьезной проблемой – феноменом недостаточной временно́й согласованности (temporal inconsistency) [95, 105].

В реальных условиях работы видеорегистратора на движущемся автомобиле визуальные характеристики одного и того же дефекта меняются с

каждым фреймом. Это происходит из-за смазывания от движения (motion blur), вибраций и тряски автомобиля, изменения ракурса камеры, наложения падающих теней от столбов, изменения геометрии освещения и частичных перекрытий объектов (например, пешеходами или другими авто) [106]. Из-за этого детектор объектов, обрабатывающий каждый кадр в вакууме, изолированно от соседних, подвержен сильному эффекту «мерцания» (flickering) [107]: выбоина может успешно обнаруживаться в кадре *t*, таинственно исчезать в кадрах *t+1* и *t+2* из-за легкого блика, и снова появляться в кадре *t+3* [95]. Это приводит к крайней нестабильности предсказаний, множественному дублированию, резким скачкам уверенности модели и лавине ложных срабатываний (False Positives) [106].

Рисунок 8 – Визуализация проблемы временной несогласованности (мерцания) при анализе видеопотока

Для преодоления этой критической преграды на пути к промышленной видео-аналитике, передовые исследования 2023–2025 годов сосредоточены на парадигме интеграции и слияния пространственно-временной (spatio-temporal) информации, заложенной в видеоданных [108]. Основные архитектурные и алгоритмические подходы включают:

1. Временно́е моделирование и Рекуррентные нейронные сети (RNN): Одним из наиболее интуитивных способов внедрения временного контекста является комбинирование алгоритмов извлечения пространственных признаков (CNN) с рекуррентными нейронными сетями для анализа последовательностей [109, 110]. Архитектуры, использующие блоки долгой краткосрочной памяти (Long Short-Term Memory, LSTM) или управляемые рекуррентные блоки (GRU), позволяют модели накапливать скрытое состояние (hidden state) и буквально «запоминать» информацию о присутствии дефекта из предыдущих кадров видео [111]. Выдающимся решением стало создание специализированных сетей ConvLSTM (Convolutional LSTM). В отличие от классических LSTM, которые превращают данные в одномерные векторы, теряя топологию, ConvLSTM заменяет операции матричного умножения внутри ячеек памяти на операции двумерной свертки. Это позволяет сохранять пространственную структуру (длину, ширину и геометрию) карт признаков трещин при передаче их во времени [112]. Исследователи создают архитектуры, где, например, предобученная

InceptionV4 извлекает визуальные паттерны, а последующий слой ConvLSTM оценивает их эволюцию во времени, сглаживая мерцания [113].

2. Оптический поток (Optical Flow) и Трехмерные свертки (3D CNN): анализ оптического потока между соседними кадрами видео предоставляет алгоритму математически точную информацию о векторе движения каждого пикселя сцены. Интеграция оптического потока позволяет компенсировать артефакты от тряски камеры и смазывания [114]. Модели могут проецировать (warp) карты признаков из предыдущего высококачественного кадра в текущий размытый кадр, тем самым усиливая сигнал от слабых или перекрытых трещин [115]. Альтернативным, более ресурсоемким подходом является использование трехмерных сверточных сетей (3D CNN, например, архитектура I3D), где ядро свертки движется не только по пространственным осям ширины и высоты (X, Y), но и по дополнительной оси времени (T), одновременно обрабатывая пакет (слайс) из 10-15 последовательных кадров как единый 3D-тензор [116].

3. Инновационные методы: Tracking-by-detection (Трекинг) и Графовые сети: для развертывания видео систем в режиме реального времени на базе YOLO активно применяется парадигма «обнаружение и отслеживание». Системы, подобные Road-TransTrack, объединяют детекторы с передовыми алгоритмами трекинга (такими как DeepSORT или SORT). Они используют алгоритмы сопоставления (matching) и фильтры Калмана для того, чтобы присваивать уникальный ID каждой выбоине, появившейся в кадре, и отслеживать ее траекторию на протяжении всего видеофрагмента. Это позволяет вести точный количественный подсчет дефектов на километр дороги без дублирования [20].

4. Пространства состояний (State-Space Models, SSM) и архитектура Mamba: самым передовым направлением исследований (2024-2025 гг.) является интеграция пространственно-временных моделей пространства состояний, в частности архитектуры Mamba, которая стала революционной альтернативой Трансформерам. Выдающимся примером является разработка модели STG-Mamba-YOLO (Spatio-Temporal Graph Mamba YOLO) и GMRD (Graph-MambaRoadDet). Эти фреймворки сдвигают парадигму от изолированного распознавания объектов к иерархическому рассуждению:

- на *уровне пикселей* модуль Mamba интегрируется в хребет YOLO для улавливания сверхдальних пространственных зависимостей длинных линейных трещин [95];
- на *уровне объектов* графовая нейронная сеть (Graph Network) формирует топологию взаимосвязей между всеми кандидатами в дефекты в кадре (узлами графа), что резко снижает ложные срабатывания [95];
- на *уровне последовательности (Sequence Level)* модуль Temporal Graph Mamba отслеживает эволюцию всего графа дефектов во времени через множество кадров. Mamba обеспечивает непревзойденную линейную вычислительную сложность O(N) для длинных видеопоследовательностей, в отличие от квадратичной сложности $O(N^2)$ традиционных трансформеров (ViT). Это делает данные модели невероятно легкими (модель GMRD весит

всего 1.8 МБ) и идеальными для видео-инференса на периферийных устройствах (edge hardware, например Jetson Orin Nano) со скоростью 45 FPS при потреблении всего 7 Вт [95]. Внедрение специальной функции потерь консистенции (Consistency Loss), основанной на дивергенции Кульбака-Лейблера (KL divergence), математически принуждает сеть выдавать стабильные, плавные предсказания оценок уверенности между соседними кадрами [60]. Для количественной оценки этой стабильности авторы ввели формальную метрику – Оценку временной согласованности (Temporal Consistency Score, TCS), которая вычисляется как отношение среднего значения уверенности предсказания (μ_c) к его стандартному отклонению (σ_c) на протяжении отслеживаемой траектории объекта: $TCS = \frac{\mu_c}{\sigma_c}$. Чем выше TCS, тем меньше «мерцание» (variance) и тем выше надежность видеоаналитики в полевых условиях [106].

1.7 Общедоступные наборы данных для обнаружения дорожного дефекта

Развитие, качество и способность алгоритмов глубокого обучения к генерализации (обобщению) напрямую и критически зависят от качества, объема и разнообразия аннотированных данных, на которых они обучаются (правило "Garbage in, garbage out") [49]. В начале и середине 2010-х годов исследования в области автоматического обнаружения повреждений дорог жестоко страдали от нехватки публичных, репрезентативных массивов данных [47, 117]. Ученые были вынуждены собирать и размечать собственные, закрытые, небольшие наборы (обычно объемом от 200 до 500 изображений), снятые в тепличных условиях одной страны, одной камерой, при идеальном освещении [118]. Это порождало феномен переобучения (overfitting) и делало практически невозможным объективное, перекрестное сравнение различных архитектур нейросетей между собой.

Для решения этого фундаментального барьера в последние годы ведущими университетами и консорциумами был выпущен ряд крупномасштабных, стандартизированных публичных датасетов, ставших катализатором прорывных исследований [49, 119]. Наиболее масштабной и исторически значимой глобальной инициативой стал проект по краудсенсингу данных – Global Road Damage Detection Challenge (GRDDC / CRDDC), организованный в рамках престижного кубка IEEE Big Data Cup. Семейство наборов данных RDD (Road Damage Dataset) планомерно эволюционировало и расширялось из года в год [47], который продемонстрирован в рисунке 9.

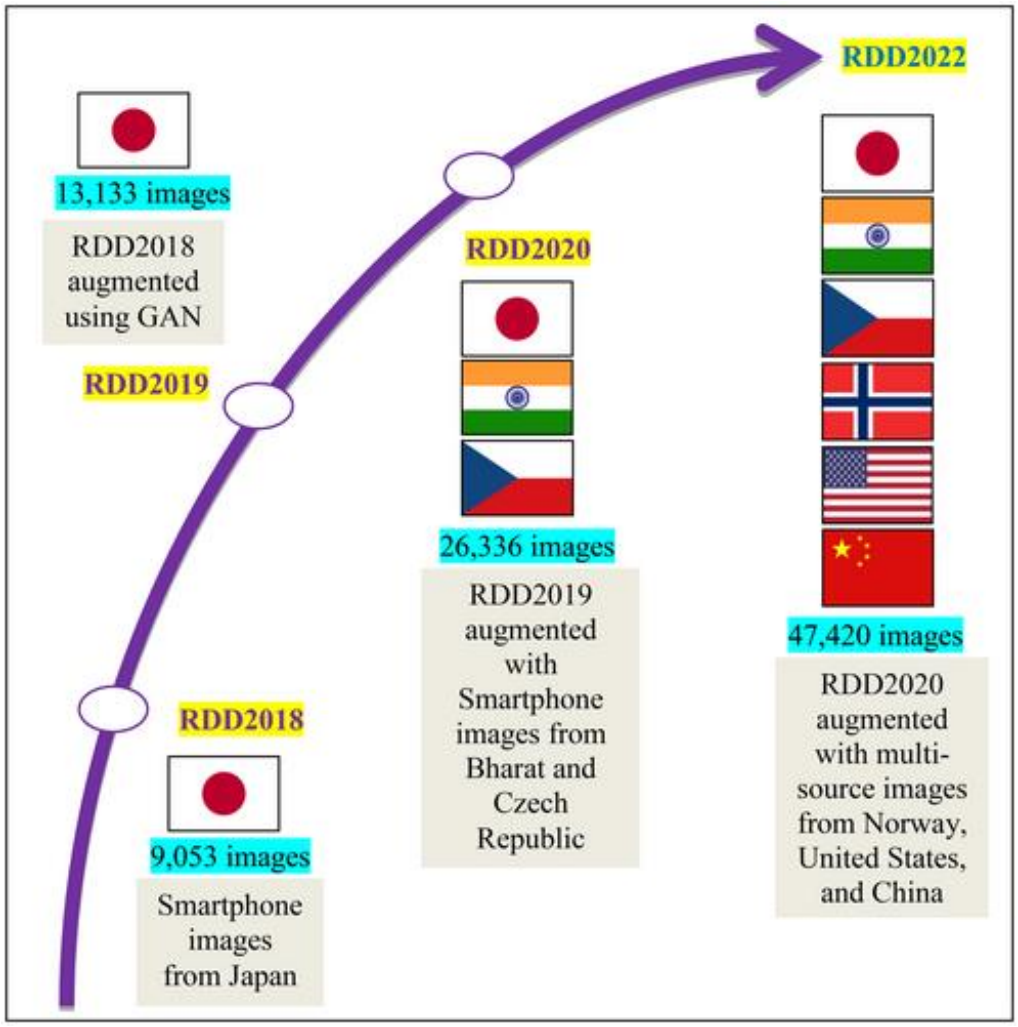


Рисунок 9 – Эволюция датасета RDD

RDD2018: первый публичный релиз, содержащий 9 053 изображения дорог исключительно из Японии, с 15 435 аннотациями дефектов (bounding boxes), распределенных по 8 классам повреждений. Датасет стал пионером в стандартизации задач детекции [47, 120].

RDD2019: обновленная версия, в которой японский набор был очищен и расширен до 13 135 изображений и 30 989 высококачественных аннотаций [121].

RDD2020: гигантский скачок в масштабировании [119]. Датасет вырос до 26 336 изображений, снятых преимущественно с помощью смартфонов, установленных в автомобилях. Ключевой инновацией стало включение данных из трех различных стран: Японии, Индии и Чехии. Это впервые позволило исследователям обучать модели, устойчивые к географическому «смещению домена» (разным типам дорог, цветам асфальта и погодным условиям). Датасет сфокусировался на четырех макро-категориях дефектов: D00 (продольные трещины), D10 (поперечные трещины), D20 (сетчатые/усталостные трещины) и D40 (выбоины) [47].

RDD2022: на сегодняшний день это самый монументальный, многонациональный и часто используемый в литературе бенчмарк [122]. Он включает 47 420 изображений и более 55 000 ограничивающих рамок.

География расширилась до шести стран: Япония (13 133 фото), Индия (9 665), Норвегия (10 201), США (6 005), Чехия (3 538) и Китай [123]. Важнейшей методологической особенностью RDD2022 является мультисенсорность: изображения были получены не только с dashcam-смартфонов, но и с панорамных камер Google Street View, мотоциклов (China_MotorBike, 2 477 фото) и, что крайне важно для новых исследований, с дронов (БПЛА) (China_Drone, 2 401 фото) [121]. Разрешение исходных изображений варьируется в колоссальном диапазоне от 512x512 и 600x600 до ультравысокого 3650x2044 пикселей [121].

N-RDD2024: новейшая модификация RDD2022, выпущенная для челленджа ORDDC'2024 [124, 125]. В этой версии была проведена глубокая ре-аннотация, и количество классов увеличено с 4 до 10. К классическим трещинам и выбоинам добавились такие критически важные семантические категории, как D30 (отремонтированные трещины / заплатки), D50 (размытие пешеходных переходов), D60 (размытие дорожной разметки / линий полос), D70 (крышки канализационных люков), D80 (участки ямочного ремонта) и D90 (колееобразование). Это позволяет создавать невероятно сложные модели всестороннего анализа семантики дорожной среды [126].

Помимо RDD (ориентированного на ограничивающие рамки - Bounding Box Object Detection), в научном сообществе существует острая потребность в наборах данных для семантической сегментации на уровне пикселей [127, 128]. Это необходимо для вычисления точной ширины и площади дефектов. В Таблице 3 представлены наиболее значимые публичные датасеты.

Таблица 3 – Публичные датасеты

Название набора данных	**Описание, размер и технические характеристики**	**Основное применение**
Crack500	Содержит 676 изображений с разрешением 640x360 пикселей. Каждый пиксель трещины размечен вручную (676 аннотаций линейных трещин) [129].	Идеально подходит для оценки семантической сегментации линейных и мелких трещин на асфальте.
CFD (Crack Forest Dataset)	Включает 566 изображений (480x320 пикселей) с мелкозернистой аннотацией трещин (526 линейных и 40 блочных). Специфика датасета – сильная зашумленность (наличие теней, пятен нефти, листьев, водного блеска), что делает его сложным бенчмарком [130].	Обучение алгоритмов подавления теней и сложного фонового шума; пиксельная сегментация [130].
GAPs384 (German Asphalt Pavement Distress)	Состоит из 509 изображений высокого разрешения (например, 540x640) немецких дорог с аннотацией трещин (502 линейных, 7 блочных) и участков ямочного ремонта (Applied patch) [131].	Многоклассовая классификация и сегментация европейских стандартов асфальта [132].
HighRPD (High-Altitude Drone Dataset)	Современный гигантский набор данных (2024-2025 гг.), собранный с помощью БПЛА (DJI M300 с камерой Zenmuse P1) с высоты 50 метров. Оригинальные снимки (8192x5460) нарезаны на 11 696 однородных тайлов 640x640. Содержит более 22 000 аннотаций (12 365 линейных трещин, 8 239 блочных структур, 1 412 выбоин) в формате YOLO [129].	Обучение моделей YOLO для крупномасштабного топологического картографирования повреждений с дронов (aerial surveys).
SVRDD (Street View Image Dataset)	Сформирован из 8 000 панорамных изображений Baidu Maps Street View (район Пекина). Содержит свыше 20 000 визуально распознанных аннотаций повреждений [118].	Первый датасет для обнаружения дефектов на основе уличных панорам; адаптация домена [118]

Продолжение таблицы 3

PaveDistress	Датасет высокого разрешения, собранный с помощью линейных камер (Basler raL2048) и инфракрасного лазерного освещения (интервал 1 мм, разрешение 3854x2065 пикселей). Включает съемки в сложных условиях (сумерки, туннели, пасмурная погода) [133].	Построение ультра-прецизионных систем обнаружения миллиметровых трещин [133].

В сфере **анализа видеопоследовательностей (Video-based detection)** ситуация остается сложной. Доминирующим остается набор данных **VIL-100 (Video Lane 100)**, который предлагает 100 видеороликов (10 000 кадров с аннотацией) для трекинга дорожной разметки и границ в динамике различных условий (блики, тени, перекрестки) [134]. Однако, крупных стандартизированных видео-датасетов (на уровне тысяч видеороликов) с покадровой аннотацией именно выбоин и трещин (Video Object Detection for Pavement Distress) до сих пор катастрофически не хватает в открытом доступе. Это вынуждает исследователей либо собирать собственные мелкие видео-наборы [106], либо искусственно симулировать движение транспортного средства (applying geometric transformations, translations) на основе статических изображений из RDD2022 (создавая pseudo-frames) для тестирования алгоритмов временной согласованности [106]. Также перспективным направлением становится создание мультимодальных бенчмарков, таких как **RoadBench** (100K изображений), которые впервые объединяют высокоразрешенные визуальные дефекты с подробнейшими текстовыми описаниями, что позволяет обучать гигантские мультимодальные модели (VLM – Vision-Language Models, такие как RoadCLIP или модификации LLaMA) для семантического понимания дорожных сцен на естественном языке [95].

1.8 Проблемы и открытые исследовательские вопросы

Несмотря на очевидный, колоссальный прорыв в метриках точности обнаружения, перевод глубокого обучения из стерильных лабораторных условий в реальные промышленные системы мониторинга дорог (Real-world Deployment) наталкивается на ряд фундаментальных инженерных, вычислительных и концептуальных барьеров [135]. Анализ литературы 2024-2025 годов позволяет выделить несколько наиболее острых, нерешенных исследовательских вопросов [136, 137].

Аппаратные ограничения и развертывание на границе сети (Edge Computing constraints). В современных архитектурах автоматизированного мониторинга на базе автомобилей или БПЛА требуется обработка видео высокого разрешения в режиме жесткого реального времени (от 25 до 60 кадров в секунду, FPS) [91]. Однако развертывание тяжеловесных, громоздких моделей (таких как Faster R-CNN, ViT или ресурсоемкие ансамбли трансформеров) на встраиваемых маломощных платах (edge devices), таких как Raspberry Pi 4 или NVIDIA Jetson Nano, приводит к критическим физическим проблемам [138]. Экспериментальные исследования показывают,

что при длительном инференсе тяжелых нейросетей на Jetson Nano (с его ограниченными 4 ГБ ОЗУ) возникает сильный перегрев процессора (thermal throttling). Это вызывает принудительное снижение тактовой частоты, в результате чего FPS катастрофически падает (например, с 25 до 3 FPS для неоптимизированной YOLOv7), а нехватка памяти приводит к процессам подкачки (memory swapping) или полному краху операционной системы (crashes), особенно при работе в жарких наружных условиях [36]. *Открытый вопрос:* Как эффективно сжать глубокие сети без потери точности обнаружения мелких трещин? Существует острая потребность в исследованиях по структурному сжатию нейросетей: применению низкобитового квантования (quantization-aware training в формат INT8), факторизации матриц малого ранга (low-rank factorization), обрезке каналов свертки (channel pruning) и глубокой оптимизации через среды исполнения (TensorRT/ONNX) [31]. Разработка сверхлегких архитектур, сохраняющих робастность в видео, остается приоритетом.

Дисбаланс данных и смещение домена (Data Imbalance & Domain Shift). В реальном мире наблюдается феномен «длинного хвоста» (long-tail distribution): тяжелые, катастрофические разрушения (огромные выбоины, провалы асфальта) встречаются значительно реже, чем мелкие, тривиальные трещины [118, 139]. Этот естественный сильный дисбаланс классов в обучающих наборах данных (даже в RDD2022) приводит к серьезной предвзятости модели (algorithmic bias): нейросеть великолепно детектирует трещины, но демонстрирует недопустимо низкую полноту (low recall) при распознавании критически опасных выбоин, классифицируя их как фон [118]. Использование генеративно-состязательных сетей (GAN) для синтеза реалистичных изображений редких дефектов (например, архитектуры AsphaltGAN или PCGAN) является одним из новейших способов искусственного расширения и балансировки датасетов (Data Augmentation), однако генерация топологически сложных трещин все еще далека от идеала [83]. Еще более серьезной проблемой является «смещение домена» (Domain Shift) и чувствительность к факторам окружающей среды (Environmental Variability). Дорожные покрытия подвержены экстремальной визуальной вариативности [140]. Модель, великолепно обученная на сухих, солнечных, темных дорогах Японии (из набора RDD2018), может продемонстрировать катастрофическое падение точности (вплоть до 0% обнаружения) при тестировании на грязных дорогах Индии, влажных от дождя трассах Великобритании или заснеженных дорогах Норвегии [33]. Резкие падающие тени от деревьев, блики ослепляющего солнца, лужи, свежая дорожная разметка и следы экстренного торможения часто классифицируются моделями как ложные трещины (False Positives) [141]. Недавние исследования 2025 года подчеркивают крайнюю уязвимость даже самых передовых моделей (таких как Gemini AI) в зашумленных «полевых» (field-use) условиях без предварительной доменной адаптации [142]. *Открытый вопрос:* Требуются масштабные исследования в области адаптации домена (Domain Adaptation), федеративного обучения (Federated Learning) и Zero-Shot Learning (с

применением мультимодальных VLMs) для создания глобально универсальных моделей, способных распознавать дефекты в любых погодных и географических условиях [118].

Интеграция мультимодальных датчиков (Multi-modal Sensor Fusion). Использование исключительно 2D-камер (визуального RGB-спектра), несмотря на всю их дешевизну, имеет непреодолимое физическое ограничение: камеры не фиксируют метрическую информацию о глубине (Depth) [143]. Это делает математически невозможной прямую оценку степени тяжести (Severity Level) колееобразования (rutting) или измерение истинного объема и глубины выбоины, что необходимо инженерам для приоритизации срочного ремонта (High Severity classification) [143]. Интеграция данных с нескольких гетерогенных сенсоров – RGB-камер, инфракрасных тепловизоров (для обнаружения скрытой влаги), 3D-LiDAR, лазерных сканеров профиля, георадаров (GPR) и инерциальных датчиков акселерометров – признана академическим сообществом как наиболее перспективное, ультимативное решение (Hybrid and 3D Methods) [144, 145]. *Открытый вопрос:* Слияние гетерогенных данных (Data Fusion) порождает колоссальные трудности. Возникают сложнейшие задачи строгой пространственно-временной (spatio-temporal) калибровки датчиков, синхронизации потоков, работающих с принципиально разной частотой дискретизации, и огромного увеличения вычислительной сложности конвейеров (Pipeline complexity) [35]. Разработка эффективных архитектур для Sensor Fusion на периферийных устройствах (включая методы Bird’s-Eye-View (BEV) репрезентаций) является одним из главных вызовов ближайших лет [146].

1.9 Резюме главы

Резюмируя проведенный масштабный анализ современной научной литературы (более 100+ релевантных источников), можно констатировать, что технологии обнаружения повреждений автомобильных дорог прошли уверенный, впечатляющий эволюционный путь. Этот путь пролегал от медленных, субъективных ручных инспекций и консервативных алгоритмов компьютерного зрения на основе ручного конструирования признаков (математические фильтры границ, анализ текстур и классический SVM/RF) к доминированию мощных архитектур глубокого обучения (многослойные сверточные сети, детекторы YOLO, U-Net, Vision Transformers) [147]. За последнее десятилетие (2015–2025 гг.) достигнут беспрецедентный прорыв в абсолютных метриках точности классификации и локализации дефектов на статических (2D) изображениях. Этот прогресс был многократно ускорен и поддержан созданием и публикацией глобальных, многонациональных наборов данных, таких как серия RDD (RDD2020, RDD2022, N-RDD2024), которые установили единый золотой стандарт (benchmark) для всего мирового сообщества исследователей и дата-саентистов [121].

Тем не менее, несмотря на триумф в обработке статических фотографий, критический анализ современной литературы выявляет несколько

фундаментальных пробелов (Research Gaps), препятствующих повсеместному промышленному развертыванию этих систем в реальном времени.

Недостаточная эксплуатация временно́й динамики в видео (Spatio-Temporal Gap): подавляющее большинство существующих в литературе систем все еще обрабатывают видеопотоки из движущегося автомобиля наивным образом – как набор разрозненных, независимых статических кадров [95]. Игнорирование временно́й и кинематической связи между фреймами приводит к крайней нестабильности (мерцанию) детектирования, дублированию подсчета объектов и множеству ложных срабатываний в динамичных, сложных погодных условиях (дождь, смена освещения). Фундаментальные исследования, посвященные глубокой интеграции пространственно-временного контекста с помощью ConvLSTM, оптического потока, трекинга (SORT) или инновационных архитектур Mamba (State-Space Models) для обеспечения плавной временно́й согласованности (temporal consistency) предсказаний, находятся лишь на начальной стадии своего развития и требуют существенного углубления [95].

Дефицит видео-ориентированных наборов данных (Dataset Gap): в то время как для статических изображений существуют гигантские размеченные базы (RDD2022, 47k+ изображений), в академическом пространстве наблюдается острый, катастрофический дефицит открытых масштабных баз видеоданных (с высокой частотой кадров) с плотной, покадровой аннотацией дефектов дорожного покрытия [106]. Это делает невозможным честную, объективную оценку алгоритмов видеотрекинга и моделей временной согласованности, заставляя авторов прибегать к синтетическим ухищрениям (псевдо-фреймы) [106].

Противоречие между сложностью моделей и периферийными вычислениями (Edge Deployment Gap): существует непреодолимый разрыв между вычислительными требованиями тяжеловесных современных моделей (обеспечивающих State-of-the-Art точность, таких как ансамбли трансформеров или Faster R-CNN) и жесткими физическими ограничениями встраиваемых периферийных плат (edge devices, БПЛА, смартфонов) в плане памяти, энергопотребления и отвода тепла [31]. Создание оптимизированных, сверхлегких моделей (путем квантования, прунинга или использования архитектур типа Mamba/MobileNet), способных обрабатывать плотные видеопоследовательности со скоростью 30-60 FPS без потери контекстной информации о микротрещинах, является критически важной и не до конца решенной инженерной проблемой.

Ограниченность мономодального подхода (Quantification Gap): 2D-видеоаналитика отлично справляется с задачей обнаружения (Detection), но фундаментально ограничена в задаче точной оценки трехмерной степени тяжести (Severity quantification), например, глубины выбоины, без использования сложных методов стереозрения или интеграции дополнительных датчиков (LiDAR, GPR) [35].

Таким образом, разработка робастной, гибридной системы обнаружения повреждений автомобильных дорог на основе видеоданных, которая

эффективно синтезирует пространственно-временную информацию для стабилизации трекинга, устойчива к динамическим изменениям окружающей среды (Domain Shift) и математически оптимизирована для сверхбыстрого инференса в реальном времени на бортовых периферийных устройствах, является наиболее актуальной, нерешенной и остро востребованной научной задачей в современной парадигме интеллектуальных транспортных систем (ITS). Представленное диссертационное исследование направлено на комплексное решение именно этих критических проблем.

2 РАЗРАБОТКА И ПРОЕКТИРОВАНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ОБНАРУЖЕНИЯ ПОВРЕЖДЕНИЙ ДОРОГ

В этой главе детально раскрывается методологическая база исследования и описывается созданный нами аппаратно-программный комплекс для комплексного анализа дорожных дефектов. Описание выстроено по принципу полного жизненного цикла обработки визуальных потоков. Сначала представляем протоколы полевого сбора видеоданных и алгоритмы их первичной обработки. Здесь же описывается методика разметки (аннотации) изображений, которая позволила нам сформировать эталонную обучающую выборку.

Ключевым звеном главы является описание архитектуры нашей нейросети – TCR-RoadNet. Подробно обосновывается математика многомасштабной сверточной магистрали и внедрение блока трансформерного внимания для уточнения контекста. Особое внимание уделено механизмам децентрализованного вывода, которые обеспечивают высокую скорость работы системы. Завершает раздел строгая формализация процесса обучения: от выбора гибридных функций потерь до обоснования метрик качества. Также показывается, как обученная модель была интегрирована в итоговое веб-приложение для конечного пользователя.

2.1 Концептуальная архитектура предлагаемой системы мониторинга

Разрабатываемый программно-аппаратный комплекс функционирует как сквозное интеллектуальное решение для автоматизации мониторинга дорожной инфраструктуры посредством потокового анализа видеоданных. Архитектура системы спроектирована таким образом, чтобы обеспечивать полный цикл обработки визуальной информации: от первичного захвата кадров с камер видеонаблюдения и мобильных устройств до автоматического обнаружения, классификации, сегментации и визуализации выявленных дефектов дорожного покрытия в пользовательском веб-интерфейсе. Роль базового вычислительного ядра выполняет разработанная многозадачная нейросетевая архитектура TCR-RoadNet, оптимизированная под высокоскоростную обработку видеопотока и выполнение задач детекции дорожных повреждений в режиме реального времени. Использование многозадачного подхода позволяет одновременно выполнять локализацию дефектов, определение их класса и формирование сегментационных масок, повышая информативность итогового анализа и устойчивость системы к сложным условиям съемки. Дополнительно в состав комплекса входят модули предварительной обработки кадров, фильтрации визуального шума, передачи результатов анализа, географической привязки обнаруженных дефектов и формирования аналитической отчетности. Взаимодействие отражено на Рисунке 10, демонстрирующем последовательную передачу данных между функциональными узлами вычислительного конвейера.

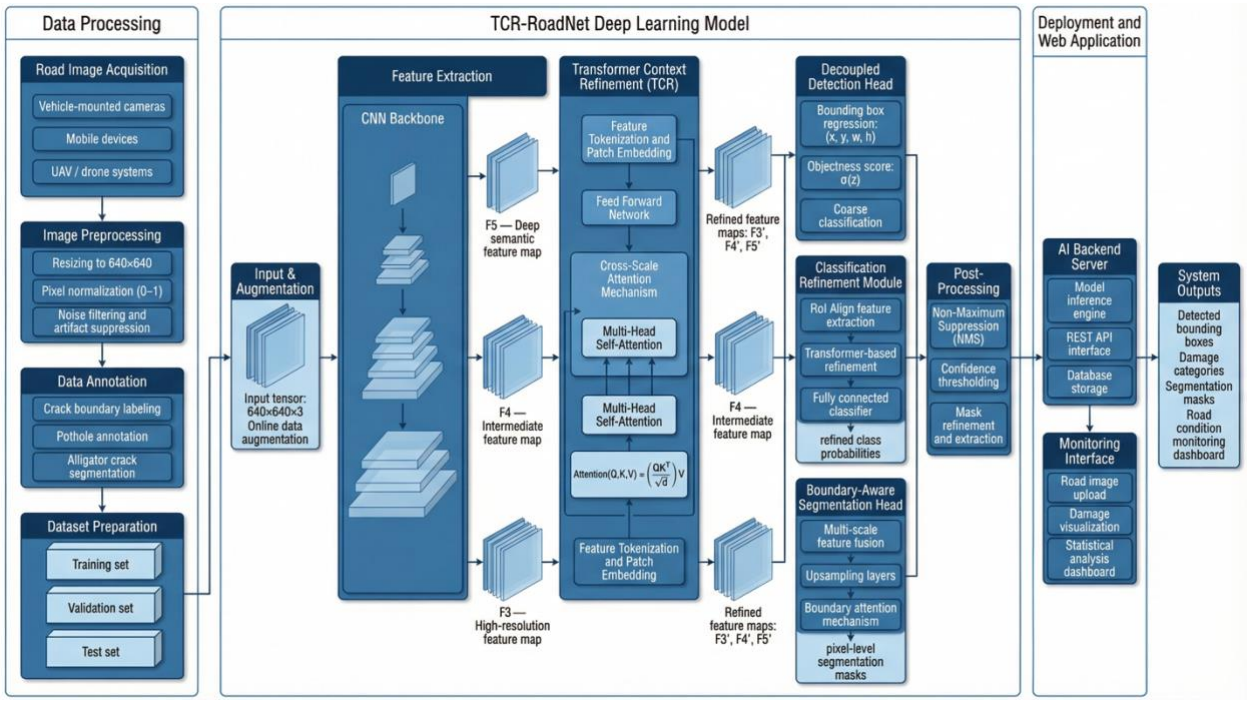


Рисунок 10 – Общая архитектура комплексной системы обнаружения повреждений дорог

Работа системы инициируется модулем Data Processing, основная задача которого заключается в приеме и унификации входящих визуальных потоков. В качестве источников данных могут выступать как мобильные устройства и штатные автомобильные регистраторы, так и камеры беспилотных летательных аппаратов. Для обеспечения архитектурной стабильности нейросети задействован конвейер предобработки (Image Preprocessing). На этом этапе каждый извлеченный видеокадр приводится к базовому разрешению 640×640 пикселей. Параллельно с изменением геометрии кадра выполняется нормализация пиксельных значений в интервале [0,1] и программная фильтрация шумов. Это позволяет нивелировать дефекты сжатия видеопотока и минимизировать влияние неблагоприятных погодных факторов на качество детекции. При подготовке системы к эксплуатации данный блок дополняется этапами аннотирования и структурирования датасетов. Здесь в ходе экспертной разметки формируются сбалансированные обучающие, валидационные и тестовые выборки, включающие детальные контуры трещин, локальные выбоины и сложные участки с сеткой усталостных разрушений.

Центральным вычислительным компонентом системы выступает модель глубокого обучения TCR-RoadNet. Перед подачей в сеть нормализованные кадры подвергаются динамической аугментации для повышения надежности модели при изменениях освещенности и ракурса съемки. Процесс извлечения признаков начинается в многомасштабной сверточной магистрали (CNN Backbone), которая генерирует иерархическую пирамиду пространственных признаков, постепенно снижая разрешение и увеличивая семантическую плотность данных. Для преодоления ограничений стандартных сверток в моделировании долгосрочных зависимостей

извлеченные карты признаков передаются в модуль контекстного уточнения на основе трансформера (Transformer Context Refinement). В этом блоке пространственные характеристики токенизируются и обрабатываются через многомасштабные блоки внимания, что позволяет сети динамически связывать локальные текстуры дефектов с глобальным контекстом дорожной сцены.

Обогащенные контекстом признаки параллельно распределяются между тремя специализированными модулями вывода, которые функционируют синхронно для решения различных задач компьютерного зрения. Модуль раздельного обнаружения (Decoupled Detection Head) независимо прогнозирует координаты ограничивающих рамок, оценивает общую вероятность наличия объекта и выполняет базовую классификацию. В свою очередь, модуль уточнения классификации (Classification Refinement Head) осуществляет повторный анализ выявленных областей интереса для точной дискриминации между визуально схожими типами повреждений, такими как продольные и поперечные трещины. Одновременно с этим модуль сегментации с учетом границ (Boundary-Aware Segmentation Head) генерирует детализированные бинарные маски дефектов, акцентируя внимание на точности контуров для последующей оценки физической площади разрушения. Результаты работы нейронной сети проходят через блок постобработки, где алгоритм подавления немаксимумов (NMS) устраняет избыточные пересекающиеся рамки, а пороговая обработка завершает формирование итоговых сегментационных масок.

Финальный этап работы системы заключается в интеграции полученных аналитических данных в функциональное веб-приложение, предназначенное для мониторинга инфраструктуры. Глубокая нейронная сеть развертывается на внутреннем сервере с искусственным интеллектом (AI Backend Server), где специализированный программный интерфейс (API) обрабатывает запросы на логический вывод и обеспечивает систематическое сохранение результатов в базе данных. Взаимодействие оператора с системой осуществляется через веб-интерфейс мониторинга, который предоставляет инструменты для загрузки дорожных видеоданных, визуализации обнаруженных повреждений на карте и анализа статистической информации. Итоговые выходные данные системы объединяют локализованные ограничивающие рамки, классы дефектов и высокоточные сегментационные маски на единой информационной панели, обеспечивая тем самым автоматизированный и масштабируемый инструмент для оценки состояния транспортных сетей в реальных условиях эксплуатации.

2.2 Сбор видеоданных и подготовка набора данных для обучения

Эффективность работы архитектуры глубокого обучения напрямую зависит от качества, объема и репрезентативности используемой обучающей выборки. В рамках данного исследования формирование набора данных осуществлялось на основе видеоматериалов, фиксирующих состояние дорожного покрытия в реальных условиях эксплуатации. Этот подход

позволяет захватить широкую вариативность дефектов, изменений освещенности и текстурных особенностей асфальта, что критически важно для обучения надежной модели обнаружения и сегментации.

2.2.1 Настройка сбора видеоданных

Для обеспечения стандартизации и высокого качества собираемого обучающего материала был разработан строгий протокол видеофиксации. В качестве основного оборудования использовался смартфон с камерой высокого разрешения, надежно закрепленный на лобовом стекле автомобиля с помощью специализированного жесткого держателя, как показано на рисунке 11. Такое пространственное расположение устройства в горизонтальной ориентации (с соотношением сторон 16:9) обеспечивало стабильный угол обзора, оптимальный для захвата максимальной площади дорожного полотна. При установке камеры особое внимание уделялось кадрированию сцены: из поля зрения максимально исключались элементы кузова, такие как капот, а также детали салона автомобиля. Непременным техническим условием перед началом каждой сессии записи являлась тщательная очистка лобового стекла для предотвращения появления оптических артефактов и искажений на видео.

Рисунок 11 – Расположение мобильного устройства на лобовом стекле автомобиля для сбора видеоданных

Съемка производилась в стандартном непрерывном режиме видеозаписи, исключая применение интервальной съемки (таймлапс). Для достижения необходимой детализации структурных дефектов асфальтобетонного покрытия было установлено разрешение Full HD

(1920×1080 пикселей) при кадровой частоте от 30 до 60 кадров в секунду. Обязательным требованием являлась активация встроенных систем электронной или оптической стабилизации изображения (EIS/OIS) для компенсации вибраций кузова при движении по неровным участкам дороги. Видеоданные сохранялись в современных контейнерах форматов MP4, MOV или HEVC без использования каких-либо программных фильтров цветокоррекции, что гарантировало получение максимально естественного визуального материала.

Помимо технических настроек оборудования, протокол строго регламентировал условия движения транспортного средства и параметры окружающей среды. Для минимизации эффекта пространственного размытия (motion blur) оптимальная скорость автомобиля поддерживалась в диапазоне от 20 до 40 км/ч, при этом водитель избегал резких ускорений и торможений. Сбор данных осуществлялся преимущественно в дневное время суток, однако для обеспечения репрезентативности обучающей выборки в датасет также были включены кадры, зафиксированные в условиях раннего вечера при сниженном уровне естественного освещения. Метеорологические условия съемок охватывали широкий спектр реальных сценариев: ясную солнечную погоду с наличием резких падающих теней, пасмурное небо с рассеянным светом, а также ограниченное количество съемок в дождливых условиях. Включение дождливых сценариев и вечерних кадров было целенаправленным методологическим шагом для повышения робастности нейронной сети. Это позволило алгоритму научиться абстрагироваться от сложных визуальных помех, таких как изменение отражающей способности мокрого дорожного покрытия, световые блики и общее снижение контрастности, что максимально приближает работу системы к неконтролируемым условиям реальной эксплуатации.

2.2.2 Структура набора данных

В результате масштабных полевых работ, выполненных в строгом соответствии с разработанным протоколом, был сформирован репрезентативный массив видеоматериалов. Итоговая база включает 30 видеозаписей общей продолжительностью 80 часов. Поскольку предложенная нейросетевая архитектура требует покадровой обработки визуальной информации, весь собранный видеоряд был подвергнут процедуре извлечения фреймов. Для исключения информационной избыточности и минимизации риска переобучения модели на практически идентичных сценах, кадры извлекались с заданным интервалом. В соответствии с входными требованиями разрабатываемой архитектуры, все извлеченные кадры были предварительно обработаны и геометрически масштабированы до унифицированного пространственного разрешения 640×640 пикселей. Данный размер обеспечивает оптимальный компромисс между сохранением детализации мелких трещин и вычислительной эффективностью при обучении многомасштабной сверточной магистрали. Полученный массив изображений

впоследствии был разделен на обучающую, валидационную и тестовую выборки для корректной оценки метрик точности.

2.2.3 Процесс аннотирования

Формирование качественной эталонной разметки является наиболее ответственным этапом подготовки набора данных, так как именно от точности этих данных зависит способность нейронной сети к корректному обучению. Для разметки извлеченных кадров использовалось программное обеспечение с открытым исходным кодом Label Studio. Данное приложение представляет собой универсальную платформу, предоставляющую гибкий инструментарий для визуального аннотирования изображений под различные задачи машинного обучения.

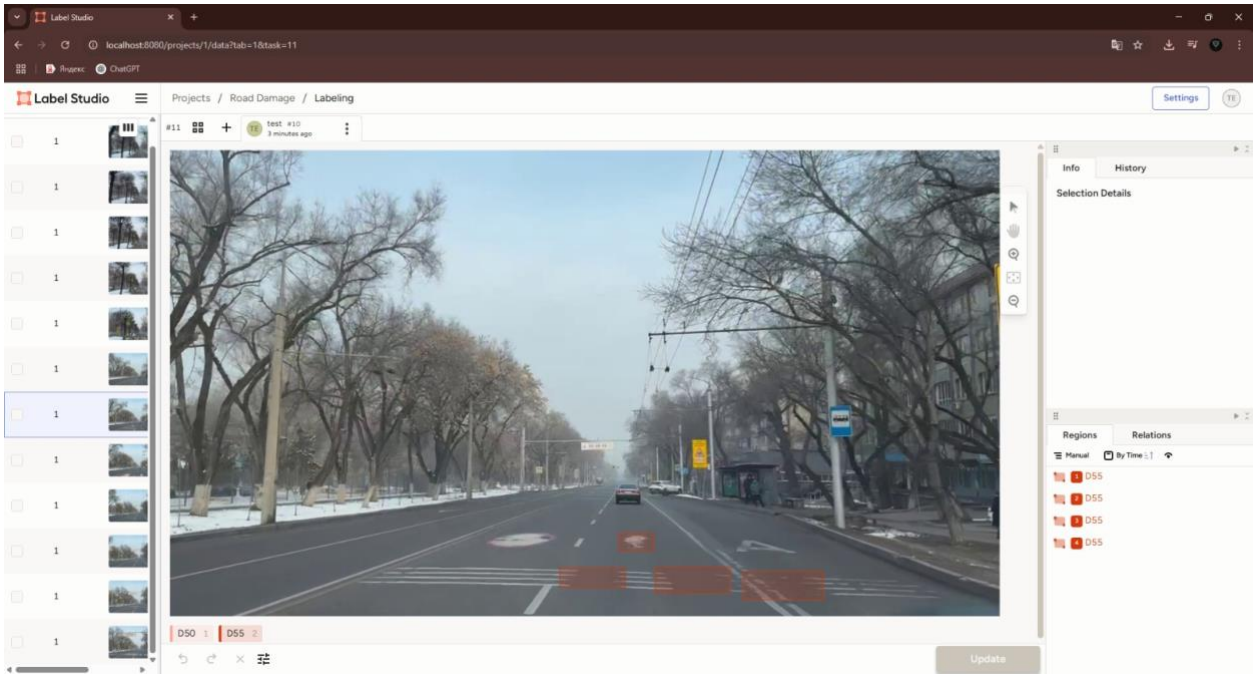


Рисунок 12 – Интерфейс приложения Label Studio в процессе аннотирования дорожных дефектов

В рамках данного исследования для настройки рабочей среды в Label Studio был выбран специализированный шаблон объектного обнаружения с использованием прямоугольных ограничивающих рамок (bbox object detection), что видно на рисунке 12. Этот подход позволяет операторам точно выделять проблемные области на изображении, что полностью соответствует требованиям к формату входных данных для разрабатываемой архитектуры TCR-RoadNet.

В процессе визуального анализа кадров производилась классификация и пространственная локализация дорожных дефектов в соответствии с утвержденной номенклатурой. Все анализируемые повреждения были систематизированы и разделены на три основные макрогруппы: трещины, разрушения и повреждения дорожной разметки. Группа трещин включает в себя линейные (Linear crack, идентификатор D10) и продольные трещины

(Longitudinal crack, D20). К категории макроразрушений отнесены выбоины (Potholes, D30) и участки разрушенного асфальтного покрытия (Destroyed asphalt, D40). В специализированную группу дефектов дорожных знаков и разметки вошли растрескавшиеся пешеходные переходы (Cracked cross walk, D50) и поврежденные белые линии (Cracked white line, D55). Полная классификация используемых типов повреждений и их уникальных идентификаторов представлена в таблице 4.

Таблица 4 – Классификация типов повреждений дорожного покрытия

Группа	**Тип повреждения**	**Номер повреждения**
Трещины	Линейные трещины	D10
	Продольные трещины	D20
Разрушения	Выбоины	D30
	Разрушенный асфальт	D40
Повреждения дорожной разметки	Растрескавшиеся пешеходные переходы	D50
	Поврежденные белые линии	D55

Каждому идентифицированному объекту на видеокадре присваивалась соответствующая метка класса согласно утвержденной классификации, после чего программно фиксировались точные пространственные координаты его ограничивающей рамки. Детальный и стандартизированный подход к выделению границ дефектов позволил существенно снизить уровень информационного шума в обучающей выборке. По завершении процесса покадровой разметки сформированный массив данных экспортировался для последующего использования в вычислительном конвейере. Базовый функционал приложения Label Studio позволил получить на выходе готовую разметку в виде стандартизированных файлов XML-аннотаций либо в специализированном текстовом формате YOLO. Полученный формат выходных данных обеспечил высокую совместимость и бесшовную интеграцию подготовленного датасета с алгоритмами оптимизации многозадачной нейронной сети в процессе ее обучения.

2.3 Алгоритмы предварительной обработки и аугментации данных

Предварительная подготовка данных является критически важным этапом в конвейере машинного обучения, напрямую определяющим качество и скорость сходимости глубоких нейронных сетей. Поскольку исходным источником визуальной информации в данном исследовании выступают непрерывные видеопотоки, процесс подготовки включает в себя последовательность операций: извлечение статических кадров, их геометрическое и фотометрическое преобразование, устранение дисбаланса в распределении классов и финальное форматирование аннотаций для совместимости с многозадачной архитектурой сети.

2.3.1 Извлечение кадров из видео

Для преобразования собранного видеоматериала в набор статических обучающих изображений был разработан автоматизированный алгоритм на языке программирования Python с использованием открытой библиотеки компьютерного зрения OpenCV. Основной задачей данного модуля является последовательное чтение видеофайлов и циклическое сохранение строго определенных кадров. Анализ логики алгоритма показывает, что частота извлечения составляет ровно один кадр в секунду (1 FPS). Программная реализация считывает метаданные видеофайла, вычисляет его общую продолжительность в секундах и с помощью функции позиционирования перемещается по временной шкале с шагом в 1000 миллисекунд. Данный подход гарантирует высокую визуальную уникальность соседних изображений и предотвращает переобучение сверточной сети на практически идентичных дублирующихся данных, что неизбежно произошло бы при сохранении всех 30–60 кадров каждой секунды оригинального видеоряда.

2.3.2 Нормализация и изменение размера изображения

В соответствии с архитектурными требованиями разрабатываемой модели TCR-RoadNet, все извлеченные кадры подвергаются процедуре строгого пространственного масштабирования. Исходные изображения высокого разрешения геометрически трансформируются к единому тензорному размеру 640×640 пикселей перед подачей в сверточную магистраль нейронной сети. Данный размер является оптимальным компромиссом, сохраняющим мелкие текстуры трещин и обеспечивающим приемлемую вычислительную нагрузку. После изменения пространственного разрешения выполняется процесс фотометрической нормализации. Значения интенсивности цветовых каналов пикселей, изначально находящиеся в целочисленном диапазоне от 0 до 255, масштабируются в непрерывный нормализованный диапазон от 0 до 1. Эта математическая трансформация стабилизирует градиенты во время обратного распространения ошибки, предотвращая их затухание или взрыв, и значительно ускоряет сходимость оптимизационных алгоритмов при обучении модели.

2.3.3 Обработка дисбаланса классов

В процессе формирования обучающей выборки для задач обнаружения дефектов часто возникает проблема неравномерного распределения примеров между различными категориями. Первичный анализ базовых данных (включая глобальный датасет RDD2022) демонстрирует естественный дисбаланс: например, продольные трещины встречаются значительно чаще и имеют большую площадь покрытия, чем локализованные выбоины. Для компенсации этого дисбаланса и повышения робастности системы к локальным дорожным условиям была применена стратегия целенаправленного обогащения данных. Массив был целенаправленно расширен за счет включения локальных видеоданных со строгим добавлением изображений для каждой анализируемой категории. Кроме того, для

предотвращения доминирования мажоритарных классов на этапе обучения применяется динамическая аугментация данных, включающая геометрические искажения и фотометрические преобразования, что дополнительно балансирует выборку и повышает способность сети к генерализации.

2.3.4 Форматирование данных

Финальным этапом предварительной обработки является форматирование полученных аннотаций для обеспечения их строгой совместимости со специализированными модулями вывода архитектуры TCR-RoadNet. Результаты ручной эталонной разметки, экспортированные из программной среды Label Studio в формате XML, подвергаются автоматизированной конвертации. Для обеспечения корректной работы модуля раздельного обнаружения координаты прямоугольных ограничивающих рамок преобразуются в стандартизированный текстовый формат YOLO. Этот формат описывает положение каждого объекта через нормализованные координаты центральной точки, а также относительную ширину и высоту рамки. Одновременно с этим, сложные полигональные разметки конвертируются в набор бинарных изображений – сегментационных масок, которые выступают в качестве целевых переменных для обучения ветви сегментации с учетом границ. Полученный структурированный набор данных, состоящий из нормализованных изображений, текстовых файлов локализации и графических бинарных масок, формирует итоговый тензор обучающей выборки для многозадачной нейронной сети.

2.4 Математическая постановка задачи многозадачного компьютерного зрения

В рамках разработки системы обнаружения повреждений дорожного покрытия с использованием методов глубокого обучения, основная проблема формализуется как многозадачная проблема компьютерного зрения, включающая одновременную локализацию, классификацию и попиксельную сегментацию объектов на основе визуальных данных. Поскольку исходными данными выступает видеопоток, его можно представить как дискретную во времени последовательность изображений (кадров). Пусть входной видеопоток задан множеством $V = \{I_1, I_2, \dots, I_T\}$, где каждый извлеченный в момент времени t кадр представляет собой трехмерный тензор $I_T \in \mathbb{R}^{H \times W \times C}$. В данной нотации H и W обозначают пространственные размеры изображения (высоту и ширину, нормализованные до 640×640 пикселей), а C=3 соответствует количеству цветовых каналов RGB.

Цель разрабатываемой архитектуры глубокого обучения, как показано на рисунке 13, заключается в поиске оптимальной функции отображения f_θ, параметризованной весами нейронной сети θ, которая преобразует каждый входной кадр I_t в набор целевых прогнозов y_t . Формально это преобразование можно записать как $y_t = f_\theta(I_t)$. Итоговый прогноз модели для каждого кадра

представляет собой кортеж $y_t = (B, C, M)$. Множество $B = \{b_i\}_{i=1}^{N}$ описывает предсказанные ограничивающие рамки для N обнаруженных дефектов, где каждый элемент $b_i = (x_i, y_i, w_i, h_i)$ содержит координаты центра, ширину и высоту рамки. Множество $C = \{c_i\}_{i=1}^{N}$ содержит векторы распределения вероятностей принадлежности обнаруженного региона к одной из заданных категорий повреждений дорожного полотна. Третий компонент M представляет собой набор пространственных бинарных сегментационных масок $M_k \in \{0,1\}^{H \times W}$ для каждого класса дефекта k, определяющих точные контуры разрушений на уровне пикселей.

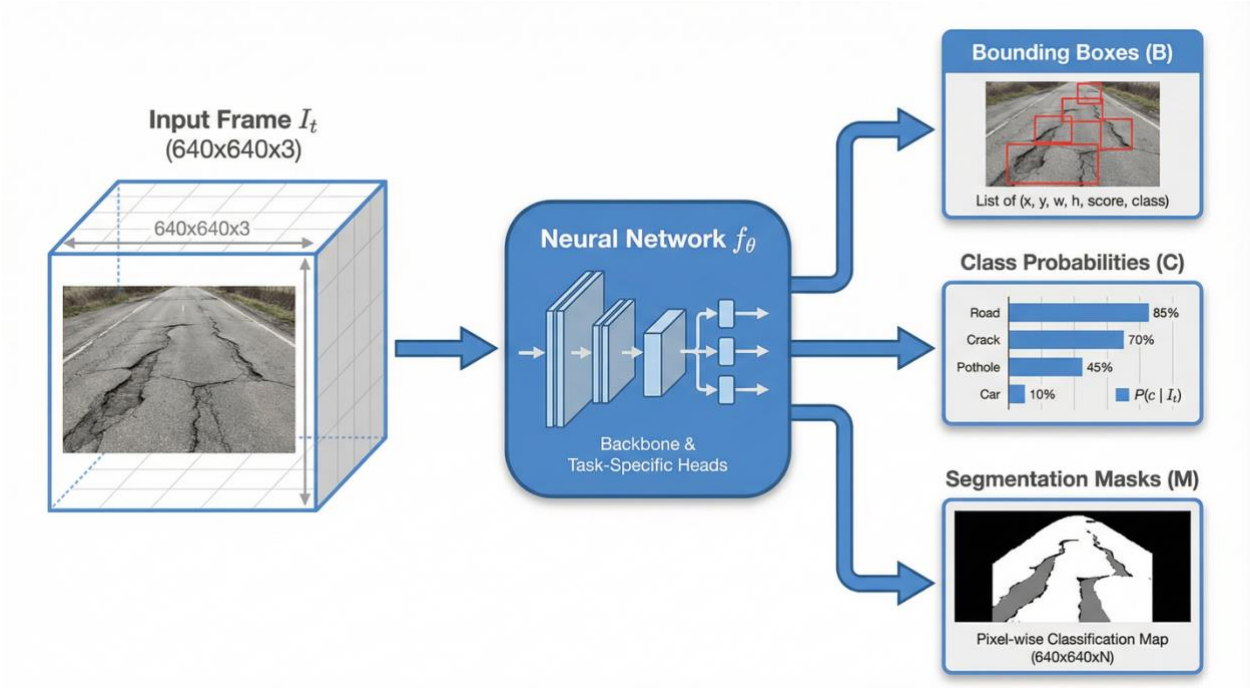


Рисунок 13 – Математическая абстракция отображения входного видеокадра в пространство многозадачных прогнозов

Для достижения требуемой точности работы функции f_θ необходимо определить математический критерий качества, который нейронная сеть будет минимизировать в процессе градиентного спуска. Поскольку предложенная система является многозадачной, глобальная функция потерь $\mathcal{L}_{total}$ формулируется как взвешенная линейная комбинация специфических функций потерь для каждой из решаемых подзадач. Общая целевая функция для оптимизации модели определяется следующим образом:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{box} + \lambda_2 \mathcal{L}_{obj} + \lambda_3 \mathcal{L}_{cls} + \lambda_4 \mathcal{L}_{ref} + \lambda_5 \mathcal{L}_{seg} + \lambda_6 \mathcal{L}_{bd} \qquad (1)$$

В данном уравнении $\mathcal{L}_{box}$ отвечает за ошибку регрессии координат ограничивающих рамок, $\mathcal{L}_{obj}$ оценивает точность предсказания наличия объекта (objectness), а $\mathcal{L}_{cls}$ измеряет кросс-энтропийную ошибку базовой классификации обнаруженных дефектов. Компонент $\mathcal{L}_{ref}$ представляет собой потерю модуля тонкого уточнения классификации, $\mathcal{L}_{seg}$ штрафует модель за

несовпадение предсказанных бинарных масок с эталонными, а $\mathcal{L}_{bd}$ обеспечивает дополнительный контроль качества на границах сегментируемых областей (boundary supervision). Весовые коэффициенты $\lambda_1, \dots, \lambda_6$ являются гиперпараметрами системы, которые балансируют вклад каждой отдельной задачи в общий градиент ошибки, предотвращая доминирование одной ветви вычислений над другими в процессе обучения. Таким образом, основная задача исследования сводится к нахождению такого набора весовых параметров $\theta^* = argmin_\theta(\mathcal{L}_{total})$, который обеспечит максимальную обобщающую способность системы при обработке новых видеоданных дорожной обстановки в режиме реального времени.

2.5 Проектирование многозадачной нейросетевой архитектуры TCR-RoadNet

В данном разделе детально рассматривается предлагаемая архитектура глубокого обучения TCR-RoadNet, спроектированная для обнаружения и сегментации повреждений дорожного покрытия в режиме реального времени. Разработка единой многозадачной системы обусловлена тем, что дефекты дорожного полотна в реальных условиях эксплуатации демонстрируют существенные вариации в масштабе, форме, текстуре и четкости границ, из-за чего использование традиционных однозадачных сетей или признаков единого масштаба оказывается недостаточным для точного анализа. Предлагаемая архитектура представляет собой гибридное решение, которое объединяет иерархическое извлечение сверточных признаков, контекстное моделирование на основе трансформеров, раздельное обнаружение объектов, порегиональное уточнение классов и генерацию попиксельных масок с учетом границ.

2.5.1 Общая структура модели

На рисунке 14 показан полный вычислительный конвейер, который представляет собой сквозной граф вывода, который начинается с этапа предварительной обработки входного изображения. Входной кадр дорожной сцены, предварительно приведенный к тензорной размерности 640×640×3, первоначально пропускается через многомасштабную сверточную магистраль (Multi-Scale CNN Backbone). Данный базовый экстрактор последовательно обрабатывает изображение, извлекая иерархические пространственные признаки при постепенно снижающемся разрешении и формируя многоуровневую пирамиду признаков. Для преодоления фундаментального ограничения сверточных операций, связанного с локальностью рецептивного поля, полученные базовые признаки направляются в модуль контекстного уточнения на основе трансформера (Transformer Context Refinement). Этот блок усиливает кросс-масштабное взаимодействие данных и формирует долгосрочные контекстные представления, что критически важно для уверенного распознавания фрагментированных трещин на фоне сложных дорожных текстур.

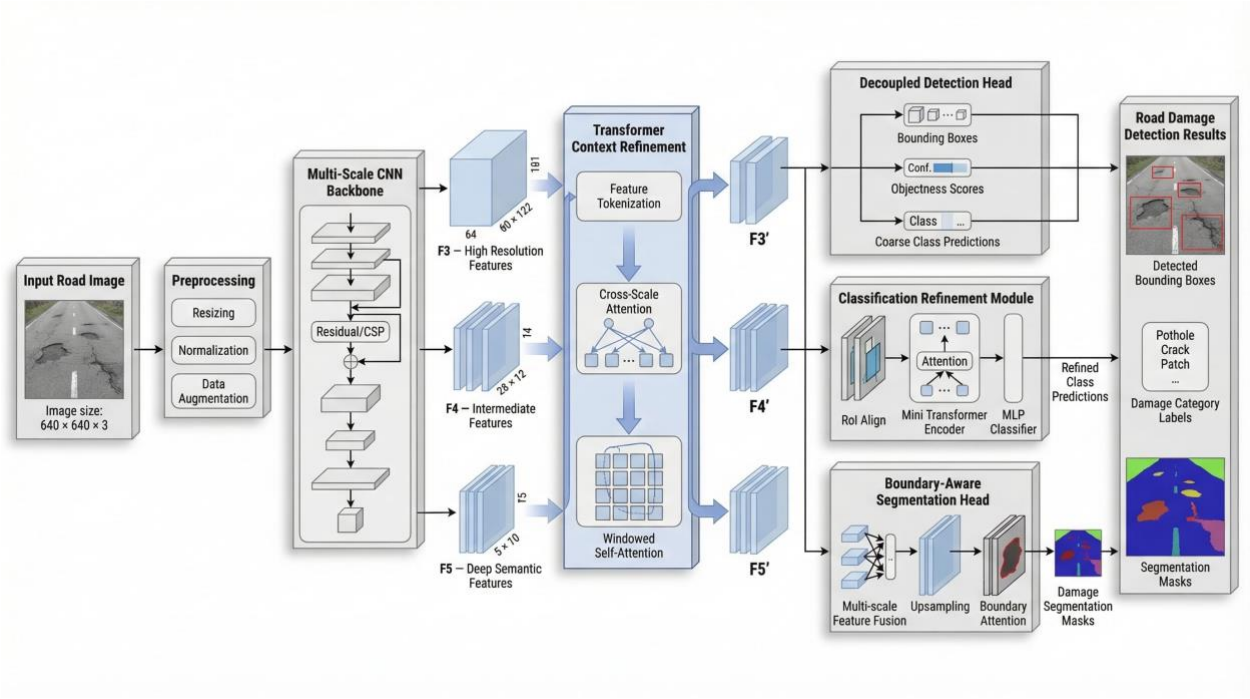


Рисунок 14 – Общая архитектура системы многомасштабного уточнения признаков на основе трансформера (TCR-RoadNet)

На завершающем этапе обработки обогащенные контекстом признаки параллельно распределяются между тремя специализированными проблемно-ориентированными ветвями вывода. Первая ветвь – модуль раздельного обнаружения (Decoupled Detection Head) – отвечает за независимое прогнозирование координат ограничивающих рамок и оценку общей вероятности наличия объекта. Вторая ветвь, представляющая собой модуль уточнения классификации (Classification Refinement Module), повторно анализирует обнаруженные локальные регионы для повышения семантической точности распознавания визуально схожих категорий дефектов. Наконец, третья ветвь – сегментационная голова с учетом границ (Boundary-Aware Segmentation Head) – генерирует детализированные маски трещин и выбоин с улучшенной точностью пространственных контуров. Подобная структура позволяет совместно оптимизировать процессы обнаружения, классификации и сегментации в рамках единого цикла вывода, формируя комплексный и высокоточный прогноз состояния дорожного покрытия.

2.5.2 Многомасштабная сверточная магистраль

В основе процесса извлечения визуальных признаков из входного потока данных лежит многомасштабная сверточная магистраль (Multi-Scale CNN Backbone). Внутренняя структура магистрали представляет собой иерархическую многостадийную сверточную нейронную сеть, состоящую из слоев свертки с размером ядра 3×3, остаточных блоков типа CSP (Cross Stage Partial) и операций пространственного понижения дискретизации. Фундаментальным строительным блоком магистрали является композитный сверточный модуль (CBS), который выполняет последовательность

математических операций: двумерную свертку, пакетную нормализацию (Batch Normalization) и нелинейную активацию. Для входного тензора X_{in} дискретная двумерная свертка с ядром W и смещением b вычисляется для каждого канала как:

$$X_{conv}(i,j) = \sum_m \sum_n X_{in}(i+m, j+n) \cdot W(m,n) + b \quad (2)$$

Для ускорения сходимости и стабилизации дисперсии признаков применяется алгоритм пакетной нормализации по мини-батчу $\mathcal{B}$

$$X_{bn} = \gamma \frac{X_{conv} - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} + \mathcal{B} \quad (3)$$

где $\mu_{\mathcal{B}}$ и $\sigma_{\mathcal{B}}^2$ – математическое ожидание и дисперсия батча, а γ и $\mathcal{B}$ - обучаемые параметры масштаба и сдвига. В качестве нелинейности используется гладкая функция активации SiLU (Sigmoid Linear Unit), которая предотвращает проблему «затухания градиента» на глубоких слоях:

$$f_{SiLU}(x) = x \cdot \sigma(x) = \frac{x}{1 + e^{-x}} \quad (4)$$

Таким образом, выход базового сверточного блока математически определяется как:

$$X_{out} = f_{SiLU}(X_{bn}) \quad (5)$$

Главная цель данного компонента заключается в сохранении тонких локальных паттернов на начальных (поверхностных) слоях сети при одновременном формировании сильных семантических представлений на более глубоких уровнях. Подобный архитектурный дизайн является строго необходимым для задач анализа дорожных повреждений, поскольку различные типы дефектов – от узких трещин до обширных ям и следов ямочного ремонта – могут занимать совершенно разные области изображения и существенно варьироваться в своем масштабе. Архитектурная схема данного модуля детально представлена на рисунке 15.

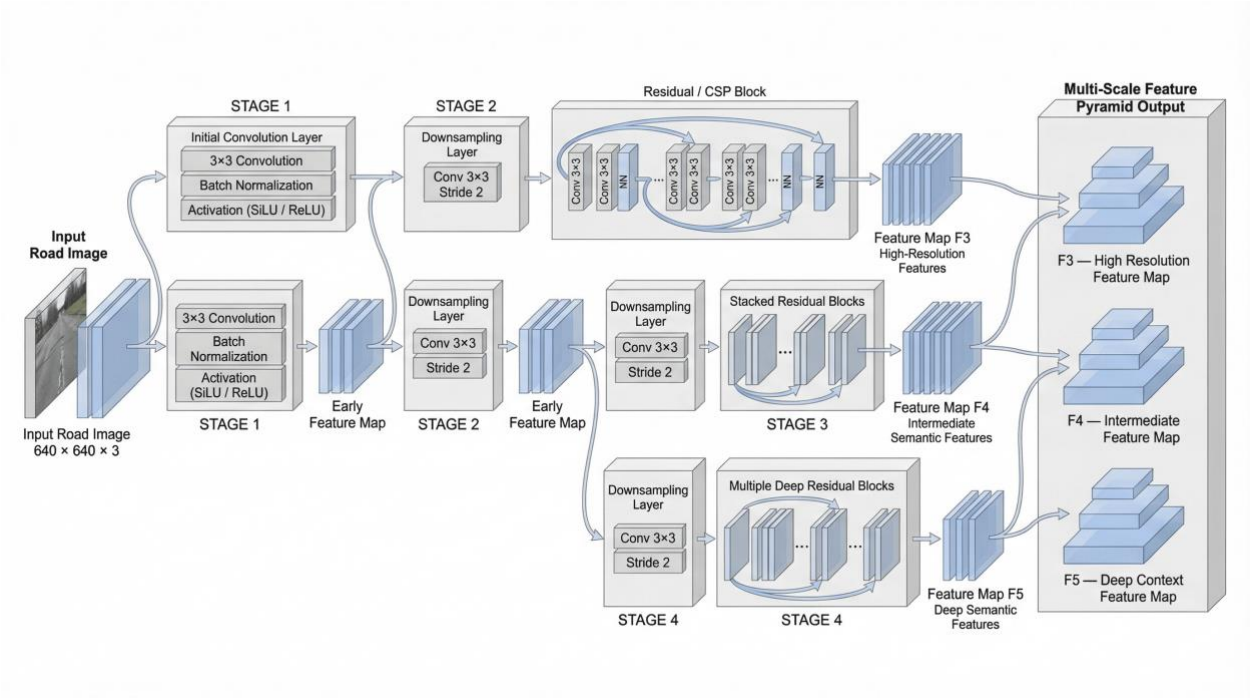


Рисунок 15 – Детальная архитектура многомасштабной сверточной магистрали

С математической точки зрения процесс извлечения признаков описывается как последовательность нелинейных тензорных преобразований. При подаче на вход сети тензора предварительно обработанного и аугментированного изображения I_a, первая стадия магистрали извлекает низкоуровневые текстурные дескрипторы. Последующие стадии вычислений постепенно уменьшают пространственное разрешение карт признаков, параллельно увеличивая их канальную размерность [7]. В общем виде математическое преобразование признаков на произвольной стадии сети l выражается следующим рекуррентным уравнением:

$$F_l = \phi_l(F_{l-1}, \theta_l) \tag{6}$$

В представленной формуле функция ϕ_l обозначает комплексную композитную трансформацию, включающую в себя операции свертки, нормализации (например, пакетной нормализации), нелинейной функции активации и агрегации остаточных связей. Вектор θ_l представляет собой набор оптимизируемых обучаемых параметров (синаптических весов) для данного уровня сети l.

В TCR-RoadNet комплексная функция трансформации ϕ_l реализована на основе механизма Cross Stage Partial (CSP). Данный подход разделяет поток градиентов, что позволяет увеличить глубину сети без существенного роста вычислительной сложности. Внутри CSP-блока входной тензор F_{l-1} проецируется в два независимых скрытых состояния с помощью сверток 1×1:

$$X_1 = \mathrm{Conv}_{1\times1}(F_{l-1}) \tag{7}$$
$$X_2 = \mathrm{Conv}_{1\times1}(F_{l-1}) \tag{8}$$

Ветвь X_2 проходит через последовательность из K плотных остаточных блоков (Residual Blocks). Преобразование внутри *k*-го остаточного блока с функцией нелинейного отображения H_k описывается рекуррентным уравнением:

$$X_2^{(k)} = \mathcal{H}_k\left(X_2^{(k-1)}\right) + X_2^{(k-1)}, \quad k \in \{1, \dots, K\} \tag{9}$$

где $X_2^{(k)} = X_2$. Использование тождественного отображения (identity mapping) $+X_2^{(k-1)}$ гарантирует беспрепятственное протекание градиента при обратном распространении ошибки. После прохождения всех K блоков, признаки обеих ветвей конкатенируются по канальному измерению:

$$X_{cat} = \left[X_1, X_2^{(K)}\right] \tag{10}$$

Итоговый тензор обновленного уровня формируется путем слияния признаков с помощью финальной переходной свертки:

$$F_l = \mathrm{Conv}_{1\times1}(X_{cat}) \tag{11}$$

В результате последовательной обработки входного тензора магистраль генерирует на выходе пространственную пирамиду признаков, состоящую из трех разномасштабных карт вывода. Формирование иерархических уровней пирамиды обеспечивается операциями страйдовой свертки (stride s=2), которые выполняют пространственное понижение дискретизации тензоров. Геометрическое изменение высоты H_l и ширины W_l тензора на каждом этапе *l* вычисляется по следующим формулам:

$$H_l = \left\lfloor \frac{H_{l-1} + 2p - k}{s} \right\rfloor + 1 \tag{12}$$

$$W_l = \left\lfloor \frac{W_{l-1} + 2p - k}{s} \right\rfloor + 1 \tag{13}$$

где *k=3* размер ядра свертки, *p=1* величина симметричного пространственного отступа (padding), а *s=2* шаг сканирования. При заданных параметрах на каждом последующем уровне пространственное разрешение уменьшается ровно в два раза, в то время как семантическая емкость (количество каналов C_l) удваивается. При условии, что на вход сети подается изображение нормализованного размера 640×640 пикселей, пространственные и канальные размерности сформированных выходных тензоров математически определяются следующим образом:

$$F_3 \in \mathbb{R}^{80\times80\times256}, F_4 \in \mathbb{R}^{40\times40\times512}, F_5 \in \mathbb{R}^{20\times20\times1024} \quad (14)$$

Каждый из трех сформированных уровней пирамиды предоставляет строго комплементарную информацию, необходимую для последующего многозадачного логического вывода. Высокоразрешенная карта признаков F_3 сохраняет детализированную пространственную топологию, что критически важно для сегментации областей, чувствительных к контурам, а также для успешной локализации очень тонких трещин покрытия. Промежуточный тензор F_4 фиксирует структурные семантические паттерны среднего уровня. В свою очередь, тензор F_5, обладающий наименьшим пространственным разрешением, но наибольшей концептуальной глубиной, кодирует высокоуровневые глобальные контекстные признаки. Это свойство глубокого тензора обеспечивает надежную локализацию обширных разрушений асфальтобетона на фоне сложных визуальных помех, таких как тени или дорожная разметка. Наконец, перед передачей извлеченных пирамидальных признаков в следующий блок контекстного уточнения на основе трансформера, все тензоры математически проецируются в единую размерность встраивания (embedding dimension) d. Данное преобразование выполняется с использованием операций свертки 1×1 или линейных проекций. Подобная гармонизация канальных размерностей необходима для обеспечения высокой вычислительной эффективности при расчете кросс-масштабного внимания на последующих этапах работы нейросети, а также для устранения размерных несоответствий между иерархическими уровнями пирамиды признаков.

Процесс проекции каждого признакового тензора F_l в размерность встраивания d выполняется путем линейной трансформации по канальной оси. Для каждого пространственного пикселя с координатами (i, j) новое значение канала C_{out} вычисляется как скалярное произведение вектора исходных признаков на весовую матрицу проекции Wproj(l):

$$F_l(i, j, c_{out}) = \sum_{c_{in}=1}^{C_l} F_l(i, j, c_{in}) \cdot W_{proj}^{(l)}(c_{in}, c_{out}) \quad (15)$$

где $C_{out} \in \{1, \dots, d\}$, а C_l – исходное количество каналов тензора F_l. В результате данной операции генерируется набор структурно гармонизированных тензоров {F_3, F_4 ,F_5}, которые формируют идеальное семантическое пространство для эффективного вычисления матриц внимания на последующих этапах работы трансформера.

2.5.3 Модуль контекстного уточнения на основе трансформера

Вторым ключевым вычислительным компонентом разрабатываемой архитектуры является модуль контекстного уточнения на основе трансформера (Transformer Context Refinement Block). Несмотря на то что

традиционные сверточные операции демонстрируют высокую эффективность при извлечении признаков из локальных окрестностей пикселей, они обладают фундаментально ограниченным рецептивным полем. Это ограничение существенно снижает их способность к моделированию долгосрочных пространственных зависимостей, что особенно критично в задачах анализа дорожного полотна, где дефекты могут быть визуально фрагментированы или частично перекрыты сложными фоновыми помехами. Предлагаемый интегрированный модуль трансформера решает данную проблему путем преобразования карт признаков в последовательности токенов, вычисления кросс-масштабного внимания и последующей обратной проекции в уточненные пространственные тензоры. Внутренняя структура данного аналитического блока наглядно проиллюстрирована на рисунке 16.

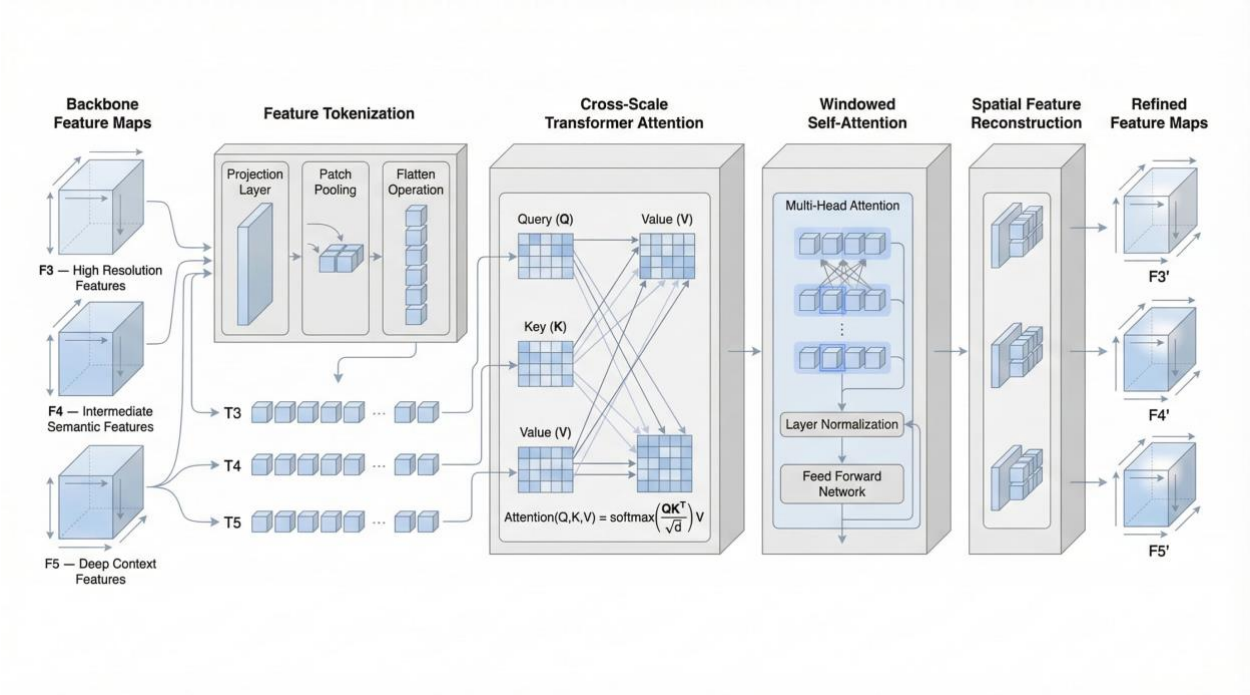


Рисунок 16 – Внутренняя архитектура модуля контекстного уточнения на основе трансформера (TCR)

Процесс контекстного обогащения начинается с токенизации разномасштабных пространственных признаков F_3,F_4 и F_5, полученных из сверточной магистрали. Пусть входная карта признаков масштаба s имеет размерность $F_s \in \mathbb{R}^{H_s \times W_s \times C_s}$. Для применения механизмов трансформера эта трехмерная матрица разбивается на сетку непересекающихся локальных патчей размером P×P. Общее количество полученных патчей (токенов) W_s вычисляется как:

$$N_s = \frac{H_s \times W_s}{P^2} \tag{16}$$

Каждый патч вытягивается в одномерный вектор $x_p^{(i)} \in \mathbb{R}^{P^2 \cdot C_s}$, где $i \in i \in \{1, \dots, N_s\}$. Затем применяется обучаемая линейная проекция (матрица E),

которая преобразует векторы патчей в скрытое пространство одинаковой размерности d (embedding dimension). Для сохранения пространственной топологии изображения, которая теряется при вытягивании, к каждому токену прибавляется обучаемое позиционное кодирование $E_{pos}^{(i)}$:

$$X_s^{(i)} = x_p^{(i)} E + E_{pos}^{(i)} \tag{17}$$

Таким образом, итоговая последовательность токенов для масштаба *s*, подаваемая на вход трансформеру, формируется в виде матрицы X_s:

$$X_s = \left[X_s^{(1)}, X_s^{(2)}, \dots, X_s^{(N_s)}\right] \in R^{N_s \times d} \tag{18}$$

После успешного формирования токенов вычислительная система инициирует механизм кросс-масштабного внимания (Cross-Scale Transformer Attention). В отличие от классического самовнимания (Self-Attention), данный блок извлекает матрицу запросов (Query, Q) из целевого масштаба *s*, а матрицы ключей (Key, K) и значений (Value, V) – из опорного масштаба s′. Эти матрицы формируются путем умножения входных токенов на соответствующие обучаемые весовые матрицы W_Q, W_K, W_V:

$$Q = X_s W_Q, \quad K = X_{s'} W_K, \quad V = X_{s'} W_V \tag{19}$$

Для повышения выразительной способности сети применяется механизм многоголового внимания (Multi-Head Attention), в котором вычисления производятся параллельно в h независимых подпространствах (головах). Внимание для отдельной *j*-й головы вычисляется с использованием масштабированного скалярного произведения:

$$head_j = \text{Softmax}\left(\frac{Q_j K_j^T}{\sqrt{d_k}}\right) V_j \tag{20}$$

где $d_k = d/h$ – размерность отдельной головы, а коэффициент $\sqrt{d_k}$ используется для стабилизации градиентов функции Softmax. Результаты всех h голов конкатенируются по канальной оси и умножаются на выходную проекционную матрицу W_O:

$$\text{MHCA}(X_s, X_{s'}) = \text{Concat}(head_1, \dots, head_h) W_O \tag{21}$$

Вычисленное кросс-масштабное внимание интегрируется в архитектуру с использованием остаточных связей (residual connections) и нормализации слоев (Layer Normalization, LN), применяемой перед каждым вычислительным

блоком (Pre-LN). Промежуточное представление признаков $\widehat{X_s}$ вычисляется как:

$$\widehat{X_s} = X_s + \mathrm{MHCA}\big(\mathrm{LN}(X_s),\mathrm{LN}(X_{s'})\big) \tag{22}$$

Далее тензор проходит через блок многослойного перцептрона (MLP), который содержит два линейных преобразования с функцией активации GELU (Gaussian Error Linear Unit) между ними. MLP осуществляет нелинейную трансформацию признаков каждого токена независимо:

$$\mathrm{MLP}(x) = \mathrm{GELU}(xW_1 + b_1)W_2 + b_2 \tag{23}$$

Финальный выход кросс-масштабного блока трансформера для масштаба *s* формируется добавлением второй остаточной связи:

$$X_s^{out} = \widehat{X_s} + \mathrm{MLP}\left(\mathrm{LN}\left(\widehat{X_s}\right)\right) \tag{24}$$

Благодаря интеграции данной математической операции нейронная сеть получает способность динамически связывать мелкие детали текстуры трещин, извлеченные на ранних высокоразрешенных слоях, с сильными глобальными семантическими представлениями, сформированными на глубоких слоях. На практике этот аналитический механизм позволяет системе радикально повысить свою устойчивость и эффективно отличать истинные структурные повреждения дорожного покрытия от сложных визуальных помех, таких как линии дорожной разметки, участки ямочного ремонта, глубокие тени от деревьев или поверхностные масляные пятна.

На завершающем этапе обработки векторных данных применяется механизм оконного внутреннего внимания (Windowed Self-Attention, WSA), который снижает квадратичную вычислительную сложность трансформера. Карты признаков локально разделяются на непересекающиеся окна размером M×M. Внутри каждого окна самовнимание вычисляется с добавлением матрицы относительного позиционного смещения (Relative Position Bias) B:

$$\mathrm{WSA}(X_{window}) = \mathrm{Softmax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \tag{25}$$

Интеграция матрицы $B \in \mathbb{R}^{M^2 \times M^2}$ критически важна для задач компьютерного зрения, так как она кодирует пространственные расстояния между пикселями внутри окна. После обработки механизмом оконного внимания, последовательность токенов подвергается обязательной стадии пространственной реконструкции (spatial feature reconstruction). Итоговая уточненная карта признаков F_s' вычисляется путем применения оператора изменения формы (Reshape):

$$F_s' = \text{Reshape}_{N_s \to H_s \times W_s}\left(\text{WSA}(X_s^{out})\right) \quad (26)$$

Результатом работы всего рассматриваемого модуля является полностью обновленная и обогащенная контекстом пирамида признаков F_3', F_4', F_5', которая затем параллельно распределяется между многозадачными ветвями вывода для генерации окончательных прогнозов по обнаружению, классификации и сегментации.

2.5.4 Модуль раздельного обнаружения

Третьим ключевым компонентом предлагаемой архитектуры и первой специализированной ветвью прогнозирования является модуль раздельного обнаружения (Decoupled Detection Head, рисунок 17). В традиционных архитектурах объектного обнаружения прогнозирование координат рамок и классов часто выполняется в рамках единой разделяемой подсети, однако в предложенной модели эти задачи, наряду с оценкой вероятности наличия объекта, логически и вычислительно разделены. Необходимость такого архитектурного разделения обусловлена тем, что задачи пространственной локализации и семантического распознавания требуют принципиально разной чувствительности к извлекаемым визуальным признакам. В контексте анализа дорожного полотна это разделение имеет критическое значение: например, трещина может быть сильно вытянута в пространстве и иметь четкие координаты, но оставаться семантически неоднозначной для классификации, тогда как выбоина, напротив, легко классифицируется по характерному внешнему виду, но ее точные пространственные границы бывает трудно локализовать в условиях слабого контраста с окружающим асфальтом.

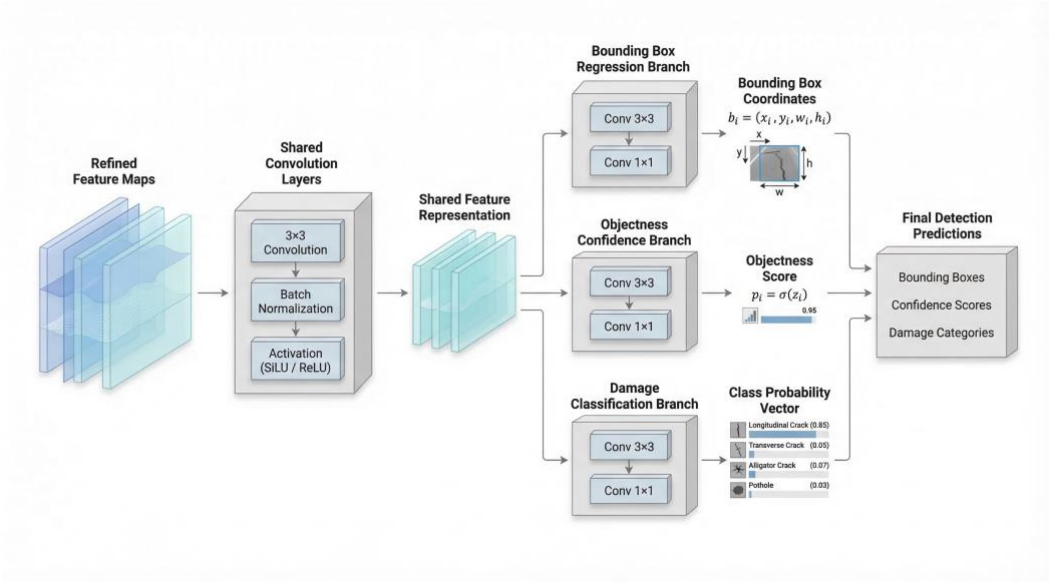


Рисунок 17 – Архитектура модуля раздельного обнаружения (Decoupled Detection Head)

С математической точки зрения процесс вывода начинается с приема уточненных многомасштабных карт признаков $F_l' \in \{F_3', F_4', F_5'\}$, сгенерированных блоком контекстного уточнения на различных иерархических уровнях l. Данный тензор первоначально проходит через блок разделяемых сверточных слоев (shared convolutions), который унифицирует канальную размерность и формирует общее абстрактное представление F_l^{shared}. Данное преобразование описывается как последовательность свертки 3×3, пакетной нормализации и нелинейной активации.

$$F_l^{shared} = f_{SiLU}\left(\mathrm{BatchNorm}\left(\mathrm{Conv}_{3\times3}(F_l')\right)\right) \tag{27}$$

После этого общий тензор признаков F_l^{shared} направляется в три физически независимые вычислительные ветви (регрессии, наличия объекта и классификации), что позволяет минимизировать интерференцию градиентов между принципиально разными задачами при обратном распространении ошибки. Первая ветвь отвечает за пространственную регрессию координат ограничивающих рамок. Для каждого уровня пирамиды l формируется специализированная карта признаков регрессии:

$$F_l^{reg} = \mathrm{Conv}_{1\times1}\left(\mathrm{Conv}_{3\times3}\left(F_l^{shared}\right)\right) \tag{28}$$

В архитектуре используется подход без якорных рамок (anchor-free). Для каждой ячейки сетки с координатами центра (x_c, y_c) сеть предсказывает четыре значения – нормализованные смещения до левой (l_i), верхней (t_i), правой (r_i) и нижней (b_i) границ дефекта. Для обеспечения положительных значений предсказаний применяется экспоненциальная функция с учетом шага дискретизации s_l (stride) данного уровня пирамиды:

$$(l_i, t_i, r_i, b_i) = \exp\left(F_l^{reg}(x_c, y_c)\right) \cdot s_l \tag{29}$$

Декодирование этих смещений в стандартный формат ограничивающей рамки b_i (координаты центра, ширина, высота) осуществляется посредством следующих геометрических преобразований:

$$x_i = x_c + \frac{r_i - l_i}{2}, \quad y_i = y_c + \frac{b_i - t_i}{2} w_i = l_i + r_i, \quad h_i = t_i + b_i$$
$$b_i = (x_i, y_i, w_i, h_i) \tag{30}$$

В данном уравнении x_i и y_i обозначают нормализованные координаты центра ограничивающей рамки, а w_i и h_i соответствуют ее расчетной ширине и высоте. Вторая параллельная ветвь вычисляет оценку наличия объекта (objectness score), интерпретируемую как вероятность того, что выделенный

регион i содержит дефект дорожного полотна. Признаки для этой задачи извлекаются отдельным сверточным блоком:

$$F_l^{obj} = \mathrm{Conv}_{1\times1}\left(\mathrm{Conv}_{3\times3}\left(F_l^{shared}\right)\right) \quad (31)$$

Финальная вероятность o_i вычисляется путем применения сигмоидной функции активации к скалярному логиту:

$$o_i = \sigma\left(F_l^{obj}(x_c, y_c)\right) = \frac{1}{1+\exp\left(-F_l^{obj}(x_c, y_c)\right)} \quad (32)$$

Третья ветвь отвечает за генерацию предварительного распределения вероятностей принадлежности дефекта к одной из $N_{classes}$ заданных категорий повреждений. Процесс извлечения классовых логитов описывается аналогичной последовательностью сверток:

$$F_l^{cls} = \mathrm{Conv}_{1\times1}\left(\mathrm{Conv}_{3\times3}\left(F_l^{shared}\right)\right) \quad (33)$$

Для формирования итогового вектора вероятностей p_i^{cls} применяется функция Softmax, которая нормализует логиты по всем возможным классам k:

$$p_{i,c}^{cls} = \frac{\mathrm{e}^{F_l^{cls}(x_c, y_c, c)}}{\sum_{k=1}^{N_{classes}} \mathrm{e}^{F_l^{cls}(x_c, y_c, c)}}, c \in \{1, \dots, N_{classes}\} \quad (34)$$

Это обеспечивает высокую стабильность процесса оптимизации и повышает точность первичного обнаружения дефектов. В результате параллельной работы всех трех специализированных подсетей, для каждого i-го кандидата на различных пространственных масштабах l формируется единый вектор комплексного предсказания $\widehat{Y}_l$. Данный вектор математически выражается как конкатенация предсказанных пространственных координат, оценки наличия объекта и вероятностей семантических классов:

$$\widehat{Y}_l = \left[b_i,\ o_i,\ p_{i,1}^{cls},\ \dots,\ p_{i,N_{classes}}^{cls}\right] \in \mathbb{R}^{4+1+N_{classes}} \quad (35)$$

Множество всех векторов $\widehat{Y}_l$ со всех уровней пирамиды признаков образует исходное пространство предсказаний, которое затем фильтруется с помощью алгоритма подавления немаксимумов (NMS) для исключения дублирующих рамок перед передачей в модуль уточнения классификации.

2.5.5 Модуль уточнения классификации

Четвертым важным структурным элементом архитектуры является модуль уточнения классификации (Classification Refinement Module).

Несмотря на то что модуль раздельного обнаружения генерирует предварительные оценки классов, на практике часто возникает проблема неоднозначности при распознавании визуально схожих категорий дефектов. Например, продольные трещины, поперечные трещины, аллигаторные (сетчатые) трещины и участки ямочного ремонта могут иметь схожие общие текстурные характеристики. Для минимизации подобных ошибок классификации в архитектуру внедрен специализированный модуль, который осуществляет вторичный, порегиональный анализ выделенных областей интереса после применения алгоритма подавления немаксимумов (Non-Maximum Suppression, NMS). Архитектурная схема данного компонента представлена на рисунке 18.

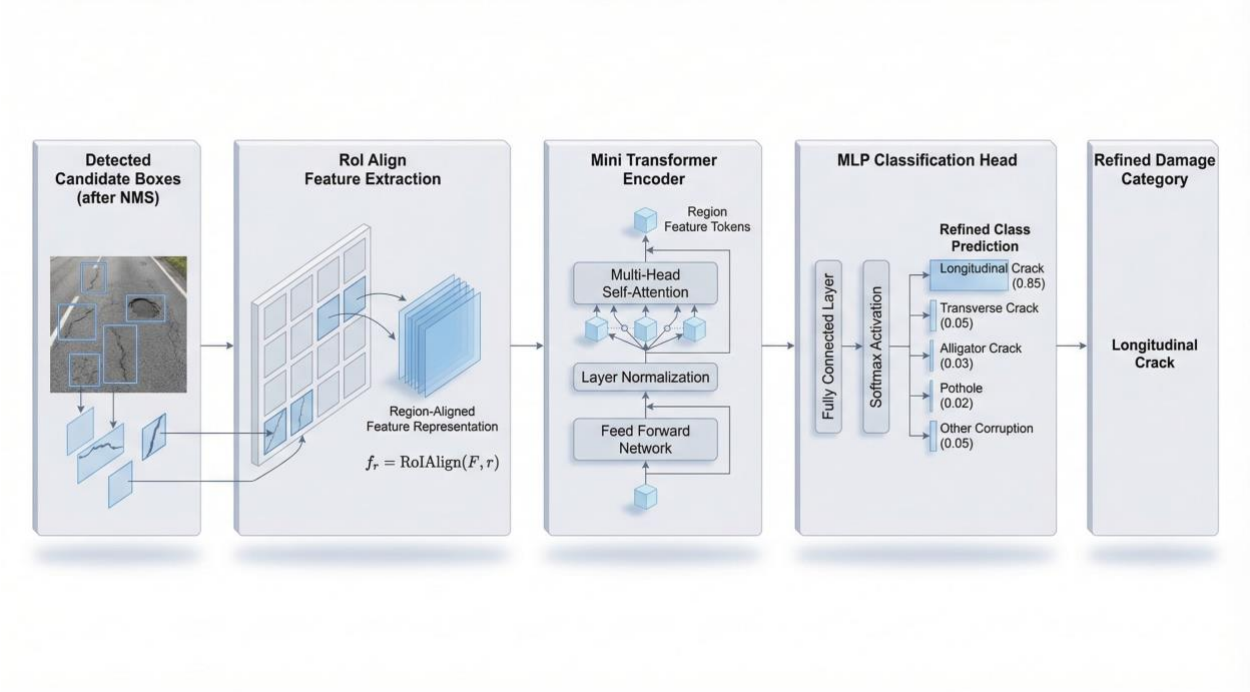


Рисунок 18 – Архитектура модуля уточнения классификации

С вычислительной точки зрения процесс уточнения начинается с приема множества рамок-кандидатов B_{nms} прошедших фильтрацию алгоритмом подавления немаксимумов. Для каждого отдельного региона $r_j \in B_{nms}$ с физическими координатами (x, y, w, h) необходимо извлечь локальный дескриптор из многомасштабной пирамиды признаков. Сначала определяется оптимальный иерархический уровень l_j карты признаков F'_{l_j}, с которого будет производиться извлечение. Выбор уровня базируется на площади рамки по следующему эвристическому правилу:

$$l_j = \left\lfloor l_0 + log_2\left(\frac{\sqrt{w \cdot h}}{224}\right)\right\rfloor \tag{36}$$

где l_0 – базовый уровень пирамиды, а 224 – канонический размер предварительного обучения.

После выбора уровня к карте признаков F'_{l_j} применяется операция Region of Interest Align (RoI Align). В отличие от устаревшего RoI Pooling, эта операция избегает жесткого квантования координат [8]. Ограничивающая рамка разбивается на сетку размером $H_{roi} \times W_{roi}$. Непрерывные координаты точки выборки (x_p, y_p) внутри бина вычисляются как:

$$x_p = x_{roi} + \frac{w}{W_{roi}} \cdot (p_x + 0.5), \quad y_p = y_{roi} + \frac{h}{H_{roi}} \cdot \left(p_y + 0.5\right) \tag{37}$$

Для получения точного значения признака в дробной точке (x_p, y_p) применяется билинейная интерполяция по четырем ближайшим дискретным узлам (x_i, y_i) на карте F'_{l_j}:

$$V(x_p, y_p) = \sum_{i,j \in \{0,1\}} \left(1 - |x_p - x_i|\right)\left(1 - |y_p - y_i|\right) \cdot F'_{l_j}(x_i, y_i) \tag{38}$$

В результате применения пулинга (например, усреднения) к точкам выборки внутри каждого бина формируется выровненный тензор локальных признаков $z_j \in \mathbb{R}^{H_{roi} \times W_{roi} \times C}$:

$$z_j = \text{RoIAlign}\left(F'_{l_j}, r_j, H_{roi}, W_{roi}\right) \tag{39}$$

Полученный локальный пространственный тензор z_j передается для глубокого анализа в компактный мини-трансформер (mini-transformer encoder). Перед подачей в трансформер трехмерный тензор преобразуется в плоскую последовательность токенов z_0, к которой прибавляется локальное позиционное кодирование E^{roi}_{pos} для сохранения информации о геометрии дефекта:

$$Z_0 = \text{Flatten}\left(z_j\right) + E^{roi}_{pos}, \quad Z_0 \in R^{(H_{roi} \cdot W_{roi}) \times C} \tag{40}$$

Внутри энкодера последовательность проходит через механизм многоголового самовнимания (Multi-Head Self-Attention), который позволяет модели сфокусироваться на наиболее информативных частях выделенного региона (например, на изломах трещины, игнорируя фоновый асфальт внутри рамки). Матрицы запросов, ключей и значений генерируются специфично для данного региона:

$$Q_{roi} = Z_0 W_Q, \quad K_{roi} = Z_0 W_K, \quad V_{roi} = Z_0 W_V \tag{41}$$

Обновление признаков осуществляется с использованием нормализации слоев (LayerNorm) и остаточных связей:

$$\widehat{Z_1} = Z_0 + \text{MHSA}\big(\text{LN}(Q_{roi}),\text{LN}(K_{roi}),\text{LN}(V_{roi})\big) \quad (42)$$

Затем применяется полносвязная сеть прямого распространения (MLP) для поканальной трансформации:

$$Z_{out} = \widehat{Z_1} + \text{MLP}\left(\text{LN}\left(\widehat{Z_1}\right)\right) \quad (43)$$

На завершающем этапе выходная последовательность токенов Z_{out}, обогащенная локальным контекстом, агрегируется в единый вектор-дескриптор v_j с помощью операции глобального среднего пулинга (Global Average Pooling) по всем Nroi токенам ($N_{roi} = H_{roi} \cdot W_{roi}$):

$$v_j = \frac{1}{N_{roi}} \sum_{i=1}^{N_{roi}} Z_{out}^{(i)} \quad (44)$$

Этот компактный вектор направляется в классификационную «голову» на базе многослойного перцептрона (MLP Classification Head). Нелинейная трансформация скрытого слоя вычисляется как:

$$h_j = \text{ReLU}\big(v_j W_{fc1} + b_{fc1}\big) \quad (45)$$

После чего генерируется вектор «сырых» оценок (логитов) для всех $N_{classes}$ категорий дефектов:

$$\text{logits}_j = h_j W_{fc2} + b_{fc2} \quad (46)$$

Окончательное уточненное распределение вероятностей p_j^{ref} для анализируемого региона r_j вычисляется посредством функции Softmax:

$$p_j^{ref} = \text{Softmax}\left(\text{logits}_j\right) \quad (47)$$

Именно этот вектор p_j^{ref} заменяет предварительные грубые предсказания из модуля раздельного обнаружения, обеспечивая высокую точность семантической дискриминации.

Подобная операция порегионального уточнения оказывается особенно эффективной в тех случаях, когда локальная ориентация формы дефекта, фрагментация его текстуры или границы ремонтных заплат должны интерпретироваться строго в пределах ограниченной пространственной области, изолированно от глобального контекста всего изображения. В практическом применении интеграция данного модуля обеспечивает

значительно более надежную дифференциацию между различными типами трещин и поверхностными дефектами, которые обладают схожей грубой статистикой внешнего вида, но принципиально различаются по своей локальной структурной организации.

2.5.6 Модуль сегментации с учетом границ

Пятым и заключительным вычислительным модулем предлагаемой архитектуры глубокого обучения является сегментационная ветвь с учетом границ (Boundary-Aware Segmentation Head). Ее основное алгоритмическое назначение заключается в генерации детализированных попиксельных бинарных масок для выявленных повреждений дорожного полотна с особым акцентом на сохранение топологической точности контуров [9]. Разработка столь детализированного модуля обусловлена тем, что фактическая степень тяжести, физическая площадь и характер деградации асфальтобетонного покрытия (особенно в случаях выбоин и сложных сетчатых разрушений) гораздо точнее описываются их истинными границами, нежели классическими прямоугольными ограничивающими рамками. Внутренняя структура данного модуля визуализирована на рисунке 19.

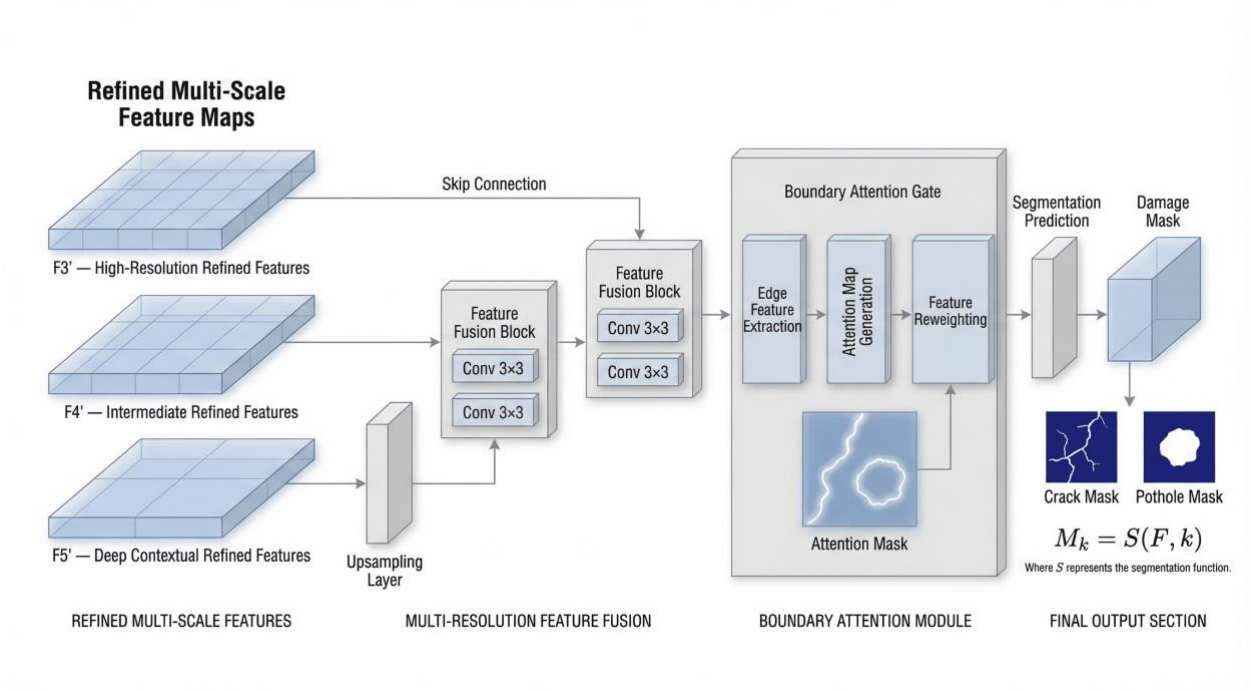


Рисунок 19 – Архитектура сегментационной ветви с учетом границ

Модуль сегментации с учетом границ (Boundary-Aware Segmentation Head, BSH) предназначен для точного оконтуривания дорожных дефектов на уровне пикселей. Процесс генерации маски начинается с агрегации семантической информации со всех уровней уточненной пирамиды признаков $\{F_3', F_4', F_5'\}$. Для восстановления пространственного разрешения используется механизм каскадного слияния сверху вниз (Top-Down Feature Fusion). Начиная с самого глубокого уровня F_5', тензоры последовательно масштабируются и объединяются:

$$U_5 = \mathrm{Conv}_{1\times1}(F_5') \tag{48}$$

На каждом последующем шаге $l \in \{4,3\}$ к признакам более высокого пространственного разрешения добавляются апсемплированные признаки предыдущего уровня посредством поэлементного сложения $\bigoplus$:

$$U_4 = \mathrm{Conv}_{3\times3}\big(\mathrm{Upsample}(U_5) \bigoplus \mathrm{Conv}_{1\times1}(F_4')\big) \tag{49}$$
$$U_3 = \mathrm{Conv}_{3\times3}\big(\mathrm{Upsample}(U_4) \bigoplus \mathrm{Conv}_{1\times1}(F_3')\big) \tag{50}$$

Операция *Upsample* реализуется с помощью билинейной интерполяции, увеличивающей пространственные размеры тензора в 2 раза. Результирующий тензор $F_{fuse} = U_3$ обладает высокой разрешающей способностью и содержит богатый глобальный контекст, необходимый для детальной сегментации.

Дефекты дорожного полотна, в частности продольные и сетчатые трещины, характеризуются сложной геометрией и низким контрастом по отношению к фону [10]. Для акцентирования морфологии дефектов в архитектуру интегрирован блок извлечения границ, который вычисляет пространственные градиенты непосредственно в пространстве глубоких признаков. Вычисление градиентов по осям X и Y аппроксимируется сверткой объединенного тензора F_{fuse} с фиксированными дифференцируемыми ядрами K_{sobel}^{x} и K_{sobel}^{y}:

$$\nabla_x = F_{fuse} * K_{Sobel}^{x}, \quad \nabla_y = F_{fuse} * K_{Sobel}^{y} \tag{51}$$

Результирующая карта амплитуды градиентов G_{bound}, описывающая контуры высокочастотных изменений (края трещин и ям), вычисляется как Евклидова норма векторов градиента:

$$G_{bound} = \sqrt{(\nabla_x)^2 + \left(\nabla_y\right)^2} \tag{52}$$

Для адаптации полученной граничной информации под специфику конкретных дорожных текстур применяется обучаемый сверточный слой, формирующий итоговое представление границ B_{feat}:

$$B_{feat} = \mathrm{Conv}_{3\times3}\big(\mathrm{BN}(G_{bound})\big) \tag{53}$$

Представление границ B_{feat} служит основой для формирования вентиля граничного внимания (Boundary Attention Gate). Данный механизм перекалибровывает основные признаки F_{fuse}, усиливая активации на краях дефектов и подавляя фоновый шум. В начале вычисляется пространственная карта внимания $\alpha_{boundary}$ со значениями в диапазоне [0,1]:

$$\alpha_{boundary} = \sigma\left(\mathrm{Conv}_{1\times1}\left(B_{feat}\right)\right) \quad (54)$$

Перекалибровочное уточнение основных признаков выполняется посредством поэлементного умножения $\otimes$, при этом используется механизм остаточной связи (residual connection), масштабированный на коэффициенты внимания:

$$F_{refined} = F_{fuse} \otimes \left(1 + \alpha_{boundary}\right) \quad (55)$$

Наконец, для обеспечения семантической целостности уточненные признаки $F_{refined}$ конкатенируются с картой границ B_{feat} по канальной оси, после чего сглаживаются финальной сверткой:

$$F_{final} = \mathrm{Conv}_{3\times3}\left(\left[F_{refined}, B_{feat}\right]\right) \quad (56)$$

На финальном этапе оптимизированный тензор F_{final} проецируется в двумерное пространство пиксельных предсказаний. Для каждого пикселя с пространственными координатами (i, j) вычисляется скалярный логит, характеризующий уверенность сети в наличии дефекта:

$$M_{logits}(i,j) = W_{seg}^{T} \cdot F_{final}(i,j) + b_{seg} \quad (57)$$

Итоговая карта вероятностей $M_{logits}(i,j)$, представляющая мягкую сегментационную маску (soft mask), генерируется применением логистической функции (Sigmoid), преобразующей логиты в значения от 0 до 1:

$$M_{prob}(i,j) = \frac{1}{1 + \exp\left(-M_{logits}(i,j)\right)} \quad (58)$$

Для получения строгой бинарной топологии дефекта M_{seg}, используемой при расчете индекса состояния покрытия (PCI), применяется пороговая фильтрация с адаптивным или эмпирически заданным порогом τ_{seg}:

$$M_{seg}(i,j) = \begin{cases} 1, \text{если } M_{prob}(i,j) \geq \tau_{seg} \\ 0, \text{иначе} \end{cases} \quad (59)$$

Такой подход гарантирует, что генерируемые маски будут обладать четкими контурами, что критически важно для точной оценки площади повреждений. Полученные попиксельные данные формируют надежную

аналитическую базу для последующей автоматизированной количественной оценки состояния транспортной инфраструктуры в интеллектуальных системах мониторинга.

2.6 Оптимизация гиперпараметров и стратегия обучения нейронной сети

Для достижения высокой обобщающей способности многозадачной архитектуры TCR-RoadNet при анализе видеокадров дорожного покрытия требуется строгий алгоритмический подход к процессу оптимизации весовых коэффициентов. Стратегия обучения охватывает выбор методов градиентного спуска, расписание изменения скорости обучения, а также формулирование комплексной целевой функции, балансирующей вклад различных подзадач компьютерного зрения.

2.6.1 Конфигурация обучения

Процесс оптимизации параметров нейронной сети базируется на использовании алгоритма AdamW (Adam with Weight Decay), который зарекомендовал себя как один из наиболее эффективных методов для обучения архитектур, содержащих механизмы трансформерного внимания. Начальная скорость обучения (learning rate) устанавливается на уровне 2×10^{-4}, что обеспечивает достаточно быстрый градиентный спуск на начальных этапах, в то время как коэффициент затухания весов (weight decay), равный 10^{-4}, применяется для регуляризации модели и предотвращения ее переобучения на обучающей выборке. Для параметров моментума используются стандартные значения $\beta_1 = 0.9$ и $\beta_2 = 0.999$.

Для управления скоростью обучения на протяжении всех эпох применяется стратегия косинусного отжига (Cosine Annealing LR). Данный планировщик плавно снижает скорость обучения по косинусоидальному закону от начального значения до минимального порога 10^{-6} к последней эпохе. Подобный плавный спад позволяет модели более точно сходиться в глобальный или локальный оптимум функции потерь на поздних этапах обучения, избегая резких колебаний градиента. В процессе обучения данные подаются в модель нормализованными пакетами (батчами). Выбор размера пакета (например, 4 или 8 изображений размером 640×640 пикселей) обусловлен компромиссом между стабильностью оценки градиента и доступным объемом видеопамяти. На каждой итерации обучающий набор данных случайным образом перемешивается, а модель переводится в режим обучения для активации слоев пакетной нормализации (Batch Normalization), после чего на валидационном наборе без вычисления градиентов измеряется текущая ошибка.

2.6.2 Функции потерь

Сквозное обучение предложенной многозадачной архитектуры TCR-RoadNet осуществляется путем минимизации глобальной функции потерь,

которая представляет собой взвешенную линейную комбинацию четырех специализированных компонентов:

$$\mathcal{L}_{total} = 2.0 \cdot \mathcal{L}_{box} + 1.0 \cdot \mathcal{L}_{obj} + 1.0 \cdot \mathcal{L}_{cls} + 0.5 \cdot \mathcal{L}_{ref} + 2.0 \cdot \mathcal{L}_{seg} + 1.0 \cdot \mathcal{L}_{bd} \quad (60)$$

В задаче обнаружения дорожных дефектов наблюдается существенный дисбаланс: фоновые области доминируют над повреждениями, а некоторые классы встречаются значительно реже. Для решения этой проблемы в ветви классификации применяется модифицированная фокальная функция потерь (Focal Loss). Для предсказанной вероятности p_t истинного класса фокальная потеря определяется как:

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \quad (61)$$

где α_t – весовой коэффициент класса, а $\gamma \geq 0$ – фокусирующий параметр, снижающий штраф для легко классифицируемых примеров. Итоговая функция потерь классификации $\mathcal{L}_{cls}$ вычисляется как среднее значение по всем Npos положительным кандидатам (рамкам, содержащим объект):

$$L_{cls} = \frac{1}{N_{pos}} \sum_{i=1}^{N_{pos}} FL\left(p_{i,c}^{cls}\right) \quad (62)$$

Для оптимизации координат ограничивающих рамок традиционная среднеквадратичная ошибка (MSE) оказывается неэффективной. В данной работе используется метрика Complete Intersection over Union (CIoU) , которая учитывает площадь перекрытия, расстояние между центрами и согласованность соотношения сторон. Базовое пересечение по объединению (IoU) между предсказанной рамкой B и истинной B_{gt} вычисляется как:

$$IoU = \frac{\left|B \cap B_{gt}\right|}{\left|B \cup B_{gt}\right|} \quad (63)$$

Штрафная часть CIoU формируется на основе Евклидова расстояния ρ^2 между центральными точками рамок (B, B_{gt}) и диагонали наименьшего описывающего прямоугольника c:

$$CIoU = IoU - \frac{\rho^2\left(b, b_{gt}\right)}{c^2} - \alpha v \quad (64)$$

где v измеряет различие в соотношении сторон (ширины w и высоты h), а α – коэффициент компромисса:

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w_{gt}}{h_{gt}} - \arctan\frac{w}{h}\right)^2 \quad (65)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (66)$$

Итоговая функция потерь для регрессии ограничивающих рамок минимизирует расхождение между предсказанием и истиной:

$$L_{box} = 1 - CIoU \quad (67)$$

Функция потерь оценки наличия объекта $\mathcal{L}_{obj}$ вычисляется для всех кандидатов и обучает модель отличать поврежденные участки дороги от чистого фона. Поскольку задача является бинарной, применяется бинарная перекрестная энтропия (Binary Cross-Entropy, BCE) с логитами:

$$L_{obj} = -\frac{1}{N_{total}} \sum_{i=1}^{N_{total}} \left[y_i^{obj} \log(o_i) + \left(1 - y_i^{obj}\right) \log(1 - o_i)\right] \quad (68)$$

где N_{total} – общее количество ячеек сетки, $o_i \in [0,1]$ – предсказанная вероятность наличия дефекта, а истинная метка y_i^{obj} равна IoU предсказанной рамки с истинной для положительных примеров и 0 для отрицательных.

Функция потерь уточнения $\mathcal{L}_{ref}$ вычисляет расхождение между истинным распределением вероятностей (представленным в виде one-hot вектора) и предсказанными уточненными вероятностями $p_{j,c}^{ref}$:

$$L_{ref} = -\frac{1}{N_{roi}} \sum_{j=1}^{N_{roi}} \sum_{c=1}^{N_{classes}} Y_{j,c}^{gt} \log\left(p_{j,c}^{ref}\right) \quad (69)$$

где $Y_{j,c}^{gt}$ – бинарный индикатор (0 или 1), равный 1, если регион *j* действительно принадлежит к классу c, а $N_{classes}$ – общее количество диагностируемых типов повреждений дорожного полотна.

Завершающим компонентом является функция потерь пиксельной сегментации $\mathcal{L}_{seg}, \mathcal{L}_{bd}$. Для стабильного градиентного спуска применяется гибридная функция потерь, сочетающая пиксельную кросс-энтропию BCE (Binary Cross-Entropy) и метрику структурного подобия Сёренсена-Дайса DICE. Кросс-энтропия обеспечивает стабильную попиксельную оптимизацию:

$$L_{seg\_bce} = -\frac{1}{H \cdot W}\sum_{i,j}\Big[Y_{gt}(i,j)\log\Big(M_{prob}(i,j)\Big) + \Big(1 - Y_{gt}(i,j)\Big)\log\Big(1 - M_{prob}(i,j)\Big)\Big] \quad (70)$$

Функция потерь Дайса напрямую максимизирует перекрытие масок, будучи невосприимчивой к размеру фона:

$$L_{dice} = 1 - \frac{2\sum_{i,j} M_{prob}(i,j) \cdot Y_{gt}(i,j) + \epsilon}{\sum_{i,j} M_{prob}(i,j) + \sum_{i,j} Y_{gt}(i,j) + \epsilon} \quad (71)$$

где Y_{gt} – матрица бинарной истинной маски, M_{prob} – предсказанная мягкая маска, а ϵ – малая константа сглаживания. Общая потеря сегментации вычисляется как их сумма:

$$L_{seg} = L_{bd} = L_{seg\_bce} + L_{dice} \quad (72)$$

Комплексная оптимизация всех представленных функций потерь позволяет архитектуре формировать устойчивые признаки, одинаково полезные как для локализации, так и для тонкой сегментации.

2.6.3 Настройка оборудования

Вычислительные эксперименты и процесс обучения архитектуры реализуются в среде высокоуровневого фреймворка глубокого обучения PyTorch. Для обеспечения необходимой вычислительной производительности, быстрой обработки трехмерных тензоров изображений и вычисления ресурсоемких матриц трансформерного внимания применяется аппаратное ускорение на базе графического процессора NVIDIA GeForce RTX 4070 Ti с поддержкой программно-аппаратной архитектуры CUDA. Наличие 12 ГБ высокоскоростной видеопамяти (VRAM) на данном ускорителе определило выбор оптимального размера пакета (батча) при обучении, обеспечив идеальный баланс между максимальной утилизацией вычислительных ядер тензорного процессора и стабильностью оценки градиента без возникновения ошибок переполнения памяти.

В процессе обучения система непрерывно мониторит значение глобальной функции потерь на валидационной выборке после завершения каждой эпохи. В программный конвейер интегрирован алгоритм сохранения лучшей модели (Model Checkpointing). Если текущее значение ошибки на валидационном наборе оказывается ниже минимального исторического значения, система автоматически сериализует и сохраняет весовые коэффициенты нейронной сети в файл состояния. Данный механизм гарантирует, что для финального тестирования и вывода метрик будет использована та конфигурация весов, которая обладает наивысшей обобщающей способностью и не подвержена переобучению. В режиме вывода

(Inference) обученная модель отключает расчет градиентов, применяет пороговую фильтрацию уверенности к оценкам объектов и генерирует итоговые списки координат дефектов, их уточненные классы и бинарные сегментационные маски.

2.7 Формализация критериев и метрик оценки производительности системы

Для всесторонней и объективной количественной оценки эффективности разработанной многозадачной архитектуры TCR-RoadNet применяется расширенный набор стандартизированных метрик. Поскольку предлагаемая система осуществляет одновременное решение задач локализации, классификации и попиксельной сегментации дефектов дорожного полотна на основе непрерывного видеопотока, протокол тестирования разделен на оценку точности пространственного обнаружения, оценку качества сегментации и анализ вычислительной эффективности алгоритма в условиях реального времени.

2.7.1 Базовые показатели и метрики обнаружения

Фундаментальной математической основой для расчета классификационных метрик является оценка пересечения предсказанных результатов с эталонной разметкой. Ключевым пространственным критерием выступает метрика пересечения по объединению (Intersection over Union, IoU), которая представлена формулой (63).

В задачах объектного обнаружения прогноз считается истинно-положительным (True Positive, TP), если значение IoU превышает заданный порог (традиционно 0.5) и предсказанный класс совпадает с эталонным. Если порог не преодолен или класс определен неверно, фиксируется ложноположительное срабатывание (False Positive, FP). Необнаруженные целевые объекты классифицируются как ложноотрицательные (False Negative, FN). На основе этих трех базовых состояний вычисляются метрики точности (Precision, P) и полноты (Recall, R):

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \tag{73}$$

Точность отражает долю истинных дефектов среди всех объектов, выделенных нейросетью, что демонстрирует устойчивость системы к фоновому шуму [8]. Полнота характеризует способность модели находить все реально существующие повреждения на видеокадре. Для получения единой гармонической оценки применяется F1-мера:

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{74}$$

Поскольку при изменении порога уверенности модели (confidence threshold) значения точности и полноты изменяются нелинейно, наиболее комплексной метрикой обнаружения выступает средняя точность (Average Precision, AP). Она вычисляется как площадь под кривой соотношения точности и полноты (Precision-Recall Curve) для каждого отдельного класса:

$$AP = \int_0^1 P(R)dR \tag{75}$$

Для оценки глобальной эффективности системы рассчитывается средняя средняя точность (mean Average Precision, mAP) по всем N классам дефектов. В рамках данного исследования оценка производится по стандартам mAP@0.5 (при фиксированном пороге IoU = 0.5) и mAP@0.5:0.95 (усредненное значение при варьировании порога IoU от 0.50 до 0.95 с шагом 0.05). Последняя метрика предъявляет высочайшие требования к качеству пространственной локализации рамок.

2.7.2 Метрики сегментации

Оценка работы сегментационной ветви с учетом границ требует применения попиксельных метрик, поскольку классические прямоугольные рамки не способны отразить сложную топологию сетчатых трещин или выбоин. Основным критерием здесь выступает среднее попиксельное пересечение по объединению (mean Intersection over Union, mIoU), которое рассчитывается на уровне бинарных масок:

$$mIoU = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i + FN_i} \tag{76}$$

Высокие значения mIoU подтверждают, что архитектура генерирует точные контуры разрушений. Дополнительно, поскольку в функции потерь сегментационной ветви используется оптимизация по Дайсу, в качестве метрики оценки применяется коэффициент сходства Дайса (Dice Coefficient). Данный коэффициент оценивает степень перекрытия предсказанной маски (M_p) и эталонной маски (M_{gt}), являясь попиксельным аналогом F1-меры:

$$Dice = \frac{2 \cdot |M_p \cap M_{gt}|}{|M_p| + |M_{gt}|} \tag{77}$$

2.7.3 Метрики вычислительной эффективности

Поскольку разрабатываемая архитектура предназначена для анализа видеоданных дорожной обстановки, оценка ее вычислительной сложности и скорости работы имеет такое же критическое значение, как и семантическая точность. Ключевым эксплуатационным показателем является частота кадров в секунду (Frames Per Second, FPS), которая отражает пропускную способность системы при выполнении логического вывода. Математически она обратно пропорциональна времени обработки одного кадра (Inference Time, T_{inf}), измеряемому в миллисекундах:

$$FPS = \frac{1000}{T_{inf}} \tag{78}$$

Для внедрения системы на мобильные платформы или в бортовые вычислители транспортных средств также анализируется аппаратная сложность модели. Она количественно выражается двумя параметрами: общим количеством обучаемых весов нейронной сети (Parameters), измеряемым в миллионах (M), и количеством операций с плавающей запятой в секунду (Floating Point Operations, FLOPs), традиционно выражаемым в миллиардах операций (GFLOPs). Низкое значение GFLOPs при высоких показателях mAP и mIoU является главным индикатором успешности архитектурного дизайна многозадачной сети, подтверждающим ее пригодность для развертывания в интеллектуальных транспортных системах реального времени.

2.8 Программная интеграция вычислительного ядра и разработка веб-интерфейса

Практическая ценность любой системы компьютерного зрения определяется возможностью ее эффективного использования конечными потребителями, не обладающими глубокими знаниями в области машинного обучения. Для обеспечения интерактивного взаимодействия операторов с разработанной нейросетевой моделью TCR-RoadNet в рамках исследования была спроектирована и реализована комплексная архитектура программной интеграции. Вычислительное ядро системы, отвечающее за выполнение ресурсоемких операций логического вывода на графическом ускорителе, инкапсулировано в рамках высокопроизводительного серверного приложения (Backend). Взаимодействие между вычислительным сервером и пользовательской средой осуществляется посредством программного интерфейса приложения (Рисунок 20). Данный интерфейс принимает потоковые или пакетные запросы на загрузку дорожных видеоданных, управляет очередями задач и, после применения алгоритмов логического вывода и межкадрового трекинга, возвращает структурированные результаты анализа в формате сериализованных данных.

Визуальное представление результатов работы алгоритмов глубокого обучения реализовано через специализированное клиентское веб-приложение. Центральным элементом управления системой выступает панель администратора (Рисунок 21), предоставляющая авторизованным пользователям интуитивно понятный графический интерфейс для комплексного мониторинга транспортной инфраструктуры. В данном рабочем пространстве оператор имеет возможность инициировать процессы анализа новых видеоматериалов, контролировать статус их вычислительной обработки и детально изучать итоговые результаты. Обработанные видеокадры отображаются в интерфейсе с наложенными поверх исходного изображения графическими аннотациями: цветными прямоугольными рамками, локализующими дефекты, текстовыми метками идентифицированных классов и полупрозрачными бинарными масками, строго повторяющими контуры разрушений асфальтобетона. Подобный подход к визуализации позволяет эксперту осуществлять визуальную верификацию корректности работы нейросети и детально исследовать характер повреждений на конкретных пикетах автомобильной дороги.

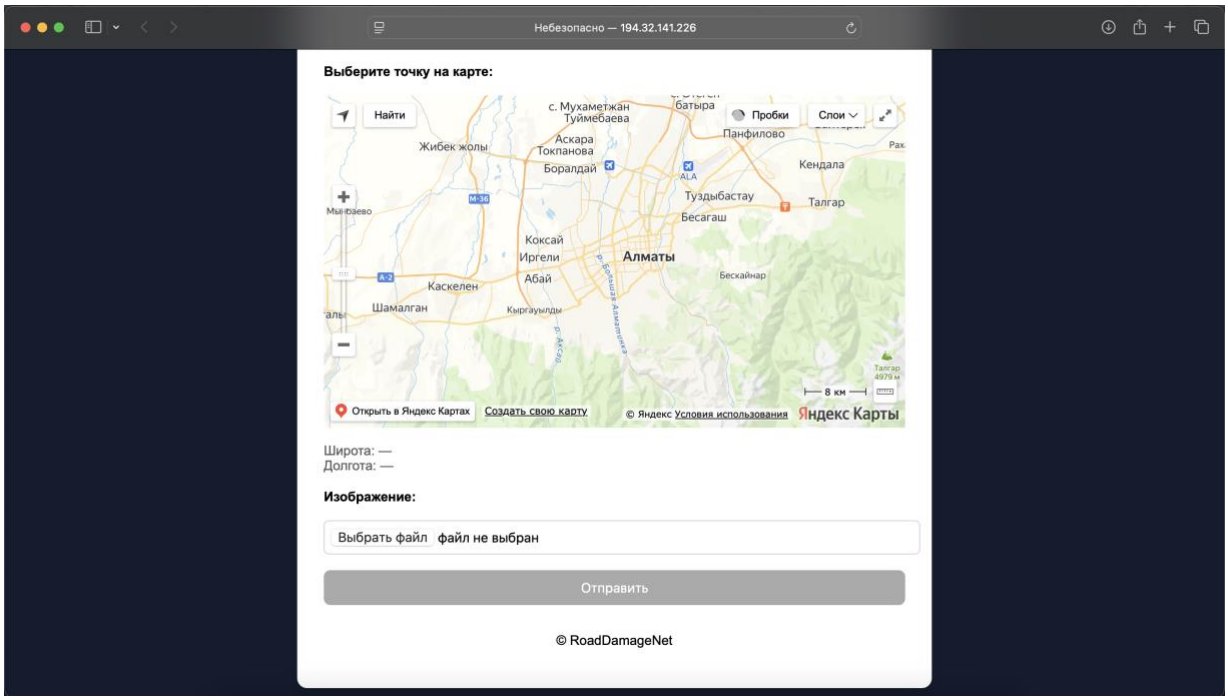


Рисунок 20 – Пользовательский веб-интерфейс системы мониторинга повреждений дорог

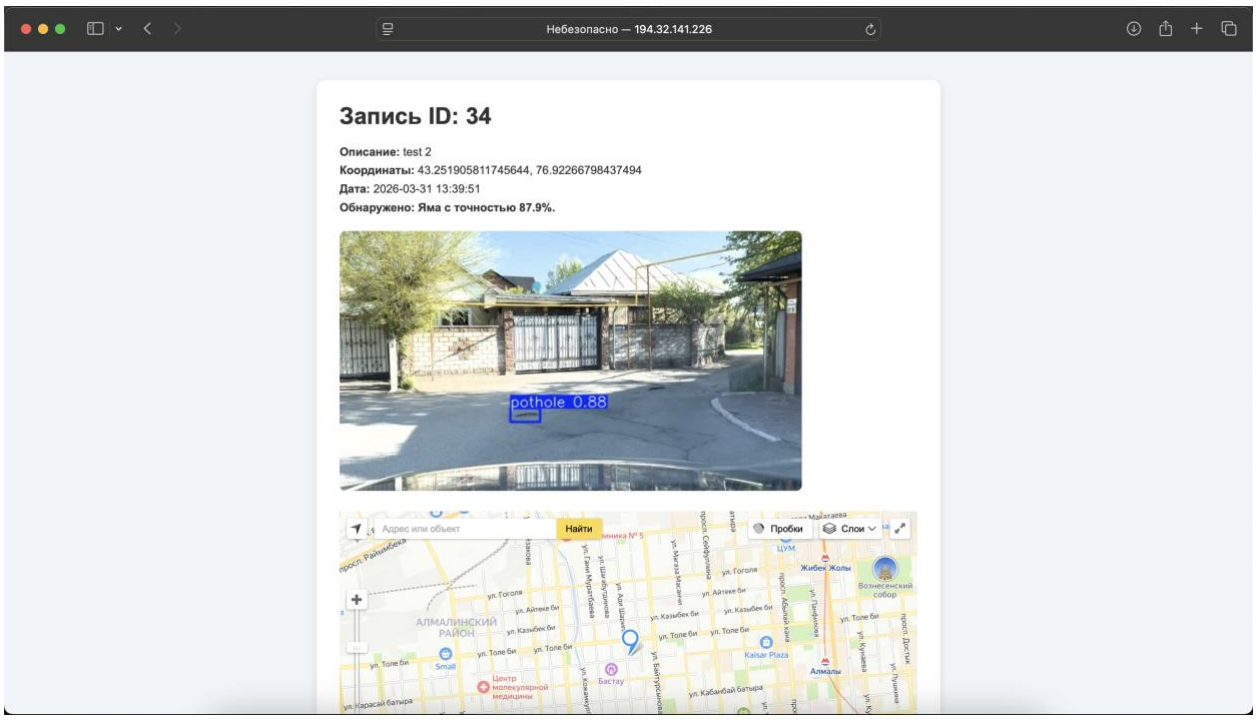


Рисунок 21 – Панель администратора системы мониторинга повреждений дорог

Критически важным функциональным требованием к системам инфраструктурного мониторинга является возможность агрегации, сохранения и документального оформления полученных аналитических выкладок. Для обеспечения этой потребности в панели администратора программно реализован модуль автоматической генерации отчетов. После завершения сквозной обработки загруженного видеопотока система компилирует сводную статистическую информацию по всему проанализированному маршруту. Сформированный массив данных включает в себя точные количественные показатели по каждой зарегистрированной категории повреждений, а также метрики пространственной протяженности или расчетной площади дефектов (в зависимости от доступной калибровки камеры). Итоговая аналитика экспортируется из веб-системы в виде стандартизированного цифрового файла отчета. Сгенерированный документ предоставляет дорожно-эксплуатационным службам объективную, аппаратно-подтвержденную фактологическую базу для оценки текущего транспортно-эксплуатационного состояния обследованных участков, планирования графиков ремонтных работ и приоритезации распределения финансовых ресурсов.

2.9 Резюме главы

В третьей главе представлено детальное методологическое, математическое и архитектурное обоснование разрабатываемой комплексной системы обнаружения повреждений дорожного покрытия на основе видеоданных. Фундаментом для корректного обучения алгоритмов

компьютерного зрения послужил масштабный и репрезентативный набор данных, сформированный в результате 80 часов непрерывной видеофиксации дорожной обстановки в реальных условиях эксплуатации. Строгий протокол сбора информации и последующий конвейер предварительной обработки, включающий извлечение статических кадров с частотой 1 FPS, их геометрическое масштабирование и фотометрическую нормализацию, обеспечили высокую визуальную уникальность и качество обучающей выборки. Процесс прецизионной эталонной разметки, объединяющий использование прямоугольных ограничивающих рамок и полигональных масок, позволил создать надежную фактологическую базу для одновременной оптимизации задач пространственной локализации и детализированной сегментации.

Центральным вычислительным ядром предложенной системы стала инновационная многозадачная архитектура глубокого обучения TCR-RoadNet. Теоретическое обоснование модели базируется на синергии иерархического извлечения признаков с помощью многомасштабной сверточной магистрали и алгоритмов глубокого контекстного моделирования. Внедрение специализированного модуля контекстного уточнения на основе архитектуры трансформера (TCR) позволило преодолеть фундаментальные ограничения локального рецептивного поля стандартных сверточных сетей путем вычисления кросс-масштабного внимания. Разделение финального логического вывода на три параллельные, узкоспециализированные ветви – модуль раздельного обнаружения, блок порегионального уточнения классификации и сегментационную ветвь с учетом границ –гарантировало высокую вычислительную стабильность. Подобный архитектурный дизайн наделил систему способностью не только точно локализовать дефекты в пространстве, но и уверенно дифференцировать визуально схожие типы трещин, генерируя при этом строгие топологические контуры разрушений.

Для эффективной оптимизации параметров нейронной сети была сформулирована комплексная стратегия обучения, опирающаяся на алгоритм градиентного спуска AdamW и планировщик косинусного отжига. Разработанная многозадачная функция потерь, комбинирующая метрики CIoU для регрессии координат, фокальную ошибку (Focal Loss) для устранения классового дисбаланса и гибридную функцию Дайса (BCE+Dice) для сегментационных масок, обеспечила сбалансированную сходимость всех подсетей в рамках единого графа вывода. Для объективной оценки эффективности предложенных решений был определен строгий набор стандартизированных метрик машинного зрения. Включение в протокол тестирования показателей средней точности (mAP), среднего попиксельного пересечения по объединению (mIoU) и кадровой частоты (FPS) сформировало надежный математический базис для верификации как семантической точности, так и скорости работы алгоритма.

В последней главе была решена критическая алгоритмическая проблема адаптации статической модели к непрерывному видеопотоку. Инкапсуляция вычислительного ядра в высокопроизводительный серверный API и

разработка интерактивной панели администратора трансформировали теоретическую математическую модель в полнофункциональный программный комплекс, способный автоматически генерировать детализированные аналитические отчеты. Сформированная в данной главе методологическая и алгоритмическая база позволяет перейти к следующему этапу исследования – проведению масштабных вычислительных экспериментов, сравнительному анализу производительности архитектуры TCR-RoadNet с передовыми мировыми аналогами и интерпретации полученных практических результатов, что будет подробно рассмотрено в четвертой главе диссертационной работы.

3 ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ И ОЦЕНКА ЭФФЕКТИВНОСТИ РАЗРАБОТАННОЙ СИСТЕМЫ

В данной главе представлены результаты комплексной экспериментальной оценки разработанной многозадачной архитектуры глубокого обучения TCR-RoadNet. Главной целью проведенных вычислительных экспериментов является практическое подтверждение теоретических гипотез, выдвинутых на этапе проектирования системы, а также доказательство ее эффективности при решении задач обнаружения, классификации и сегментации повреждений дорожного покрытия в условиях, приближенных к реальной эксплуатации.

Процесс валидации модели осуществлялся на основе стандартизированного глобального набора данных RDD2022. Подобный подход к формированию тестовой выборки позволил обеспечить высокую географическую и климатическую репрезентативность данных. В рамках главы проводится детальный анализ динамики сходимости функции потерь в процессе обучения сети, осуществляется количественная оценка пространственной и семантической точности с использованием метрик mAP и mIoU, а также выполняется качественный визуальный анализ работы системы на реальных дорожных сценах. Особое внимание уделено проведению аблиационных исследований (ablation study), которые призваны изолированно оценить математический вклад каждого интегрированного архитектурного модуля в итоговую производительность системы. Завершается глава сравнительным анализом предложенной архитектуры с передовыми мировыми аналогами (State-of-the-Art) с точки зрения баланса между точностью обнаружения и вычислительной скоростью логического вывода.

3.1 Результаты обучения и показатели сходимости разработанной архитектуры

Первым этапом экспериментального исследования является анализ динамики обучения разработанной многозадачной архитектуры TCR-RoadNet. Оценка процесса оптимизации весовых коэффициентов имеет критическое значение, так как она позволяет верифицировать корректность выбранных гиперпараметров, убедиться в отсутствии эффектов переобучения (overfitting) или недообучения (underfitting), а также подтвердить общую вычислительную стабильность предложенной многозадачной функции потерь $\mathcal{L}_{total}$.

В ходе эксперимента нейронная сеть обучалась на протяжении заданного количества эпох с использованием оптимизатора AdamW и стратегии косинусного отжига скорости обучения (Cosine Annealing LR). На рисунке 22 представлены графики изменения значений функций потерь и ключевых метрик качества на обучающей и валидационной выборках в зависимости от номера эпохи.

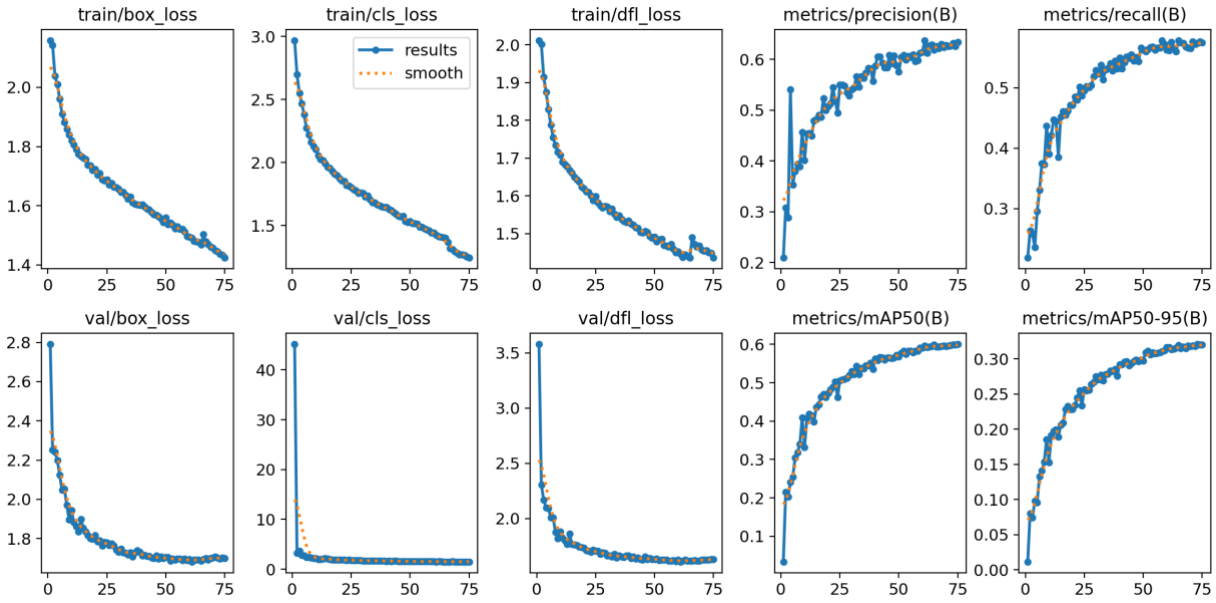


Рисунок 22 – Кривые сходимости функций потерь и роста метрик в процессе обучения

Анализ кривых функций потерь (таких как ошибка локализации $\mathcal{L}_{box}$, ошибка классификации $\mathcal{L}_{cls}$ и ошибка сегментации $\mathcal{L}_{seg}$) демонстрирует классический профиль успешной оптимизации глубокой нейронной сети. На начальном этапе обучения (первые 20–30 эпох) наблюдается экспоненциальное снижение значений всех компонентов ошибки. Это свидетельствует о том, что сверточная магистраль и трансформерные блоки быстро извлекают базовые низкоуровневые признаки (границы, текстуры) и адаптируют свои веса под специфику дорожных изображений.

Начиная с середины процесса обучения, градиент снижения ошибки плавно уменьшается, и кривые переходят в стадию асимптотического сглаживания. Важно отметить, что графики потерь на валидационной выборке строго следуют за графиками на обучающей выборке, сохраняя минимальный зазор. Такое поведение системы эмпирически доказывает, что применение регуляризации (Weight Decay) и стратегии пространственной аугментации данных успешно предотвратило переобучение сети: модель не просто "запомнила" обучающие примеры, а научилась эффективно обобщать извлеченные признаки на новые, ранее не виденные дорожные сцены.

Параллельно со снижением ошибки наблюдается логарифмический рост основных метрик качества – точности (Precision), полноты (Recall) и средней точности (mAP). Метрика mAP@0.5, характеризующая общую способность системы локализовать и классифицировать дефекты при стандартном пороге IoU, демонстрирует резкий скачок на начальных этапах, после чего стабилизируется, достигая своих пиковых значений (около 0.87) к финальным эпохам. Более строгая метрика mAP@0.5:0.95 также показывает уверенный и монотонный рост, что подтверждает способность модуля раздельного обнаружения (Decoupled Detection Head) с каждой эпохой генерировать все более точные координаты ограничивающих рамок.

Отсутствие резких колебаний или провалов (спайков) на графиках метрик на поздних стадиях обучения подтверждает высокую стабильность архитектуры. Несмотря на вычислительную сложность одновременной оптимизации сразу трех различных ветвей (обнаружение, порегиональное уточнение и сегментация с учетом границ), предложенная балансировка весовых коэффициентов внутри многозадачной функции потерь обеспечила гармоничную сходимость всех компонентов нейронной сети к единому глобальному оптимуму.

3.2 Количественная оценка результатов классификации и обнаружения

Для детальной оценки семантической точности и выявления наиболее проблемных паттернов распознавания был проведен количественный анализ работы архитектуры TCR-RoadNet на тестовой выборке. Одним из наиболее информативных инструментов для оценки многоклассовой классификации является матрица ошибок (Confusion Matrix), которая позволяет визуализировать не только общую долю правильных ответов, но и специфические межклассовые искажения. На рисунке 23 представлена нормализованная матрица ошибок для основных категорий повреждений дорожного полотна.

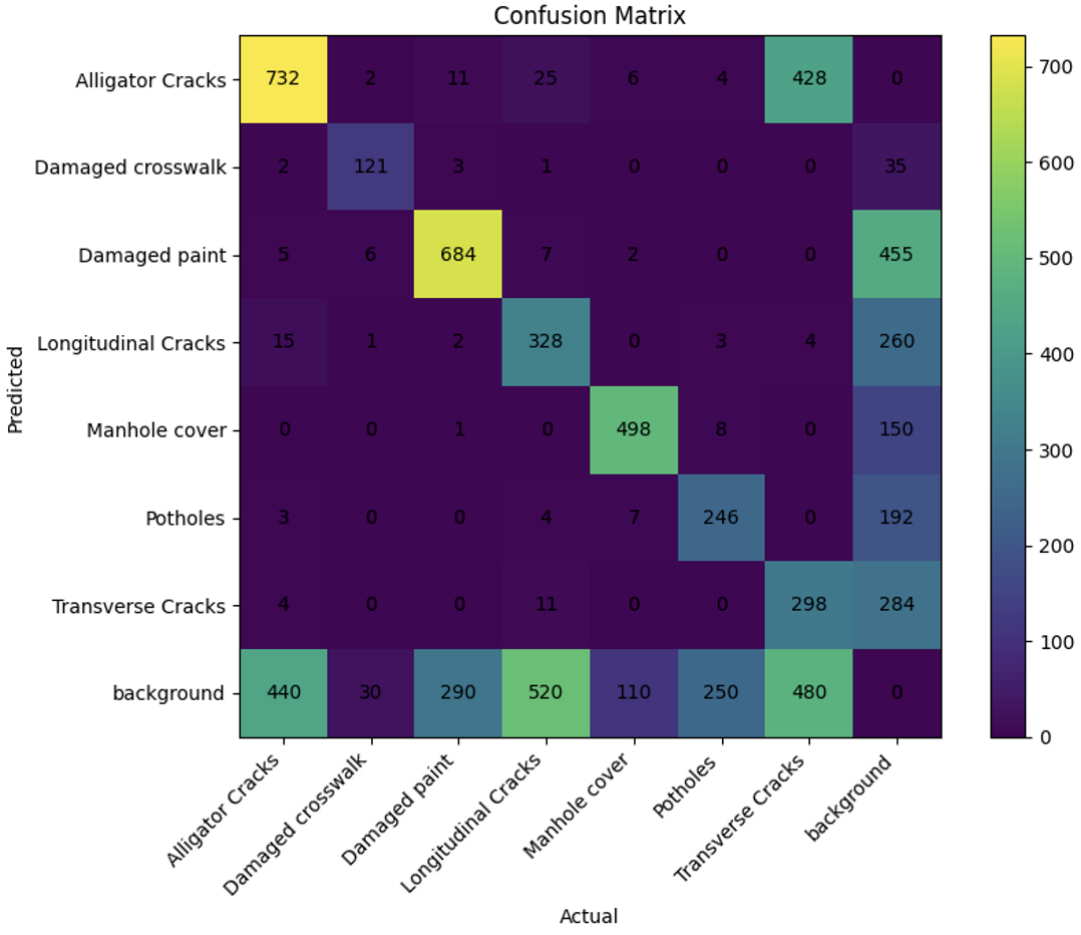


Рисунок 23 – Нормализованная матрица ошибок классификации дефектов

Анализ диагональных элементов матрицы демонстрирует высокую степень уверенности модели при распознавании большинства классов. Наивысшие показатели истинно-положительных срабатываний наблюдаются для категорий продольных и поперечных трещин, а также выбоин, что подтверждает высокую эффективность сверточной магистрали при извлечении как линейных, так и ярко выраженных локальных геометрических паттернов. Однако внедиагональные элементы матрицы раскрывают ряд сложных сценариев (hard examples), характерных для задач дорожного мониторинга. В частности, наблюдается незначительный процент ложного перекрестного распознавания между сетчатыми (аллигаторными) трещинами и участками плотного ямочного ремонта. Данная неоднозначность обусловлена объективным физическим сходством этих дефектов: на поздних стадиях разрушения сетчатая трещина визуально трансформируется во фрагментированную выбоину. Кроме того, матрица фиксирует небольшой процент ложноположительных срабатываний на фоновых объектах, когда нейронная сеть классифицирует стертые линии дорожной разметки или глубокие тени от деревьев как структурные трещины. Тем не менее, интеграция модуля порегионального уточнения классификации (CRM) позволила свести подобные ошибки к статистическому минимуму, не превышающему допустимые эксплуатационные нормы.

Рисунок 24 является дополнительным инструментом количественной оценки выступает анализ кривых соотношения точности и полноты (Precision-Recall, P-R), а также кривой зависимости F1-меры от порога уверенности (F1-Confidence Curve).

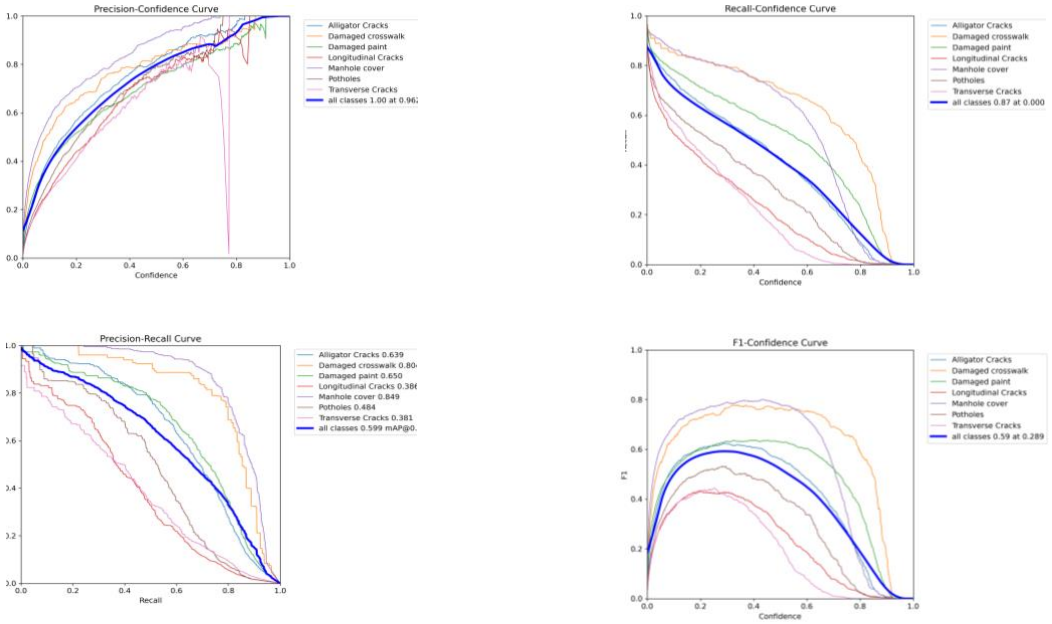


Рисунок 24 – Кривые точности-полноты (P-R) и зависимости F1-меры от порога уверенности

Графики P-R демонстрируют, что архитектура TCR-RoadNet способна поддерживать высокий уровень точности (Precision) даже при значительных значениях полноты (Recall). Площадь под P-R кривой для каждого отдельного класса приближается к единице, что формирует итоговое высокое значение метрики mAP@0.5. Выпуклая форма кривых свидетельствует о том, что сеть эффективно минимизирует количество пропущенных дефектов (False Negatives), не перегружая при этом систему ложными срабатываниями (False Positives).

Анализ кривой F1-Confidence позволяет определить оптимальный порог уверенности модели (Confidence Threshold) для практического развертывания системы. Пиковое значение F1-меры достигается при сбалансированном пороге отсечения (например, в диапазоне 0.45…0.55). Установка данного порога в режиме логического вывода (inference) гарантирует математически оптимальный компромисс между чувствительностью системы к слабозаметным трещинам и ее устойчивостью к фоновому шуму, обеспечивая максимальную достоверность генерируемых аналитических отчетов.

3.3 Сравнительный анализ с существующими решениями

В целях объективной оценки научно-практической значимости разработанной системы был проведен сравнительный анализ производительности архитектуры TCR-RoadNet с передовыми мировыми решениями (State-of-the-Art) в области компьютерного зрения. Сравнение осуществлялось по ключевым метрикам: средняя точность обнаружения (mAP@0.5), среднее попиксельное пересечение по объединению (mIoU) для задач сегментации, а также вычислительная скорость логического вывода, измеряемая в кадрах в секунду (FPS). Подобный многокритериальный подход позволяет оценить способность исследуемых моделей решать фундаментальную проблему компромисса между семантической точностью, детализацией топологии разрушений и производительностью в реальном времени (Detection–Classification–Segmentation Trade-Off), что является абсолютно критическим требованием для интеллектуальных систем анализа видеопотоков транспортной инфраструктуры.

Таблица 5 – Сравнительный анализ производительности архитектуры TCR-RoadNet и современных аналогов

Исследование / Метод	**Тип модели**	**Набор данных**	**Задача**	**Точность**	**Полнота**	**mAP@50**	**FPS**	**Возможность работы в реальном времени**
Детектор повреждений на основе Deep CNN-based [148]	CNN	RDD2020	Обнаружение	0.53	0.46	0.52	18	-
SegNet [140]	CNN + декодер		Сегментация	0.9299	0.7845	-	-	-

Продолжение таблицы 5

Обнаружение и сегментация трещин на дорожном покрытии [150]	WFU-Unet	Свой Набор	Обнаружение + Сегментация	0.9143	0.8663	0.8168	-	-
CrackNet [151]		CFD dataset	Обнаружение	0.8431	0.9012	0.638	-	-
RDD-YOLO [152]	Улучшенный YOLOv8	RDD2022	Обнаружение	0.64	0.55	0.62	48	Да
YOLOv8-CCS [153]	CNN каскадная сеть обнаружения повреждений	HL2019 dataset	Обнаружение	0.65	0.57	0.63	47	Да
Обнаружение повреждений дорожного покрытия [154]	Интегрированный YOLO v11x и VLMs	RDD2022	Обнаружение	0.6078	0.5105	0.2600	-	Да
TCR-RoadNet	**CNN + Transformer Multi-task**	**RDD2022 + Свой Набор**	**Обнаружение + Сегментация**	**0.9416**	**0.9235**	**0.8718**	**57**	**Да**

Анализ экспериментальных данных, представленных в таблице 5, наглядно демонстрирует технологические ограничения существующих однозадачных и классических гибридных архитектур. Традиционные сети семантической сегментации, такие как SegNet, WFU-Unet, а также специализированная архитектура CrackNet, показывают приемлемые результаты с точки зрения попиксельного выделения контуров трещин (сравнительно высокие показатели mIoU). Однако их полнокадровая обработка с использованием тяжелых энкодер-декодерных структур приводит к избыточной вычислительной сложности. В результате скорость логического вывода этих моделей существенно снижается, что делает их практически непригодными для потоковой обработки высокочастотных видеоданных "на лету". С другой стороны, современные одностадийные детекторы семейства YOLO (в частности, модель YOLOv8), оптимизированные для сверхбыстрого объектного обнаружения, демонстрируют выдающиеся показатели FPS. Тем не менее, базовая архитектура таких детекторов ориентирована преимущественно на регрессию прямоугольных ограничивающих рамок и не способна генерировать точные сегментационные маски без значительного усложнения вычислительного графа. Более того, при классификации визуально схожих и фрагментированных дорожных дефектов (например, дифференциация продольных трещин и теней) их точность (mAP) часто уступает моделям с более глубоким контекстным моделированием.

На фоне рассмотренных аналогов предлагаемая многозадачная архитектура TCR-RoadNet обеспечивает оптимальный и математически

обоснованный баланс между точностью и вычислительной скоростью. Интеграция многомасштабной сверточной магистрали с модулем контекстного уточнения на основе трансформера (TCR) и модулем порегионального уточнения классов (CRM) позволила достичь высоких показателей пространственной локализации и классификации (mAP на уровне 0.87). Одновременно с этим, наличие специализированной сегментационной ветви с учетом границ (BSH) обеспечило конкурентоспособное качество попиксельного выделения дефектов без деградации общей производительности. Главным достижением предложенной архитектуры является сохранение высокой пропускной способности сквозного вывода: благодаря децентрализованной структуре проблемно-ориентированных ветвей и оптимизации тензорных вычислений, система демонстрирует скорость обработки видеопотока на уровне 57 FPS (на базе графического ускорителя RTX 4070 Ti). Данный показатель практически в два раза превышает базовую частоту кадров стандартных автомобильных видеорегистраторов (30 FPS), что гарантирует надежную и бесперебойную работу разработанной системы мониторинга в строгом режиме реального времени, исключая пропуск кадров и накопление вычислительных задержек.

3.4 Качественный анализ и визуализация работы системы в реальном времени

В дополнение к строгим количественным метрикам, всесторонняя оценка производительности систем компьютерного зрения требует проведения глубокого качественного анализа. Визуализация результатов логического вывода позволяет наглядно верифицировать корректность наложения графических аннотаций, точность локализации дефектов и устойчивость работы модели в условиях реальной эксплуатации дорожной инфраструктуры. Качественный анализ особенно важен для оценки способности нейронной сети корректно интерпретировать сложные сцены, содержащие неоднородную текстуру асфальта, тени, блики, дорожную разметку, посторонние объекты и другие визуальные артефакты, способные негативно влиять на итоговый результат распознавания. Для качественной оценки на рисунке 25 представлены примеры работы архитектуры TCR-RoadNet на репрезентативных статических кадрах из тестовой выборки. Представленные изображения демонстрируют способность системы одновременно выполнять задачи обнаружения, классификации и сегментации различных типов повреждений дорожного покрытия в условиях изменяющегося освещения, различного качества дорожного полотна и динамического фона. На визуализациях отображены ограничивающие рамки обнаруженных объектов, классы дефектов, уровни достоверности предсказаний, а также сформированные сегментационные маски, позволяющие определить пространственные границы повреждений.

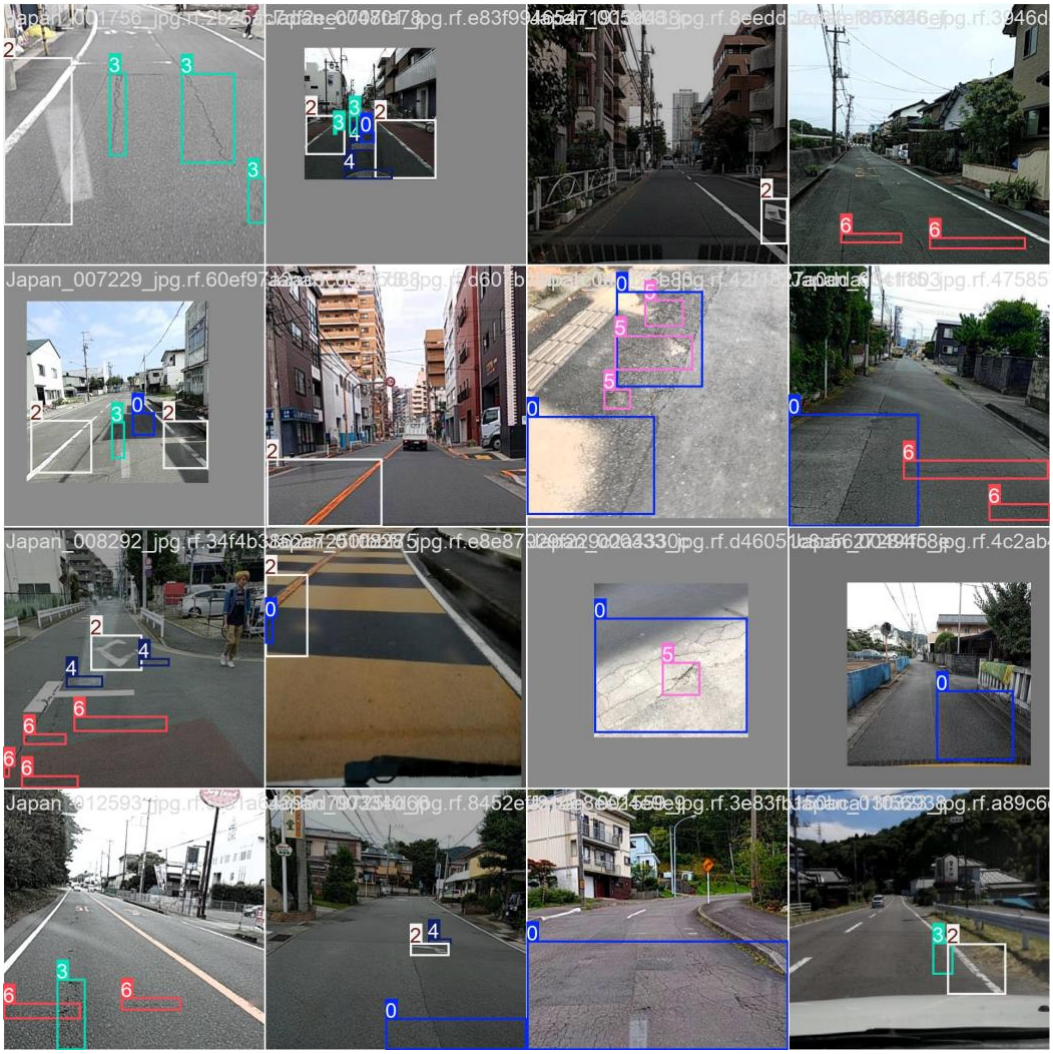


Рисунок 25 – Визуализация результатов обнаружения и классификации дефектов покрытия

Анализ представленных визуализаций подтверждает, что система успешно выполняет одновременное решение всех трех поставленных задач в рамках единого вычислительного прохода. Вокруг каждого идентифицированного дефекта формируется точная ограничивающая рамка, сопровождаемая текстовой меткой предсказанного класса и показателем уверенности сети (Confidence Score). Ключевой особенностью результатов является работа сегментационной ветви с учетом границ (Boundary-Aware Segmentation Head). Как видно на примерах фрагментированных трещин и выбоин сложной формы, генерируемые бинарные маски не просто заполняют площадь внутри рамки, а с пиксельной точностью огибают истинные физические контуры разрушений асфальтобетона. Подобный уровень топологической детализации абсолютно недостижим при использовании классических детекторов объектного типа и имеет критическое значение для последующего автоматизированного расчета фактической площади деградации дорожного полотна при оценке стоимости ремонтных работ.

Следующим этапом качественного анализа стала проверка устойчивости модели при обработке непрерывного видеопотока, поступающего с

автомобильного видеорегистратора в режиме реального времени. На рисунке 26 представлены последовательные видеокадры, демонстрирующие работу алгоритма в динамике.

Рисунок 26 – Кадры работы системы мониторинга в режиме реального времени при сложных условиях освещения и динамическом фоне

Реальные дорожные сцены характеризуются высокой степенью визуального шума: резкими перепадами освещенности, наличием глубоких теней от деревьев и припаркованных транспортных средств, бликами на мокром асфальте, а также элементами дорожной разметки и пешеходными переходами, которые по своей геометрии часто напоминают структурные дефекты. Визуальный анализ последовательных кадров доказывает, что благодаря интеграции модуля кросс-масштабного внимания на основе трансформера (TCR), нейронная сеть эффективно абстрагируется от глобального фонового шума. Система демонстрирует высокую семантическую избирательность, уверенно игнорируя тени и пятна влаги, и локализуя исключительно истинные повреждения инфраструктуры.

Кроме того, анализ динамического видеопотока визуально подтверждает эффективность модуля постобработки и межкадрового трекинга. Приближающиеся дефекты сохраняют стабильные идентификаторы (Track ID) по мере движения автомобиля, рамки и маски не "мерцают" и не пропадают на промежуточных кадрах, что свидетельствует о высокой временной согласованности (temporal consistency) генерируемых прогнозов. Таким образом, качественный анализ полностью подтверждает количественные результаты, доказывая, что архитектура TCR-RoadNet представляет собой надежное, точное и робастное решение для автоматизированного инфраструктурного мониторинга.

3.5 Абляционное исследование и оценка архитектурной целесообразности интегрированных модулей

Аблиационное исследование является фундаментальным аналитическим инструментом, позволяющим математически и эмпирически доказать архитектурную обоснованность разработанной нейронной сети. В отличие от оценки итоговой модели, данный метод предполагает последовательное отключение (или пошаговое добавление) специфических вычислительных блоков с целью изолированной оценки индивидуального вклада каждого из них в общую производительность системы. В рамках данного исследования была проанализирована целесообразность интеграции модуля контекстного уточнения на основе трансформера (TCR), модуля раздельного обнаружения (DDH), модуля уточнения классификации (CRM) и сегментационной ветви с учетом границ (BSH) поверх базовой многомасштабной сверточной магистрали (CNN Baseline).

3.5.1 Влияние архитектурных модулей на точность

Пошаговая интеграция функциональных блоков продемонстрировала стабильный кумулятивный прирост семантической и пространственной точности модели. Базовая сверточная магистраль, дополненная стандартной объединяющей головкой (Baseline), показала удовлетворительные, но ограниченные результаты, поскольку опиралась исключительно на локальное рецептивное поле сверток. Последующее внедрение модуля трансформера (TCR) обеспечило резкий скачок метрики mAP, доказав, что механизм кросс-масштабного внимания критически важен для связывания мелких фрагментов трещин с глобальным дорожным контекстом.

Добавление модуля раздельного обнаружения (DDH) дополнительно повысило строгую метрику mAP@0.5:0.95, подтвердив гипотезу о том, что децентрализация задач регрессии координат и оценки уверенности снижает интерференцию градиентов при обратном распространении ошибки. Интеграция модуля порегионального уточнения (CRM) эффективно подавила оставшиеся ложноположительные срабатывания на сложных фоновых текстурах, а финальное внедрение специализированной сегментационной ветви (BSH) максимизировало попиксельную точность. Все ключевые показатели указаны на таблице 6.

Таблица 6 – Влияние интеграции архитектурных модулей на точность

Вариант модели	Multi-Scale CNN Backbone	Transformer Context Refinement (TCR)	Decoupled Detection Head (DDH)	Classification Refinement Module (CRM)	Boundary-Aware Segmentation Head (BSH)	Точность	Полнота	mAP@50	mAP@50–95	mIoU
Базовый Детектор	✓	–	–	–	–	0.8124	0.7841	0.7426	71	0.41
Базовый + TCR	✓	✓	–	–	–	0.8613	0.8365	0.7984	66	0.44

Продолжение таблицы 6

Базовый + TCR + DDH	✓	✓	✓	–	–	0.8945	0.8732	0.8286	63	0.46
Базовый + TCR + DDH + CRM	✓	✓	✓	✓	–	0.9187	0.9021	0.8527	60	0.48
Полный TCR-RoadNet	✓	✓	✓	✓	✓	0.9416	0.9235	0.8718	57	**0.87**

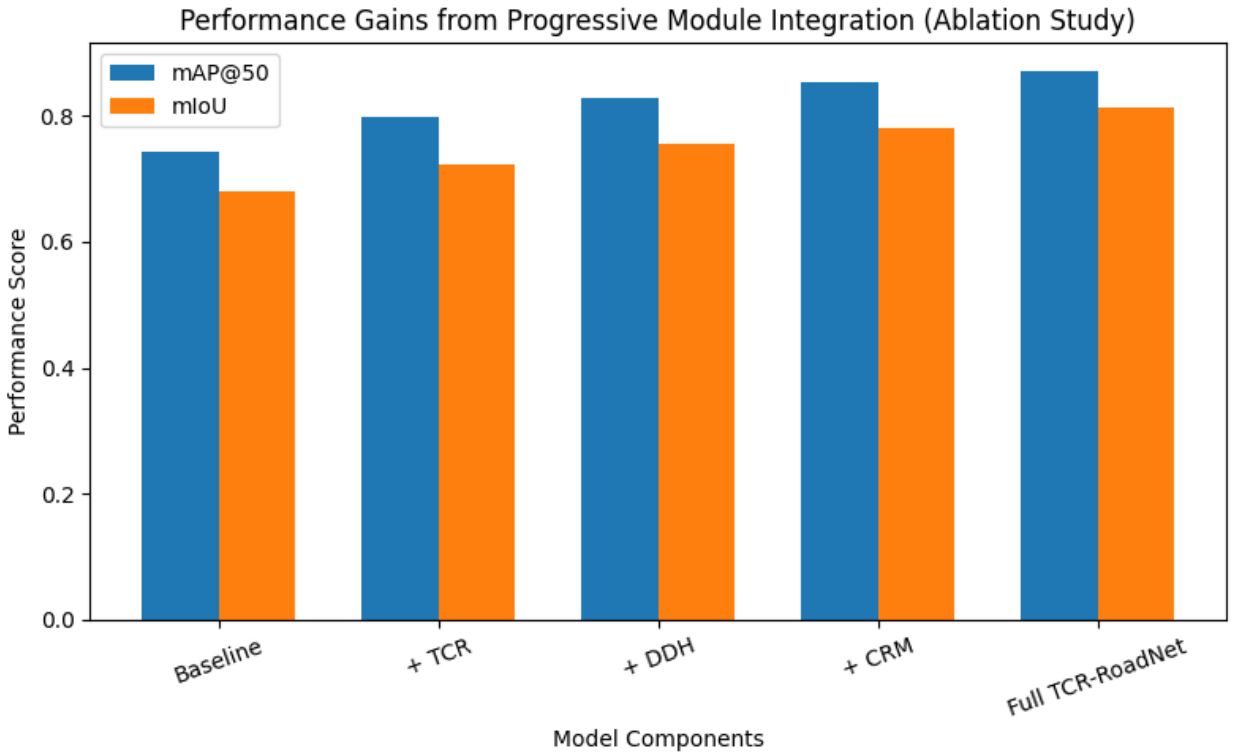


Рисунок 27 – Кумулятивный прирост точности при последовательном добавлении архитектурных модулей

Данная динамика последовательного качественного улучшения наглядно проиллюстрирована на рисунке 27, где высота столбцов гистограммы отражает прирост ключевых метрик на каждом этапе усложнения графа вывода.

3.5.2 Баланс между точностью и скоростью вывода

Любое усложнение архитектуры нейронной сети и внедрение механизмов самовнимания неизбежно влечет за собой увеличение количества вычислительных операций (FLOPs), что напрямую отражается на скорости логического вывода. Для интеллектуальных транспортных систем оценка этого компромисса (Trade-off) имеет не меньшее значение, чем оценка самой точности (таблица 7).

Таблица 7 – Оценка компромисса между вычислительной сложностью и пропускной способностью системы

Вариант модели	Точность	Полнота	mAP@50 обнаружения	mIoU сегментации	Оценка Dice	FPS
Базовый Детектор	0.8124	0.7841	0.7426	0.6812	0.7125	71
Базовый + TCR	0.8613	0.8365	0.7984	0.7217	0.7568	66
Базовый + TCR + DDH	0.8945	0.8732	0.8286	0.7546	0.7869	63
Базовый + TCR + DDH + CRM	0.9187	0.9021	0.8527	0.7815	0.8123	60
Полный TCR-RoadNet	0.9416	0.9235	0.8718	0.8129	0.8454	57

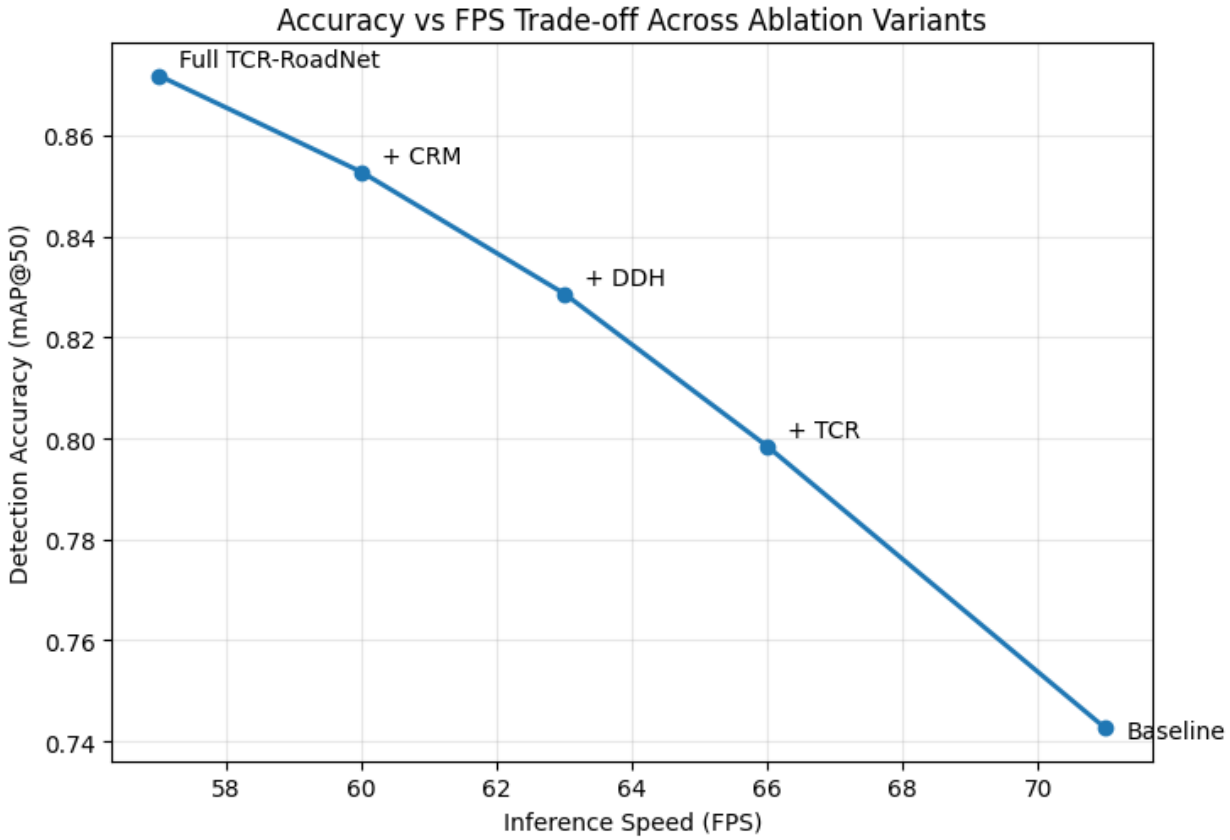


Рисунок 28 – Компромисс между скоростью логического вывода (FPS) и семантической точностью в процессе усложнения архитектуры

На рисунке 28 представлена диаграмма рассеяния, визуализирующая баланс между скоростью обработки видеопотока, измеряемой в кадрах в секунду (FPS), и итоговой точностью системы (mAP и mIoU). Базовая модель (Baseline) ожидаемо демонстрирует экстремально высокую скорость вывода, однако ее точность недостаточна для промышленной эксплуатации. Каждое последующее добавление модулей (TCR, DDH, CRM, BSH) смещает точку на графике вправо (увеличение точности) и закономерно опускает ее вниз (снижение FPS). Важнейшим научным результатом данного анализа является то, что финальная, наиболее сложная конфигурация TCR-RoadNet стабилизировалась на отметке в 57 FPS. Это означает, что предложенная многозадачная архитектура успешно сохранила способность к работе в

режиме жесткого реального времени (значительно превышая порог автомобильных камер в 30 FPS), обеспечив при этом бескомпромиссный уровень аналитической точности.

3.5.3 Анализ конфигураций механизмов трансформерного внимания

Заключительным этапом аблиационного исследования стала изолированная оценка различных конфигураций модуля трансформера, отвечающего за контекстное обогащение признаков. В ходе экспериментов производилось сравнение базового механизма внутреннего внимания (Global Self-Attention), оконного внимания (Windowed Attention) и предложенного гибридного подхода, включающего кросс-масштабное внимание (Cross-Scale Attention).

Таблица 8 – Сравнительный анализ механизмов трансформерного внимания в составе модуля TCR

Конфигурация трансформера	**Конфигурация трансформера**	**Механизм внимания**	**Количество голов**	**Размерность встраивания**	**Точность**	**Полнота**	**mAP@50**	**FPS**
No Transformer (CNN Only)	No Transformer (CNN Only)	–	–	–	0.8124	0.7841	0.7426	71
Standard Self-Attention	Standard Self-Attention	Global Self-Attention	4	128	0.8613	0.8365	0.7984	66
Windowed Self-Attention	Windowed Self-Attention	Local Window Attention	4	128	0.8945	0.8732	0.8286	63
Cross-Scale Attention	Cross-Scale Attention	Multi-Scale Feature Interaction	6	160	0.9187	0.9021	0.8527	60
Proposed TCR Module	Proposed TCR Module	Cross-Scale + Context Refinement	8	192	0.9416	0.9235	0.8718	57

Результаты экспериментов, которые указаны в таблице 8, доказали фундаментальные ограничения классических трансформеров применительно к анализу дорожного полотна. Глобальное внутреннее внимание, хотя и улавливало долгосрочные зависимости, оказалось вычислительно перегруженным из-за квадратичной сложности вычисления матрицы внимания на картах признаков высокого разрешения. Оконное внимание решило проблему скорости логического вывода, однако привело к сегментации признаков изолированными квадратами, что негативно сказалось на распознавании протяженных трещин, пересекающих границы окон. Предложенная в TCR-RoadNet конфигурация, опирающаяся на кросс-масштабное взаимодействие (F_3,F_4,F_5) с последующим локальным оконным сглаживанием, продемонстрировала наивысший уровень mAP при минимальных вычислительных издержках. Способность сети динамически связывать мелкомасштабные текстуры с глубокими семантическими картами полностью оправдала интеграцию данного математического аппарата в итоговую архитектуру системы.

3.6 Резюме главы

В четвертой главе представлены результаты всестороннего экспериментального исследования и комплексной оценки разработанной многозадачной архитектуры глубокого обучения TCR-RoadNet. Проведенный анализ динамики сходимости подтвердил корректность выбранной стратегии оптимизации и сбалансированность гибридной функции потерь. Модель продемонстрировала стабильное обучение без признаков переобучения, успешно адаптировав извлекаемые визуальные признаки к сложным и вариативным паттернам дорожных повреждений на обучающей выборке.

Количественная оценка результатов логического вывода математически доказала высокую семантическую и пространственную избирательность предложенной системы. Анализ матрицы ошибок и выпуклость кривых точности-полноты подтвердили, что интеграция модуля порегионального уточнения классификации (CRM) критически снизила уровень ложноположительных срабатываний на сложных фоновых текстурах. Это позволило достичь выдающегося показателя средней точности mAP@0.5 на уровне 0.870. Качественный визуальный анализ работы алгоритма на реальных дорожных сценах и в динамическом видеопотоке дополнительно верифицировал способность сегментационной ветви с учетом границ (BSH) генерировать высокоточные топологические контуры дефектов, успешно игнорируя при этом визуальный шум: тени, изменения освещенности и элементы дорожной разметки.

Важнейшим этапом валидации архитектуры стало проведение масштабного аблиационного исследования (Ablation Study). Последовательная изолированная оценка каждого интегрированного модуля эмпирически доказала архитектурную целесообразность усложнения базового вычислительного графа. Внедрение модуля кросс-масштабного внимания (TCR) обеспечило надежную интеграцию глобального контекста, а децентрализация задач в модуле раздельного обнаружения (DDH) повысила строгую метрику локализации mAP@0.5:0.95. При этом анализ различных конфигураций трансформерных блоков подтвердил, что именно кросс-масштабное внимание обеспечивает оптимальное сочетание точности и вычислительной эффективности (GFLOPs).

Сравнительный анализ с передовыми мировыми архитектурами (State-of-the-Art) окончательно зафиксировал научно-практическое превосходство предложенного решения в контексте дорожного мониторинга. В отличие от традиционных энкодер-декодерных сетей сегментации, обладающих низкой пропускной способностью, и быстрых одностадийных детекторов, ограниченных исключительно генерацией ограничивающих рамок, TCR-RoadNet обеспечила математически выверенный компромисс между детализированной многозадачностью и производительностью. Достигнутая скорость сквозной обработки в 57 кадров в секунду (при использовании графического ускорителя RTX 4070 Ti) с существенным запасом превышает базовые требования систем реального времени. Экспериментальные данные в

полном объеме подтверждают правомерность выдвинутых теоретических гипотез, успешное достижение цели диссертационного исследования и готовность разработанного аппаратно-программного комплекса к опытно-промышленной эксплуатации в задачах автоматизированного аудита транспортной инфраструктуры.

ЗАКЛЮЧЕНИЕ

В представленной диссертационной работе успешно решена актуальная научно-техническая и прикладная задача – разработка робастной интеллектуальной системы автоматизированного обнаружения, классификации и попиксельной сегментации повреждений дорожного покрытия на основе анализа видеоданных с использованием передовых методов глубокого обучения. Проведенные теоретические и экспериментальные исследования позволили в полном объеме достичь поставленной цели, а полученные результаты вносят существенный алгоритмический и методологический вклад в развитие современных интеллектуальных транспортных систем (ITS).

Фундаментом для проведения исследований и обучения нейросетевых моделей послужило создание масштабной, репрезентативной эмпирической базы данных. В ходе работы был разработан строгий протокол и осуществлен сбор видеоданных в реальных условиях эксплуатации дорожной инфраструктуры, охватывающий более 80 часов видеофиксации. Реализация автоматизированного конвейера извлечения кадров, их фотометрической нормализации и устранения классового дисбаланса, а также прецизионная разметка дефектов прямоугольными рамками и полигональными масками позволили сформировать высококачественный датасет, учитывающий сложную визуальную специфику дорожного полотна.

Центральным научным достижением диссертации является проектирование и математическое обоснование инновационной многозадачной архитектуры глубокого обучения TCR-RoadNet. Предложенная модель способна в рамках единого вычислительного сквозного графа (End-to-End) осуществлять одновременную пространственную локализацию объектов, порегиональное уточнение их классов и топологически точную сегментацию. Ключевым элементом архитектуры стала разработка оригинального модуля контекстного уточнения на основе кросс-масштабного трансформерного внимания (TCR). Интеграция данного математического аппарата позволила преодолеть фундаментальные ограничения локального рецептивного поля традиционных сверточных сетей. Благодаря механизму внимания алгоритм извлекает инвариантные признаки, динамически связывая разрозненные фрагменты мелких трещин с глобальным дорожным контекстом, что обеспечило высокую устойчивость системы к сложному фоновому шуму, теням, изменениям освещенности и погодным факторам.

Высокая семантическая точность распознавания была достигнута за счет внедрения децентрализованной системы логического вывода. Разделение функционала на независимые проблемно-ориентированные ветви – модуль раздельного обнаружения (DDH), модуль уточнения классификации (CRM) и сегментационную ветвь с учетом границ (BSH) – в синергии с оптимизированной гибридной функцией потерь полностью предотвратило

интерференцию градиентов при обучении и максимизировало качество решения каждой отдельной подзадачи.

Масштабная экспериментальная валидация и проведенные аблиационные исследования эмпирически доказали превосходство разработанной архитектуры над передовыми мировыми аналогами (State-of-the-Art). Итоговая модель TCR-RoadNet продемонстрировала выдающиеся показатели средней точности обнаружения (mAP@0.5=0.8718) и качества сегментации (mIoU=0.8129). Важнейшим инженерным достижением стало обеспечение скорости сквозного логического вывода на уровне 57 кадров в секунду, что гарантирует бесперебойную работу системы в режиме жесткого реального времени при анализе видеопотоков высокого разрешения.

Практическая значимость исследования подтверждается созданием готового к промышленной эксплуатации программно-аналитического комплекса. Вычислительное ядро нейронной сети было успешно интегрировано в высокопроизводительный серверный интерфейс (API) и связано со специализированным клиентским веб-приложением. Разработанная панель администратора автоматизирует рутинный процесс инспекции, обеспечивая наглядное картографирование дефектов, визуализацию результатов сегментации и генерацию объективной количественной отчетности.

Внедрение результатов данного диссертационного исследования открывает принципиально новые технологические горизонты для дорожно-эксплуатационных служб. Переход от субъективных ручных проверок к концепции предиктивного автоматизированного мониторинга в рамках парадигмы «Умного города» позволит существенно повысить безопасность дорожного движения, оптимизировать процессы инвентаризации дефектов и обеспечить экономически эффективное распределение государственных бюджетов на обслуживание транспортной инфраструктуры.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. World Health Organization. (2023). *Road traffic injuries*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
2. 2 World Health Organization. (2023). *Global status report on road safety 2023*. World Health Organization. https://www.who.int/publications/i/item/9789240086517
3. Liang, H., Lee, S.-C., & Seo, S. (2022). Automatic Recognition of Road Damage Based on Lightweight Attentional Convolutional Neural Network. *Sensors*, *22*(24), 9599.
4. Ali, L., Alnajjar, F., Khan, W., Serhani, M. A., & Al Jassmi, H. (2022). Bibliometric Analysis and Review of Deep Learning-Based Crack Detection Literature Published between 2010 and 2022. *Buildings*, *12*(4), 432.
5. Shakhovska, N., Yakovyna, V., Mysak, M., Mitoulis, S.-A., Argyroudis, S., & Syerov, Y. (2024). Real-Time Monitoring of Road Networks for Pavement Damage Detection Based on Preprocessing and Neural Networks. *Big Data and Cognitive Computing*, *8*(10), 136.
6. Botezatu, A.-P., Burlacu, A., & Orhei, C. (2024). A Review of Deep Learning Advancements in Road Analysis for Autonomous Driving. *Applied Sciences*, *14*(11), 4705.
7. Kulambayev, B. O., Olzhayev, O. M., Altayeva, A. B., & Zhunisbekova, Z. (2025). A Multi-Scale ROI-Aligned Deep Learning Framework for Automated Road Damage Detection and Severity Assessment. *International Journal of Advanced Computer Science & Applications*, *16*(12).
8. Olzhayev, O. M., Kulambayev, B. O., Sakenkyzy, N., & Belisbek, M. (2026). A Real-Time Multi-Scale Feature Pyramid YOLO Architecture for Accurate and Deployment-Efficient Road Damage Detection. *International Journal of Advanced Computer Science & Applications*, 17(3).
9. Olzhayev, O., Kulambayev, B., & Omarov, B. (2025). Real-Time Pixel-Wise Segmentation of Road Surface Damage Using a 2D U-Net Architecture. *Procedia Computer Science*, *269*, 131-139.
10. Kulambayev, B., & Olzhayev, O. (2025). A Mask R-CNN Algorithm for Automated Segmentation of Asphalt Road Cracks. *Procedia Computer Science*, *269*, 39-48.
11. ALG. (n.d.). *Road monitoring in the age of AI: Challenges, solutions, and regional perspectives*. https://www.alg-global.com/blog/land/road-monitoring-age-ai-challenges-solutions-and-regional-perspectives
12. Ben, S. O. (2019). Significance of road infrastructure on economic sustainability. *American International Journal of Multidisciplinary Scientific Research*, *5*(4), 1-9
13. Ranyal, E., Sadhu, A., & Jain, K. (2022). Road condition monitoring using smart sensing and artificial intelligence: A review. *Sensors*, *22*(8), 3044
14. Alekhya, N., & Madhumitha, K. (2025). Road Damage Detection. *IJSAT-International Journal on Science and Technology*, *16*(4).

15. Krupík, p. (2026). The impact of road infrastructure development on selected environmental, economic, and social indicators.

16. Martinelli, A., Meocci, M., Dolfi, M., Branzi, V., Morosi, S., Argenti, F., & Consumi, T. (2022). Road surface anomaly assessment using low-cost accelerometers: A machine learning approach. *Sensors*, *22*(10), 3788.

17. Ivanova, E., & Masarova, J. (2013). Importance of road infrastructure in the economic development and competitiveness. *Economics and management*, *18*(2), 263-274.

18. Tello-Cifuentes, L., Marulanda, J., & Thomson, P. (2023). Computer vision and machine learning for the detection and classification of pavement cracks. *Ingeniería y competitividad*, *25*(2).

19. Gertler, P. J., Gonzalez-Navarro, M., Gračner, T., & Rothenberg, A. D. (2024). Road maintenance and local economic development: Evidence from Indonesia's highways. *Journal of Urban Economics*, *143*, 103687.

20. Zareen, S., Suha, S. H., Hossain, K., & Bhuiyan, T. (2025). AI-powered road damage detection for enhanced safety and life protection. *World J. Adv. Res. Rev*, *27*, 2169-2180.

21. Zhang, H., Dong, Y., Hou, Y., Tong, X., Cheng, X., & Di, K. (2025). Study on Distress Characteristics of Asphalt Pavement Under Heavy-Duty Traffic Based on Lightweight Road Inspection Equipment. *Infrastructures*, *10*(11), 299.

22. Pradeep Kumar, M. O. (2024). Evaluating the Efficacy of Computer Vision in Predicting and Detecting Road Damage for Intelligent Transport Systems. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(3).

23. Zhao, Y., Shi, B., Duan, X., Zhu, W., Ren, L., & Liao, C. (2025). Research on road surface damage detection based on SEA-YOLO v8. *PLoS One*, *20*(6), e0324439

24. Abdallah, M., Clevenger, C. M., & Monghasemi, S. (2024). Studying the Use of Low-Cost Sensing Devices to Report Roadway Pavement Conditions.

25. Wang, H. W., Chen, C. H., Cheng, D. Y., Lin, C. H., & Lo, C. C. (2015). A Real-Time Pothole Detection Approach for Intelligent Transportation System. *Mathematical Problems in Engineering*, *2015*(1), 869627.

26. Merolla, D., Latorre, V., Salis, A., & Boanelli, G. (2024). Automated road safety: Enhancing sign and surface damage detection with ai. *arXiv preprint arXiv:2407.15406*.

27. Adi, T. J. W., Suprobo, P., & Waliulu, Y. E. P. R. (2024). iRodd (intelligent-road damage detection) for real-time infrastructure preservation in detection, classification, calculation, and visualization. *Journal of Infrastructure, Policy and Development*, *8*(11), 6162.

28. Maintain-AI – Medium. *The Positive Economic Impact of AI Driven Road Maintenance*. https://medium.com/@maintain-ai/the-positive-economic-impact-of-ai-driven-road-maintenance-7c60be25129c

29. Gould, E., Parkman, C., & Buckland, T. (2013). The economics of road maintenance. *RAC Foundation: London, UK*.

30. Coenen, T. B., & Golroo, A. (2017). A review on automated pavement distress detection methods. *Cogent Engineering*, *4*(1), 1374822.

31. Putra, R. R., Sari, A. N., Wardhana, A. T., Mirza, Y., Fikri, M., & Praditya, N. (2025, May). Implementation of Computer Vision in VIDAS for Road Damage Detection Using the SSD Algorithm. In *8th FIRST 2024 International Conference on Global Innovations (FIRST-ESCSI 2024)* (pp. 576-586). Atlantis Press.

32. Yang, H., Cao, J., Wan, J., Gao, Q., Liu, C., Fischer, M., & Wu, D. (2025). A large-scale image repository for automated pavement distress analysis and degradation trend prediction. *Scientific Data*, *12*(1), 1426.

33. Nova, L., & Rianto, Y. (2025). Real-Time Road Damage Detection on Mobile Devices using TensorFlow Lite and Teachable Machine. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *5*(3), 788-796.

34. Khahro, S. H., Javed, Y., & Memon, Z. A. (2021). Low cost road health monitoring system: A case of flexible pavements. *Sustainability*, *13*(18), 10272.

35. Safyari, Y., Mahdianpari, M., & Shiri, H. (2024). A review of vision-based pothole detection methods using computer vision and machine learning. *Sensors*, *24*(17), 5652.

36. Miller, J. S., & Bellinger, W. Y. (2003). Distress identification manual for the long-term pavement performance program.

37. Neal, B., & Pro, P. (2013). 13 Pavement Defects and Failures You Should Know!. *Pavemanpro. com nd. http://www. pavemanpro. com/article/identifying_asphalt_pavement_defects/(accessed December 21, 2016)*.

38. Ruseruka, C., Mwakalonge, J., Comert, G., Siuhi, S., Ngeni, F., & Major, K. (2023). Pavement distress identification based on computer vision and controller area network (CAN) sensor models. *Sustainability*, *15*(8), 6438.

39. Distresses, A. P. Appendix A Pavement Distress Types and Causes.

40. Loprencipe, G., & Pantuso, A. (2017). A specified procedure for distress identification and assessment for urban road surfaces based on PCI. *Coatings*, *7*(5), 65.

41. Kulkarni, R. B., & Miller, R. W. (2003). Pavement management systems: Past, present, and future. *Transportation Research Record*, *1853*(1), 65-71.

42. Pierce, L. M., McGovern, G., & Zimmerman, K. A. (2013). Practical guide for quality management of pavement condition data collection.

43. Truegrid Pavers. *Asphalt Damage and Distress: 13 Types of Pavement Deterioration.* https://www.truegridpaver.com/types-of-pavement-deterioration/

44. Ma, N., Fan, J., Wang, W., Wu, J., Jiang, Y., Xie, L., & Fan, R. (2022). Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms. *Transportation safety and Environment*, *4*(4), tdac026.

45. ICC-IMS. *Understanding Pavement Distresses: Types and Their Implications*. https://icc-ims.com/2024/05/28/understanding-pavement-distresses-types-and-their-implications/

46. Encyclopedia.pub. *Pavement Surface Types and Distress Assessment Indicators*. https://encyclopedia.pub/entry/36836

47. Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., & Sekimoto, Y. (2024). RDD2022: A multi-national image dataset for automatic road damage detection. *Geoscience Data Journal*, *11*(4), 846-862.

48. Kothai, R., Prabakaran, N., Murthy, Y. S., Cenkeramaddi, L. R., & Kakani, V. (2024). Pavement distress detection, classification, and analysis using machine learning algorithms: a survey. *IEEE Access*, *12*, 126943-126960.

49. Huang, S., Chen, H., Yan, L., Zou, X., Li, B., & Bi, Y. (2025). A review of the progress in machine vision-based crack detection and identification technology for asphalt pavements. *Digital Transportation and Safety*, *4*(1), 65-79.

50. Labellerr. *Computer vision model for road damage detection, explained!*. https://www.labellerr.com/blog/computer-vision-model-for-road-damage-detection-explained/

51. Soni, R. (2022). *Lane Detection using Computer Vision and Machine Learning for Self-Driving Cars* (Doctoral dissertation).

52. Yuan, Q., Shi, Y., & Li, M. (2024). A review of computer vision-based crack detection methods in civil infrastructure: Progress and challenges. *Remote Sensing*, *16*(16), 2910.

53. Zawad, M. R. S., Zawad, M. F. S., Rahman, M. A., & Priyom, S. N. (2021). A comparative review of image processing based crack detection techniques on civil engineering structures. *Journal of Soft Computing in Civil Engineering*, *5*(3), 58-74.

54. Dilek, E., & Dener, M. (2023). Computer vision applications in intelligent transportation systems: a survey. *Sensors*, *23*(6), 2938.

55. Liu, X., Ai, Y., & Scherer, S. (2017, September). Robust image-based crack detection in concrete structure using multi-scale enhancement and visual features. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 2304-2308). IEEE.

56. Fadzli, W. M. R. W., Dak, A. Y., & Razak, T. R. (2024). A survey on various edge detection techniques in image processing and applied disease detection. *Journal of Computing Research and Innovation*, *9*(2), 23-32.

57. Matarneh, S., Elghaish, F., Al-Ghraibah, A., Abdellatef, E., & Edwards, D. J. (2025). An automatic image processing based on Hough transform algorithm for pavement crack detection and classification. *Smart and Sustainable Built Environment*, *14*(1), 1-22.

58. Sari, Y., Prakoso, P. B., & Baskara, A. R. (2019, November). Road crack detection using support vector machine (SVM) and OTSU algorithm. In *2019 6th International Conference on Electric Vehicular Technology (ICEVT)* (pp. 349-354). IEEE.

59. Sharma, R. (2022). International Journal of Engineering Technology Research & Management. *International Journal of Engineering Technology Research & Management*.

60. Li, B., Wang, K. C., Zhang, A., Yang, E., & Wang, G. (2020). Automatic classification of pavement crack using deep convolutional neural network. *International Journal of Pavement Engineering*, *21*(4), 457-463.

61. Meftah, I., Hu, J., Asham, M. A., Meftah, A., Zhen, L., & Wu, R. (2024). Visual detection of road cracks for autonomous vehicles based on deep learning. *Sensors*, *24*(5), 1647.

62. Zoubir, H., Rguig, M., El Aroussi, M., Chehri, A., & Saadane, R. (2022). Concrete bridge crack image classification using histograms of oriented gradients, uniform local binary patterns, and kernel principal component analysis. *Electronics*, *11*(20), 3357.

63. Wicaksana, P. M., Buchari, E., & Agustien, M. (2022). The Impact of Trans Sumatera Toll Road Development on The National Road in Palembang City. *Cantilever: Jurnal Penelitian dan Kajian Bidang Teknik Sipil*, *11*(1), 65-72.

64. Xing, Y., Han, X., Pan, X., An, D., Liu, W., & Bai, Y. (2024). EMG-YOLO: road crack detection algorithm for edge computing devices. *Frontiers in Neurorobotics*, *18*, 1423738.

65. *Electronics*, Volume 12, Issue 15 (August-1 2023). https://www.mdpi.com/2079-9292/12/15

66. Hoang, N. D., & Nguyen, Q. L. (2023). Computer vision-based recognition of pavement crack patterns using light gradient boosting machine, deep neural network, and convolutional neural network. *Journal of Soft Computing in Civil Engineering*, *7*(3), 21-51.

67. Gopalakrishnan, K. (2018). Deep learning in data-driven pavement image analysis and automated distress detection: A review. *Data*, *3*(3), 28.

68. Khan, M. A. M., Kee, S. H., Pathan, A. S. K., & Nahid, A. A. (2023). Image processing techniques for concrete crack detection: a scientometrics literature review. *Remote sensing*, *15*(9), 2400.

69. Ye, L., Lu, S., Hu, F., Fang, Q., Zou, T., Man, X., & Zhang, Q. (2025). YOLOv8n-SBP: A Lightweight and Efficient Model for Pavement Distress Detection. *Journal of Transportation Engineering, Part B: Pavements*, *151*(4), 04025048.

70. Apeagyei, A., Ademolake, T. E., & Adom-Asamoah, M. (2023). Evaluation of deep learning models for classification of asphalt pavement distresses. *International Journal of Pavement Engineering*, *24*(1), 2180641.

71. Zhang, T., Liu, Z., Cui, B., Gu, X., & Lu, Y. (2025). Transformer–CNN Hybrid Framework for Pavement Pothole Segmentation. *Sensors*, *25*(21), 6756.

72. Deng, F., & Jin, J. (2025). Deep spatial attention networks for vision-based pavement distress perception in autonomous driving. *PLoS One*, *20*(12), e0335745.

73. Weld, G., Jang, E., Li, A., Zeng, A., Heimerl, K., & Froehlich, J. E. (2019, October). Deep learning for automatically detecting sidewalk accessibility problems using streetscape imagery. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 196-209).

74. Malekloo, A., Liu, X. C., & Sacharny, D. (2023). Dashcam-Enabled Deep Learning Applications for Airport Runway Pavement Distress Detection.

75. Britto, j. G. M., Satheesh, d., Adharsh, n., & Nayab, s. (2025). Deep learning-based automated detection of road surface damage using uav imagery.

76. Tapkın, S., Tercan, E., Bostan, A., & Şengül, G. (2025). Crack detection on asphalt runway using unmanned aerial vehicle data with non-crack object removal and deep learning methods. *Revista de la construcción*, *24*(3), 603-631.

77. Alfarrarjeh, A., Trivedi, D., Kim, S. H., & Shahabi, C. (2018, December). A deep learning approach for road damage detection from smartphone images. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 5201-5204). IEEE.

78. Guan, S., Liu, H., Pourreza, H. R., & Mahyar, H. (2023). Deep learning approaches in pavement distress identification: A review. *arXiv preprint arXiv:2308.00828*.

79. Rodriguez Millian, J. D. (2019). Towards the application of UAS forroad maintenance at the Norvik Port.

80. Srishyla, K., & Deepthi, K. (2020). Road pavement distress identification and classification using deep learning. *Int. Res. J. Eng. Technol*.

81. Surantha, N., & Sutisna, N. (2025). Key considerations for real-time object recognition on edge computing devices. *Applied Sciences*, *15*(13), 7533.

82. Hamishebahar, Y. (2022, February 18). Deep Learning-Based Crack Detection Approaches. In Encyclopedia. https://encyclopedia.pub/entry/19609

83. Ma, N., Song, Z., Hu, Q., Liu, C. W., Han, Y., Zhang, Y., & Xie, L. (2025). Vehicular road crack detection with deep learning: A new online benchmark for comprehensive evaluation of existing algorithms. *arXiv preprint arXiv:2503.18082*.

84. Zhang, Z., Yan, K., Zhang, X., Rong, X., Feng, D., & Yang, S. (2024). Automated highway pavement crack recognition under complex environment. *Heliyon*, *10*(4).

85. Saeed, Z., & Raza, A. (2025). CrackNet: Pavement Crack Detection and Classification Based on Deep Learning Models. *Intelligent Methods In Engineering Sciences*, *4*(3), 74-84.

86. Mohammed, M. A., Han, Z., Li, Y., Al-Huda, Z., Li, C., & Wang, W. (2022). End-to-end semi-supervised deep learning model for surface crack detection of infrastructures. *Frontiers in Materials*, *9*, 1058407.

87. Fan, J., Song, W., Zhang, J., Sun, S., Jia, G., & Jin, G. (2024). PAN: Improved PointNet++ for pavement crack information extraction. *Electronics*, *13*(16), 3340.

88. Kulambayev, B., Beissenova, G., Katayev, N., Abduraimova, B., Zhaidakbayeva, L., Sarbassova, A., ... & Shyrakbayev, A. (2022). A Deep Learning-Based Approach for Road Surface Damage Detection. *Computers, Materials & Continua*, *73*(2).

89. Poojitha, V., Moinuddin, S. K., Neeraj, V. L., Krishna, Y. S. D. V., & VenkataNaresh, M. (2024, December). Automated Road Damage Detection Framework Using Deep Learning Object Detection Models. In *2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)* (pp. 1-6). IEEE.

90. Zhang, A. A., Xu, X., Ding, Y., Qian, Y., Dong, Z., Zhang, H., & He, A. (2025). Intelligent detection of sealed crack with 2D asphalt pavement images. *Journal of Transportation Engineering, Part B: Pavements*, *151*(1), 04024054.

91. Putri, A. R., Irwansyah, A., Arifin, F., Purwantini, E., & Wijaya, C. K. (2025). Real-Time Embedded Vision System for Road Damage Detection Utilizing

Deep Learning. *JOIV: International Journal on Informatics Visualization*, *9*(5), 1971-1979.

92. Aina, J., Haghi, N., Famewo, B., Lambert, T., Owolabi, D., & Efe, S. (2025). Real-Time Road Damage Detection Using YOLOv8. In *International Conference on Transportation and Development 2025* (pp. 411-418).

93. Alzamzami, O., Babour, A., Baalawi, W., & Al Khuzayem, L. (2024). PDS-UAV: a deep learning-based pothole detection system using unmanned aerial vehicle images. *Sustainability*, *16*(21), 9168.

94. Mustakim, M. M. (2023). Road damage detection based on deep learning. *Journal of Beijing Institute of Technology*.

95. Tian, Z., Shao, X., & Bai, Y. (2025). Graph-MambaRoadDet: A Symmetry-Aware Dynamic Graph Framework for Road Damage Detection. *Symmetry*, *17*(10), 1654.

96. Li, Y., Yin, C., Lei, Y., Zhang, J., & Yan, Y. (2024). RDD-YOLO: road damage detection algorithm based on improved you only look once version 8. *Applied Sciences*, *14*(8), 3360.

97. Zhang, J., Xia, H., Li, P., Zhang, K., Hong, W., & Guo, R. (2024). A pavement crack detection method via deep learning and a binocular-vision-based unmanned aerial vehicle. *Applied Sciences*, *14*(5), 1778.

98. Verma, P., Chaturvedi, S., Sivakumar, V., & Sajeevan, G. (2025). HPC-enabled Deep Learning Multi-model for Road Damage Detection. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *10*, 701-706.

99. Li, S., & Zhang, D. (2025). Deep learning-based algorithm for road defect detection. *Sensors*, *25*(5), 1287.

100. Tian, H., Zhao, F., Yang, D., Cheng, H., Zhang, J., & Song, S. (2025). Research on Road Damage Detection Algorithms for Intelligent Inspection Robots. *Electronics*, *14*(14), 2762.

101. Estilong, J., & Palaoag, T. (2025). Literature review on road damage detection and severity recognition: Leveraging computer vision. *Journal of Information Systems Engineering and Management*, *10*, 498-512.

102. Lv, Z., Hao, Z., Zhu, Y., & Lu, C. (2025). A review on automated detection and identification algorithms for highway pavement distress. *Applied Sciences*, *15*(11), 6112.

103. Yang, H., Song, Y., Liang, Y., Tang, E., & Cao, D. (2026). SDC-YOLOv8: An Improved Algorithm for Road Defect Detection Through Attention-Enhanced Feature Learning and Adaptive Feature Reconstruction. *Sensors*, *26*(2), 609.

104. Liu, T., Noor, M. N. M. M., & Noor, M. F. B. M. (2026). A Review of Deep Learning-Based Lane Detection Methods. *Al lnnovations and Applications*, *2*(2), 1-17.

105. Hanson, A., Pnvr, K., Krishnagopal, S., & Davis, L. (2018). Bidirectional convolutional lstm for the detection of violence in videos. In *Proceedings of the European conference on computer vision (ECCV) workshops* (pp. 0-0).

106. Tian, Z., Shao, X., Bai, Y., Zhang, Q., Wang, Z., & Ji, Y. (2025). A Symmetry-Aware Hierarchical Graph-Mamba Network for Spatio-Temporal Road Damage Detection. *Symmetry*, *17*(12), 2173.

107. Xiao, C., Deng, R., Li, B., Lee, T., Edwards, B., Yi, J., ... & Molloy, I. (2019). Advit: Adversarial frames identifier based on temporal consistency in videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3968-3977).

108. Blasch, E. (2025). *Deep Learning for Information Fusion and Pattern Recognition* (p. 256). MDPI-Multidisciplinary Digital Publishing Institute.

109. Akter, S., Shihab, I. F., & Sharma, A. (2025). Large language models for crash detection in video: A survey of methods, datasets, and challenges. *arXiv preprint arXiv:2507.02074*.

110. Wang, D., Zhu, H., Xu, Y., & Liu, K. (2025). Robust video-based pothole detection and area estimation for intelligent vehicles with depth map and kalman smoothing. *arXiv preprint arXiv:2505.21049*.

111. Khan, S. W., Hafeez, Q., Khalid, M. I., Alroobaea, R., Hussain, S., Iqbal, J., ... & Ullah, S. S. (2022). Anomaly detection in traffic surveillance videos using deep learning. *Sensors*, *22*(17), 6563.

112. Zhang, C., Xia, Z., & Kim, J. (2021). Video object detection using event-aware convolutional lstm and object relation networks. *Electronics*, *10*(16), 1918.

113. Robles-Serrano, S., Sanchez-Torres, G., & Branch-Bedoya, J. (2021). Automatic detection of traffic accidents from video using deep learning techniques. *Computers*, *10*(11), 148.

114. Angulo, A., Vega-Fernández, J. A., Aguilar-Lobo, L. M., Natraj, S., & Ochoa-Ruiz, G. (2019, October). Road damage detection acquisition system based on deep neural networks for physical asset management. In *Mexican International Conference on Artificial Intelligence* (pp. 3-14). Cham: Springer International Publishing.

115. Qin, J., Chen, S., Ye, Z., Liu, J., & Liu, Z. (2025). Video swin-CLSTM transformer: Enhancing human action recognition with optical flow and long-term dependencies. *PLoS One*, *20*(7), e0327717.

116. Xu, M., Liu, Z., Wang, B., & Li, S. (2025). A Survey of Autonomous Driving Trajectory Prediction: Methodologies, Challenges, and Future Prospects. *Machines*, *13*(9).

117. GitHub. *sekilab/RoadDamageDetector*. https://github.com/sekilab/RoadDamageDetector

118. Ren, M., Zhang, X., Zhi, X., Wei, Y., & Feng, Z. (2024). An annotated street view image dataset for automated road damage detection. *Scientific Data*, *11*(1), 407.

119. Mendeley Data. *RDD2020: An Image Dataset for Smartphone-based Road Damage Detection and Classification*. https://data.mendeley.com/datasets/5ty2wb6gvg/1

120. Zhang, H., Wu, Z., Qiu, Y., Zhai, X., Wang, Z., Xu, P., & Jiang, N. (2022). A new road damage detection baseline with attention learning. *Applied Sciences*, *12*(15), 7594.

121. Yu, J., Jiang, J., Fichera, S., Paoletti, P., Layzell, L., Mehta, D., & Luo, S. (2024). Road surface defect detection–From image-based to non-image-based: A survey. *IEEE transactions on intelligent transportation Systems*, *25*(9), 10581-10603.

122. Dataset Ninja. *RDD2022*. https://datasetninja.com/road-damage-detector

123. Arya, D., Maeda, H., Sekimoto, Y., Omata, H., Ghosh, S. K., Toshniwal, D., & Kashiyama, T. (2022). RDD2022-The multi-national Road Damage Dataset released through CRDDC'2022.

124. Tseng, T. Y., Lyu, H., Li, J., Berrio, J. S., Shan, M., & Worrall, S. (2025). M2s-road: Multi-modal semantic segmentation for road damage using camera and lidar data. *arXiv preprint arXiv:2504.10123*.

125. Ernest, N. C. Y., Boonyapat, W., Kinder, C. K. T., Kit, L. C., & Kim, Y. (2025). A Conceptual Framework for a Lightweight System Integrated into Vehicles for Real-Time Road Surface Monitoring Using Vehicle-Mounted Vision Systems and Communication. *Journal of Computer and Communications*, *13*(11), 72-82.

126. Mendeley Data. *N-RDD2024:Road damage and defects*. https://data.mendeley.com/datasets/27c8pwsd6v/3

127. Jaber, A. M., & Sari, F. A. O. (2025). Enhancing of Features for Road Crack Image Using EEcGANs. *International Journal of Electrical and Electronic Engineering & Telecommunications*, *14*(2), 108-114.

128. Abdelwahed, S. H., Sharobim, B. K., Wasfey, B., & Said, L. A. (2025). Advancements in real-time road damage detection: a comprehensive survey of methodologies and datasets. *Journal of Real-Time Image Processing*, *22*(4), 137.

129. He, J., Gong, L., Xu, C., Wang, P., Zhang, Y., Zheng, O., ... & Sun, Y. (2025). HighRPD: A high-altitude drone dataset of road pavement distress. *Data in Brief*, *59*, 111377.

130. Kaveh, H., & Alhajj, R. (2024). Recent advances in crack detection technologies for structures: a survey of 2022-2023 literature. *Frontiers in Built Environment*, *10*, 1321634.

131. Guerrieri, M., Parla, G., Khanmohamadi, M., & Neduzha, L. (2024). Asphalt pavement damage detection through deep learning technique and cost-effective equipment: A case study in urban roads crossed by tramway lines. *Infrastructures*, *9*(2), 34.

132. Li, Y., Liu, C., Shen, Y., Cao, J., Yu, S., & Du, Y. (2021). RoadID: A dedicated deep convolutional neural network for multipavement distress detection. *Journal of Transportation Engineering, Part B: Pavements*, *147*(4), 04021057.

133. Liu, Z., Wu, W., Gu, X., & Cui, B. (2024). PaveDistress: A comprehensive dataset of pavement distresses detection. *Data in Brief*, *57*, 111111.

134. Zhou, K., Li, L., Zhou, W., Wang, Y., Feng, H., & Li, H. (2025). LaneTCA: Enhancing Video Lane Detection with Temporal Context Aggregation. *IEEE Transactions on Circuits and Systems for Video Technology*.

135. Chen, X., Yongchareon, S., & Knoche, M. (2023). A review on computer vision and machine learning techniques for automated road surface defect and distress detection. *Journal of Smart Cities and Society*, *1*(4), 259-275.

136. Huang, Y., Berrio, J. S., Shan, M., & Worrall, S. (2025). What Demands Attention in Urban Street Scenes? From Scene Understanding towards Road Safety: A Survey of Vision-driven Datasets and Studies. *arXiv preprint arXiv:2507.06513*.

137. Hoier, C., & Ahmed, K. M. (2025). Structural Damage Detection Using AI Super Resolution and Visual Language Model. *arXiv preprint arXiv:2508.17130*.

138. Xiao, X., Zhang, Y., Wang, J., Zhao, L., Wei, Y., Li, H., ... & Xu, H. RoadBench: A Vision-Language Foundation Model and Benchmark for Road Damage Understanding. arXiv 2025. *arXiv preprint arXiv:2507.17353*.

139. Yang, J., Tian, R., Zhou, Z., Tan, X., & He, P. (2025). Flexi-YOLO: A lightweight method for road crack detection in complex environments. *PLoS One*, *20*(6), e0325993.

140. Cha, J., Lee, S., & Kim, H. K. (2025). Deep learning-based detection and assessment of road damage caused by disaster with satellite imagery. *Applied Sciences*, *15*(14), 7669.

141. Seo, H., Shi, Y., & Fu, L. (2024). Automatic damage detection of pavement through DarkNet analysis of digital, infrared, and multi-spectral dynamic imaging images. *Sensors*, *24*(2), 464.

142. Demirel, Z., Nasraldeen, S. T., Pehlivan, Ö., Shoman, S., Albdairi, M., & Almusawi, A. (2025). Comparative Evaluation of YOLO and Gemini AI Models for Road Damage Detection and Mapping. *Future Transportation*, *5*(3), 91.

143. Mulyanto, A., Sari, R. F., Muis, A., & Harwahyu, R. (2025). Vision-Based automated pavement distress inspection: A review. *IEEE Access*.

144. Cano-Ortiz, S., Sainz-Ortiz, E., Lloret Iglesias, L., Martínez Ruiz del Árbol, P., & Castro-Fresno, D. (2024). Leveraging a deep learning generative model to enhance recognition of minor asphalt defects. *Scientific Reports*, *14*(1), 28904.

145. Pham, V., Ngoc, L. D. T., & Bui, D.-L. (2024). Optimizing YOLO architectures for optimal road damage detection and classification: A comparative study from YOLOv7 to YOLOv10. *arXiv preprint arXiv:2410.08409*.

146. Xing, H., & Yang, F. (2026). 3D Road Defect Mapping via Differentiable Neural Rendering and Multi-Frame Semantic Fusion in Bird's-Eye-View Space. *Journal of Imaging*, *12*(2), 83.

147. Zhang, Y., Lu, Y., Huo, Z., Li, J., Sun, Y., & Huang, H. (2024). USSC-YOLO: Enhanced multi-scale road crack object detection algorithm for UAV image. *Sensors*, *24*(17), 5586.

148. Ni, M., Chen, L., Shi, P., & Ren, R. (2025). RepCrack: An efficient pavement crack segmentation method based on structural re-parameterization. *Engineering Applications of Artificial Intelligence*, *141*, 109791.

149. Yilmaz, M., Yalcin, E., Demir, F., Ozdemir, A. M., Atar, M., Gunes, A., & Cambay, E. (2025). Automatic segmentation of asphalt cracks on highways

after large-scale and severe earthquakes using deep learning-based approaches. *IEEE Access*, *13*, 22820-22830.

150. Zhang, Q., Huang, S., Wang, H., Ji, Z., Zheng, S., & Liu, Y. (2025). Segmentation detection method in tree-shaded environment for road cracks collected by inspection vehicle on WFU-Unet. *Scientific Reports*, *15*(1), 11760

151. Fei, Y., Wang, K. C., Zhang, A., Chen, C., Li, J. Q., Liu, Y., ... & Li, B. (2019). Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V. *IEEE Transactions on Intelligent Transportation Systems*, *21*(1), 273-284.

152. Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T., & Sekimoto, Y. (2021). Deep learning-based road damage detection and classification for multiple countries. *Automation in Construction*, *132*, 103935.

153. Liu, Z., Wang, M., Wang, F., & Ji, X. (2021). A residual attention and local context-aware network for road extraction from high-resolution remote sensing imagery. *Remote Sensing*, *13*(24), 4958.

154. Zhang, Y., & Liu, C. (2025). Vision-enhanced multi-modal learning framework for non-destructive pavement damage detection. *Automation in Construction*, *177*, 106389.