ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ

УДК 004.852 Научная статья
DOI: 10.35330/1991-6639-2025-27-6-89-108
EDN: HREYDL

Применение мультимодальных нейросетевых методов для определения качества дорожного полотна

М. Г. Городничев✉, К. А. Полянцева, И. Д. Разумовский

Московский технический университет связи и информатики
111024, Россия, Москва, ул. Авиамоторная, 8А

Аннотация. В статье рассматривается задача автоматического определения дефектов дорожного покрытия с использованием мультимодальных нейросетевых методов.

Цель исследования. Разработка и экспериментальная оценка мультимодального нейросетевого метода автоматического определения дефектов дорожного покрытия с использованием совмещенного анализа визуальных и трехмерных данных.

Методы исследования. Для детекции областей повреждений применялась модель Faster R-CNN, для классификации визуальных фрагментов – Swin Transformer Small, а для анализа геометрии поверхности по данным лидара – модель PointNet. Предсказания от каждой модальности объединялись методом взвешенного суммирования (веса 0.1, 0.6 и 0.4 соответственно). Обучение и тестирование проводились на мультимодальном наборе данных RSRD, включающем RGB-изображения и облака точек, полученные в различных дорожных и погодных условиях.

Результаты. Экспериментальные исследования показали, что мультимодальный подход обеспечивает прирост точности классификации до 95.57 %, а также значительное улучшение метрик детекции дефектов. Для класса «выбоины» полнота увеличилась на 27 %, а F1-score – на 20 % по сравнению с использованием отдельных моделей.

Выводы. Разработанная архитектура демонстрирует высокую устойчивость и точность в задачах анализа дорожного полотна. Полученные результаты подтверждают эффективность интеграции визуальных и пространственных данных и целесообразность применения мультимодальных методов для построения интеллектуальных систем мониторинга дорожной инфраструктуры.

Ключевые слова: машинное обучение, нейронные сети, качество дорожного покрытия, детекция дефектов, компьютерное зрение, лидар, облака точек, сверточные нейронные сети, трансформеры, интеллектуальные транспортные системы

Поступила 13.10.2025, одобрена после рецензирования 11.11.2025, принята к публикации 14.11.2025

Для цитирования. Городничев М. Г., Полянцева К. А., Разумовский И. Д. Применение мультимодальных нейросетевых методов для определения качества дорожного полотна // Известия Кабардино-Балкарского научного центра РАН. 2025. Т. 27. № 6. С. 89–108. DOI: 10.35330/1991-6639-2025-27-6-89-108

MSC: 68T07 Original article

Use of multimodal neural network techniques to assess quality of roadways

M.G. Gorodnichev✉, K.A. Polyantseva, I.D. Razumovsky

Moscow Technical University of Communications and Informatics
8A, Aviamotornaya street, Moscow, 111024, Russia

Abstract. The article discusses the problem of automatic detection of pavement defects using multimodal neural network methods.

Aim. To develop and experimentally evaluate a multimodal neural network method for automatically detecting pavement defects using combined analysis of visual and three-dimensional data.

Methods. The Faster R-CNN model is used for detecting damage areas, the Swin Transformer Small model for classifying visual fragments, and the PointNet model for analyzing surface geometry based on lidar data. The predictions from each modality are combined by weighted summation (weights 0.1, 0.6, and 0.4, respectively). The training and testing are conducted on the RSRD multimodal dataset, which includes RGB images and point clouds obtained in various road and weather conditions.

Results. Experimental studies have shown that the multimodal approach provides an increase in classification accuracy of up to 95.57%, as well as a significant improvement in defect detection metrics. For the pothole class, completeness increased by 27% and F1-score by 20% compared to using individual models.

Conclusions. The developed architecture demonstrates high stability and accuracy in the tasks of analyzing the roadway. The results obtained confirm the effectiveness of the integration of visual and spatial data and the expediency of using multimodal methods to build intelligent monitoring systems for road infrastructure.

***Keywords**:* machine learning, neural networks, pavement quality, defect detection, computer vision, lidar, point clouds, convolutional neural networks, transformers, intelligent transport systems

Submitted 13.10.2025, approved after reviewing 11.11.2025, accepted for publication 14.11.2025

For citation. Gorodnichev M.G., Polyantseva K.A., Razumovsky I.D. Use of multimodal neural network techniques to assess quality of roadways. *News of the Kabardino-Balkarian Scientific Center of RAS.* 2025. Vol. 27. No. 6. Pp. 89–108. DOI: 10.35330/1991-6639-2025-27-6-89-108

Введение

Своевременное выявление и классификация дефектов дорожного покрытия, таких как трещины и выбоины, являются критически важной задачей для обеспечения безопасности дорожного движения и эффективного управления транспортной инфраструктурой [1]. Традиционные методы мониторинга часто требуют значительных человеческих ресурсов и не могут обеспечить непрерывный масштабируемый анализ. Современные достижения в области компьютерного зрения и глубокого обучения открывают новые возможности для автоматизации этого процесса.

Однако системы, основанные исключительно на обработке двумерных изображений, сталкиваются с рядом ограничений, включая чувствительность к изменению освещенности, погодным условиям и сложность текстурных особенностей дорожного полотна от реальных дефектов. В этой связи перспективным направлением является использование мультимодальных подходов, которые комбинируют данные из различных источников.

Определение качества дорожного покрытия с помощью мультимодальных нейросетевых методов начинается с выбора правильных сенсорных технологий и организации процесса сбора данных. Эффективность всей последующей системы напрямую зависит от качества, разнообразия и репрезентативности исходных данных. Современные исследования демонстрируют широкий спектр доступных датчиков, каждый из которых обладает уникальными преимуществами и недостатками, что делает их применение контекстуально зависимым.

Основным источником данных является визуальная информация, получаемая с различных типов камер [2]. Камеры высокого разрешения обеспечивают высокую точность детекции дефектов, достигая 98,95 % в идеальных условиях, однако они уязвимы к погодным условиям, таким как дождь или туман, и плохому освещению. Для решения этой проблемы исследователи используют различные подходы. Например, в одном из проектов [3] применялись тепловизионные камеры, которые показали точность 97 % при классификации дефектов, что делает их перспективными для использования в условиях низкой видимости.

Другой подход заключается в использовании стереокамер [4, 5], которые позволяют получать плотные 3D-облака точек с точностью реконструкции более 3 мм, что полезно для оценки объема выбоин [6].

Второй важной модальностью являются данные, получаемые с инерциальных измерительных блоков (IMU), содержащих акселерометры и гироскопы. Эти сенсоры регистрируют вибрацию, создаваемую проездом над дефектами дорожного покрытия [7, 8]. Преимущество этого метода заключается в его независимости от условий освещения, что позволяет проводить обследования в любое время суток. Однако вибрационные методы чувствительны к шуму от самого транспортного средства и могут обнаруживать дефекты только при прямом проезде по ним [9]. Для повышения точности данных часто применяются алгоритмы предварительной обработки, такие как фильтрация Баттерворта для подавления помех [10] и использование нескольких датчиков для сглаживания сигнала [11]. В качестве примера можно привести систему RoadSense, которая использует датчик MPU-6050 и достигает точности 99,07 % при обнаружении выбоин [12].

Третьей все более популярной модальностью становится трехмерное сканирование с помощью LiDAR и радаров. LiDAR обеспечивает высокоточную 3D-реконструкцию дорожного полотна, позволяя точно измерять глубину и площадь выбоин [13, 14]. Однако данные LiDAR и облака точек требуют значительных вычислительных ресурсов для обработки [15]. Радары, особенно миллиметроволные (mmWave), представляют собой перспективную альтернативу. Они работают в диапазоне 76–81 ГГц, что делает их устойчивыми к дождю, туману и темноте, превосходя по этому параметру камеры и LiDAR. Прототип mmWave-радара, разработанный в Университете Аризоны, успешно сегментирует облако точек на объекты, такие как пешеходы и автомобили, что указывает на потенциал его применения для анализа дорожной обстановки и состояния покрытия. В одном из исследований было показано, что даже мобильный лазерный сканер (LiDAR) с низкой плотностью точек может эффективно обнаруживать выбоины при оптимальной плотности.

В некоторых исследованиях экспериментируют с менее очевидными модальностями, такими как звук [17]. В работе по оценке состояния гравийных дорог был предложен метод слияния визуальных и аудиальных данных, где аудиозаписи преобразовывались в спектрограммы. Этот подход показал высокую производительность, особенно при позднем слиянии признаков с использованием логических операций OR и AND, достигнув точности 97 %. Это открывает новые горизонты для создания робастных систем, способных работать в сложных условиях.

Выбор нейросетевой архитектуры является центральным элементом в разработке систем для определения качества дорожного покрытия. Исследовательское сообщество активно экспериментирует с различными моделями, адаптируя существующие архитектуры для задач компьютерного зрения и предлагая собственные решения. Основой большинства современных систем служат сверточные нейронные сети (CNN), которые демонстрируют высокую эффективность в задачах классификации, обнаружения объектов и семантической сегментации [18].

Для задачи детекции объектов, такой как поиск трещин и выбоин, наиболее популярными являются архитектуры семейства YOLO (You Only Look Once). Модификации YOLOv3/v5/v7/v8/v10 регулярно появляются в научных публикациях благодаря их высокому соотношению скорости и точности [19, 20]. Например, модель SCB-AF-Detector на основе YOLOv5s достигла 90,8 % точности на датасете IRRDD, а YOLOv7 была применена для обнаружения дефектов с точностью 94,5 % [21]. Для повышения эффективности этих моделей авторы разрабатывают специализированные модификации бэкбона (например, SCB-Darknet53 с интегрированными трансформерами) и головного блока (AFPN), что позволяет улучшить извлечение признаков и повысить общую производительность. Другие популярные архитектуры включают Faster R-CNN, RetinaNet и SSD.

Для более точной локализации дефектов, особенно при необходимости не просто обвести их рамкой, а выделить все пиксели, относящиеся к повреждению, используются архитектуры семантической сегментации. Наиболее известными здесь являются U-Net и DeepLabv3+ [22]. Модель U-Net, в частности, показала точность до 92,23 % в задаче сегментации трещин. В одном из исследований лучшие результаты на датасете ISTD-PDS7 показала модель SegFormer, достигнув F1-меры 94,23 %. Эти модели позволяют не только находить дефекты, но и точно измерять их площадь и форму.

Помимо CNN, в последнее время все большее внимание уделяется трансформерам и графовым нейросетям, хотя их применение в данной области еще не так широко распространено, как в других областях ИИ. Трансформеры демонстрируют высокую производительность в задачах, связанных с длинными зависимостями и глобальным контекстом, что может быть полезно для анализа больших участков дороги. Например, модель Pavement-DETR на основе RT-DETR для детекции шести типов дефектов показала mAP@0.5 = 87,1 % на датасете UAV-PDD2023, превзойдя базовую модель на 7,7% [23].

Архитектуры, сочетающие CNN и трансформеры, также исследуются для детекции выбоин на дорогах.

Наборы данных

Качество и масштабность наборов данных являются основополагающими факторами, определяющими развитие области мониторинга качества дорожного покрытия. Хорошо размеченный, разнообразный и большой датасет позволяет обучать более точные и робастные модели. Однако текущее состояние исследовательского поля характеризуется наличием множества небольших, часто закрытых датасетов, что затрудняет воспроизводимость результатов и сравнение методов.

Существует множество открытых наборов данных, каждый из которых имеет свои особенности. Самые крупные и популярные – это датасеты, собранные в ходе международных соревнований. Например, RDD2022 содержит 47420 изображений из шести стран мира, размеченных по четырем типам дефектов (продольные и поперечные трещины, крокодиловое растрескивание, выбоины), что делает его ценным для обучения моделей, применяющихся для разных типов дорог [24]. Датасет TD-RD содержит 7088 высококачественных изображений с 12882 размеченными экземплярами трещин и выбоин [25]. Другой крупный датасет, EGY_PDD, включает 14612 2D-изображений и 4323 сцены с 3D-данными (глубина и облако точек), собранными в Египте, что позволяет работать с мультимодальными данными [26]. Существуют также специализированные датасеты, например, ISTD-PDS7, содержащий 18527 изображений для задачи сегментации с 7 типами дефектов и большим количеством негативных примеров (теней, пятен), что повышает робастность моделей.

Несмотря на наличие подобных датасетов, исследовательская практика показывает, что большинство моделей обучается на приватных данных, собранных исследователями для своих целей. Это создает проблему, поскольку результаты, полученные на одном ограниченном датасете, могут не быть репрезентативными для других условий. Еще одной проблемой является недостаток стандартизированных 3D-датасетов, что ограничивает развитие методов, использующих LiDAR и 3D-сканирование.

Однако существуют и другие датасеты. Например, RoadBench, самый крупный в своей области датасет, содержащий 100 000 пар «изображение-текст» высокого разрешения [27]. Его уникальность заключается в том, что каждое изображение сопровождается подробным текстовым описанием, сгенерированным GPT-4o. Это позволяет развивать визуально-языковые модели, которые могут лучше понимать контекст и детали дефектов.

RoadBench включает 10 типов повреждений, включая смешанные и неопределенные случаи, что делает его более реалистичным и сложным для анализа.

Датасет TD-RD специально создан для детекции повреждений и содержит 715 088 изображений с 12882 разметками трещин, выбоин и заплат. Кроме того, были разработаны датасеты, сфокусированные на конкретных технологиях. Например, датасет, который включает данные от десяти типов датчиков (акселерометр, гироскоп, GPS и др.) и разделен на две категории: данные об аномалиях и данные о стиле вождения, что позволяет проводить более глубокий анализ 30 влияний внешних факторов [28].

Таким образом, развитие методов для определения качества дорог напрямую связано с появлением и улучшением наборов данных. Будущие исследования будут все больше зависеть от доступности крупных, разнообразных и хорошо размеченных мультимодальных датасетов, которые позволят создавать действительно универсальные и надежные системы мониторинга.

В нашем исследовании для обучения и тестирования моделей использовался датасет RSRD – Road Surface Dataset [29]. Используемый набор данных содержит 8000 файлов четырех типов (модальностей):

1. Изображения высокого разрешения, полученные в результате видеосъемки дорожного покрытия в различных погодных условиях.

2. Карты глубины, полученные на основе слияния нескольких кадров LiDAR-данных, представляют собой двумерные изображения, в которых каждый пиксель соответствует расстоянию от датчика до объекта в сцене. Такие карты строятся на основе данных, полученных с помощью LiDAR-сенсора – активного дальномера, работающего на основе лазерного сканирования.

3. Карты диспарантности – изображения, в которых каждый пиксель содержит информацию о разности положения одной и той же точки сцены на двух изображениях, снятых с разных ракурсов.

4. Облака точек, полученные с использованием многокадрового объединения данных LiDAR с компенсацией движения.

5. Видеосъемка проводилась на асфальтовых и бетонных покрытиях в городских и сельских районах. Скорость автомобиля была ограничена 40 км/ч для повышения качества данных.

Разметка изображений на классы «трещины» и «выбоины» производилась с использованием платформы Roboflow. Аннотация разметки была в формате coco.json.

Архитектура предлагаемого решения

Для задачи выявления дефектов дорожного полотна предлагается мультимодальная архитектура, объединяющая визуальные и лидарные данные. Схема работы представлена на рис. 1.

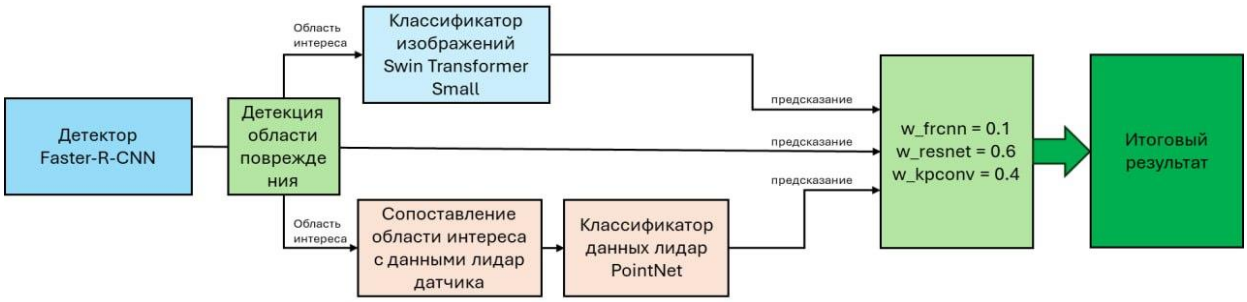

***Рис. 1**. Схема архитектуры мультимодального метода определения повреждений дорожного покрытия*

***Fig. 1**. Schematic diagram of the architecture of the multimodal method for determining road surface damage*

В качестве основного блока используется детектор Faster R-CNN, который отвечает за выделение областей интереса – потенциальных участков повреждений дорожного покрытия. Эти области далее подвергаются двунаправленной обработке:

1. **Визуальный канал.**

Области интереса подаются на вход классификатору изображений Swin Transformer Small. Этот модуль анализирует текстуру, цвет и форму дорожного полотна, что позволяет зафиксировать такие признаки повреждений, как трещины или выбоины.

2. **Лидарный канал.**

Параллельно с визуальной обработкой происходит сопоставление области интереса с данными облака точек, полученными от лидар-датчика. Для анализа трехмерной геометрии используется PointNet, который классифицирует объекты на основе пространственной структуры поверхности дороги. Такой подход позволяет выявлять дефекты, слабо выраженные на изображении, но отчетливо различимые в геометрии.

Оба канала формируют независимые предсказания, которые затем объединяются с учетом весовых коэффициентов:

- **w_frcnn = 0,1** – вклад базового детектора,
- **w_resnet = 0,6** – основной вес, присвоенный визуальному классификатору,
- **w_kpconv = 0,4** – вес предсказаний по данным лидара.

Итоговое решение формируется путем взвешенного суммирования результатов, что обеспечивает баланс между различными модальностями и повышает надежность классификации. Такой мультимодальный подход позволяет не только повысить точность детекции дефектов, но и уменьшить вероятность ложных срабатываний за счет перекрестной проверки признаков в разных источниках данных.

Детекция повреждений дорожного покрытия на изображениях, полученных в результате видеосъемки

Детекция и выделение повреждений дорожного покрытия на изображении является одним из основных этапов при анализе результатов видеосъемки. Для детекции повреждений дорожного покрытия на изображениях в исследовании были использованы архитектуры Faster R-CNN, YOLO8 Nano и YOLO8 Small.

В качестве метрик, которые влияли на выбор наилучших весов для модели при ее обучении, использовали функцию потерь и полноту (recall). Выбор полноты в качестве метрики был обусловлен важностью детекции положительных объектов на изображении. Для сравнения разных моделей использовали метрики точность (precision), полнота (recall), f1-score и точность предсказания границ (Intersection over Union, IoU).

Модель Faster R-CNN представляет собой детектор с архитектурой ResNet-50 в качестве backbone и функционалом Feature Pyramid Network (FPN). Модель Faster R-CNN является двухстадийным детектором, который на первом этапе извлекает регионы интереса, а затем определяет их класс и уточняет расположение [30].

Архитектура Faster R-CNN была инициализирована с использованием предобученных весов на COCO и дообучалась на задаче с пользовательскими классами дефектов дорожного полотна, а также с учетом фонового класса. В оригинальной модели была заменена головная часть классификатора box_predictor на новую, соответствующую числу целевых классов.

В результате обучения модели Faster R-CNN на валидационной выборке значение функции потерь составило 0,2432, а полнота (recall) составляла 0,8526.

Следует отметить, что метрика AP@[0,5:0,95], представляющая собой среднее значение точности по различным уровням перекрытия (IoU), является достаточно строгим критерием. Однако в контексте рассматриваемой прикладной задачи автоматического мониторинга дорожного покрытия эта метрика не является определяющей. Целью же предлагаемой системы является своевременное обнаружение повреждений дорожного полотна, в том числе потенциальных, с возможностью последующего анализа их характера и степени опасности. В этой связи особую важность приобретает полнота (recall) модели – способность фиксировать максимальное число случаев наличия дефекта. Напротив, ложноположительные срабатывания (false positives) в рамках такой системы представляют меньшую угрозу: они могут быть отфильтрованы на последующих этапах обработки, в том числе с использованием дополнительных модальностей, таких как облака точек, карты глубины и диспарантности. Таким образом, допустимость ложных срабатываний компенсируется отсутствием пропусков, что критически важно для обеспечения безопасности дорожного движения.

Модель YOLOv8 относится к одноэтапным детекторам объектов и реализует подход «end-to-end» – от входного изображения сразу к предсказаниям классов и координат объектов. В данной работе использовались облегченные модификации YOLOv8: YOLOv8 Nano и YOLOv8 Small, разработанные с целью повышения скорости обработки и уменьшения требований к вычислительным ресурсам. Архитектура YOLOv8 основана на полностью сверточной нейросети без явно выделенного backbone и включает в себя улучшенные механизмы извлечения признаков, а также адаптивную иерархическую структуру обнаружения объектов. YOLOv8 использует современный механизм декодирования предсказаний на основе anchor-free подхода, что позволяет повысить точность и стабильность при работе с различными масштабами объектов. Благодаря высокой скорости и компактности модели YOLOv8 Nano и Small хорошо подходят для задач реального времени и применения на встраиваемых устройствах [31].

Сравнение основных метрик моделей, используемых для детекции повреждений дорожного покрытия, представлено в таблице 1.

Таблица 1. Сравнение основных метрик моделей

Table 1. Comparison of the main metrics for the models

Наименование	**precision**		**recall**		**F1-score**		**mAP50-95**
	трещины	**выбоины**	**трещины**	**выбоины**	**трещины**	**выбоины**	
Faster-R-CNN	**0,93**	0,95	0,82	**0,61**	0,87	**0,74**	**0,166**
YOLOv8 Nano	0,85	**1**	**1**	0.18	0,92	0,31	0,104
YOLOv8 Small	0,89	**1**	**1**	0,42	**0,94**	0,60	0,143

Из рассмотренных моделей Faster R-CNN демонстрирует наиболее сбалансированные значения основных метрик, особенно по полноте и F1-score. При этом для моделей семейства YOLO характерны низкие значения метрики recall для класса «выбоины». Наибольшее значение метрики IoU было также при анализе тестовой выборки моделью Faster R-CNN. При этом данное значение все равно было достаточно низким, однако это не яв-

ляется критичным для задачи мониторинга, где ключевым показателем является recall – важно не пропустить дефект. Ошибки в виде ложного выделения исправных участков (низкий precision) менее значимы, поскольку они не приводят к реальным потерям, а увеличивают объем последующего анализа.

Классификация повреждений дорожного покрытия на изображениях, полученных при видеосъемке

Для повышения точности определения типа повреждения дорожного покрытия (выбоины или трещины) после их детекции на изображении применяли бинарную классификацию. При решении данной задачи сравнивали модели глубокого обучения на основе сверточных нейронных сетей и трансформеров. Среди сверточных нейронных сетей были выбраны такие модели, как ResNet-18, Res-Net-34, ResNet-50 и EfficientNet-B0 – B7, а среди трансформеров – SwinTransformer tiny, SwinTransformer base, SwinTransformer small.

В процессе обучения для выбора наилучшей модели оценивали функцию потерь и метрику Macro F1 на валидационной выборке. Данная метрика позволяет оценить эффективность работы классификатора на несбалансированных выборках. Для выбора наилучшей модели сравнивались такие метрики, как precision (точность для положительных предсказаний), recall (полнота), f1-score, macro average (среднее значение для несбалансированных классов), weighted average (взвешенное среднее значение), accuracy (точность).

Сверточная нейронная сеть ResNet, предобученная на датасете ImageNet, представляет собой глубокую остаточную сеть, которая использует механизм остаточных связей (residual connections). Данный механизм позволяет эффективно обучать глубокие нейронные сети, предотвращая затухание градиента и способствуя более стабильной и быстрой сходимости. Данная архитектура была выбрана из-за ее высокой устойчивости к переобучению, способности эффективно извлекать иерархические признаки с различных уровней абстракции, а также доказанной эффективности в задачах классификации изображений. Благодаря использованию остаточных блоков ResNet демонстрирует высокую обобщающую способность даже при ограниченном объеме обучающих данных, что делает ее особенно подходящей для задач, где важно точно классифицировать мелкие и разнообразные дефекты на дорожных покрытиях. Кроме того, наличие предварительно обученных весов на ImageNet позволяет ускорить обучение и повысить начальное качество модели за счет переноса знаний из широкой области компьютерного зрения [32].

На рисунке 2 показаны графики обучения моделей ResNet-18, ResNet-34, ResNet-50.

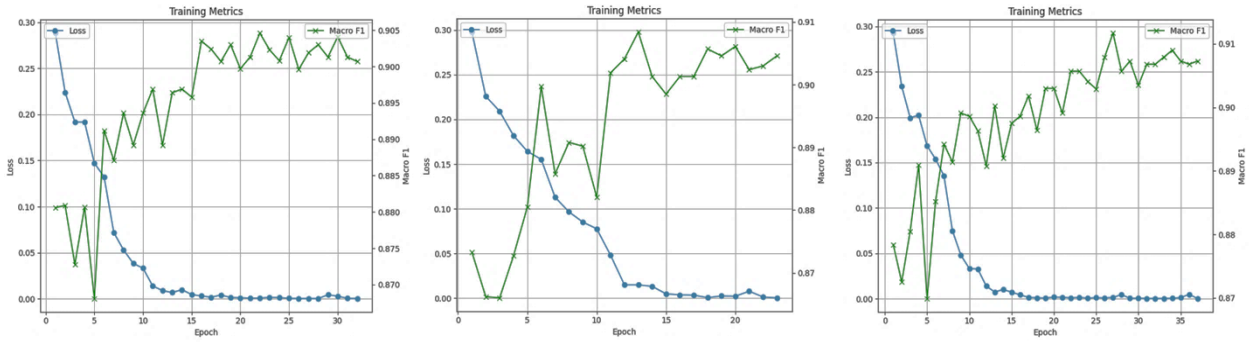

***Рис. 2**. Графики обучения на тренировочной выборке ResNet-18, ResNet-34, ResNet-50*

***Fig. 2**. Training graphs on the training set ResNet-18, ResNet-34, ResNet-50*

Среди исследованных моделей семейства ResNet (ResNet-18, ResNet-34, ResNet-50) сравнимые результаты были показаны для моделей ResNet-18 и ResNet-34. Однако наибольшее среднее значение F1-меры (0,8752) по двум классам и общее значение точности (0,8756) показала модель ResNet-18. Общее значение точности классификации модели ResNet-50 на тестовой выборке было на 2 % ниже, чем у двух других архитектур.

EfficientNet – это семейство сверточных нейронных сетей, предложенное с целью достижения оптимального баланса между точностью и вычислительной эффективностью. Архитектура EfficientNet была разработана с использованием метода автоматического масштабирования модели (compound scaling), который одновременно увеличивает глубину, ширину и разрешение входных изображений в заданных пропорциях.

Базовая модель EfficientNet-B0 была разработана с использованием метода автоматизированного подхода, позволяющего оптимизировать структуру нейронной сети для достижения баланса между точностью и вычислительной эффективностью. Эта модель послужила фундаментом для создания более крупных вариантов EfficientNet (B1–B7), которые получили путем масштабирования базовой архитектуры с сохранением ее сбалансированности. Такой подход обеспечил существенное улучшение качества предсказаний без экспоненциального роста вычислительных затрат, характерного для традиционных методов увеличения глубины или ширины сети. Архитектура EfficientNet построена на основе блоков MBConv (Mobile Inverted Bottleneck Convolution), заимствованных из MobileNetV2, и использует эффективные разделяемые сверточные операции (depthwise separable convolutions). Кроме того, каждая операция дополнена механизмом Squeeze-and-Excitation (SE), который адаптивно перенастраивает весовые коэффициенты каналов, что способствует улучшению выделения информативных признаков [33].

Ключевые особенности EfficientNet:

1. Эффективное использование вычислительных ресурсов за счет сбалансированного масштабирования.

2. Повышенная точность для задач классификации изображений.

3. Компактность и быстродействие, особенно в младших версиях (B0–B2), что делает их подходящими для мобильных и встраиваемых систем.

В исследовании были использованы модели EfficientNet-B0 – EfficientNet-B7, которые различаются между собой по глубине, ширине и разрешению входного изображения. С возрастанием поколения EfficientNet также увеличивается количество параметров. Например, в EfficientNet-B0 – 5,3 млн параметров, тогда как в EfficientNet-B7 – 66 млн.

На рисунке 3 показаны графики обучения моделей EfficientNet-B0 – EfficientNet-B7.

В семействе EfficientNet (варианты от B0 до B7) максимальные показатели основных метрик продемонстрировала модель EfficientNet-B6. Среднее значение F1-мера для обоих классов составило 0,9003, а значение точности (accuracy) – 0,9005, что делает ее наилучшей архитектурой среди исследуемых из данного семейства. При этом модель лишь незначительно превосходит по метрикам версии B4 (0,8980 и 0,8982 соответственно) и B5 (0,8956 и 0,8959 соответственно), однако все же обеспечивает наивысшую точность классификации.

Нейронная сеть на основе трансформерной архитектуры – Swin Transformer, предобученная на датасете ImageNet, представляет собой иерархическую архитектуру визуального трансформера, в которой используется оконный механизм привлечения внимания (window-based self-attention), который перемещается по изображению для выделения объектов. В отличие от классического глобального механизма внутреннего внимания оконный подход позволяет существенно сократить вычислительную сложность, что делает модель более эффективной при работе с изображениями высокого разрешения.

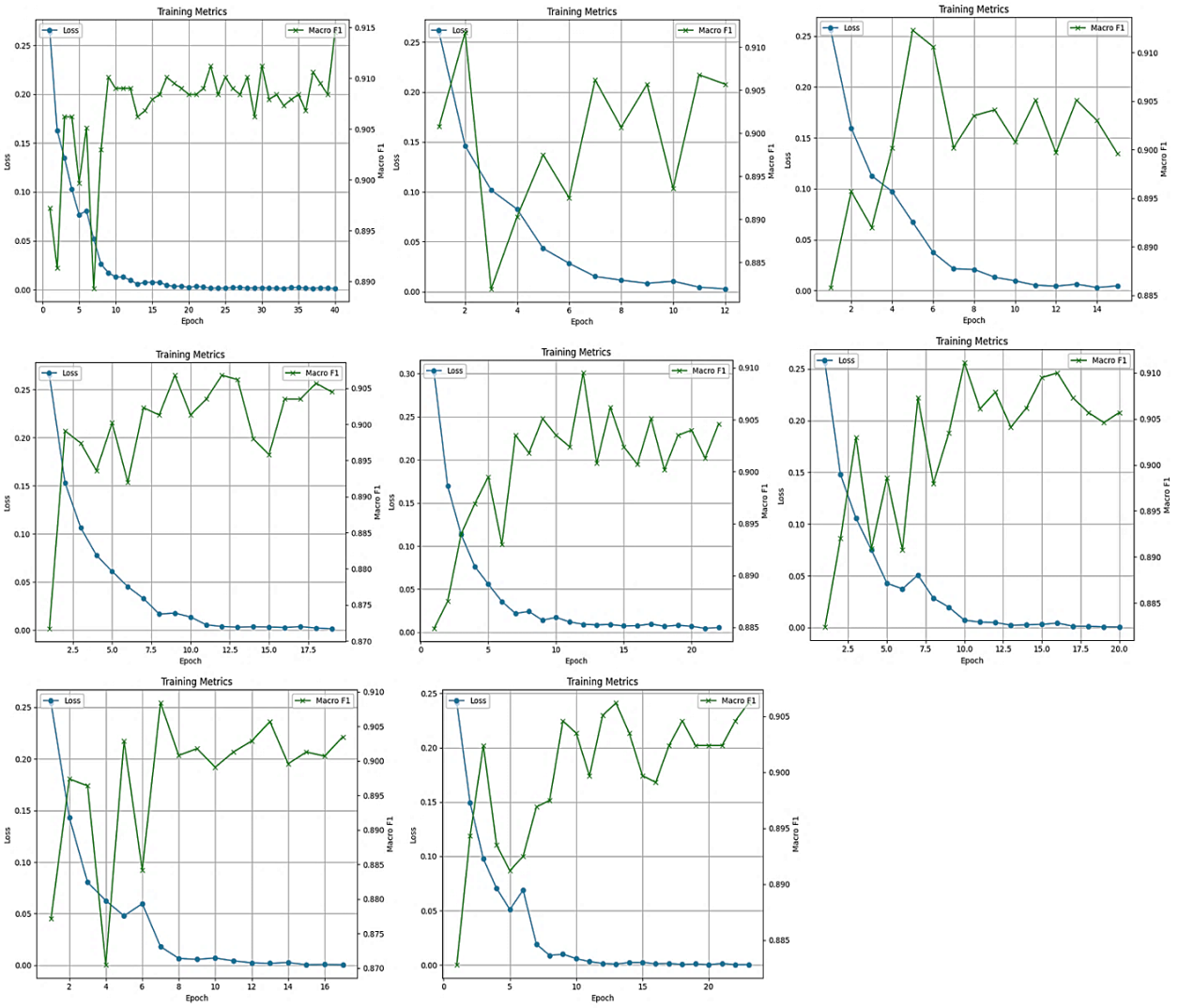

Рис. 3. *Графики обучения на тренировочной выборке EfficientNet-B0 – EfficientNet-B7*

Fig. 3. *Training graphs on the training set EfficientNet-B0 - EfficientNet-B7*

Ключевой особенностью Swin Transformer является использование сдвинутых окон (shifted windows), обеспечивающих перекрытие между соседними окнами. Это позволяет эффективно объединять локальную и глобальную информацию, что особенно важно для задач, требующих пространственного контекста, таких как классификация, сегментация и детекция объектов.

Архитектура построена иерархически: на каждом уровне обработки пространственное разрешение уменьшается, а размерность признаков увеличивается аналогично принципам, применяемым в сверточных нейронных сетях. Это позволяет Swin Transformer сочетать преимущества трансформеров и сверточных моделей, обеспечивая высокую точность при умеренной вычислительной нагрузке [34].

Модели Swin различаются по размеру (Tiny, Small, Base, Large) и числу параметров, что позволяет подобрать подходящий вариант в зависимости от доступных ресурсов и требований к качеству распознавания.

На рисунке 4 показаны графики обучения моделей Swin Tiny, Swin Small, Swin Base.

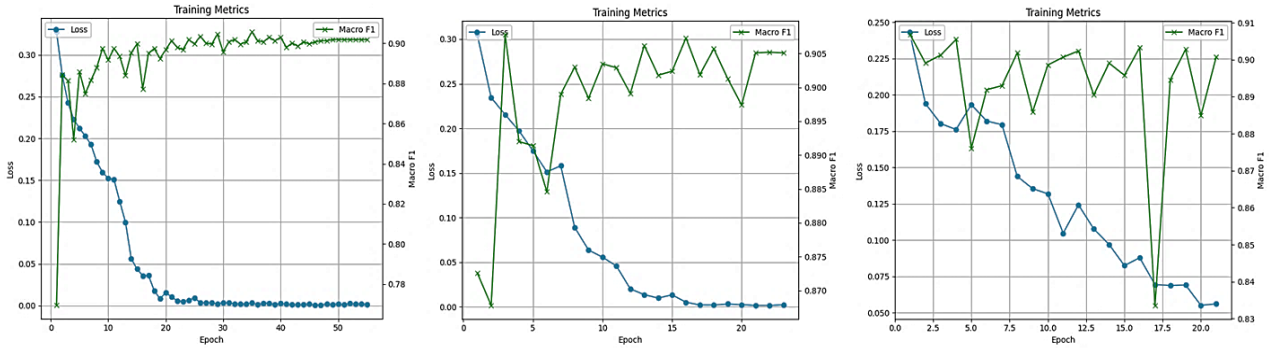

Рис. 4. *Графики обучения на тренировочной выборке Swin Tiny, Swin Small, Swin Base*

Fig. 4. *Training graphs on the training set Swin Tiny, Swin Small, Swin Base*

Среди исследованных моделей на базе Swin Transformer (Tiny, Small, Base) лучшую производительность продемонстрировала модель Swin Small. Так, среднее значение F1-меры составило 0,9004, а accuracy – 0,9005. Указанная модель существенно превосходит более легкие (Swin Tiny) и тяжелые варианты (Swin Base), что делает ее предпочтительной для рассматриваемой задачи (табл. 2).

Таблица 2. Результаты итогового сравнения моделей для бинарной классификации

Table 2. Results of the final comparison of models for binary classification

Наименование	**precision**		**recall**		**F1-score**		**accuracy**
	трещины	**выбоины**	**трещины**	**выбоины**	**трещины**	**выбоины**	
ResNet-18	0,8388	0,9213	0,9299	0,8213	0,8820	0,8684	0,8756
ResNet-34	0,8277	0,9367	0,9457	0,8032	0,8828	0,8648	0,8744
ResNet-50	0,8124	0,9086	0,9208	0,7873	0,8632	0,8436	0,8541
EfficientNet-B0	0,8686	0,9223	0,9276	0,8597	0,8972	0,8899	0,8937
EfficientNet-B1	0,8214	0,9263	0,9367	0,7964	0,8753	0,8564	0,8665
EfficientNet-B2	0,8182	**0,9482**	**0,9570**	0,7873	0,8822	0,8603	0,8722
EfficientNet-B3	0,8090	0,9272	0,9389	0,7783	0,8691	0,8462	0,8586
EfficientNet-B4	0,8636	0,9400	0,9457	0,8507	0,9028	0,8931	0,8982
EfficientNet-B5	0,8739	0,9207	0,9253	0,8665	0,8989	0,8928	0,8959
EfficientNet-B6	0,8688	0,9381	0,9434	0,8575	**0,9046**	0,8960	**0,9005**
EfficientNet-B7	0,8416	0,9171	0,9253	0,8258	0,8815	0,8690	0,8756
Swin Tiny	0,8463	0,9022	0,9095	0,8348	0,8768	0,8672	0,8722
Swin Base	0,8351	0,9073	0,9163	0,8190	0,8738	0,8609	0,8676
Swin Small	**0,8899**	0,9116	0,9140	**0,8869**	0,9018	**0,8991**	**0,9005**

По результатам сравнительного анализа моделей наилучшие показатели точности классификации (accuracy) продемонстрировали трансформер Swin Small и сверточная нейронная сеть EfficientNet-B6. Кроме того, модель Swin Small обладала наилучшими

значениями метрик recall и F1-score для класса «выбоины». Стоит отметить, что среди моделей семейства ResNet наилучшие значения точности были у моделей ResNet-18 и ResNet-34.

Классификация повреждений дорожного покрытия с использованием данных, полученных с датчиков лидар

Помимо изображений с видеокамер, для определения наличия и типа повреждения дорожного покрытия можно использовать данные, полученные с датчиков лидар (LiDAR). Данный тип датчиков использует лазерные импульсы для определения расстояния до объекта. Одним из типов данных, получаемых с датчиков лидар, является «облако точек», которое представляет собой набор точек в трехмерном пространстве и отражает геометрические особенности поверхности, карты глубины и диспарантности, которые соответствуют расстоянию от датчика до объекта [35].

Для классификации данных типа «облако точек» необходимо использовать модели, которые будут работать в трехмерном пространстве и учитывать зависимости между точками. В данном исследовании для задачи классификации «облаков точек» были использованы следующие модели: Point Transformer, PointNet и PointNet++.

Архитектура Point Transformer использует модифицированный механизм внимания (attention), адаптированный под трехмерные пространственные данные, и позволяет эффективно учитывать локальные и глобальные зависимости между точками, что критически важно при анализе сложных объектов дорожной инфраструктуры, таких как выбоины, трещины или искусственные неровности.

Входной слой модели представлял собой полносвязную проекцию координат каждой точки из пространства $\mathbf{R}^3$ в пространственное признаковое представление размерности 64. Далее следовали два блока Point Transformer, каждый из которых осуществлял агрегацию признаков с учетом ближайших соседей каждой точки (по евклидовому расстоянию) и учитывал относительные пространственные сдвиги между точками. Attention-механизм в этих блоках позволял каждой точке адаптивно взаимодействовать со своими соседями на основе как геометрической близости, так и содержательных признаков, что существенно усиливало выразительную способность модели.

После извлечения признаков из всех точек производилась глобальная агрегация с использованием адаптивного слоя max pooling, что позволяло получить компактное представление всего облака точек. Далее это представление передавалось в классификационный блок, состоящий из нескольких полносвязных слоев с функцией активации ReLU и выходным softmax-слоем. Он формировал вероятностное распределение по классам, соответствующим типам состояния дорожного покрытия (трещины или выбоины).

Таким образом, предложенная модель сочетает в себе возможности глубоких сверточных структур и гибкость внимания, адаптированного под 3D-точечные облака, что делает ее эффективным инструментом для анализа и классификации дорожных объектов на основе пространственных данных [36].

Архитектура PointNet предназначена для классификации облаков точек в 3D-пространстве. На вход модели подается тензор размерности (B, N, 3), где B – размер батча, N – количество точек в облаке, а 3 – координаты каждой точки (X, Y, Z). Для обработки данных используется последовательность сверточных слоев с ядром размера 1, что эквивалентно применению многослойного персептрона (MLP) к каждой точке отдельно. Сначала точки транспонируются в формат (B, 3, N), после чего проходят через три сверточных блока с

увеличением размерности признаков: с 3 каналов до 64, затем до 128 и, наконец, до 1024. Каждый сверточный слой сопровождается нормализацией батча (BatchNorm) и нелинейной активацией ReLU, что позволяет стабилизировать обучение и повысить выраженность признаков. После обработки всех точек осуществляется глобальный max-pooling по точкам, который агрегирует информацию по всему облаку, создавая единый вектор признаков размерности 1024, представляющий все облако целиком. Далее этот глобальный вектор подается на несколько полносвязных слоев, также снабженных нормализацией, функциями ReLU и регуляризацией Dropout для предотвращения переобучения. На выходе сети формируется вектор с числом элементов, равным количеству классов, для которых проводится классификация [37].

Модель PointNet++ является следующим поколением архитектуры PointNet за счет иерархической многоуровневой обработки точечных данных с учетом локальных геометрических структур. На вход подается тензор координат точек размерности (B, N, 3), где B – размер батча, N – число точек в облаке, а 3 – пространственные координаты каждой точки. В основе модели лежит модуль PointNetSetAbstraction, реализующий этап абстракции множества точек. В этом модуле сначала производится выбор подмножества ключевых точек (центроидов) из исходного облака с помощью алгоритма farthest point sampling (FPS). FPS последовательно выбирает точки, максимально удаленные от уже выбранных, что обеспечивает равномерное покрытие пространства точек. Затем для каждого центроида определяется множество его k ближайших соседей, формируя локальные регионы. Далее координаты соседних точек нормализуются путем вычитания координат соответствующего центроида, что позволяет модели быть инвариантной к локальному сдвигу. Если доступны дополнительные признаки точек (features), они группируются и конкатенируются с нормализованными координатами, формируя расширенные локальные дескрипторы.

Для извлечения признаков из локальных групп используется последовательность сверточных слоев с ядром 1x1, за которыми следуют операции пакетной нормализации (BatchNorm) и нелинейной активации ReLU. Такая свертка применяется к каждому фрагменту отдельно, после чего применяется операция агрегации – max pooling, которая сводит информацию о локальном регионе в единый вектор признаков на каждый центроид.

Архитектура строится иерархически: на первом уровне выбирается 512 центроидов с 32 соседями, затем на втором уровне – 128 центроидов и 32 соседа, где входными каналами второго уровня являются признаки, полученные на первом уровне. На третьем уровне реализована глобальная абстракция без подвыборки, которая агрегирует признаки по всему облаку точек, формируя глобальный вектор размерности 1024. Полученный глобальный вектор подается на несколько последовательно соединенных полносвязных слоев с пакетной нормализацией, функциями активации ReLU и Dropout для борьбы с переобучением. На выходе формируется логит-вектор с размерностью, равной числу классов задачи классификации. Таким образом, архитектура PointNet++ эффективно интегрирует локальные и глобальные геометрические особенности «облаков точек», обеспечивая высокую точность и устойчивость к вариациям пространственного расположения точек.

На рисунке 5 показаны графики обучения моделей Point Transformer, PointNet, PointNet++.

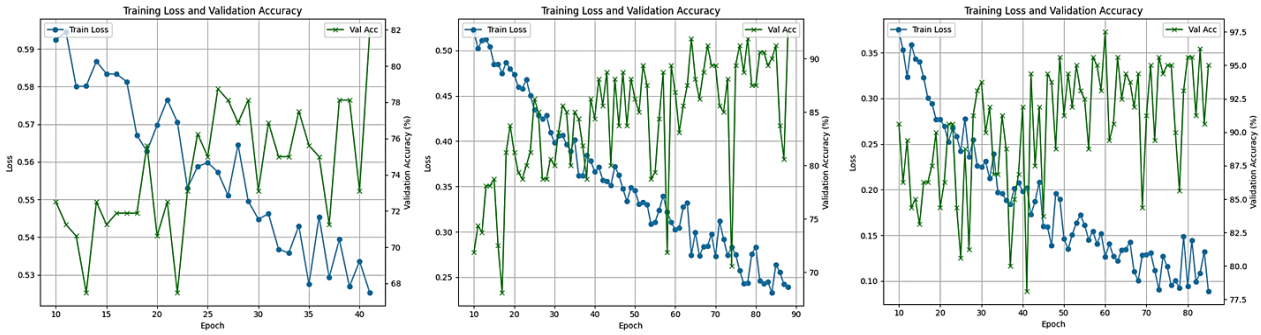

***Рис. 5**. Графики обучения на тренировочной выборке Point Transformer, PointNet, PointNet++*

***Fig. 5**. Training graphs on the training set of Point Transformer, PointNet, PointNet++*

Сравнение значений основных метрик, полученных при классификации повреждений дорожного покрытия с использованием моделей Point Transformer, PointNet и PointNet++, представлено в таблице 3.

Таблица 3. / Table 3.

Наименование	**precision**		**recall**		**F1-score**		**accuracy**
	трещины	**выбоины**	**трещины**	**выбоины**	**трещины**	**выбоины**	
Point Transformer	0,9091	0,3188	0,7006	**0,6667**	0,7914	0,4314	0,6947
PointNet	**0,9290**	0,6286	0,9172	**0,6667**	0,9231	**0,6471**	0,8737
PointNet++	0,8960	**0,8824**	**0,9873**	0,4545	**0,9394**	0,6000	**0,8947**

На основании представленных результатов можно сделать обоснованный выбор в пользу модели PointNet. Несмотря на то, что PointNet++ демонстрирует наивысшие значения F1-меры для класса «трещина», ее отзывчивость (recall) по классу «выбоины» значительно ниже (0,4545), чем у двух других моделей, что говорит о слабой эффективности определения данного типа дефектов. Архитектура Point Transformer показывает высокую точность определения класса «трещины» (0,9091) при крайне низкой точности (0,3188) и F1-мере (0,4314) для класса «выбоина». Тогда как модель PointNet, напротив, демонстрирует наиболее сбалансированные метрики бинарной классификации. Данная модель достигает наибольших значений метрик recall и F1-score для класса «выбоины» по сравнению с другими моделями. При этом значения метрик, отражающих качество классификации трещин, снижаются незначительно относительно Point Transformer и PointNet++. При этом общая точность (accuracy) модели PointNet всего на 2,1 % ниже, чем у PointNet++, и на 17,9 % выше, чем у Point Transformer.

Для задачи классификации карт глубин и диспарантности были обучены модели ResNet-18 и ResNet-34, которые показывают высокую эффективность при работе с изображениями. Такой выбор моделей был обусловлен тем, что данные карт глубин и диспарантности были представлены в формате png.

Так, модель ResNet-18 на тестовой выборке ResNet-18 показала следующие метрики для классов «трещина» и «выбоина»: precision составлял 0,72 и 1, recall – 1 и 0,22, а f1-score –

0,84 и 0,36 соответственно. При этом точность (accuracy) работы модели составила 0,74. Модель ResNet-34 показала значение precison – 0,67, recall – 1, f1-score – 0,8 для класса «трещины». Для класса «выбоина» данная модель показала нулевой результат, а общая точность (accuracy) классификации составила 0,67.

В связи с этими результатами данные модальности не использовались в дальнейшей работе, так как они показали низкий recall для класса «выбоина», что свидетельствует о низкой способности определять данный класс по данным карт глубины и диспарантности.

В конечном итоге для реализации мультимодального метода была выбрана модель PointNet, которая классифицирует данные типа «облака точек», полученные с датчиков лидар.

Классификация повреждений дорожного покрытия с использованием мультимодального подхода

Мультимодальный метод оценки дефектов дорожного покрытия сочетает в себе предсказания трех различных моделей, каждая из которых использует свою модальность: изображение RGB, фрагмент изображения (обрезка детектированного объекта) и облако точек. Целью данного подхода является повышение точности классификации объектов, обнаруженных на изображениях, за счет объединения информации из разных источников.

В процессе тестирования изображения последовательно обрабатываются при помощи модели обнаружения объектов на базе Faster R-CNN. Модель предсказывает координаты ограничивающих рамок, метки классов и значения точности в предсказаниях. Для каждого объекта, детектированного моделью Faster R-CNN с высокой точностью, осуществляется извлечение фрагмента изображения, соответствующего рамке. Полученное изображение передается в предварительно обученную модель Swin Small для классификации фрагментов дефектов дороги. Результатом является распределение вероятностей по классам для определяемых дефектов (трещины или выбоины).

В качестве еще одной модальности используются данные типа «облако точек», полученные с датчиков лидар и соответствующие входному изображению. Оно проецируется в изображение на основе калибровочных параметров камеры, после чего выделяются те точки, которые попадают в пределы каждой из обнаруженных рамок. Если число таких точек достаточно, то они нормализуются и передаются в модель PointNet для классификации на данных типа «облако точек». Данная модель также формирует свое предсказание, представленное в виде распределения вероятностей по классам.

На заключительном этапе для каждой обнаруженной области осуществляется агрегация предсказаний от всех трех моделей. Для этого соответствующие вероятностные распределения масштабируются с учетом заранее выбранных весов и суммируются. Наиболее вероятный класс в объединенном распределении считается итоговым предсказанием. Такое взвешенное объединение позволяет компенсировать недостатки отдельных модальностей, обеспечивая более устойчивую и точную классификацию.

Каждое финальное предсказание сопоставляется с соответствующей аннотацией из COCO-аннотаций на основе коэффициента перекрытия IoU. Если пересечение превышает заданный порог (например, 0.5), то объект считается корректно обнаруженным. Далее рассчитываются ключевые метрики: средняя точность (average precision), точность классификации (accuracy) по финальным меткам, а также полнота обнаружения (recall) без учета ложных срабатываний.

В процессе объединения предсказаний от трех различных моделей была реализована процедура автоматического подбора оптимальных весовых коэффициентов, обеспечива-

ющих наилучшую итоговую классификацию объектов. В качестве целевой метрики было выбрано значение усредненной F1-меры.

Оптимальной комбинацией весов, обеспечившей наивысшую среднюю F1-меру по целевым классам («трещины» и «выбоины»), оказались значения: 0,1 для предсказаний модели детекции Faster R-CNN; 0,6 для классификатора на основе Swin Transformer и 0,4 для PointNet, обрабатывающего «облака точек».

В таблице 4 представлено сравнение классификации мультимодального подхода с индивидуальными архитектурами, входящими в ее состав.

Таблица 4. / Table 4.

Наименование	**precision**		**recall**		**F1-score**		**accuracy**
	трещины	**выбоины**	**трещины**	**выбоины**	**трещины**	**выбоины**	
Swin Small	0,8899	0,9116	0,9140	0,8869	0,9018	0,8991	0,9005
PointNet	0,9290	0,6286	0,9172	0,6667	0,9231	0,6471	0,8737
Мультимодальный метод	**0,98**	0,87	**0,97**	**0,90**	**0,97**	0,88	**0,9557**

Так, значение метрики полноты выросло на 1,31 % относительно модели Swin Small и 23,33 % относительно PointNet. Значение общей точности выросло на 5,52 % и 8,20 % относительно моделей Swin Small и PointNet соответственно.

В таблице 5 представлено сравнение детекции мультимодального подхода с архитектурой Faster-R-CNN.

Таблица 5. / Table 5.

Наименование	**precision**		**recall**		**F1-score**		**mAP [0.5:0.95]**
	трещины	**выбоины**	**трещины**	**выбоины**	**трещины**	**выбоины**	
Faster-R-CNN	0,93	0,95	0,82	0,61	0,87	0,74	0,166
Мультимодальный метод	**1**	**1**	0,82	**0,88**	**0,90**	**0,94**	**0,3271**

Из представленных данных можно сделать вывод, что метрика IoU при использовании мультимодального подхода выросла более чем в 2 раза. Кроме того, выросли показатели точности, полноты и F1-score, особенно для детекции класса «выбоины». Так, значение метрики полноты для данного класса выросло на 27 %, а F1-score – на 20 %.

Таким образом, мультимодальный подход демонстрирует не только наиболее высокую точность классификации относительно базовых моделей, но и значительное улучшение детекции дорожных дефектов, тем самым подтверждая актуальность использования комплексных данных в интеллектуальных транспортных системах.

Заключение

В работе представлен мультимодальный нейросетевой метод для определения качества дорожного полотна, сочетающий анализ визуальных данных и трехмерных облаков точек. Проведенное исследование демонстрирует, что каждая из рассмотренных модальностей в отдельности – будь то детекция на изображениях с помощью Faster R-CNN или YOLO, классификация визуальных патчей с помощью Swin Transformer или анализ геометрии с помощью PointNet – обладает своими сильными и слабыми сторонами.

Ключевым результатом работы является подтверждение гипотезы о синергетическом эффекте при объединении различных модальностей. Предложенная мультимодальная архитектура, использующая взвешенное суммирование предсказаний, позволила не только достичь высокой общей точности классификации (95,57 %), но и значительно улучшить ключевые для прикладной задачи метрики детекции, особенно для сложноконтрастных дефектов, таких как выбоины. Наблюдаемый рост полноты (recall) и F1-score для этого класса свидетельствует о способности системы минимизировать пропуски дефектов, что является критически важным в контексте безопасности.

Таким образом, мультимодальный подход, представленный в статье, подтверждает свою эффективность и перспективность для внедрения в интеллектуальные транспортные системы для автоматизированного мониторинга состояния дорожной инфраструктуры.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. *Козырев С. В., Полянцева К. А.* Комплексный анализ и сравнение передовых алгоритмов дефектовки дорожного покрытия с использованием различных систем сбора данных // Инженерный вестник Дона. 2024. № 11(119). С. 72–116. EDN: JHKKTB

Kozyrev S.V., Polyantseva K.A. Comprehensive analysis and comparison of advanced road surface defect detection algorithms using various data collection systems. *Inzhenernyy vestnik Dona* [Engineering Bulletin of the Don]. 2024. No. 11(119). Pp. 72–116. EDN: JHKKTB. (In Russian)

2. Ranyal E., Sadhu A., Jain K. Road condition monitoring using smart sensing and artificial intelligence: a review. *Sensors.* 2022. Vol. 22. No. 8. P. 3044. DOI: 10.3390/s22083044

3. Abdelwahed S.H., Sharobim B.K., Wasfey B. et al. Advancements in real-time road damage detection: a comprehensive survey of methodologies and datasets. *Journal of Real-Time Image Processing*. 2025. Vol. 22. P. 137. DOI: 10.1007/s11554-025-01683-1

4. Polyantseva K.A., Gorodnichev M.G. Neural network approaches in the problems of detecting and classifying roadway defects. *Wave Electronics and Its Application in Information and Telecommunication Systems*. 2022. Vol. 5. No. 1. Pp. 364–370. EDN: CFBLOQ

5. *Полянцева К. А.* Разработка алгоритмов накопления данных посредством стереопары и детектирования дефектов дорожного полотна // Современные наукоемкие технологии. 2022. № 5-1. С. 107–112. DOI: 10.17513/snt.39156

Polyantseva K.A. Development of data accumulation algorithms using a stereo pair and detection of road surface defects. *Sovremennye naukoemkie tekhnologii* [Modern High Technologies]. 2022. No. 5-1. Pp. 107–112. DOI: 10.17513/snt.39156. (In Russian)

6. Ma N., Fan J., Wang W. et al. Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms. *Transportation Safety and Environment*. 2022. Vol. 4. No. 4. P. tdac026. DOI: 10.1093/tse/tdac026

7. Toral V., Krushangi T., Varia Harishkumar R. Automated potholes detection using vibration and vision-based techniques. *World Journal of Advanced Engineering Technology and Sciences*. 2023. Vol. 10. No. 1. Pp. 157–176.

8. Wu C., Wang Z., Hu S. et al. An automated machine-learning approach for road pothole detection using smartphone sensor data. *Sensors*. 2020. Vol. 20. No. 19. P. 5564. DOI: 10.3390/s20195564

9. Sholevar N., Golroo A., Esfahani S.R. Machine learning techniques for pavement condition evaluation. *Automation in Construction*. 2022. Vol. 136. P. 104190. DOI: 10.1016/j.autcon.2022.104190

10. Dong D., Li Z. Smartphone sensing of road surface condition and defect detection. *Sensors.* 2021. Vol. 21. No. 16. P. 5433. DOI: 10.3390/s21165433

11. Raslan E., Alrahmawy M.F., Mohammed Y.A. et al. Evaluation of data representation techniques for vibration based road surface condition classification. *Scientific Reports*. 2024. Vol. 14. P. 11620. DOI: 10.1038/s41598-024-61757-1

12. Jahan I.A., Huq A.S., Mahadi M.K. et al. RoadSense: a framework for road condition monitoring using sensors and machine learning. *IEEE Transactions on Intelligent Vehicles*. 2024. DOI: 10.1109/TIV.2024.3486020

13. Gu J., Lind A., Chhetri T.R. et al. End-to-end multimodal sensor dataset collection framework for autonomous vehicles. *Sensors*. 2023. Vol. 23. No. 15. P. 6783. DOI: 10.3390/s23156783

14. Faisal A., Gargoum S. Cost-effective LiDAR for pothole detection and quantification using a low-point-density approach. *Automation in Construction*. 2025. Vol. 172. P. 106006. DOI: 10.1016/j.autcon.2025.106006

15. Yang C., Yang L., Duan H. et al. A review of pavement defect detection based on visual perception. *International Journal of Mechatronics and Applied Mechanics*. 2024. No. 17. Pp. 131–146.

16. Mkrtchian G., Polyantseva K. On the use of an acoustic sensor in the tasks of determining defects in the roadway. *Systems of Signals Generating and Processing in the Field of on Board Communications*. 2024. Vol. 7. No. 1. Pp. 276–280. DOI: 10.1109/IEEECONF60226.2024.10496721

17. Safyari Y., Mahdianpari M., Shiri H. A review of vision-based pothole detection methods using computer vision and machine learning. *Sensors*. 2024. Vol. 24. No. 17. P. 5652. DOI: 10.3390/s24175652

18. Chen W., Yang J.S., Xia C. et al. Road surface damage detection based on enhanced YOLOv8. *Computers in Industry*. 2025. Vol. 173. P. 104363. DOI: 10.1016/j.compind.2025.104363

19. Lincy A., Dhanarajan G., Kumar S.S., Gobinath B. Road pothole detection system. *ITM Web of Conferences*. 2023. Vol. 53. P. 01008. DOI: 10.1051/itmconf/20235301008

20. Yang L., Deng J., Duan H. et al. An efficient fusion detector for road defect detection. *Scientific Reports*. 2025. Vol. 15. P. 27959. DOI: 10.1038/s41598-025-01399-z

21. Song W., Zhang Z., Zhang B. et al. ISTD-PDS7: A benchmark dataset for multi-type pavement distress segmentation from ccd images in complex scenarios. *Remote Sensing*. 2023. Vol. 15. No. 7. P. 1750. DOI: 10.3390/rs15071750

22. Zuo C., Huang N., Yuan C., Li Y. Pavement-DETR: a high-precision real-time detection transformer for pavement defect detection. *Sensors*. 2025. Vol. 25. No. 8. P. 2426. DOI: 10.3390/s25082426

23. Arya D., Maeda H., Ghosh S.K. et al. RDD2022: a multi-national image dataset for automatic road damage detection. *Geoscience Data Journal*. 2024. Vol. 11. Pp. 846–862. DOI: 10.1002/gdj3.260

24. Xiao X., Li Zh., Wang W. et al. TD-RD: a top-down benchmark with real-time framework for road damage detection. *2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* Hyderabad, India, 2025. Pp. 1–5. DOI: 10.1109/ICASSP49660.2025.10888616

25. Abdelkader M.F., Hedeya M.A., Samir E. et al. EGY_PDD: a comprehensive multi-sensor benchmark dataset for accurate pavement distress detection and classification. *Multimedia Tools and Applications*. 2025. Vol. 84. Pp. 38509–38544. DOI: 10.1007/s11042-025-20700-w

26. Xiao X. et al. Roadbench: A vision-language foundation model and benchmark for road damage understanding. *arXiv preprint arXiv:2507.17353*. 2025. URL: https://arxiv.org/abs/2507.17353. (accessed 09/01/2025)

27. Khandakar A., Michelson D.G., Naznine M. et al. Harnessing smartphone sensors for enhanced road safety: a comprehensive dataset and review. *Scientific Data*. 2025. Vol. 12. P. 418. DOI: 10.1038/s41597-024-04193-0

28. Polyantseva K., Gorodnichev M. On the applicability of multimodal neural network methods for determining the quality of the road surface. *2025 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO)*. Tyumen, Russian Federation, 2025. Pp. 1–6. DOI: 10.1109/SYNCHROINFO65403.2025.11079337

29. Ren S., He K., Girshick R., Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *arXiv preprint arXiv:1506.01497*. 2016. DOI: 10.48550/arXiv.1506.01497

30. Terven J., Cordova-Esparza D. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*. 2024. Vol. 5. Pp. 1680–1716. DOI: 10.48550/arXiv.2304.00501

31. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. *arXiv preprint arXiv:1512.03385*. 2015. DOI: 10.48550/arXiv.1512.03385

32. Tan M., Le Q.V. EfficientNet: rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*. 2019. DOI: 10.48550/arXiv.1905.11946

33. Liu Z., Lin Y., Cao Y. et al. Swin transformer: hierarchical vision transformer using shifted Windows. *arXiv preprint arXiv:2103.14030*. 2021. DOI: 10.48550/arXiv.2103.14030

34. Ma L., Li Y., Li J. et al. Mobile laser scanned point-clouds for road object detection and extraction: a review. *Remote Sensing*. 2018. Vol. 10. No. 10. P. 1531. DOI: 10.3390/rs10101531

35. Zhao H., Jiang L., Jia J. et al. Point Transformer. *arXiv preprint arXiv:2012.09164*. 2021. DOI: 10.48550/arXiv.2012.09164

36. Qi C.R., Su H., Mo K., Guibas L.J. PointNet: deep learning on point sets for 3d classification and segmentation. *arXiv preprint arXiv:1612.00593*. 2017. DOI: 10.48550/arXiv.1612.00593

37. Qi C.R., Yi L., Su H., Guibas L.J. PointNet++: deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*. 2017. DOI: 10.48550/arXiv.1706.02413

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflict of interest.

Финансирование. Исследование проведено без спонсорской поддержки.

Funding. The study was performed without external funding.

Информация об авторах

Городничев Михаил Геннадьевич, кан. техн. наук, доцент, декан факультета «Информационные технологии», Московский технический университет связи и информатики;
111024, Россия, Москва, ул. Авиамоторная, 8А;
m.g.gorodnichev@mtuci.ru, ORCID: https://orcid.org/0000-0003-1739-9831, SPIN-код: 4576-9642
Полянцева Ксения Андреевна, кан. техн. наук, доцент кафедры «Интеллектуальный анализ данных», Московский технический университет связи и информатики;
111024, Россия, Москва, ул. Авиамоторная, 8А;
k.a.poliantseva@mtuci.ru, ORCID: https://orcid.org/0000-0002-7102-4208, SPIN-код: 8112-8560
Разумовский Игорь Денисович, студент, Московский технический университет связи и информатики;
111024, Россия, Москва, ул. Авиамоторная, 8А;
igor.raz@list.ru

Information about the authors

Mikhail G. Gorodnichev, Candidate of Engineering Sciences, Associate Professor, Dean of the Faculty of Information Technology, Moscow Technical University of Communications and Informatics;
8A, Aviamotornaya street, Moscow, 111024, Russia;
m.g.gorodnichev@mtuci.ru, ORCID: https://orcid.org/0000-0003-1739-9831, SPIN-code: 4576-9642
Ksenia A. Polyantseva, Candidate of Technical Sciences, Associate Professor of the Department of Data Mining, Moscow Technical University of Communications and Informatics
8A, Aviamotornaya street, Moscow, 111024, Russia;
k.a.poliantseva@mtuci.ru, ORCID: https://orcid.org/0000-0002-7102-4208, SPIN-code: 8112-8560
Igor D. Razumovsky, Student, Moscow Technical University of Communications and Informatics;
8A, Aviamotornaya street, Moscow, 111024, Russia;
igor.raz@list.ru