

КЛАССИФИКАЦИЯ ВИДОВ ТРАНСПОРТА С ИСПОЛЬЗОВАНИЕМ АКУСТИЧЕСКИХ ПРИЗНАКОВ

Януцевич Д.В., Марковский И.М.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Леванцевич В.А. – ст. преподаватель

Аннотация. Рассматривается система классификации уличных видов транспорта на основе одномерной сверточной нейронной сети. Исследована эффективность использования спектрограммы, мел-спектрограммы и мел-частотных кепстральных коэффициентов (МЧКК) в качестве признаков для классификации транспорта. Для имитации реальных городских условий дорожных ситуаций использована аугментация данных во временной и частотно-временной областях.

Ключевые слова: Классификация, сверточная сеть, спектрограмма, мел-спектрограмма, МЧКК, аугментация.

Акустические системы классификации представляют ряд преимуществ перед системами, основанными на зрении (vision-based systems), особенно в контексте классификации уличных видов транспорта, где методы, основанные на зрении, неэффективны. В городских условиях возможность контролировать и классифицировать транспортные средства по звуку обеспечивает экономически эффективную альтернативу сетям камер, особенно в ночное время или во время неблагоприятных погодных условий.

Предлагаемая система фокусируется на классификации следующих видов транспортных средств: легковые автомобили, автобусы, троллейбусы, трамваи, метро. В отдельный класс («Иное») был вынесен случай отсутствия транспорта на дороге.

Извлекаемые акустические признаки

Акустическая классификация транспортных средств основана на извлечении информативно-значимых акустических признаков из аудиосигналов. Для исследования эффективности определенного вида признаков исследовались три наиболее распространенных способа представления аудиосигналов в виде: спектрограммы, мел-спектрограммы (спектрограммы в шкале мелов) и мел-частотных кепстральных коэффициентов (МЧКК).

Спектрограмма получается путем применения кратковременного преобразования Фурье (КВПФ) к исходному сигналу. Сигнал предварительно разделяется на небольшие отрезки (окна) и взвешиваются при помощи оконной функции (чаще окном Хэмминга). Затем к каждому окну применяется быстрое преобразование Фурье. Размер спектрограмм превышает другие рассматриваемые частотно-временные представления, и это может сказаться на эффективности работы системы, т.к. необходимо рассматривать больший объем данных. Несмотря на это, спектрограммы дают четкое представление о распределении энергии (рисунок 1а) по диапазонам частот [1].

В это же время мел-спектрограмма лучше отражает распределение энергии в частотно-временном представлении (рисунок 1б). Мел-спектрограмма получается путем наложения на спектрограмму фильтров треугольной формы, которые используются для усреднения спектра вблизи центральных частот, расположенных равномерно в шкале мелов [2]. Анализ мел-спектрограмм звука двигателя автомобиля показывает, что энергия звукового сигнала распространяется на широкий частотный диапазон. Мел-спектрограмма богаче текстурой, область концентрации энергии сигнала имеет более сложную геометрическую структуру, что делает мел-спектрограмму подходящей под задачу акустическим признаком [3].

МЧКК широко используются в процессе извлечения признаков за счет их декоррелирующего свойства. Декорреляция достигается путем применения дискретного косинусного преобразования (ДКП) к логарифму мел-спектрограммы. К тому же при анализе обычно из всех МЧКК выбираются первые 13-20 коэффициентов, что делает их эффективными в качестве входных данных системы классификации [4].

Однако МЧКК очень чувствительны к фоновым шумам, в то время как спектрограмма, полученная с использованием КВПФ, может использоваться даже для аудиообразцов, имеющих шумы, и обеспечивает более эффективные результаты.

Для разрабатываемой системы классификации были выбраны следующие значения параметров, используемых при выделении акустических признаков: частота дискретизации сигнала – 22.05 кГц, размер окна – временных отсчетов, временной шаг – 512 временных отсчетов, количество мелов – 128.

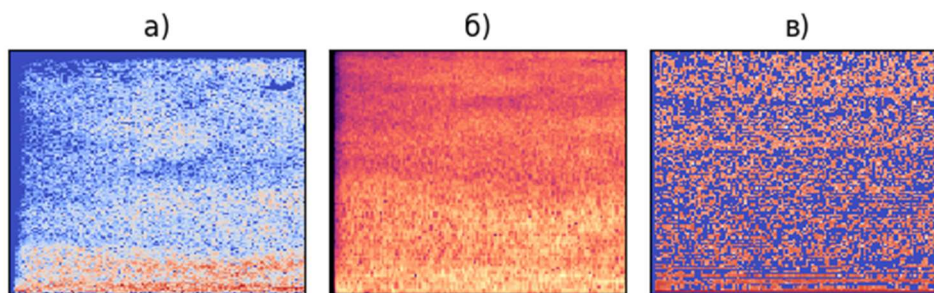


Рисунок 1 – Пример акустических признаков: а – спектрограмма, б – мел-спектрограмма, в – мел-частотные кепстральные коэффициенты

Задача классификации звука можно рассматривать близкой к задаче классификации изображений, поэтому многие подходы анализа изображений применимы к этой задаче. Для классификации звуков хорошо подходят сверточные нейронные сети. Они способны определять частотные характеристики различных источников звука во времени, даже если часть звука перекрывается другими источниками [5].

Сбор данных для обучения нейронной сети

В качестве устройства аудиозаписи был выбран телефон iPhone 11, частота дискретизации аудиозаписей – 44.1 кГц. Запись производилась в течение 12 дней в городских условиях. Средняя температура воздуха составляла -1°C , средняя скорость ветра – 7 м/с.

После записи аудио и ручного отбора файлов было собрано 668 записей, общей длительностью в 73 мин. Из них примерно 35% относится к классу «Машина», 30% – к классу «Иное», 15% – к классу «Трамвай». Остальные классы – «Автобус», «Троллейбус», «Метро» – составили по 6% от всей обучающей выборки.

Подготовка данных для обучения

Из-за того, что собранные записи имеют разную длину (что не является приемлемым для обучения нейронной сети), записи были разделены на равные отрезки в 2.0 сек. При этом необходимо отметить, что при недостатке аудиоданных в файле для создания очередного отрезка они не дополнялись нулями в конце или начале (как это делается в специализированных библиотеках для расчета аудиопризнаков), а с целью избежания потери данных выбирались отрезки с перекрытием в 1.5 сек. Таким образом максимально возможная потеря данных становится меньше, т.к. в таком случае шаг выбора отрезка составляет 0.5 сек.

После разделения файлов на отрезки равной длительности они подвергались разделению на три набора данных, используемых при обучении:

1. Тренировочный набор (training set) – непосредственно используется для обучения модели и изучения скрытых особенностей/закономерностей в данных.
2. Валидационный набор (validation set) – настраивает гиперпараметры (например, скорость обучения) и предотвращает переобучение.
3. Тестовый набор (testing set) – используется для итоговой оценки модели после завершения обучения.

Было выбрано разделение наборов в соотношении 70:15:15 [6].

Расширение обучающей выборки

Так как распределение отрезков по классам не является равномерным и не все они отражают возможные реальные ситуации (например, ускорение автомобиля), обучающая выборка была расширена путем применения аугментации данных. Аугментация позволяет улучшить результаты обучения модели путем формирования новых образов [7].

Аугментация применялась в двух представлениях: во временном (модификация исходного сигнала) и частотно-временном (модификация спектрограммы). Во временном представлении использовались следующие подходы: сдвиг высоты тона, усиление, добавление шума. В частотно-временном представлении использовались следующие методы: случайное стирание, частотное и временное маскирование, размытие [8-10].

Для задачи классификации транспорта изменение высоты тона может позволить не только охватить более широкий спектр частот транспорта (например, для имитации звука разных двигателей или скоростей), но и симитировать разные состояния транспорта (например, ускорение, замедление) [11].

Усиление сигнала используется с целью имитации разных уровней слышимости. Если коэффициент усиления непостоянный, такое усиление может позволить симитировать ситуацию приближения или отдаления транспорта к записывающему устройству. Добавление шума происходит с целью устойчивости системы к естественным шумам, может использоваться для имитации шума улицы, различных условий окружающей среды.

С целью имитации реальных условий для изменения высоты тона и усиления был реализован механизм изменяющихся параметров аугментации. Это означает, что для аугментации высота тона или коэффициент усиления может изменяться по определенному закону. Пример сценария изменения параметра представлен на рисунке 2а. На рисунке изображены сценарии изменения высоты тона или усиления в положительном направлении – от 0 до 1. Стоит отметить, что изменение также может происходить в отрицательном направлении (значения от -1 до 0) или в обоих направлениях (значения от -1 до 1). К тому же сценарии изменения могут вести и к росту, и к убыванию параметра изменения. Сценарий с убыванием параметра изменения представлен на рисунке 2б.

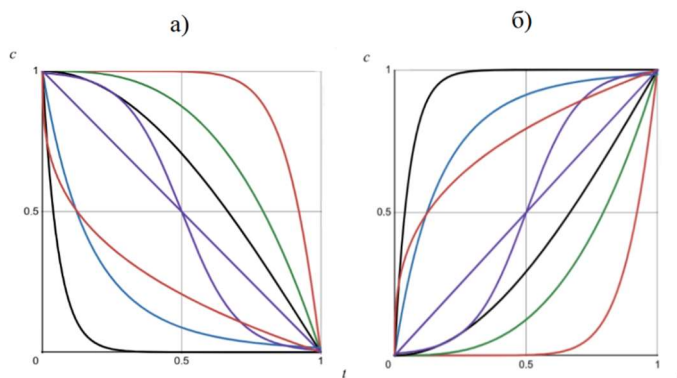


Рисунок 2 – Сценарии изменения параметров аугментации: а – с убыванием параметра, б – с возрастанием параметра

На рисунке 3 представлены примеры аугментации данных в разных представлениях. На рисунке 3б высота тона изменяется динамически в форме половины периода синуса, амплитуда изменения – 6 полутонов. На рисунке 3в случайным образом были маскированы диапазон в 20 частотных бинов и 10 временных окон.

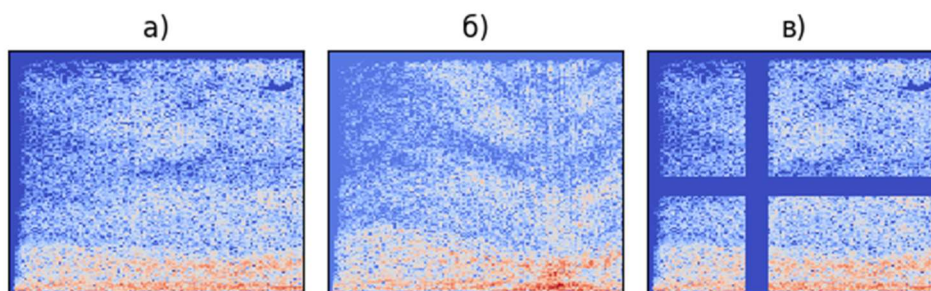


Рисунок 3 – Спектрограмма сигнала; а – исходный сигнал, б – динамическое изменение высоты тона, в – частотное и временное маскирование

Архитектура нейронной сети

Для описания структуры модели и ее обучения были выбраны язык программирования Python и фреймворк для машинного обучения PyTorch. За основу была взята архитектура модели из четырех сверточных слоев. После всех сверточных слоев также следует четыре полносвязных слоя (Dense layer) [12].

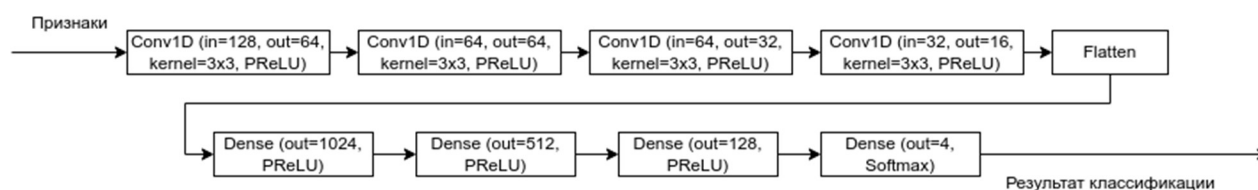


Рисунок 4 – Архитектура модели

После экспериментов с параметрами модели наиболее эффективной оказалась модель с функцией активации PReLU и пакетной нормализацией. Была использована регуляризация путем исключения нейронов с вероятностью 0,2. Модель обучалась с максимальным количеством эпох в 50,

максимальное количество эпох без улучшения обучения модели (patience) – 8. Минимальная «скорость обучения» (learning rate) – 0,001, используемый оптимизатор – Adam.

Результаты обучения модели

Ниже приведены результаты обучения модели с использованием мел-спектрограммы в качестве акустического признака. Ранее были отмечены особенности разделения аудиоотрезков по разным наборам данных, используемых при обучении. Точность модели со «строгим» разделением данных на тренировочном наборе оказалась равной 85,2%, значение функции потерь – 0,46. Как и ожидалось, точность модели с «нестрогим» разделением аудиоотрезков оказалась выше – 95,4%. И, соответственно, значение функции потерь оказалось ниже – 0,16.

В связи с тем, что звук двигателей автобусов и троллейбусов схож со звуком некоторых автомобилей, было принято решение объединить классы «Легковой автомобиль», «Автобус» и «Троллейбус». Таким образом, точность модели, предсказывающей четыре класса транспорта, со «строгим» разделением данных на тренировочном наборе оказалась равной 93,2%, значение функции потерь – 0,24. Точность модели с «нестрогим» разделением аудиоотрезков оказалась равной 97,2%, значение функции потерь – 0,08.

Тестирование программного средства

С целью тестирования приложения в реальном времени было реализовано программное средство, которое использует микрофон в качестве источника данных для аудиопотока и вычисляет акустические параметры, которые используются моделью для предсказания класса. Предсказание осуществляется каждые 250 мс и отображает результаты предсказания на временной шкале.

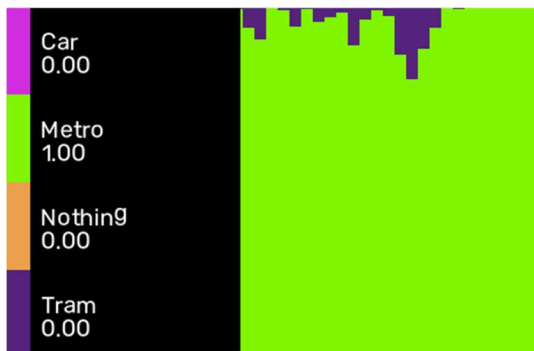


Рисунок 5 – Внешний вид программного средства

В качестве записывающего устройства при тестировании использовался телефон, при помощи которого производилась запись обучающей выборки, что дало возможность в полной мере оценить результат обучения. Для того, чтобы телефон можно было использовать в качестве микрофона, использовалось программное средство WO MIC.

Выводы об эффективности акустических признаков приведены для моделей, обученных на «строгим» разделенных данных с объединением классов «Автобус», «Троллейбус» и «Легковой автомобиль». Тестирование в режиме реального времени показало, что наиболее эффективным акустическим признаком оказалась мел-спектрограмма, так как модель, обученная на этом признаке, дает более детерминированный результат.

Список использованных источников:

1. Cheng K.W., Chow H.M., Li S.Y., Tsang T.W., Ng H.L.B., Hui C.H., Lee Y.H., Cheng K.W., Cheung S.C., Lee C.K. Spectrogram-based classification on vehicles with modified loud exhausts via convolutional neural networks [Электронный ресурс] – Режим доступа: <https://doi.org/10.1016/j.apacoust.2023.109254> – Дата доступа: 07.02.2025
2. Вашкевич М.И., Бурак А.А., Конойко Н.С., Долдова В.С. Анализ акустических параметров голоса для выявления заболеваний гортани [Электронный ресурс] – Режим доступа: <https://inf.grid.by/jour/article/view/893/938> – Дата доступа: 12.02.2025
3. Wu Z., Wan Z., Ge D., Pan L. Car engine sounds recognition based on deformable feature map residual network [Электронный ресурс] – Режим доступа: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8854583/> – Дата доступа: 07.02.2025
4. Gourisaria M.K., Agrawal R., Sahni M., Singh P.K. Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques [Электронный ресурс] – Режим доступа: <https://link.springer.com/article/10.1007/s43926-023-00049-y> – Дата доступа: 09.02.2025
5. Salamon J., Bello J.P. Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/1608.04363> – Дата доступа: 07.02.2025
6. Train Test Validation Split: How To & Best Practices [Электронный ресурс] – Режим доступа: <https://www.v7labs.com/blog/train-validation-test-set> – Дата доступа: 20.02.2025
7. Alomar K., Aysel H.I., Cai X. Data Augmentation in Classification and Segmentation: A Survey and New Strategies [Электронный ресурс] – Режим доступа: <https://www.mdpi.com/2313-433X/9/2/46> – Дата доступа: 15.02.2025
8. Zhong Z., Zheng L., Kang G., Li S., Yang Y. Random Erasing Data Augmentation [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/1708.04896> – Дата доступа: 15.02.2025

9. Daniel S. Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D. Cubuk, Quoc V. Le. *SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition* [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/1904.08779> – Дата доступа: 16.02.2025

10. Halvdansson S. *On a Time-frequency Blurring Operator With Applications In Data Augmentation* [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/2405.12899> – Дата доступа: 18.02.2025.

11. Mešić A. *Evaluating the Use of Pitch Shifting to Improve Automatic Speech Recognition Performance on Southern Dutch Accents* [Электронный ресурс] – Режим доступа: <https://repository.tudelft.nl/record/uuid:f1f54596-adbc-436a-87ac-a40394689b92> – Дата доступа: 16.02.2025

12. Ashhad M., Goenka U., Jagetia A., Akhtari P., Ambat S.K., Samuel M. *Improved Vehicle Sub-type Classification for Acoustic Traffic Monitoring* [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/2302.02945> – Дата доступа: 21.02.2025

STREET VEHICLE TYPE CLASSIFICATION BASED ON ACOUSTIC FEATURES

Yanutsevich D.V., Markouski I.M.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Levantsevich V.A. – senior lecturer

Annotation. A system for classifying street vehicles based on a one-dimensional convolutional neural network is developed. The effectiveness of spectrograms, mel-spectrograms, and mel-frequency cepstral coefficients (MFCC) as features for vehicle classification task is investigated. To simulate such situations, data augmentation in time and time-frequency domains is used.

Keywords. Classification, convolutional network, spectrogram, mel-spectrogram, MFCC, augmentation.