ПОЛИТЕХ
Санкт-Петербургский политехнический университет Петра Великого

На правах рукописи

Филатов Николай Сергеевич

МЕТОД УСТОЙЧИВОГО КОМПЛЕКСИРОВАНИЯ ДАННЫХ ДЛЯ ОБНАРУЖЕНИЯ И ОЦЕНКИ ПРОСТРАНСТВЕННОГО ПОЛОЖЕНИЯ ОБЪЕКТОВ В СИСТЕМАХ ТЕХНИЧЕСКОГО ЗРЕНИЯ МОБИЛЬНЫХ РОБОТОВ

2.5.4. Роботы, мехатроника и робототехнические системы

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата технических наук

Санкт-Петербург

2026

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Санкт-Петербургский политехнический университет Петра Великого».

Научный руководитель кандидат технических наук, доцент
Бахшиев Александр Валерьевич

Официальные оппоненты:

доктор технических наук, профессор **Карпов Алексей Анатольевич**	Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук», Лаборатория речевых и многомодальных интерфейсов, главный научный сотрудник, г. Санкт-Петербург
кандидат физико-математических наук **Кульминский Данил Дмитриевич**	Автономная некоммерческая образовательная организация высшего образования «Научно-технологический университет «Сириус», Научный центр информационных технологий и искусственного интеллекта, доцент направления «Математическая робототехника», федеральная территория «Сириус»

Ведущая организация

федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)», г. Санкт-Петербург

Защита состоится «15» мая 2026 года в 12:00 часов на заседании диссертационного совета У.2.5.4.16 федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский политехнический университет Петра Великого» (195251, г. Санкт-Петербург, ул. Политехническая, 29, главный корпус, аудитория 118).

С диссертацией можно ознакомиться в библиотеке и на сайте https://elib.spbstu.ru федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский политехнический университет Петра Великого».

Автореферат разослан «___»_____ 202_ г.

Ученый секретарь
диссертационного совета У.2.5.4.16
кандидат технических наук, доцент О.В. Кочнева

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Трёхмерное обнаружение объектов является ключевым компонентом мобильных робототехнических платформ и автономного транспорта. От качества и задержки восприятия напрямую зависят безопасность движения, стабильность траекторного планирования и общий уровень автономности. При этом на безопасность автономной мобильной робототехнической системы влияет не только качество и временная задержка обнаружения объектов, но и отказоустойчивость системы к возможным сбоям датчиков. В реальных условиях часть облака точек сенсора LiDAR (лидар) может отсутствовать или содержать очень мало информации из-за окклюзий, технических неполадок, плохих погодных условий. Аналогично видеокамеры могут выходить из строя, загрязняться или предоставлять малое количество информации при плохом освещении. Дополнительным вызовом становится требование к высокому быстродействию на ограниченном аппаратном обеспечении, типичном для мобильных роботов. Вклад в исследование методов трехмерного обнаружения объектов и их ограничений вносят ряд зарубежных и отечественных ученых, таких как: Wang H., Yan J., Bai X., Юдин Д.А., Мелехин А.А.

Анализ существующих решений показывает, что нередко быстродействие достигается ценой снижения качества, а высокая точность – за счёт сложных схем, неустойчивых к сценариям частичных сбоев датчиков. Таким образом, не существует универсального способа, одинаково удовлетворяющего требованиям высокой скорости работы, высокого качества и устойчивости к сбоям в работе датчиков. В связи с этим разработка устойчивого метода комплексирования мультимодальных данных для обнаружения и оценки пространственного положения объектов, который бы обеспечивал высокое качество, быстродействие и отказоустойчивость к частичной потере информации с датчиков, представляет высокую научную и прикладную значимость.

Степень разработанности темы исследования

Задачи трёхмерного обнаружения объектов в робототехнических системах технического зрения традиционно решаются на основе данных LiDAR, которые обеспечивают точные измерения глубины и высокую устойчивость к изменению освещённости (Шибаев В. М., Зуев М.А.). Для обработки облаков точек применяются два базовых подхода: непосредственная работа с точками и дискретизация в воксельные представления. Развиваются и методы, учитывающие разрежённость облака точек: модели на механизме внимания с учётом разрежённости (Wang H.), разрежённые трёхмерные свёртки (Graham B., Engelcke M.), а также свёртки с увеличенным эффективным ядром при разрежённой выборке (Chen Y.).

Параллельно активно исследуются методы, позволяющие получать трехмерные прогнозирования объектов только по данным видеокамеры (Маньшин И.М., Красноперов Н.С.), без использования стереопар, некоторые из таких методов также используют временные последовательности кадров для улучшения точности прогнозирований. Основным преимуществом таких методов является удешевление оборудования для создания автономных мобильных платформ за счет более низкой стоимости видеокамер по сравнению с лидарами.

Наиболее высокие результаты демонстрируют мультимодальные системы, объединяющие преимущества нескольких сенсоров (LiDAR, камеры, радар), поэтому разработка методов комплексирования мультимодальных данных является активным полем исследований. Например, предложены методы, обогащающие облако точек дополнительными точками на основе данных видеокамеры, методы, приводящие признаки разных модальностей в одно пространство (Liu Z.), методы, в которых модальности могут быть использованы независимо, но стимулируется обмен признаками через перекрестное внимание (Yang Z.).

Кроме того, в ряде работ исследуется отказоустойчивость методов трехмерного обнаружения объектов к сбоям датчиков, что дополнительно показывает преимущества мультимодальных систем, поскольку они обеспечивают повышенную отказоустойчивость за

счет избыточности в данных, получаемых с различных датчиков (Liang T.). Показано, что для достижения устойчивости к частичным отказам датчиков при использовании нескольких модальностей необходимо применять специализированные методы обучения и подходы к проектированию нейросетевой архитектуры.

Цель и задачи работы

Целью работы является разработка и исследование мультимодального метода трёхмерного обнаружения объектов, обладающего повышенным быстродействием и увеличенной устойчивостью к отказам датчиков. Для достижения поставленной цели были решены следующие задачи:

1. Анализ быстродействия отдельных модулей открытых методов трехмерного обнаружения для выявления существующих ограничений и неэффективных подходов.
2. Разработка метода комплексирования мультимодальных данных.
3. Разработка архитектурных решений, повышающих отказоустойчивость системы к частичной потере информации с датчиков, а также методов их обучения.
4. Проведение исследований и оценка показателей точности обнаружения объектов, быстродействия, отказоустойчивости к потере данных с датчиков, сравнение с существующими открытыми методами мультимодального трехмерного обнаружения объектов.

Объектом исследования являются алгоритмы и программно-аппаратные средства системы технического зрения в робототехнических системах мобильных платформ для трехмерного обнаружения объектов.

Предметом исследования являются методы эффективного комплексирования мультимодальных данных, а также методы, повышающие отказоустойчивость системы к сценариям частичной потери данных с датчиков мобильного робота.

Научная новизна

1. Разработана архитектура нейронной сети, отличающаяся непрерывным набором признаков в последовательности с использованием радиального и зигзагообразного разбиений и демонстрирующая одновременно высокое быстродействие и высокие показатели качества трехмерного обнаружения объектов.
2. Предложен метод комплексирования данных, отличительной чертой которого является использование слабого соответствия между признаками камеры и облака точек с помощью полярного угла, что позволяет значительно сократить время, затрачиваемое на комплексирование мультимодальных данных в системе технического зрения мобильного робота.
3. Предложен метод обучения, повышающий устойчивость к частичному отказу датчиков мобильного робота с использованием маскированного автоэнкодера в скрытом пространстве и четырех, ранее не предложенных, сценариев маскирования признаков, поощряющих выучивание соответствия позиций и признаков для данных разных модальностей.

Теоретическая и практическая значимость исследования

Теоретическая значимость исследования заключается в развитии методов комплексирования мультимодальных данных для трехмерного обнаружения объектов в мобильных робототехнических платформах, а также в расширении опыта применения маскированных автоэнкодеров для улучшения отказоустойчивости к частичной потере данных с датчиков за счет предложения новых сценариев маскирования и восстановления признаков при использовании маскированного автоэнкодера.

Практическая значимость исследования состоит в разработке архитектуры, которая обеспечивает высокое быстродействие на графических ускорителях потребительского уровня и высокие показатели качества трёхмерного обнаружения объектов, что позволяет использовать систему в задачах навигации и планирования движения автономных мобильных платформ.

Разработанный метод повышения устойчивости к частичной потере данных с датчиков

повышает надежность и безопасность работы системы в ряде нештатных ситуаций.

Таким образом, результаты данной работы применимы в реальных системах мобильной робототехники за счет невысоких требований к вычислительным ресурсам и повышенной отказоустойчивости.

Методология и методы исследования

Экспериментальное исследование проводилось с использованием программных реализаций методов на языке программирования Python с использованием библиотеки глубокого обучения PyTorch. Реализованные модели были обучены и протестированы на открытом наборе данных NuScenes, для оценки качества и сравнения с аналогичными решениями использовались общепринятые метрики обнаружения объектов mAP и NDS. Для обучения моделей и тестирования их быстродействия использовались графические процессоры: NVIDIA RTX 3060, NVIDIA A100.

Положения, выносимые на защиту

1. Архитектура нейронной сети, основанная на предложенном методе набора признаков в последовательности с использованием радиального и зигзагообразного разбиений, демонстрирующая высокое быстродействие и высокие показатели качества трехмерного обнаружения объектов.
2. Метод комплексирования мультимодальных данных датчиков мобильного робота для трехмерного обнаружения объектов, использующий приблизительное сопоставление признаков облака точек, обеспечивающий высокое быстродействие.
3. Метод повышения отказоустойчивости трехмерного обнаружения объектов к частичной потере информации с датчиков мобильного робота с помощью применения маскированного автоэнкодера в скрытом пространстве с использованием предложенных стратегий маскирования и восстановления признаков.

Степень достоверности полученных результатов

Достоверность и научная обоснованность результатов, полученных в диссертации, обеспечена корректным и обоснованным применением существующих методик оценки качества с использованием открытого набора данных NuScenes и сравнением результатов с другими передовыми методами на единой фиксированной выборке данных, не участвовавшей в обучении метода. При выдвижении гипотез и внесении гиперпараметров в разработанную систему проводились серии экспериментов, подтверждающих оптимальность выбранных параметров. Проведенные исследования сопоставляются с результатами отечественных и зарубежных авторов, а также подтверждаются научной экспертизой на конференциях и при публикации материалов в рецензируемых научных изданиях.

Публикации и апробация результатов

Результаты диссертации были представлены в качестве докладов на международных научно-технических конференциях: «Нейроинформатика-2024», «Нейроинформатика-2021».

Результаты работы опубликованы в журналах: «Системы. Методы. Технологии» (2025), «Робототехника и техническая кибернетика» (2025), «Наука и бизнес: пути развития» (2025), «Advances in Neural Computation, Machine Learning, and Cognitive Research VIII» (2024), «Advances in Neural Computation, Machine Learning, and Cognitive Research V» (2022). По теме диссертации опубликовано 5 печатных работ, в том числе 3 в изданиях, входящих в перечень ВАК, и 2 в изданиях, индексируемых в базе Scopus.

Личный вклад автора

Автором лично получены основные результаты диссертационной работы, в том числе разработан метод комплексирования данных и оптимизированные радиальные и зигзагообразные способы набора признаков для эффективной обработки. Автором предложена архитектура с маскированным автоэнкодером в скрытом пространстве, повышающая отказоустойчивость к сценариям частичной потери данных с датчиков. Автор играл ключевую роль в реализации архитектур и получении результатов всех экспериментов.

Соответствие диссертации паспорту научной специальности

Диссертационная работа выполнена в соответствии с паспортом специальности 2.5.4.

Роботы, мехатроника и робототехнические системы по пунктам:

- п.5. Методы, алгоритмы, программные и аппаратные средства управления роботами, робототехническими и мехатронными системами, включая адаптивное, оптимальное, супервизорное управление.

- п.6. Математическое и программное обеспечение, компьютерные методы и средства обработки информации в реальном времени в роботах, робототехнических и мехатронных системах.

Структура и объем работы

Диссертация состоит из введения, четырех глав, заключения, списка терминов, списка литературы и одного приложения. Содержание работы изложено на 120 страницах, включает 27 рисунков, 17 таблиц, 108 источников цитируемой литературы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обозначается актуальность решаемой проблемы, определяется цель и задачи исследования, краткая характеристика научной литературы из данной области, раскрыты научная новизна, теоретическая и практическая значимость работы, представлены основные положения, выносимые на защиту.

В первой главе рассматривается роль трехмерного обнаружения объектов в системах технического зрения автономных мобильных роботов и транспортных средств. Описана функциональная схема системы обеспечения автономности (сбор данных, восприятие, локализация, планирование, управление), приведена связь с уровнями автоматизации SAE J3016 и показаны ожидаемые эффекты (снижение аварийности, повышение доступности и эффективности перевозок) при сохранении требований к безопасности. Обоснована ключевая роль трёхмерного обнаружения объектов как связующего звена между сенсорной системой и высокоуровневыми модулями. Систематизированы основные датчики (лидар, радар, камеры, ультразвуковые средства), их сильные и слабые стороны, вопросы внутренней/внешней калибровки и синхронизации. Рассмотрены области эксплуатации (город, автомагистрали) и компоновки платформ (ретрофит и шатл), а также смежные классы роботов (складские, доставки, сельскохозяйственные, БПЛА) с типовыми конфигурациями сенсоров. Отмечены ограничения вычислительных ресурсов и требования работы в реальном времени. Перечислены инженерные пути их преодоления (облегчённые модели, специализированные процессоры и ПЛИС, событийные камеры и спайковые сети, распределённые вычисления). На основании обзора сформулированы ключевые проблемы для последующих глав: обеспечить сочетание точности, быстродействия и отказоустойчивости за счёт эффективного объединения данных разных сенсоров и рациональной организации вычислений.

Во второй главе представлен аналитический обзор и анализ тенденций развития методов трёхмерного обнаружения объектов. Рассмотрены методы, использующие только лидар и только камеру, затем мультимодальные методы. Приведено сопоставление метрики mAP и числа параметров для репрезентативных методов: методы, использующие только камеры, заметно продвинулись (лучшие метрики сопоставимы с лидарными методами уровня 2022 года), однако по совокупности показателей мультимодальные системы стабильно превосходят унимодальные, а методы, использующие камеры, часто требуют большего числа параметров из-за обработки нескольких изображений, собранных по периметру транспортного средства, и необходимости разрешения глубинной неоднозначности.

Проведён обзор ключевых эталонных наборов данных (KITTI, NuScenes, Waymo, Argoverse и др.), показан переход сообщества от малых и однородных к масштабным и разнообразным наборам данных с подробными аннотациями и атрибутами объектов.

Выполнен анализ публикационной активности (2014-2024), отобрано 2867 релевантных публикаций. Зафиксирован рост интереса к монокулярным методам и смещение фокуса на набор данных NuScenes, доминирование набора данных KITTI уменьшается.

Третья глава посвящена созданию метода комплексирования мультимодальных данных и архитектуре трехмерного обнаружения объектов на его базе, которая обладает высоким быстродействием и точностью, сопоставимой с аналогичными методами трехмерного обнаружения объектов.

В отличие от задач обработки изображений и естественного языка, в которых данные изначально представлены в структурированном формате, обработка облака точек представляет больше сложностей из-за неструктурированного формата данных и разреженности данных. Еще более сложной задачей является совместное использование данных лидара и камеры. Также сложно решить эту задачу при сохранении высокого быстродействия, поскольку требуется не только обработать больше данных, но и произвести операции для их сопоставления.

Высокая точность и надежность имеют первостепенное значение при трехмерном обнаружении объектов, особенно потому, что его основное применение в автономном вождении напрямую влияет на безопасность дорожного движения. Мультимодальные методы имеют решающее значение, поскольку они повышают надежность систем обнаружения, обеспечивая критически важные отказоустойчивые решения против неисправностей LiDAR или неблагоприятных погодных условий, которые могут скрыть критические данные.

Другим критическим аспектом систем обнаружения 3D-объектов является требование высокого быстродействия. Эти системы должны работать в режиме реального времени, часто в рамках ограниченных вычислительных ресурсов автономного транспортного средства. Несмотря на то, что некоторые недавние работы были посвящены повышению быстродействия, в них используются высококлассные серверные графические процессоры, которые малопригодны для реального развертывания на мобильных устройствах.

В данной работе показано, что существующие модули комплексирования данных в современных фреймворках обнаружения 3D-объектов демонстрируют значительную неэффективность. Для решения этой проблемы предложен новый подход к объединению 3D-данных с датчика LiDAR с 2D-изображениями камеры через модуль трансформера, который использует мультимодальные токены, организованные в последовательности с помощью инновационных методов радиального и зигзагообразного разбиений. Эти методы обеспечивают минимальные вычислительные затраты и сокращают количество последовательностей, необходимых для обработки.

При разработке эффективного мультимодального алгоритма трехмерного обнаружения объектов неизбежно возникает вопрос сопоставления признаков полученных с датчиков разных модальностей: лидара, камеры, радара. Наиболее логичным и распространенным подходом является сопоставление признаков в трехмерном пространстве, что требует решения неоднозначной задачи восстановления трехмерных координат из двухмерных координат изображения. Несмотря на то, что устранить неоднозначность возможно с помощью прогнозирования распределения глубин для пикселей, это требует дополнительных затрат на прогнозирование карт глубины. Альтернативные подходы точного сопоставления признаков в трехмерном пространстве также занимают значительное время. Учитывая данные факты, в этой работе выдвигается гипотеза, что точное сопоставление признаков в трехмерном пространстве для методов с высоким быстродействием неприменимо, и может быть заменено приблизительным сопоставлением на основе известных данных о внутренних и внешних параметрах камеры.

Предлагается сопоставлять колонны признаков, полученные с облака точек, с набором пикселей по полярному углу в плоскости вида сверху. Таким образом, для любого полярного угла возможно точно сопоставить набор вертикальных колонн признаков с облака точек и пиксели с изображения, используя для этого внутренние параметры камеры, а также знания о компоновке датчиков на мобильной автономной платформе (Рисунок 1), а именно: соответствующие матрицы перехода из одной системы координат в другую. Таким образом, полярный угол любого пикселя изображения вычисляется согласно (1).

Данные вычисления проводятся для всех шести камер автономного транспортного

средства с учетом их параметров. Таким образом, возможно сопоставить элементы облака точек каждой из камер и также определить области, в которых элементы облака точек присутствуют сразу на двух камерах (Рисунок 2).

$$\begin{aligned} fov_x &= \operatorname{arctg}\left(\frac{w}{2\,f_x}\right), \\ \theta_{cam} &= arctg\left(T_{1,0}, T_{0,0}\right) \cdot T_{2,1}, \\ \theta_i &= \frac{arctg(y_i, x_i) \cdot fov_x}{\pi} + \theta_{cam} - \frac{fov_x}{2}, \end{aligned} \tag{1}$$

где fov_x – горизонтальное поле зрения камеры;

f_x – горизонтальное фокусное расстояние камеры;

w – ширина изображения;

θ_{cam} – угол поворота камеры вокруг оси z в системе координат лидара;

T – матрица перехода из системы координат камеры в систему координат лидара;

θ_i – угол пикселя i в полярных координатах;

y_i, x_i – координаты пикселя i в декартовой системе координат изображения, имеющей начало снизу в середине изображения и осью y направленной вверх.

Иллюстрация найденных полярных углов приведена на рисунке 1.

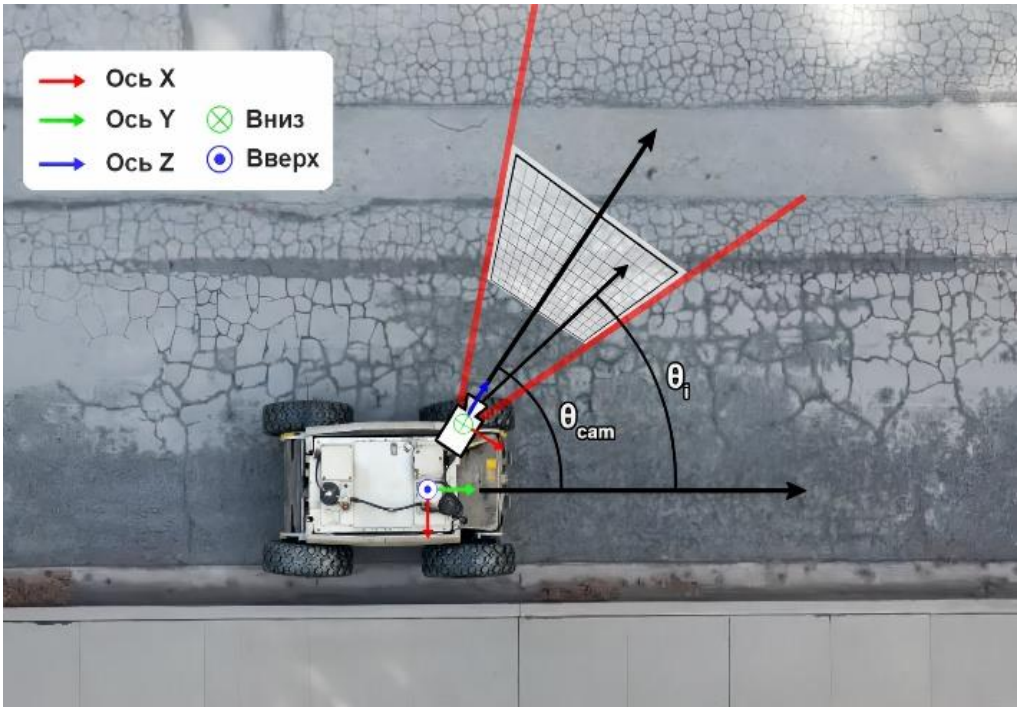

Рисунок 1 – Определение вращения камеры в системе координат лидара и вычисление полярного угла пикселя изображения

Принимая во внимание успех и распространенность архитектур, основанных на трансформерах, а также логичную необходимость параллельной обработки признаков лидара и камеры для максимизации быстродействия, была построена единая модель для извлечения признаков данных лидара и камеры аналогично UniTR подходу, это предполагает три основных этапа:

1. Токенизация облака точек на колонны признаков с использованием динамической вокселизации или аналога.

2. Токенизация изображений (например, на равномерные квадратные патчи со стороной 8 пикселей).

3. Совместная обработка токенов разных модальностей трансформерной архитектурой.

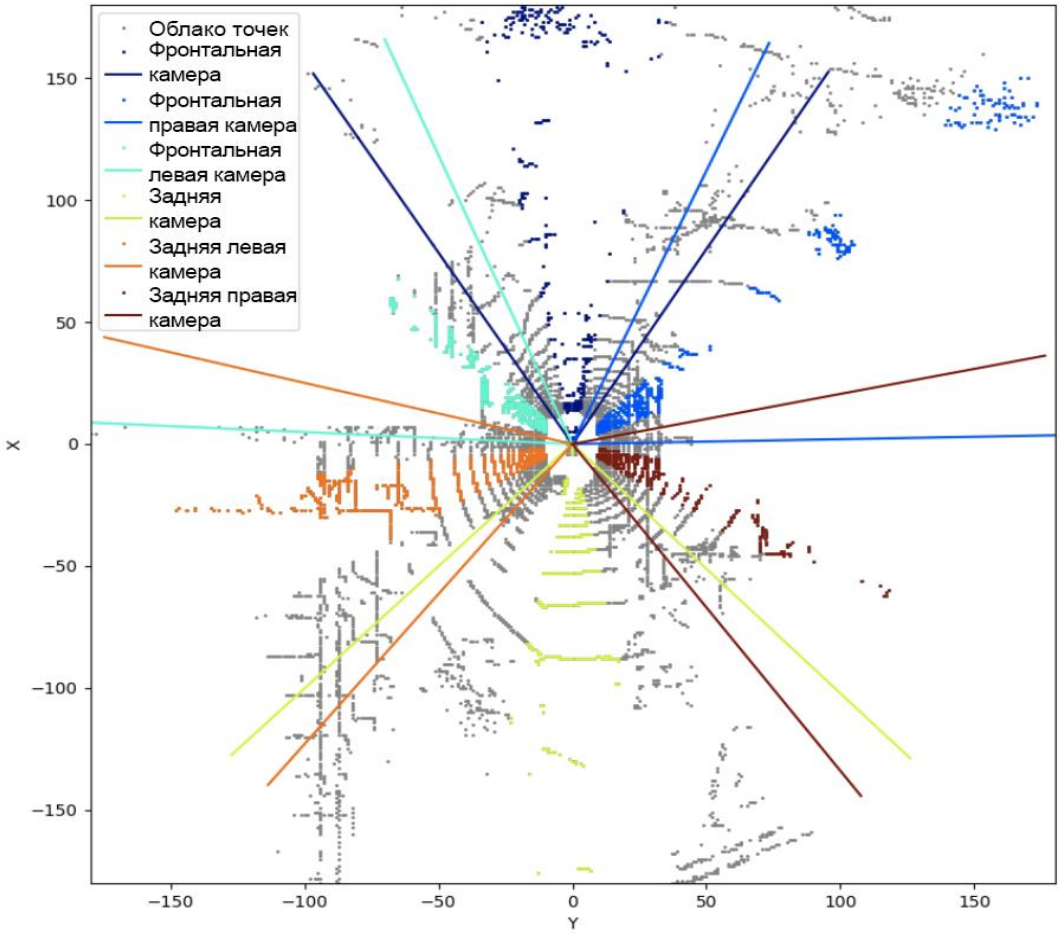

Рисунок 2 – Соответствие облака точек и областей видимостей камер

Данный подход позволяет обрабатывать данные разных модальностей единой архитектурой, производить зависимое от контекста обогащение признаков одной модальностью признаками другой модальности за счет модуля внимания, являющегося основным компонентом трансформерной архитектуры. Однако для корректного и эффективного применения модуля внимания необходимо рационально создавать последовательности токенов для обработки, поскольку модуль внимания имеет квадратичную вычислительную сложность от длины последовательности, были сформированы три гипотезы для рационального формирования последовательностей:

1. Формируем радиальные окна, представляющие собой сектор между двух полярных углов, для каждого окна набираем мультимодальные последовательности, набирая последовательность из сортированных по полярному углу значений до тех пор, пока не достигнута желаемая длина последовательности или не закончились токены в данном радиальном окне. Если достигнута желаемая длина последовательности, последующие токены следует определить в следующую последовательность. Если желаемая длина последовательности еще не достигнута, но уже закончились токены внутри текущего радиального окна, окно необходимо заполнить фиктивными токенами и создать маску наличия таких токенов, которая позволит маскировать их при выполнении модуля внимания и фактически исключать из значений, влияющих на результат. Для выучивания контекста вне рассчитанных окон, возможно проводить сдвиг радиальных окон. Также для улучшения взаимодействия токенов может быть полезно инвертировать сортировку по полярному углу внутри окон для формирования другого набора последовательностей.

2. Известно, что токенов облака точек существенно меньше, чем токенов изображений, а также они обладают наибольшей ценностью. Так в методе UniTR после выполнения нескольких этапов модуля внимания в дальнейших вычислениях используются только токена лидара, а токены изображений отбрасываются. Исходя из аналогичной мотивации можно

применять модуль внимания в режиме кросс-внимания, используя V и Q как последовательность токенов облака точек и K как последовательность токенов изображений, чтобы максимизировать обогащение токенов лидара информацией с токенов изображений. Возможно использовать такие же радиальные окна как в гипотезе 1, но нужно формировать несколько групп токенов облака точек и раздельно формировать несколько групп токенов изображений. Затем необходимо параллельно выполнять модуль внимания для всех комбинаций групп токенов облака точек и групп токенов изображений. Это будет гарантировать, что внутри радиального окна все токены облака точек получат информацию о всех токенах изображения. Недостатком данной гипотезы является необходимость дополнительной агрегации результатов вычисления для токенов облака точек, которые участвовали в вычислениях более одного раза. Также этот подход исключает выучивание важности токенов облака точек в зависимости от контекстной информации других токенов облака точек. Для решения этой проблемы можно добавить итерации применения модуля внимания только для токенов облака точек, или выбирать их в качестве K при формировании некоторых последовательностей.

3. Выбор токенов по номеру окна, к которому они принадлежат, может быть неэффективной операцией, поскольку это требует итеративного выполнения команд и не может быть выполнено параллельно на графическом процессоре без написания специализированного кода на языке CUDA. Однако нет необходимости набирать в одну последовательность для выполнения внимания только токены, строго принадлежащие одному окну. Вполне логично, что при нехватке токенов в текущем окне следующий токен может быть взят из соседнего окна, при условии, что оно также является соседним не только по номеру, но и по пространственному положению. То есть, фактически, возможно провести сортировку токенов по полярному углу и набирать последовательности подряд для всего массива токенов. Этот подход не только устраняет потенциально неэффективные вычисления при выборе токенов для конкретного окна, но также сокращает потребность в заполнении последовательностей фиктивными токенами при их нехватке в текущем окне. Заполнение последовательности может потребоваться только один раз при достижении конца массива. Более того, такой подход может быть обобщен на произвольные окна, например, на равномерные окна в декартовой системе координат, поскольку потребуется всего лишь присвоить номер окна, что легко выполняется параллельно, а затем произвести сортировку по номерам окон. Окна в виде сетки в декартовых координатах вида сверху или радиально-секторные окна могут быть полезны для обработки отдельно лидарных токенов, поскольку для них доступна точная информация не только о полярном угле, но также и о радиусе – расстоянии до точки в пространстве.

Таким образом, в предложенном подходе, разработанном для работы в реальном времени, утверждается, что точное сопоставление для эффективного комплексирования датчиков не является необходимостью. Вместо того, чтобы пытаться прогнозировать глубины или использовать сложные обходные пути для отображения двумерных данных на трехмерные структуры, используется легкодоступная информация – полярный угол.

Разработанная система использует колонны признаков облака точек и патчи изображений для того, чтобы создать последовательности смежных токенов по полярному углу. Однако для оптимизации набора токенов в последовательностях и улучшения их представления для обработки модулем внимания необходимо назначить некоторые значения полярного радиуса пикселям, которые изначально обладают только данными о полярном угле. Для этой цели были реализованы две методологии:

1. Евклидов метод для расчета полярного радиуса (2): этот метод вычисляет полярный радиус как евклидово расстояние от нижнего центра изображения до координат каждого пикселя. Хотя этот подход естественным образом предполагает, что пиксели, расположенные дальше от центра, представляют более удаленные объекты, он вносит искажения, особенно в углах изображения (Рисунок 3 б).

2. Гибридный евклидов метод (3): для уменьшения искажений и достижения более

равномерного радиального распределения гибридный евклидов метод вычисляет радиус как среднее значение евклидова расстояния и постоянного радиуса для каждого пикселя. (Рисунок 3 в).

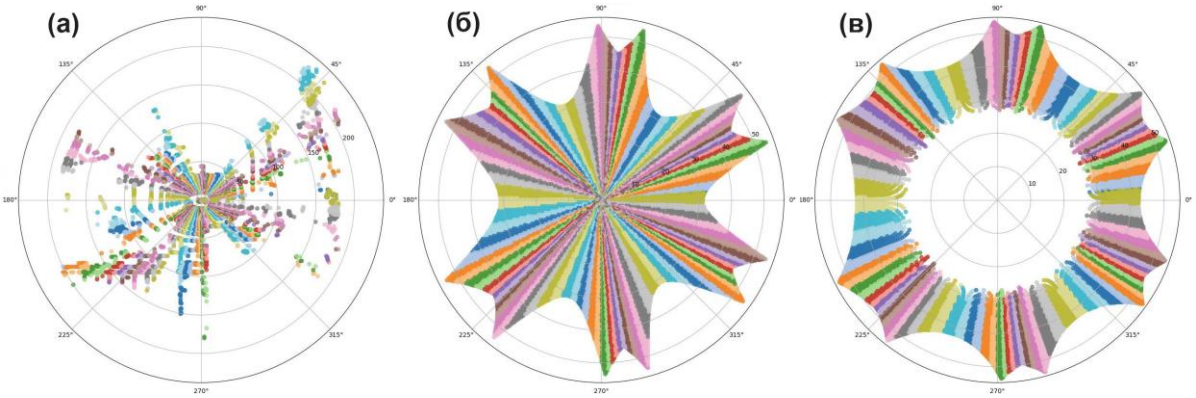

Рисунок 3 – Визуализация радиального разбиения в виде сверху: а) колонны признаков данных LiDAR; б) патчи изображений с полярным радиусом, рассчитанным по евклидовому методу; в) патчи изображений с полярным радиусом рассчитанным по гибридному евклидовому методу

$$R_i^{Euc} = \sqrt{x_i^2 + y_i^2}, \tag{2}$$

$$R_i^{Hyb} = \frac{\sqrt{x_i^2+y_i^2}+r_c}{2}, \tag{3}$$

где R_i^{Euc}– полярный радиус, вычисленный по евклидовому методу для i-го пикселя (патча) изображения;

x_i, y_i – координаты i-го пикселя или патча изображения;

R_i^{Hyb}– полярный радиус, вычисленный по гибридному евклидовому методу;

r_c – константный радиус.

Этот подход обеспечивает надежное соответствие между токенами LiDAR и камеры по полярному углу и пропорциональное приблизительное соответствие по полярному радиусу. В отличие от существующих фреймворков, таких как DSVT и UniTR, которые ограничивают последовательности определенными областями в пространстве вида сверху, предложенный метод непрерывно создает последовательности токенов, используя отсортированные токены сначала по полярному углу, а затем по радиусу. Эта стратегия значительно сокращает избыточные вычисления, поскольку она снижает необходимость в дополнении последовательностей пустыми значениями и маскировании последовательностей для достижения одинаковых размеров, все последовательности полностью заполнены существующими токенов. Единственным исключением может быть конечная последовательность, где циклическое свойство радиального углового разбиения используется для устранения любого дефицита токенов. Например, если последняя последовательность токенов берется с пределом 2π радиан, то недостаток токенов можно устранить путем повторного включения токенов с начала последовательности, около 0 радиан. Это обеспечивает эффективное, непрерывное разбиение токенов, поддерживая соответствие один к одному между индексами входов и выходов в процессе применения модуля внимания, поскольку используются почти все токены и не применяется маскирование токенов. Концепция проиллюстрирована на рисунке 4 б. Более того, для обеспечения обмена информацией между соседними последовательностями производится множественное разбиение на последовательности с использованием различных начальных полярных углов.

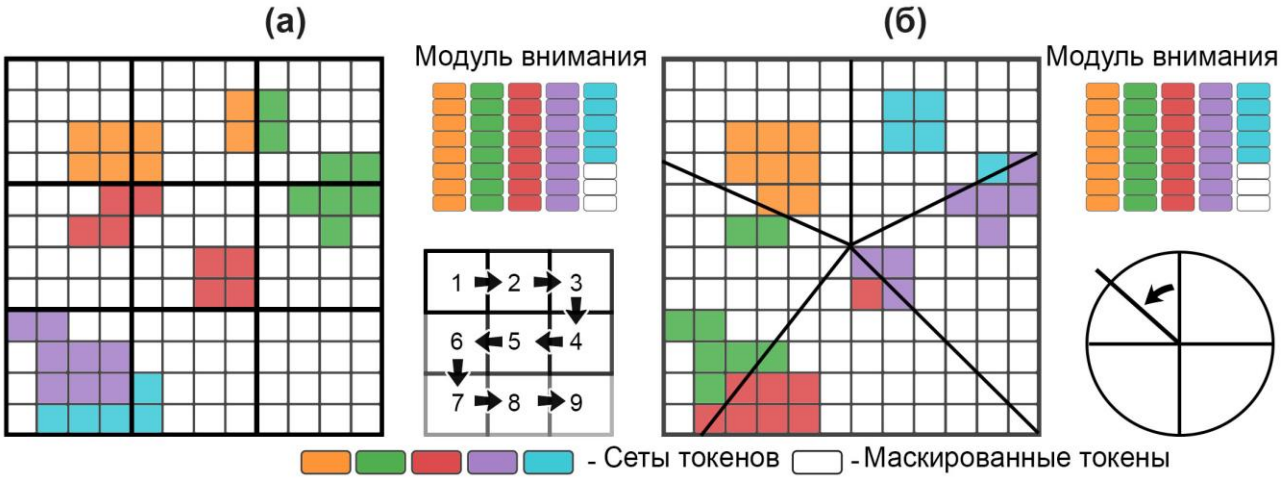

Рисунок 4 – Визуализация непрерывных способов набора токенов: а) зигзагообразный набор токенов облака точек; б) радиальный набор мультимодальных токенов

Данный подход представляет предлагаемую стратегию радиального разбиения токенов на последовательности, повышая эффективность обработки данных и улучшая общую производительность трехмерного обнаружения объектов за счет использования присущих им данных без использования сложных механизмов сопоставления.

Однако обучение нейросети в пространстве вида сверху с использованием только радиальных окон может быть неоптимальным для объектов на близких расстояниях. Интуитивно возникает необходимость включить разбиения по равномерной сетке декартовых координат. Поскольку признаки камеры могут быть представлены в декартовых координатах только с использованием грубо рассчитанного полярного радиуса, то разбиение токенов в декартовых координатах выполняется только для токенов облака точек, полученных с LiDAR датчика.

Прямое назначение токенов из данных LiDAR в последовательности на основе их присутствия в локальных декартовых окнах может привести к неэффективности, в первую очередь, из-за необходимости дополнения (паддинга) и маскирования неполных последовательностей и сложности, связанной с возвращением исходной размерности данных после параллельной обработки частично маскированных последовательностей. Чтобы обойти эти проблемы, вводится метод зигзагообразного непрерывного разбиения токенов (Рисунок 4 а). Суть этого метода заключается в вычислении номеров окон для каждого токена с помощью нескольких простых векторизованных операций, а затем в непрерывном создании последовательностей, используя токены, отсортированные по номеру окна. Чтобы гарантировать, что каждый токен в последовательности является пространственно-смежным, сортировка выполняется таким образом, что после достижения границы пространства вида сверху следующее окно соседствует с предыдущим. Таким образом, номера окон образуют зигзагообразный узор в декартовой сетке пространства BEV. Этот метод гарантирует, что все токены организованы в полностью заполненные последовательности, минимизируя необходимость дополнения, за исключением (потенциально) конечной последовательности. Для перекрестного распространения информации между локальными окнами аналогично DSVT сортировка чередуется между осью X и осью Y в разных наборах последовательностей. Другой используемый способ получения различных наборов последовательностей заключается в том, чтобы начать сортировку локальных окон в обратном порядке. Однако отличие от DSVT заключается в том, что из-за сортировки окон декартовой сетки по зигзагообразному шаблону, токены набираются в последовательности непрерывно из всего пространства вида сверху, а не только из локальных окон декартовой сетки.

Используя эти новые методы набора токенов, определен мультимодальный трансформерный модуль комплексирования (Рисунок 5) признаков. Он состоит из четырех трансформерных блоков, два трансформерных блока обрабатывают мультимодальные токены

модулем внимания, используя радиальное разделение, а другие два трансформерных блока применяют модуль внимания для токенов LiDAR, используя зигзагообразный набор токенов. В каждом трансформерном блоке модуль внимания выполняется дважды с использованием различных наборов последовательностей токенов.

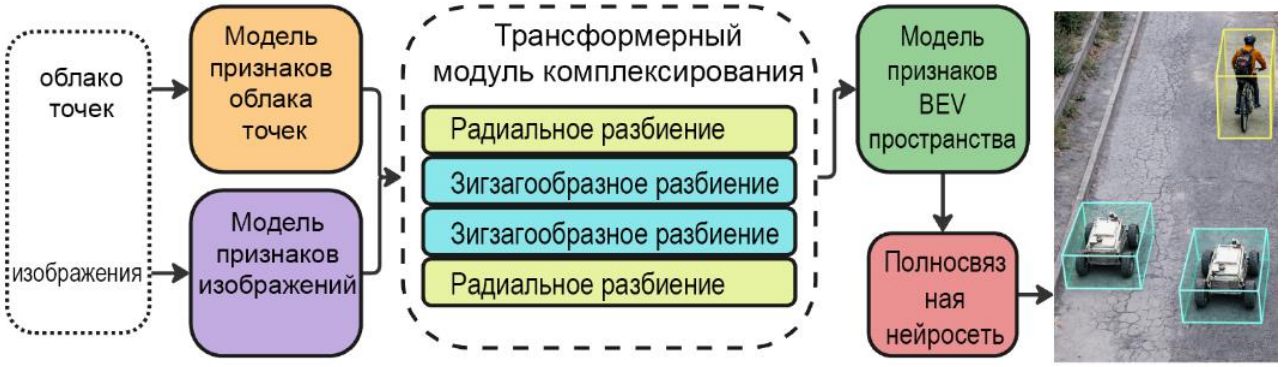

Рисунок 5 – Предложенная архитектура CTP-net (Continuous Token Partitioning)

Вся система включает в себя легковесную нейросеть для извлечения признаков-колонн из облака точек LiDAR, модуль токенизации изображений, модуль комплексирования, двухмерную сверточную нейросеть для обработки признаков в пространстве вида сверху, а также полносвязную нейросеть для прогнозирования объектов. Токенизатор изображений нужен только для разрезания изображений на квадратные участки способом, аналогичным ViT, без дополнительного извлечения признаков. Нейросеть для извлечения колонн-признаков облака точек LiDAR возвращает колонны уже в пространстве вида сверху, поэтому после комплексирования с модальностью камеры результирующие признаки могут быть эффективно обработаны двухмерной сверточной нейросетью, работающей в плоскости вида сверху.

Для проверки предлагаемого подхода и исследования его гиперпараметров был проведен ряд экспериментов (Таблица 1). В этом ряде экспериментов предложенный модуль комплексирования был дообучен в течение шести эпох с использованием замороженных обученных весов других модулей системы, которые были доступны в открытом исходном коде. Для всех исследованных конфигураций использовались оптимизатор AdamW и скорость обучения 0,001 с расписанием одного цикла.

Все модели сравнивались с использованием средней точности (mAP) для трехмерного обнаружения объектов, а также с помощью метрики NDS (NuScenes Detection Score), которая определяется как взвешенная сумма mAP и набора из пяти метрик ошибок TP (4), где каждая метрика ошибок mTP представляет собой усредненное по классам значение следующих ошибок: ошибка сдвига, ошибка масштаба, ошибка ориентации, ошибка скорости, ошибка атрибута.

$$NDS = \frac{1}{10}[5\text{mAP} + \sum_{mTP \in \mathbb{TP}}(1 - \min(1, mTP))], \quad (4)$$

где mAP – средняя точность трехмерного обнаружения объектов,

$mTP \in \mathbb{TP}$ – одна из метрик ошибок усредненная по классам набора данных среди следующих метрик ошибок: ошибка сдвига, ошибка масштаба, ошибка ориентации, ошибка скорости, ошибка атрибута.

Полученные результаты показали, что расчет евклидового полярного радиуса для участков изображения уступает гибридной евклидовой стратегии, поскольку первая вносит больше неестественных искажений в данные. Визуализация последовательностей токенов, полученных с помощью этих различных стратегий, показана на рисунке 3.

Таблица 1 – Тестирование влияния гиперпараметров, L – LiDAR, C – Камера

Модальность	Длина последовательности	Метод вычисления полярного радиуса	mAP	NDS	FPS RTX 3060
L	90	–	59,97	66,68	9,14
LC	90	Euclidian	59,89	66,86	8,50
LC	192	Euclidian	62,59	68,29	7,84
LC	256	Euclidian	62,80	68,31	7,71
LC	90	Hybrid Euclidian	62,82	68,38	8,50
LC	192	Hybrid Euclidian	63,21	68,71	7,83
LC	256	Hybrid Euclidian	64,00	69,00	7,73

Поскольку увеличение длины последовательностей, обрабатываемых модулем внимания, может сократить количество мультимодальных последовательностей, в которых были размещены только токены камеры, и потенциально позволяет получить наиболее релевантный контекст, были обучены модели с длинами последовательности 90, 192 и 256. Увеличение длины последовательности улучшает метрику в каждом эксперименте, и, таким образом, максимальный NDS 69,0 достигается при длине последовательности 256 токенов на последовательность. Для этих конфигураций также измерялась задержка с использованием графической карты GeForce RTX 3060, снижение скорости от модели с длиной последовательности 90 до модели с длиной последовательности 256 составляет менее одного FPS (кадров в секунду) с текущей архитектурой. Как было показано ранее, модуль комплексирования занимает относительно низкий процент времени по сравнению со всей системой, это объясняет небольшое снижение скорости.

Таблица 2 – Показатели качества трехмерного обнаружения объектов на наборе данных NuScenes

Метод	mAP	NDS	FPS RTX 3060	FPS A100	$NDS \times FPS_{3060}$
DeepInteraction	69,9	73,4	0,67	1,85	49,2
BEVFusion	68,5	71,4	1,64	7,20	117,1
CMT	67,9	70,8	5,47	14,31	387,3
UniTR	70,0	73,1	3,98	11,23	290,9
CTP-net	62,8	68,4	8,50	20,72	581,4

Предложенная архитектура CTP-net сравнивалась с существующими мультимодальными фреймворками трехмерного обнаружения объектов (Таблица 2). Для этого сравнения вычислены метрики на валидационной выборке набора данных NuScenes, также приводится скорость работы в кадрах в секунду, измеренная на видеокартах GeForce RTX 3060 и Tesla A100. Показано, что при наличии сопоставимых метрик качества CTP-net достигает самого высокого быстродействия и имеет приемлемую скорость прогнозирования в 8,5 FPS даже при использовании GPU потребительского уровня.

В четвертой главе предлагается использование маскированного автоэнкодера для улучшения обобщающей способности метода комплексирования данных, а также для улучшения его устойчивости к частичной потере данных с датчиков.

При использовании нейросетей для трехмерного обнаружения объектов в автономных роботах критически важным является скорость прогнозирования. Так в данной работе уже представлена архитектура CTP-net, в которой предлагаются упрощенные архитектурные решения, значительно повышающие быстродействие системы.

Однако еще одним важным свойством для мультимодального метода трехмерного обнаружения объектов является устойчивость к возможным ситуациям отказа датчиков. Анализ показал, что архитектуры UniTR и CTP-net проявляют низкую устойчивость, если часть области лидарных данных перекрыта, например, если доступен только диапазон от $-\pi/2$

до π /2, или от - π/3 до + π/3 (Таблица 5). Это связано с тем, что в данных архитектурных решениях после обработки мультимодальных токенов в пространство вида сверху передаются только обновленные токены лидарных признаков, соответственно, если признаки изначально отсутствовали в данной области, то мультимодальное взаимодействие токенов не помогает решить проблему. Более того, CTP-net демонстрирует еще меньшую устойчивость и показывает, что фактически признаки камеры почти не используются для прогнозирований, поскольку сценарии с потерей камеры почти не показывают изменения в метриках.

Исходя из названных недостатков, была разработана новая архитектура, вдохновлённая эффективным радиально-зигзагообразным разбиением, но дополненная модулями для повышения устойчивости к сценарию отказа датчиков. В частности, были внесены следующие изменения:

- полноценный энкодер изображений на базе ResNet-50, что повышает значимость и информативность токенов от камеры;

- эффективный трансформер-декодер для непосредственного прогнозирования трехмерных объектов в стиле DETR. Такой подход одновременно избавляет от явной проекции в пространство вида сверху и позволяет обрабатывать любые мультимодальные последовательности, а также устраняет необходимость в алгоритме подавления немаксимумов (NMS).

На рисунке 6 представлена общая архитектура системы. Исходные признаки для изображения извлекаются энкодером ResNet с иерархической пирамидой признаков, а для облака точек используется энкодер на основе разрежённых свёрток (Sparse Convolution), аналогичный SECOND. Полученные признаки разбиваются на радиальные окна в полярных координатах относительно пространства вида сверху и обрабатываются трансформерными блоками (модулями внимания) аналогично CTP-net. По завершении мультимодального взаимодействия для токенов вычисляются новые обучаемые позиционные эмбеддинги. Затем эти токены вместе с обучаемыми векторами-запросами (query) передаются в трансформерный декодер, аналогичный архитектуре в CMT. Результаты последнего слоя декодера напрямую используются для регрессии координат и размеров трехмерных ограничивающих рамок.

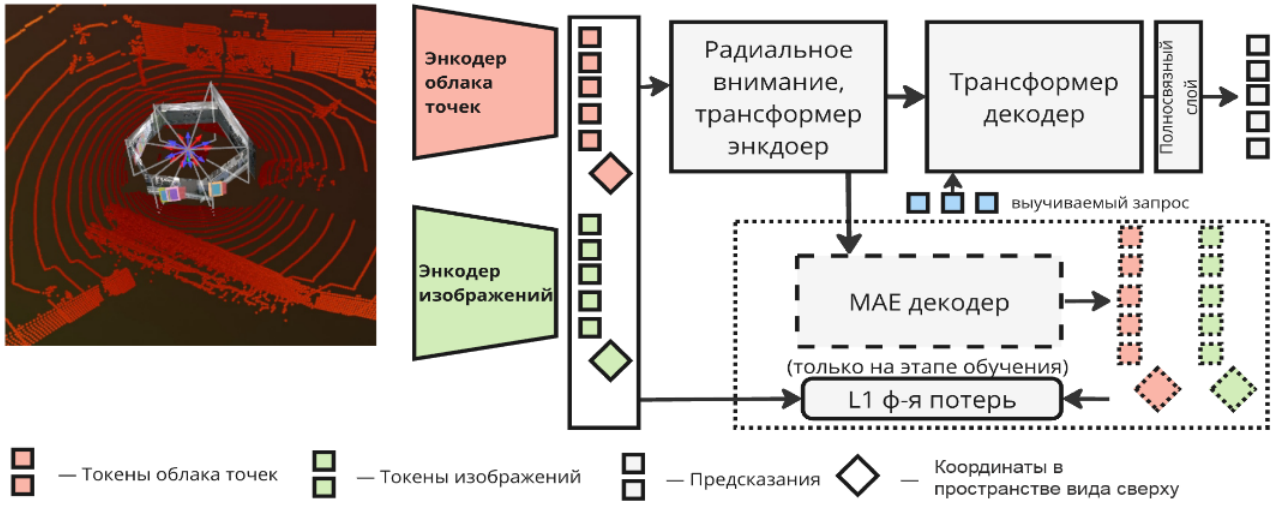

Рисунок 6 – Архитектура для трехмерного обнаружения объектов с использованием мультимодального маскированного автоэнкодера

Поскольку на примере архитектур UniTr и CTP-net видно, что модель может предпочитать использовать признаки лидара, содержащие точную информацию о положении объекта в пространстве и пренебрегать признаками изображения, было принято решение опробовать мультимодальный маскированный автоэнкодер, разработанный специально под задачу поощрения использования токенов изображения. В качестве признаков, подвергающихся маскировке и реконструкции, используются представления предварительно

обученных энкодера облака точек и энкодера изображений, что соответствует предыдущим работам по MAE, в которых показано, что MAE в скрытом пространстве способствует обучению более высокоуровневым признакам.

Для целей поощрения использования токенов изображения для прогнозирования объектов с использованием MAE были разработаны четыре специализированные стратегии:

1. Маскирование всех признаков токенов лидара (без изменения их позиционных эмбеддингов).

2. Маскирование только позиционных эмбеддингов лидара (при сохранении самих лидарных признаков).

3. Полное маскирование токенов лидара (и признаков, и эмбеддингов).

4. Маскирование 60% всех мультимодальных токенов (лидара и камер) одновременно, сохраняя их позиционные эмбеддинги.

Интуитивно, смысл задач 1-3 состоит в том, чтобы реконструировать признаки и/или положение токенов облака точек в пространстве вида сверху, используя токены изображений. При этом маскирование позиционных эмбеддингов для того, чтобы модель определяла их положение по контексту, является принципиально новым подходом. Важным нюансом является то, что для расчета функции потерь оценки положения используются исходные немаскированные координаты в пространстве вида сверху и координаты, спрогнозированные при помощи MLP для каждого реконструированного токена. Пример реконструированных координат для признаков облака точек изображен на рисунке 7. Задача 4 представляет собой традиционную задачу реконструкции данных в MAE, но в мультимодальной постановке.

Архитектура декодера при обучении MAE является вариацией исходного радиального энкодера на трансформерах, адаптированной для решения задачи восстановления и сопоставления токенов в скрытом пространстве. Таким образом, моделируются мультимодальные представления более высокого уровня, что, как ожидается, позитивно сказывается на способности детектора использовать камерные признаки.

Разработанная архитектура была использована в серии экспериментов. Использовался набор данных NuScenes, являющийся крупным мультимодальным набором данных для трехмерного обнаружения объектов. В качестве метрик качества используется усредненная точность mAP и комбинированная метрика набора данных NDS, выражающая точность обнаружений и точность прогнозирования атрибутов объектов.

Все эксперименты проводились в режиме дообучения в течение 3 эпох, с использованием начальных весов модулей из энкодеров облака точек и изображений из CTP-net, веса которых не обновлялись во время обучения, а извлекаемые признаки служили целевым выходом для MAE.

Сравнение метрик качества и быстродействия приведено в таблице 3. Данные результаты показывают, что разработанный метод существенно повысил метрики качества по сравнению с CTP-net, сохранив сопоставимое быстродействие.

Таблица 3 – Показатели качества трехмерного обнаружения объектов на наборе данных NuScenes

Метод	mAP	NDS	FPS RTX 3060	FPS A100	$NDS \times FPS_{3060}$
DeepInteraction	69,9	73,4	0,67	1,85	49,2
BEVFusion	68,5	71,4	1,64	7,20	117,1
CMT	67,9	70,8	5,47	14,31	387,3
UniTR	70,0	73,1	3,98	11,23	290,9
CTP-net	62,8	68,4	8,50	20,72	581,4
CTP2-MAE	66,0	70,0	8,23	20,11	576,1

Для анализа робастности разработанного решения рассматриваются пять сценариев нарушения работы датчиков:

1. Видимость лидара от $-\pi/2$ до $\pi/2$ (половина обзора перекрыто).

2. Видимость лидара от - π/3 до + π/3.

3. Потеря точек объекта (с вероятностью 50 % все точки внутри ограничивающей рамки объекта удаляются, что симулирует возможную потерю лучей лидара от отражающих поверхностей).

4. Потеря фронтальной камеры.

5. Частичное перекрытие трех камер (перекрытие левой половины передней камеры и правой половины для задней правой и задней левой камер).

В результате анализа устойчивости методов составлена таблица метрики NDS при различных сценариях отказа датчиков (Таблица 4) и аналогичная таблица для метрики mAP (Таблица 5). Анализ метрик качества в случае различных неисправностей датчиков показал, что реализованная архитектура устраняет недостатки UniTR и CTP-net, которые имеют сильное падение точности при ограниченных данных лидара.

Таблица 4 – Метрика качества NDS на наборе данных NuScenes при моделировании различных сценариев отказа датчиков

Метод	Видимость лидара π/2	Видимость лидара π/3	Потеря точек объекта	Потеря фронтальной камеры	Частичное перекрытие трех камер
DeepInteraction	54,5	48,3	67,3	72,9	72,9
BEVFusion	53,1	47,6	62,6	72,2	70,8
CMT	53,9	47,2	66,9	70,5	70,4
UniTR	53,2	47,8	62,6	72,2	72,2
CTP-net	49,5	44,5	55,9	68,2	68,2
CTP2-MAE	51,8	47,0	65,8	68,9	68,9

Таблица 5 – Метрика качества mAP на наборе данных NuScenes при моделировании различных сценариев отказа датчиков

Метод	Видимость лидара π/2	Видимость лидара π/3	Потеря точек объекта	Потеря фронтальной камеры	Перекрытие трех камер
DeepInteraction	38,3	35,1	61,0	69,6	69,7
BEVFusion	37,4	33,9	54,4	68,7	68,2
CMT	37,9	33,1	59,8	66,9	67,7
UniTR	31,4	21,9	52,4	68,6	68,8
CTP-net	26,6	18,2	41,5	62,5	62,5
CTP2-MAE	38,0	33,3	59,7	63,9	63,9

Дополнительно были проведены эксперименты, посвящённые оптимальному режиму включения маскированного автоэнкодера (MAE) в процесс обучения (Таблица 6). В качестве базовой контрольной модели рассматривался вариант, где не использовался маскированный автоэнкодер. Далее проверялся способ, при котором на каждой итерации с вероятностью 0,5 производилось маскирование части последовательностей и одновременно вычислялись функции потерь как для обнаружения объектов, так и для их реконструкции. Наблюдалось некоторое ухудшение итоговой точности, обусловленное тем, что учёт ошибки обнаружения по замаскированным последовательностям оказывался не вполне корректным.

Таблица 6 – Анализ влияния различных способов использования MAE

Способ использования MAE	mAP	NDS
Не используется	64,1	68,6
Задачи реконструкции и обнаружения объектов обучаются на каждой итерации	63,3	67,8
Задачи реконструкции и обнаружения объектов обучаются на различных итерациях	65,4	69,5
Донастройка после обучения MAE	66,0	70,0

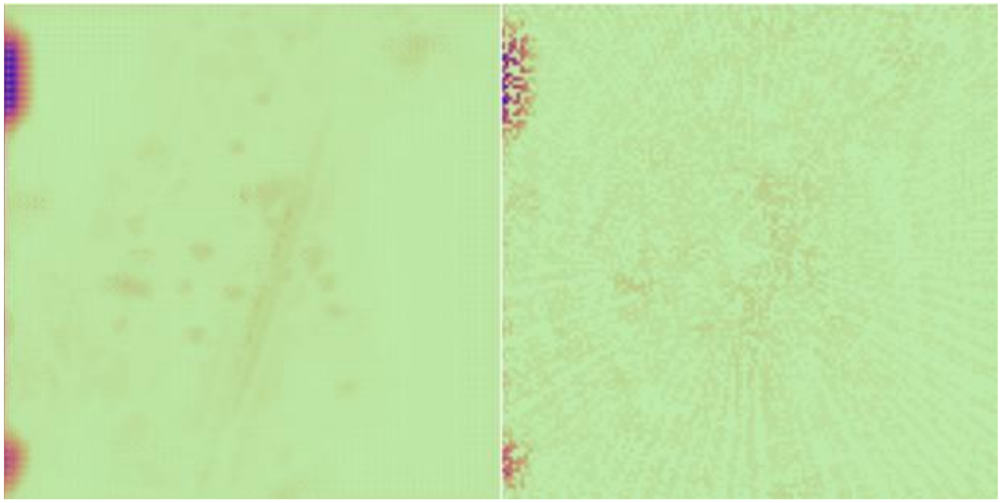

Рисунок 7 – Сравнение исходной карты признаков лидара (слева) и карты признаков с восстановленными координатами токенов лидара по данным признакам лидара без позиционных эмбеддингов и по признакам камеры с позиционными эмбеддингами (справа)

Альтернативный подход заключался в раздельном учёте функций потерь: задача обнаружения оптимизировалась только на немаскированных последовательностях, а задача реконструкции – на остальных. В этом случае негативное влияние оказывалось меньше, и итоговая точность повышалась. Однако наиболее высокие метрики (mAP и NDS) удалось получить при схеме, когда модель предварительно обучалась исключительно на задаче реконструкции (MAE) до сходимости, а затем отдельным этапом дообучалась на задачу обнаружения объектов по полному набору данных NuScenes.

ЗАКЛЮЧЕНИЕ

В работе решена задача трёхмерного обнаружения объектов для автономных мобильных платформ при ограниченных вычислительных ресурсах и возможных отказах сенсоров. Обзор показал устойчивое превосходство мультимодальных методов по сравнению с методами, использующими единственную модальность, по метрикам качества и смещение исследовательского фокуса к крупным и разнообразным корпусам (NuScenes).

Предложены два взаимодополняющих решения. Во-первых, CTP-net – архитектура эффективного комплексирования признаков лидара и камер в составе системы технического зрения мобильного робота с приблизительным согласованием по полярному углу (на основе внутренних/внешних параметров камер) и рациональной организацией последовательностей признаков: радиальное разбиение в плоскости вида сверху и зигзагообразное упорядочивание в декартовой сетке для лидарных признаков. Это снижает вычислительную стоимость операций внимания и долю вспомогательных (не нейросетевых) вычислений. Модульный анализ показал значительное ускорение модуля комплексирования при сохранении точности. Во-вторых, мультимодальный маскированный автоэнкодер в скрытом пространстве (CTP2-

MAE), ориентированный на усиление вклада признаков камеры в прогнозирование объектов и повышение отказоустойчивости. Реализованы четыре задачи маскирования (включая маскирование позиционных представлений лидарных признаков с восстановлением их положения по контексту). Наилучшие результаты даёт двухэтапная схема: предварительное обучение задаче реконструкции, затем дообучение для трехмерного обнаружения объектов.

Эксперименты на NuScenes подтвердили практическую применимость: CTP-net достигла mAP 62,8 и NDS 68,4, обеспечив скорость прогнозирования 8,50 Гц на NVIDIA RTX 3060 и 20,72 Гц на NVIDIA A100, что является на 55 % и 45 % быстрее аналогов и подтверждает практическую применимость подхода на графических процессорах потребительского уровня, недоступную аналогичным подходам ранее. Вариант с CTP2-MAE улучшил качество до mAP 66,0, NDS 70,0 при близком быстродействии (8,23 Гц на RTX 3060) и заметно повысил устойчивость к потере части данных датчиков: частичная утрата углового сектора лидара, выпадение точек внутри объектов, потеря фронтальной камеры и частичное закрытие нескольких камер. В сценариях нарушения полноты облака точек разработанный метод показал наименьшее относительное падение метрик по сравнению с аналогами.

Практическая значимость результатов обусловлена возможностью применения разработанного метода в реальных мобильных робототехнических системах с маломощным аппаратным обеспечением за счет низких требований к вычислительным ресурсам и отказоустойчивости к ряду сценариев отказа датчиков.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, входящих в перечень ВАК:

1. **Филатов Н.С.** Тенденции развития методов трехмерного обнаружения объектов // Системы. Методы. Технологии. – 2025 – № 2. – С. 167-174.

2. **Филатов Н.С.** Мультимодальный маскированный автоэнкодер в скрытом пространстве для трехмерного обнаружения объектов // Робототехника и техническая кибернетика. – Т. 13. – № 4. – Санкт-Петербург: ЦНИИ РТК. – 2025. – С. 301-308.

3. **Филатов Н.С.**, Бахшиев. А.В. Составление оптимальных последовательностей токенов и их позиционных признаков для трёхмерного обнаружения объектов // Наука и бизнес: пути развития. – 2025 – № 11. – С. 67-76 .

Публикации в изданиях, индексируемых в базах Scopus:

4. **Filatov N**., Isakov T., Bakhshiev A. Research on the Applicability of Monocular 3D Object Detection Using CARLA Simulator // *Advances in Neural Computation, Machine Learning, and Cognitive Research V* : proceedings. – Cham : Springer, 2022. – (Studies in Computational Intelligence; Vol. 1008). – P. 224–229.

5. **Filatov N**., Potekhin R. Continuous Token Partitioning for Real-Time Multi-modal 3D Object Detection // *International Conference on Neuroinformatics* : proceedings. – Cham : Springer Nature Switzerland, 2025. – (Studies in Computational Intelligence; Vol. 1179). – P. 426–437.