

СРАВНЕНИЕ МЕТОДОВ АКУСТИЧЕСКОГО ДЕТЕКТИРОВАНИЯ, КЛАССИФИКАЦИИ И ЛОКАЛИЗАЦИИ ДЛЯ РЕШЕНИЯ ЗАДАЧ ОБНАРУЖЕНИЯ СОБЫТИЙ И СЦЕН

М.Г. Городничев, к.т.н., доцент, Московский технический университет связи и информатики, m.g.gorodnichev@mtuci.ru;

А.В. Тимчук, Московский технический университет связи и информатики, a.v.timchuk@mtuci.ru;

Г.М. Мкртчян, Московский технический университет связи и информатики, g.m.mkrtchyan@mtuci.ru.

УДК 004.023

Аннотация. Использование акустических методов позволяет решить многие проблемы с определением окружающей обстановки, которые присущи другим методам: камеры, ручное наблюдение. В статье разбираются и сравниваются решения, реализующие акустическое детектирование, классификацию и локализацию, и выделяются методы, которые используются для решения данных задач.

Ключевые слова: акустическое детектирование; акустическая классификация; акустическая локализация; нейронные сети; акустическая сцена.

COMPARISON OF ACOUSTIC DETECTION, CLASSIFICATION AND LOCALIZATION METHODS FOR SOLVING PROBLEMS OF DETECTING EVENTS AND SCENES

М.Г. Городничев, Candidate of Technical Sciences, associate professor, Moscow technical university of communications and informatics;

A.V. Timchuk, Moscow technical university of communications and informatics;

G.M. Mkrtchian, Moscow technical university of communications and informatics.

Annotation. The use of acoustic methods allows you to solve many problems with determining the environment that other methods have: cameras, manual surveillance. The article analyzes and compares solutions that implement acoustic detection, classification and localization, and identifies the methods that are used to solve these problems.

Keywords: acoustic detection; acoustic classification; acoustic localization; neural networks; acoustic scene.

Введение

Потребность в акустических методах детектирования, классификации и локализации событий и сцен существует во множестве областей, что приводит к активному развитию и эволюции применяемых методов [1, 2]. Поскольку задача детектирования сцены довольно широка, нет единого верного метода решения, который бы применялся для всех возможных условий использования.

Можно выделить следующие условия, влияющие на решение данной задачи:

- Характеристика окружающего фона: естественный шум окружения, в котором должна работать система. Для ряда задач некоторые события, должны к обнаружению, являются естественным фоном и не должны учитываться. Примером подобного события может быть человеческая речь: если у нас стоит задача обнаружения людей, это интересное нам событие, если мы ищем аномалии в работе оборудования – речь будет фоном.

- Акустические характеристики событий: каждое акустическое событие имеет уникальный набор гармоник. В зависимости от их частоты и амплитуды можно выделить общие характеристики для интересующих нас событий и отсеять лишнее. В этом случае возможно применить методы, специализированные для конкретных акустических параметров.
- Акустические искажения: искажения звукового сигнала из-за эха, эффекта Доплера, ревербераций, интерференции. Существенность влияния данных эффектов зависит от окружающей среды и акустических характеристик событий.
- Полнота задачи: какие из трех подзадач обнаружения сцены требуется решить: только детектирование, детектирование и классификация, только локализация, или все вместе. Решение каждой из подзадач требует своего подхода.

Поскольку постановка задач довольно сильно различается, необходимо выделить критерии, по которым необходимо проанализировать и сравнить существующие решения, их условия использования, преимущества и недостатки. Сравнимые решения будут рассматриваться как алгоритмы обработки, состоящие из трех фаз: предобработка, обработка, постобработка. Решения будут сравниваться по: решаемой задаче, универсальности применения, возможности адаптации к другим задачам, по эффективности решения задачи. Перед анализом решений следует рассмотреть четыре варианта интерпретации аудиоданных, применимых для решения подобных задач.

Варианты интерпретации аудио

Первый вариант интерпретации аудиоданных предполагает использование обычной вейв формы (рис. 1), характеризующейся временем и амплитудой сигнала. Это самый простой вариант, предоставляющий минимум информации о сигнале. Как правило, этот вариант может использоваться как выходная форма, но редко, как входная.

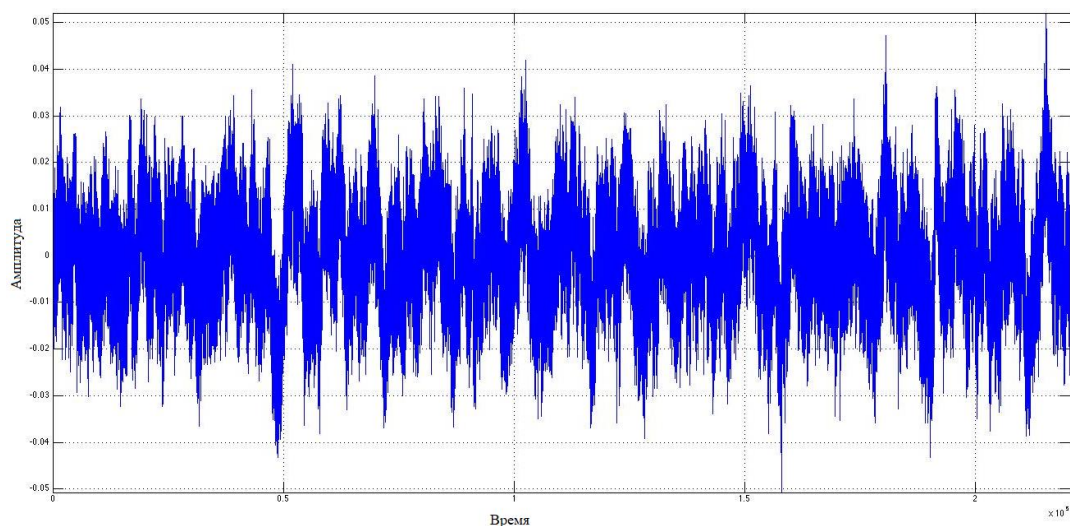


Рисунок 1

Второй вариант – преобразование Фурье [3], которое является классическим для обработки цифровых сигналов.

При его применении из амплитудно-временного представления мы переходим к амплитудно-частотному (рис. 2). Данная интерпретация полезна, так как любое акустическое событие характеризуется определенным набором частот, однако потеря временной характеристики лишает нас возможности определить момент времени события. По этой причине на практике применяется краткосрочное преобразование, позволяющее получить дискретную характеристику времени, что в дальнейшем позволяет сформировать спектрограмму частота-время-амплитуда (рис. 3).

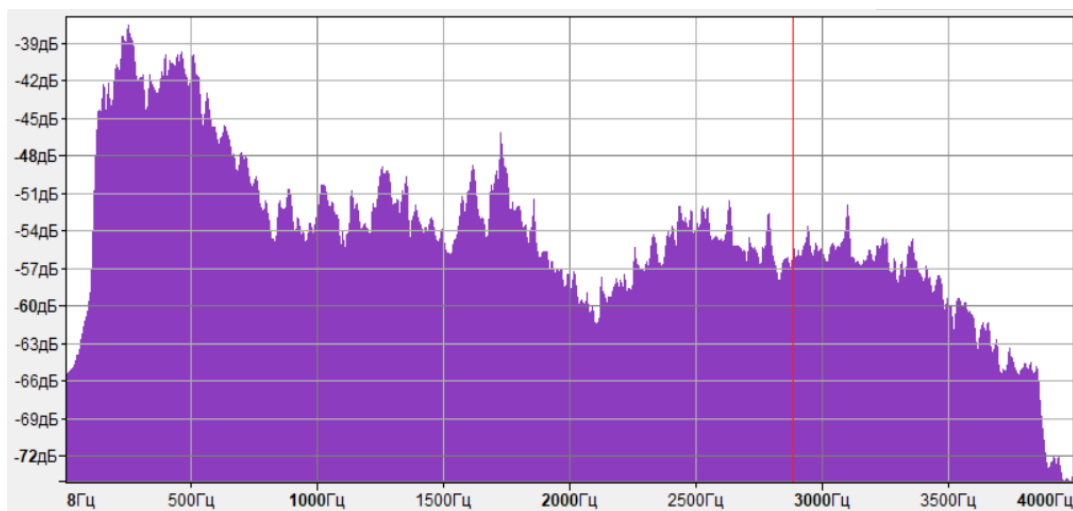


Рисунок 2

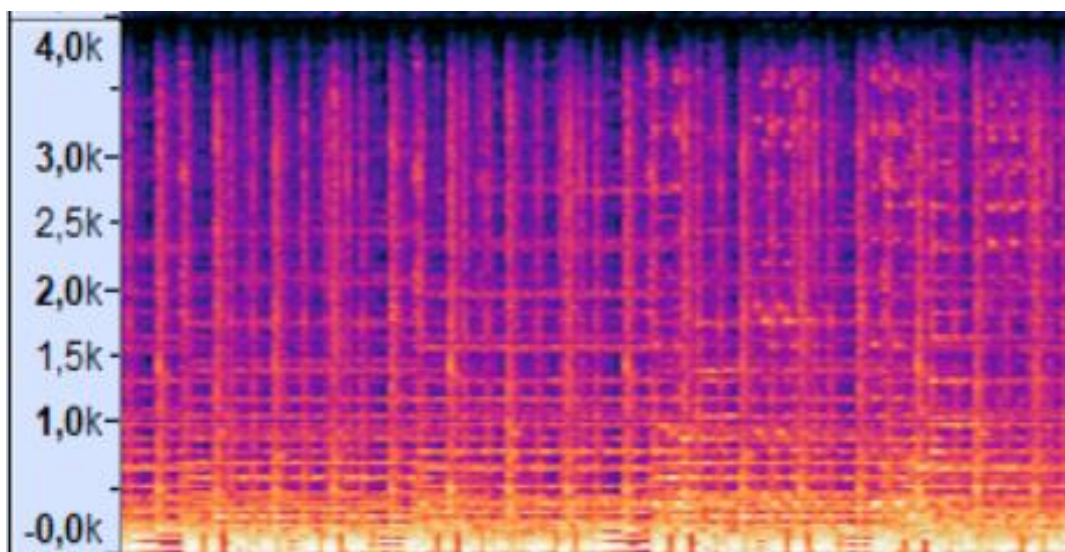


Рисунок 3

Третий вариант – *mel*-спектрограмма [4], которая является обычной спектрограммой (рис. 3), но где частоты переведены в *MEL*-шкалу (рис. 4), больше соответствующая тому, как человек воспринимает звук. Такой способ интерпретации особенно полезен для анализа голоса.

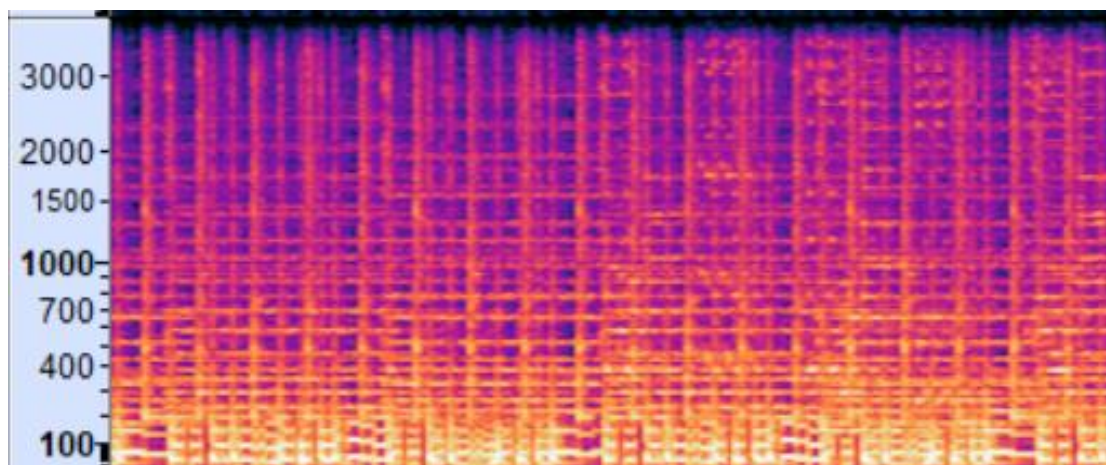


Рисунок 4

Четвертым вариантом является *MFCC* [5], *mel*-частотные кепстральные коэффициенты, которые используются в задачах анализа речи. Для формирования необходимо выполнить следующие операции с *mel*-спектрограммой: взять натуральный логарифм амплитуды (энергии) для каждого значения *mel* и применить дискретное косинусное преобразование отдельно для каждого набора амплитуд по времени. В результате получится спектрограмма вида: номер коэффициента – время – значение коэффициента.

Решения для детекции акустических событий

Первое решение [6] было создано Томасом Гриллом для решения задачи детектирования птиц. Целью исследования было получить модель, способную эффективно работать в рамках неоднородности исходных данных [7], что в случае с аудио означает: разные фоновые шумы, разная громкость, разные частотные искажения. Решение представляет из себя сверточную нейронную сеть, с двухэтапным алгоритмом обучения, призванным адаптировать модель к новым данным. На этапе предобработки задействован алгоритм создания *mel*-спектрограммы, в результате которой формируются картинки, поступающие на вход модели. На этапе обработки выполняется сверточная модель, на выходе которой получается значение от 0 до 1, как вероятность наличия птицы на данном фрагменте. Постобработка не применяется.

Алгоритм тренировки данной модели двухступенчатый и предполагает наличие двух и более в дальнейшем блоков исходных данных. Первым блоком является изначальный тренировочный размеченный датасет, вторым блоком является тестовый датасет без разметки. На первой стадии тренировки получается первичная модель на первом блоке данных. С ее помощью производится разметка второго блока данных, если значение вероятности достаточно близко к 0 или 1, размеченные данные прибавляются к первому блоку. После чего выполняется тренировка модели на расширенном наборе данных. Расширение датасета и дообучение может проводиться для каждого новых по характеристикам аудиоданных.

Плюсом подобного подхода является работа с ненормализованными данными без необходимости полного обучения с нуля для новых акустических условий, соответственно, модель не накладывает жестких ограничений на используемые технические средства (микрофоны), что характерно для решаемой задачи.

Второе решение для детектирования [8] создано группой исследователей для определения аномалий в аудиопотоке. Под аномалией в данном случае подразумевается любое нетипичное для окружения событие. Оценка точности модели велась по двум метрикам: *ROC* и *F1*.

Предобработка использует *mel*-спектрограммы для интерпретации аудиосигнала, после которой применяется автоэнкодер. Модель автоэнкодера представляет из себя объединение сверточного энкодера и сверточного декодера, которые обучаются на восстановление изначальной картинки, в качестве такой картинки определяются *mel*-спектрограммы. На этапе постобработки происходит расчет ошибки расхождения восстановленной и изначальной спектрограмм, и, если ошибка превышает установленный рубеж, значит на данном отрезке аудиосигнала присутствует аномалия.

Для тренировки данной модели требуется использовать чистый датасет с максимальным наполнением по нейтральным событиям, так как любые события, не представленные в датасете, будут определяться как аномалии.

Неоспоримым плюсом подобного подхода является обнаружение любых событий, изначально не включенных в раздел допустимых, однако это также и приводит к осязаемому недостатку – датасет должен содержать достаточно образцов окружения, и в случае смены рабочего окружения требуется собирать новый датасет и производить повторное обучение модели с нуля.

Третье решение [9], рассматриваемое в рамках детектирования, решает аналогичную предыдущему задачу, но использует другой подход.

Вместо автоэнкодера применяется модель *WaveNet*, предназначение которой – генерация волновой формы. Сверка ошибки результатов сгенерированной волновой формы относительно исходной позволяет определить аномалию в конкретный момент времени. Оценка точности велась с использованием двух *F1* метрик: одна для оценки обнаружения аномалий и вторая для оценки обнаружения аномалии во времени (оценка задетектированной длительности события с реальной).

На этапе предобработки применяется логарифмическое *mel*-преобразование с 40 значениями (полная спектрограмма не строится) с окном в 25 мс и наложением в 4%, результат которой в дальнейшем нормализуется к интервалу от 0 до 1, на базе статистики по тренировочному датасету, и в заключение расширяется, чтобы сохранить временную характеристику согласованной с изначальной волновой формой. После этого на этапе обработки выполняется *WaveNet* и выполняется оценка сгенерированной формы. На этапе постобработки выполняется три действия для сглаживания результатов детектирования: применяется медианный фильтр, заполняются небольшие разрывы в событиях и убираются события, чья длина меньше установленного уровня.

Данное решение наследует плюсы и минусы предыдущего, но дополнительно приобретает еще один плюс: точная привязка ко времени входящего сигнала, что позволяет добавить в характеристику события точного времени начала и завершения.

Решения для классификации акустических событий

Четвертое решение относится к категории классификаторов и решает задачу обнаружения событий в автодорожном тоннеле [10]. Данная задача подразумевает выполнение классификации событий, т.е., переход от бинарной классификации к классам. У этой задачи есть особые акустические условия: шум и искажения, присущие тоннелю. Для решения проблемы шумов используется довольно сложный алгоритм предобработки.

Первым этапом предобработки выполняется преобразование Фурье на коротком промежутке времени. Следующим этапом идет обработка шума. Обработка шума может производиться разными методами, но в данном случае применяется алгоритм разделения сигнала и шума на базе *NTF* (*non-negative tensor factorization*). Для работы данного алгоритма нам необходимо заранее получить два базиса: базис, характеризующий сигнал, и базис, характеризующий шум. После применения данного алгоритма будет получено две частотные характеристики. Помимо разделения сигнал/шум был добавлен алгоритм, который должен корректировать базис шума для адаптации к изменяемым характеристикам фона, например, изменению погоды. На этом этап предобработки завершается.

На этапе обработки используется комбинированная сверточно-рекуррентная модель с разделенным входом. На первый вход подается ранее извлеченный сигнал, на второй – извлеченный шум. На первом входе и на втором происходит сначала формирование *mel*-спектрограммы, которая затем проходит через сверточную модель. Параметры формирования спектрограммы и веса сверточной модели различны для сигнала и шума. Результаты их работы конкатенируются и подаются на рекуррентную модель, после которой выполняется классификация. В данном случае извлеченный на этапе предобработки шум не отбрасывается, а также используется, чтобы снизить ошибку классификации. Этапа постобработки нет.

Плюсом данного решения является хорошая устойчивость к акустическим шумам тоннеля, что позволяет с большей точностью классифицировать события. Минусом же данного решения является сложность подготовки датасета и сильная привязанность к условиям использования: при использовании в других акустических условиях требуется собирать датасет и тренировать модель с нуля.

Пятое решение [11] создано для решения акустического обнаружения и классификации дронов. Особенностью данного решения является использование метода трансферного обучения для компенсации объективно малого датасета. Архитектура модели базируется на *CNN14* с изменением количества выходных классов на три: дрон, вертолет и шум.

На этапе предобработки создается *mel*-спектрограмма из исходного сигнала. На этапе обработки выполняется модель, а этап постобработки отсутствует.

Для тренировки модели были взяты веса от *PANN* модели, которая является *CNN14*, натренированной на *AudioSet* датасете.

Применение трансферного обучения позволило добиться высокого качества классификации при обучении на датасете с малым количеством данных, так как модель, взятая за основу, была обучена на схожую задачу классификации, что демонстрирует выгоду такого подхода.

Шестое решение [12] решает задачу определения акустической сцены. Отличие от определения акустических событий заключается в том, что сцена может определяться совокупностью различных событий, которые рассматриваются в группе. Рассматриваемое решение использует довольно сложную структуру, состоящую из большого этапа предобработки, обработки и задействующее постобработку.

На первом этапе предобработки происходит преобразование входного аудио в семь каналов. Исходный сигнал должен быть стерео. Первая пара сигналов, стандартные левый и правый, преобразуются в *mel*-спектрограмму и подаются на двухканальную сверточную модель, где свертка применяется отдельно к каждому каналу, а результат конкатенируется и классифицируется. Вторая пара сигналов «*mid*» и «*side*» формируются как $L+R$ и $L-R$ и также подаются на свою двухканальную сверточную модель. Третья пара сигналов образуется путем

разделения одного из исходных каналов по алгоритму *HPSS* (*harmonic-percussive sound separation*), которые также подаются на свою двухканальную сверточную модель. Последний сигнал формируется из одного исходного канала, преобразуется в *mel*-спектрограмму, после чего применяется алгоритм извлечения фона и результат передается в одноканальную сверточную модель.

На этапе обработки применяются все сверточные модели и их результат объединяется.

На этапе постобработки полученный результат классификации фильтруется через средние вероятности за 10 с, а потом применяется поиск максимального, результат которого и определяет обнаруженную сцену.

Применение такого количества различных вариантов обработки сигнала позволяет делать акценты на особые характеристики звука, в результате чего решение показывает довольно большую точность при тестировании классификации сцен.

Минусами же подобного подхода можно назвать тяжеловесность решения и длительность обработки, так как требуется применить довольно много алгоритмов и моделей.

Решения для локализации акустических событий

Седьмое решение [13] ориентировано на локализацию дронов для обеспечения точной посадки. Поскольку предполагается использование подобной системы для заранее известных дронов, задача классификации заменяется задачей идентификации и сводится к использованию системы меток, которые однозначно могут идентифицировать летающий аппарат. Таким образом, важным аспектом данного решения является аппаратная составляющая, определяемая оборудованием, принимающим сигналы на посадочной площадке, и оборудованием, устанавливаемым на дроны, для отправки сигналов.

Идея данного решения заключается в том, что каждый дрон излучает специальным образом модулируемый сигнал, который является и сигналом идентификации, однозначно определяя аппарат, и опорным сигналом для алгоритма *TDoA*.

Первый этап фазы предобработки – это детектирование смодулированного сигнала на аудиопотоках, записываемых с группы микрофонов посадочной площадки. Второй этап предобработки – расчет *TDoA*, то есть, задержки сигнала между двумя парами микрофонов, расположенных на максимальном удалении друг от друга, рассчитанные значения передаются на дрон, согласно идентификации, по беспроводной сети.

На этапе обработки на дроне выполняется его локализация по полученным данным.

Данное решение применимо в довольно узких рамках и обладает рядом ограничений, соблюдение которых необходимо для обеспечения работы системы. Также стоит заметить высокую чувствительность системы к уровню шума и эффекту Доплера, которые приводят к затруднению распознавания передаваемого кода.

Следующее, восьмое решение [14], посвящено гибридной задаче, задействующей все три аспекта: детектирование, классификация и локализация. Цель созданной системы – локализация и классификация выстрелов. Следует отметить, что система требует определенного аппаратного обеспечения, а именно, записывающих модулей, на уровне которых и выполняется детектирование.

На этапе предобработки записывающий модуль проводит двухэтапное детектирование выстрела, в первую очередь, срабатывает детектор пиков,

сигнализирующий о резком звуке, после чего применяется медианный фильтр. В случае, если подтвержден звук, похожий на выстрел, выполняется отправка на центральный вычислительный модуль записанного фрагмента с прикрепленными *GPS* координатами и меткой времени.

На этапе обработки на центральном модуле выполняется более точное детектирование и классификация, по результатам которой выполняется локализация, посредством триангуляции по полученным координатам и временным меткам. Точность локализации, в данном случае, зависит от количества модулей, которые среагировали на выстрел.

Достоинством подобного решения можно назвать применимость и в других сценариях, требующих детектирования, классификации и локализации. Недостатками – необходимость использования множества аппаратных модулей для достижения качественной работы системы.

Последнее, девятое, рассматриваемое решение [15] также выполняет все задачи, и предназначено для обеспечения социальной помощи. Аппаратная часть для данной системы состоит из нескольких сферических массивов и одиночных микрофонов, расположенных по всему помещению.

На этапе предобработки в первую очередь применяется подсистема *VAD* или ее аналог *SAD*. Предназначение данных подсистем – детектирование возможности наличия событий для активации дальнейшей обработки для оптимизации вычислений. Вторым этапом идет отделение шума с применением пробалистической маски шума (*cochleogram*), отделение шума выполняется, так как он обычно снижает точность работы классификаторов. После отделения шума происходит извлечение устойчивых к шуму признаков через применение восьми ориентированных краевых детекторов. На этом этап предобработки завершается.

На этапе обработки выполняется детектирование события путем сравнения энергии отделенных шума и сигнала, и, если эта энергия выше определенного уровня, фиксируется наличие события, после чего применяется классификатор.

На этапе постобработки выполняется локализация событий путем расчета множества *DOA* (*direction of arrival*) и их последующей комбинации. Поскольку вычислить *DOA* можно только, зная расположение одного и того же сигнала на записях с разных микрофонов, используется выделение пика и кросс-корреляция (*GCC-PHAT*). Выделение пика применимо в данном случае, так как необходимо обнаружить местоположение человека и опорная частота для мужчин и женщин известна.

Приведенная система является специализированной и заточена на особые условия работы с жестко заданными анализируемыми событиями. Ее использование вне установленных рамок как есть – невозможно, однако отдельные подходы и идеи применимы и в других условиях.

Сравнение используемых методов

Сравним решения в каждой из категорий. Первой рассмотрим категорию решений для детектирования.

Подходы, примененные для трех сравниваемых решений, можно разделить на два типа: поиск конкретных событий и поиск аномалий. Первый подход подразумевает формирование набора данных из двух категорий. В первую категорию попадают образцы всех интересующих нас событий, во вторую – образцы шума, фона и т.д. При подобном подходе требуется приложить достаточно усилий для формирования сбалансированного качественного набора данных. При втором подходе выполняется проверка входных данных на соответствие

нормальным, и любое отклонение помечается как аномалия. Датасет, в этом случае, должен быть сформирован из всех вариантов нормального окружения.

Выбор подхода из этих двух зависит от следующих факторов: известно ли заранее, какие события мы ищем, постоянны ли они; постоянен ли окружающий фон; количество событий, которые необходимо детектировать.

Применяемые подходы этих решений делают их достаточно универсальными в рамках задачи детектирования, что позволит их использовать на разных наборах событий.

Решение по поиску событий устойчивее к смене акустического окружения, в то время как системы по поиску аномалий потребуют переобучения.

Механизм дообучения, представленный в первом решении, может приводить к снижению точности, но зато позволяет адаптироваться к разнородным исходным данным.

Задача классификации является расширением задачи детектирования, так как второе сводится по своей сути к классификации с одним классом. По этой причине методы, используемые в предыдущих решениях, применимы и здесь.

Рассмотренные решения можно разделить по объекту классификации: одиночное событие и сцена. Сцена может включать в себя множество событий, присутствие которых не постоянно, а природа различна.

Методы предобработки, используемые в решении для автодорожного тоннеля, позволяют обеспечить работу в довольно сложных акустических условиях, что выгодно отличает от второго, однако в более простых случаях это избыточно. Особенностью пятого решения является переносное обучение, что позволяет упростить сбор датасета, в остальном же нет каких-то особенных решений. Оба решения не учитывают природу событий для своей работы.

Шестое решение отличается одновременным использованием нескольких представлений исходных аудиоданных, каждое из которых обрабатывается отдельно и учитывается в общем результате. Подобный подход позволяет учитывать события с разными акустическими свойствами, что необходимо для классификации сцены.

Задача локализации существенно отличается от двух предыдущих. Первое отличие – аппаратное ограничение: требуется несколько входных аудиопотоков, записанных с микрофонов, причем взаимное расположение этих микрофонов должно быть известно, а их минимальное количество три. Второе отличие – для локализации используются алгоритмы, а не нейронные сети.

Действия, необходимые для локализации, следующие: определить время происхождения события на каждом аудиопотоке, выполнить локализацию.

Первый этап довольно важен, в разобранных моделях он был решен следующими способами: заранее известный сигнал, детектирование с разных точек, кросс-корреляция и поиск пика.

Заранее известный сигнал предполагает, что этот сигнал был сгенерирован в рамках вашей системы, как в случае с посадкой дронов, данный подход довольно легок и эффективен, но недопустим для классификации других сигналов. Такой сигнал легко находится на аудио потоке путем сравнения с паттерном. Однако стоит учитывать влияние эффекта Доплера и прочих искажений.

Детектирование с разных точек – наиболее удобный вариант в случае, если задача совмещена с детектированием и/или классификацией и имеется возможность разнести записывающие устройства по площади. Детектирование обеспечит получение времени происхождения сигнала. Акустические искажения учитываются на этапе детектирования.

Кросс-корреляция и поиск пика применимы, если нет возможности использовать предыдущую схему. В этом случае не выполняется поиск сходных паттернов на аудиодорожках, что позволяет выделить одинаковые события. При таком подходе сложно определить класс события.

Этап локализации выполняется либо через триангуляцию, либо с использованием *TDoA* нескольких пар микрофонов.

Заключение

Рассмотрев и сравнив решения в области детектирования, классификации и локализации можно выделить методы, использование которых позволит создать решение под поставленную задачу, а также особенности, которые стоит учитывать.

Для первичной обработки сигнала используется преобразование Фурье и *mel*-шкала, но построение спектрограмм выполняется в том случае, если модель на вход требует изображение, как *CNN*, например. Чистая вейв форма или *MFCC* не используются, так как у первой недостаточно признаков, а вторая имеет выгоду для анализа речи, а не событий.

В зависимости от акустических условий могут использоваться фильтрующие и разделяющие алгоритмы для решения проблем шума, отделения фона или акустических искажений.

Задача детектирования может быть сформулирована как классификация с одним классом, либо как поиск аномалии. Для первого варианта можно применять различные алгоритмические или нейросетевые подходы, в зависимости от типа обнаруживаемых событий, требований к производительности и т.д. Во втором варианте используется подход с восстановлением исходной картины с отсечением аномалий, благодаря чему последующее сравнение входа и выхода позволяет обнаружить расхождения.

Классификация является расширением детектирования, как классификации с одним классом. Если классифицируются события, то отличий от детектирования нет, и, возможно, вместо детектирования следует использовать сразу классификацию. Однако, если стоит задача классификации сцен, то следует учитывать различную природу всех событий, что составляют сцену, а также то, что не все события постоянно присутствуют на сцене.

Локализация имеет аппаратные требования для обеспечения нескольких точек получения сигнала, минимум трех и может выполняться самостоятельно. В этом случае необходимо использование алгоритмов, позволяющих найти одно и то же событие на разных потоках, например, поиск по образцу, поиск пика, кросс-корреляция. Либо как дополнительный этап после детектирования или классификации, и в этом случае, на предыдущем этапе необходимо получить время и координаты точки записи с каждого устройства.

Аппаратная часть также может влиять и на детектирование с классификацией, это зависит от поставленной задачи и выбранного способа ее решения и должно учитываться при выборе методов.

Литература

1. Daniele Barchiesi, Dimitrios Giannoulis, Dan Stowell and Mark D. Plumbley. Acoustic Scene Classification: Classifying environments from the sounds they produce // IEEE Signal Processing Magazine, 2015. – № 32 (3). – С. 16-34.
2. Jakob Abeßer A. Review of Deep Learning Based Methods for Acoustic Scene Classification // Applied Sciences, 2020. – № 10 (6).
3. Зорич В. А. Математический анализ. – М.: Физматлит, 1984. – 544 с.

4. Smith S.S., Volkmann John, Newman E.B. A scale for the measurement of the psychological magnitude pitch // Journal of the Acoustical Society of America, 1937. – № 8 (3). – С. 185-190.
5. Vibha Tiwari Tiwari MFCC and its applications in speaker recognition, 2010.
6. URL https://github.com/OFAI/bird_audio_detection_challenge_2017 (Дата обращения – декабрь 2023).
7. Dan Stowell, Mike Wood, Hanna Pamuła, Yannis Stylianou, Hervé Glotin Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge, 2018.
8. Duman T.B., Bayram B., İnce G. Acoustic Anomaly Detection Using Convolutional Autoencoders in Industrial Processes // 14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019). SOCO 2019. Advances in Intelligent Systems and Computing, vol 950.
9. Hayashi T., Komatsu T., Kondo R., Toda T., Takeda K. Anomalous Sound Event Detection Based on WaveNet // 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 2018. – pp. 2494-2498.
10. Kim N.K., Jeon K.M., Kim H.K. Convolutional Recurrent Neural Network-Based Event Detection in Tunnels Using Multiple Microphones // Sensors, 2019. – № 19 (12):2695.
11. Yaacoub M., Younes H., Rizk M. Acoustic Drone Detection Based on Transfer Learning and Frequency Domain Features // 2022 International Conference on Smart Systems and Power Management (IC2SPM), Beirut, Lebanon, 2022. – pp. 47-51.
12. Yoonchang Han, Jeongsoo Park, Kyogu Lee Convolutional Neural Networks with Binaural Representations and Background Subtraction for Acoustic Scene Classification // В книге: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop. Сборник материалов Detection and Classification of Acoustic Scenes and Events (DCASE) 2017, Munich, Germany. – pp. 46-50.
13. Weiguo Wang, Luca Mottola, Yuan He, Jinming Li, Yimiao Sun, Shuai Li, Hua Jing, and Yulei Wang MicNest: Long-Range Instant Acoustic Localization of Drones in Precise Landing // In Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems (SenSys '22). Association for Computing Machinery, New York, – pp. 504-517.
14. Jakub Svatos, Jan Holub, Jan Belak System for an acoustic detection, localisation and classification // 24th IMEKO TC4 International Symposium 22nd International Workshop on ADC and DAC Modelling and Testing IMEKO TC-4 2020. – pp. 62-69.
15. Goetze S., Schroder J., Gerlach S., Hollosi D., Appell J.-E., Wallhoff F. Acoustic Monitoring and Localization for Social Care // Journal of Computing Science and Engineering. – V. 6. no. 1. – pp. 40-50.