Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ)

На правах рукописи

Мкртчян Грач Маратович

Разработка методов и средств нейросетевой обработки акустической информации

Специальность 2.3.8 —
«Информатика и информационные процессы»

Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:
кандидат технических наук, доцент
Городничев Михаил Геннадьевич

Москва — 2025

Оглавление

Стр.

Стр.

Введение

Современные и перспективные технические системы требуют информационной поддержки, обеспечивающей обработку информации об их состоянии для принятия решений по управлению, развитию и оптимизации.

С каждым годом наблюдается значительный рост числа автотранспортных средств, увеличение загрузки дорог и возрастание интеллектуальной нагрузки на водителей при управлении транспортным средством. Эти изменения подчеркивают актуальность разработки и внедрения передовых методов и технологий обеспечения безопасности, соответствующих современным направлениям развития автотранспорта и организации дорожного движения. Одним из ключевых инструментов в этой области являются системы помощи водителю (ADAS). Однако такие системы в основном опираются на визуальные данные, поступающие с камер и лидаров. Их эффективность существенно снижается в условиях плохой видимости, неблагоприятных погодных явлений или при наличии препятствий, затрудняющих обзор.

Использование акустических данных даёт возможность анализировать текущую обстановку на дороге, идентифицируя акустические сигналы, исходящие от различных объектов и событий. Это могут быть акустические сигналы приближающихся транспортных средств, сирены экстренных служб, шумы аварийных ситуаций и другие акустические сигналы.

Современные исследования подтверждают перспективность применения акустических данных в системах безопасности. Они включают разработку методов классификации транспортных средств на основе акустических сигналов, анализ акустических сцен с использованием спектральных характеристик и технологий машинного обучения. Такие методы помогают более точно классифицировать различные дорожные ситуации и окружающую среду.

Одним из наиболее перспективных подходов в данной области является использование нейросетевых технологий. Нейронные сети демонстрируют высокую эффективность при обработке акустических данных, включая классификацию акустических сцен, распознавание транспортных средств на основе их акустических подписей и оптимизацию обработки данных за счёт снижения их размерности. Эти технологии подтверждают свою значимость и перспективность для создания интеллектуальных систем оценки дорожной обстановки и

принятия решений при управлении транспортными средствами и дорожным движением.

Степень разработанности темы исследования. Своевременность темы подтверждается большим количеством исследований в этой области. Задачи анализа акустических сигналов окружающей среды представлены в работах: Ю. Леженин, Н. Богач, Ю. Фурлетов, С. Шадрин, Ли, Шваб, Ашхад, Барчиези, Шао, Море, Ибаньес-Гусман, Суноу, Перкус, Тоффа, Миньот, Нанни, Чжао, Инь, Чжан, Лю, Линь, а также Заммана и их соавторов. Эти авторы внесли значительный вклад в разработку методов и технологий анализа акустической информации, разработку алгоритмов глубокого обучения для решения задач классификации акустических сцен, транспортных средств и экологических шумов, полученные результаты могут служить основой для дальнейших исследований. Несмотря на достигнутые успехи, в области анализа акустических сигналов остаются нерешенные задачи и перспективные направления для дальнейших исследований:

- **Улучшение качества классификации в условиях акустического шума:** создание устойчивых к помехам моделей, способных эффективно работать в реальных условиях с высокой степенью фонового шума.
- **Анализ многоканальных акустических данных:** разработка методов обработки пространственных признаков, позволяющих более точно локализовать источники акустических сигналов и анализировать акустические сцены.
- **Интеграция методов мультисенсорного анализа:** комбинирование акустических данных с визуальными или вибрационными данными для повышения точности классификации.
- **Энергоэффективные алгоритмы для встроенных систем:** разработка легковесных моделей глубокого обучения, пригодных для работы на мобильных устройствах и IoT-устройствах.

Целью диссертационной работы является разработка методов и средств нейросетевой обработки акустической информации о дорожных событиях для повышения безопасности дорожного движения посредством добавления дополнительного модуля цифровой обработки сигнала в существующие системы помощи водителям.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Исследовать существующие методы и алгоритмы классификации акустических данных применительно к повышению безопасности движения транспортных средств.
2. Разработать метод сбора и аннотирования акустической информации о дорожно-транспортной обстановке.
3. Спланировать эксперимент сбора, аннотирования и исследования нейросетевых методов классификации акустических данных дорожных событий.
4. Разработать устойчивый алгоритм обучения нейронной сети в условиях выбросов и шумов в обучающем наборе данных за счёт применения робастных функций потерь совместно с дистилляций знаний.
5. Разработать алгоритм классификации акустических данных дорожных событий позволяющий достигать необходимой точности в рамках предметной области.
6. Разработать архитектуру программно-аппаратного комплекса сбора и цифровой обработки акустических данных дорожных событий.

Объектом исследования является математические и технические средства и методы анализа и классификации акустической информации дорожных событий в условиях городской среды.

Предметом исследования являются алгоритмическое и техническое обеспечение нейросетевой системы классификации акустической информации дорожных событий.

Методы исследования. Для решения указанных задач применялись методы автоматической обработки акустических данных, статистического анализа, цифровой обработки сигналов и программирования.

Научная новизна результатов диссертации заключается в разработке совокупности взаимосвязанных алгоритмических, программных, технических и организационных решений, направленных на повышение безопасности дорожного движения путем применения методов обработки акустической информации с использованием нейросетей.

В процессе выполнения диссертационной работы получены следующие оригинальные научные результаты:

1. Метод сбора и аннотирования акустической информации о дорожно-транспортной обстановке, *отличающийся* внедрением предобученной модели распознавания, *позволяющий* повысить скорость аннотирования данных не менее чем на 30%, а также минимизировать человеческий фактор (2.3.8, п.7).
2. Алгоритм повышения устойчивости при обучении нейронной сети, предназначенной для классификации акустических данных дорожных событий, основанный на применении робастной функции потерь совместно с дистилляцией знаний, *позволяющий* минимизировать влияние выбросов и шумов в обучающем наборе данных при добавлении до 15% зашумленных данных, без значимой потери качества (2.3.8, п.4).
3. Алгоритм классификации акустической информации дорожных событий, *отличающийся* от существующих применением слоев Колмогорова-Арнольда, *позволяющий* достигнуть точности не менее 95% в условиях городского шума (2.3.8, п.4).
4. Архитектура программно-аппаратного комплекса сбора, хранения и классификации акустической информации дорожных событий, обладающая возможностью непрерывной обработки цифрового сигнала на борту транспортного средства, *позволяющая* интегрировать в существующие информационные системы помощи водителя дополнительный модуль цифровой обработки акустического сигнала для повышения точности определения дорожной обстановки (2.3.8, п.9).

Теоретическая и практическая значимость определяется возможностью повышения безопасности дорожного движения путем интеграции разработанных методов и алгоритмов классификации акустического окружения в системы помощи водителю (ADAS). Такой подход позволяет дополнить информацию от визуальных сенсоров акустическими данными, что повышает объективность оценки реальной обстановки, точность обнаружения потенциальных источников опасности, особенно в условиях плохой видимости или ограниченного поля зрения камер. Создание и испытания действующего прототипа бортовой системы обработки акустической информации позволяют сделать вывод о возможности практической реализации системы в рамках подсистемы ADAS, что может ускорить распространение и применение подобных

систем на дорогах, делая вождение более безопасным и прогнозируемым. Результаты диссертационной работы могут применяться в отраслях, где требуется классификация акустических сигналов, например, для обеспечения безопасности в общественных местах, на производстве.

Основные положения, выносимые на защиту:

1. Метод сбора акустической информации дорожных событий, *позволяющий* повысить эффективность подготовки набора данных и минимизировать влияние человеческого фактора, что достигается за счёт использования предобученной модели, исключающей вероятность пропуска событий из-за человеческой невнимательности или утомляемости.
2. Впервые представлен набор данных об акустической информации дорожных событий, состоящий из 5 классов общим размером 2600 образцов, собранный в реальных условиях дорожного движения.
3. Алгоритм повышения устойчивости процесса обучения нейронной сети классификации акустических данных, *позволяющий* осуществить перенос информации из крупной модели в компактную, уменьшив её размер до 0.19 млн параметров при сохранении высокой точности (не менее 92%). Это предоставляет возможность использовать модель на устройствах с ограниченными вычислительными ресурсами.
4. Алгоритм классификации акустических данных о дорожных событиях, позволяющий повысить точность компактных нейросетевых моделей не менее чем 3% в условиях зашумленной обстановки.
5. Архитектура нейросетевого программно-аппаратного комплекса сбора, хранения и обработки цифрового сигнала, позволяющего повысить безопасность передвижения транспортных средств на дорогах общего пользования за счет интеграции разработанных методов и средств обработки акустической информации в существующие информационные системы помощи водителям, тем самым при принятии решения анализируется большое количество информации.

Степень достоверности и апробации результатов работы обеспечиваются использованием в качестве базы современных методов и моделей, применяемых для классификации и распознавания акустических данных. Математическую основу исследования составляют адаптированные для решения поставленных задач методы теории обработки сигналов, машинного обучения,

математической статистики и спектрального анализа. Результаты были представлены и обсуждались на ряде значимых международных конференций, в том числе Core A, посвящённых обработке сигналов, телекоммуникациям и применению электроники в информационных системах. Результаты работы докладывались и осуждались на Российских и международных конференциях:

- 2024 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF);
- 2024 Systems of Signals Generating and Processing in the Field of on Board Communications;
- 2023 Systems of Signals Generating and Processing in the Field of on Board Communications;
- 2023 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF);
- 2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO);
- 2024 INTERSPEECH: Conference of the International Speech Communication Association

Результаты также обсуждались на заседании кафедры, а также на научных семинарах в МТУСИ.

Личный вклад.

В ходе исследования автором лично были разработаны и реализованы ключевые подходы, направленные на повышение эффективности и надежности анализа акустических данных в системах помощи водителю :

- обеспечение устойчивости модели нейронной сети для классификации акустических данных, что дало возможность улучшить способность модели сохранять точность предсказаний при наличии внешних возмущений и шумов, характерных для дорожной среды;
- применение метода дистилляции для оптимизации работы модели без потери точности, что позволило уменьшить объем вычислений и ресурсов, необходимых для работы модели, что особенно важно для её применения в условиях ограниченных вычислительных мощностей в реальном времени;
- программно-аппаратный комплекс (прототип) для практического применения и испытания предложенной модели, включающий необходимое программное обеспечение и специализированное оборудование, что поз-

воляет интегрировать решение в системы помощи водителю и другие приложения;

– тестирование и оценка эффективности предлагаемых решений, как в условиях симуляции, так и в реальных условиях для оценки точности и устойчивости модели к различным внешним факторам, оценки её надежности и эффективности при различных сценариях эксплуатации.

Реализация и внедрение. Алгоритмы и архитектура программно-аппаратного комплекса, разработанные в настоящей работе, внедрены в следующих организациях:

– «МКАД» (ООО) (г. Гудермес) и «ЭР СИ ТЕХНОЛОДЖИС» (ООО) (г. Москва) как модуль общего комплекса оценки дорожной ситуации;

– в учебный процесс кафедры «Математическая кибернетика и информационные технологии» Московского технического университета связи и информатики (МТУСИ).

Подтверждается соответствующими актами внедрения результатов диссертационной работы.

Соответствие специальности. Тематика и результаты диссертации соответствуют следующим направлениям специальности: 2.3.8 — «Информатика и информационные процессы».

п.4. «Разработка методов и технологий цифровой обработки аудиовизуальной информации с целью обнаружения закономерностей в данных, включая обработку текстовых и иных изображений, видео контента. Разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения требуемой информации из текстов» .

п.7. «Разработка методов обработки, группировки и аннотирования информации, в том числе, извлеченной из сети интернет, для систем поддержки принятия решений, интеллектуального поиска, анализа» .

п.9. «Разработка архитектур программно-аппаратных комплексов поддержки цифровых технологий сбора, хранения и передачи информации в инфокоммуникационных системах, в том числе, с использованием «облачных» интернет-технологий и оценка их эффективности».

Публикации. Основные результаты по теме диссертации изложены в 12 печатных изданиях, 3 из которых изданы в журналах, рекомендованных

ВАК, 9 — в периодических научных журналах, индексируемых Web of Science и Scopus, в том числе Q2. Зарегистрированы 3 программы для ЭВМ.

Объем и структура работы. Диссертация состоит из введения, 4 глав, заключения и 2 приложений. Полный объём диссертации составляет 135 страниц, включая 44 рисунка и 9 таблиц. Список литературы содержит 102 наименования.

Глава 1. Анализ современных методов и средств нейросетевой обработки акустических данных

В первой главе диссертационного исследования обсуждается необходимость повышения устойчивости нейронных сетей в задачах классификации акустических сигналов. Подчеркивается важность разработки стабильных и надежных алгоритмов, способных эффективно работать в реальных условиях. Рассматриваются уязвимости существующих алгоритмов к внешним шумам и возмущениям, что особенно актуально для приложений в области автоматического анализа дорожных сцен. Описываются различные подходы и алгоритмы предобработки данных, которые способствуют улучшению точности и устойчивости алгоритмов в условиях изменяющейся акустической среды.

Классификация в машинном обучении заключается в построении функции $f : \mathbb{R}^n \to \{1, 2, \ldots, K\}$, которая на основе вектора признаков $\mathbf{x} \in \mathbb{R}^n$ предсказывает класс $y \in \{1, 2, \ldots, K\}$. Для этого используется обучающая выборка $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$, где $\mathbf{x}_i$ — вектор признаков, а y_i — метка класса. Обучение модели заключается в минимизации функции потерь $\mathcal{L}(\boldsymbol{\theta})$, которая измеряет расхождение между предсказанными и истинными классами.

1.1 Современное состояние проблемы обеспечения безопасности дорожного движения

Проблема управления безопасностью сложных технологических процессов, к которым относится управление транспортными средствами и дорожным движением, постоянно находится в центре внимания, так как её решение связано с обеспечением безопасности граждан и объектов инфраструктуры.

Управление безопасностью любой системы связано с принятием решений на основе собранной информации. Информация может различаться по физическим принципам возникновения и представления, длительности существования и качеству фиксации, затратам на сбор и обработку. Кроме того, количество видов информации, требуемой для принятия решений, постоянно увеличивается, что требует создания новых средств и методов её обработки.

В связи с этим требуется решение целого ряда задач сбора и обработки информации для принятия решений. Это в полной мере относится к информации, необходимой для принятия решений при управлении безопасностью дорожного движения. Ранее было достаточно только визуальных данных, которые формировались при наблюдении за ситуацией на дороге, имели незначительный объём и обрабатывались в течение достаточно долгих промежутков времени. Однако, с увеличением видов транспортных средств, загрузки автомагистралей и повышением требований к качеству и количеству обрабатываемых данных возникает необходимость расширения видов информации и сокращения сроков её обработки.

Современные системы помощи водителю, более известные под аббревиатурой ADAS (англ. Advanced Driver Assistance Systems), играют всё более значимую роль в автомобильной промышленности и существенно влияют на безопасность дорожного движения. Поддержка, которую такие системы оказывают водителю, может варьироваться от простых предупреждений (например, сигнал при выходе из полосы движения) до почти полного управления автомобилем в определённых условиях. Международно признанная классификация уровней автоматизации, сформулированная Обществом автомобильных инженеров (англ. Society of Automotive Engineers, SAE), выделяет шесть ключевых уровней автоматизации (SAE J3016)[1]. Ниже приводится детальное описание каждого из этих уровней, включая типичные функциональные возможности, требования к водителю, а также примеры современных реализаций.

Уровень 0: Отсутствие автоматизации (No Automation)

Характеристика. На этом уровне водитель полностью контролирует транспортное средство. Все функции управления — рулевое управление, торможение, ускорение, смена полосы движения и так далее — осуществляются человеком. Технологические решения, доступные в автомобиле (если они вообще присутствуют), не вмешиваются в процесс управления, а лишь предупреждают водителя о возможных опасностях.

Примеры систем.

– *Сигнализация выхода из полосы* (Lane Departure Warning, LDW), которая может подавать акустический или визуальный сигнал, если автомобиль начинает отклоняться от полосы движения. На Уровне 0 система не умеет автоматически корректировать траекторию.

– *Система обнаружения объектов в слепых зонах* (Blind Spot Monitoring, BSM), когда датчики лишь оповещают водителя о присутствии транспортного средства в мёртвой зоне, но не вмешиваются в управление.

Ответственность водителя. Водитель должен полностью концентрироваться на управлении, поскольку система не обладает функционалом для активного вмешательства. Всё принятие решений и реагирование на дорожные события остаются исключительно на человеке.

Уровень 1: Вспомогательные системы (Driver Assistance)

Характеристика. Уровень 1 предполагает наличие систем, способных взять на себя выполнение одной из основных функций управления транспортным средством, таких как рулевое управление или поддержание заданной скорости (круиз-контроль). При этом водитель должен постоянно контролировать обстановку на дороге и быть готовым в любой момент взять управление в свои руки.

Типичные технологии.

– *Адаптивный круиз-контроль* (Adaptive Cruise Control, ACC), который позволяет автоматически поддерживать заданную скорость и дистанцию до впередиидущего транспортного средства.
– *Подруливание* (Lane Keeping Assistance, LKA), позволяющее корректировать рулевое управление, чтобы автомобиль оставался в пределах своей полосы.

Особенности применения. Система на Уровне 1 требует постоянного мониторинга, поскольку не может полностью самостоятельно управлять автомобилем. Она поддерживает лишь одну основную функцию, хотя в современных реализациях могут присутствовать сразу несколько информационных систем (например, предупреждение о препятствиях, помощь при парковке), которые, однако, не синхронизируют между собой действия.

Уровень 2: Частичная автоматизация (Partial Automation)

Характеристика. На данном уровне системы могут одновременно управлять несколькими функциями автомобиля. Как правило, это совмещённое управление рулём и ускорением/торможением. Водитель по-прежнему должен быть готов вмешаться в любой момент и нести полную ответственность за безопасную эксплуатацию транспортного средства.

Примеры современных реализаций.

– *Система удержания полосы* (Lane Centering) совместно с адаптивным круиз-контролем, способная удерживать автомобиль в центре полосы и адаптировать скорость в зависимости от окружающего трафика.
– *Автоматизированная парковка* (Autopark), когда автомобиль может сам управлять рулевым управлением и скоростью во время манёвра парковки, но водитель контролирует ситуацию и при необходимости останавливает процесс.

Роль водителя. Хоть система способна взять на себя сразу несколько функций, например, удержание в полосе и поддержание расстояния до впередиидущего автомобиля, водителю не рекомендуется отвлекаться, поскольку в случае возникновения критической ситуации человек должен незамедлительно взять управление в свои руки.

Уровень 3: Условная автоматизация (Conditional Automation)

Характеристика. Уровень 3 предполагает, что система способна полностью управлять автомобилем в ограниченных условиях, без участия водителя. Эти условия обычно включают определённый тип дорог (например, автомагистраль), скорость движения, а также благоприятные погодные условия. При возникновении нестандартной или сложной дорожной ситуации система может запросить водителя взять управление на себя.

Примеры применения.

– *Высокоавтоматизированное движение по автомагистрали*, когда автомобиль способен самостоятельно осуществлять разгон, торможение, смену полосы, оценивая при этом дорожную обстановку с помощью камер, радаров и лидаров.
– *Управление в пробках* (Traffic Jam Pilot), позволяющее двигаться в плотном медленном потоке без постоянного участия водителя, но при превышении определённой скорости или появлении препятствий система передаёт управление обратно водителю.

Технические и социальные аспекты. Система Уровня 3 требует высокого уровня надёжности сенсоров и алгоритмов принятия решений. При этом законодательные ограничения во многих странах всё ещё не полностью определены для ситуаций, когда автомобиль движется в полностью автоматическом режиме, а водитель в этот момент формально перестаёт контролировать обстановку. Нормативные и этические вопросы, связанные с безопасностью и ответственностью за ДТП, становятся крайне актуальными.

Уровень 4: Высокая автоматизация (High Automation)

Характеристика. На данном уровне автомобиль способен выполнять все основные функции управления в определённых сценариях (так называемых ODD — Operational Design Domain), без участия водителя. Система может продолжать движение даже в большинстве нештатных ситуаций, однако может существовать ряд внешних ограничений, например, определённые погодные условия, отсутствие некоторых дорожных знаков или разметки, нестандартная инфраструктура.

Примеры сценариев использования.

- *Роботакси* (Robotaxi) в пределах заранее определённой зоны городской среды или кампуса, где система чётко знает все маршруты, имеет подробные карты местности и может обходить большинство сложных ситуаций за счёт прогнозирования и планирования траектории.
- *Автоматизированные грузовые перевозки* (Autonomous Trucks), когда движение осуществляется по автомагистралям между ограниченным количеством логистических центров, а водитель участвует только при въезде в городскую среду или в критических случаях.

Ограничения и вызовы. Хотя автомобиль на Уровне 4 может функционировать без участия человека во многих ситуациях, существуют обстоятельства, при которых система может быть не в состоянии безопасно продолжать движение (например, при резком ухудшении погодных условий: сильный туман, снегопад, гололёд). В таких случаях система должна либо безопасно остановить автомобиль, либо запросить вовлечение водителя (если он есть).

Уровень 5: Полная автоматизация (Full Automation)

Характеристика. Наивысший уровень автоматизации предполагает, что автомобиль способен самостоятельно выполнять абсолютно все функции управления в любых условиях, доступных человеку. При Уровне 5 роль водителя, по сути, исчезает: человек может выступать только как пассажир, без необходимости когда-либо брать на себя управление.

Перспективы и сложность.

- Для достижения Уровня 5 необходимы прорывные решения в области искусственного интеллекта, сенсорики и взаимодействия с внешней инфраструктурой (Car-to-Car, Car-to-Infrastructure).
- Требуется существенная доработка законодательной базы, этических норм и стандартов безопасности, поскольку при Уровне 5 возникает

множество новых вопросов ответственности, лицензирования и сертификации.

Текущее состояние разработок. Несмотря на бурное развитие автопилотов, полной сертификации транспортных средств на Уровень 5 в массовом сегменте пока не существует. Отдельные проекты (*Waymo, Cruise, Baidu Apollo*) демонстрируют высокую степень автономности, но их деятельность зачастую ограничена тестовыми зонами с заранее известной инфраструктурой.

Примечание по целевому уровню автоматизации в диссертации

В рамках данной диссертационной работы, где основное внимание уделяется применению нейросетевых методов для классификации акустических данных дорожных событий, целевым является **Уровень 3** (условная автоматизация). Данный выбор обусловлен следующими факторами:

- **Практическая применимость.** На Уровне 3 возможен более широкий спектр реальных сценариев, чем на уровнях 0–2, поскольку автомобиль уже может автономно вести себя в определённых условиях. При этом система всё ещё обращается к водителю в сложных ситуациях, что делает решение технически и юридически более реалистичным на современном этапе развития отрасли.
- **Задачи классификации дорожных событий.** Для работы систем Уровня 3 необходимо эффективно и быстро распознавать объекты (другие автомобили, пешеходов, мотоциклистов и т. д.) и анализировать сложные дорожные ситуации (перестройка, авария, дорожные работы, погодные условия). Нейросетевые методы, обладающие высокой точностью распознавания и прогнозирования траекторий, являются ключевым инструментом для реализации данной функциональности.
- **Безопасность и реакция в критических ситуациях.** Системы Уровня 3 предполагают критически важные механизмы взаимодействия между человеком и машиной (HMI, Human-Machine Interface). Водитель может находиться в состоянии ограниченного внимания, поэтому нужны надёжные алгоритмы детектирования опасных событий и своевременного оповещения водителя о необходимости немедленного вмешательства.

Интерес к использованию акустических данных для анализа состояния автотранспортных средств и дорожных событий обусловлен стремлением повысить эффективность систем принятия решений при управлении транспортом.

Системы помощи водителю транспортных средств, как правило, используют визуальные данные, получаемые с камер и лидаров [2]. Естественно, что такие системы могут сталкиваться с ограничениями при неблагоприятных погодных условиях, при наличии препятствий и различных эффектов, затрудняющих обзор [3; 4]. Акустические сенсоры предоставляют дополнительную информацию, способствующую более полному и объективному восприятию окружающей среды.

Использование акустических данных для анализа дорожных событий позволяет выявлять акустические сигналы, исходящие от различных объектов и событий на дороге (приближающиеся транспортные средства, сирены экстренных служб, акустические сигналы аварийных ситуаций и другие шумы), которые могут свидетельствовать о потенциальной опасности. Интеграция акустической информации с визуальными данными в комплексную информационную систему способствует созданию более надёжных и эффективных систем мониторинга и управления дорожным движением.

Существует множество исследований, посвящённых классификации акустических данных в контексте дорожных событий. Например, в работе [5] исследованы возможности классификации транспортных средств на основе акустических сигналов, предлагаются методы повышения точности классификации. В статье [6] обсуждаются методы классификации акустических сцен, включая использование спектральных признаков и машинного обучения для распознавания различных типов окружающей среды.

Однако анализ акустических данных имеет свои специфические особенности, связанные с ограничением на длительность периода обработки, необходимостью предварительной подготовки и созданием массива сценариев для принятия решений, большой размерностью. Часто требуется применение статистических методов обработки. Также следует учитывать возможность изменения и появления новых акустических данных, связанных с развитием транспортных средств, изменением правил движения, конфигурации и состояния дорог. При этом правила классификации акустических данных и выработки управляющих решений плохо формализуются.

Для обработки таких данных применение традиционных вычислительных методов, использующих высокопроизводительные компьютеры с классической архитектурой и языки программирования, не всегда позволяет получить желае-

мые результаты, так как перечисленные особенности приводят к необходимости постоянно менять правила вычислений (программы).

Более перспективным видится подход, основанный на применении методов, использующих обучение, для проведения анализа и выработки решающих правил. Реализация данного подхода достаточно хорошо отработана с применением нейросетевых технологий и широко применяется рядом авторов.

Так, в работе [7] демонстрируется эффективность нейронных сетей для классификации акустических сцен. Методы безнадзорного обучения для распознавания транспортных средств на основе акустических подписей описаны в статье [8]. Статья [9] содержит результаты применения способов снижения размерности данных для повышения эффективности обработки. Актуальность темы подтверждается большим количеством исследований в этой области [10—14].

1.2 Обзор существующих методов и алгоритмов классификации акустических сигналов окружающей среды

Процесс классификации акустических сигналов играет ключевую роль в анализе акустической информации, поскольку позволяет определить и отнести акустические события к конкретным категориям или классам на основе заранее определенных критериев. Классификация акустических сигналов способствует не только структурированию и систематизации аудиоданных, но и снижает сложность дальнейшей обработки, облегчая интерпретацию информации и принятие решений на её основе. Кроме того, классификация акустических сигналов является важным этапом в решении ряда практических задач, таких как распознавание событий, обнаружение аномалий, улучшение качества записи и выявление источников акустического сигнала.

Существует широкий спектр методов для классификации акустических сигналов, что позволяет учесть многообразие аспектов акустического анализа [15]. Эти методы включают исследование акустических характеристик сигнала (например, громкость, высоту тона и тембр), временные и частотные параметры, а также контекстное окружение акустической сцены, что особенно важно при анализе сложных и многослойных акустических сред [16]. Современные

методы классификации часто используют представления сигнала в различных формах, таких как временные ряды, спектрограммы, мел-спектрограммы и другие производные признаки [17].

Классические (или «традиционные») методы классификации акустических сигналовых сигналов основаны на цифровой обработке (DSP), простых статистических критериях, пороговых алгоритмах и шаблонном сопоставлении. В таких подходах не требуется обучение на больших выборках, поскольку используются «ручные» (hand-crafted) признаки и эвристические правила. Ниже кратко описаны наиболее распространённые методы и приведены примеры их применения.

Пороговое детектирование (Threshold-based Detection). Данный метод основывается на сравнении одного или нескольких параметров сигнала (энергия, амплитуда) с заранее заданным порогом. Например, если энергетическая огибающая превышает порог, фиксируется наличие сигнала (или его переход в иной класс). Такой подход широко применяется для детектирования начала и окончания звучания (Voice Activity Detection) [18] и классификации коротких импульсных акустических сигналов (хлопок, удар). **Спектральные признаки и фильтрация.** Спектральный анализ (FFT, STFT) позволяет получить распределение энергии по частотам. Сигнал можно пропустить через набор полосовых фильтров, оценить энергию в каждой полосе и на её основе принять решение. Например, преобладание высоких частот (выше 2–3 кГц) может указывать на «свист» или «сигнал тревоги»; низкочастотный пик — на «басовый» или «гулкий» акустический сигнал. Подобный подход применяется в простых системах безопасности [19]. **Dynamic Time Warping (DTW).** Алгоритм динамического выравнивания по времени применяется для сопоставления временных последовательностей (например, речевых фрагментов) без обучения больших статистических моделей. DTW позволяет «растягивать» или «сжимать» сигнал по времени, чтобы найти оптимальное соответствие с эталоном. Метод популярен в ранних системах распознавания изолированных слов [20]. Классические методы (пороговые, корреляционные, шаблонные и эвристические) заложили основу для многих исторических систем классификации акустических сигналов и до сих пор используются в узкоспециализированных приложениях. Они не требуют больших обучающих выборок и сравнительно легко реализуются. Однако их возможности ограничены статической природой

«ручных» признаков и зависимостью от порогов, поэтому для более сложных и изменчивых сигналов предпочтительнее применять обучаемые алгоритмы.

Подходы к классификации акустических сигналов сегодня в значительной степени основаны на применении методов машинного обучения, которые предоставляют возможности для точной классификации даже сложных акустических событий. Основная идея этих подходов заключается в том, чтобы обучить модель распознавать определённые закономерности и характеристики, присущие каждому классу акустических сигналов, используя обширные и разнообразные наборы данных. Это позволяет моделям учитывать различные аспекты акустических данных, включая их спектральные, временные и амплитудные характеристики, что критически важно для достижения высокой точности классификации.

Традиционные методы машинного обучения, такие как решающие деревья, метод опорных векторов, случайные леса и градиентный бустинг, представляют собой фундаментальный подход к анализу и классификации акустических данных. Эти алгоритмы работают с заранее извлечёнными признаками, такими как мел-кепстральные коэффициенты (MFCC), спектральная плотность мощности и статистические характеристики сигнала. Такие методы эффективны для базового анализа акустических сигналов и могут быть успешно применены для решения ограниченного круга задач классификации, например, разделения аудиосигналов на простые категории или обнаружения специфических шумов.

Однако их использование имеет свои ограничения. Традиционные алгоритмы требуют тщательной и часто трудоёмкой процедуры извлечения признаков, которая зависит от экспертизы разработчика системы. Кроме того, они менее эффективно справляются с задачами, связанными с классификацией сложных или слаборазличимых акустических сигналов, особенно в условиях высокого уровня шума или при наличии множества наложенных друг на друга источников акустического сигнала. В таких сценариях традиционные подходы нередко уступают более современным методам, основанным на глубоких нейронных сетях.

В рамках машинного обучения также широко применяются глубокие нейронные сети, которые обладают значительными преимуществами в обработке и классификации акустических сигналов, особенно сложных и многокомпонентных. Существует несколько основных архитектур нейронных сетей, таких как

сверточные нейронные сети (CNN) [21], рекуррентные нейронные сети (RNN) [22], капсульные сети [23] и трансформеры [24], каждая из которых имеет свои уникальные особенности и преимущества в обработке акустических данных. Сверточные сети, например, оптимально подходят для работы с двухмерными представлениями акустического сигнала, такими как спектрограммы, и успешно используются для извлечения сложных пространственно-временных признаков. Рекуррентные сети применяются для анализа временной последовательности событий и помогают распознавать акустические сигналы, структура которых изменяется во времени. Капсульные сети, в свою очередь, направлены на распознавание сложных паттернов и взаимосвязей между элементами акустической сцены.

Трансформеры, изначально разработанные для обработки текстовых данных, зарекомендовали себя как универсальный инструмент, который может быть эффективно адаптирован для решения задач в различных областях, включая обработку аудиоданных. Их способность учитывать глобальные взаимосвязи между элементами входных данных делает их незаменимыми в задачах, связанных с анализом временных последовательностей. В аудиосфере трансформеры нашли широкое применение в решении таких задач, как классификация акустических сигналов, сегментация аудиозаписей, идентификация источников акустического сигнала и генерация новых аудиоданных.

Одним из ключевых преимуществ трансформеров является использование механизма внимания (attention), который позволяет моделям концентрироваться на наиболее важных элементах входной последовательности. Это особенно полезно при анализе сложных аудиосигналов, где значимые акустические события могут быть распределены по всей длине записи. Например, в случае обработки музыкальных композиций трансформеры могут выявлять ключевые паттерны, связанные с мелодией, ритмом или гармонией, а также учитывать взаимодействия между различными музыкальными инструментами.

Трансформеры также демонстрируют свою эффективность при работе с аудиозаписями, содержащими наложенные акустических сигналовые слои. В таких ситуациях традиционные методы могут сталкиваться с трудностями в выделении отдельных компонентов акустического сигнала. Благодаря способности учитывать контекст на глобальном уровне, трансформеры способны анализировать аудиосигналы в их полной целостности, идентифицируя как основные, так и второстепенные слои акустического сигнала. Это открывает возможности для

их использования в таких приложениях, как распознавание голоса в многолюдных местах, анализ сложных шумовых сред или извлечение отдельных голосов из хоровой записи.

Использование этих разнообразных методов машинного обучения и глубоких нейронных сетей позволяет улучшить точность и надежность классификации акустических сигналов. В результате, каждый из этих подходов, или их комбинация, может быть применен для достижения наиболее полного анализа и классификации акустических данных[25—30]. На рисунке 1.1 представлена иерархическая схема методов классификации акустических сигналов, иллюстрирующая многообразие существующих подходов и их взаимосвязь.

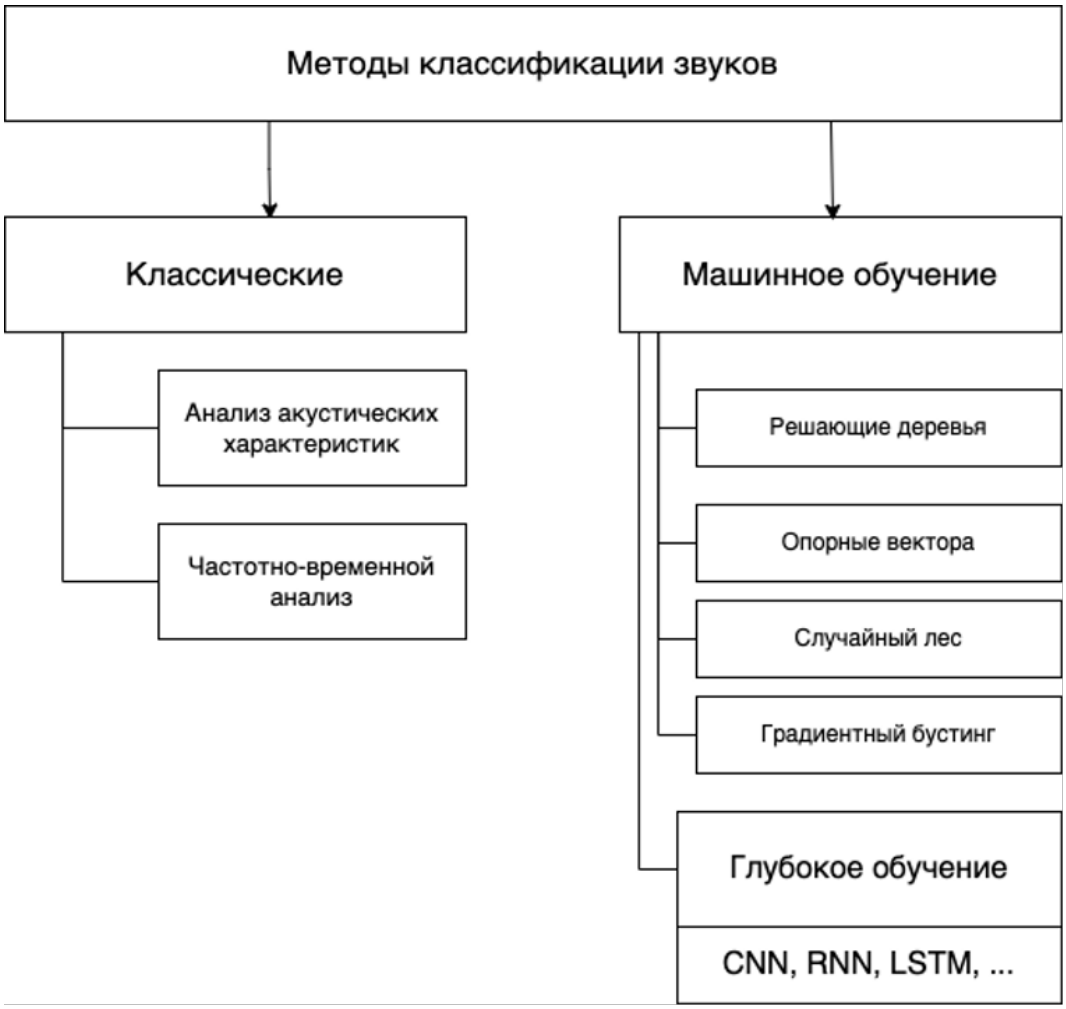


Рисунок 1.1 — Существующие подходы для классификации акустических сигналов.

Представление акустического сигнала для обработки и анализа может быть выполнено с использованием множества различных методов. акустических сигналовые сигналы можно представить в виде временно-амплитудной зависимости, а также совокупности синусоид разных частот. На рисунке 1.2

представлено амплитудно-временное представление акустического сигнала сирены.

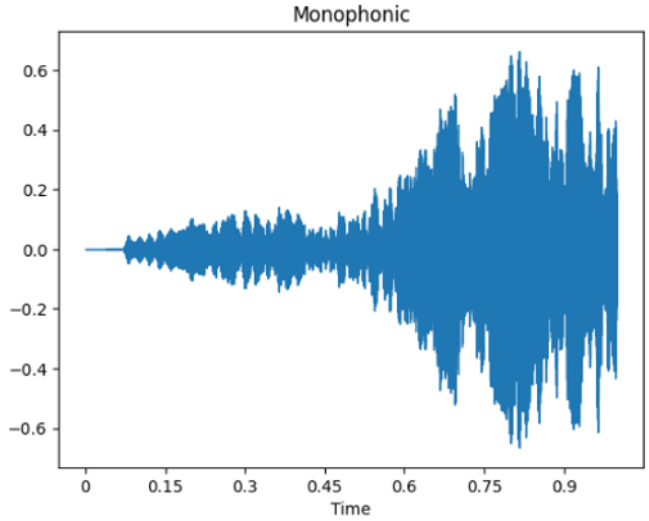


Рисунок 1.2 — Амплитудно-временное представление акустического сигнала сирены.

Спектральное представление использует преобразование Фурье для перехода от временного представления к частотному. Спектрограмма может быть создана путем вычисления преобразования Фурье для небольших временных участков сигнала, получаемых с помощью оконных функций. Это дает набор срезов спектра частот, каждый из которых соответствует конкретному временному интервалу. Оконное преобразование Фурье можно представить следующим образом:

$$F(t, \omega) = \int_{-\infty}^{\infty} f(\tau) W(\tau - t) e^{-i\omega\tau} \, d\tau \tag{1.1}$$

где $W(\tau - t)$ — некоторая оконная функция. В практических условиях нам чаще всего приходится исследовать акустические сигналы, ограниченные во времени. Например, речь не является статичным сигналом — ее спектр изменяется со временем. В связи с этим при спектральном анализе внимание направлено на отдельные короткие отрезки сигнала. Для анализа таких отрезков цифрового аудиосигнала используется дискретное преобразование Фурье, которое описывается следующим образом:

$$X_k = \sum_{n=0}^{N-1} x(n) e^{-i 2\pi \frac{k}{N} n} \tag{1.2}$$

где X_k — комплексные коэффициенты после преобразования Фурье, $x(n)$ — отсчёты входного дискретного сигнала, N — количество измерений сигнала $x(n)$ в анализируемом промежутке, а k — индекс частоты (от 0 до $N-1$).

Во многих случаях возникает потребность отслеживать изменение спектра сигнала со временем. Такое представление сигнала получило название спектрограммы. Для ее создания используется оконное преобразование Фурье: спектр рассчитывается для последовательных окон сигнала, и каждый из этих спектров формирует столбец на спектрограмме. На рисунке 1.3 представлена спектрограмма акустического сигнала сирены.

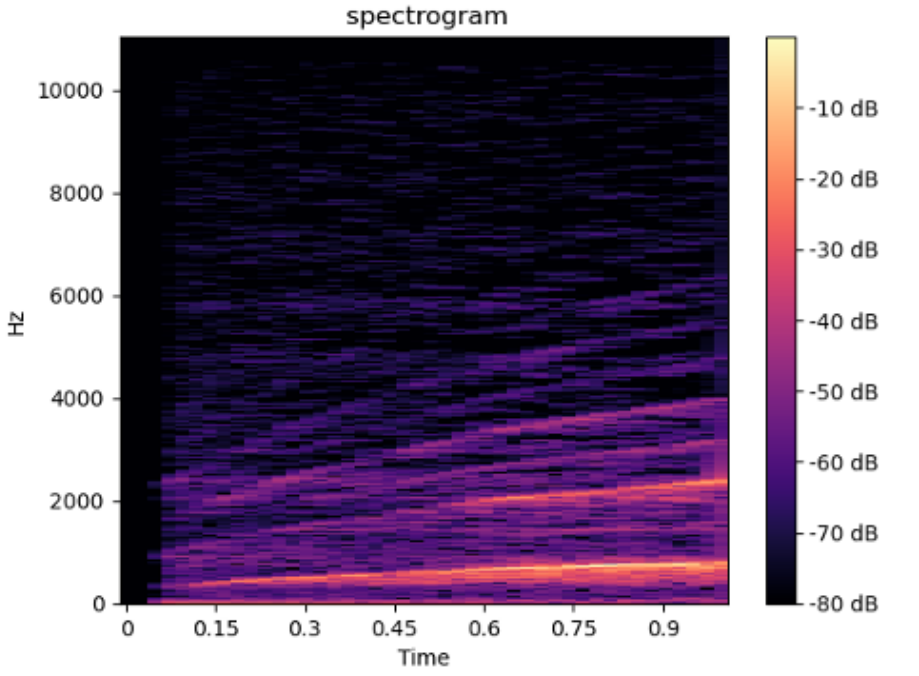


Рисунок 1.3 — Спектрограмма сирены полицейской машины длительностью в 1 секунду.

Дискретное преобразование Фурье упрощает анализ благодаря простоте интерпретации его результатов, но оно обеспечивает лишь обобщенный обзор спектральных характеристик сигнала. Для решения этих проблем используется кратковременное преобразование Фурье.

Одним из ключевых ограничений анализа Фурье является неспособность обеспечить локализацию частоты и времени одновременно. Также оно создает сложное параметрическое представление сигнала, требующее дополнительных шагов обработки. Поэтому проводятся исследования, направленные на применение вейвлет-преобразования и кепстрального анализа.

Кепстральный анализ стал основой для множества алгоритмов, таких как метод мел-частотных кепстральных коэффициентов (MFCC), основанный на мел-шкале:

$$M(f) = 1127.01048 \ln(1 + \frac{f}{700}) \tag{1.3}$$

Мел-спектрограмма производит преобразования над спектром акустических сигналового сигнала, чтобы представить его так, что он лучше соответствует восприятию человека. Процесс вычисления MFCC сводится к выполнению дискретного косинусного преобразования (DCT) на логарифме энергетического спектра, масштабированного на мел-частотную шкалу.

Мел-спектрограмма представляет спектрограмму в мел-шкале, в то время как MFCC являются компактным представлением этих мел-спектрограммы с использованием DCT. На рисунке 1.4 представлена мел - спектрограмма и MFCC.

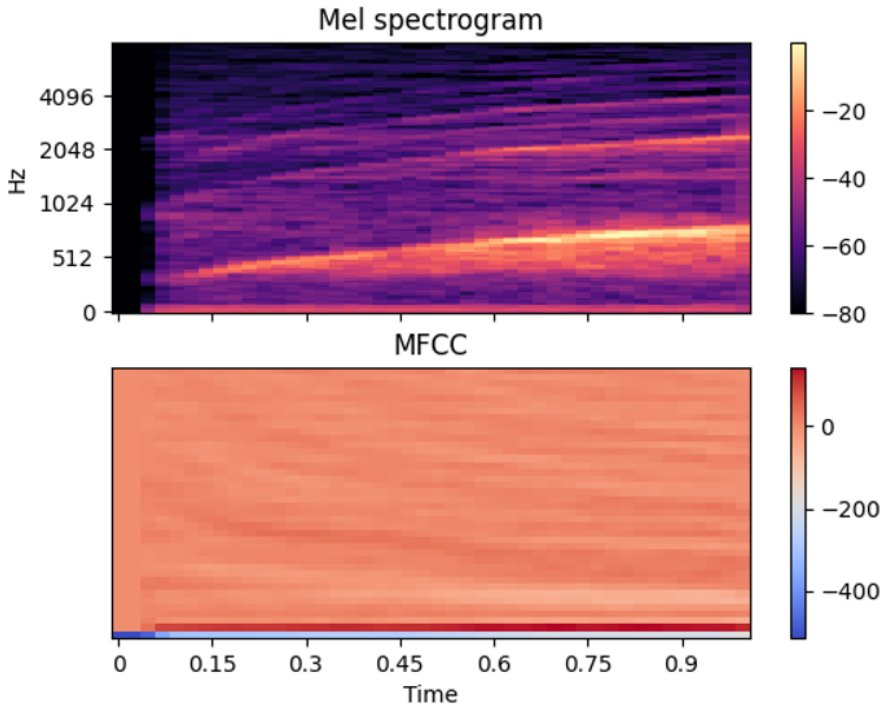


Рисунок 1.4 — Представление мел-спектрограммы и MFCC.

Процесс расчёта MFCC включает несколько этапов. Сигнал разбивается на фреймы (20-40 мс), в течение которых он стабилен. Для уменьшения эффекта растекания используется оконная функция Хемминга:

$$w(n) = 0.53836 - 0.46164 \cos\left(\frac{2\pi n}{N-1}\right) \tag{1.4}$$

К каждому фрейму применяется дискретное преобразование Фурье, после чего рассчитывается периодограмма:

$$P_j(k) = \frac{|X_j(k)|^2}{N} \tag{1.5}$$

Затем вычисляется блок мел-фильтров, который суммирует энергию для каждого фильтра, после чего происходит логарифмическая обработка энергий. Завершающий шаг — применение дискретного косинусного преобразования для получения мел-кепстральных коэффициентов:

$$MFCC(n) = \sum_{m=1}^{M} \log(E_m) \cos\left(\frac{\pi n(m-0.5)}{M}\right) \tag{1.6}$$

Применение дискретного косинусного преобразования (DCT) для MFCC помогает устранить избыточность информации, создавая независимые коэффициенты, которые отражают основные акустические особенности сигнала. Это упрощает дальнейшую обработку данных и повышает эффективность алгоритмов машинного обучения, использующих MFCC для анализа акустических сцен.

1.3 Современное состояние нейросетевых методов классификации акустических сигналов окружающей среды

Значительный прогресс в области искусственного интеллекта и машинного обучения открыл новые возможности для анализа и интерпретации акустических сигналов окружающей среды. В работе рассмотрены различные подходы к обработке и анализу аудиоданных, включая сверточные нейронные сети (CNN), рекуррентные нейронные сети (RNN) и их комбинации. Важной частью исследования станет оценка эффективности этих методов.

End-to-end 1DCNN [31] - метод, основанный на использовании одномерной сверточной нейронной сети для классификации аудио, предполагает применение техники кадрирования аудиосигнала для достижения фиксированной длины аудио. В отличие от большинства подходов к классификации акустических сигналов, каждый кадр аудио передается непосредственно в нейронную

сеть без использования предварительного, необучаемого этапа для сокращения размерности и выделения признаков.

Данная сеть нацелена на изучение набора параметров Θ, которые обеспечивают соответствие между входным вектором X и целевым прогнозом T. Это достигается посредством иерархической обработки признаков, регламентируемой заданным уравнением:

$$T = F(X|) = f_L(...f_2(f_1(X|1)|2)|L) \tag{1.7}$$

где L - количество скрытых слоев в сети. Для сверточных слоев работа i-го слоя может быть выражена следующим образом:

$$T_i = f_i(X_i|i) = h(W \otimes X_i + b), i = [W, b] \tag{1.8}$$

где $\otimes$ обозначает операцию свертки, X_i - двумерная входная матрица из N наборов признаков, W - набор из N одномерных ядер, используемых для извлечения нового набора признаков из входного вектора, b - вектор смещения, а $h()$ - функция активации. Размеры X_i, W, T_i равны (N, d), (N, m) и $(N, d - m + 1)$ соответственно. Несколько объединяющих слоев также применяется между свёрточными слоями для увеличения площади, покрываемой ядрами следующих слоев.

Выходной тензор финального сверточного слоя переходит в одномерный вектор и используются в качестве входных данных нескольких полносвязных слоев, которые можно описать следующим образом:

$$T_L = f_L(X_L|L) = h(WX_L + b), L = [W, b] \tag{1.9}$$

Для многоклассовой классификации количество нейронов выходного слоя равно количеству классов K. После softmax в качестве функции активации для выходного слоя, выходной нейрон i указывает вероятность принадлежности кадра к классу i:

$$\text{Softmax}(t_{L,i}) = \frac{e^{t_{L,i}}}{\sum_{j=1}^{K} e^{t_{L,j}}} \tag{1.10}$$

В процессе обучения параметры сети корректируются в соответствии с обратным распространением ошибки для минимизации функции потерь. Интересной особенностью этой модели, является то, что при инициализации весов первого слоя может использоваться банк фильтров Gammatone. Gammatone фильтр -

это линейный фильтр, описываемый импульсной характеристикой на произведение гамма-распределения и синусоидального сигнала. Одна из конфигураций сети 1DCNN представлена на рисунке 1.5.

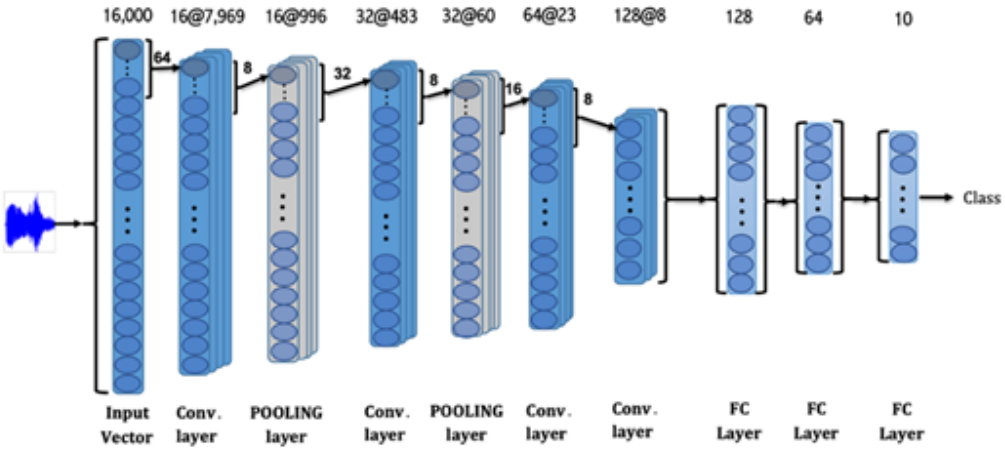


Рисунок 1.5 — Архитектура 1DCNN [31].

Основное преимущество таких моделей заключается в их относительной простоте и небольшом количестве параметров. При обработке аудиозаписей нет необходимости в расчёте мел-спектрограмм или других вычислительно сложных признаков. Эти модели подходят для задач классификации акустических сигналов, различимых по временным шаблонам, но их точность обычно ниже по сравнению с моделями, использующими более сложные архитектуры. Также, успех интеграции результатов классификации во многом зависит от характеристик данных. Если аудиозаписи включают в себя фрагменты, не относящиеся к классифицируемым акустического сигналам, такие как тишина или посторонние шумы, разбиение на кадры может снизить качество набора данных, поскольку эти кадры могут быть ошибочно отнесены к другим классам.

Несмотря на некоторые ограничения, этот метод нашел применение в ряде исследований. Например, в работе [32] используется аналогичная модель с добавленными остаточными соединениями для классификации пола и предсказания роста и возраста человека по его голосу. В исследовании [33] применяется похожая модель в системе для определения видов птиц, которая состоит из непрерывно действующей компактной 1DCNN модели для обнаружения пения птиц и другой модели, требующей вычисления спектрограмм, для классификации самих видов птиц. Этот подход позволяет сократить объем вычислений и, соответственно, энергопотребление системы.

ESResNet [34] (Environmental Sound Residual neural Network) - архитектура остаточной сверточной нейронной сети для классификации акустических сигналов. Предобработка аудиозаписи x для данной сети заключается в приведении к фиксированному размеру t первым подходом и расчётом лог спектрограммы S, с помощью оконного преобразования Фурье:

$$S = 10 \log_{10} |X(\omega, \tau)|^2 \tag{1.11}$$

где

$$X(\omega, \tau) = \sum_{n=-\infty}^{+\infty} x(n) \cdot w(n - \tau) \cdot e^{-j\omega n} \tag{1.12}$$

где w - оконная функция Блэкмана-Харриса:

$$w(k) = a_0 - a_1 \cos\left(\frac{2\pi k}{N}\right) + a_2 \cos\left(\frac{4\pi k}{N}\right) - a_3 \cos\left(\frac{6\pi k}{N}\right) \tag{1.13}$$

с размером окна 37,5 мс и сдвигом 12,7 мс. В результате спектрограмма имеет форму $\left(F \times \frac{t}{0.0127}\right)$, где F - частотное разрешение. В целях переиспользования архитектуры ResNet, спектрограмма разделяется на 3 частотные полосы, которые интерпретируются как каналы изображения, после чего входной тензор имеет форму $(3 \times \frac{F}{3} \times \frac{t}{0.0127})$.

Сама ResNet призвана решить проблему затухающих градиентов путем добавления соединений быстрого доступа между сверточными слоями. Рассмотрим функцию $F(x)$, которая представляет собой последовательность сверточных слоев, примененных к тензору x. Тогда блок с соединением быстрого доступа определяется как функция $H(x)$:

$$H(x) = F(x) + x \tag{1.14}$$

Также в блоках присутствуют слои пакетной нормализации:

$$BN_i(x_i) = \gamma \frac{(x_i - B)}{\sqrt{B^2 + \varepsilon}} \tag{1.15}$$

где B - математическое ожидание пакета, B^2 - дисперсия пакета, γ, β - обучаемые параметры для каждой размерности.

Параллельно некоторым блокам ResNet добавлены блоки внимания, которые состоят из последовательных слоев объединения (по максимуму), сверток и пакетной нормализации.

$$H'(x) = H_2(H_1(H_0(x))) \odot BN'(F_1'(F_0'P(x))) \tag{1.16}$$

где $\odot$ - поэлементное умножение, $P(x)$ - объединяющий слой.Архитектура сети EsResNet представлена на рисунке 1.6

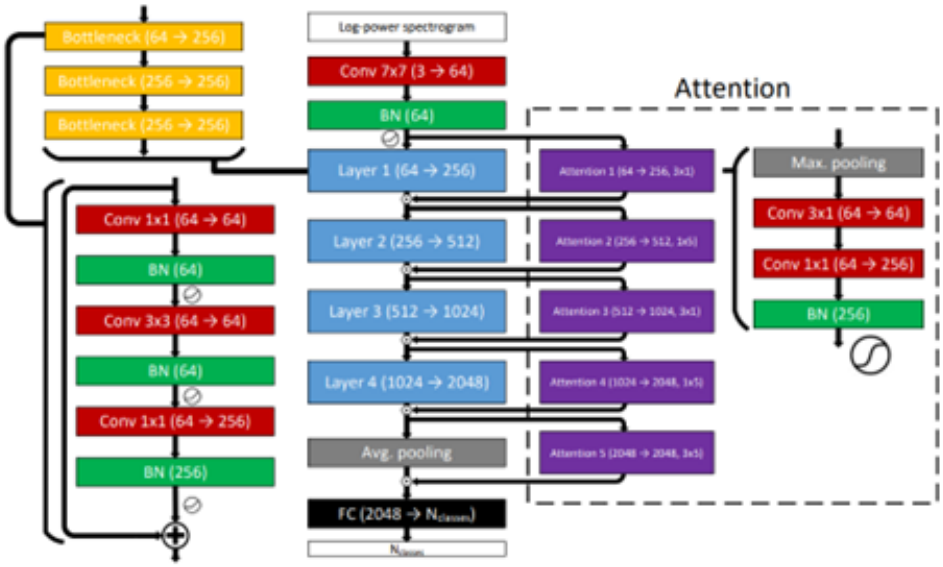


Рисунок 1.6 — Архитектура EsResNet [34].

Благодаря сходству большой части архитектуры с оригинальной ResNet, возможно инициализировать веса совпадающих блоков используя веса предобученной модели ResNet на наборе данных для классификации изображений ImageNet [35]. Этот перенос знаний позволяет повысить точность модели на несколько процентов. Для обработки стерео аудиозаписей сеть до последнего полносвязного слоя применяется на каждом канале, два вектора размером 2048 суммируются и подаются в полносвязный слой. Такая обработка 2-канальной аудио записи, вместо 1-канальной, улучшает метрику качества на наборе данных UrbanSonund8K [36]. ESResNet использовали в ряде исследований. В [37] изменив преобразование Фурье на вейвлет-преобразование и заменив базовую архитектуру ResNet на ResNeXt [38], добились улучшения точности в задаче классификации акустических сигналов городской среды. В [39] используют ESResNet как основу для модели, обученной без учителя. В [40] описывают метод улучшения качества модели путем обучения ESResNet с дополнительным параллельным блоком MLP и гибридной функцией потерь.

AST [41] (Audio Spectrogram Transformer) - архитектура нейронной сети для классификации акустических сигналов, основанная на Vision Transformer (ViT [42]). В качестве входного тензора используется Мел-спектрограмма, для

получения которой применяется окно Хемминга размером 25 мс с шагом 10 мс и 128 Мел фильтров. Соответственно, при длительности аудиозаписи t секунд, форма входного тензора равна $(128 \times 100t)$.

Этот входной тензор делится с помощью двумерного скользящего окна на патчи размером (16×16) с шагом 10 (перекрытием 6 в оба измерения). Следовательно, количество патчей $N = 12 \left\lceil \frac{(100t-16)}{10} \right\rceil$. Каждый патч с помощью линейной проекции преобразуется в одномерный вектор E_d, называемый эмбеддингом патча. Затем, к каждому патчу в соответствии с его позицией добавляется вектор, называемый эмбеддингом позиции [43]. Эмбединги позиций в данной архитектуре являются абсолютными и обучаемыми, таким образом существует матрица $P_{N \times d}$, которая корректируется в процессе обучения, и итоговый эмбеддинг патча с позицией i равен $E'_i = E_i + P_i$.

На практике существует жесткое ограничение на длину аудиозаписей из-за необходимости определения максимального числа патчей во время тренировки сети. Однако, использование билинейной интерполяции вдоль временной оси позволяет адаптировать матрицу P для разного количества патчей. Затем, эмбеддинги патчей вместе с эмбеддингом специального токена отправляются через последовательные слои энкодера трансформера [24]. Вывод, соответствующий этому специальному токену, направляется в классифицирующие слои.

Архитектура AST представлена на рисунке 1.7.

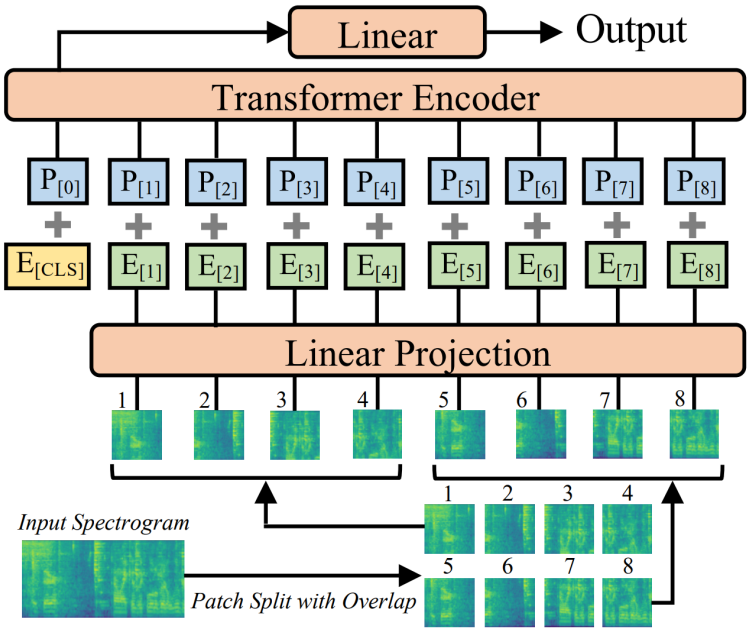


Рисунок 1.7 — Архитектура AST [41].

Применение классической архитектуры энкодеров трансформера, идентичной используемой в некоторых визуальных трансформерах, дает возможность применить технику трансферного обучения. Таким образом, веса аудио-спектрального трансформера (AST) могут быть инициализированы с использованием весов, полученных от обучения визуального трансформера на классификации изображений. Это способствует уменьшению необходимого объема данных для обучения AST. В целом, AST обычно превосходит сверточные нейронные сети по показателям качества, но уступает в производительности из-за квадратичного роста вычислительной сложности и требований к памяти с увеличением числа патчей. AST лежит в основе ряда других моделей [42; 44]. Например, в исследовании [43] блоки внимания в энкодере были заменены на блоки Фурье, что привело к улучшению качества и сокращению размера модели в задачах классификации медицинских акустических сигналов, таких как кашель, насморк и другие.

PaSST [45] (Patchout faSt Spectrogram Transformer) - архитектура нейронной сети для классификации акустических сигналов, основанная на AST. Эмбеддинги позиций разделены на эмбеддинги времени и частоты - каждый патч имеет номер оси времени i, и номер по оси частоты j, итоговый эмбеддинг патча, вместо (10), равен:

$$E'_{i,j} = E_{i,j} + T_i + F_j \tag{1.17}$$

где $T_{i,d}$, $F_{j,d}$ - обучаемые параметры. При увеличении длительности аудиозаписи, меняется только i, а следовательно достаточно интерполировать только матрицу T.

В целях оптимизации количества патчей применяется структурированное выпадение патчей: случайным образом выбирается несколько строк и столбцов, и все патчи, принадлежащие им, не попадают в трансформер. Исключение половины патчей не только увеличивает скорость работы и уменьшает затраты памяти в 4 раза, но и аналогично механизму исключения нейронов [18] улучшает метрики качества.

Также в процессе обучения применяется множество аугментаций аудиоданных:

– Смешивание [46] аудиозаписей и смешивание спектрограмм;
– Маскировка [47] частотных и временных интервалов;
– Смещение аудиосигнала по оси времени;

– Случайное изменение громкости: умножение амплитуды сигнала на -7 до $+7$ дБ.

Архитектура PaSST представлена на рисунке 1.8.

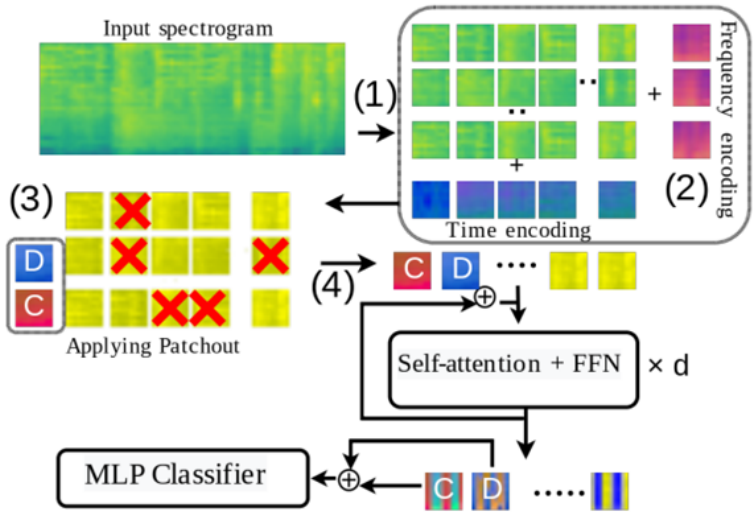


Рисунок 1.8 — Архитектура PaSST [45].

Используя архитектуру PaSST, создана модель для классификации акустической сцены по акустическому сигналу. В [48] PaSST изучают как модель выделения эмбеддингов, на основе которой уже строят классификатор стилей музыки. В [45] подробно сравнивают использование эмбеддингов, генерируемых PaSST на разных уровнях энкодера. С использованием PaSST в качестве учителя созданы модели [49] и [50].

mn_40as [49] - пример модели, использующей подход Transformer-to-CNN [51] в контексте классификации акустических сигналов. Он позволяет добиться точности моделей с большим числом параметров, основанных на AST, используя модели, основанные на двумерных сверточных нейронных сетях. Для получения фиксированной длины аудио, аудиозапись дополняется нулями, выделение признаков состоит в расчёте мел-спектрограмм аудиозаписи. Далее эта спектрограмма рассматривается как изображение и подается в модель с архитектурой MobileNetV3-Large [52].

Для повышения точности малой модели-ученика используется метод дистилляции знаний из модели-учителя архитектуры PaSST. Для этого, помимо обычных потерь при классификации, модель-ученик минимизирует потери при дистилляции, основанные на прогнозах модели-учителя. Итоговую функцию потерь можно сформулировать как взвешенную сумму потерь классификации

L_l и потерь дистилляции L_{kd}:

$$\text{Loss} = L_l((z_S), y) + (1 - \alpha)L_{kd}((z_S), (z_T/\tau)) \tag{1.18}$$

где z_S, z_T - выходные векторы классификатора модели-ученика и модели-учителя, y - действительные метки, σ - функция активации, α - весовой коэффициент и τ - температура обучения.

В данном подходе потери классификации и дистилляции рассчитывали как CrossEntropyLoss:

$$L(p, y) = -\sum_{i=1}^{C} y_i \log(p_i) \tag{1.19}$$

Такой метод позволяет улучшить метрику качества и увеличить производительность, как показано в [49]. Анализ такого подхода показывает, что метод дистилляции знаний демонстрирует высокую эффективность, позволяя сохранить точность сложных моделей при снижении их вычислительной сложности. Этот подход оправдывает дальнейшее его использование и исследование в задачах классификации акустических сигналов.

1.4 Постановка задачи исследования

Акустические данные являются незаменимыми в ряде ситуаций, когда другие методы анализа, такие как визуальные сенсоры или лидары, оказываются недостаточными. Одним из ключевых примеров является распознавание акустических сигналов сирен экстренных служб, таких как полиция, скорая помощь или пожарные машины. В условиях ограниченной видимости (туман, дождь, ночь) или при наличии препятствий (здания, деревья, повороты дороги), камеры и лидары не могут своевременно обнаружить приближающийся автомобиль экстренных служб. Однако акустический сигнал сирены, распространяющийся по воздуху, позволяет системе классифицировать ситуацию задолго до того, как визуальные сенсоры получат информацию.

Аналогично, акустические сигналы, сопровождающие аварийные ситуации, такие как громкие хлопки, треск стекла или скрип тормозов, являются важными индикаторами происшествий. Эти акустические сигналы могут быть

обнаружены даже в тех случаях, когда событие происходит за пределами поля зрения камеры, например, за поворотом или в соседнем ряду. Таким образом, акустическая информация предоставляет уникальные возможности для оперативного анализа дорожной обстановки и принятия решений, невозможные при использовании исключительно визуальных данных.

Исходя из проведенного анализа предметной области, задача классификации акустических событий представляется возможной и актуальной. Для успешной реализации поставленных задач необходимо определить круг ключевых классов акустических сигналов, которые будут использованы в процессе обучения модели. Эти классы должны учитывать как часто встречающиеся акустические события, так и критически важные для анализа дорожной обстановки. Такой подход обеспечит формирование репрезентативного набора данных, отражающего реальную акустическую среду, и позволит адаптировать модель к разнообразным условиям эксплуатации.

Наиболее часто встречающимися и значимыми акустического сигналами в дорожной обстановке являются автомобильные гудки, сирены экстренных служб, акустические сигналы резкого торможения, аварийные шумы и шумы проезжающих транспортных средств. Автомобильные гудки используются водителями для предупреждения и привлечения внимания, их высокая частота обнаружения делает этот класс акустических сигналов важным для систем мониторинга дорожной обстановки. Сирены экстренных служб, таких как полиция, скорая помощь или пожарные машины, имеют среднюю частоту обнаружения, однако их значимость для безопасности движения крайне высока, так как они требуют от водителей оперативного реагирования. акустические сигналы резкого торможения, сопровождающиеся скрипом шин, сигнализируют о возможных опасных ситуациях, таких как аварии или экстренные остановки, и имеют среднюю частоту обнаружения. К аварийным шумам, которые включают акустические сигналы ударов, треска разбитого стекла и повреждений, относятся важные индикаторы ДТП, что делает их критически важными для анализа. Эти классы акустических событий были выбраны из-за их значимости для обеспечения безопасности и мониторинга дорожной обстановки.

Настоящее исследование нацелено на повышение безопасности движения транспортных средств путем разработки и внедрения устойчивых методов классификации акустических данных дорожных событий. Для достижения данной цели сформулированы следующие задачи:

- Исследовать существующие наборы данных применительно к предметной области. На этом этапе предполагается провести обзор обучающих наборов данных соответствующих предметной области. Особое внимание уделяется их сравнению, а также оценке их применимости.
- Спланировать эксперимент сбора, аннотирования и исследования нейросетевых методов классификации акустических данных дорожных событий. На данном этапе определяется методика сбора акустических данных: выбираются места и условия записи, способы снижения ошибок аннотирования, а также разрабатывается план опытов для количественной оценки эффективности различных моделей.
- Разработать систему аннотирования акустических данных дорожных событий и собрать уникальный датасет. Для получения репрезентативной выборки необходимо обеспечить корректное разметку собранных аудиозаписей. Разработанная система аннотирования должна учитывать широкий спектр акустических сигналовых источников (сирены, автомобильные сигналы, шумы двигателей и т.п.) и обеспечивать корректность и воспроизводимость меток.
- Разработать устойчивый алгоритм обучения нейронной сети в условиях выбросов и шумов в обучающем наборе данных за счёт применения устойчивых функций потерь совместно с дистилляцией знаний(knowledge distillation). Предполагается исследовать влияние выбросов и шумов в обучающем наборе, а также внедрить механизмы, повышающие устойчивость модели. К таким механизмам относятся устойчивые функции потерь (способные снижать вклад искаженных данных) и методы дистилляции знаний (позволяющие передавать ключевые признаки от более крупной «учительской» модели к компактной «ученической»).
- Разработать алгоритм классификации акустических данных дорожных событий, позволяющий достигать необходимой точности в рамках предметной области. Итоговое решение должно обладать высокой точностью и адаптируемостью в условиях динамично меняющейся акустической среды. Важно обеспечить низкий уровень ложных тревог и пропусков в реальном времени.
- Разработать комплекс сбора и цифровой обработки акустических данных дорожных событий. Завершающим этапом является создание

интегрированной системы, включающей модули фильтрации, предварительной обработки и классификации акустических данных. Эта система должна быть готова к эксплуатации в реальных условиях, с учетом ограничений по производительности и надежности.

1.5 Выводы по главе

В данной главе полностью выполнена первая задача данного исследования, а именно выполнен обзор методов классификации акустических сигналов окружающей среды с использованием нейросетевых подходов. Рассмотрены существующие подходы к классификации акустических данных, что соответствует поставленной задаче исследования.

Ключевые выводы включают следующие аспекты:

- **Анализ методов классификации акустических сигналов.** Рассмотрены современные подходы к классификации акустических сигналов, включая методы на основе сверточных нейронных сетей (CNN), остаточных сетей (ResNet, ESResNet), а также архитектуры с трансформерами (AST, PaSST). Выявлено, что трансформеры обеспечивают более высокую точность благодаря механизму внимания, учитывающему глобальные взаимосвязи.
- **Представление акустических данных.** Исследованы способы представления акустических сигналов, такие как спектрограммы, мел-спектрограммы и кепстральные коэффициенты (MFCC), которые позволяют извлекать ключевые характеристики сигнала для классификации.
- **Методы дистилляции знаний.** Особое внимание уделено дистилляции знаний, позволяющей снизить вычислительные затраты при сохранении точности. Эффективность метода продемонстрирована с использованием модели-учителя PaSST и модели-ученика MobileNetV3-Large.
- **Сравнительный анализ моделей.** Трансформеры (AST, PaSST) показали более высокую точность по сравнению с моделями на основе CNN, особенно при сложной структуре данных.

Обзор подтверждает перспективность использования трансформеров и методов дистилляции знаний для задач классификации акустического сигнала. На основании сделанных выводов сформулирована постановка задачи исследования, включающая разработку программно-аппаратного комплекса для анализа акустической обстановки. Полученные результаты в данной главе позволяют провести исследование нейросетевых алгоритмов классификации акустических данных дорожных событий во второй главе.

Глава 2. Разработка метода сбора и аннотирования акустических данных о дорожных событиях

В данной главе описывается процесс анализа существующих данных, а также эксперимент, связанный со сбором акустических данных дорожных событий и анализ нейросетевых методов классификации, применённых к этим данным. Отдельно рассматриваются вопросы организации и проведения выездов, включая выбор маршрута и условий записи, что позволило охватить широкий спектр дорожных ситуаций и акустических особенностей [53; 54].

Отдельное внимание уделено разработке протоколов проверки, синхронизации и записи акустической информации, помогающих поддерживать высокое качество и надёжность результирующего набора данных.

Формирование разнообразного набора акустических записей стало ключевым шагом для обучения и тестирования нейронных моделей, рассмотренных в данной главе. Приводится сравнительный анализ нескольких архитектур глубокого обучения, позволяющий определить наиболее эффективные подходы к автоматической классификации дорожных аудиособытий [55]. Такой комплексный экспериментальный подход к сбору данных и их последующей обработке делает описанные методы важным инструментом дальнейших исследований.

2.1 Исследование обучающих наборов данных, постановка эксперимента по сбору акустических данных дорожных событий

Существует множество наборов данных, предназначенных для обучения и оценки моделей классификации акустических сигналов окружающей среды. В исследовательских целях были выбраны 3 открытых, доступных набора данных — ESC-50, UrbanSound8K, FSD50K, которые максимально охватывают те классы которые для предметной области необходимы.

– **ESC-50** [56] (Environmental Sound Classification - 50 class) — Набор данных, состоящий из аудиозаписей продолжительностью 5 секунд, организованных в 50 классов (по 40 файлов в каждом классе), каждый класс принадлежит одной из 5 категории.

– **UrbanSound8K** [57] — представляет собой набор аудиоданных, содержащий 8732 помеченных акустических фрагмента длительностью до 4 секунд акустического сигнала городской среды, разделенных на 10 классов.
– **FSD50K** [58] (Freesound Database 50K) — это открытый набор из 51197 акустических файлов длитель ительностью от 0.3 до 30 секунд, неравномерно распределенных по 200 классам, взятым из набора данных AudioSet [59].

Таблица 1 — Сравнение наборов данных

Наборы данных	ESC-50	UrbanSound8K	FSD50K
Классы	50	10	200
Формат данных	wav, PCM 16bit	wav, PCM 16bit	wav, PCM 16bit
Объем (шт.)	2000	8732	51197
Длительность (файла)	5 с	$\leqslant 4$ с	0.3–30 с
Длительность (общая)	2.7 часа	27 часов	108.3 часа
Разделение	5-секций	10-секций	train/val

Оценка алгоритмов классификации важна для определения их сильных и слабых сторон, выбора моделей для решения определенной задачи и сравнения результатов разных моделей. Ниже приведены некоторые из используемых метрик:

– **Средняя доля правильных ответов** при K секционной кросс-валидации при классификации на N классов:

$$\text{acc}_{K\text{-fold}} = \frac{1}{K}\sum_{j=1}^{K}\sum_{i=1}^{N}\frac{y_i = f_j(x_i)}{N} \tag{2.1}$$

– **Средняя точность** (mAP) — это широко используемый показатель для оценки производительности моделей классификаторов:

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N}(\text{AP}_i) \tag{2.2}$$

где N — количество классов объектов на изображении, AP_i — Average Precision для класса i.

Столь низкая доля правильных ответов 1DCNN на наборе данных ESC-50 по сравнению с UrbanSound8K, может быть объяснена тем, что в наборе ESC-50

Таблица 2 — Оценки перечисленных моделей на наборах данных UrbanSound8K, ESC-50, FSD50K

Модель	UrbanSound8K, $acc_{10\text{-fold}}$	ESC-50, $acc_{5\text{-fold}}$	FSD50K, mAP
1DCNN	0.89	0.35	-
EsResNet	0.854	0.915	-
AST	-	0.957	-
PaSST	-	0.968	0.6555
mn_40as	-	0.9745	0.656

аудиозаписи не состоят целиком из акустических событий, в связи с чем подход 1DCNN при делении на кадры получает много кадров, не содержащих акустических событий, но размеченные как таковые. Также, исходя из представленного авторами статьи [31] исходного кода, видно, что при обучении и проверке этого метода на наборе данных UrbanSound8K секции были сформированы случайным образом, что могло привести к утечке данных из тестового набора в обучающий. В совокупности с фактором малого числа аудиозаписей на один класс в наборе ESC-50, метод 1DCNN демонстрирует отстающие результаты по сравнению с другими подходами.

Рассмотренные открытые наборы данных обладают рядом существенных ограничений: от несбалансированности и малого количества примеров в каждом классе до неоднородности содержимого аудиозаписей. Это приводит к заметным погрешностям при обучении моделей и снижает их надежность в реальных условиях эксплуатации. В результате оптимальным выходом из ситуации стало формирование собственного набора данных, наиболее полно отражающего необходимые классы акустических сигналов и позволяющего обеспечить контроль качества на всех этапах сбора.

Подробное описание архитектуры и реализации системы сбора акустических данных, применённой в данном эксперименте, приведено в разделе 4.1. Настоящий раздел посвящён планированию полевых выездов, маршрутов и методике сбора, а также предварительной обработке полученных аудиозаписей.

Для решения задач исследования потребовалась тщательная организация эксперимента, включающего планирование маршрутов и временных промежутков, разработку протоколов проверок и контроль качества собираемых данных. Данный эксперимент проводился в реальных условиях дорожного движения,

чтобы максимально приблизить акустическую обстановку к типичным городским ситуациям.

Планирование временных промежутков. Исходя из анализа трафика мегаполиса, был составлен график выездов, охватывающий три основных периода дня: утренний (8:00–10:00), дневной (13:00–15:00) и вечерний (17:00–19:00). Такой выбор обусловлен изменением интенсивности движения и характерных акустических событий в разное время суток [60; 61]. Каждый выезд длился два часа, что позволило собрать статистически репрезентативные данные и учесть влияние погодных и сезонных факторов в течение трёх месяцев эксперимента.

Организация маршрутов. Для проведения сбора акустических данных был разработан специальный маршрут, охватывающий различные районы города Москвы. Маршрут был спланирован таким образом, чтобы обеспечить максимальное разнообразие дорожных условий, характерных для крупного мегаполиса. Ключевыми точками маршрута стали пересечения крупных улиц и проспектов, представленные в таблице 3. Эти локации выбраны как наиболее репрезентативные с точки зрения интенсивности дорожного движения, сложности организации транспортных потоков и широкого разнообразия акустических событий.

Таблица 3 — Пересечения улиц по маршруту сбора данных.

№	Пересечение улиц
1	Авиамоторная улица
2	Солдатская улица
3	Госпитальная набережная
4	Гольяновская улица
5	Проспект Ветеранов
6	Ростокинский проезд
7	Рижская эстакада
8	Садовая-Черногрязская улица
9	Площадь Крестьянская Застава
10	Шоссе Энтузиастов
11	Улица Лапина

Для наглядного представления маршрута сбора акустических данных в Москве на рисунке 2.1 приведена карта с указанием ключевых точек — пересечений улиц, обозначенных в таблице 3.

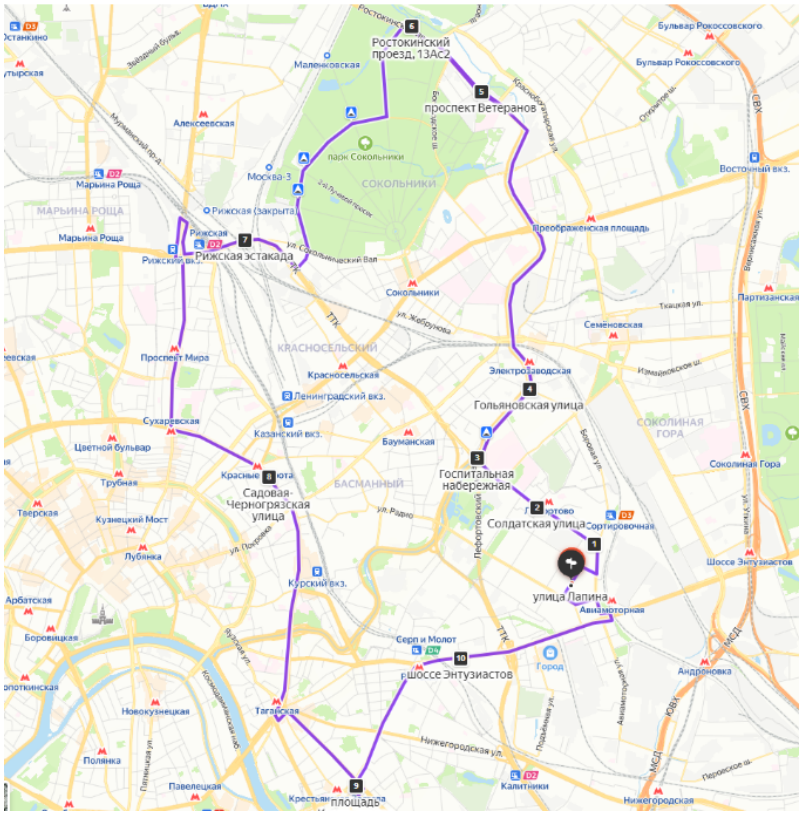


Рисунок 2.1 — Карта маршрута сбора акустических данных в Москве с указанием ключевых точек.

Контроль качества и чек-листы. Перед каждым выездом проводился регламентный осмотр всех компонентов системы: **микрофонного массива**, камеры и программного обеспечения. Была разработана подробная пошаговая инструкция (чек-лист), включающая:

- Проверку надёжности крепления микрофонного массива и корректности подключения аудиокабелей;
- Тестовую запись короткого аудио- и видеофайла для оценки уровня сигнала и работоспособности оборудования;
- Фиксацию параметров погоды и дорожной обстановки (температура, осадки, интенсивность трафика) для последующего анализа;
- Контроль заполнения базы данных после каждого выезда (записи и метаданные должны корректно выгружаться в хранилище).

Такой подход минимизировал риск пропуска акустических событий и позволял своевременно обнаруживать и устранять возможные сбои оборудования.

Сбор акустических данных. В процессе выездов непрерывно велась запись аудио и видеопотоков на протяжении всего маршрута. Запись производилась с частотой дискретизации 44,1 кГц в монофоническом режиме, с использованием кодека Ogg Opus при битрейте 192 кбит/с. Исследования показывают, что Opus демонстрирует хорошие результаты по параметрам сжатия и качества [62]. На рисунке 2.2 представлен график искажений при использовании различных аудиокодеков.

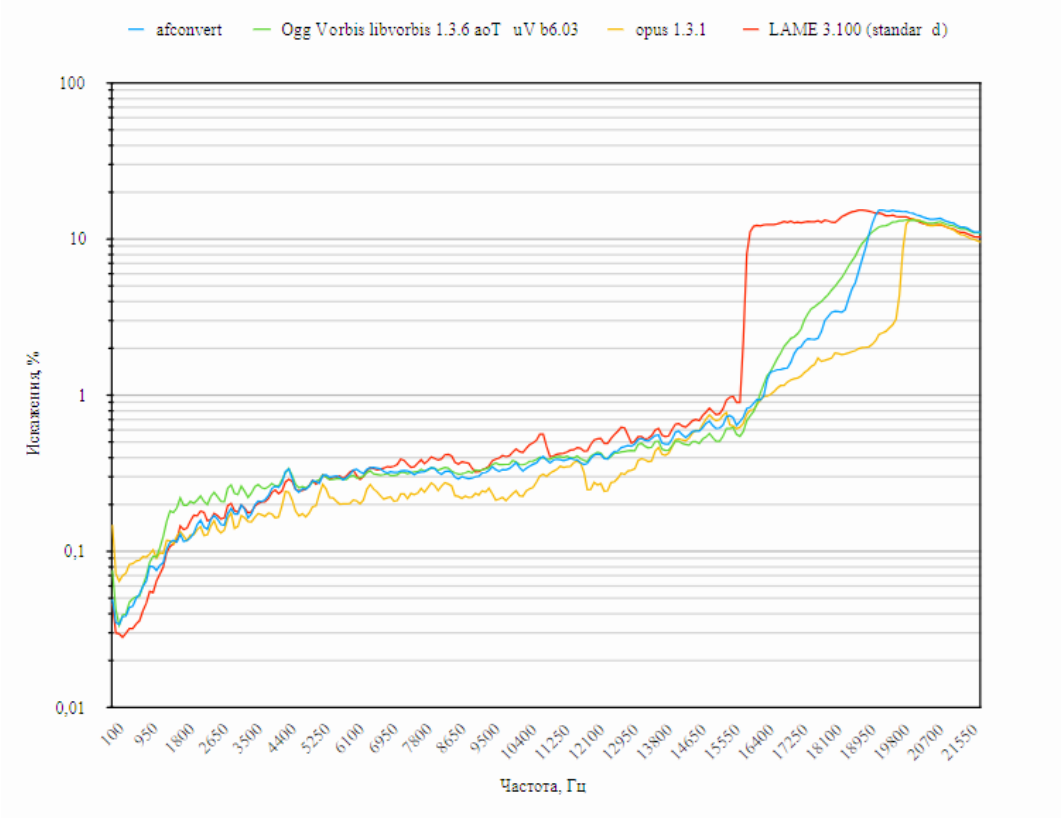


Рисунок 2.2 — График искажений при применении разных кодеков.

Использование Ogg Opus позволило существенно сократить размер хранимых данных, сохраняя при этом высокое качество записей, что критично при больших объёмах акустических данных.

Предобработка и выгрузка данных. По окончании каждого выезда все собранные акустические данные выгружались в специально созданную базу. Каждая запись сопровождалась метаданными, включая дату, время, погодные условия и объекты на маршруте. Это упрощало поиск и отбор нужных фрагментов при последующем анализе, а также позволяло учитывать влияние разных факторов на акустическую обстановку [55].

Объём собранной информации. Благодаря системному планированию эксперимента и регулярным выездам удалось собрать более 300 часов аудиозапи-

Рисунок 2.3 — Процесс сбора и первичной проверки данных.

сей, охватывающих широкий спектр дорожных ситуаций (пробки, перекрёстки, пешеходные переходы, магистрали), а также различные погодные условия (ясно, дождь, снег, переменная облачность).

2.2 Система сбора и аннотирования акустических данных

Процесс сбора и аннотирования акустических данных является сложной задачей, обусловленной рядом технических и организационных факторов. Одной из ключевых проблем является высокий объем необработанных аудиоданных, требующий значительных временных и вычислительных ресурсов для разметки. Ручная аннотация, несмотря на высокую точность при небольших объемах данных, становится нецелесообразной при масштабировании, так как сопряжена с большими временными затратами, субъективностью оценок и увеличением количества ошибок вследствие монотонности работы оператора.

Дополнительные сложности возникают при обеспечении согласованности разметки между разными аннотаторами, так как интерпретация акустических событий может варьироваться в зависимости от индивидуального опыта специалиста. Это создает необходимость в методах стандартизации аннотации и автоматизации процесса, обеспечивающих воспроизводимость результатов.

Для решения данной проблемы была разработана система автоматизированной разметки, основанная на использовании предобученной модели BEATs в сочетании с инструментом LabelTool

В ходе исследования был собран обширный массив акустических данных, суммарный объём которого превысил 300 часов аудиозаписей. Первичный анализ показал, что ручная обработка и разметка такого объёма информации являются крайне неэффективными по нескольким ключевым причинам:

– **Высокие временные затраты.** Разметка одного часа аудиозаписи в среднем требовала 4,7 часа (±0,5 часа) работы специалиста. Обработка всего датасета заняла бы порядка 1410 человеко-часов, что делало данный подход неприемлемым в рамках диссертационной работы.
– **Рост количества ошибок.** Длительная монотонная работа приводила к увеличению числа разметочных ошибок, что снижало точность и достоверность итоговых данных.
– **Отсутствие эффективного механизма контроля качества.** Проверка и верификация вручную размеченных данных требовала дополнительных ресурсов, сравнимых с первичной разметкой, что существенно снижало общую эффективность процесса.

С учётом указанных ограничений было принято решение об автоматизации процесса аннотации данных. Для этого использовалась комбинация специализированного инструмента **LabelTool** и предобученной модели **BEATs**, адаптированной к специфике задачи посредством дообучения на собранных данных. Такой подход позволил значительно сократить трудозатраты, повысить точность аннотации и обеспечить более высокую воспроизводимость результатов.

LabelTool — это специализированное программное обеспечение, разработанное для ускорения процесса ручной разметки аудиоданных. Инструмент предоставляет удобный интерфейс с возможностями быстрой навигации по аудиофайлам, использования горячих клавиш и функций массового редактирования меток.

В дополнение к LabelTool в процессе разметки применялась предобученная трансформерная модель **BEATs**, предназначенная для обработки аудиосигналов. Модель была частично дообучена на собранных данных, что позволило ей более точно распознавать характерные для исследования акустические события.

Интеграция модели BEATs в процесс разметки осуществлялась следующим образом:

1. **Автоматическая предварительная разметка.** Модель BEATs использовалась для автоматического выделения потенциально значимых фрагментов в аудиоданных и присвоения им соответствующих меток.

2. **Использование в LabelTool.** Результаты автоматической разметки загружались в LabelTool, где разметчики могли просматривать и корректировать предложенные моделью метки в удобном интерфейсе.
3. **Итеративное дообучение модели.** После проверки и корректировки разметки экспертами обновлённые данные использовались для дальнейшего дообучения модели BEATs, что способствовало повышению её точности и адаптации к специфике исследуемых данных.

Комбинированный подход продемонстрировал значительное сокращение времени, затрачиваемого на разметку. Благодаря автоматической предварительной разметке и удобному интерфейсу LabelTool время аннотации одного часа аудиозаписи сократилось до **2,1 часа** ($\pm$0,3 часа), что повысило производительность на **55%** по сравнению с полностью ручной разметкой.

Кроме того, предложенные моделью BEATs метки служили ориентиром для разметчиков, снижая когнитивную нагрузку и вероятность ошибок, обусловленных усталостью и человеческим фактором. Совместное использование BEATs и LabelTool позволило эффективно выявлять сложные акустические паттерны и их взаимосвязи, что существенно обогатило результаты исследования.

Предобученная модель BEATs. В качестве базового решения использовалась версия модели BEATs, обученная на обширном наборе аудиоданных различных типов. Для повышения точности идентификации характерных акустических событий было проведено частичное дообучение модели на собранных данных с применением техник трансферного обучения.

Предобработка данных. Для обеспечения корректной работы модели аудиофайлы проходили предварительную обработку, включающую нормализацию сигнала, преобразование в формат, совместимый с BEATs, а также выделение ключевых фрагментов для повышения точности предсказаний.

Интеграция с LabelTool. Разработан модуль, обеспечивающий связь выходных данных модели BEATs с интерфейсом LabelTool. Это позволило реализовать бесшовное взаимодействие автоматической и ручной разметки, обеспечивая удобство работы разметчиков и минимизацию временных затрат.

Обратная связь для модели. Скорректированные разметчиками данные возвращались в модель BEATs для последующего дообучения, что способствовало её адаптации к специфике задачи и повышало точность разметки на следующих итерациях.

LabelTool представляет собой распределённую систему с микросервисной архитектурой, разработанную для эффективной разметки аудиоданных. Развёртывание осуществляется с использованием технологий контейнеризации Docker (версия 23.0.5) и оркестрации Docker Compose (версия 2.17.3), что обеспечивает гибкость, масштабируемость и удобство управления зависимостями.

База данных. В качестве системы управления базами данных используется PostgreSQL (версия 13). Она предназначена для хранения метаданных разметки, информации о пользователях и статистики выполнения задач. PostgreSQL обеспечивает высокую надёжность, отказоустойчивость и эффективность при работе с большими объёмами данных за счёт поддержки транзакционной целостности (ACID), индексации и оптимизированных механизмов параллельной обработки запросов. На рисунке 2.4 представлена схема базы данных системы.

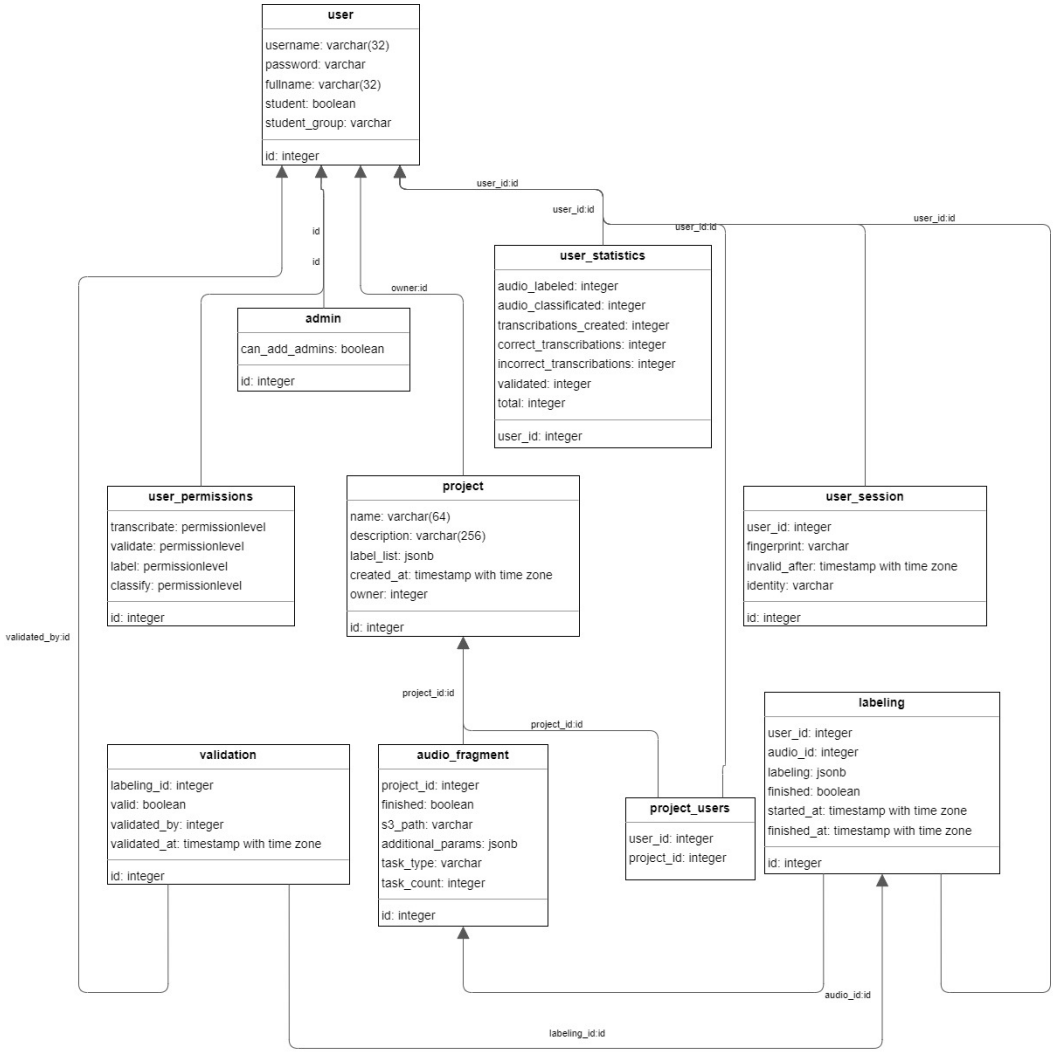


Рисунок 2.4 — Схема базы данных Labeltool.

Серверное приложение. Серверный компонент реализован на языке Python с использованием фреймворка FastAPI. Архитектура основана на REST API, что позволяет организовать взаимодействие между клиентским приложением, базой данных и файловым хранилищем. Выбор FastAPI обусловлен его высокой производительностью, встроенной поддержкой асинхронных операций (ASGI) и автоматической генерацией документации API (OpenAPI). Серверный модуль также реализует механизмы аутентификации, авторизации пользователей и контроля доступа к данным.

Хранилище данных. Для хранения аудиофайлов используется MinIO — S3-совместимое объектное хранилище, обеспечивающее отказоустойчивость и горизонтальную масштабируемость. MinIO позволяет организовать локальное облачное хранилище с поддержкой расширенных механизмов контроля доступа, версионности файлов и распределённого хранения данных.

Веб-сервер. В качестве веб-сервера используется Nginx, который выполняет несколько критически важных функций: обработку и кеширование статических ресурсов, проксирование запросов к API и балансировку нагрузки между сервисами. Благодаря поддержке многопоточной обработки запросов и механизму кэширования, Nginx повышает общую производительность системы и снижает задержки при взаимодействии с клиентским приложением.

Клиентское приложение. Фронтенд-система представлена одностраничным веб-приложением (SPA), разработанным на React.js. Интерфейс оптимизирован для работы с аудиоданными и включает следующие функциональные возможности:

- Визуализация аудиосигнала с использованием Web Audio API, что позволяет разметчикам работать с акустическими событиями в удобном графическом представлении.
- Поддержка горячих клавиш и настраиваемых комбинаций для ускорения процесса аннотации.
- Асинхронное взаимодействие с серверным API для загрузки данных, сохранения результатов разметки и динамического обновления пользовательского интерфейса.
- Адаптивный дизайн, обеспечивающий корректное отображение на различных устройствах.

Взаимодействие между компонентами системы организовано через REST API с передачей данных в формате JSON. Использование асинхронных меха-

низмов на всех уровнях архитектуры позволяет минимизировать задержки при обработке запросов и обеспечить высокую отзывчивость интерфейса. На рисунке 2.5

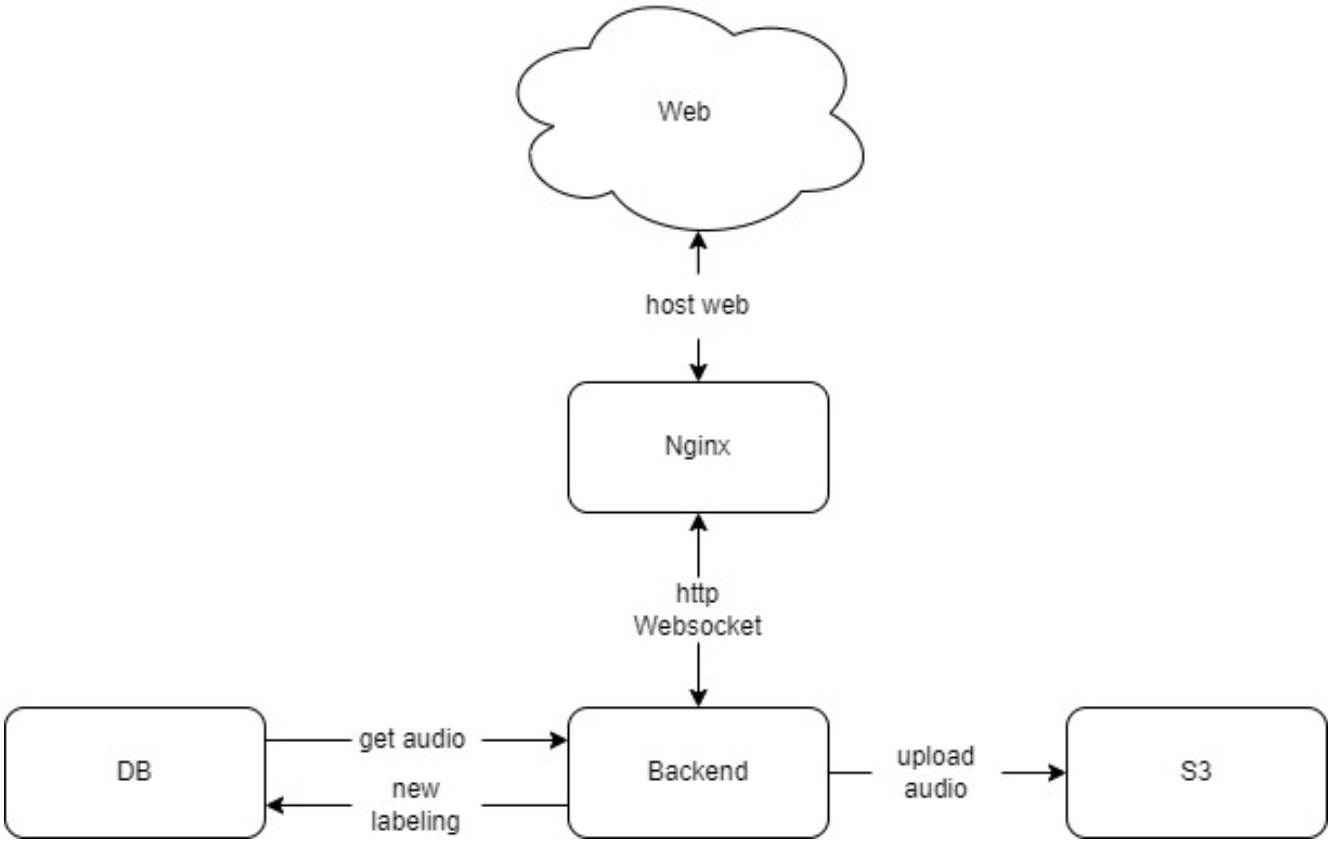


Рисунок 2.5 — Архитектура ПО Labeltool.

Предложенная архитектура LabelTool обеспечивает гибкость развертывания, возможность масштабирования системы при увеличении объёма данных, а также устойчивость к отказам за счёт контейнеризации и распределённой обработки информации.

Ключевые возможности LabelTool

Аутентификация и авторизация. Система безопасности реализована с использованием разграничения прав доступа и ролевой модели. Поддерживаются следующие роли пользователей:

- **Администратор** — управляет пользователями, распределяет задачи, контролирует процесс разметки и анализирует статистику.
- **Разметчик** — выполняет первичную аннотацию аудиофайлов с использованием интерфейса LabelTool.
- **Верификатор** — проверяет и корректирует разметку, обеспечивая контроль качества данных.

Многопользовательская работа. Система поддерживает одновременную работу нескольких пользователей, что позволяет эффективно распределять задачи и повышать производительность процесса аннотации. Распределение задач осуществляется с учётом производительности и специализации каждого пользователя, что способствует равномерной нагрузке и минимизации простоев.

Интерфейс разметки. Веб-интерфейс LabelTool оптимизирован для быстрого и удобного аннотирования аудиоданных. Основные функциональные возможности:

- Поддержка горячих клавиш для ускоренной работы.
- Визуализация аудиосигнала с возможностью масштабирования и детального анализа фрагментов.
- Функции паузы, перемотки и точной отметки времени, позволяющие разметчикам быстро ориентироваться в данных.

Алгоритм распределения задач. Система автоматически назначает задачи пользователям на основе их текущей загруженности, скорости работы и качества ранее выполненных аннотаций. Такой подход позволяет оптимизировать процесс разметки и снизить вероятность перегрузки отдельных пользователей.

Система контроля качества. Реализован механизм перекрёстной проверки разметки, при котором один и тот же фрагмент аудио может быть размечен и проверен несколькими пользователями. Количество проверяющих определяется параметром `MAX_USERS_WITH_FRAGMENT`, который настраивается в зависимости от требований к качеству разметки.

Модуль статистики и аналитики. Встроенные инструменты сбора и анализа данных позволяют отслеживать ключевые метрики:

- Производительность пользователей (количество размеченных файлов, скорость работы).
- Качество разметки (сравнение результатов первичной аннотации и верификации).
- Общая статистика выполнения задач, что позволяет выявлять узкие места и оптимизировать процесс аннотации.

API для интеграции. LabelTool предоставляет открытый API, позволяющий интегрироваться с внешними системами. Например, эндпоинт

`/api/file/upload` используется для автоматизированной загрузки новых аудиофайлов в систему.

Комплексная архитектура и развитые функциональные возможности LabelTool обеспечивают высокую эффективность процесса разметки аудиоданных, автоматизируя ключевые этапы аннотации и контроля качества.

Разработанная система разметки акустических данных на основе LabelTool и модели BEATs продемонстрировала высокую эффективность по сравнению с традиционной ручной аннотацией. Использование автоматической предварительной разметки в сочетании с инструментами корректировки позволило значительно сократить временные затраты, повысить точность аннотации и снизить влияние человеческого фактора на процесс разметки.

Применение микросервисной архитектуры обеспечило гибкость и масштабируемость системы, а контейнеризация с использованием Docker упростила процесс развёртывания и управления компонентами. Взаимодействие между серверным приложением (FastAPI), базой данных (PostgreSQL), объектным хранилищем (MinIO) и клиентским интерфейсом (React.js) позволило создать удобную и отказоустойчивую платформу для разметки аудиофайлов.

Интеграция модели BEATs, дообученной на специфичных данных, повысила точность выделения значимых акустических событий. Реализация механизма итеративного дообучения модели на основе данных, размеченных экспертами, способствовала адаптации алгоритма к особенностям исследуемого домена.

Дополнительно разработанная система контроля качества с перекрёстной проверкой разметки и модуль статистики позволили оптимизировать процесс аннотации, обеспечивая детальный анализ производительности пользователей и точности аннотированных данных.

В результате предложенный метод автоматизированной разметки продемонстрировал увеличение производительности на 55% по сравнению с полностью ручной разметкой, а также улучшение согласованности аннотаций. Полученные данные могут быть использованы для дальнейшего обучения нейросетевых моделей и разработки более точных систем детекции акустических событий.

2.3 Исследование нейросетевых методов в задаче классификации акустических данных дорожных событий

В данном разделе рассматриваются современные подходы и архитектуры нейронных сетей, которые использовались для классификации акустических данных дорожных событий. Каждая из описанных моделей предлагает уникальный подход к обработке и анализу акустических данных, а их результаты были тщательно проанализированы. Рассмотрение архитектурных особенностей, методов предобработки и достигнутых показателей точности позволило определить наиболее подходящие решения для задач классификации акустических сигналов в дорожной обстановке.

Особенности задачи классификации акустических данных

Задачи классификации акустических данных связаны с необходимостью учитывать высокую изменчивость акустических сигналов в реальных условиях. Влияние внешних факторов, таких как шум, наложение акустических сигналов друг на друга, а также различные погодные условия, значительно усложняет обработку и классификацию данных [63]. Современные нейронные сети используют механизмы глубокого обучения, позволяющие выявлять скрытые зависимости в сложных акустических данных.

Для изучения применимости различных подходов были выбраны архитектуры от относительно простых (1DCNN) до сложных (BEATs, EsResNet), что позволяет оценить их возможности и ограничения в задаче классификации акустических сигналов.

Методы предобработки данных

Важным этапом подготовки данных является их предобработка. Для этого применяются такие методы, как:

- Разбиение аудиофайлов на короткие временные фреймы с последующим преобразованием в спектральное представление (логарифмическая спектрограмма или мел-спектрограмма) [64].
- Применение оконных функций для уменьшения эффектов краевых шумов.
- Нормализация сигналов для снижения влияния амплитудных вариаций, вызванных разной громкостью акустических событий.

Эти методы позволяют преобразовать исходные сигналы в формат, пригодный для анализа с использованием современных нейросетевых архитектур.

Архитектурные особенности нейросетей

Используемые модели различаются по сложности и подходам к обработке данных:

- Модели с одномерными сверточными слоями (1DCNN) ориентированы на анализ временных рядов и извлечение локальных зависимостей.
- Полносвязные сети с расширенными признаковыми пространствами (PIPMN) обеспечивают гибкость за счет анализа различных спектральных характеристик.
- Сложные сверточные сети (EsResNet) и трансформеры (BEATs) способны извлекать глубинные признаки, что особенно важно для сложных акустических паттернов [24].

Каждая из архитектур имеет свои сильные стороны, что делает их подходящими для разных аспектов анализа акустических данных.

Модель 1DCNN

Одной из первых рассмотренных моделей стала одномерная сверточная нейронная сеть (1DCNN), которая применялась для анализа временных последовательностей. Данная архитектура используется для извлечения локальных зависимостей в акустических сигналах и хорошо зарекомендовала себя в задачах обработки речи [65].

Обучение модели проводилось на наборе данных, содержащем акустические сигналы дорожной обстановки, такие как шум автомобилей, сигналы сирен и фоновые акустические сигналы окружающей среды. Максимальная точность модели составила 47,89%, что отражает её ограниченные возможности при анализе сложных акустических паттернов. На рисунке 2.6 представлен график обучения.

Такая точность обусловлена ограничениями архитектуры. Модель не включает механизм внимания, что снижает её способность выделять значимые участки сигнала. Кроме того, разбиение аудиофайлов на равные фрагменты привело к созданию большого числа пустых сегментов, которые оказывали негативное влияние на процесс обучения [63]. Эти сегменты, не содержащие полезной информации, вносили шум в выборку и ухудшали результаты классификации.

Для повышения точности планируется использовать следующие улучшения:

– Добавление механизма внимания (attention) для фокусировки на наиболее значимых участках сигнала [24].
– Аугментация данных для создания большего разнообразия обучающей выборки, включая изменения громкости, частоты и временные сдвиги.

Несмотря на свои ограничения, модель 1DCNN продемонстрировала потенциал для обработки акустических данных с низкой вычислительной сложностью. Однако её использование ограничивается задачами, где акустические события имеют относительно простую структуру.

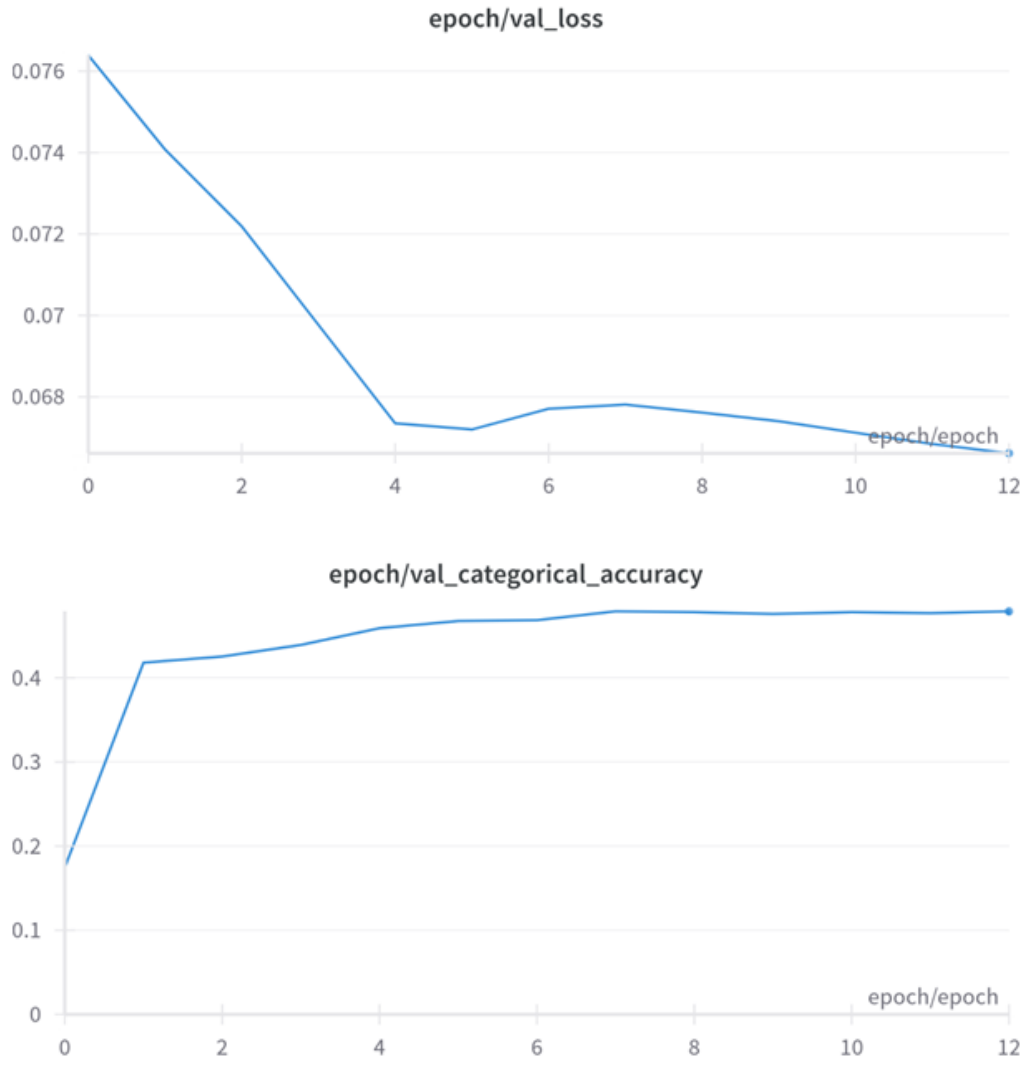


Рисунок 2.6 — Графики обучения модели 1DCNN.

Модель PIPMN

Parallelly Interconnected Perceptron-based Multilayer Network (PIPMN) представляет собой архитектуру, ориентированную на извлечение широкого

набора спектральных признаков. Данная модель позволяет эффективно работать с аудиосигналами, используя многообразие кепстральных коэффициентов [64]. Точность модели составила 84,04%, что значительно выше по сравнению с 1DCNN. На рисунке 2.7 представлен график обучения.

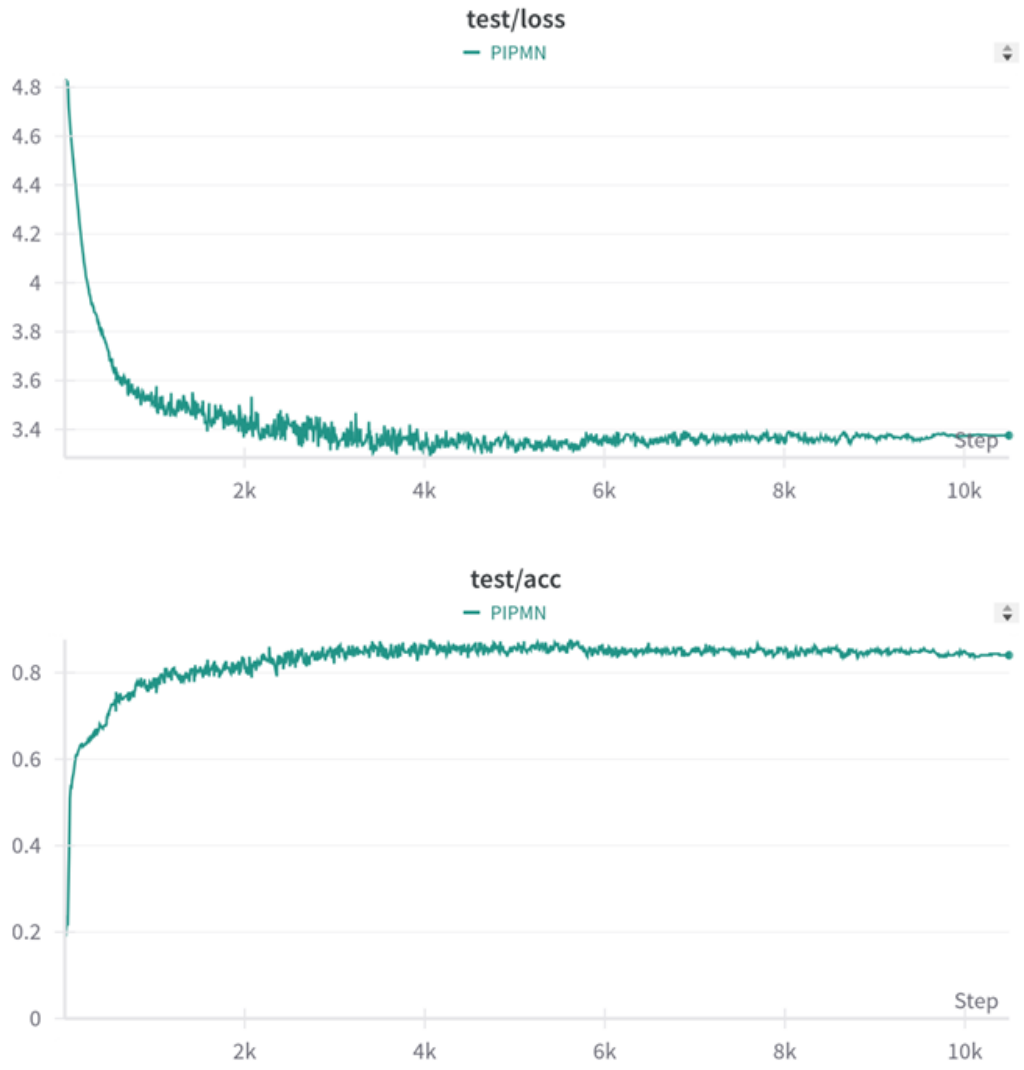


Рисунок 2.7 — Графики обучения модели PIPMN.

Ключевым элементом архитектуры является анализ различных типов коэффициентов:

- **LFCC:** линейно-частотные кепстральные коэффициенты.
- **MFCC:** мел-частотные кепстральные коэффициенты, широко применяемые в обработке речи и акустического сигнала.
- **PNCC:** нормализованные по мощности коэффициенты, устойчивые к шуму.
- **BFCC:** коэффициенты шкалы Барк, имитирующие человеческое восприятие частот.

– **GFCC:** коэффициенты гамматонового фильтра, адаптированные для анализа речевых сигналов [66].

Модель использует полносвязные слои и слои нормализации, что делает её менее ресурсоёмкой. Однако отсутствие сверточных слоёв снижает её способность учитывать локальные пространственно-временные особенности, что ограничивает точность в задачах с более сложной структурой данных.

Модель показала хорошие результаты на данных с относительно низким уровнем шума. Однако в условиях сложной акустической среды ей недостает возможностей, присущих сверточным сетям. Для улучшения её работы рекомендуется исследовать комбинации PIPMN с более сложными архитектурами, включая сверточные сети и трансформеры.

Модель FACE

FACE (Feature Aggregation and Classification Ensemble) — это специально оптимизированная архитектура, разработанная для классификации акустических событий с использованием сверточных нейронных сетей. Модель сочетает в себе использование компактных признаков, таких как мел-кепстральные коэффициенты (MFCC), и дополнительных спектральных характеристик, включая спектральный контраст. Максимальная достигнутая точность модели составила 90,23% [66]. График обучения FACE приведён на рисунке 2.8.

Основные характеристики модели включают:

– **Мел-кепстральные коэффициенты (MFCC):** Широко используемые признаки, которые отражают восприятие частот человеческим ухом, особенно в задачах распознавания речи [64].

– **Спектральный контраст:** Мера разницы между пиковыми и минимальными значениями спектра, полезная для анализа текстурных особенностей акустического сигнала.

– **Одномерные сверточные слои:** Используются для извлечения локальных временных зависимостей в сигнале.

– **Нормализация по батчам:** Увеличивает стабильность обучения, ускоряя сходимость модели.

Одним из ключевых преимуществ FACE является её способность эффективно обрабатывать сложные и разнообразные паттерны данных, благодаря чему она особенно хорошо подходит для анализа акустических событий в шумной, динамичной и изменчивой среде. Однако для её успешного и устойчивого обучения требуется значительный объём данных с высокой степенью точности

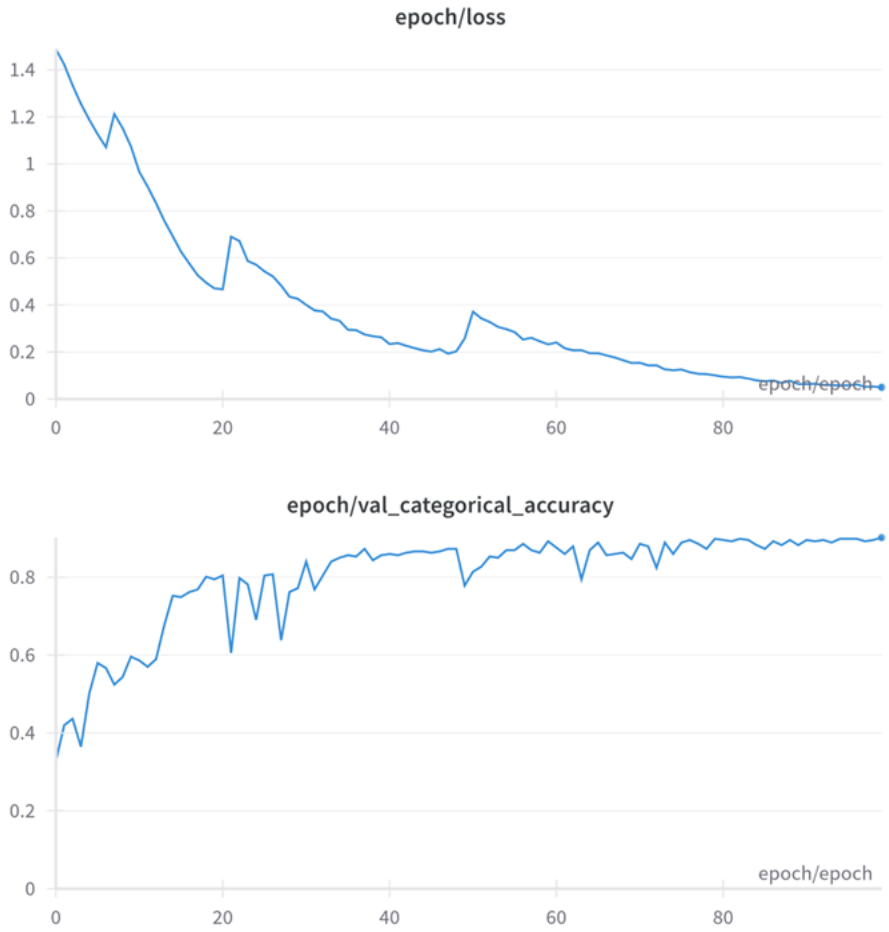


Рисунок 2.8 — Графики обучения модели FACE.

и детализации разметки, что увеличивает затраты на подготовку обучающего набора.

Модель EsResNet

Enhanced Spectrogram-based ResNet (EsResNet) использует адаптированную архитектуру ResNet, которая была специально модифицирована для работы с аудиоданными в виде двумерных спектрограмм. Спектрограмма представляет сигнал как изображение, где одна ось соответствует времени, а другая — частоте, при этом интенсивность цвета отражает амплитуду [67]. Максимальная точность модели достигла 91,48% после применения трансфера обучения. На рисунке 2.9 показаны результаты обучения модели.

Основные аспекты архитектуры EsResNet:

– **Двумерные сверточные слои:** Применяются для анализа временно-частотных характеристик акустических данных.

– **Резидуальные блоки:** Уменьшают эффект затухания градиента, что особенно полезно для глубоких сетей [67].

Рисунок 2.9 — Графики обучения модели EsResNet.

– **Трансфер обучения:** Модель инициализируется предобученными весами из ResNet50, обученной на наборе изображений ImageNet [68].

Трансфер обучения позволил значительно сократить время тренировки и повысить точность модели, адаптируя её к акустическим данным. Однако для использования EsResNet требуется значительный объём вычислительных ресурсов, что может ограничивать её применение в реальных системах.

Модель BEATs

BEATs (Bidirectional Encoder representation from Audio Transformers) представляет собой современную архитектуру, основанную на трансформерах. Эта модель была адаптирована для анализа акустических данных, представленных в виде логарифмических мел-спектрограмм, что позволяет учитывать как временные, так и частотные зависимости сигнала. Максимальная точность модели составила 97,06%, что является наивысшим результатом среди рассмотренных моделей. График обучения BEATs представлен на рисунке 2.10.

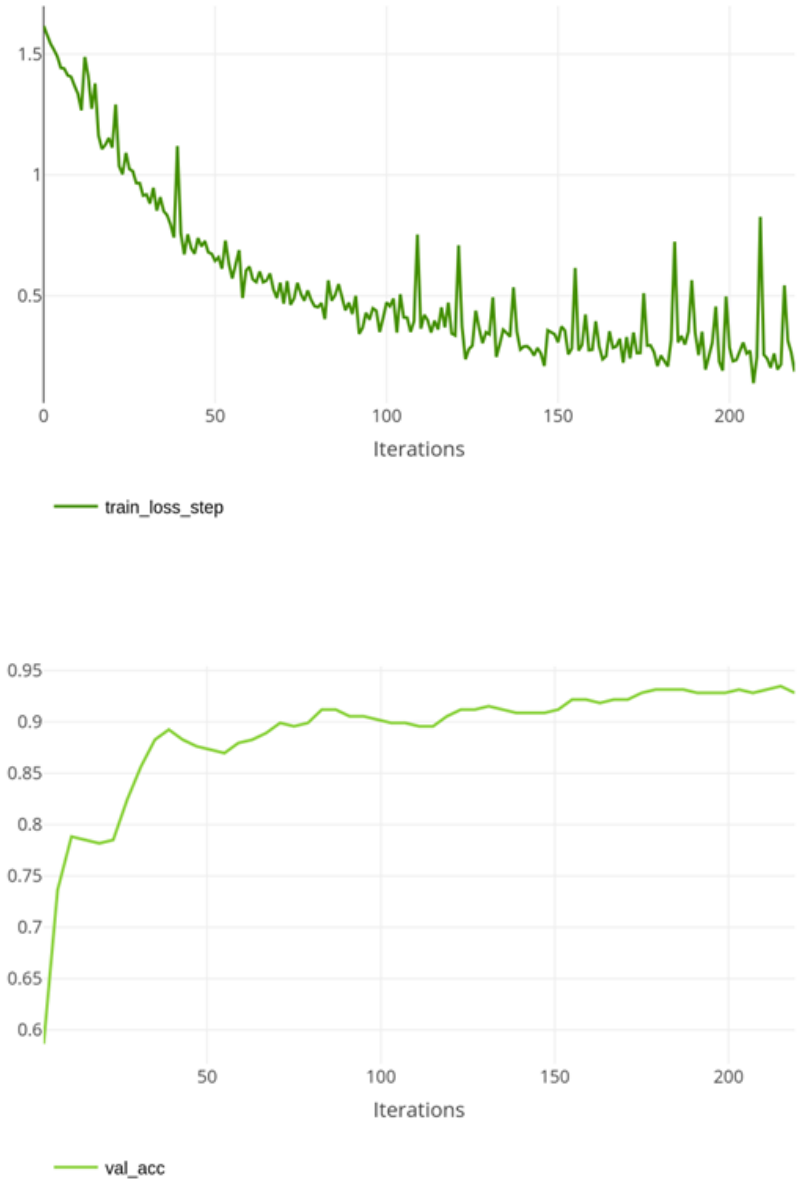


Рисунок 2.10 — Графики обучения модели BEATs.

Ключевые особенности BEATs:

– **Архитектура трансформера:** Использование механизма внимания позволяет выделять наиболее значимые части входного сигнала [24].

– **Предобучение на больших наборах данных:** Модель использует методы обучения без учителя (self-supervised learning), что улучшает её способность извлекать информативные признаки [69].

– **Дообучение на специализированных данных:** Для адаптации к дорожным акустическим событиям модель была дополнительно обучена на специализированном наборе данных.

– **Эффективная обработка сложных паттернов:** Механизмы трансформера позволяют классифицировать акустические события даже в условиях высокой зашумлённости.

Применение BEATs в дорожных условиях

Высокая точность модели делает её подходящей для задач мониторинга дорожной обстановки и классификации акустических сигналов в сложной акустической среде. Однако модель требует значительных вычислительных ресурсов, что может ограничивать её использование в устройствах с ограниченной мощностью, например, в системах встроенного анализа данных.

Сравнительный анализ рассмотренных моделей

Сравнение моделей показало, что каждая из них имеет свои сильные и слабые стороны. Модели 1DCNN и PIPMN являются менее ресурсоёмкими и подходят для задач с ограниченными требованиями к точности. FACE и EsResNet предоставляют более глубокий анализ данных, что делает их подходящими для сложных задач, однако они требуют большего объёма данных для обучения. Модель BEATs продемонстрировала наивысшую точность, но её применение ограничивается высокими вычислительными затратами.

В результате анализа нейросетевых моделей было установлено, что BEATs является наиболее эффективной архитектурой для задач классификации акустических событий дорожной обстановки. Однако для практического использования в реальных системах необходимы исследования, направленные на уменьшение её вычислительной сложности.

Для оптимизации модели и дальнейшего повышения качества классификации акустических сигналов предлагается рассмотреть два основных подхода: применение робастных методов оптимизации и дистилляция знаний.

Применение робастных функций потерь

При обучении моделей на реальных данных часто присутствуют выбросы и шумы, которые могут негативно влиять на их качество. Чтобы повысить устойчивость модели BEATs к таким аномальным данным, планируется исследовать использование робастных функций потерь, устойчивых к выбросам. Вместо стандартной функции потерь на основе перекрёстной энтропии предполагается протестировать следующие робастные функции:

– Функция Хьюбера;

– Функция Эндрюса;

– Биквадратная функция Тьюки.

Эти функции способны уменьшать влияние выбросов на процесс обучения за счёт снижения чувствительности к большим ошибкам. Ожидается, что применение робастных функций потерь позволит модели BEATs сохранять высокую точность даже при наличии значительного количества шумов и выбросов в обучающих данных.

В рамках этого подхода планируется:

– Подбор оптимальных параметров для каждой функции, чтобы достичь наилучшей производительности;
– Анализ влияния доли выбросов на качество модели при использовании различных функций потерь;
– Сравнительное исследование с использованием стандартной функции перекрёстной энтропии для оценки эффективности робастных функций.

Предполагается, что результатом станет улучшение обобщающей способности модели и повышение её устойчивости к аномалиям в данных.

Дистилляция знаний

Дистилляция знаний — метод, позволяющий передать знания от большой, высокопроизводительной модели (учителя) к более компактной и эффективной модели (ученику). В данном контексте предлагается перенести знания предобученной модели BEATs в компактную сверточную нейронную сеть.

Основные этапы предлагаемой дистилляции включают:

– **Дистилляция на основе ответов:** Модель-ученик будет обучаться на основе выходных распределений вероятностей (логитов) модели-учителя. Это позволит ученику перенять неявные знания об особенностях данных и взаимосвязях между классами, которые извлекла модель-учитель.
– **Использование методов аугментации данных:** В процессе обучения модели-ученика планируется применять различные техники аугментации данных, такие как:
    – Изменение скорости воспроизведения;
    – Сдвиг по времени;
    – Добавление шума.

Эти методы должны способствовать улучшению робастности и обобщающей способности модели-ученика.

Планируется провести эксперименты по обучению модели-ученика с различными коэффициентами температуры в функции потерь при дистилляции, чтобы оптимизировать процесс передачи знаний. Также предполагается оценить производительность модели-ученика на тестовых данных для сравнения с исходной моделью BEATs. Анализ компромисса между размером модели и точностью позволит определить оптимальную архитектуру модели-ученика.

2.4 Выводы по главе

В данной главе рассмотрены экспериментальные мероприятия по сбору акустических данных и созданию уникального набора (датасета), что позволило решить задачу №2 диссертационного исследования. Проведен анализ наборов данных (ESC-50, UrbanSound8K, FSD50K), включая их ключевые характеристики: количество классов, объем данных и метрики оценки. В течение трёх месяцев был проведён сбор данных в различных условиях (время суток, погода, дорожные ситуации). Специально разработанный маршрут в Москве обеспечил репрезентативность данных. Был проведён анализ и формирование наборов акустических данных дорожных событий. На основе собранной информации создан уникальный датасет, использованный для формирования обучающей и тестовой выборок, что стало важным этапом исследования.

В рамках решения задачи № 2 диссертации проведено исследование нейросетевых методов акустических данных дорожных событий. В рамках данного исследования получены следующие результаты:

- **1DCNN:** точность 47,89%, ограничение — слабый учёт контекста коротких событий;
- **PIPMN:** точность 84,04%, высокая эффективность, но недостаточная точность;
- **FACE:** точность 90,23%, благодаря MFCC и спектральному контрасту;
- **EsResNet:** точность 91,48% при трансферном обучении;
- **BEATs:** наилучшая точность 97,06%, благодаря трансформеру и обучению без учителя.

В данной главе решена третья задача диссертации, а именно: разработана система аннотирования акустических данных дорожных событий LabelTool. Ис-

пользование предобученной модели BEATs в процессе автоматической разметки позволило сократить время аннотирования одного часа аудиозаписей с 4,7 до 2,1 часа, одновременно улучшив качество разметки. Итеративное дообучение модели обеспечило дополнительный рост точности и более тонкую адаптацию к специфике акустических дорожных событий.

Также были предложены следующие методы повышения качества классификации акустических данных дорожных событий:

- **Робастные функции потерь:** использование робастных функций для устойчивости к шумам и выбросам.
- **Дистилляция знаний:** перенос знаний от модели BEATs к компактной модели для снижения вычислительных затрат.

Разработка алгоритма на базе данных методов рассматривается в главе 3.

Глава 3. Разработка метода и алгоритмического обеспечения нейросетевой обработки акустических данных дорожных событиях

В данной главе рассматриваются методы и подходы, направленные на разработку устойчивых и эффективных нейросетевых алгоритмов для классификации акустических данных дорожных событий. Учитывая специфические требования задач анализа акустической среды в реальном времени, таких как высокая точность и ограниченные вычислительные ресурсы, предложены решения, которые позволяют повысить точность моделей, сократить их вычислительные затраты и обеспечить устойчивость к шумам и выбросам в данных.

3.1 Исследование методов оптимизации нейросетевых алгоритмов классификации акустических данных дорожных событий

Специфика задачи классификации акустических сигналов дорожной среды для систем помощи водителю (ADAS, Advanced Driver Assistance Systems) требует покрытия больших территорий и работы в условиях интенсивного акустического фона. Такие системы должны обеспечивать обработку акустических сигналов в режиме реального времени для предупреждения водителей о потенциальных угрозах, таких как аварии, акустические сигналы сигнализации или экстренные акустические сигналы. Реализация таких функций предполагает использование множества устройств, которые либо локально обрабатывают аудиоданные, либо передают их на центральный сервер для анализа.

Передача данных на центральный сервер требует высокой пропускной способности сети, что предъявляет строгие требования к инфраструктуре и безопасности системы. Локальная обработка, в свою очередь, зависит от вычислительных возможностей самих устройств. Модель *BEATs*[70], являющаяся высокоточной системой для классификации акустических сигналов, включает в себя более 90 млн параметров в формате `float32`, что требует более 360 мегабайт оперативной памяти. Это ограничивает возможности использования модели на компактных устройствах, таких как встроенные модули автомобилей, и увеличивает стоимость оборудования.

Для оптимизации вычислительных ресурсов и снижения требований к оборудованию, при сохранении высокой точности классификации, в работе был применён метод дистилляции знаний. Этот подход позволил значительно уменьшить размер модели, адаптируя её для работы в условиях ограниченных вычислительных мощностей, что делает её более подходящей для задач встраивания в системы ADAS.

Дистилляция знаний (*Knowledge Distillation*) — это набор методов, позволяющих извлекать некоторое знание из модели, называемой учителем, и переносить их в другую модель, называемую учеником. Этот метод впервые был формализован в статье [71]. Изначально целью метода являлось уменьшение вычислительной стоимости применения модели, так как модель-ученик может иметь архитектуру, отличающуюся от модели-учителя, с минимизацией падения точности. Однако некоторые практические случаи показывают, что применение методов дистилляции знаний может даже улучшить метрики качества модели [72].

Методы дистилляции знаний можно классифицировать следующим образом:

- по типу знаний,
- по принципу дистилляции,
- по разнице архитектур учителя и ученика.

Знания, основанные на ответах. Метод дистилляции знаний заключается в обучении модели-ученика имитировать предсказание последнего слоя учителя[73]. Для этого вводится функция потерь дистилляции:

$$\mathcal{L}_{\text{dist}} = \mathcal{L}(\text{softmax}(\mathbf{z}_T/T), \text{softmax}(\mathbf{z}_S/T)), \tag{3.1}$$

где $\mathbf{z}_T$ и $\mathbf{z}_S$ — выходы последнего слоя моделей учителя и ученика соответственно, T — температура, контролирующая плавность распределения вероятностей классов, а $\mathcal{L}$ — функция расстояния, которая может быть расстоянием Кульбака-Лейблера или средней квадратической ошибкой. Применение *softmax* с температурой T определяется формулой:

$$p(z_i, T) = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}, \tag{3.2}$$

где z_i — выход соответствующий i-му классу. Температура позволяет модели-ученику узнать распределение вероятностей классов для каждого примера входных данных.

Такой подход часто комбинируют с техникой сглаживания меток:

$$T_{\text{smoothed}} = T(1 - \alpha) + \frac{\alpha}{C}, \tag{3.3}$$

где C — количество классов, α — коэффициент неуверенности, увеличивая который, можно поднимать штраф за чрезмерную уверенность модели в ответе.

Знания, основанные на признаках. Более общий подход заключается в обучении модели-ученика с использованием промежуточных признаков модели-учителя[74]. Функция потерь для этого вида дистилляции выглядит следующим образом:

$$\mathcal{L}_{\text{dist}} = \mathcal{L}(\mathcal{T}(f_T(x)), \mathcal{S}(f_S(x))), \tag{3.4}$$

где $f_T(x)$ и $f_S(x)$ — признаки, полученные из промежуточных слоев учителя и ученика, $\mathcal{T}$ и $\mathcal{S}$ — функции трансформации, применяемые для согласования размерностей признаков, а $\mathcal{L}$ — функция расстояния, которая может быть выбрана из множества вариантов: расстояние Кульбака-Лейблера, среднеквадратичная ошибка, L1-норма, перекрестная энтропия или MMD-loss.

Знания, основанные на связях. Этот подход использует информацию о связях между признаками, извлеченными моделью для различных входных данных[75]. Функция потерь в данном случае формулируется следующим образом:

$$\mathcal{L}_{\text{dist}} = \sum_{i,j} \mathcal{L}(\mathcal{S}(z_{i,S}, z_{j,S}), \mathcal{T}(z_{i,T}, z_{j,T})), \tag{3.5}$$

где $z_{i,S}$ и $z_{j,S}$ — признаки модели ученика, а $\mathcal{S}$ и $\mathcal{T}$ — функции сходства признаков, например, косинусное расстояние или евклидова метрика.

Методы дистилляции могут быть классифицированы по принципу их применения:

- **Использование предобученной модели-учителя.** Этот подход является самым распространенным и используется для упрощения архитектуры конечной модели.
- **Совместное обучение учителя и ученика.** Обучение обеих моделей одновременно позволяет повысить эффективность дистилляции.

– **Дистилляция внутри одной модели.** Например, передача знаний из глубоких слоев модели в более поверхностные.

Архитектуры ученика могут отличаться от архитектуры учителя:

– **Идентичные архитектуры.** Ученик использует ту же архитектуру, но работает с меньшей точностью данных, например, `float16`[76].

– **Упрощенные архитектуры.** Модель-ученик имеет меньше слоев, уменьшенные размеры слоев или блоков[77].

– **Сильно отличающиеся архитектуры.** Например, дистилляция между различными типами моделей, такими как сверточные сети и трансформеры[78].

Дистилляция знаний позволяет существенно сократить вычислительные затраты, сохраняя при этом высокую точность модели и обеспечивая ее адаптацию к различным условиям применения.

3.2 Устойчивый алгоритм обучения нейронной сети в условиях выбросов и шумов в обучающем наборе данных

Аудиозаписи постоянной длины t_{len} секунд, закодированные импульсно-кодовой модуляцией с частотой дискретизации f_d, представляют собой вектора X длины $m = \lfloor t_{\text{len}} \cdot f_d \rfloor$. Существует множество классов $C = [1 : N]$, и каждой аудиозаписи сопоставлен вектор $T = (0\ 1\ \ldots\ 0\ 1_i\ 0\ \ldots\ 0_N)$, где i соответствует классу аудиозаписи.

Нейронная сеть предназначена для изучения набора параметров Θ для сопоставления входного вектора X с предсказанием T в соответствии с иерархическим извлечением признаков, заданным уравнением:

$$T = F(X|\Theta) = f_L(...f_2(f_1(X|\Theta_1)|\Theta_2)|\Theta_L), \tag{3.6}$$

где f_L — функция L-го слоя сети.

В процессе обучения параметры сети корректируются в соответствии с обратным распространением ошибки для минимизации функции потерь. Задача может быть описана как оптимизация:

$$\mathcal{L} = \min \sum_{i=1}^{N} \sum_{j=1}^{M} \text{loss}(z_{ij}, t_{ij}), \tag{3.7}$$

где $\mathcal{L}$ — суммарная функция потерь, t_{ij} — требуемый выходной j-го нейрона, соответствующий входным данным x_i, а z_{ij} — действительный выходной j-го нейрона, соответствующий тем же входным данным x_i.

Классический метод оптимизации параметров нейронной сети обеспечивает высокую степень точности только при условии, что анализируемые данные достаточно качественные. Однако, в реальной практике исследователи часто сталкиваются с ситуацией, когда данные не соответствуют этому критерию, так как они могут включать аномальные наблюдения (выбросы), которые сильно отличаются от остальных.

Такие ситуации возникают из-за того, что данные, как правило, являются результатом реальных измерений. Выбросы могут появляться из-за грубых ошибок в процессе измерений или их присутствие может быть обусловлено самим характером данных.

Один из наиболее очевидных путей решения этой проблемы — это предварительная обработка данных и выбраковка выбросов, однако такой подход может привести к утрате важной информации.

Альтернативой может выступить устойчивый подход, применение которого позволяет учитывать аномальные наблюдения и минимизировать их отрицательное воздействие в процессе обучения нейросети. В этом случае в алгоритме обучения нейронной сети предполагается использование функции потерь, устойчивой к большим изменениям данных.

Необходимо учесть, что функция потерь, используемая в алгоритме обратного распространения ошибки, должна быть гладкой функцией первого порядка или непрерывно-дифференцируемой, то есть, имеющей непрерывную производную.

Биквадратная функция потерь Тьюки предназначена для снижения влияния выбросов в обучающих данных за счет ограничения изменений весов при больших отклонениях $z - t$. Эта функция применяется в задачах регрессии и классификации, где требуется высокая устойчивость[79]. Если отклонение меньше порогового значения λ, то используется шестая степень разности, чтобы смягчить влияние на обучение. При больших отклонениях используется квадратичная аппроксимация, что приводит к игнорированию выбросов.

$$\text{loss}(z_j, t_j) = \begin{cases} \frac{(z_j - t_j)^6}{6\lambda^2}, & |z_j - t_j| < \lambda, \\ \frac{(z_j - t_j)^4}{2\lambda^2} + \frac{(z_j - t_j)^2}{2}, & |z_j - t_j| \geqslant \lambda \end{cases} \tag{3.8}$$

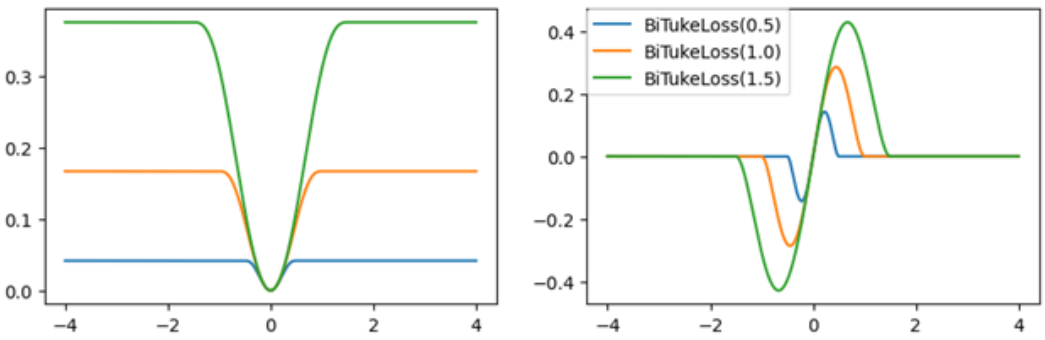


Рисунок 3.1 — Биквадратные функции потерь Тьюки с параметрами $(0.5, 1.0, 1.5)$, и их производные от $z - t$.

На рисунке 3.1 представлены графики биквадратной функции потерь Тьюки с разными значениями λ, а также производные, показывающие влияние отклонений на изменение весов.

Функция потерь Коши используется для регрессии и классификации, где требуется учитывать выбросы, но не игнорировать их полностью[80]. Она вводит плавный логарифмический рост потерь при увеличении отклонения $|z - t|$, что позволяет сохранять некоторую информацию о выбросах без сильного их влияния на итоговую модель. Это делает её полезной для задач с ограниченным количеством шумных данных.

$$\text{loss}(z_j, t_j) = \ln\left(\frac{1}{2}\left(\frac{z_j - t_j}{\lambda}\right)^2 + 1\right) \tag{3.9}$$

На рисунке 3.2 показаны графики функции потерь Коши для разных значений λ, а также производные, демонстрирующие её более мягкое воздействие на изменение весов по сравнению с классическими квадратичными функциями.

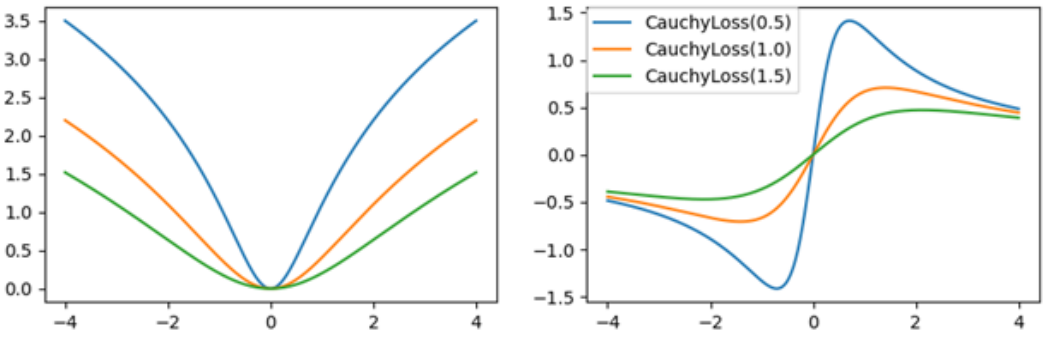


Рисунок 3.2 — Функции потерь Коши с параметрами $(0.5, 1.0, 1.5)$, и их производные от $z - t$.

Функция потерь Geman–McCluer предназначена для обработки данных с выбросами за счёт плавного уменьшения влияния отклонений, когда $|z - t|$

становится достаточно большим[81]. Geman–McCluer ограничивает потери при увеличении выбросов, что делает её подходящей для задач с высокой степенью шума.

$$\text{loss}(z_j, t_j) = \frac{(z_j - t_j)^2}{(z_j - t_j)^2 + \lambda} \tag{3.10}$$

На рисунке 3.3 представлены графики функции потерь Geman–McCluer для различных значений параметра λ, а также производные, которые иллюстрируют её устойчивость к выбросам.

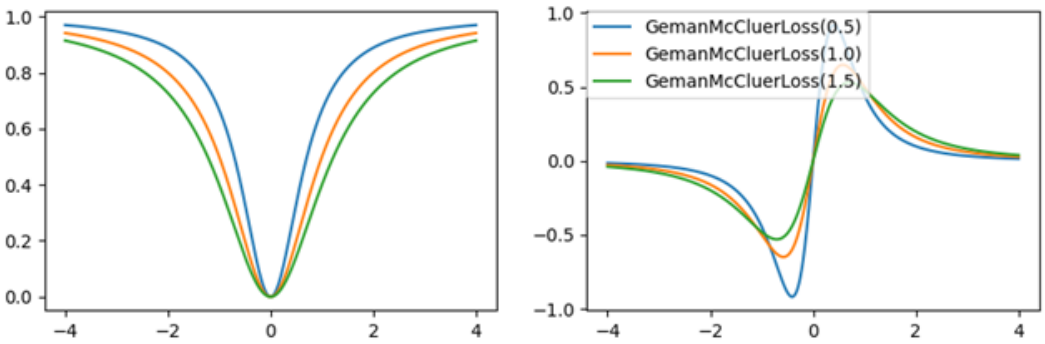


Рисунок 3.3 — Функции потерь Geman–McCluer с параметрами $(0.5, 1.0, 1.5)$, и их производные от $z - t$.

Функция Charbonnier представляет собой модифицированную версию абсолютной ошибки, обеспечивающую гладкость за счёт корня из квадратичной формы. Она является дифференцируемой и часто применяется для задач, где требуется минимизация выбросов[81].

$$\text{loss}(z_j, t_j) = \sqrt{\left(\frac{z_j - t_j}{\lambda}\right)^2 + 1} \tag{3.11}$$

На рисунке 3.4 приведены графики функции потерь Charbonnier для разных значений λ, а также производные, показывающие её поведение при больших отклонениях $|z - t|$.

Функция потерь Мешалкина предназначена для минимизации выбросов в данных с использованием экспоненциального подавления ошибок[82]. Она демонстрирует плавное поведение и снижает влияние больших отклонений $|z-t|$, что делает её полезной для устойчивого обучения.

$$\text{loss}(z_j, t_j) = \frac{1}{\lambda}\left(1 - \exp\left(-\frac{\lambda}{2}(z_j - t_j)^2\right)\right) \tag{3.12}$$

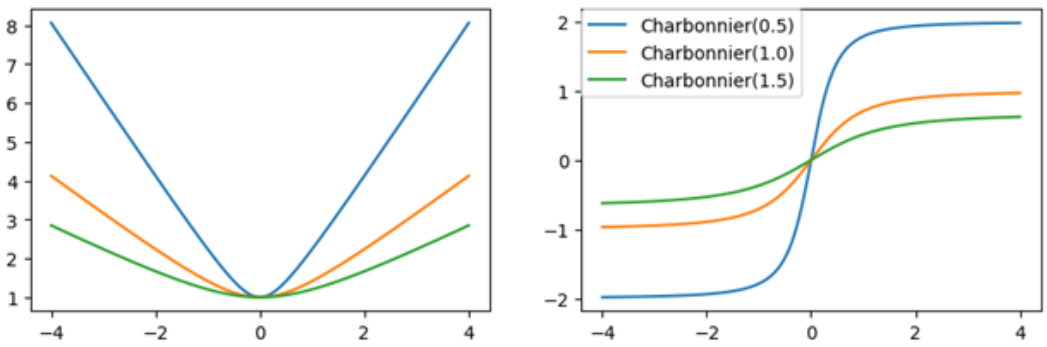


Рисунок 3.4 — Функции потерь Charbonnier с параметрами $(0.5, 1.0, 1.5)$, и их производные от $z - t$.

На рисунке 3.5 представлены графики функции потерь Мешалкина для различных значений параметра λ, а также их производные.

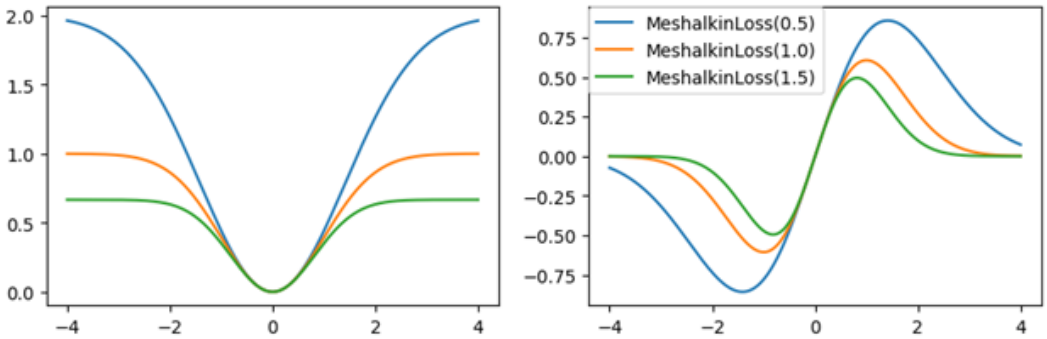


Рисунок 3.5 — Функции потерь Мешалкина с параметрами $(0.5, 1.0, 1.5)$, и их производные от $z - t$.

Функция Хьюбера является гибридной, сочетая квадратичную функцию для небольших ошибок и линейную функцию для больших ошибок[83]. Это делает её полезной для обработки данных с выбросами, где требуется сбалансированное влияние больших и малых ошибок.

$$\text{loss}(z_j, t_j) = \begin{cases} \frac{1}{2}(z_j - t_j)^2, & \text{если } |z_j - t_j| \leqslant \lambda, \\ \lambda|z_j - t_j| - \frac{1}{2}\lambda^2, & \text{если } |z_j - t_j| > \lambda. \end{cases} \tag{3.13}$$

На рисунке 3.6 приведены графики функции потерь Хьюбера для различных значений параметра λ, а также производные, демонстрирующие плавный переход между квадратичной и линейной частью.

Функция Эндрюса эффективно подавляет влияние выбросов благодаря использованию косинусного компонента, который ограничивает рост ошибки для больших отклонений $|z - t|$[84].

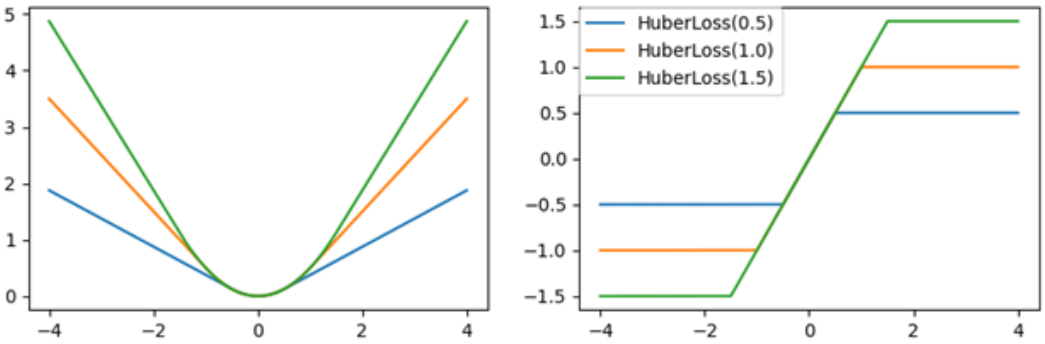


Рисунок 3.6 — Функции потерь Хьюбера с параметрами $(0.5, 1.0, 1.5)$, и их производные от $z - t$.

$$\text{loss}(z_j, t_j) = \begin{cases} \lambda\left(1 - \cos\left(\frac{z_j - t_j}{\lambda}\right)\right), & \text{если } |z_j - t_j| < \pi\lambda, \\ 2\lambda, & \text{если } |z_j - t_j| \geqslant \pi\lambda. \end{cases} \tag{3.14}$$

На рисунке 3.7 представлены графики функции потерь Эндрюса для различных значений параметра λ, а также их производные.

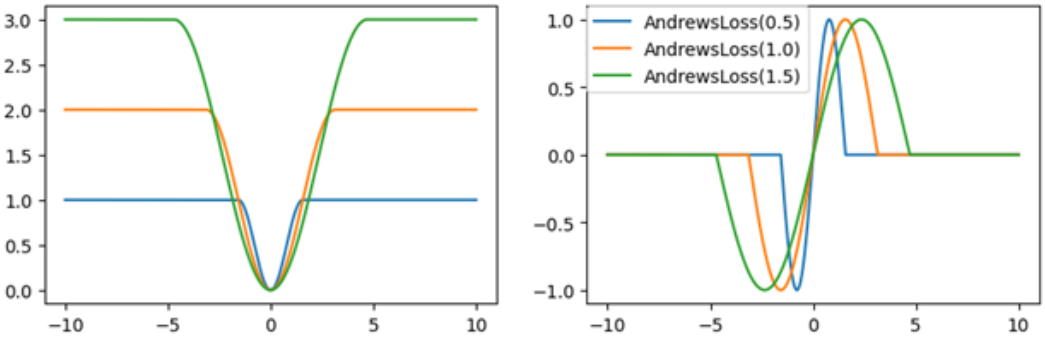


Рисунок 3.7 — Функции потерь Эндрюса с параметрами $(0.5, 1.0, 1.5)$, и их производные от $z - t$.

Функция Рамсея использует экспоненциальное подавление ошибок и логарифмическое сглаживание, что делает её устойчивой к выбросам[85].

$$\text{loss}(z_j, t_j) = \frac{1 - (1 + \lambda|z_j - t_j|)\exp(-\lambda|z_j - t_j|)}{\lambda^2} \tag{3.15}$$

На рисунке 3.8 приведены графики функции потерь Рамсея и их производные для различных значений параметра λ.

Функция потерь Уэлша подавляет большие отклонения ошибки $z - t$ с помощью экспоненциального сглаживания, что делает её устойчивой к выбросам и шумам[86].

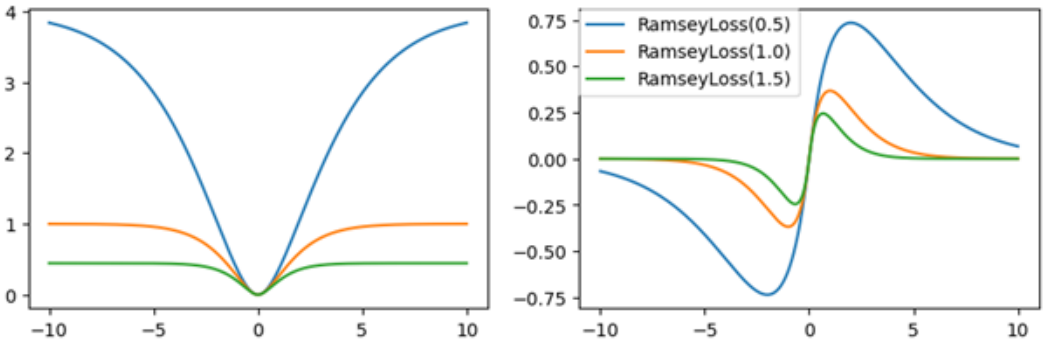


Рисунок 3.8 — Функции потерь Рамсея с параметрами $(0.5, 1.0, 1.5)$, и их производные от $z - t$.

$$\text{loss}(z_j, t_j) = 1 - \exp\left(-\frac{1}{2}\left(\frac{z_j - t_j}{\lambda}\right)^2\right) \tag{3.16}$$

На рисунке 3.9 приведены графики функции потерь Уэлша и их производные для различных значений параметра λ.

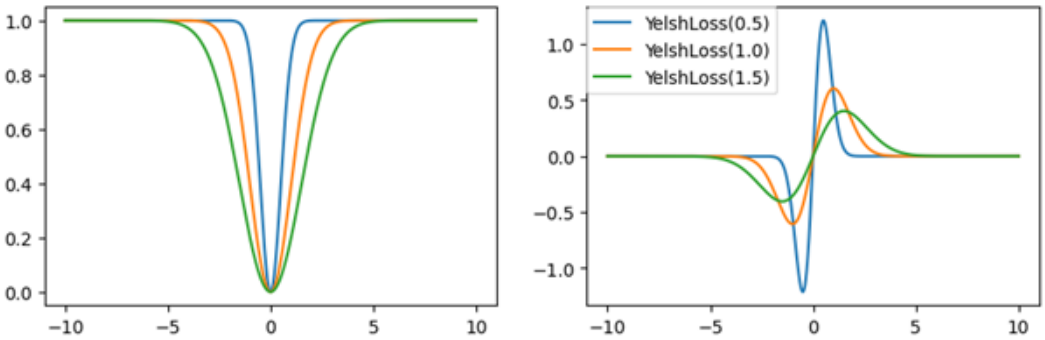


Рисунок 3.9 — Функции потерь Уэлша с параметрами $(0.5, 1.0, 1.5)$, и их производные от $z - t$.

Все производные представленных функций имеют предел, у всех за исключением функции Хьюбера и Charbonnier он равен 0. В контексте алгоритма обратного распространения ошибки это означает:

– для функций Хьюбера и Charbonnier — одинаковое изменение весов при большом отклонении z от t, что не даёт сильно испортить модель при обучении на выбросах;
– для остальных функций — отсутствие изменения весов при большом отклонении z от t, выбросы (пока они считаются таковыми моделью, что может измениться в процессе обучения) не учитываются.

В данной работе, в исследовательских целях, мы применили каждую из выше описанных функций с параметром $\{0.9, 1.0, 1.1\}$ при обучении модели BEATs на собранном наборе данных дорожных событий.

Набор данных состоит из 2600 аудиозаписей длительностью до 5 секунд (в среднем 2.5 секунды), распределённых на 5 классов. Из-за нехватки данных в определенных классах, вручную были добавлены примеры для уменьшения дисбаланса. Набор данных был разделён на обучающую и валидационную выборки в отношении 9 : 1. В целях изучения влияния количества выбросов на качество итоговой модели, мы добавили в обучающую выборку вне доменных аудиозаписей речи из набора данных "Голос" [87], так, чтобы количество выбросов составляло 5%, 10% и 15% от общего числа в обучающей выборке. Аудиозаписи для добавления выбирались случайно, причём записи, уже выбранные в датасет с меньшим процентом выбросов, исключались при последующих добавлениях.

Распределение классов в наборе данных представлено в таблице 4.

Таблица 4 — Распределение классов в наборе данных

Выборка	Сигнализация	Разбитое стекло	Гудок	Резкая остановка	Авария
Валидационная	68	73	56	58	52
Обучающая	530	532	571	466	660
Обучающая (с выброс., 5%)	598	605	627	524	712

Алгоритм обучения сети был реализован с использованием фреймворка `Pytorch-Lightning` и MLOps платформы `ClearML`. В целях ускорения процесса подбора гиперпараметров мы использовали предобученную на наборе данных `AudioSet`[88] модель, часть параметров которой не изменялась в процессе обучения. Обучение происходило с оптимизатором `AdamW`[89], с параметром `lr=1e-3` в течение 5 эпох. Размер батча составил 64 аудиозаписи, каждые 4 батча пересчитывалась точность на валидационной выборке. В качестве стандартной функции потерь была выбрана функция перекрестной энтропии.

Для обеспечения детерминированности во всех экспериментах использовалась одинаковая псевдослучайная последовательность. Максимальная точность на валидационной выборке для каждого эксперимента представлена в таблице 5.

Таблица 5 — Максимальная точность на валидационной выборке

Функция потерь	λ	**0**	**0,05**	**0,1**	**0,15**
Charbonnier	0,9	0,944	0,885	0,895	0,895
	1	0,944	0,885	0,895	0,895
	1,1	0,944	0,885	0,892	0,895
Geman–McCluer	0,9	0,941	0,892	0,899	0,899
	1	0,941	0,892	0,899	0,895
	1,1	0,941	0,889	0,895	0,895
Коши	0,9	0,941	0,885	0,895	0,895
	1	0,941	0,885	0,895	0,895
	1,1	0,944	0,885	0,895	0,895
Мешалкина	0,9	0,944	0,885	0,895	0,895
	1	0,944	0,885	0,895	0,895
	1,1	0,944	0,885	0,895	0,895
Перекрестной энтропии	-1	0,934	0,879	0,885	0,895
Рамсея	0,9	0,941	0,882	0,892	0,895
	1	0,941	0,882	0,892	0,895
	1,1	0,941	0,882	0,892	0,895
Тьюки (Биквадратная)	0,9	0,345	0,328	0,286	0,302
	1	0,928	0,908	0,905	0,899
	1,1	0,934	0,889	0,905	0,905
Уэлша	0,9	0,941	0,885	0,895	0,895
	1	0,944	0,885	0,895	0,895
	1,1	0,944	0,885	0,895	0,895
Хьюбера	0,9	0,944	0,882	0,882	0,863
	1	0,944	0,882	0,892	0,892
	1,1	0,944	0,882	0,892	0,892
Эндрюса	0,9	0,944	0,882	0,892	0,892
	1	0,944	0,882	0,892	0,892
	1,1	0,944	0,882	0,892	0,892

Анализ данных в таблице позволяет сделать несколько важных выводов. Прежде всего, добавление выбросов в обучающую выборку приводит к заметному снижению точности модели на валидационной выборке при первых уровнях добавления выбросов (5%). Это свидетельствует о высокой чувствительности модели к присутствию вне доменных данных, что может указывать на недостаточную устойчивость модели к шумам и искажениям, несвязанным с целевыми классами данных, что представлено на графике рисунка 3.10.

Однако при дальнейшем увеличении доли выбросов до 10–15% снижение точности оказывается менее значительным. Этот эффект можно объяснить несколькими факторами, связанными со свойствами предобученной модели и её способностью сохранять извлечённые признаки даже при увеличении количества шумовых данных.

Во-первых, в модели использовалась предобученная сеть, большая часть весов которой была заморожена. Такое замораживание весов ограничивает степень адаптации модели к новым данным, предотвращая значительные изменения параметров на этапе обучения. Этот механизм замораживания сыграл роль стабилизатора, сохранив изначально полученные характеристики модели, обученной на чистых данных. Как следствие, модель демонстрировала относительную устойчивость к добавленным выбросам, что помогало избежать резкого ухудшения её результатов на валидационной выборке. Это особенно важно в условиях, когда выбросы занимают значительную часть обучающей выборки, поскольку модель сохраняет способность извлекать полезные паттерны из оставшихся данных высокого качества.

Во-вторых, выбранный детерминированный процесс обучения значительно уменьшил вариативность результатов. Детерминированность обеспечивает предсказуемость поведения модели, так как при одних и тех же входных данных и параметрах обучения она выдаёт одинаковые результаты. Это позволяет избежать случайных колебаний в точности, которые могли бы быть вызваны шумами или другими внешними факторами. Благодаря этому модели удавалось достичь высокой степени повторяемости результатов, что также способствовало её стабильности, даже несмотря на присутствие выбросов в обучающих данных. Такой подход снижает вероятность того, что модель "переобучится"на случайные шумы, что особенно важно при работе с данными, содержащими выбросы.

Влияние выбросов на качество обучения существенно зависело от выбора функции потерь. Наиболее устойчивой оказалась биквадратная функция

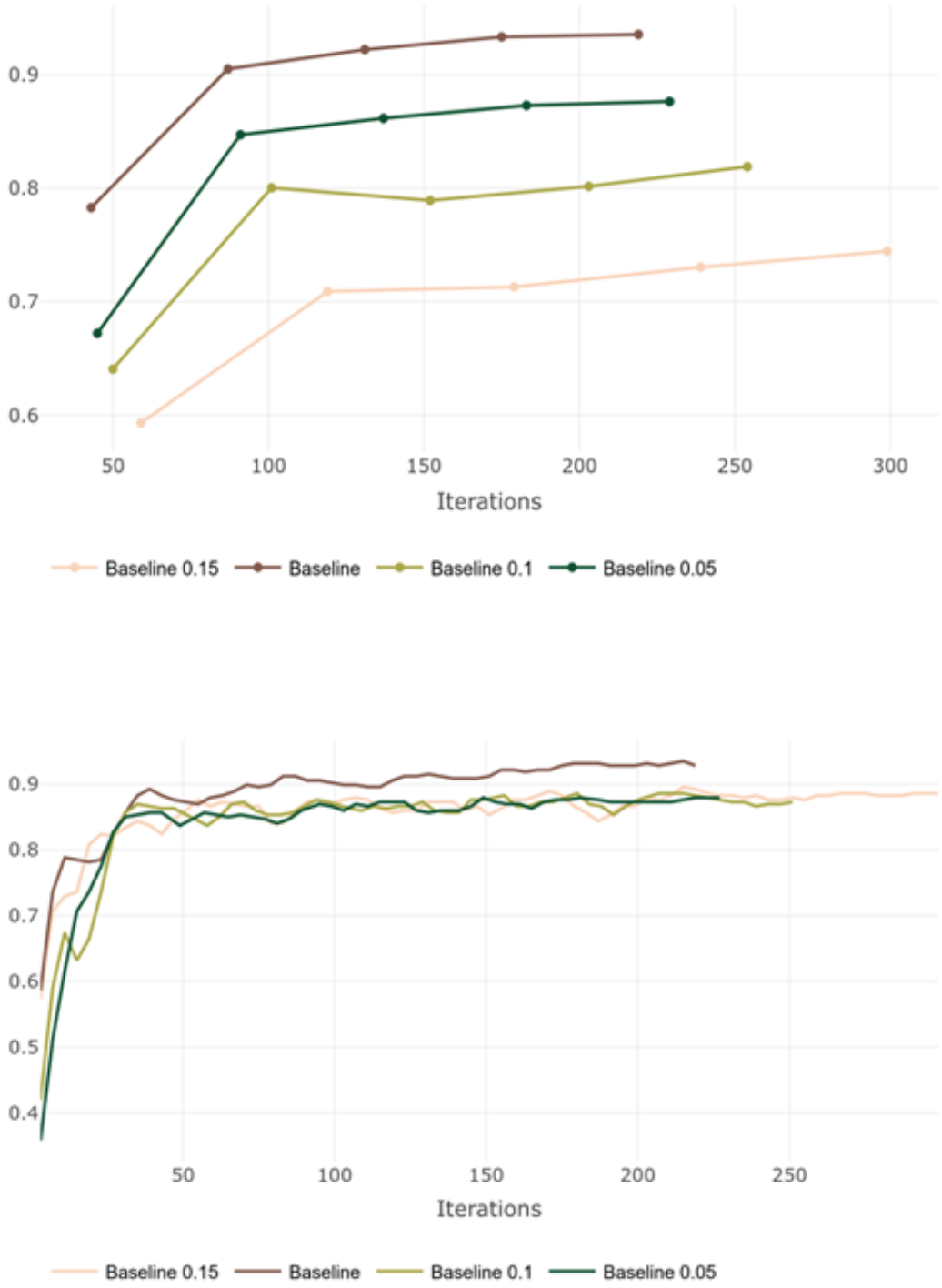


Рисунок 3.10 — График точности на валидационной и обучающей выборке с функцией перекрёстной энтропии.

Тьюки, при которой модель сохраняла высокую точность даже при увеличении доли выбросов до 15%. Благодаря свойству «игнорировать» крупные аномальные наблюдения, функция Тьюки эффективно справлялась с шумными и вне доменными данными.

В то же время другие функции потерь, такие как функция Хьюбера и функция Эндрюса, показали свои ограничения. Они продемонстрировали высокую точность классификации в условиях отсутствия выбросов, достигая значений точности до 0.944. Однако их эффективность значительно снижалась с увеличением доли выбросов, особенно когда она превышала 5%. Это указывает на их чувствительность к шумам и выбросам, что делает их менее подходящими для использования в задачах, где данные содержат значительное количество аномальных наблюдений. Это может быть связано с тем, что данные функции потерь менее агрессивно обрабатывают выбросы, позволяя им вносить вклад в обновление весов модели, что приводит к ухудшению её способности обобщать.

Такое поведение функций Хьюбера и Эндрюса подчёркивает необходимость их осторожного применения, особенно в задачах, где точность критически важна, а уровень шума и выбросов в данных высок. Тем не менее, их высокая эффективность в условиях чистых данных делает их полезными в случаях, когда качество обучающих данных может быть гарантировано.

Таким образом, результаты анализа функций потерь демонстрируют, что выбор подходящей функции потерь является важным аспектом для обеспечения устойчивости модели в условиях зашумленных данных. Биквадратная функция Тьюки оказалась наиболее универсальным решением для работы с выбросами, обеспечивая высокую стабильность и точность модели даже при значительном уровне шума. Это делает её предпочтительным выбором для реальных приложений, где данные часто бывают неидеальными.

Для задач, предполагающих работу с данными, содержащими выбросы, наиболее подходящим выбором является функция потерь биквадратная Тьюки, поскольку она демонстрирует наибольшую устойчивость к шумам и минимальные потери в точности.

3.3 Разработка нейросетевого алгоритма классификации акустических данных дорожных событий

Для существенного снижения размера модели была проведена дистилляция знаний из модели BEATs[70] в модель MobileNetv3[52] . BEATs использует

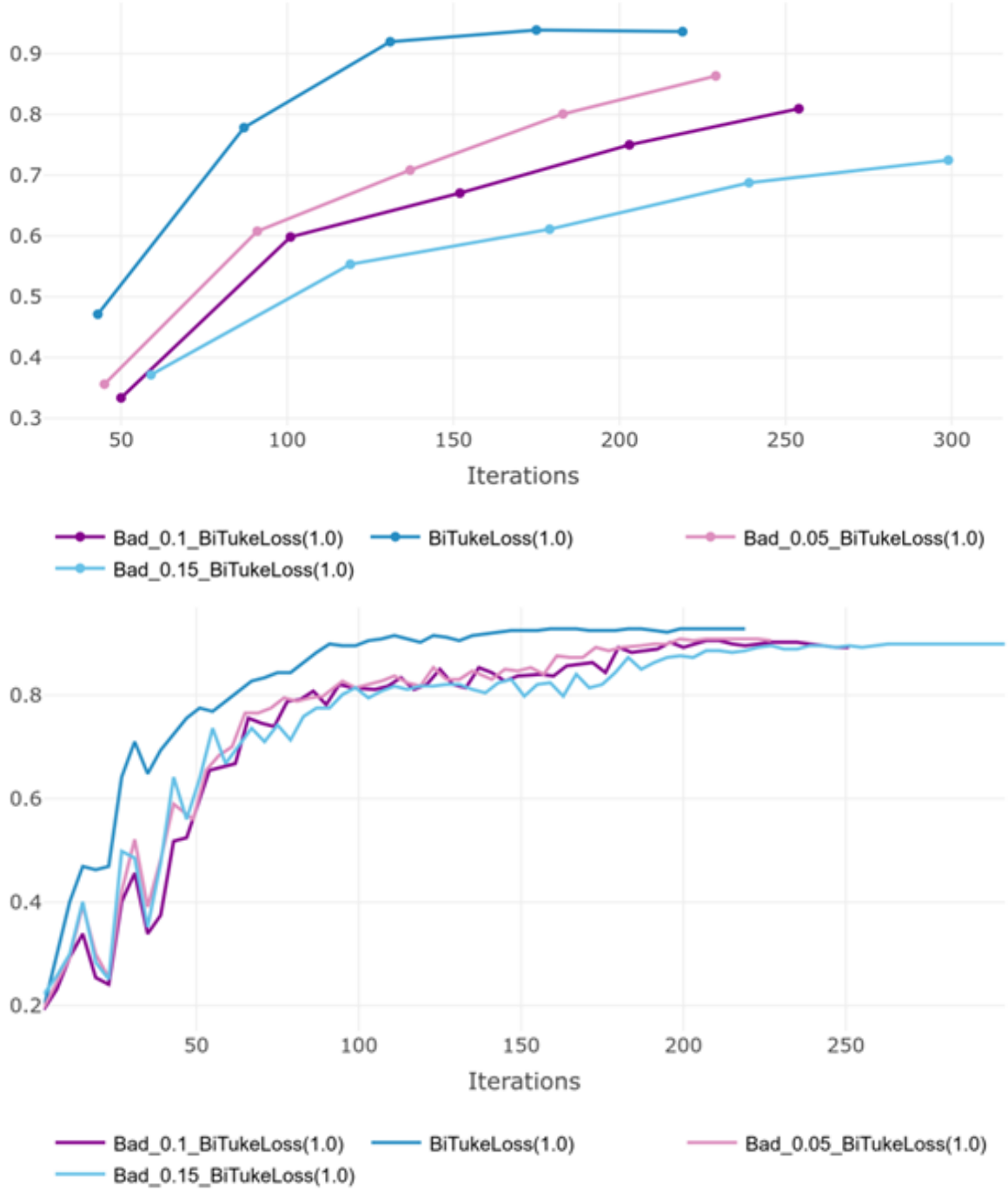


Рисунок 3.11 — График точности на валидационной и обучающей выборке с биквадратной функцией Тьюки.

архитектуру аудиоспектрограммного трансформера, тогда как MobileNetv3 представляет собой сверточную нейронную сеть, разработанную для процессоров мобильных устройств.

В таблице 6 приведены характеристики моделей BEATs и MobileNetv3, включая количество параметров, среднее время работы и объем используемой оперативной и видеопамяти при обработке одной аудиозаписи длиной 5 секунд

с частотой дискретизации 16 кГц. Тесты проводились на ПК со следующими характеристиками:

– Операционная система: Ubuntu 23.10;
– Процессор: AMD Ryzen 7 6800H;
– Видеокарта: GeForce RTX 3050 Mobile с драйвером NVIDIA 535.171.04;
– Оперативная память: DDR5 4800MHz SO-DIMM.

Таблица 6 — Характеристики работы моделей BEATs и MobileNetv3

Модель	Кол-во парам., млн	ОЗУ, МБ	VRAM, МБ	Время, мс
BEATs (CPU)	90,3	318,5	0,0	123,0
BEATs (GPU)	90,3	13,2	321,5	23,7
MNv3 (CPU)	0,19	29,6	0,0	7,0
MNv3 (GPU)	0,19	1,0	28,7	6,3

Для обучения модели MobileNetv3 использовалась предобученная модель BEATs (точность на валидационной выборке составила 0.97). Тестировались два подхода дистилляции знаний: основанная на ответе и на связях между признаками. Используемая функция потерь имеет вид:

$$\text{Loss} = \alpha \cdot \text{Loss}_1 + (1 - \alpha) \cdot \text{Loss}_2, \tag{3.17}$$

где Loss_1 — функция перекрестной энтропии, а Loss_2 — функция потерь дистилляции. Для дистилляции, основанной на связях между данными, в качестве функции сходства использовалось евклидово расстояние, а в качестве функции расстояния между парами аудиозаписей — функция Хьюбера с параметром $\lambda = 1$.

Эксперименты проводились как с аугментацией данных, так и без неё. Использовались следующие методы аугментации:

– Изменение амплитуды на ±10 дБ;
– Сдвиг по оси времени;
– Инверсия полярности.

Для сокращения времени обучения предсказания и признаки модели BEATs для обучающей выборки были заранее вычислены. Для аугментированных данных использовались признаки, полученные из неизмененных аудиозаписей. Это сократило время обучения одной эпохи с 55 секунд до 5 секунд.

На рисунке 3.12 представлена диаграмма с параллельными координатами, иллюстрирующая результаты подбора гиперпараметров на протяжении 60 эпох.

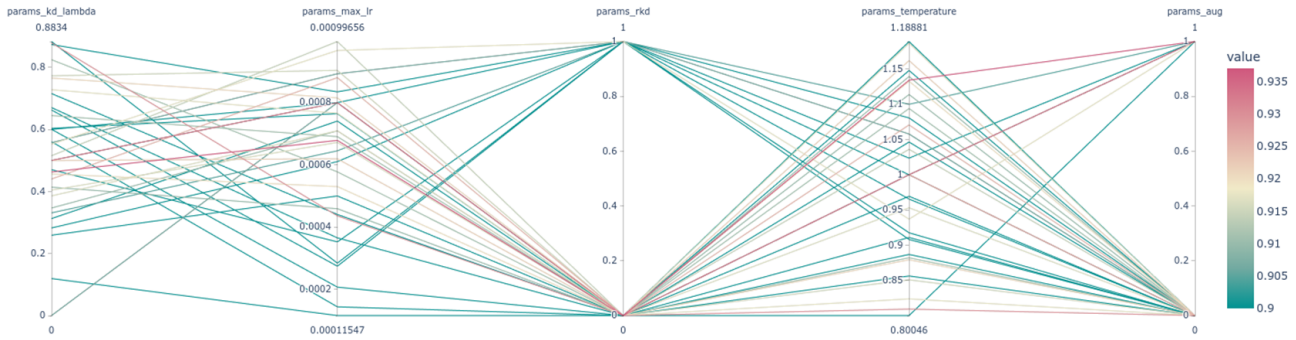


Рисунок 3.12 — Результаты подбора гиперпараметров.

Наилучший результат показала дистилляция знаний, основанная на ответе, с параметрами $G = 1.134$, $\alpha = 0.462897$ и максимальной скоростью обучения 6.79×10^{-4}. Разница между итоговой точностью на валидационном наборе данных при обучении с дистилляцией и без неё составила 3%.

Используя оптимальные параметры, модель была обучена на 80 эпох. Графики точности на валидационном наборе данных для вариантов с аугментацией и без неё, а также с дистилляцией и без неё, представлены на рисунке 3.13.

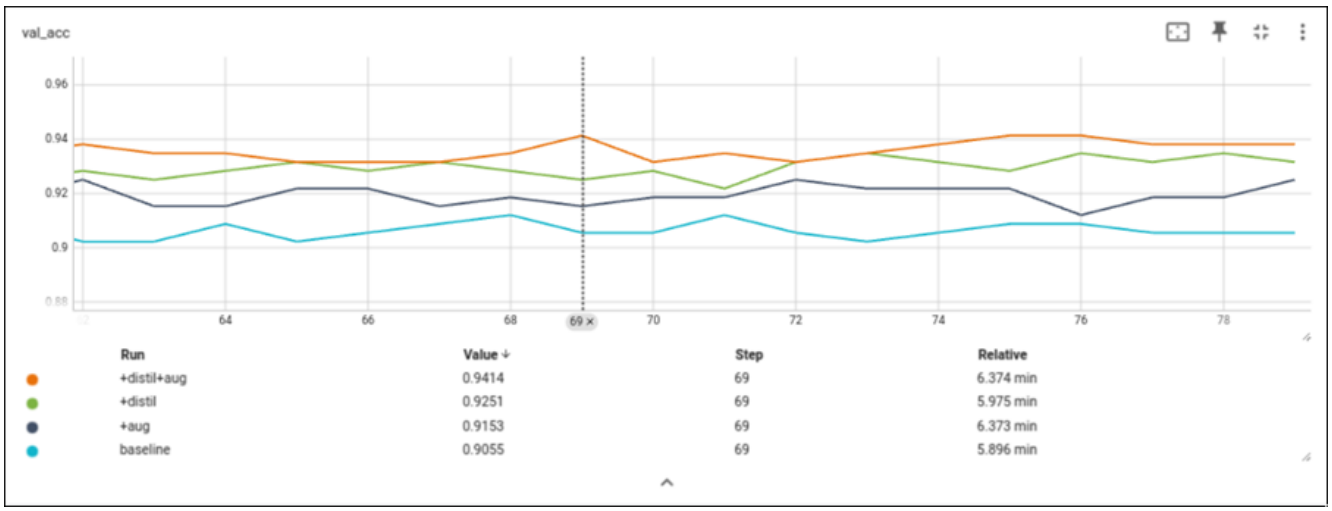


Рисунок 3.13 — Точность на валидационном наборе данных при использовании аугментации.

Из рисунка 3.13 видно, что аугментация может приводить к кратковременным падениям точности, однако итоговый результат с её использованием в среднем на 1%-2% лучше.

В таблице 7 представлены результаты максимальной и итоговой точности на валидационном наборе данных при различных комбинациях аугментации и дистилляции.

Таблица 7 — Максимальная и итоговая точность на валидационном наборе данных

Дистилляция	Аугментация	Максимальная точность, %	Итоговая точность, %
Нет	Нет	91,42	90,5
Нет	Да	93,48	92,5
Да	Нет	94,14	93,1
Да	Да	94,14	93,8

В результате исследований дистилляции знаний и её влияния на производительность модели MNv3 была выявлена проблема: несмотря на уменьшение вычислительных затрат и компактность модели, её точность на валидационной выборке не достигла уровня модели BEATs, показав 94,14% против 97,06%. Это стало стимулом для дальнейших экспериментов с использованием сетей Колмогорова — Арнольда (KAN)[90] в целях повышения точности модели и улучшения её способности обрабатывать сложные зависимости между признаками аудиоданных.

Сети Колмогорова — Арнольда (Kolmogorov — Arnold Networks, KAN) основаны на теореме Колмогорова — Арнольда, которая утверждает, что любую многомерную непрерывную функцию можно представить в виде суперпозиции непрерывных функций одной переменной [91]. Это свойство открывает возможность аппроксимации сложных зависимостей между признаками, используя набор функций меньшей размерности. Математическая формулировка теоремы выражается следующим образом:

$$f(X) = \sum_{q=1}^{2n+1} \sum_{p=1}^{n} \psi_{q,p}(x_p), \tag{3.18}$$

где $f(X) : [0,1]^n \to \mathbb{R}$ — многомерная функция, $\psi_{q,p}$ — одномерные непрерывные функции. Данное представление позволяет заменить многомерные преобразования на композиции одномерных функций.

В задачах классификации аудиоданных, таких как определение дорожных событий, использование KAN открывает возможность улучшить обработку

высокоуровневых признаков, извлечённых из аудиосигналов. Однако есть ряд проблем, которые необходимо учитывать при использовании KAN:

1. **Число необходимых функций.** Для аппроксимации функций в n-мерном пространстве требуется $2n+1$ функций для каждого класса, что приводит к значительным вычислительным затратам.
2. **Отсутствие дифференцируемости.** Теорема Колмогорова — Арнольда не требует дифференцируемости функций, что затрудняет использование метода градиентного спуска.

Для решения этих проблем в данной работе была реализована модифицированная архитектура KAN. В качестве предварительного этапа использовалась модель MNv3 для извлечения признаков, которые затем обрабатывались KAN.

Для повышения вычислительной эффективности были предложены следующие модификации KAN:

1. **Сжатие входных данных.** Исходные данные предварительно сокращались до векторов меньшей размерности с помощью модели MNv3. Это позволило снизить вычислительную сложность задачи.
2. **Использование сплайнов.** Для обеспечения гладкости функций активации использовались B-сплайны, определяемые рекурсивной формулой Кокса—де Бура [92]:

$$B_{i,0}(x) = \begin{cases} 1, & t_i \leqslant x < t_{i+1}, \\ 0, & \text{в противном случае,} \end{cases} \tag{3.19}$$

$$B_{i,k}(x) = \frac{x - t_i}{t_{i+k} - t_i} B_{i,k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} B_{i+1,k-1}(x), \tag{3.20}$$

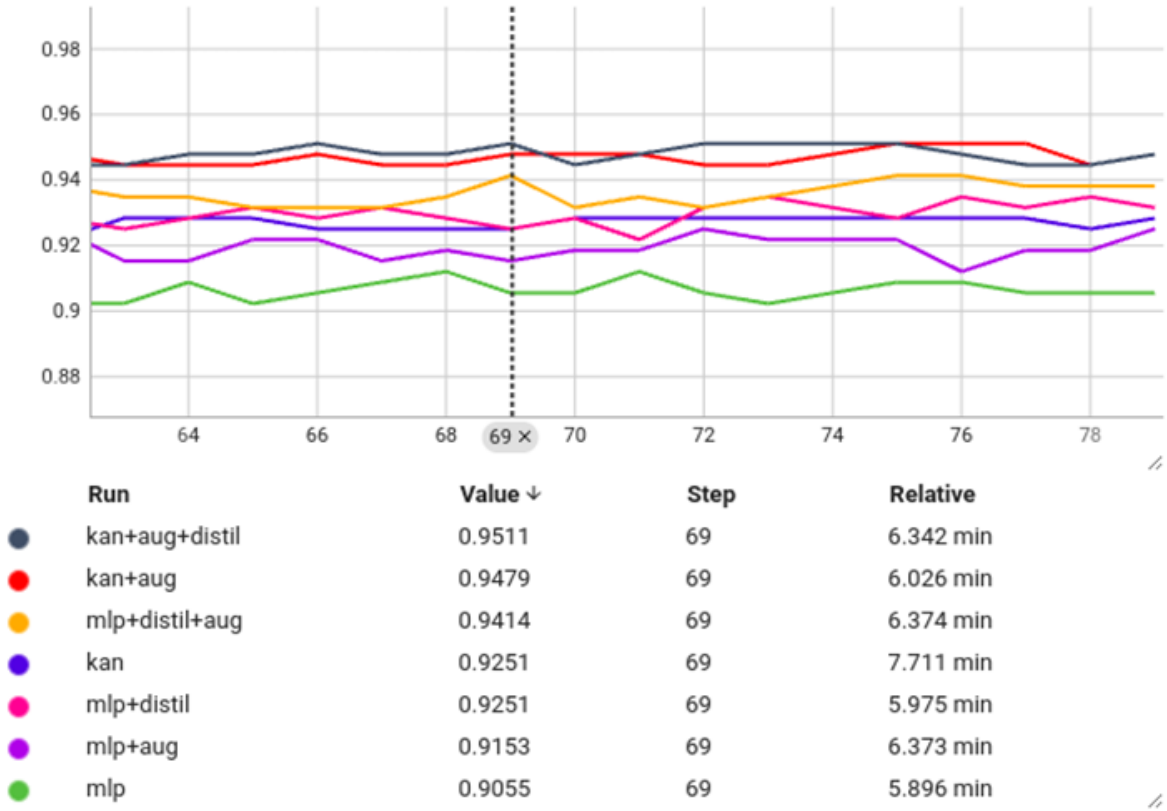
   где t_i — узлы сплайнов.
3. **Перенос активаций на рёбра.** Функции активации сети были перенесены на рёбра, что позволило аппроксимировать их взвешенной суммой базисной функции $b(x)$ и B-сплайнов $\mathrm{spline}(x)$.

Итоговая архитектура слоя KAN описывалась следующим образом:

$$x_{l,p} = \sum_{q=1}^{r_{l-1}} \psi_{l,q,p}(x_{l-1,q}), \tag{3.21}$$

где $x_{l,p}$ — выходной сигнал нейрона p на слое l, $\psi_{l,q,p}$ — функция активации.

Для проверки эффективности KAN модель MNv3 была модифицирована путём замены её финальных слоёв на сеть KAN. Модель обучалась с использованием аугментации данных и дистилляции знаний из модели BEATs. На рисунке 3.14 представлено сравнение точности модели с KAN и с многослойным персептроном (MLP). После внедрения KAN была достигнута точность 95,11% на валидационной выборке. Наилучшие результаты наблюдались при использовании аугментации данных совместно с дистилляцией знаний.



Run	Value ↓	Step	Relative
kan+aug+distil	0.9511	69	6.342 min
kan+aug	0.9479	69	6.026 min
mlp+distil+aug	0.9414	69	6.374 min
kan	0.9251	69	7.711 min
mlp+distil	0.9251	69	5.975 min
mlp+aug	0.9153	69	6.373 min
mlp	0.9055	69	5.896 min

Рисунок 3.14 — График точности на валидационной выборке с аугментацией.

Наилучшие результаты были достигнуты при использовании KAN совместно с аугментацией и дистилляцией знаний, что позволило достичь точности 95,11%.

3.4 Выводы по главе

В данной главе были изучены и реализованы три ключевых подхода для повышения точности и эффективности классификации акустических данных: устойчивые методы оптимизации, дистилляция знаний и внедрение сетей

Колмогорова — Арнольда (KAN). Полученные результаты демонстрируют соответствие поставленной задаче по разработке устойчивого алгоритма анализа и классификации акустических данных для определения характеристик и источников акустических сигналов на дороге.

Решена задача № 4 диссертации, а именно: разработан устойчивый алгоритм нейронной сети в условиях выбросов и шумов в обучающем наборе данных за счет применения функций потерь совместно с дистиляцией знаний. Анализ функций потерь Хьюбера, Эндрюса и Тьюки показал их эффективность для обработки данных с шумами и выбросами, характерными для городской среды. Эти методы минимизировали влияние аномалий, повысив устойчивость модели. Например, при увеличении выбросов до 15% устойчивые функции сохраняли точность модели, где стандартные подходы значительно теряли в производительности. Применение дистилляции знаний из модели *BEATs* в *MobileNetV3* сократило размер модели с 90,3 млн до 0,19 млн параметров, сохранив высокую точность (94,14% против 97,06% у *BEATs*). Использование аугментации данных, таких как временные сдвиги и изменение амплитуды, позволило улучшить обобщающую способность модели, сохранив её производительность даже на шумных данных. Оптимизация вычислений снизила время обработки на *CPU* с 123 мс до 7 мс, а на *GPU* — с 23,7 мс до 6,3 мс, что делает модель подходящей для систем реального времени.

В данной главе решена задача № 5, разработан алгоритм классификации акустических данных дорожных событий. Замена финальных слоёв *MobileNetV3* на сеть KAN повысила точность до 95,11%. Эти сети эффективно моделируют сложные зависимости, улучшая производительность модели при сохранении её компактности. Адаптация методов аппроксимации функций активации с использованием B-сплайнов обеспечила стабильность и эффективность в условиях городской среды.

Эти методы обеспечивают высокую точность и ресурсоэффективность, создавая основу для применения в системах мониторинга дорожной обстановки, где необходима надёжность в условиях динамической городской среды.

Глава 4. Разработка архитектуры программно-аппаратного комплекса сбора и цифровой обработки акустических данных дорожных событиях

В данной главе описывается комплексный подход к решению задачи акустического обнаружения и классификации дорожных событий, начиная с разработки системы сбора акустических данных (выбора аппаратной платформы, конфигурации микрофонного массива и методов масштабируемого сбора информации) и завершая описанием системы классификации и постобработки, включающей алгоритмы пространственной фильтрации (beamforming), нейросетевую модель для определения типов событий и серверную инфраструктуру, необходимую для анализа и хранения результатов.

В целом рассматриваются методы и алгоритмы, реализованные при создании программно-аппаратного комплекса акустического обнаружения. Подробно описывается выбор аппаратной платформы и конфигурации микрофонного массива, алгоритмы предобработки аудиоданных, процессы классификации и постобработки результатов, а также приводятся результаты теоретических и практических испытаний системы.

4.1 Архитектура комплекса сбора акустических данных

В данном разделе основное внимание уделяется разработке и реализации **системы сбора акустических данных**. В качестве ключевого элемента системы используется **микрофонный массив**, состоящий из восьми микрофонов. Размещение микрофонов осуществляется по схеме, позволяющей охватить круговую область вокруг автомобиля. Основная задача данной конфигурации — обеспечить всесторонний сбор данных, который необходим для точной классификации и анализа акустических событий в различных условиях дорожного движения [93]. Эффективность сбора данных и их последующая обработка являются ключевыми факторами для достижения высокой точности в работе нейросетевых алгоритмов, применяемых в данном исследовании. На рисунке 4.1 представлена схема микрофонного массива.

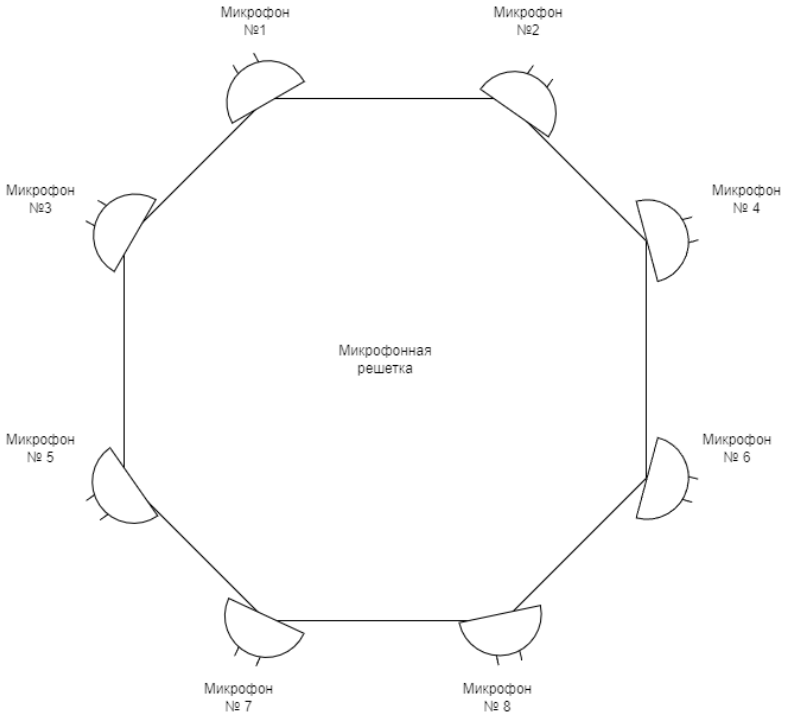


Рисунок 4.1 — Схема микрофонного массива.

Применение микрофонного массива, состоящего из восьми микрофонов, позволяет реализовать методы пространственной локализации источников акустического сигнала. Это достигается за счёт анализа разницы во времени поступления акустических волн на каждый микрофон массива [94]. При наличии информации о геометрическом расположении микрофонов становится возможным вычисление направления на источник акустического сигнала. Запись акустического сигнала с микрофонов осуществляется с применением 8-канальной аудиокарты, что позволяет решить задачу синхронизации аудиодорожек между собой [95].

Для реализации системы сбора данных были выбраны следующие комплектующие:

- Микрофоны sE electronics 8 pair [96];
- Аудиокарта Behringer UMC 1820 [95];
- Балансные XLR кабели для подключения микрофонов [53];
- Компьютер с программным обеспечением Reaper для записи акустического сигнала [97];
- 4 камеры GoPro (360°) [98].

Выбранные микрофоны обладают кардиоидной диаграммой направленности, что обеспечивает оптимальный баланс между направленностью и широтой охвата. Эта характеристика выгодно отличает их от суперкардиоидных и ги-

перкардиоидных микрофонов, которые имеют более узкую направленность. На рисунке 4.2 представлена диаграмма направленности микрофона sE electronics 8 pair.

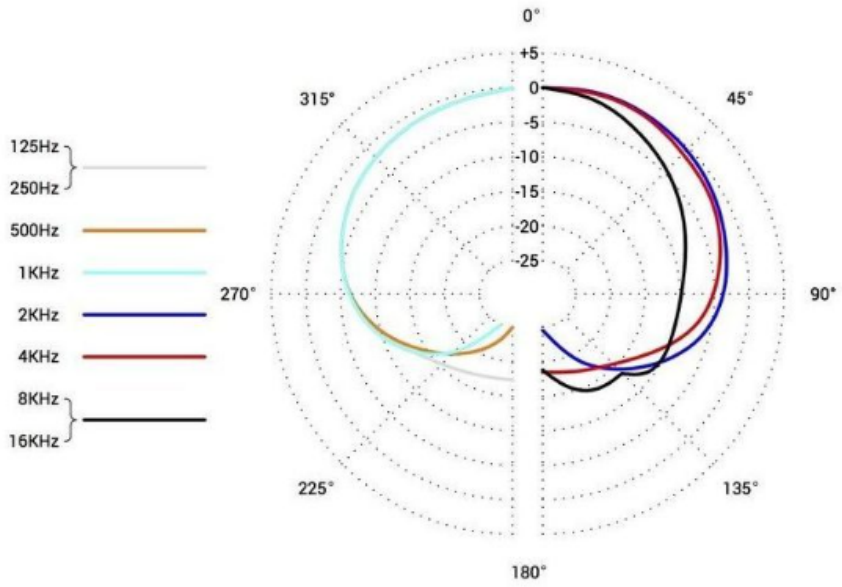


Рисунок 4.2 — Диаграмма направленности микрофона.[96]

При установке микрофонного массива и камеры в системе сбора акустических данных было уделено внимание их оптимальному расположению и ориентации. Микрофоны были равномерно распределены по периметру автомобиля, обеспечивая максимальный охват пространства и минимизируя влияние шумов, создаваемых автомобилем. Камеры 360° были установлены на крыше автомобиля в центральной точке, обеспечивая наилучший обзор [99].

Для синхронизации записей с микрофонов и камеры был разработан протокол, основанный на использовании акустических меток. Перед началом каждой записи подавался характерный акустической сигнал, который одновременно записывался микрофонами и камерой. Этот сигнал служит точкой отсчёта для последующей синхронизации аудио- и видеопотоков методом кросс корреляции. Такой подход минимизирует риск аппаратных сбоев, обеспечивая точность синхронизации.

Для обеспечения надёжности работы системы разработан протокол проверки, выполняемый перед каждым сеансом сбора данных:

1. Визуальный осмотр микрофонов и камеры на предмет механических повреждений [95].
2. Проверка надёжности подключения микрофонов к аудиокарте и камеры к системе записи.

3. Тестовая запись короткого аудио с каждого микрофона для оценки качества акустического сигнала.
4. Тестовая запись короткого видео для оценки качества изображения.
5. Проверка работы системы синхронизации с помощью акустического сигнала.

В случае выявления неисправностей принимаются меры по устранению, вплоть до замены неисправных компонентов. Только после успешного прохождения всех этапов проверки начинается полноценный сбор данных [60; 61]. Схема программного комплекса представлена на рисунке 4.3.

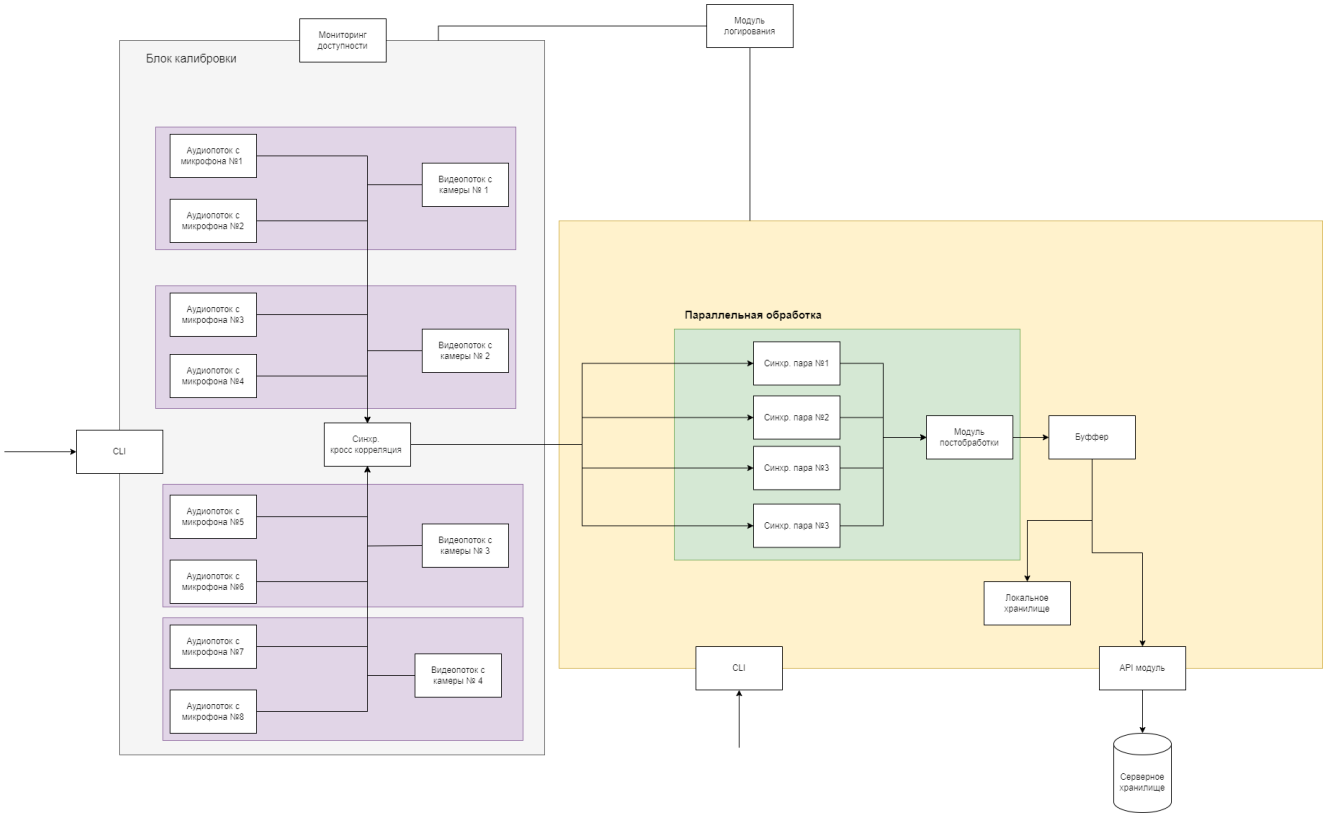


Рисунок 4.3 — Архитектура ПО сбора акустических данных.

В настоящем разделе описана система, позволяющая эффективно собирать акустические данные в реальном дорожном окружении. Однако для последующего анализа и классификации акустических событий одной лишь надёжной аппаратуры сбора недостаточно. Необходимо учесть ограничения вычислительных ресурсов (особенно при мобильном размещении в автомобиле), а также требования к компактности микрофонного массива и объёму передаваемых данных. В следующих разделах будет описаны уже модули система классификации, которая должна работать в режиме реального времени, обраба-

тывать сигналы с ограниченных аппаратных ресурсов и обеспечивать высокую точность детектирования различных дорожных событий.

4.2 Выбор аппаратной основы и конфигурации микрофонного массива

В целях увеличения дистанции детектирования акустических событий и обеспечения возможности точного определения их пространственного расположения в горизонтальной плоскости, был выбран микрофонный массив в качестве основного компонента системы. Микрофонный массив представляет собой совокупность приёмников акустического сигнала, работающих в координации друг с другом [94]. Сигналы от этих приёмников обрабатываются с учётом определённых фазовых задержек, что позволяет формировать направленную диаграмму чувствительности. Такой подход даёт возможность выделять акустические сигналы, приходящие из заданного направления, и подавлять шумы и сигналы, поступающие из других направлений. Это значительно повышает точность и качество детектирования акустических событий.

Применение микрофонного массива обеспечивает несколько ключевых преимуществ. Во-первых, использование множественных микрофонов позволяет повысить чувствительность системы за счёт объединения сигналов от нескольких приёмников. Во-вторых, возможность регулировать параметры обработки сигналов, такие как задержки и весовые коэффициенты, открывает широкие возможности для динамической настройки направленной диаграммы чувствительности массива. Это особенно важно в условиях городской среды, где источники акустических событий могут быть подвижными или находиться в разных направлениях относительно системы.

Существуют различные конфигурации расположения приёмников акустического сигнала в пространстве, каждая из которых обладает своими преимуществами и ограничениями. Поскольку задачей являлось определение направления на акустическое событие в горизонтальной плоскости без необходимости трёхмерной локализации, была выбрана плоская двумерная круговая конфигурация микрофонного массива. Такая конфигурация представляет собой расположение микрофонов по окружности с равными угловыми

интервалами между ними. Круговая структура массива обладает рядом преимуществ: она обеспечивает равномерное покрытие по азимуту, что позволяет обнаруживать акустические сигналы независимо от их направления относительно системы. Кроме того, такая конфигурация упрощает математические расчёты фазовых задержек, необходимых для формирования направленной диаграммы чувствительности.

Решение о выборе круговой конфигурации также было обусловлено компактностью и простотой её реализации. В условиях городской среды, где пространство для размещения аппаратуры может быть ограничено, использование кругового массива позволяет эффективно использовать доступное место. Это делает его подходящим для установки на транспортных средствах или стационарных объектах с ограниченными размерами.

Схематическое изображение круговой конфигурации микрофонного массива представлено на рисунке 4.4. Такая структура массива закладывает основу для эффективного решения задачи акустического мониторинга, обеспечивая высокую точность и гибкость системы в различных условиях эксплуатации.

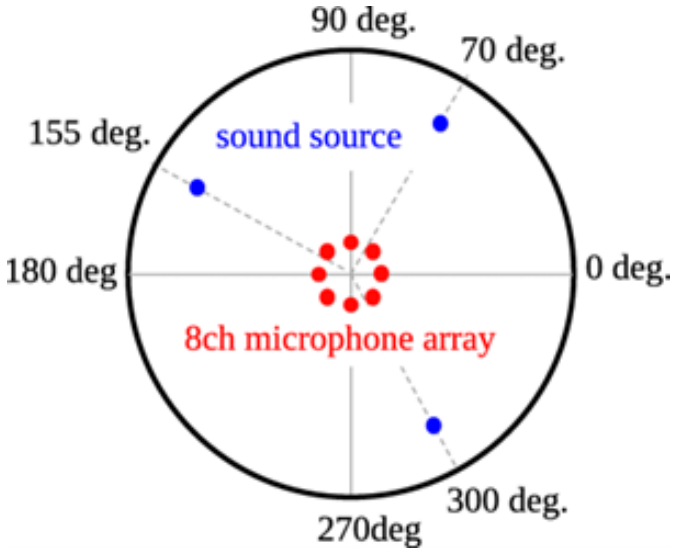


Рисунок 4.4 — Круговая конфигурация микрофонного массива.

Микрофонный массив был установлен на транспортное средство для тестирования системы помощи водителю (ADAS) в условиях реального дорожного движения. Такое размещение обеспечивает мобильность системы и её работу в различных сценариях городской среды. Установка на автомобиле позволяет собирать данные в реальных условиях, что важно для проверки функциональности системы.

Использование встроенного источника питания автомобиля упрощает эксплуатацию, устраняя необходимость в дополнительных элементах питания. Тестирование в реальных условиях дорожного движения помогает учитывать влияние факторов, таких как шум ветра, движение других транспортных средств и окружающие акустические сигналы, что делает испытания максимально информативными. Этот подход позволяет оценить работоспособность системы и её потенциал для интеграции в полноценные ADAS-решения.

Основой массива является конфигурация из восьми высокоточных микрофонов, которые обеспечивают качественный захват акустического сигнала в широком частотном диапазоне. Такая конфигурация микрофонов позволяет достичь высокой чувствительности и точности, что особенно важно для обнаружения и локализации акустических событий в условиях городской среды, где присутствует множество фоновых шумов. Фотография собранного микрофонного массива, установленного на транспортное средство, представлена на рисунке 4.5.

Рисунок 4.5 — Фотография микрофонного массива из 8 микрофонов.

Для управления микрофонами и обработки их сигналов используется микроконтроллер STM32 NUCLEO-L476RG[100] и многоканальная аудиокарта, которые обеспечивают стабильную и надёжную работу системы. Микроконтроллер реализует функцию опроса микрофонов, обрабатывает полученные данные и формирует многоканальный аудиопоток. Выбор STM32 NUCLEO-L476RG обусловлен его высокой производительностью и низким энергопотреблением.

Многоканальный аудиопоток формируется в формате, совместимом с WAV-файлами, что делает его удобным для дальнейшего анализа. Заголовки фиксированного размера содержат информацию о количестве каналов (в данном случае 8), частоте дискретизации (48 000 Гц) и формате аудиоданных (32-битное целое число со знаком). Аудиоданные передаются в виде непрерывного потока сэмплов, где данные с каждого микрофона идут последовательно в фиксированной последовательности. Это обеспечивает синхронизацию сигналов и облегчает их обработку алгоритмами формирования направленных диаграмм чувствительности.

Этот массив, интегрированный с автомобильной платформой, представляет собой гибкую и эффективную систему для сбора акустических данных в сложных условиях городской среды.

4.3 Метод предобработки акустических данных

Для формирования диаграммы направленности микрофонного массива и выделения акустческих сигналов из определённых направлений применялся стандартный алгоритм формирования луча *Delay And Sum* (DAS), что в переводе означает «задержка и суммирование» [101]. Этот алгоритм является одним из наиболее широко используемых подходов в акустической обработке сигналов и позволяет адаптивно усиливать акустические сигналы, поступающие из заданного направления, одновременно подавляя сигналы из других направлений. Преимущество метода DAS заключается в его простоте реализации, универсальности и низких вычислительных затратах, что делает его подходящим выбором.

Основной принцип работы алгоритма заключается в том, что сигналы с разных микрофонов микрофонного массива обрабатываются с учётом временных задержек, компенсирующих разницу во времени прихода акустической волны к каждому микрофону [102]. После этого обработанные сигналы суммируются, что позволяет усиливать сигналы, поступающие из заданного направления, за счёт когерентного сложения их фаз, и ослаблять сигналы из других направлений, где их фазы не совпадают.

Обозначим m_i как i-й микрофон из массива, состоящего из N элементов, а $y_i(t)$ как выходной сигнал этого микрофона в момент времени t. Каждый микро-

фон записывает сигнал с учётом своего расположения относительно источника акустического сигнала. Для корректной обработки алгоритм DAS применяет временную задержку τ_i, которая компенсирует разницу во времени прихода акустической волны к i-му микрофону, и амплитудный весовой коэффициент w_i, позволяющий учитывать индивидуальные особенности каждого микрофона, такие как чувствительность и направленность. После применения задержек и весовых коэффициентов сигналы с N микрофонов суммируются, формируя результирующий сигнал.

Сигнал на выходе алгоритма DAS можно представить следующим выражением:

$$z(t) = \sum_{i=1}^{N} w_i \cdot y_i(t - \tau_i), \tag{4.1}$$

где $z(t)$ — результирующий сигнал, w_i — весовой коэффициент для i-го микрофона, а τ_i — временная задержка для сигнала i-го микрофона. Это выражение описывает процесс преобразования сигналов микрофонного массива в сигнал, преимущественно отражающий акустический сигнал, поступающий из заданного направления. Где τ_i рассчитываются таким образом, чтобы обеспечить максимальную чувствительность массива к сигналам, поступающим из определённого направления. Регулируя эти задержки, можно изменять диаграмму направленности массива без изменения его положения в пространстве. Управление весовыми коэффициентами w_i позволяет усилить приём сигналов с нужного направления и ослабить влияние сигналов из других направлений.

Схематическое представление алгоритма DAS показано на рисунке 4.6.

Для расчёта задержек τ_i была разработана математическая модель кругового микрофонного массива. Целью расчёта является определение разницы в пути, которое акустическая волна проходит от источника до каждого микрофона, чтобы компенсировать соответствующие временные задержки.

Предполагая, что задача заключается в определении направлений на события в двумерной плоскости и учитывая физические размеры микрофонного массива, введено допущение о параллельном распространении акустических сигналов от удалённых источников. Это позволяет упростить расчёты, принимая фронт акустической волны за плоский.

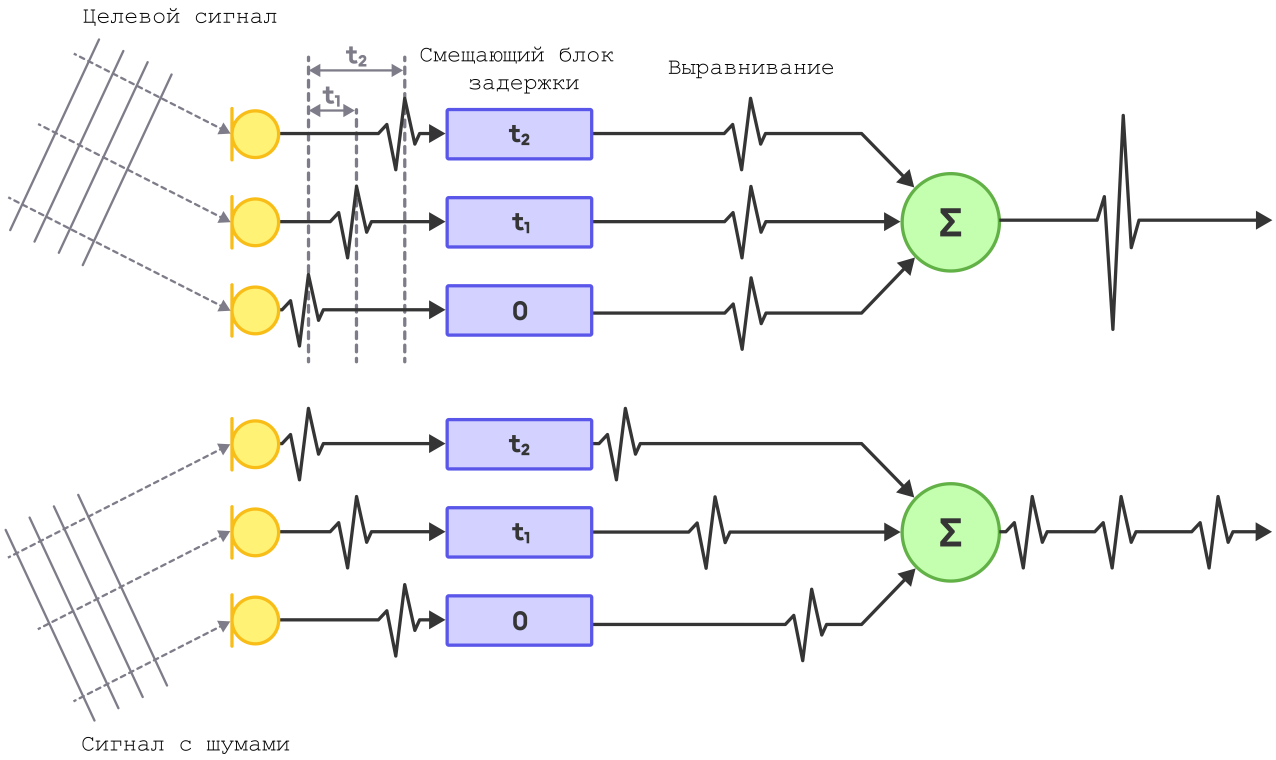


Рисунок 4.6 — Схема алгоритма *Delay And Sum*.

Координаты i-го микрофона в двумерной плоскости можно выразить как:

$$x_i = R\cos\left(\frac{2\pi i}{N}\right), \quad y_i = R\sin\left(\frac{2\pi i}{N}\right), \tag{4.2}$$

где R — радиус микрофонного массива. Вектор, направленный вдоль распространения акустического сигнала от источника под углом θ к оси абсцисс, задаётся как $(\cos\theta, \sin\theta)$. Тогда прямая волновой поверхности описывается уравнением:

$$\cos\theta \cdot (x - R\cos\theta) + \sin\theta \cdot (y - R\sin\theta) = 0. \tag{4.3}$$

Расстояние d_i от волновой поверхности до i-го микрофона равно:

$$d_i = R \cdot (1 - \cos(\theta - \frac{2\pi i}{N})). \tag{4.4}$$

Задержка τ_i для каждого микрофона вычисляется как разность расстояний до ближайшего микрофона:

$$\tau_i = \frac{d_i - d_{\min}}{c}, \tag{4.5}$$

где c — скорость акустического сигнала в воздухе.

В первой версии алгоритма весовые коэффициенты w_i использовались только для нормализации амплитуды акустического сигнала:

$$z(t) = \frac{1}{N} \sum_{i=1}^{N} y_i \left(t - \frac{d_i - d_{\min}}{c} \right). \tag{4.6}$$

Для обеспечения возможности локализации акустических событий была выбрана программная реализация задержек. Такой подход позволяет параллельно обрабатывать аудиоданные с различных направлений без необходимости физического перемещения или изменения конфигурации микрофонного массива.

Для улучшения точности обнаружения запись аудиоданных производится с использованием временного окна длительностью T и перекрытием P. Алгоритм обработки аудиоданных включает следующие шаги:

1. Запись аудиоданных с N микрофонов длительностью P в буфер (поток 1).
2. Запись аудиоданных с N микрофонов длительностью $T - P$ в конец буфера (поток 1).
3. Формирование K направлений алгоритмом DAS (поток 2).
4. Перезапись аудиоданных длительностью P из конца буфера в начало (поток 2).
5. Обработка K аудиозаписей длиной T секунд батчами в нейронной сети (поток 2).
6. Объединение результатов классификации с каждого направления в один отчёт (поток 2).
7. Повторение цикла с шага 2.

Количество направлений K напрямую влияет на вычислительную нагрузку системы. С увеличением K растёт время выполнения шагов обработки, а также возрастает объём используемой оперативной памяти, так как данные обрабатываются для каждого направления. Однако увеличение K позволяет добиться большей точности локализации источников акустического сигнала, что особенно важно в сложных акустических условиях. Для сохранения непрерывности работы время, затраченное на выполнение шагов 3–6, должно быть меньше или равно $T - P$ секунд, где T — длительность временного окна, а P — время перекрытия окон.

В разработанной системе классификация аудиозаписей осуществляется с использованием нейронной сети, подробно описанной в предыдущей главе. Модель была предварительно обучена с применением метода дистилляции знаний из предобученной модели, что позволило существенно сократить вычислительные ресурсы, необходимые для работы классификатора. Это сокращение вычислительных затрат сделало возможным обработку аудиоданных, поступающих одновременно из множества направлений, в режиме низкой задержки.

Дополнительно, постобработка результатов классификации играет важную роль в повышении надёжности системы. Алгоритм объединяет результаты классификации из нескольких направлений в единый отчёт, выполняя следующие шаги:

- Для каждого направления выбирается класс с наибольшей вероятностью.
- Если вероятность класса ниже порогового значения, предсказание заменяется на «отсутствие событий» или «фон».
- События из соседних направлений с одинаковыми классами объединяются в одно, а угол обнаружения рассчитывается как среднее значение углов или направление с максимальной вероятностью.

Этот подход позволяет учитывать пространственную информацию и минимизировать вероятность ложных срабатываний, особенно в условиях наличия фонового шума или множественных источников акустического сигнала. Кроме того, интеграция пространственных данных делает систему более надёжной и точной в реальных условиях.

В совокупности, компактная архитектура, применение метода дистилляции знаний и алгоритмы постобработки обеспечивают высокую производительность и точность системы даже в сложных акустических сценариях. Это делает её подходящей для использования в задачах, требующих оперативного обнаружения и классификации акустических событий.

После обработки аудиоданных алгоритмом *Delay And Sum* (DAS) сигналы из различных направлений передаются на классификатор. Результаты классификации каждого направления объединяются в общий отчёт с использованием следующего алгоритма:

1. Для каждого направления определяется класс с наибольшей вероятностью. Это значение соответствует наиболее вероятному событию, зафиксированному в данном направлении.

2. Если вероятность класса меньше установленного порога, результат классификации заменяется на «отсутствие событий» или «фон».
3. Одинаковые классы событий, обнаруженные в соседних направлениях, объединяются в одну детекцию. При этом угол детекции определяется одним из двух методов:
    а) как среднее значение углов направлений, в которых были обнаружены идентичные события;
    б) как угол направления с максимальной вероятностью детекции.

Данный подход позволяет учитывать пространственную информацию, предоставляемую микрофонным массивом, что повышает точность и надёжность обнаружения событий. Он минимизирует количество ложных детекций и корректно группирует одинаковые события, происходящие в разных направлениях.

Эффективная работа системы требует постоянного анализа её производительности и адаптации к реальным условиям эксплуатации. Сбор данных является ключевым шагом для улучшения модели классификации, так как собранные аудиозаписи и результаты классификаций позволяют:

- анализировать поведение модели в реальных дорожных условиях;
- выявлять ошибки классификации, такие как ложные срабатывания или пропуски событий;
- идентифицировать новые типы событий, которые могут быть добавлены в систему для расширения её функциональности;
- оценивать метрики качества, такие как точность, полнота и устойчивость модели к шумам и выбросам.

Для автоматизации процесса сбора, хранения и анализа данных была разработана серверная инфраструктура, обеспечивающая эффективную обработку больших объёмов информации. Собранные аудиоданные и результаты классификаций передаются на сервер для дальнейшего анализа и хранения. Эти данные используются не только для мониторинга текущей работы системы, но и закладывают основу для постоянного обучения нейронной модели. Благодаря этой стратегии система сможет адаптироваться к изменяющимся условиям городской среды, учитывать новые типы акустических событий и поддерживать высокую точность классификации без необходимости полного переобучения.

Такой подход обеспечивает постоянное улучшение работы модели и её соответствие реальным условиям эксплуатации.

Серверная инфраструктура включает в себя следующие компоненты:

1. **Сервер для приёма данных.** Реализован на базе фреймворка *FastAPI*. Устройства, оборудованные микрофонными массивами, отправляют на сервер следующие данные:
   - аудиоданные с каждого микрофона;
   - результаты классификации для каждого направления;
   - сводные отчёты;
   - метаданные, такие как временные метки, координаты местоположения устройства, а также аппаратные метрики (загрузка процессора, использование памяти).

   Передача данных осуществляется по защищённому протоколу *HTTPS*, что обеспечивает конфиденциальность и целостность информации. Для оптимизации сетевого трафика применяются алгоритмы сжатия: *FLAC* для аудиоданных и *gzip* для остальных типов данных.
2. **База данных.** Для хранения структурированных данных используется система управления базами данных *PostgreSQL*. В базе данных хранятся:
   - временные метки событий;
   - типы событий и вероятности предсказаний;
   - результаты классификации и ссылки на аудиофайлы в файловом хранилище.
3. **Хранилище аудиозаписей.** Аудиозаписи сохраняются в распределённой системе хранения данных *Ceph*, которая обеспечивает отказоустойчивость, масштабируемость и высокий уровень доступности данных. Использование *Ceph* позволяет обрабатывать большие объёмы информации, поступающей от нескольких устройств.
4. **Система мониторинга.** Для визуализации метрик работы системы используется *Grafana*. Мониторинг включает в себя:
   - загрузку серверов и устройств;
   - количество поступающих событий;
   - использование сетевого и дискового пространства.

Это позволяет в с низкой задержкой отслеживать производительность системы и быстро реагировать на возникновение узких мест или неисправностей.

Схема программного комплекса представлена на рисунке 4.7.

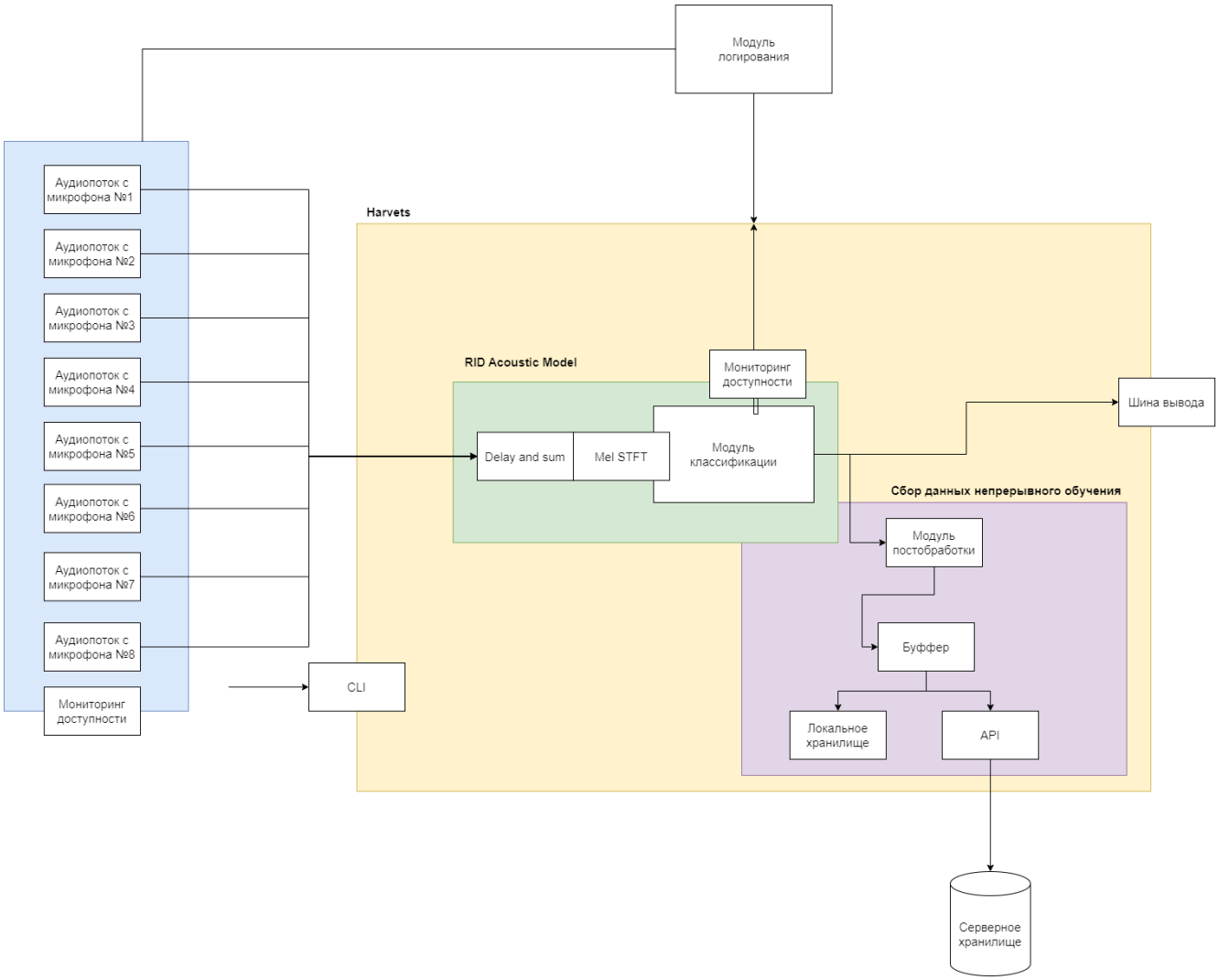


Рисунок 4.7 — Архитектура программного обеспечения цифровой обработки акустических данных дорожных событий.

В случае временной недоступности связи с сервером устройство реализует механизм локальной буферизации данных. Этот механизм позволяет сохранить все критически важные данные о событиях, произошедших в период отсутствия связи, обеспечивая их последующую передачу на сервер при восстановлении соединения. Буферизация выполнена в виде циклического лога, который сохраняет данные о последних событиях. Такой подход предотвращает потерю информации даже в условиях нестабильного сетевого соединения и минимизирует вероятность пропуска значимых событий.

Циклический лог представляет собой ограниченный по размеру буфер, в который записываются события в порядке их поступления. В случае переполнения буфера старые данные автоматически заменяются новыми. Это гарантирует, что устройство всегда сохраняет актуальную информацию, готовую к передаче на сервер.

Алгоритм работы механизма буферизации включает следующие этапы:

– **Запись данных о событиях.** В циклический лог записываются данные о каждом событии, включая аудиофайлы, результаты классификации, временные метки, координаты, аппаратные метрики и другие метаданные. Это обеспечивает полный набор информации для последующего анализа.
– **Контроль размера буфера.** При добавлении нового события система проверяет, достигнут ли предел размера буфера. В случае переполнения самые старые записи удаляются, чтобы освободить место для новых данных. Это позволяет устройству продолжать работу без нарушения функциональности.
– **Обнаружение восстановления связи.** Устройство периодически проверяет статус соединения с сервером. Как только соединение восстанавливается, система инициирует процесс передачи данных.
– **Автоматическая отправка данных.** Все данные, накопленные в буфере, автоматически отправляются на сервер. Перед отправкой используется сжатие данных (FLAC для аудио и gzip для метаданных), что позволяет минимизировать нагрузку на сеть.
– **Подтверждение передачи.** После успешной отправки сервер отправляет подтверждение. Получив его, устройство удаляет переданные данные из буфера, освобождая место для новых записей.

Такой подход обладает несколькими ключевыми преимуществами:

– **Надёжность.** Сохранение данных о событиях в условиях нестабильного соединения обеспечивает их полное восстановление после восстановления связи.
– **Непрерывность работы.** Алгоритм предотвращает прерывание работы устройства в случае длительных перебоев связи, позволяя ему продолжать сбор данных.

– **Оптимизация использования памяти.** Циклический лог эффективно управляет ограниченным объёмом памяти, гарантируя сохранение наиболее актуальной информации.
– **Эффективность передачи данных.** Использование алгоритмов сжатия снижает объём передаваемой информации, что особенно важно в условиях ограниченной пропускной способности сети.

Механизм локальной буферизации данных обеспечивает устойчивость системы к внешним факторам, таким как нестабильность сетевого соединения, и позволяет устройству сохранять высокую производительность и надёжность работы в реальных условиях. Этот подход делает систему более адаптивной и устойчивой, особенно в сценариях, где перебои связи неизбежны.

4.4 Бортовая система классификации акустических данных

Данный раздел посвящён системе классификации акустических событий. Здесь рассматриваются как теоретические аспекты работы алгоритмов, так и практические результаты испытаний.

Для проверки работоспособности алгоритма и оценки эффективности системы была проведена симуляция работы микрофонного массива в условиях получения акустического сигнала от нескольких источников. Расположение микрофонного массива и источников акустического сигнала (A и B) в одной из симуляций показано на рисунке 4.8.

В симуляции источником A служила аудиозапись аварии из тестовой выборки, а источником B – аудиозапись проезжающей фуры с включённым автомобильным сигналом. Осциллограммы этих акустческих сигналов представлены на рисунке 4.9.

Осциллограмма, снимаемая с одного микрофона, и осциллограммы аудиозаписей, полученных в результате работы алгоритма DAS для углов 0° и 90°, представлены на рисунке 4.10.

На рисунке 4.11 показано распределение вероятностей классов событий, предсказанных моделью на основе аудиозаписей, полученных после обработки алгоритмом DAS. Видно, что применение алгоритма DAS позволяет модели

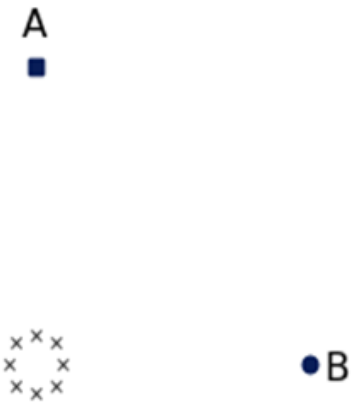
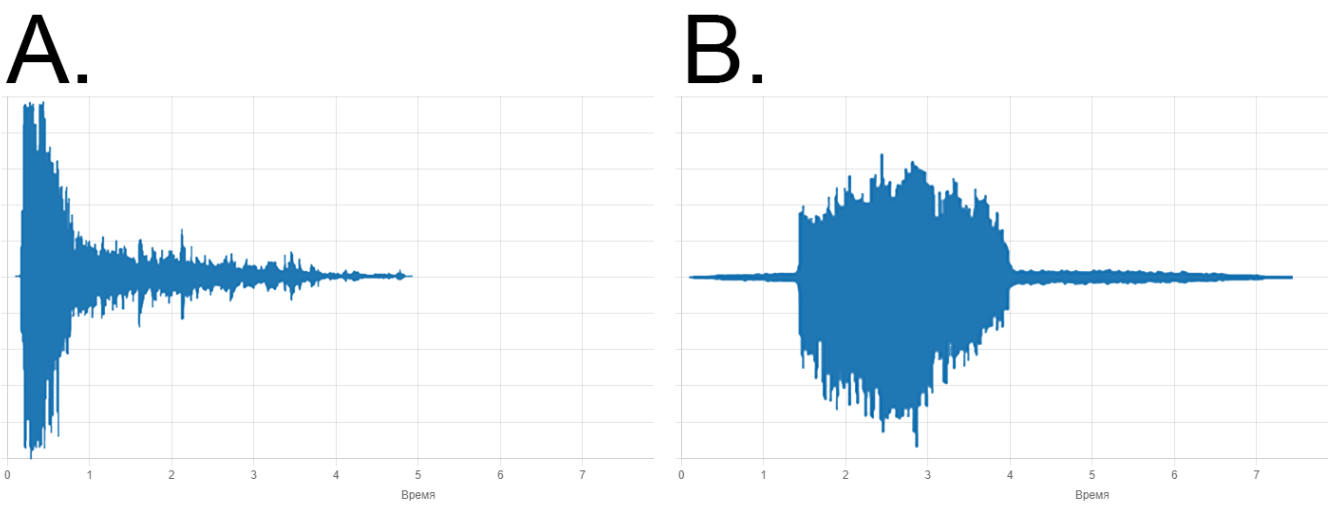


Рисунок 4.8 — Расположение микрофонного массива и источников акустического сигнала (масштаб изменён; расстояние между источниками и массивом составляет 40 м).

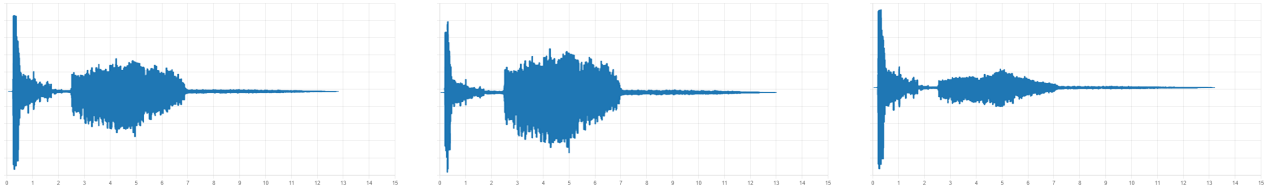


а) б)

Рисунок 4.9 — Осциллограммы акустических сигналов, воспроизводимых в источниках A и B .

повышать вероятность правильной классификации событий, источник которых находится в направлении обработки.

Для каждой аудиозаписи из тестового набора данных было проведено по две аналогичные симуляции для анализа влияния различных типов событий на точность классификации. В одной симуляции в источнике B воспроизводилась аудиозапись события другого класса из тестового набора данных, что моделировало ситуацию смешивания акустческих сигналов разных категорий, присутствующих в обучающем наборе. Во второй симуляции в источнике B



а) с первого микрофона б) для угла 0° в) для угла 90°

Рисунок 4.10 — Осциллограммы акустического сигнала до и после обработки алгоритмом DAS

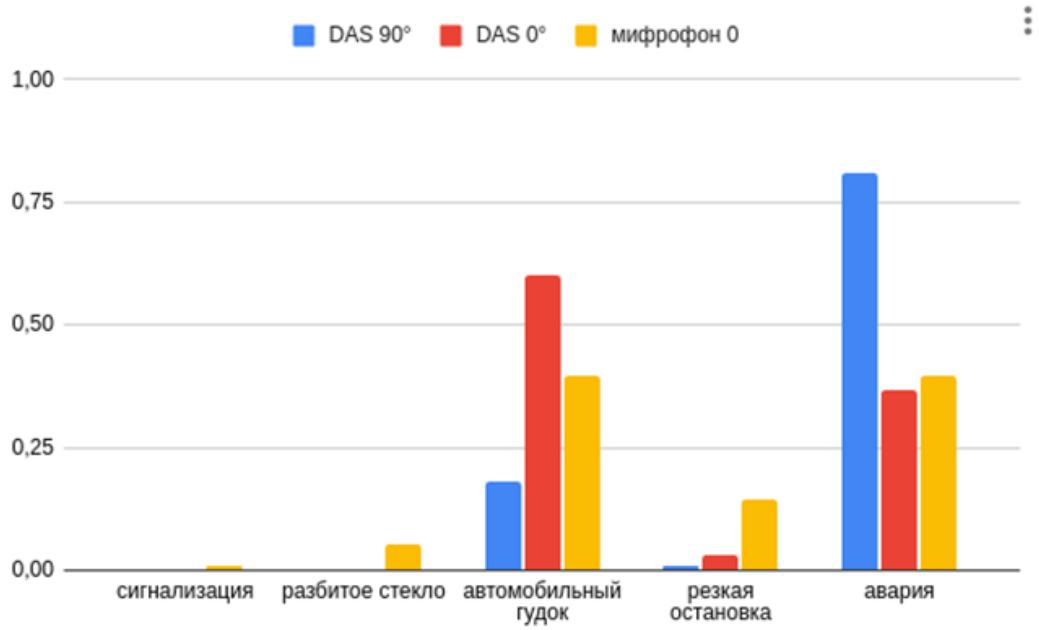


Рисунок 4.11 — Диаграмма распределения вероятностей классов событий для аудиозаписей, полученных алгоритмом DAS под углами 0° и 90°, а также для записи напрямую с первого микрофона.

воспроизводилась аудиозапись события, класс которого не входил в обучающий набор данных, имитируя ситуацию с появлением ранее неизвестного типа акустического сигнала. Это позволило оценить устойчивость модели к новым, неучтённым в обучении данным и её способность избегать ошибочных классификаций.

После проведения симуляций все собранные аудиоданные были обработаны алгоритмом DAS с настройкой на угол 90°. Алгоритм позволял сосредоточиться на акустическом сигнале, поступающем из заданного направления, тем самым минимизируя влияние фонового шума или акустческих сигналов из других направлений. Полученные результаты в виде направленных аудиозаписей были поданы на вход ранее обученной модели.

Для оценки работы системы на основе предсказанных классов была рассчитана метрика точности. Она определялась как доля правильно классифицированных аудиозаписей относительно общего числа записей в тестовом наборе данных.

Результаты экспериментов представлены в таблице 8. В ней указаны точность классификации для двух случаев: при использовании необработанных данных, снятых напрямую с первого микрофона, и для данных, обработанных алгоритмом DAS под углом 90°. Также отдельно выделены два типа событий в точке B: события, класс которых был представлен в обучающем наборе, и события, относящиеся к неизвестным для модели категориям.

Расширенный анализ показал, что использование алгоритма DAS существенно улучшает точность классификации, особенно в случае смешанных акустических событий. Это объясняется способностью алгоритма выделять акустические сигналы из заданного направления, что снижает влияние шумов и пересекающихся сигналов. Проведённые симуляции подтверждают эффективность предлагаемого подхода для задач классификации акустических событий в условиях сложной акустической обстановки.

Таблица 8 — Точность классификации аудиозаписей в симуляции.

Предобработка	**Напрямую с микрофона**		**DAS** (90°)	
Тип аудиозаписи в точке B	**1**	**2**	**1**	**2**
Точность	0.516	0.874	0.853	0.921

Полученные результаты показывают, что применение алгоритма DAS помогает нейронной модели более точно классифицировать события, происходящие в интересующем направлении. Особенно заметна разница в случае без использования алгоритма DAS, когда в точке B воспроизводилась аудиозапись класса, присутствующего в обучающей выборке. В таких случаях две аудиозаписи разных классов смешивались, что затрудняло однозначную классификацию.

Для проведения практических испытаний в реальных условиях была организована поездка по тестовому маршруту. Маршрут протяжённостью 30 км проходил в пределах города Москвы, в северо-восточной части. Карта маршрута представлена на рисунке 4.12.

Маршрут включал различные дорожные условия: тихие улицы, загруженные перекрёстки, участки с интенсивным движением и зоны строительных

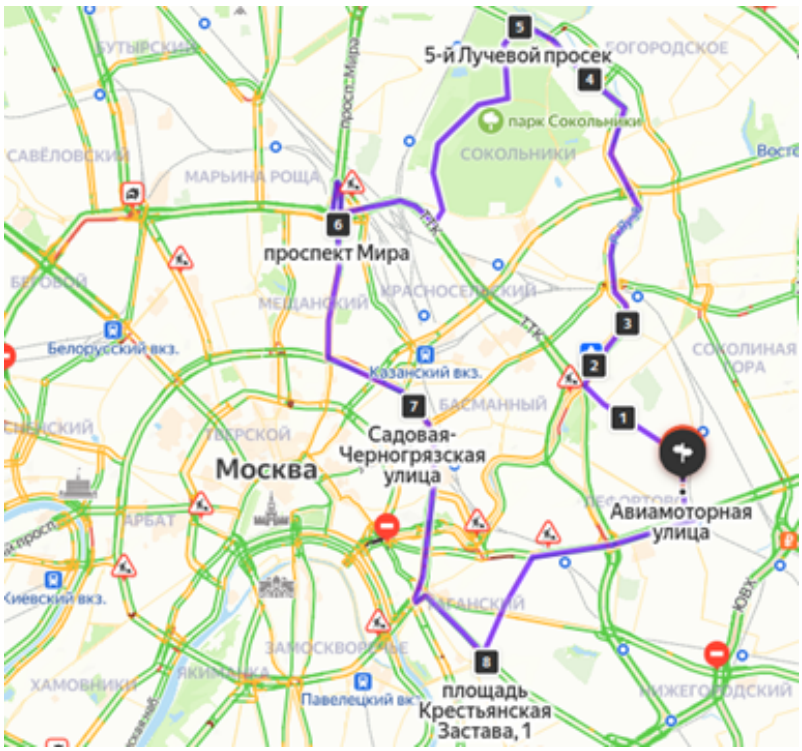


Рисунок 4.12 — Карта маршрута практических испытаний.

работ. Целью испытаний была проверка работоспособности детектора в реальных условиях городской среды.

По итогам испытаний была рассчитана точность классификации по каждому классу событий. Результаты представлены в таблице 9.

Таблица 9 — Точность классификации аудиозаписей в ходе практических испытаний.

Класс событий	**Количество событий**	**Точность**
Сигнализация	3	1.0
Разбитое стекло	0	-
Автомобильный гудок	21	0.85
Резкая остановка	6	0.66
Авария	0	-

Следует отметить, что некоторые классы событий, такие как «авария» и «разбитое стекло», не были зафиксированы в ходе испытаний, что объясняется их редкостью в повседневной городской среде в течение ограниченного времени тестирования.

Все аудиозаписи и данные о классификациях сохранялись в системе сбора данных, что позволило выявить слабые места классификатора и подготовить данные для дальнейшего дообучения модели.

Практические испытания подтвердили работоспособность системы в реальных условиях и продемонстрировали её потенциал для применения в задачах мониторинга и оперативного реагирования на акустические события в городской среде.

Эффективная работа программно-аппаратного комплекса в реальном времени оценивается по трём основным параметрам: латентности, джиттеру и соблюдению мягкого дедлайна (250мс). Для этого экспериментально замеряли время выполнения каждого шага обработки (предобработки, инференса и постобработки).

На практике эти параметры были измерены на разработанном программно-аппаратном комплексе с помощью замеров времени выполнения каждой операции. Полученные данные использовались для расчёта средней латентности, вариативности времени обработки (джиттера) и оценки соответствия мягкому дедлайну.

Латентность (T_{total}) — это общее время задержки, которое проходит от поступления сигнала до получения результата. Она рассчитывается как сумма времён всех этапов обработки:

$$T_{\text{total}} = T_{\text{preprocessing}} + T_{\text{inference}} + T_{\text{postprocessing}}. \tag{4.7}$$

Джиттер (J) характеризует вариативность задержек и рассчитывается как стандартное отклонение времён обработки:

$$J = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_i - \overline{T})^2}, \tag{4.8}$$

где T_i — время обработки в i-м цикле, $\overline{T}$ — среднее время обработки.

Дедлайн в системе установлен на уровне 250 мс. Система должна работать в режиме реального времени, чтобы вовремя предупреждать о потенциально опасных ситуациях. В рамках данного исследования установлено, что дедлайн на уровне 250 мс является оптимальным балансом между вычислительной сложностью анализа и скоростью оповещения водителя. Ниже приведены основные аргументы в пользу такого выбора.

1. **Время реакции водителя.** Среднее время реакции водителя на внештатные дорожные события зачастую превышает 700–1000 мс. Задержка в 250 мс позволяет системе сформировать сигнал предупреждения

задолго до того, как водитель предпримет физические действия (нажмёт на педаль тормоза, повернёт руль). Таким образом, даже четверть секунды является достаточно «быстрой» для предупреждения, не снижая пользу системы.

2. **Характер дорожных событий.** Большинство опасных ситуаций на дороге развивается не мгновенно, а в течение нескольких секунд (приближение спецтранспорта, появление сигнала от дальнего автомобиля, посторонний шум, указывающий на неисправность). При дедлайне в 250 мс система успевает адекватно обработать входящий акустический поток, классифицировать событие и сформировать своевременное оповещение.
3. **Стабильная работа при пиковых нагрузках.** Городской трафик часто характеризуется пиковыми ситуациями, когда одновременно присутствуют несколько источников шума: сигналы от нескольких машин, шум двигателя, дорожные работы и т. д. При слишком коротких ограничениях система могла бы пропускать важные события или неверно классифицировать сигналы. Дедлайн в 250 мс допускает кратковременные пиковые нагрузки, обеспечивая качественный анализ без постоянных «проседаний» по времени.

Таким образом, установка дедлайна в 250 мс основана на реальных дорожных сценариях, физиологических характеристиках (время реакции водителя), потребности в «глубоком» анализе акустческих сигналов где задержка порядка нескольких сотен миллисекунд считается приемлемой. Система в этих условиях способна эффективно обнаруживать критические события и своевременно информировать водителя, сохраняя высокий уровень точности и надёжности обработки.

Поскольку система работает в условиях мягкого дедлайна, допустимо превышение целевого времени обработки в редких случаях, но с ограничениями:

– Средняя латентность должна быть значительно ниже 250 мс, чтобы оставить запас для возможных нарушений.

В ходе практики были выполнены замеры времён обработки для каждого этапа:

– **Предобработка**:
  Замеры (мс): [56.4, 56.7, 57.0, 56.3, 56.8, 56.5, 57.1]

Среднее время: $\overline{T_{\text{preprocessing}}} = 56.64\,\text{мс}$

Джиттер: $J_{\text{preprocessing}} \approx 0.31\,\text{мс}$

– **Классификация (MobileNetV3)**:

Замеры (мс): [159.7, 160.2, 159.8, 160.0, 159.9, 160.1, 159.5]

Среднее время: $\overline{T_{\text{inference}}} = 159.93\,\text{мс}$

Джиттер: $J_{\text{inference}} \approx 0.25\,\text{мс}$

– **Постобработка**:

Замеры (мс): [7.40, 7.60, 7.50, 7.58, 7.45, 7.60, 7.52]

Среднее время: $\overline{T_{\text{postprocessing}}} = 7.55\,\text{мс}$

Джиттер: $J_{\text{postprocessing}} \approx 0.08\,\text{мс}$

Общая латентность и джиттер:

$$T_{\text{total}} = 56.64 + 159.93 + 7.55 = 224.12\,\text{мс},$$

$$J_{\text{total}} \approx 0.31 + 0.25 + 0.08 = 0.64\,\text{мс}.$$

Рассчёты показали следующие выводы:

– Средняя латентность системы составляет 224.12 мс, что заметно ниже целевого дедлайна в 250 мс.

– Общий джиттер равен 0.64 мс, что свидетельствует о достаточной стабильности обработки данных на каждом этапе.

Поскольку система работает в условиях мягкого дедлайна, редкие превышения времени обработки допустимы. Это позволяет обрабатывать критические события (например, аварии или сирены) даже при пиковых нагрузках, сохраняя надёжность и оперативность. Учет этих характеристик подтверждает, что система способна эффективно функционировать в реальном времени, укладываясь в установленные требования к латентности.

4.5 Выводы по главе

В данной главе была решена задача № 6 диссертации, а именно разработка сбора и цифровой обработки дорожных событий.

В данной главе была рассмотрена двухкомпонентная структура программно-аппаратного комплекса для детектирования и классификации акустических событий на дороге. Первая часть посвящена системе сбора

акустических данных, включающей круговой микрофонный массив, аудиоинтерфейс и протоколы синхронизации. Вторая часть описывает систему классификации и постобработки, включающую алгоритмы beamforming (Delay And Sum), нейросетевую модель, механизмы локальной буферизации и серверную инфраструктуру.

Основные результаты и выводы:

- Выбор круговой конфигурации микрофонного массива (8 микрофонов) позволил добиться равномерного охвата по азимуту и гибко управлять задержками сигналов.
- Реализация алгоритма Delay And Sum продемонстрировала эффективность при выделении акустических событий из заданного направления, что повысило точность работы классификатора.
- Использование заранее обученной нейросети (с дистилляцией знаний) дало возможность обрабатывать несколько пространственных направлений акустического сигнала в режиме низкой задержки.
- Разработанная серверная инфраструктура обеспечивает отказоустойчивое хранение данных, их сжатие и мониторинг производительности.
- Теоретические симуляции и практические выезды показали, что комплекс корректно определяет ряд дорожных событий (сигнализация, автомобильный гудок и др.) даже при высоком уровне шумов.

Перспективы дальнейших исследований включают:

- Расширение набора классифицируемых акустческих сигналов (сирены, скрежет и другие специфические шумы);
- Исследование более сложных beamforming-алгоритмов и интеграцию с другими сенсорными системами (видео, лидары);
- Разработку методов самообучения модели на новых данных без полного переобучения;
- Оценку масштабирования системы для мониторинга на больших территориях или автопарках.

Полученные результаты свидетельствуют о том, что разработанная система может стать основой для реализации продвинутых ADAS-компонентов и систем акустического мониторинга в городской среде, повышая безопасность и комфорт дорожного движения.

Заключение

В рамках данного исследования была разработана и реализована комплексная система акустического обнаружения и классификации акустических событий для применения в городской среде. Работа охватила широкий спектр задач: от сбора и обработки акустических данных до разработки и оптимизации алгоритмов машинного обучения и их практического применения в реальных условиях.

Ключевые достижения исследования включают:

1. Изучены методы классификации акустических данных с целью повышения безопасности движения транспортных средств. Проведён глубокий анализ существующих алгоритмов, таких как методы на основе свёрточных нейронных сетей, архитектуры на базе трансформеров и традиционные статистические подходы. Это позволило выявить преимущества и ограничения каждого метода и выбрать наиболее эффективные подходы для дальнейших исследований. **(Задача № 1)**
2. Предложен метод сбора и аннотирования акустической информации дорожных событий, позволяющий повысить эффективность подготовки набора данных и минимизировать влияние человеческого фактора, что достигается за счёт использования предобученной модели, исключающей вероятность пропуска событий из-за человеческой невнимательности или утомляемости. **(Задача № 2)**
3. Разработана система сбора и аннотирования данных о дорожных событиях в реальных условиях городской среды, что позволило сформировать уникальный набор данных. Система включает микрофонную решётку из 8 микрофонов, размещённых по периметру транспортного средства, с обеспечением 360° покрытия камеры и оборудование для синхронизации данных. Также была создана специальная система разметки данных, основанная на предобученной модели BEATs, что минимизировало вероятность пропуска событий из-за человеческого фактора. Набор данных, включающий 5 классов акустических сигналов и состоящий из 2600 сэмплов, стал основой для выявления новых закономерностей в аудиоданных и значительного повышения точности классификации. Внедрение инструмента LabelTool позволило

сократить время разметки почти вдвое и повысить качество аннотаций. **(Задача № 3, Положения № 1, 2)**

4. Разработано алгоритмическое обеспечение нейросетевой обработки акустических данных, включающее алгоритм устойчивого обучения нейронной сети и алгоритм классификации акустических данных дорожных событий. Алгоритм устойчивого обучения нейронной сети учитывает выбросы и шумы в данных. Применение робастных функций потерь, таких как функции Хьюбера и биквадратная функция Тьюки, обеспечило устойчивость моделей к аномалиям. Использование метода дистилляции знаний позволило значительно уменьшить размер модели с 90,3 млн до 0,19 млн параметров, сохранив точность классификации(92%). Это открывает возможность применения модели на устройствах с ограниченными вычислительными ресурсами, таких как транспортные системы. Разработанный алгоритм классификации акустических данных дорожных событий с применением модификаций KAN, обеспечивает точность не менее 95% в условиях городской среды. Алгоритм демонстрирует устойчивость к шумам и сложным акустическим условиям благодаря использованию современных нейронных сетей на основе архитектур трансформеров. Проведён сравнительный анализ архитектур нейронных сетей, в результате которого последняя модель показала наилучшую точность на собранном наборе данных. **(Задача № 4,5, Положение № 3,4)**
5. Разработана архитектура и реализован комплекс для сбора и цифровой обработки акустических данных, включающий специализированное оборудование и программное обеспечение. Реализация метода Delay And Sum (DAS) для формирования направленной диаграммы чувствительности микрофонной решётки позволила локализовать источники акустического сигнала. Оптимизация алгоритмов обработки в режиме низкой задержки обеспечила непрерывность классификации и возможность обработки нескольких направлений одновременно. Разработанная архитектура программно-аппаратного комплекса включает сервер приёма данных, базу данных PostgreSQL и распределённое хранилище аудиозаписей на базе Ceph. Практические испытания комплекса в городских условиях подтвердили его эффективность: точность

классификации ряда акустических событий достигла не менее 95%. **(Задачи № 6, Положение № 5)**

В заключение можно сказать, что проведённое исследование вносит значительный вклад в развитие технологий акустического мониторинга и открывает новые возможности для создания интеллектуальных систем управления и безопасности. Разработанная система демонстрирует высокий потенциал для практического применения и развития, что может существенно повлиять на уровень безопасности и эффективность управления городской инфраструктурой.

Предложенные методы и средства, полученные в рамках диссертационного исследования создают основу для дальнейших инноваций в области акустического анализа и мониторинга . Объединение технологий машинного обучения, обработки сигналов и распределённых вычислений открывает перспективы для создания более умных, безопасных и экологически устойчивых городских пространств будущего.

Список литературы

1. *Serban, A.* A Standard Driven Software Architecture for Fully Autonomous Vehicles [Текст] / A. Serban, E. Poll, J. Visser // Journal of Automotive Software Engineering. — 2020. — Янв. — Т. 1.
2. Composition and Application of Current Advanced Driving Assistance System: A Review [Текст] / X. Li [и др.]. — 2021. — arXiv: 2105.12348 [cs.AI]. — URL: https://arxiv.org/abs/2105.12348.
3. *Li, Y.* Emergent Visual Sensors for Autonomous Vehicles [Текст] / Y. Li, J. Moreau, J. Ibanez-Guzman. — 2023. — arXiv: 2205.09383 [cs.CV]. — URL: https://arxiv.org/abs/2205.09383.
4. Adverse Weather Conditions in the Validation of ADAS/AD Sensors [Текст] / G. Schwab [и др.] // ATZelectronics worldwide. — 2022. — Февр. — Т. 17. — С. 46—49.
5. Improved Vehicle Sub-type Classification for Acoustic Traffic Monitoring [Текст] / M. Ashhad [и др.]. — 2023. — arXiv: 2302.02945 [cs.SD]. — URL: https://arxiv.org/abs/2302.02945.
6. Acoustic Scene Classification: Classifying environments from the sounds they produce [Текст] / D. Barchiesi [и др.] // IEEE Signal Processing Magazine. — 2015. — Май. — Т. 32, № 3. — С. 16—34. — URL: http://dx.doi.org/10.1109/MSP.2014.2326181.
7. Deep semantic learning for acoustic scene classification [Текст] / Y. Shao, X. Ma, Y. Ma [и др.] // Journal of Audio, Speech, and Music Processing. — 2024. — Т. 1. — С. 1—2024. — URL: https://doi.org/10.1186/s13636-023-00323-5.
8. *Sunu, J.* Unsupervised vehicle recognition using incremental reseeding of acoustic signatures [Текст] / J. Sunu, B. Hunter, A. G. Percus. — 2018. — arXiv: 1802.06287 [stat.ML]. — URL: https://arxiv.org/abs/1802.06287.
9. *Sunu, J.* Dimensionality reduction for acoustic vehicle classification with spectral embedding [Текст] / J. Sunu, A. G. Percus. — 2018. — arXiv: 1705.09869 [stat.ML]. — URL: https://arxiv.org/abs/1705.09869.

10. *Toffa, O. K.* Environmental Sound Classification Using Local Binary Pattern and Audio Features Collaboration [Текст] / O. K. Toffa, M. Mignotte // IEEE Transactions on Multimedia. — 2021. — Т. 23. — С. 3978—3985.

11. An Ensemble of Convolutional Neural Networks for Audio Classification [Текст] / L. Nanni [и др.] // ArXiv. — 2020. — Т. abs/2007.07966.

12. *Zhao, W.* Environmental sound classification based on pitch shifting [Текст] / W. Zhao, B. Yin // 2022 International Seminar on Computer Science and Engineering Technology (SCSET). — 2022. — С. 275—280.

13. *Zhang, Y.* The Classification of Environmental Audio with Ensemble Learning [Текст] / Y. Zhang, D. jv Lv, Y. Lin // Proceedings of the 2013 International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013). — Atlantis Press, 2013/08. — С. 368—371. — URL: https://doi.org/10.2991/icacsei.2013.93.

14. A Survey of Audio Classification Using Deep Learning [Текст] / K. Zaman [и др.] // IEEE Access. — 2023. — Т. 11. — С. 106620—106649.

15. *Abeßer, J.* A Review of Deep Learning Based Methods for Acoustic Scene Classification [Текст] / J. Abeßer // Applied Sciences. — 2020. — Т. 10, № 6. — URL: https://www.mdpi.com/2076-3417/10/6/2020.

16. *McAdams, S.* The Perceptual Representation of Timbre [Текст] / S. McAdams // Timbre: Acoustics, Perception, and Cognition / под ред. K. Siedenburg [и др.]. — Cham : Springer International Publishing, 2019. — С. 23—57. — URL: https://doi.org/10.1007/978-3-030-14832-4_2.

17. *Kiktova, E.* Feature selection for acoustic events detection [Текст] / E. Kiktova, J. Juhár, A. Čižmár // Multimedia Tools and Applications. — 2013. — Июнь. — Т. 74.

18. *Yiming, S.* Voice Activity Detection Based on the Improved Dual-Threshold Method [Текст] / S. Yiming, W. Rui // 2015 International Conference on Intelligent Transportation, Big Data and Smart City. — 2015. — С. 996—999.

19. *VOCAL Technologies Ltd.* Voice Activity Detection with Adaptive Thresholding [Текст] / VOCAL Technologies Ltd. — n.d. — URL: https://vocal.com/voice-quality-enhancement/voice-activity-detection-with-adaptive-thresholding/.

20. *Amin, T. B.* Speech Recognition using Dynamic Time Warping [Текст] / T. B. Amin, I. Mahmood // 2008 2nd International Conference on Advances in Space Technologies. — 2008. — С. 74—79.

21. Gradient-based learning applied to document recognition [Текст] / Y. LeCun [и др.] // Proceedings of the IEEE. — 1998. — Т. 86, № 11. — С. 2278—2324.

22. *Rumelhart, D. E.* Learning representations by back-propagating errors [Текст] / D. E. Rumelhart, G. E. Hinton, R. J. Williams // Nature. — 1986. — Т. 323, № 6088. — С. 533—536.

23. *Sabour, S.* Dynamic Routing Between Capsules [Текст] / S. Sabour, N. Frosst, G. E. Hinton // Advances in neural information processing systems. — 2017. — Т. 30. — С. 3856—3866.

24. Attention is all you need [Текст] / A. Vaswani [и др.] // Advances in neural information processing systems. — 2017. — С. 5998—6008.

25. *Mushtaq, Z.* Environmental sound classification using a regularized deep convolutional neural network with data augmentation [Текст] / Z. Mushtaq, S. Su // Applied Acoustics. — 2020.

26. *Sharma, R.* Listening to the Environment: Applying Deep Learning Techniques for Robust Environmental Sound Classification [Текст] / R. Sharma, M. Nagpal // 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT). — 2024. — Т. 1. — С. 1012—1016.

27. Environmental Sound Classification Based on Continual Learning [Текст] / Y. Sun [и др.] // 2023 International Conference on New Trends in Computational Intelligence (NTCI). — 2023. — Т. 1. — С. 155—159.

28. Classifying environmental sounds using image recognition networks [Текст] / V. Boddapati [и др.] // Procedia Computer Science. — 2017. — Т. 112. — С. 2048—2056. — URL: https://www.sciencedirect.com/science/article/pii/S1877050917316599 ; Knowledge-Based and Intelligent Information Engineering Systems: Proceedings of the 21st International Conference, KES-20176-8 September 2017, Marseille, France.

29. Sound Source Direction of Arrival Estimation for Autonomous Driving Applications [Текст] / Y. Furletov [и др.] //. — 11.2022. — С. 1—5.

30. *Lezhenin, I.* Urban Sound Classification using Long Short-Term Memory Neural Network [Текст] / I. Lezhenin, N. Bogach, E. Pyshkin //. — 09.2019. — С. 57—60.

31. *Abdoli, S.* End-to-End Environmental Sound Classification using a 1D Convolutional Neural Network [Текст] / S. Abdoli, P. Cardinal, A. L. Koerich. — 2019. — arXiv: 1904.08990 [cs.SD]. — URL: https://arxiv.org/abs/1904.08990.

32. *Hameed Jaid, U.* End-to-End Speaker Profiling Using 1D CNN Architectures and Filter Bank Initialization [Текст] / U. Hameed Jaid, A. Karim // International Journal of Online and Biomedical Engineering (iJOE). — 2023. — Авг. — Т. 19. — С. 65—81.

33. *Zabidi, M.* Fowl Play: Identifying Birds by Bioacoustics and Deep Learning [Текст] / M. Zabidi. — 05.2023.

34. ESResNet: Environmental Sound Classification Based on Visual Domain Models [Текст] / A. Guzhov [и др.] // 2020 25th International Conference on Pattern Recognition (ICPR). — 2020. — С. 4933—4940. — URL: https://api.semanticscholar.org/CorpusID:215786556.

35. ImageNet Large Scale Visual Recognition Challenge [Текст] / O. Russakovsky [и др.]. — 2015. — arXiv: 1409.0575 [cs.CV]. — URL: https://arxiv.org/abs/1409.0575.

36. *Salamon, J.* A Dataset and Taxonomy for Urban Sound Research [Текст] / J. Salamon, C. Jacoby, J. P. Bello // 22nd ACM International Conference on Multimedia (ACM-MM'14). — Orlando, FL, USA, 11.2014. — С. 1041—1044.

37. ESResNeXt-fbsp: Learning Robust Time-Frequency Transformation of Audio [Текст] / A. Guzhov [и др.] // 2021 International Joint Conference on Neural Networks (IJCNN). — IEEE. 2021. — С. 1—8.

38. Aggregated Residual Transformations for Deep Neural Networks [Текст] / S. Xie [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — С. 1492—1500.

39. Unsupervised Discriminative Learning of Sounds for Audio Event Classification [Текст] / S. Hornauer [и др.] // ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). — IEEE. 2021. — С. 3035—3039.

40. *Nasiri, A.* SoundCLR: Contrastive Learning of Representations For Improved Environmental Sound Classification [Текст] / A. Nasiri, J. Hu. — 2021. — arXiv: 2103.01929 [eess.AS]. — arXiv preprint arXiv:2103.01929.

41. *Gong, Y.* AST: Audio Spectrogram Transformer [Текст] / Y. Gong, Y.-A. Chung, J. Glass. — 2021. — arXiv: 2104.01778 [cs.SD]. — URL: https://arxiv.org/abs/2104.01778.

42. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [Текст] / A. Dosovitskiy [и др.]. — 2021. — arXiv: 2010.11929 [cs.CV]. — URL: https://arxiv.org/abs/2010.11929.

43. *Wang, Y.* What Do Position Embeddings Learn? An Empirical Study of Pre-Trained Language Model Positional Encoding [Текст] / Y. Wang, Y. Kim, A. Rush. — 2020. — arXiv: 2010.04903 [cs.CL]. — URL: https://arxiv.org/abs/2010.04903.

44. Efficient Training of Audio Transformers with Patchout [Текст] / K. Koutini [и др.]. — 2021. — arXiv: 2110.05069 [cs.SD]. — URL: https://arxiv.org/abs/2110.05069.

45. Efficient Training of Audio Transformers with Patchout [Текст] / K. Koutini [и др.]. — 2021. — arXiv: 2110.05069 [cs.SD]. — arXiv:2110.05069.

46. mixup: Beyond Empirical Risk Minimization [Текст] / H. Zhang [и др.]. — 2017. — arXiv: 1710.09412 [cs.LG]. — arXiv:1710.09412.

47. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition [Текст] / D. S. Park [и др.]. — 2019. — arXiv: 1904.08779 [eess.AS]. — arXiv:1904.08779.

48. *Alonso-Jiménez, P.* Efficient Supervised Training of Audio Transformers for Music Representation Learning [Текст] / P. Alonso-Jiménez, X. Serra, D. Bogdanov // ISMIR 2023 Hybrid Conference. — 2023.

49. Hint-dynamic Knowledge Distillation [Текст] / Y. Liu [и др.]. — 2022. — arXiv: 2211.17059 [cs.LG]. — arXiv:2211.17059.

50. *Schmid, F.* Dynamic Convolutional Neural Networks as Efficient Pretrained Audio Models [Текст] / F. Schmid, K. Koutini, G. Widmer. — 2023. — arXiv: 2310.15648 [cs.SD]. — arXiv:2310.15648.

51. *Chia, Y. K.* Transformer to CNN: Label-scarce Distillation for Efficient Text Classification [Текст] / Y. K. Chia, S. Witteveen, M. Andrews. — 2019. — arXiv: 1909.03508 [cs.CL]. — arXiv:1909.03508.

52. Searching for MobileNetV3 [Текст] / A. Howard [и др.] // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2019. — С. 1314—1324.

53. *Eargle, J.* Audio Engineering for Sound Reinforcement [Текст] / J. Eargle, C. Foreman. — Springer, 2015.

54. *Ballou, G.* Handbook for Sound Engineers [Текст] / G. Ballou. — Taylor & Francis, 2008.

55. *Rossing, T. D.* Springer Handbook of Acoustics [Текст] / T. D. Rossing. — Springer, 2007.

56. *Piczak, K. J.* ESC: Dataset for Environmental Sound Classification [Текст] / K. J. Piczak // Proceedings of the 23rd ACM International Conference on Multimedia. — 2015. — С. 1015—1018.

57. *Salamon, J.* A Dataset and Taxonomy for Urban Sound Research [Текст] / J. Salamon, C. Jacoby, J. P. Bello // Proceedings of the 22nd ACM International Conference on Multimedia. — 2014. — С. 1041—1044.

58. FSD50K: An Open Dataset of Human-Labeled Sound Events [Текст] / E. Fonseca [и др.] // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2021. — Т. 30. — С. 829—852.

59. Audio Set: An Ontology and Human-Labeled Dataset for Audio Events [Текст] / J. F. Gemmeke [и др.] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). — 2017. — С. 776—780.

60. *Hollosi, D.* Classification of Environmental Sounds Using Time-Domain Features and Supervised Learning [Текст] / D. Hollosi, S. Sigg, G. Tröster // IEEE Transactions on Audio, Speech, and Language Processing. — 2012.

61. *Wang, D.* Computational Auditory Scene Analysis: Principles, Algorithms, and Applications [Текст] / D. Wang, G. J. Brown. — Wiley-IEEE Press, 2006.

62. High-Quality, Low-Delay Music Coding in the Opus Codec [Текст] / J.-M. Valin [и др.]. — 2016. — arXiv: 1602.04845 [cs.MM]. — URL: https://arxiv.org/abs/1602.04845.

63. Deep learning for audio signal processing [Текст] / H. Purwins [и др.] // IEEE Journal of Selected Topics in Signal Processing. — 2019. — Т. 13, № 2. — С. 206—219.

64. *Logan, B.* Mel frequency cepstral coefficients for music modeling [Текст] / B. Logan // Proceedings of ISMIR. — 2000. — Т. 2000. — С. 1—11.

65. Convolutional networks for images, speech, and time series [Текст] / Y. LeCun [и др.] // The handbook of brain theory and neural networks. — 1995. — Т. 3361. — С. 255—257.

66. CNN architectures for large-scale audio classification [Текст] / S. Hershey [и др.] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). — 2017. — С. 131—135.

67. Deep residual learning for image recognition [Текст] / K. He [и др.] // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2016. — С. 770—778.

68. Learning transferable features with deep adaptation networks [Текст] / M. Long [и др.] // International Conference on Machine Learning (ICML). — 2015. — С. 97—105.

69. An image is worth 16x16 words: Transformers for image recognition at scale [Текст] / A. Dosovitskiy [и др.] // arXiv preprint arXiv:2010.11929. — 2020.

70. *Palanisamy, K.* BEATs: A Bidirectional Encoder from Audio Transformers for Audio Understanding [Текст] / K. Palanisamy, R. Das, R. Krishnan // ArXiv. — 2022. — Т. abs/2203.00041.

71. *Hinton, G.* Distilling the Knowledge in a Neural Network [Текст] / G. Hinton, O. Vinyals, J. Dean. — 2015. — arXiv: 1503.02531 [stat.ML]. — URL: https://arxiv.org/abs/1503.02531.

72. Knowledge Distillation: A Survey [Текст] / J. Gou [и др.] // International Journal of Computer Vision. — 2021. — Март. — Т. 129, № 6. — С. 1789—1819. — URL: http://dx.doi.org/10.1007/s11263-021-01453-z.

73. Knowledge Distillation from A Stronger Teacher [Текст] / T. Huang [и др.] // ArXiv. — 2022. — Т. abs/2205.10536.

74. Knowledge Distillation via Multi-Teacher Feature Ensemble [Текст] / X. Ye [и др.] // IEEE Signal Processing Letters. — 2024. — Т. 31. — С. 566—570.

75. Multilevel Attention-Based Sample Correlations for Knowledge Distillation [Текст] / J. Gou [и др.] // IEEE Transactions on Industrial Informatics. — 2023. — Т. 19. — С. 7099—7109.

76. *Xie, Z.* Throughput-oriented and Accuracy-aware DNN Training with BFloat16 on GPU [Текст] / Z. Xie, S. Raskar, M. Emani // 2022 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). — 2022. — С. 1084—1087.

77. TutorNet: Towards Flexible Knowledge Distillation for End-to-End Speech Recognition [Текст] / J. W. Yoon [и др.] // IEEE/ACM Transactions on Audio, Speech, and Language Processing. — 2021. — Т. 29. — С. 1626—1638.

78. DTCNet: Transformer-CNN Distillation for Super-Resolution of Remote Sensing Image [Текст] / C. Lin [и др.] // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. — 2024. — Т. 17. — С. 11117—11133.

79. Robust Optimization for Deep Regression [Текст] / V. Belagiannis [и др.]. — 2015. — arXiv: 1505.06606 [cs.CV]. — URL: https://arxiv.org/abs/1505.06606.

80. *Mlotshwa, T.* Cauchy Loss Function: Robustness Under Gaussian and Cauchy Noise [Текст] / T. Mlotshwa, H. van Deventer, A. S. Bosman. — 2023. — arXiv: 2302.07238 [cs.LG]. — URL: https://arxiv.org/abs/2302.07238.

81. *Barron, J. T.* A General and Adaptive Robust Loss Function [Текст] / J. T. Barron. — 2019. — arXiv: 1701.03077 [cs.CV]. — URL: https://arxiv.org/abs/1701.03077.

82. *Айвазян, С. А.* Прикладная статистика. Исследование зависимостей: справочное издание [Текст] / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин ; под ред. С. А. Айвазян. — Москва : Финансы и статистика, 1985. — С. 487. — Библиогр.: с. 459–471.

83. *Huber, P. J.* Robust Estimation of a Location Parameter [Текст] / P. J. Huber // The Annals of Mathematical Statistics. — 1964. — Т. 35, № 1. — С. 73—101. — URL: https://doi.org/10.1214/aoms/1177703732.

84. *Rukhin, A. L.* Loss Functions for Loss Estimation [Текст] / A. L. Rukhin // The Annals of Statistics. — 1988. — Т. 16, № 3. — С. 1262—1269. — URL: https://doi.org/10.1214/aos/1176350960.

85. *Chatelain, J.-B.* Wealth in the quadratic loss function of the Ramsey-Malinvaud-Cass-Koopmans model of optimal savings [Текст] / J.-B. Chatelain, K. Ralf // Revue d'économie politique. — 2024. — Т. 134, № 3. — С. 371—390.

86. Laplacian Welsch Regularization for Robust Semisupervised Learning [Текст] / J. Ke [и др.] // IEEE Transactions on Cybernetics. — 2022. — Т. 52, № 1. — С. 164—177.

87. *Karpov, N.* Golos: Russian Dataset for Speech Research [Текст] / N. Karpov, A. Denisenko, F. Minkin // Proc. Interspeech 2021. — 2021. — С. 1419—1423.

88. Audio Set: An ontology and human-labeled dataset for audio events [Текст] / J. F. Gemmeke [и др.] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). — 2017. — С. 776—780.

89. *Loshchilov, I.* Decoupled Weight Decay Regularization [Текст] / I. Loshchilov, F. Hutter. — 2019. — arXiv: 1711.05101 [cs.LG]. — URL: https://arxiv.org/abs/1711.05101.

90. KAN: Kolmogorov-Arnold Networks [Текст] / Z. Liu [и др.] // arXiv preprint arXiv:2404.19756. — 2024. — URL: https://arxiv.org/abs/2404.19756.

91. *Колмогоров, А. Н.* О представлении непрерывных функций нескольких переменных суперпозицией функций одной переменной и сложения [Текст] / А. Н. Колмогоров, В. И. Арнольд // Доклады АН СССР. — 1957. — Т. 114, № 5. — С. 953—956. — Оригинальная работа, в которой излагается основа теоремы Колмогорова-Арнольда.

92. *de Boor, C.* On calculating with B-splines [Текст] / C. de Boor // Journal of Approximation Theory. — 1972. — Т. 6, № 1. — С. 50—62. — URL: https://www.sciencedirect.com/science/article/pii/0021904572900809.

93. *Blauert, J.* Spatial Hearing: The Psychophysics of Human Sound Localization [Текст] / J. Blauert. — MIT Press, 1997.

94. *Brandstein, M.* Microphone Arrays: Signal Processing Techniques and Applications [Текст] / M. Brandstein, D. Ward. — Springer, 2001.

95. *Behringer.* Behringer UMC1820 User Manual [Текст] / Behringer. — 2020. — URL: https://www.behringer.com/product.html?modelCode=P0B2J.

96. *Electronics, sE.* sE Electronics 8 Pair Datasheet [Текст] / sE Electronics. — 2023. — URL: https://www.seelectronics.com/se8-pair.

97. *Inc., C.* Reaper User Manual [Текст] / C. Inc. — 2023. — URL: https://www.reaper.fm/userguide.php.

98. *GoPro.* GoPro MAX User Guide [Текст] / GoPro. — 2021. — URL: https://gopro.com.

99. *Bovik, A. C.* Handbook of Image and Video Processing [Текст] / A. C. Bovik. — Academic Press, 2010.

100. *STMicroelectronics.* STM32L476RG Ultra-low-power ARM Cortex-M4 32-bit MCU with FPU, 1 Mbyte of Flash memory, 128 Kbytes of SRAM [Текст] / STMicroelectronics. — 2024. — URL: https://www.st.com/en/microcontrollers-microprocessors/stm32l476rg.html.

101. So you think you can DAS? A viewpoint on delay-and-sum beamforming [Текст] / V. Perrot [и др.] // Ultrasonics. — 2021. — Март. — Т. 111. — С. 106309. — URL: http://dx.doi.org/10.1016/j.ultras.2020.106309.

102. *Leman, R.* Beamforming using Digital Piezoelectric MEMS Microphone Array [Текст] / R. Leman, B. Travaglione, M. Hodkiewicz. — 2021. — arXiv: 2111.10087 [eess.SP]. — URL: https://arxiv.org/abs/2111.10087.

Список рисунков

Список таблиц

Приложение А. Свидетельства о государственной регистрации программ для ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2024690295

AudioHarvest - программный комплекс для сбора аудио данных

Правообладатель: ***Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (RU)***

Автор(ы): ***Мкртчян Грач Маратович (RU)***

Заявка № **2024688923**

Дата поступления **28 ноября 2024 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***13 декабря 2024 г.***

Руководитель Федеральной службы по интеллектуальной собственности

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 0692e7c1a6300bf54f240f670bca2026
Владелец **Зубов Юрий Сергеевич**
Действителен с 10.07.2024 по 03.10.2025

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2024690558

RID Acoustic Model — система аудиодетекции дорожных событий в реальном времени

Правообладатель: ***Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (RU)***

Автор(ы): ***Мкртчян Грач Маратович (RU)***

Заявка № **2024688949**
Дата поступления **28 ноября 2024 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ ***16 декабря 2024 г.***

Руководитель Федеральной службы по интеллектуальной собственности

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 0692e7c1a6300bf54f240f670bcca2026
Владелец Зубов Юрий Сергеевич
Действителен с 10.07.2024 по 03.10.2025

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

RU2023681411

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):
2023681411
Дата регистрации: **13.10.2023**
Номер и дата поступления заявки:
2023669946 30.09.2023
Дата публикации и номер бюллетеня:
13.10.2023 Бюл. № 10

Автор(ы):
Мкртчян Грач Маратович (RU),
Мосева Марина Сергеевна (RU),
Павликов Артем Евгеньевич (RU),
Заликов Руслан Артурович (RU)
Правообладатель(и):
Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (RU)

Название программы для ЭВМ:
LabelSpeech - программный комплекс, предназначенный для аннотации различных видов данных, включая аудио, видео, текст и изображения

Реферат:
Программный комплекс (ПК) предназначен для аннотации различных видов данных, включая аудио, видео, текст и изображения. ПК позволяет аннотировать большое количество данных. ПК позволяет администрировать группу людей, которые занимаются аннотированием данных. Также помогает пользователям тем, что предварительно размечает данные с использованием встроенной нейронной сети. Проверка данных осуществляется самими пользователями.

Язык программирования: Python, JavaScript

Объем программы для ЭВМ: 4,9 МБ

Приложение Б. Акты о внедрении

«УТВЕРЖДАЮ»
Ректор ордена Трудового Красного
Знамени федерального государственного
бюджетного образовательного учреждения
высшего образования «Московский
технический университет связи и
информатики» (МТУСИ)

_____
д.т.н., доцент Ерохин С.Д.
«___» _____ 2025 г.

АКТ

об использовании результатов диссертационной работы Мкртчяна Г.М. на тему: «Разработка методов и средств нейросетевой обработки акустической информации» в учебном процессе кафедры «Математическая кибернетика и информационные технологии»

Комиссия в составе:

- проректора по учебной работе, к.э.н., доц. Аджиковой Алтынай Султахановны;
- руководителя Департамента организации и управления учебным процессом, к.э.н., доц. Краснова Евгения Владимировича,

удостоверяет, что в учебном процессе кафедры «Математическая кибернетика и информационные технологии» при выполнении лабораторных и практических работ по дисциплинам: «Машинное обучение» и «Методы интеллектуальной обработки аудиосигналов» для бакалавров направления 09.03.01 «Информатика и вычислительная техника» используются результаты диссертации Мкртчяна Грача Маратовича, а именно: проведенный соискателем анализ современных методов и средств нейросетевой обработки акустических данных, а также разработанный алгоритм повышения устойчивости при обучении нейронной сети, предназначенной для классификации акустических данных. Эффективность внедрения заключается в приобретении студентами знаний по перспективным направлениям развития науки и техники.

Проректор по учебной работе А.С. Аджикова

Руководитель Департамента организации и управления учебным процессом Е.В. Краснов

ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ
«МКАД»
366200, Чеченская Республика, Гудермесский район,
г. Гудермес, ул. А.Кадырова, д. 38, оф. 13/1
ИНН 2632083929 КПП 200501001
m.mkad@mail.ru

УТВЕРЖДАЮ:
Директор ООО «МКАД»
_____Ю.И. Хахонин
«_9_» декабря 2024 г.

АКТ

о внедрении результатов диссертационной работы
Мкртчяна Грача Маратовича,
представленной на соискание ученой степени
кандидата технических наук

Настоящим актом подтверждается, что основные результаты диссертационного исследования Мкртчяна Грача Маратовича «Разработка методов и средств нейросетевой обработки акустической информации» в настоящее время используется в работе ООО «МКАД», а именно:

- Метод сбора и аннотирования акустической информации, отличающийся внедрением предобученной модели распознавания, позволяющий повысить скорость аннотирования данных не менее, чем на 30%;
- Архитектура программно-аппаратного комплекса сбора, хранения и классификации акустической информации, обладающая возможностью непрерывной обработки цифрового сигнала.

Результаты диссертационного исследования позволили осуществить выбор эффективных решений при разработке программного комплекса натурных акустических и виброакустических измерений на разных этапах проектирования, строительства и реконструкции уникальных зданий и сооружений.

Директор ООО «МКАД» Хахонин Ю.И.

LIMITED LIABILITY COMPANY
«RC TECHNOLOGIES»
(Road Construction Technologies)

ООО «Эр Си Технолоджис»
111024, г. Москва, Авиамоторная улица, д. 50, стр. 2, этаж 2, помещ. 11, комн. 10, офис 210.
ИНН 7722465745, КПП 772201001,
ОГРН 1187746742842.
Email: RC-Technologies@yandex.ru

Утверждаю
Генеральный директор ООО "Эр Си Технолоджис"
_____ Мамедов К.Ю.
24 ноября 2024г.

АКТ

о внедрении результатов диссертационной работы
Мкртчяна Грача Маратовича,
представленной на соискание ученой степени
кандидата технических наук

Настоящим актом подтверждается, что основные результаты диссертационного исследования Мкртчяна Грача Маратовича «Разработка методов и средств нейросетевой обработки акустической информации» в настоящее время используется в работе ООО «ЭР СИ ТЕХНОЛОДЖИС», а именно:

- алгоритм классификации акустических данных, демонстрирующий высокую точность работы (не менее 94%) в условиях городской среды, обеспечивающий устойчивость к шумам и сложным акустическим условиям за счёт применения современных нейросетевых архитектур на основе трансформеров.

Результаты диссертационного исследования находятся в стадии пробной эксплуатации, а по ее окончании будут внедрены в систему комплекса безопасности, которая фиксирует и анализирует аудиопоток, получаемый через подключённые или встроенные микрофоны, что позволяет выявлять потенциальные ситуации повышенной опасности, например, звон разбитого стекла, крики людей и т.д.

Генеральный Директор
ООО «Эр Си Технолоджис» _____ Мамедов К. Ю.