

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.04.02 «Информатика и вычислительная техника»
 Отделение школы (НОЦ) Отделение информационных технологий

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Применение генеративно-сопоставительных сетей в задаче генерации сложных музыкальных произведений

УДК 004.7.032:004.312.26:781

Студент:

Группа	ФИО	Подпись	Дата
8ВМ93	Колобов Ростислав Сергеевич		

Руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ ИШИТР	Спицын Владимир Григорьевич	Д.Т.Н.		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Верховская Марина Витальевна	К.Э.Н.		

По разделу «Социальная ответственность»:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ООД ШБИП	Федоренко Ольга Юрьевна	Д.М.Н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ ИШИТР	Спицын Владимир Григорьевич	Д.Т.Н.		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

по направлению 09.04.01 Информатика и вычислительная техника

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий
УК(У)-2	Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	Способен организовывать и руководить работой команды, выработывая командную стратегию для достижения поставленной цели
УК(У)-4	Способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
Общепрофессиональные компетенции	
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
ОПК(У)-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач
ОПК(У)-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями
ОПК(У)-4	Способен применять на практике новые научные принципы и методы исследований
ОПК(У)-5	Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ОПК(У)-6	Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования
ОПК(У)-7	Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий
ОПК(У)-8	Способен осуществлять эффективное управление разработкой программных средств и проектов
Профессиональные компетенции	
ПК(У)-1	Способен разрабатывать и администрировать системы управления базами данных
ПК(У)-2	Способен проектировать сложные пользовательские интерфейсы
ПК(У)-3	Способен управлять процессами и проектами по созданию (модификации) информационных ресурсов
ПК(У)-4	Способен осуществлять руководство разработкой комплексных проектов на всех стадиях и этапах выполнения работ
ПК(У)-5	Способен проектировать и организовывать учебный процесс по образовательным программам с использованием современных образовательных технологий

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.04.02 «Информатика и вычислительная техника»
 Отделение школы (НОЦ) Отделение информационных технологий

УТВЕРЖДАЮ:
 Руководитель ООП

 (Подпись) (Дата)

Спицын В.Г.
 (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации

Студенту:

Группа	ФИО
8ВМ93	Колобов Ростислав Сергеевич

Тема работы:

Применение генеративно-состязательных сетей в задаче генерации сложных музыкальных произведений	
Утверждена приказом директора	46-3/с от 15.02.2021
Срок сдачи студентом выполненной работы:	01.06.2020

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	Данные ряда съемок территорий пихтовых насаждений, поврежденных уссурийским полиграфом, и кедровых насаждений, поврежденных союзным короедом, а также соответствующая экспертная разметка в виде точек различных цветов на кронах деревьев.
Перечень подлежащих исследованию, проектированию и разработке вопросов	<ul style="list-style-type: none"> – Исследование современных подходов к генерации музыки с помощью генеративно-состязательных сетей; – автоматизация процесса получения и подготовки данных; – исследование возможностей нейронных сетей в задаче сжатия музыкальных данных; – проведение экспериментов с различными типами генеративно-состязательных сетей – финансовый менеджмент; – социальная ответственность; – раздел на иностранном языке.
Перечень графического материала	<ul style="list-style-type: none"> – Схемы реализованных моделей; – примеры сгенерированных данных

Консультанты по разделам выпускной квалификационной работы	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Верховская Марина Витальевна, доцент ОСГН ШБИП, к.э.н.
Социальная ответственность	Федоренко Ольга Юрьевна, профессор ООД ШБИП, д.м.н.
Раздел на иностранном языке	Розанова Яна Викторовна, Старший преподаватель ОИЯ
Названия разделов, которые должны быть написаны на русском и иностранном языках:	
Разделы на русском: обзор литературы; получение и подготовка данных для обучения; эксперименты с АЕ и ВАЕ; эксперименты с сверточными генеративно-сопоставительными сетями; финансовый менеджмент, ресурсоэффективность и ресурсосбережение; социальная ответственность.	
Раздел на английском: Downloading and preparing data for training	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику:	10.03.2020
---	------------

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ ИШИТР	Спицын Владимир Григорьевич	д.т.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ93	Колобов Ростислав Сергеевич		

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.04.02 «Информатика и вычислительная техника»
 Уровень образования Магистратура
 Отделение школы (НОЦ) Отделение информационных технологий
 Период выполнения _____ 2020/2021 учебный год

Форма представления работы:

Магистерская диссертация

**КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
выполнения выпускной квалификационной работы**

Срок сдачи студентом выполненной работы:	01.06.2020
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
01.04.2021	Автоматизация получения и подготовки данных	20
15.04.2021	Проведение экспериментов с АЕ и ВАЕ	10
27.05.2021	Проведение экспериментов со сверхточными генеративно-состязательными сетями	40
01.06.2021	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
01.06.2021	Социальная ответственность	10
01.06.2021	Раздел на иностранном языке	10

СОСТАВИЛ:

Руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ ИШИТР	Спицын Владимир Григорьевич	Д.Т.Н.		

СОГЛАСОВАНО:

Руководитель ООП:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ ИШИТР	Спицын Владимир Григорьевич	Д.Т.Н.		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
8ВМ93	Колобов Ростислав Сергеевич

Школа	Инженерная школа информационных технологий и робототехники	Отделение (НОЦ)	Информационных технологий
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	<i>Использовать действующие ценники и договорные цены на потребленные материальные и информационные ресурсы, а также указанную в МУ величину тарифа на эл. энергию</i>
2. Планирование и формирование бюджета научных исследований	<i>Районный коэффициент 30 %; Коэффициент дополнительной заработной платы 12 %; Накладные расходы 16 %.</i>
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	<i>Действующие ставки единого социального налога и НДС</i>

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Планирование и формирование бюджета научных исследований	<i>Формирование плана и графика разработки: - определение структуры работ, - определение трудоемкости работ, - создание диаграммы Ганта. Формирование бюджета затрат на разработку: - затраты на специальное оборудование; - заработная плата (основная и дополнительная), - социальные отчисления, накладные расходы.</i>
2. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности разработки	<i>Определение потенциального эффекта разработки.</i>

Перечень графического материала (с точным указанием обязательных чертежей):

1. Диаграмма Ганта
2. Таблицы

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Верховская М. В.	к. э. н.		01.03.2021

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ93	Колобов Ростислав Сергеевич		01.03.2021

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8ВМ93	Колобов Ростислав Сергеевич

Школа	Инженерная школа информационных технологий и робототехники	Отделение (НОЦ)	Информационных технологий
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

Тема ВКР:

Применение генеративно-состязательных сетей в задаче генерации сложных музыкальных произведений	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Целью работы является применение генеративно-состязательных сетей в задаче генерации музыки. Оборудованием, на котором осуществляется работа является персональный компьютер. Рабочим местом является жилое помещение с площадью в 18 кв.м. Помещение оборудовано отоплением, вентиляцией и соответствует нормам освещенности. Имеется одноместный стол с настольной светодиодной лампой мощностью 11 Вт.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности: - специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; - организационные мероприятия при компоновке рабочей зоны.	<ul style="list-style-type: none"> - ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования; - Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 27.12.2018); - СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95; - ГОСТ 12.0.003-2015 ССБТ. Опасные и вредные производственные факторы. Классификация; - СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений; - СН 2.2.4/ 2.1.8.562-96 Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки. Санитарные нормы; - ГОСТ 12.1.038-82 ССБТ Электробезопасность; - ГОСТ 12.1.004-91 ССБТ Пожарная безопасность; - СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания.
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	<p>Анализ выявленных вредных факторов:</p> <ul style="list-style-type: none"> - Недостаточная освещенность рабочей зоны; - Отклонение параметров микроклимата; - Повышенный уровень шума; - Психофизиологические факторы (монотонность труда, нервно-психические перегрузки, перенапряжение зрительных анализаторов). <p>Анализ выявленных опасных факторов:</p> <ul style="list-style-type: none"> - Поражение электрическим током; - Статическое электричество; - Короткое замыкание.

3. Экологическая безопасность:	Атмосфера: воздействие отсутствует; Гидросфера: воздействие отсутствует; Воздействие на литосферу происходит при утилизации: <ul style="list-style-type: none"> – компьютера и периферийных устройств (принтеры, МФУ, веб-камеры, наушники, колонки, телефоны); – люминесцентных ламп; – макулатуры.
4. Безопасность в чрезвычайных ситуациях:	Возможные ЧС: аварии на системах жизнеобеспечения населения; – Наиболее типичная ЧС: возникновение пожара на рабочем месте.

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ООД ШБИП	Федоренко О. Ю.	Д.М.Н.		01.03.2021

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ93	Колобов Ростислав Сергеевич		01.03.2021

Реферат

Выпускная квалификационная работа выполнена на 85 с., содержит 24 рис., 16 табл., 49 источников, 2 приложения.

Ключевые слова: генеративно-состязательная сеть, машинное обучение, автоэнкодер, генерация музыки, автоматизация получения и подготовки данных.

Целью данной работы является применение генеративно-состязательных сетей в задаче генерации сложных музыкальных произведений.

В выпускной квалификационной работе рассмотрены современные подходы в генерации музыки с использованием генеративно-состязательных сетей. Разработан ряд скриптов, позволяющих автоматизировать получение и подготовку данных для последующего обучения нейронных сетей. Проведены эксперименты по понижению размерности данных с применением автоэнкодеров и вариационных автоэнкодеров. Были собраны и обучены различные типы генеративно-состязательных сетей, а также приведены результаты их обучения.

Содержание

Введение.....	12
1. Обзор литературы.....	13
1.1 Генеративно-сопязательные сети.....	13
1.2 MuseGAN	14
1.3 WaveGAN.....	15
1.4 GANSynth.....	17
2. Получение и подготовка данных для обучения	20
2.1 Загрузка неподготовленных данных из видеохостинга Youtube	20
2.2 Подготовка данных	21
2.2.1 Преобразование аудио в одинаковый формат	21
2.2.2 Разделение аудио на музыкальные составляющие	21
2.2.3 Удаление шумов из вокальной партии	22
2.2.4 Разделение аудио на отрезки, содержащие вокал	22
2.2.5 Извлечение акустических характеристик. Спектрограммы..	23
3. Эксперименты с АЕ и ВАЕ	27
3.1 Обучение сети автоэнкодера	27
3.2 Обучение сети вариационного автоэнкодера.....	29
4. Эксперименты с сверточными генеративно-сопязательными сетями	32
4.1 Глубокая сверточная генеративно-сопязательная сеть	32
4.2 Эксперименты с генеративно-сопязательной сетью, использующей функцию потерь Вассерштейна и градиентный штраф .	33
5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	39
6. Социальная ответственность.....	54

Заключение	68
Список источников	69
Приложение А (Downloading and preparing data for training)	73
Приложение Б.....	85

Введение

Синтез звука для определенных областей имеет множество практических применений в творческом звуковом дизайне для музыки и кино. Музыканты и артисты просматривают большие базы данных звуковых эффектов, чтобы найти конкретные аудиозаписи, подходящие для конкретных сценариев. Эта стратегия кропотлива и может не дать результатов, если идеальный звуковой эффект отсутствует в библиотеке.

Несмотря на огромное, постоянное растущее количество музыкального контента в интернете, создание новых музыкальных произведений все еще остается актуальным, как актуально создание произведений других видов искусства. Появляются и исчезают новые жанры, исполнители и музыканты. Вместе с тем становится все более заметным рост интереса к автоматической генерации музыкальных произведений. Однако существующие на текущий момент результаты в большинстве случаев являются монофоническими или используют синтезированные звуки музыкальных инструментов. Это в свою очередь оставляет большой простор для исследований.

В данной работе рассматриваются современные подходы к генерации музыкальных произведений с помощью генеративно-состязательных сетей, их особенности. Рассматривается процесс автоматизированного получения, и подготовки данных для обучения. Также, исследуется возможность представления данных, посредством извлечения скрытого вектора из обученного вариационного автоэнкодера. Проводятся эксперименты по созданию музыки с помощью генеративно-состязательных сетей.

Объектом исследования является изучение моделей генеративно-состязательных сетей для генерации аудиопоследовательностей.

Предмет исследования: исследование возможностей генеративно-состязательных сетей создавать реалистичные многоканальные аудиопоследовательности.

1. Обзор литературы

1.1 Генеративно-сопязательные сети

Генеративно-сопязательные сети [1] – класс нейронных сетей являющихся альтернативой вариационным автоэнкодерам, использующий обучение без учителя, позволяя выделять скрытые пространства данных[2]. Они реализованы, как система из двух нейронных сетей – генератора и дискриминатора, соревнующихся друг с другом в условиях игры с нулевой суммой. Первая сеть, называемая генератором, получает на вход случайный вектор (точку скрытого пространства), декодируя его в искусственный экземпляр данных. Вторая сеть, называемая дискриминатором, получает на вход экземпляры реальных и искусственных данных и определяет, какие экземпляры взяты из обучающего набора, а какие созданы генератором. [3]

Генеративно-сопязательная сеть является системой, в которой минимум функции оптимизации не фиксирован, в противоположность другим типам нейронных сетей. Это динамическая система, в которой процесс оптимизации стремится к равновесию потерь генератора и дискриминатора.[2]

В начале процесса обучения генератор выводит искаженные изображения, а дискриминатор оценивает их случайным образом. Главная особенность генеративно-сопязательной сети – попеременное обучение. В процессе обучения генератора веса дискриминатора фиксируются, а генератор постепенно совершенствуется в создании более похожих на реальные образцов. Далее, фиксируются уже веса генератора, а дискриминатор, адаптируясь, учится правильно определять сгенерированные наблюдения, что вынуждает сеть генератора повышать качество генерируемых образцов. [5]

Целевая функция дискриминатора имеет следующий вид:

$$L_D = \frac{1}{m} \sum_{i=1}^m [\log D(x^i) + \log (1 - D(G(z^i)))] \quad (1.1)$$

А целевая функция генератора:

$$L_G = \frac{1}{m} \sum_{i=0}^m \log D(G(z^i)) \quad (1.2)$$

Тогда, исходя из теории игр, генеративно-сопоставительную сеть можно определить как «минимаксную» игру двух игроков со значением функции $V(G, D)$ [1]:

$$\min_G \max_D V(G, D) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1.3)$$

Архитектурно, генератор похож на декодер автоэнкодера, а дискриминатор на энкодер, за исключением того, что выходом дискриминатора является не скрытое представление, а вероятность того, что поданное на его вход изображение является реальным.

В ходе исследования литературы был обнаружен ряд исследований, использующих различные подходы и наборы данных для создания музыкальных произведений с помощью генеративно-сопоставительных сетей. Представленный ряд работ не является исчерпывающим, однако отражает большую часть подходов к обучению генеративно-сопоставительных моделей, используемых для обучения данных и их представлений.

1.2 MuseGAN

В статье [5] представлены модели генеративно-сопоставительных сетей для генерации аудио, состоящего из нескольких музыкальных партий. Генерация осуществлялась с использованием трех моделей. В первой (Jamming model), каждая музыкальная партия генерировалась отдельной генеративно-сопоставительной сетью. Во втором случае (Composer model), генерация всех партий осуществлялась из одного вектора шума. Третья модель (Hybrid model) является объединением подходов двух предыдущих. При этом, помимо использования отдельных векторов шума для каждой партии использовался общий для всех партий вектор, который задавал общую тему аудио, а на вход дискриминатора подавались все партии одновременно. На рисунке 1 можно увидеть общее представление моделей, предложенных авторами статьи.

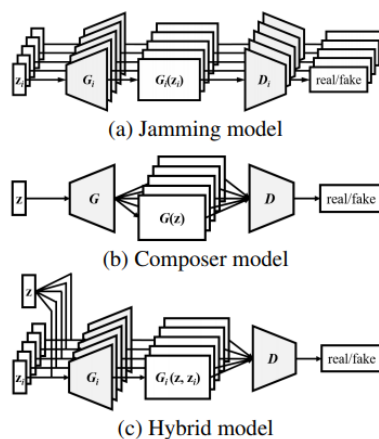


Рисунок 1 – Общее представление моделей

Особенностью работы является обучение на наборе данных состоящем из более чем ста тысяч тактов рок музыки в формате MIDI [6], полученных из открытых источников. Модели могут генерировать связную музыку из четырех тактов с нуля. Примеры сгенерированных MIDI дорожек приведены на рисунке 2.

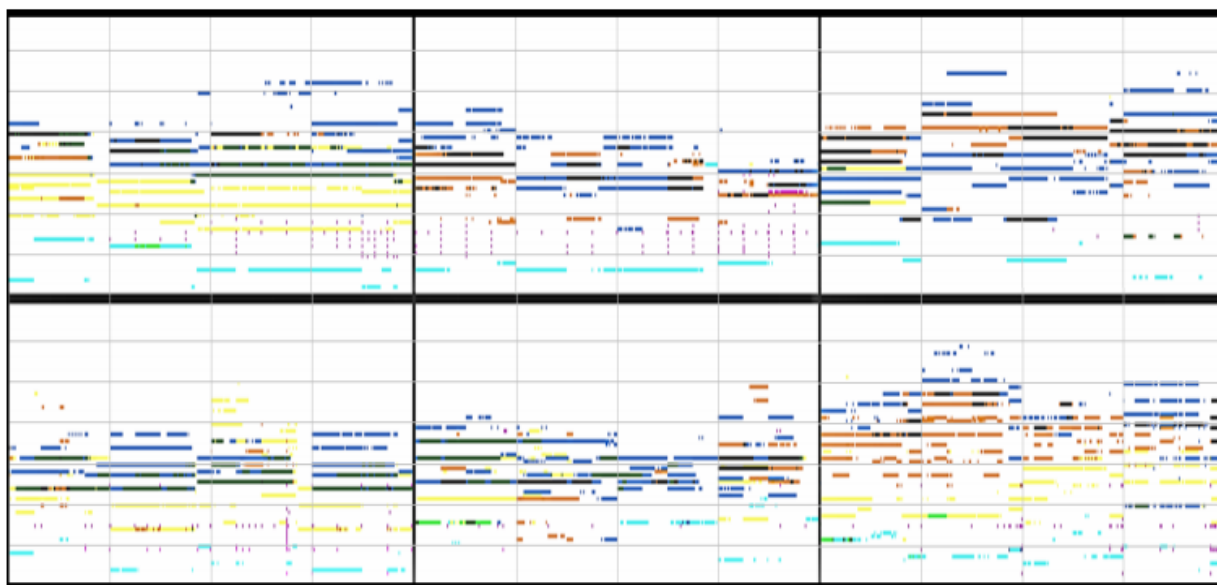


Рисунок 2 – Пример результатов генерации второй и третьей модели

1.3 WaveGAN

WaveGAN [7] – генеративно-состязательная сеть для генерации музыки, основанная на архитектуре DCGAN [8], генератор которой представлен на рисунке 3.

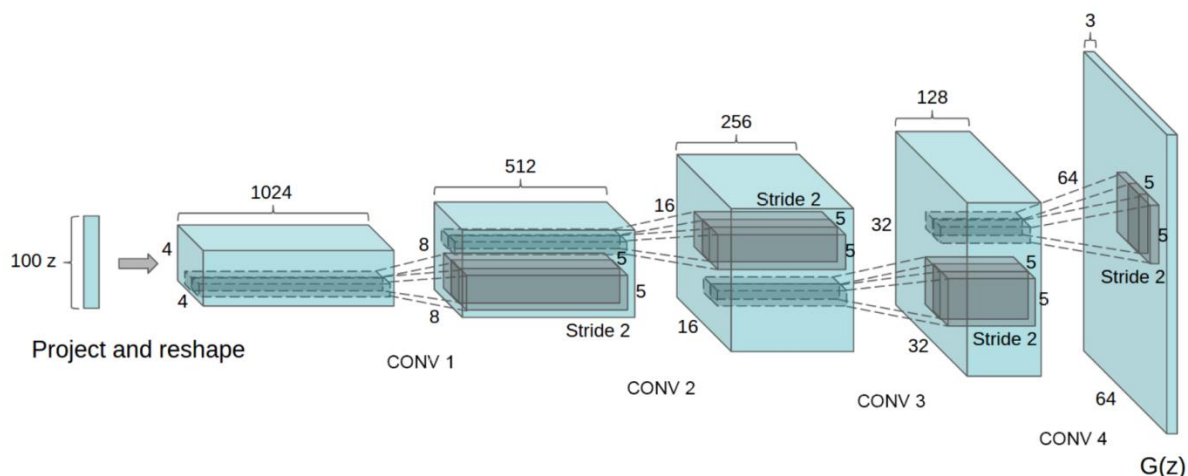


Рисунок 3 - Архитектура генератора DCGAN

Сеть использует сверточные слои как в генераторе, так и дискриминаторе, позволяя генерировать аудио частотой 16 кГц. Так как представление аудио, в отличие от изображений уже не является двумерным, исходные двумерные свертки заменяются одномерными: 2D-фильтры 5x5 в DCGAN теперь становятся одномерными 25x1 фильтрами в WaveGAN.

Чтобы решить проблему, свойственную транспонированным сверткам, а именно появление артефактов в виде шахматной доски, которые упрощают процесс классификации дискриминатором, что значительно затрудняет процесс обучения генератора, авторы вводят решение под названием Phase Shuffle – метод удаления артефактов тонального шума. Данный метод случайным образом изменяет фазу активации каждого слоя перед передачей его в следующий слой [8] (рисунок 4).

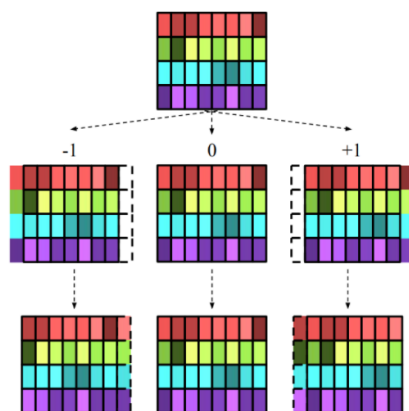


Рисунок 4 – Работа метода Phase Shuffle

Сдвиг фазы применяется только к дискриминатору, усложняя классификацию, требуя неизменности фазы входного сигнала.

Также WaveGAN:

- не использует пакетную нормализацию;
- обучение производится по сценарию WGAN-GP [9].

Сеть обучалась генерации голосовых команд, а также воспроизведению пения птиц, звуков барабанов и пианино. Результаты генерации, взятые из оригинальной статьи приведены на рисунке 5.

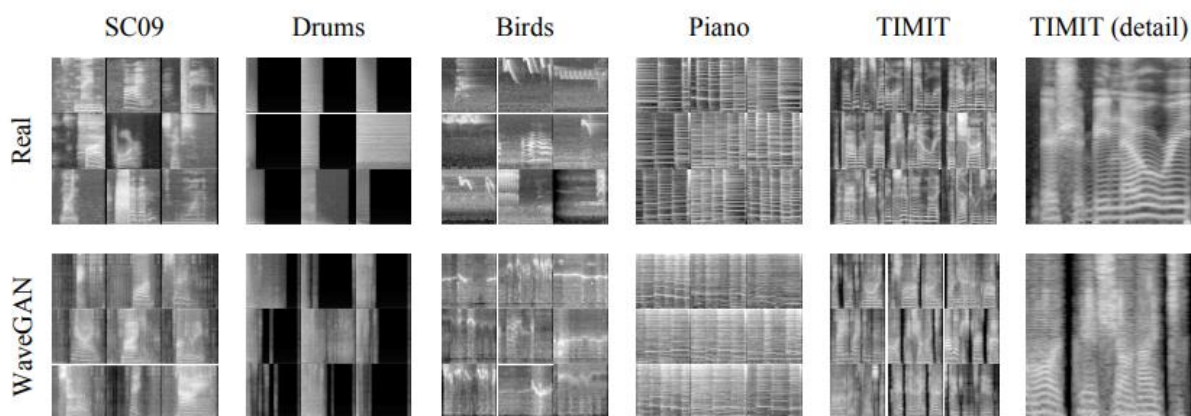


Рисунок 5 – Результаты работы WaveGAN

1.4 GANSynth

GANSynth [10] является адаптацией модели WaveGAN. При этом используется подход, применяемый в PGGAN [4], который заключается в постепенном увеличении количества и размера слоев, что позволяет сети сначала изучать крупные структуры, а в дальнейшем переходить к более мелким деталям (рисунок 6). Поэтапное изменение размера сети делает процесс обучения более стабильным, а результаты более качественными.

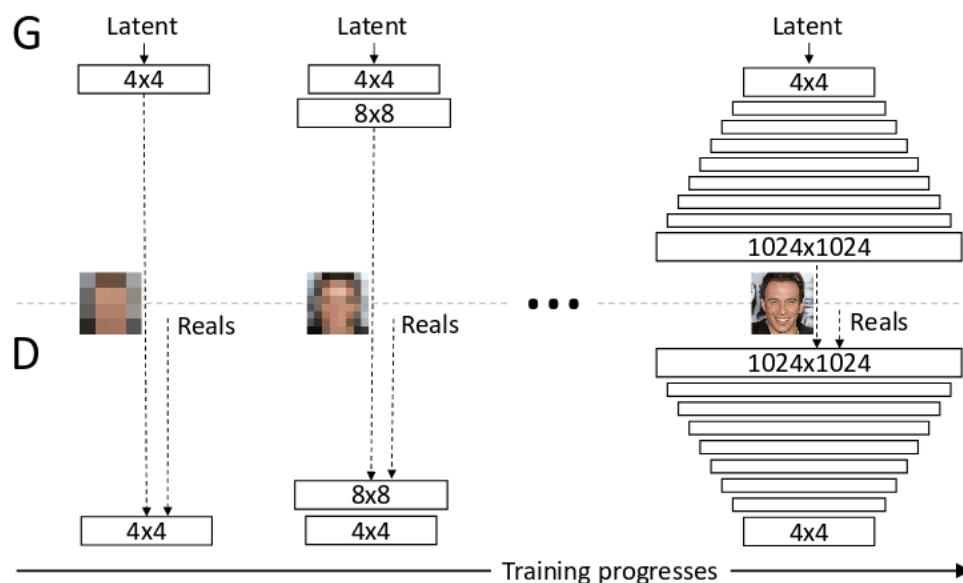


Рисунок 6 – Увеличение количества слоев во время обучения PGGAN

Главной особенностью GANSynth является использование спектрограмм мгновенных частот (instantaneous frequencies). Для результатов быстрого преобразования Фурье производится разворот фазы по границе 2π , после чего берется ее производная для получения мгновенной радиальной частоты, которая представляет постоянную взаимосвязь между частотой звука и частотой кадра[10]. Графическое изображение данного процесса, предоставленное авторами оригинальной статьи можно увидеть на рисунке 7. Данное преобразование выполняется по причине того, что большая часть звука состоит из локально когерентных волн с локальной периодичностью. При этом, как транспонированные свертки, так и кратковременное преобразование Фурье имеют заданный размер окна и его шаг смещения, в результате чего имеет место смещение фаз окна преобразования и фазы звука. Поэтому, транспонированные свертки вынуждены охватывать все возможные частоты и фазы для сохранения фазовой когерентности, что в свою очередь значительно усложняет процесс обучения.

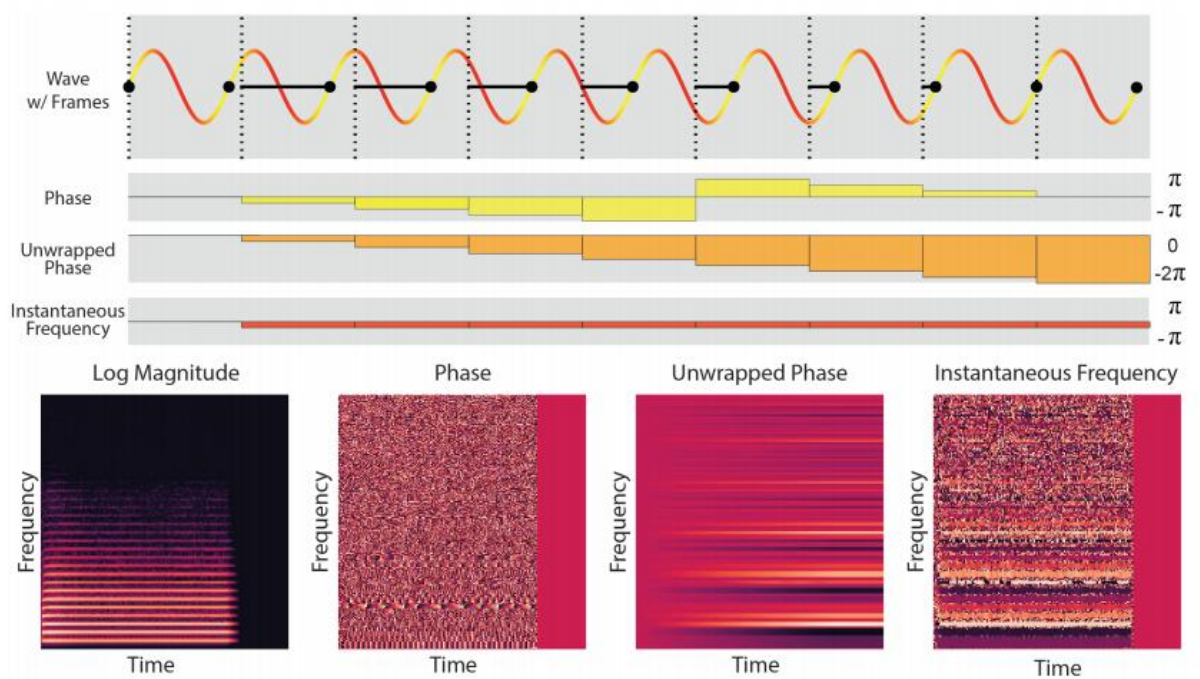


Рисунок 7 – Преобразование спектрограммы в спектрограмму мгновенных частот

Обучение GANSynth производилось на наборе данных NSynth [11], который состоит из 300 тысяч аудио длительностью четыре секунды, содержащих музыкальные ноты, выровненные относительно длительности аудио, произведенные 1000 различных инструментов. Пример сгенерированной спектрограммы, который приводят авторы статьи, представлен на рисунке 8.

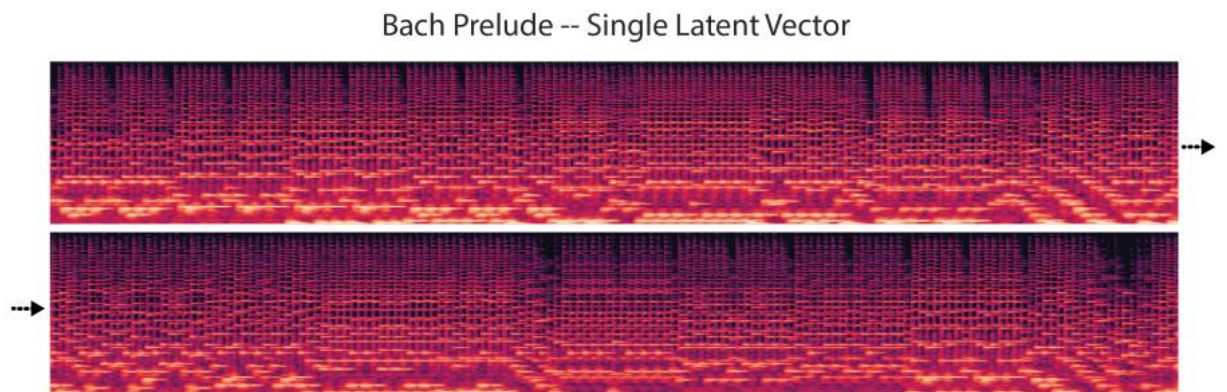


Рисунок 8 – Пример сгенерированного аудио

2. Получение и подготовка данных для обучения

В целях сбора обучающего набора данных был разработан ряд скриптов позволяющих автоматизировать получение и подготовку данных. В работе, в качестве данных для обучения используются, аудиодорожки полученные из открытого видеохостинга Youtube[2].

2.1 Загрузка неподготовленных данных из видеохостинга Youtube

Для загрузки неподготовленных данных разработано два скрипта, использующих в своей основе библиотеку youtube_dl [13]. Скрипты позволяют выполнять следующие операции:

- Получение списка URL каждого видео;
- Непосредственная выгрузка данных, на основе полученного на предыдущем этапе списка.

Получить список URL видео можно:

- Указав URL конкретного канала Youtube, тогда будут получены URL всех видео, относящихся к данному каналу;
- Указав ключевое слово, на основе которого будет производиться поиск среди выложенных на Youtube видео.

Далее, используя полученные URL производится непосредственная выгрузка аудиодорожек. При этом извлекается только аудиосоставляющие каждого видео и сохраняются в форматах webm или mp4. Однако, при большом количестве запросов видеохостинг может ограничивать количество запросов в день вплоть до ограничения доступа к сервису с конкретного IP-адреса. Поэтому скрипт выгрузки данных также предоставляет возможность указать список прокси адресов, использование которых позволяет выгружать аудиодорожки несколько потоков, используя одновременно несколько прокси.

В данном случае использовалась загрузка с помощью указания URL таких Youtube-каналов, как:

- фольк-рок группы «Мельница»[14];
- Диана Арбенина и «Ночные Снайперы» [15];

- «Город 312» [16];
- «Земфира» [17].

Во избежание попадания в обучающий набор аудио из промо видео или интервью производится дополнительное отсеивание, основанное на длительности видео. Сохраняются только те дорожки, длительность которых составляет не менее двух с половиной и не более пяти минут.

2.2 Подготовка данных

Чтобы облегчить процесс обучения, необходимо упростить обучающие данные и привести их к общему виду.

2.2.1 Преобразование аудио в одинаковый формат

Необходимо преобразовать полученные на предыдущем этапе данные в аудиоформат, а также изменить частоту дискретизации. Преобразование осуществляется с помощью утилиты командной строки `ffmpeg`[18] со следующими параметрами:

- частота дискретизации: 16 КГц;
- 16 бит;
- моно;
- формат `wav`.

2.2.2 Разделение аудио на музыкальные составляющие

Для того, чтобы упростить обучение сети, производится разделение аудиозаписи на музыкальные партии:

- вокал;
- ударные;
- остальные инструменты.

Предполагается, что подобное разделение упростит изучение сложных композиционных распределений.

Разделение аудио на музыкальные партии осуществляется с помощью применения библиотеки с открытым исходным кодом `Spleeter`[19]. Данная библиотека предоставляет возможность использовать предобученные нейросетевые модели, позволяющие разделять исходные аудио на две, четыре и

пять музыкальных партий. Одним из достоинств данной библиотеки является высокая скорость работы, разделение аудио на партии с использованием графического процессора может занимать до 0,01 от исходной длительности аудио. Так как библиотека не предоставляет возможности разделить только на три вышеописанные партии, разделение осуществляется с помощью четырехпартийной модели, после чего, партии, не относящиеся к вокалу и ударным объединяются с помощью утилиты ffmpeg.

2.2.3 Удаление шумов из вокальной партии

В ходе изучения разделенных на партии аудиоданных было обнаружено, что в вокальных партиях часто встречаются артефакты, похожие на эхо от других партий. Так как подобные артефакты не имеют отношения к вокалу, появляется необходимость избавиться от них.

Для удаления артефактов из вокальных партий используется разработка, основанная на нейросетевом подходе, Facebook Denoiser [20], которая является реализацией статьи [21]. Используемая в разработке модель основана архитектуре энкодер-декодер LSTM [22]. Модель обучена специально для удаления шумов в аудио, содержащих голос, позволяя удалять различные виды фонового шума, включая стационарные и нестационарные шумы, а также комнатную реверберацию. Удаление шумов используется исключительно для вокальных партий.

2.2.4 Разделение аудио на отрезки, содержащие вокал

Еще одним упрощением является обучение не на всем аудио, а на некоторой его части, в обучении будут использоваться только отрезки аудио, содержащие речь не менее 50% от своей длительности. В целях разделения аудио на отрезки, содержащие вокал, используется библиотека librosa [23], а в частности ее метод split, разбивающий аудио на промежутки, не содержащие тишину. Для каждого аудио создается дополнительный массив, такого же размера, содержащий единицы в тех промежутках, где в аудио присутствует вокал. Далее, на основе полученного массива производится отбор участков, содержащих речь не менее 50% от своей длительности, с помощью задаваемого

размера окна и шага его смещения. В данной работе используется окно размером 330000 с шагом 32000, что соответствует примерно 20 секундам аудио с шагом в две секунды. Если длительность речи в окне составляет не менее 10 секунд, то из аудиофайлов партий вокала, ударных и остальных инструментов извлекаются окна с соответствующими координатами и передаются на следующий этап подготовки, в противном случае окно отбрасывается. Таким образом, результатом данного процесса является тензор формой $N \times 330000 \times 3$, где N – число подходящих окон, извлеченных из всех аудио.

2.2.5 Извлечение акустических характеристик. Спектрограммы.

Полученные на предыдущем шаге данные все еще представлены в виде последовательности мгновенных значений, считанных через равные промежутки времени (рисунок 9). Проводить обучение на таких данных довольно сложное и вычислительно затратное, поэтому имеет смысл извлечь частотные признаки меньшей размерности, но достаточно хорошо сохраняющие информативность.

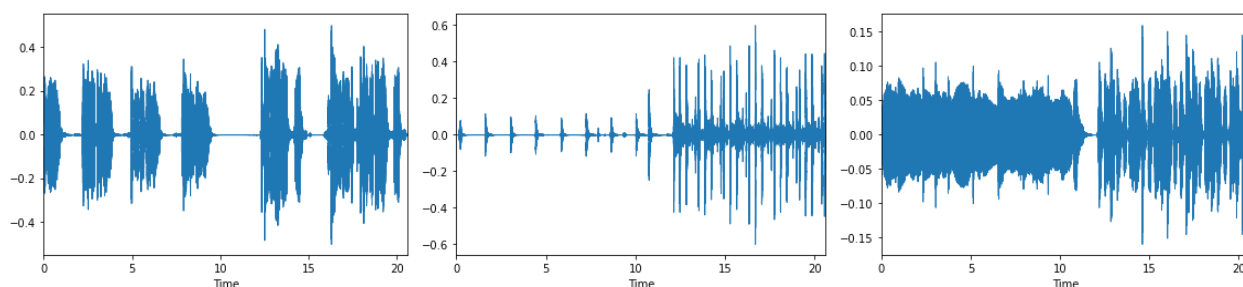


Рисунок 9 – Данные в виде мгновенных значений (вокал, ударные, остальное)

Спектрограммы являются результатом вычисления быстрого преобразования Фурье, применяемого над перекрывающимися оконными сегментами. Данное представление позволяет визуально представить амплитуду сигнала, меняющегося с течением времени на разных частотах. При этом, ось ординат преобразуется в логарифмическую шкалу, значение в конкретной точке преобразуются в децибелы [24]. На рисунке 10 приводится пример гистограмм.

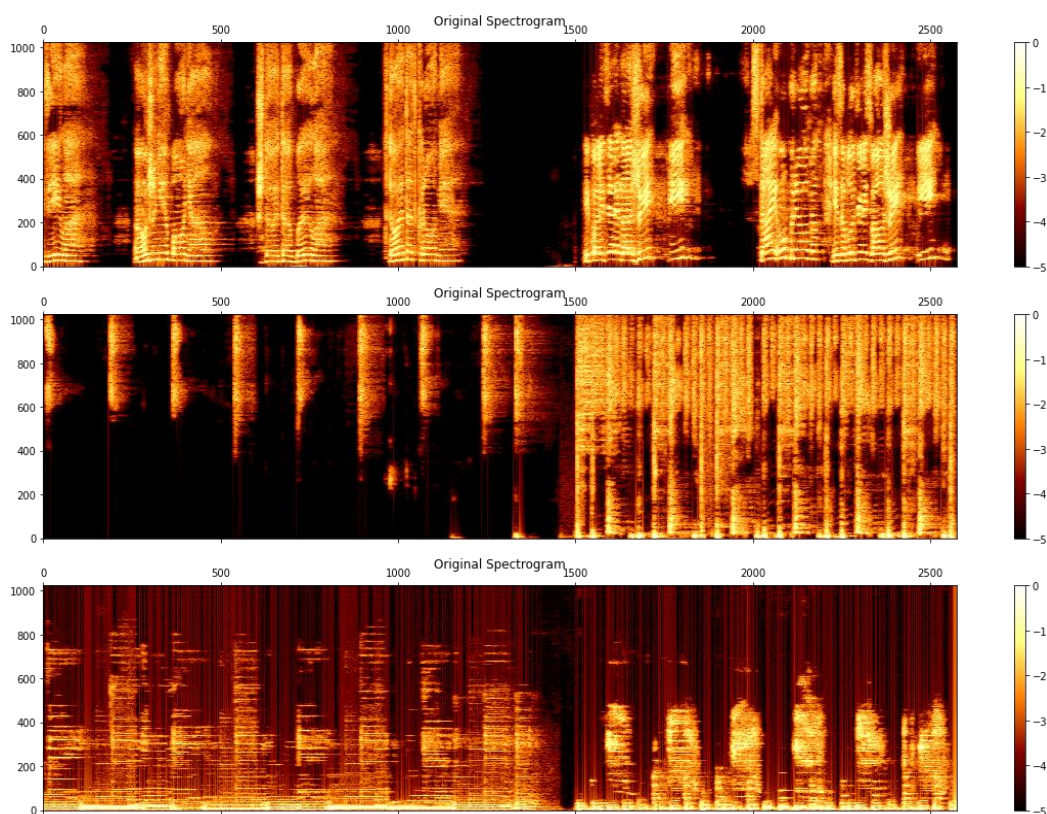


Рисунок 10 – Пример спектрограмм (вокал, ударные, остальное)

Однако люди не воспринимают частоты в линейном масштабе и лучше обнаруживают различия на более низких частотах, чем на более высоких. Также использование гистограмм в качестве данных для обучения все еще является вычислительно затратным, а наличие мелких деталей значительно усложнит обучение. Поэтому было решено упростить данные, ценой потери качества и преобразовать частотные значения спектрограмм в Мел-шкалу.

Мел спектрограмма является спектрограммой, в которой шкала частот преобразована в масштабе Мел – психофизической единицы высоты звука [25].

Преобразованные в Мел-шкалу спектрограммы представлены на рисунке 11.

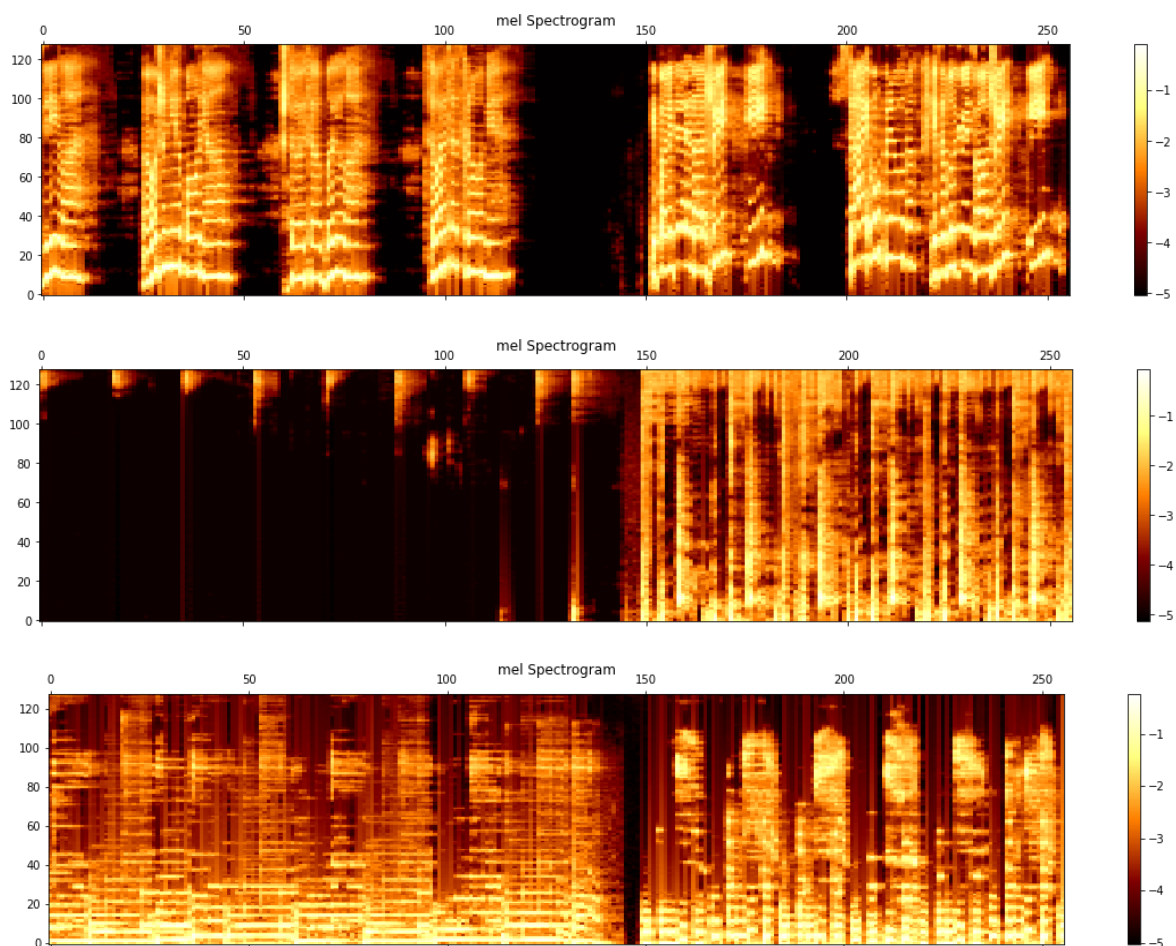


Рисунок 11 – Пример Мел-спектрограмм (вокал, ударные, остальное)

Так как преобразование заметно сжимает данные, происходит потеря некоторой информационной составляющей, однако восстановленные из Мел-спектрограмм аудиоданные все еще достаточно хорошо отображают исходные (рисунок 12).

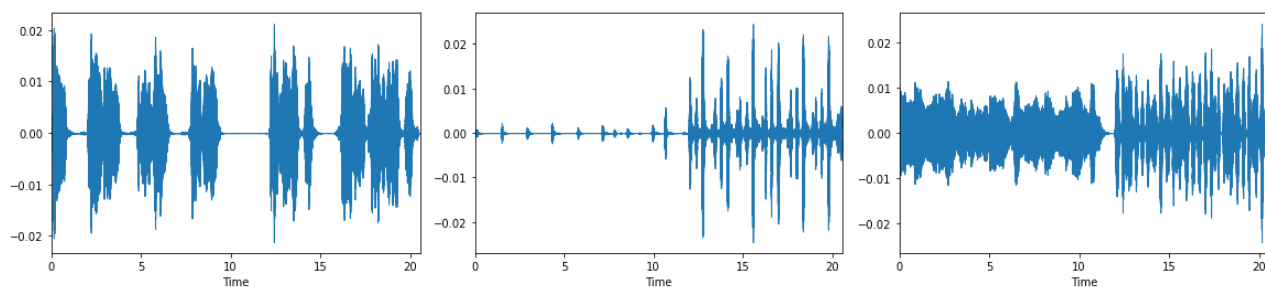


Рисунок 12 – Восстановленные аудио (вокал, ударные, остальное)

Преобразования производились со следующим набором параметров:

- размер окна для быстрого преобразования Фурье: 2048;
- размер шага окна: 128;

- число Мел-частотных компонент: 128;
- коэффициент сокращения: 10;
- начальная частота сэмплирования: 300;
- конечная частота сэмплирования: 16000.

Весь процесс преобразования основан на применении библиотек `numru` [26], `scipy` [27] и `librosa`.

Таким образом, обучающий набор представлен тензором формой $N \times 128 \times 256 \times 3$.

Всего было загружено 726 аудио, после процесса обработки которых размер обучающего набора составил 46270 экземпляров. Графическое представление процесса получения и подготовки данных представлено в приложении Б.

3. Эксперименты с АЕ и ВАЕ

Чтобы изучить возможность еще сильнее уменьшить размер данных для обучения имеет смысл обучить нейронные сети типа кодер-декодер и извлечь полученное скрытое представление.

3.1 Обучение сети автоэнкодера

В первую очередь необходимо изучить возможность нейронной сети создавать скрытое представление, позволяющее реконструировать данные.

Автоэнкодеры (АЕ) [28] относятся к нейронным сетям прямого распространения, позволяющим восстанавливать входной сигнал на выходе. Внутри нейронной сети имеется скрытый слой, в общем виде описывающий данные. Данный тип нейронных сетей конструируется в первую очередь таким образом, чтобы нейронная сеть не имела возможности идеально скопировать входной сигнал на выход, для этого размер скрытого представления должен быть меньше размерности входных данных. Однако, ввиду данного ограничения, входной сигнал восстанавливается с ошибками из-за потери информации при кодировании, что вынуждает нейронную сеть отбирать наиболее информативные признаки [4].

Нейронная сеть автоэнкодера состоит из двух частей: энкодера и декодера. Задачей энкодера является сжатие входных данных до размерности скрытого представления, а декодера восстановить исходный сигнал из полученного скрытого представления.

На данном этапе используется сверточный автоэнкодер, архитектура которого представлена на рисунке 13. При этом используется только часть обучающего набора, а именно только вокальная составляющая, ограниченная 64 значениями по оси времени и 64 значениями по оси частот.

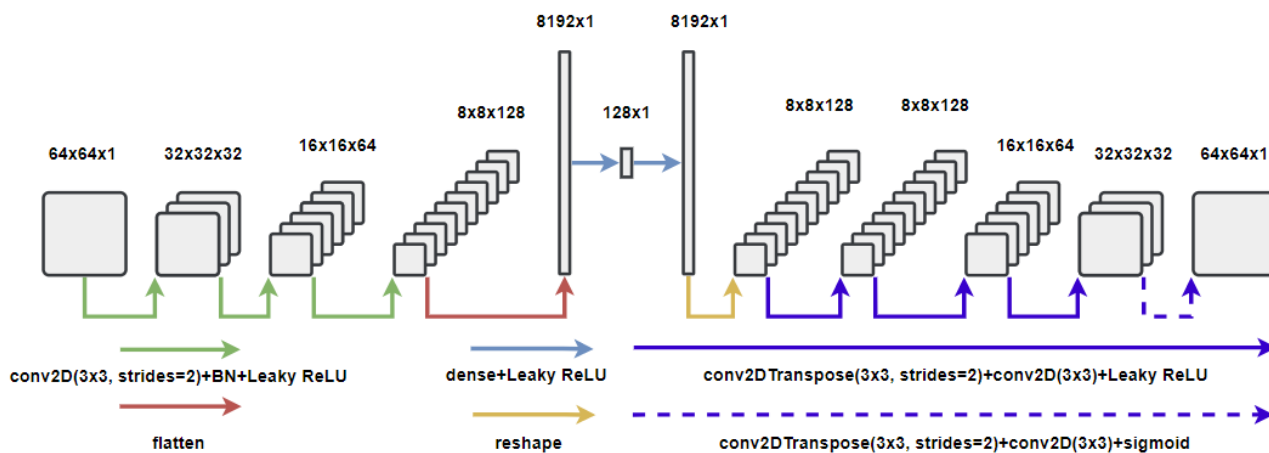


Рисунок 13 – Архитектура автоэнкодера

Данная архитектура была реализована и обучена с помощью TensorFlow [29] реализации с использованием высокоуровневого API Keras [30]. На выходном слое используется функция активации Sigmoid. Обучение производилось с использованием в качестве функции потерь бинарной кросс-энтропии, а в качестве оптимизатора adam с параметром скорости обучения равным 0,001. Обучение велось на протяжении 64 эпох, однако было остановлено на 50-й эпохе по причине отсутствия уменьшения функции потерь, достигнув значения 0,5595 на обучающей и 0.5471 на валидационной выборке

Результаты реконструкции входных данных представлены на рисунке 14.

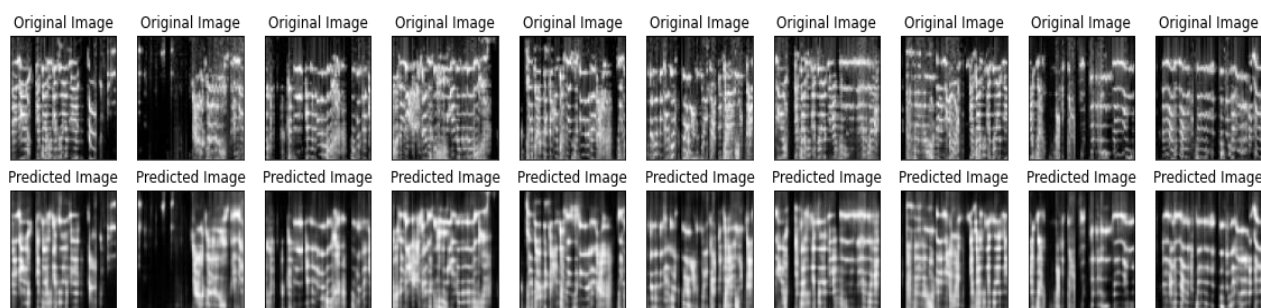


Рисунок 14 – Результаты реконструкции входных данных

Как видно из рисунка, сеть достаточно хорошо справилась с задачей реконструкции выходных данных. Однако, так как скрытое пространство сетей автоэнкодеров не является непрерывным, его нельзя использовать как полноценное представление данных, которые он не видел, а также для генерации новых. Поэтому имеет смысл перейти к следующему этапу, а

именно, обучение нейронной сети вариационного автоэнкодера, непрерывное скрытое представление которого исправляет недостатки текущей нейронной сети.

3.2 Обучение сети вариационного автоэнкодера

Нейронные сети автоэнкодеров отображают каждое изображение в точку в скрытом пространстве, которое не является непрерывным. В вариационном автоэнкодере [31] (VAE) каждое изображение отображается в многомерное нормальное распределение вокруг точки в скрытом пространстве. Непрерывность скрытого пространства достигается использованием двух скрытых векторов: вектора средних значений и вектора стандартных отклонений, которые вместе определяют многомерное нормальное распределение в скрытом пространстве. Часто также выбирается отображение в логарифм дисперсии, потому что он может принимать любые действительные значения в диапазоне $(-\infty, \infty)$, что соответствует естественному диапазону выходных значений узла нейронной сети, тогда как значения дисперсии всегда положительны. В итоге кодировщик будет преобразовывать каждое входное изображение в два вектора, μ – средних значений и $\log(\sigma^2)$ – логарифм дисперсии, которые вместе определяют многомерное нормальное распределение в скрытом пространстве. Преобразовать изображение в определенную точку z в скрытом пространстве можно путем выборки из этого распределения с помощью следующего уравнения [4]:

$$z = \mu + \sigma * \varepsilon, \quad (3.1)$$

где $\sigma = \exp(\log(\sigma^2)/2)$, а ε – точка, выбранная из стандартного нормального распределения.

Так как случайное значение выбирается из области вокруг μ , то все соседние точки должны декодироваться в похожие изображения. В таком случае, потери связанные с реконструкцией должны оставаться небольшими. Данная операция называется семплированием, а ее результат является выходом кодировщика.

Для генерации новых данных недостаточно только сопоставления векторов с распределением. Поскольку нет ограничений на то, какие значения принимают векторы среднего и стандартного отклонения, кодировщик может возвращать распределения с разными средними значениями для разных классов или кластеров с небольшим отклонением от среднего так, что кодировки не будут сильно отличаться от одного и того же образца. В результате закодированные распределения находятся далеко друг от друга. Поэтому в функцию потерь автоэнкодера вводится расстояние Кульбака-Лейблера (KL) [32], измеряющее то, насколько одно распределение вероятности отличается от другого. В данном случае нам необходимо знать, насколько полученное распределение отличается от нормального, функция потерь, использующая расстояние KL имеет следующий вид:

$$KL_{loss} = -0.5 * \sum(1 + \log(\sigma^2) - \mu^2 - \exp(\log(\sigma^2))) \quad (3.2)$$

Такая функция потерь вынуждает кодировщик распределять значения близко к стандартному нормальному распределению (среднее значение 0 и стандартное отклонение 1). В таком случае гарантируется, что распределения расположены близко и фактически перекрываются. Это повышает шанс того, что скрытое представление действительно непрерывно. Однако оптимизируя только расхождение KL, кодировщик будет случайным образом строить точки, близкие к центру. Необходимо чтобы он был более организован по классам, поэтому также нужно использовать потерю реконструкции. Сумма этих двух членов составляет функцию потерь вариационных автоэнкодеров.

Используемая в эксперименте сеть вариационного автоэнкодера имеет архитектуру, представленную на рисунке 15. Функция потерь, включает в себя не только ошибку реконструкции, но и расстояние Кульбака-Лейбнера.

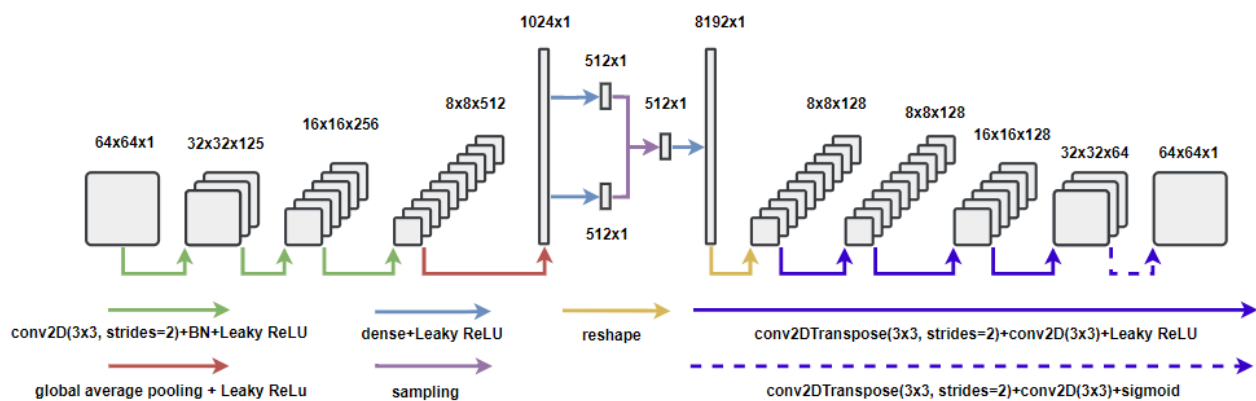


Рисунок 15 – Архитектура вариационного автоэнкодера

Обучение было остановлено на 150-й эпохе по причине отсутствия уменьшения функции потерь, при этом с уменьшением значения потерь при реконструкции происходил рост значения потерь KL, что говорит о том, что сеть не справляется с задачей генерации непрерывного представления.

Результаты реконструкции после обучения модели представлены на рисунке 16.

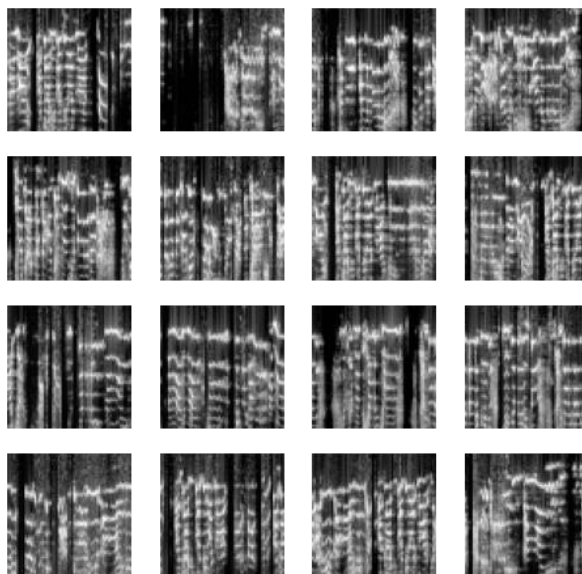


Рисунок 16 – Результаты реконструкции вариационным автоэнкодером

Как видно из рисунка результаты реконструкции похожи на реальные изображения спектрограмм, однако здесь же проявляется негативная для нашей задачи особенность вариационных автоэнкодеров, а именно размытость изображений, что вносит дополнительный шум в реконструируемое аудио. Данная особенность может заметно усложнить обучение последующих

генеративных моделей, что в совокупности с высоким значением потерь KL не позволяет использовать скрытое представление для генерации новых данных, поэтому было принято решение продолжить эксперименты с генеративно-состязательными сетями, используя данные спектрограмм.

4. Эксперименты с сверточными генеративно-состязательными сетями

Чтобы убедиться, что нейронная сеть может изучить распределения, которые характеризуют музыкальные партии, имеет смысл произвести обучение на наборе данных с некоторыми ограничениями. В данном случае для обучения используется только вокальная партия, при этом мы ограничиваем размерность каждого обучающего примера до $64 \times 64 \times 1$, что соответствует примерно пяти секундам аудио и 64 Мел-частотным компонентам. Нормализация данных производится до значений от -1 до 1.

4.1 Глубокая сверточная генеративно-состязательная сеть

В данном эксперименте используется сверточная генеративно-состязательная сеть, архитектура которой представлена на рисунке 17. Обучение было запущено с оптимизатором Адам, в качестве функции потерь использовалась бинарная кросс-энтропия. Планировалось обучать нейронную сеть на протяжении 200 эпох, однако, уже на 12 эпохе значения функций потерь генератора и дискриминатора начали неконтролируемо расти, а уже к 20 эпохе наблюдалось явление взрывного градиента. По этой причине обучение было остановлено. Пример сгенерированных изображений, полученных на 12-й эпохе, приведен на рисунке 18.

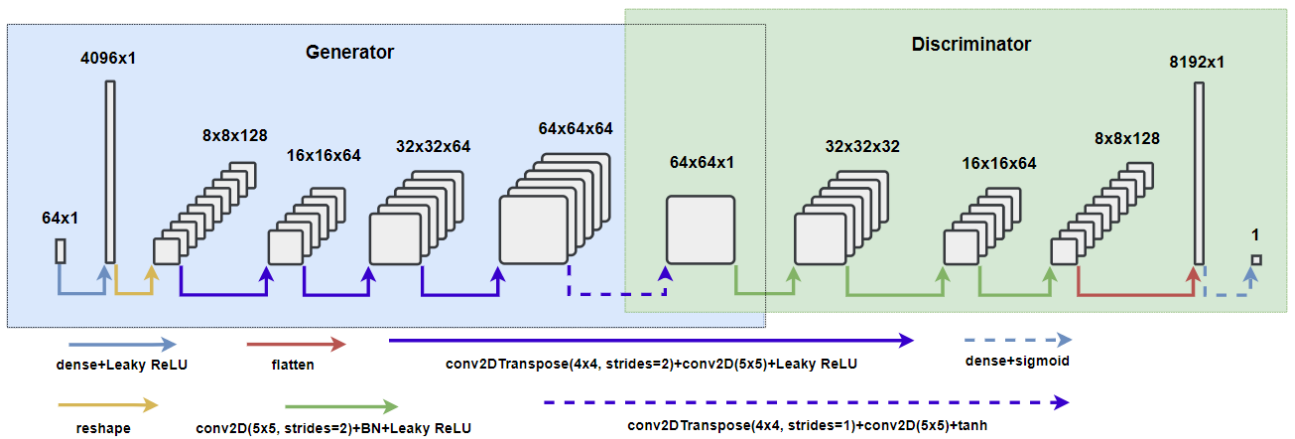


Рисунок 17 - Архитектура генеративно-сопоставительной сети

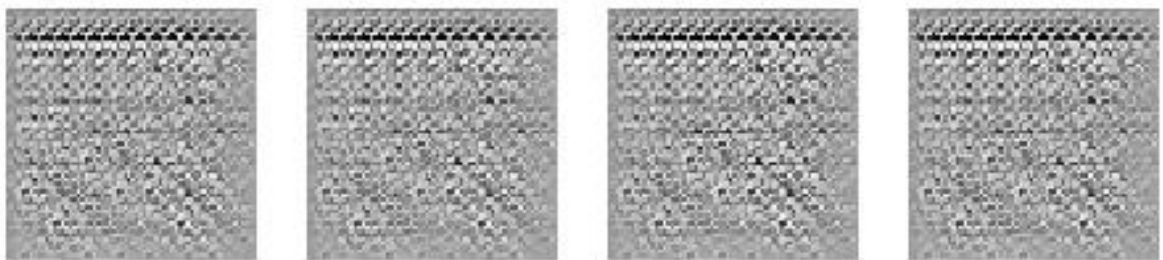


Рисунок 18 – Примеры сгенерированных изображений

Полученные изображения даже отдаленно не напоминают спектрограммы вокальной партии. Многократный перезапуск обучения приводил к аналогичным результатам, поэтому, было принято решение изменить архитектуру нейронной сети, а также воспользоваться более стабильными методами обучения.

4.2 Эксперименты с генеративно-сопоставительной сетью, использующей функцию потерь Вассерштейна и градиентный штраф

Расстояние Вассерштейна показывает минимальные затраты на транспортировку при преобразовании одного распределения к другому. Расстояние между распределением P_r реальных данных и распределением сгенерированных данных P_g математически определяется как наибольшая нижняя граница для любого транспортного плана[33]:

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} E_{(x,y) \sim \gamma} [\|x - y\|] \quad (4.1)$$

где, $P(P_r, P_g)$ – два неких распределения, а $\gamma(x, y)$ указывает, сколько «массы» необходимо перенести из точки x в точку y , чтобы преобразовать распределение P_r в распределение P_g .

При использовании функции потерь Вассерштейна необходимо отказаться от функции активации сигмоида в дискриминаторе, чтобы преобразование больше не ограничивалось значениями от нуля до единицы и могло принимать любые значения. Однако по этой причине величина потери может быть очень большой. Авторы оригинальной статьи WGAN [35] пишут о необходимости наложения дополнительного ограничения на дискриминатор, а именно, он должен быть одномерной липшицевой непрерывной функцией, что значит, для любых двух входных изображений должен удовлетворять следующему неравенству:

$$\frac{|D(x_1) - D(x_2)|}{|x_1 - x_2|} \leq 1 \quad (4.2)$$

Тогда, целевая функция потерь дискриминатора будет иметь вид:

$$L_c = \frac{1}{m} \sum_{i=1}^m [f(x^i) - f(G(z^i))], \quad (4.3)$$

функция потерь генератора:

$$L(G) = \frac{1}{m} \sum_{i=1}^m f(G(z^i)), \quad (4.4)$$

Чтобы реализовать ограничение Липшица производится усечение весов дискриминатора до диапазона от -0,01 до 0,01. Однако в работе [35] был представлен более надежный способ сделать функцию липшицевой – штраф за градиент. Ключевой идеей все также остается сохранение весов в компактном пространстве, но вместо усечения, авторы ограничивают норму градиента вывода дискриминатора относительно его выхода. Иными словами, общая функция потерь Вассерштейна дополняется штрафом за градиент между реальными и сгенерированными изображениями.

Функция потерь штрафа за градиент измеряет квадрат разности между нормой градиента прогнозов для входных изображений и 1. Естественно, модель будет стремиться найти веса, гарантирующие минимальный штраф за градиент, стараясь тем самым соответствовать ограничению Липшица. Однако,

вычисление подобного градиента повсюду во время обучения значительно повышало бы сложность вычислений, поэтому он оценивается только по нескольким точкам[4]. Для сохранения баланса используется набор интерполированных изображений, лежащих в случайно выбранных точках вдоль линий, соединяющих пары реальных и сгенерированных изображений.

Архитектура используемой генеративно-состязательной сети с функций потерь Вассерштейна и штрафом градиента представлена на рисунке 19.

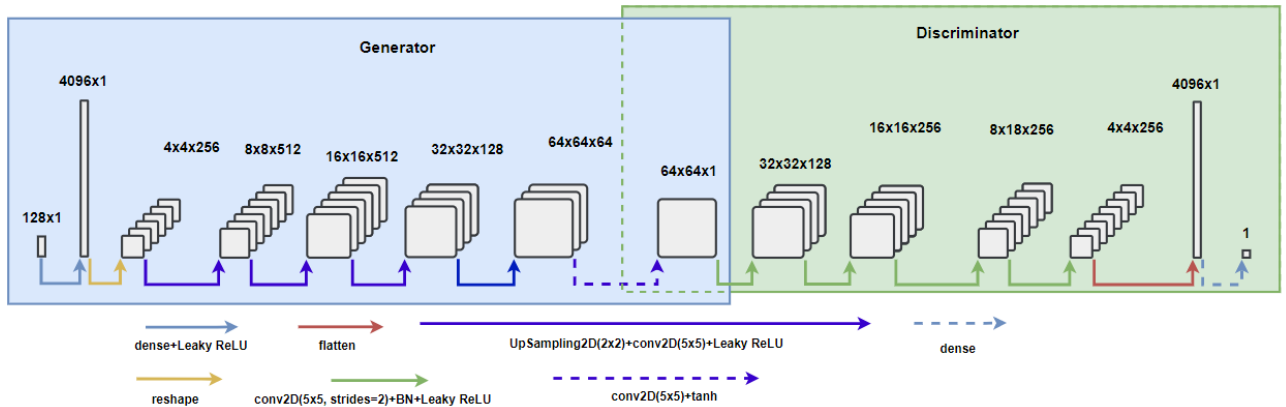


Рисунок 19 – Архитектура нейронной сети

Как видно из рисунка, по сравнению с предыдущей архитектурой было увеличено количество слоев, а слои транспонированной свертки были заменены на слои увеличения масштаба. Также на выходе дискриминатора отсутствует функция активации.

Обучение происходило на протяжении 140 эпох. Результаты генерации представлены на рисунке 20.

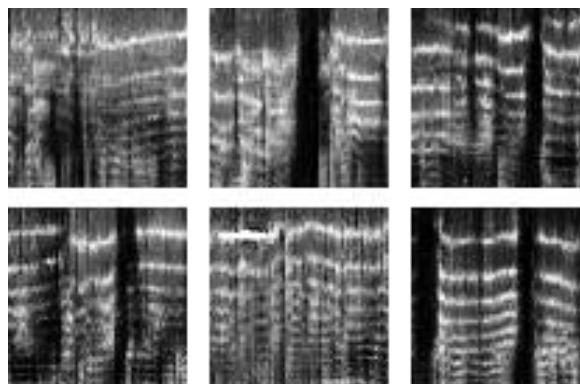


Рисунок 20 – Сгенерированные спектрограммы вокальной партии

Восстановленное из одной спектрограммы аудио, в виде мгновенных значений представлено на рисунке 21.

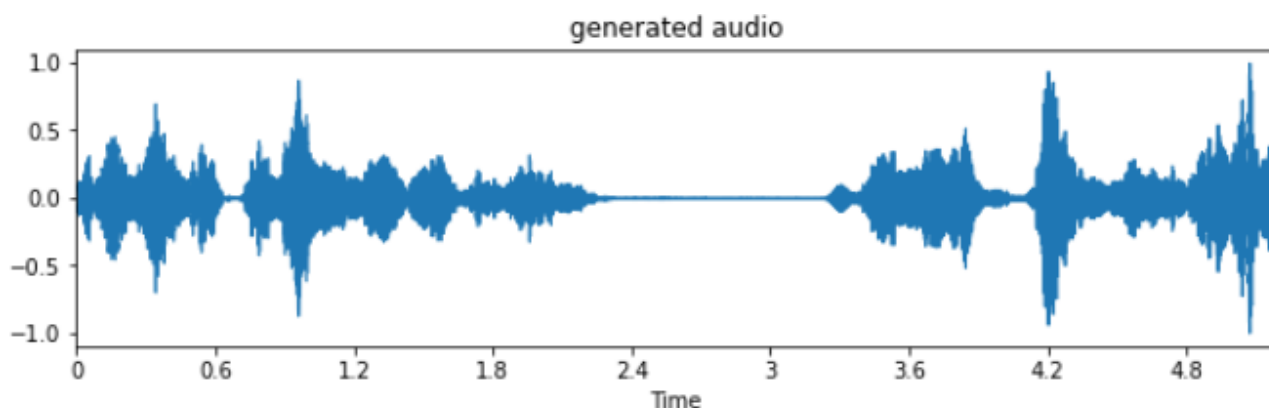


Рисунок 21 – Результат восстановления сгенерированного аудио

При прослушивании сгенерированного аудио невозможно разобрать слова, но достаточно хорошо можно отследить мелодию. Вероятно, это связано с тем, что при генерации не учитывается содержание произносимых слов. Также, в аудио присутствует шум, похожий на эхо.

Теперь имеет смысл обучить нейронную сеть генерировать одновременно все три вышеописанные музыкальные партии. При этом также, размер спектрограмм для обучения увеличен до 128x128, что соответствует примерно десяти секундам аудио и 128 Мел-частотных составляющих. Для этого, в сети генератора и дискриминатора добавляется по одному дополнительному слою, чтобы сгенерированные генератором спектрограммы и спектрограммы реальных аудио, подающиеся на вход дискриминатора, имели размерность 128x128x3. Обучение проводилось на протяжении 176 эпох и было остановлено, по причине отсутствия визуальных улучшений генерируемых спектрограмм на протяжении последних 26 эпох. Пример результатов генерации представлен на рисунке 22.

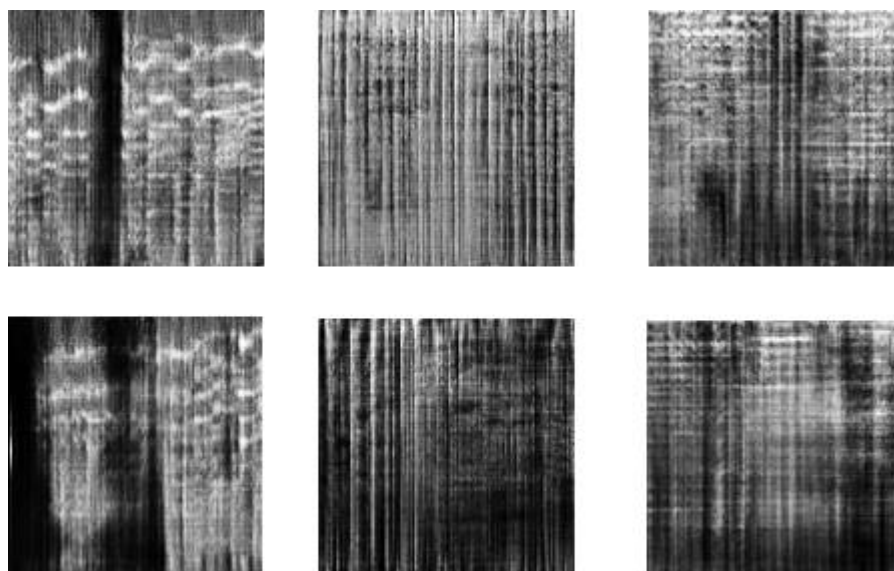


Рисунок 22 – Два примера сгенерированных музыкальных партий (вокал, ударные, остальное)

Как видно из рисунка выше, сгенерированные спектрограммы вокальной партии и партии ударных инструментов похожи на спектрограммы реальных аудио, чего нельзя сказать о спектрограмме остальных инструментов. Вероятно, причиной этого является слишком большое многообразие звучаний и инструментов, которыми представлены обучающие данные, поэтому выявление общей для них структуры является слишком сложной задачей для текущей архитектуры сети. Также стоит отметить, что качество сгенерированных спектрограмм вокальных партий понизилось, они стали содержать больше шума, в результате чего проследить мелодию стало сложнее.

Для улучшения качества получаемых результатов в дальнейшем следует провести ряд следующих экспериментов:

- Увеличить объем данных для обучения и повторить обучение с целью изучения влияния объема данных на качество генерируемых спектрограмм;
- Заменить спектрограммы, представляющие данные на спектрограммы мгновенных частот [7] изучить влияние их использования на качество генерируемого аудио в нашей задаче;

- Добавить в обучающие данные метки, тем самым предоставив дополнительную информацию о данных сети, что является одной из частых рекомендаций при обучении генеративных моделей;
- Изучить влияние размера сети на качество генерируемых данных
- Обучить сети вариационных автоэнкодеров или сетей класса U-net повышению качества генерируемых генеративно-состязательной сетью спектрограмм;
- Обучить сеть типа VAEGAN [36] – архитектуры, объединяющей в себе вариационный автоэнкодер и генеративно-состязательную сеть.

5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

Цель раздела – комплексное описание и анализ финансово-экономических аспектов выполненной работы. Необходимо оценить полные денежные затраты на исследование, а также дать хотя бы приближенную экономическую оценку результатов ее внедрения. Это в свою очередь позволит с помощью традиционных показателей эффективности инвестиций оценить экономическую целесообразность осуществления работы. Раздел должен быть завершен комплексной оценкой научно-технического уровня ВКР на основе экспертных данных.

5.1 Анализ конкурентных технических решений

В процессе поиска конкурентных технических решений был выявлен ряд аналогов:

- WaveGAN[8]
- MuseGAN[5]
- GANSynth[7]

В качестве факторов конкурентоспособности были выбраны: «Качество звучания», «Скорость генерации», «обучение на необработанных аудиоданных», «Сложность обучения» и «Разнообразие звучания», которым были присвоены соответствующие значения важности фактора путем осуществления экспертной оценки. Все разработки, взятые для сравнения, были оценены по выбранным факторам, на основании чего были рассчитаны итоговые оценки конкурентоспособности.

В таблице 1 представлена оценочная карта конкурентных технических решений. Анализируя данные таблицы можно сделать вывод о перспективности разработки.

Таблица 1 – Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Баллы				Конкурентоспособность			
		Б _р	Б _{WaveGAN}	Б _{MuseGAN}	Б _{GANSynth}				
1	2	3	4	5	6	7	8	9	10
Технические оценки ресурсоэффективности									
1. Качество звучания	0,3	3	4	2	5	0,9	1,2	0,6	1,5
2. Скорость генерации	0,05	3	4	3	5	0,15	0,2	0,15	0,25
3. Простота обучения	0,15	4	3	2	3	0,6	0,45	0,3	0,45
4. Разнообразие звучания	0,2	5	4	5	3	1	0,8	1	0,6
5. Обучение на неподготовленных данных	0,3	5	1	2	3	1,5	0,3	0,6	0,9
ИТОГО	1	20	16	14	19	4,15	2,95	2,65	3,7

На основании данных таблицы можно сделать вывод о конкурентоспособности разработки, наибольший вклад в которую дает возможность обучения на неподготовленных данных.

5.2 Структура работ в рамках научного исследования

В процессе выполнения работ по исследованию для конкретной задачи необходимо определить список проводимых работ, количество исполнителей, а также продолжительность данных работ. Так как число исполнителей не превышает двух, линейный график работ является наиболее удобным и компактным способом представления данных планирования. Список исполнителей включает в себя научного руководителя работ (НР) и непосредственного исполнителя (И). График выполнения работ с указанием перечня задач приведен в Таблице 2.

Таблица 2 – Перечень этапов работ и распределение исполнителей

Основные этапы	№ раб	Содержание работ	Должность исполнителя
Разработка технического задания	1	Составление и утверждение технического задания	Руководитель темы
Выбор направления исследований	2	Подбор и изучение материалов по теме	Инженер
	3	Выбор направления исследований	Руководитель, инженер
	4	Календарное планирование работ	Руководитель, инженер
	5	Проведение патентных исследований	Инженер
Теоретические и экспериментальные исследования	6	Разработка и реализация алгоритма подготовки данных	Инженер
	7	Анализ полученных данных	Инженер
	8	Эксперименты с различными архитектурами и типами нейронных сетей	Инженер
	9	Выбор базовой архитектуры нейронной сети	Руководитель, инженер
	10	Тонкая настройка параметров нейронной сети	Инженер
	11	Анализ результатов работы сети	Инженер
Теоретические и экспериментальные исследования	13	Технико-экономические расчеты	Инженер
	14	Вопросы безопасности и экологичности проекта	Инженер
Обобщение и оценка результатов	15	Анализ полученных результатов	Руководитель, инженер
	16	Согласование выполненной работы с научным руководителем	Руководитель, инженер
Проведение ОКР			
Оформление отчета по НИР	17	Составление пояснительной записки	Инженер

5.3 Определение трудоемкости выполнения работ

Для расчёта трудоемкости этапов работ рассмотрены два основных метода: технико-экономический и опытно-статический.

В связи с тем, что в работе задействована не развитая нормированная база трудоёмкости планируемых процессов, что в первую очередь обуславливается низкой степенью повторяемостью процессов, то в качестве основного метода расчёта трудоёмкости работ выбран опытно-статический метод. Расчёт по опытно-статическому методу реализуется на основе аналогового способа для определения вероятных значений трудоёмкости работ $t_{ож}$.

$$t_{ожi} = \frac{3*t_{mini} + 2*t_{maxi}}{5} \quad (5.1)$$

где $t_{ожi}$ – ожидаемая трудоёмкость выполнения i -ой работы чел.-дн.;

t_{mini} – минимальная продолжительность работы, дн;

t_{maxi} – максимальная продолжительность работы, дн.

Для выполнения перечисленных в таблице 1 работ требуются специалисты:

- инженер – в его роли действует исполнитель НИР (ВКР);
- научный руководитель.

Исходя из ожидаемой трудоёмкости работ, определяется продолжительность каждой работы в рабочих днях T_p , учитывающая параллельность выполнения работ несколькими исполнителями. Такое вычисление необходимо для обоснованного расчета заработной платы, так как удельный вес зарплаты в общей сметной стоимости научных исследований составляет около 65 %.

$$T_{pi} = \frac{t_{ожi}}{Ч_i} \quad (5.2)$$

где T_{pi} – продолжительность одной работы, раб. дн.;

$t_{ожi}$ – ожидаемая трудоёмкость выполнения одной работы, чел.-дн;

$Ч_i$ – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

5.4 Разработка графика проведения научного исследования

Диаграмма Ганта является наиболее удобным и наглядным способом представления графика проведения работ.

Диаграмма Ганта представляет собой отрезки, размещенные на горизонтальной шкале времени. Каждый отрезок соответствует отдельной задаче или подзадаче. Начало, конец и длина отрезка на шкале времени соответствуют началу, концу и длительности задачи.

Для построения графика Ганта, следует, длительность каждой из выполняемых работ из рабочих дней перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой, для каждого исполнителя расчеты производятся индивидуально:

$$T_{ki} = T_{pi} \cdot k_{\text{кал}}, \quad (5.3)$$

где T_{ki} – продолжительность выполнения i -й работы в календарных днях;

T_{pi} – продолжительность выполнения i -й работы в рабочих днях;

$k_{\text{кал}}$ – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}} = \frac{365}{365 - 66} = 1,22 \quad (5.4)$$

где $T_{\text{кал}}$ – количество календарных дней в году;

$T_{\text{вых}}$ – количество выходных дней в году;

$T_{\text{пр}}$ – количество праздничных дней в году.

$$\text{Для инженера } k_{\text{кал}} = \frac{365}{365 - 118} = 1,48 \quad (5.5)$$

$$\text{Для научного руководителя } \frac{365}{365 - 66} = 1,22 \quad (5.6)$$

Рассчитанные значения в календарных днях по каждой работе T_{ki} необходимо округлить до целого числа.

Все значения, полученные при расчетах по вышеприведенным формулам, были сведены в таблице 3.

Таблица 3 – Временные показатели проведенного исследования

Название работы	Трудоёмкость работ						Длительность работ в рабочих днях T_p		Длительность работ в календарных днях T_k	
	t_{\min} , чел-дни		t_{\max} , чел-дни		$t_{\text{ож}}$, чел-дни		Руководитель	Инженер	Руководитель	Инженер
	Руководитель	Инженер	Руководитель	Инженер	Руководитель	Инженер				
Составление и утверждение технического задания	1		1		1	0	1	0	1	0
Подбор и изучение материалов по теме		6		12	0	8,4	0	8,4	0	12
Выбор направления исследований	1	2	2	4	1	2,8	0,5	1,4	1	2
Календарное планирование работ	1	2	2	4	1,4	2,8	0,7	1,4	1	2
Проведение патентных исследований		2		3	0	2,4	0	2,4	0	4
Разработка и реализация алгоритма подготовки данных		5		10	0	7	0	7	0	10
Анализ полученных данных		5		10	0	7	0	7	0	10
Эксперименты с различными архитектурами и типами нейронных сетей	1	15	1	25	1	19	0,5	9,5	1	14
Выбор базовой архитектуры нейронной сети		1		3	0	1,8	0	1,8	0	3
Тонкая настройка параметров нейронной сети		15		25	0	19	0	19	0	28

Продолжение таблицы 3

Анализ результатов работы нейронной сети	1	3	2	5	1,4	3,8	0,7	1,9	1	3
Технико-экономические расчеты		5		7	0	5,8	0	5,8	0	9
Вопросы безопасности и экологичности проекта		5		7	0	5,8	0	5,8	0	9
Анализ полученных результатов	1	4	2	6	1	4,8	0,5	2,4	1	4
Согласование выполненной работы с научным руководителем	1	4	2	6	1	4,8	0,5	2,4	1	4
Составление пояснительной записки		1		3	0	1,8	0	1,8	0	3

На основе таблицы 3 строится календарный план-график. График строится для максимального по длительности исполнения работ в рамках научно-исследовательского проекта и представлен на рисунке 25 с разбивкой по месяцам и неделям за период времени дипломирования.

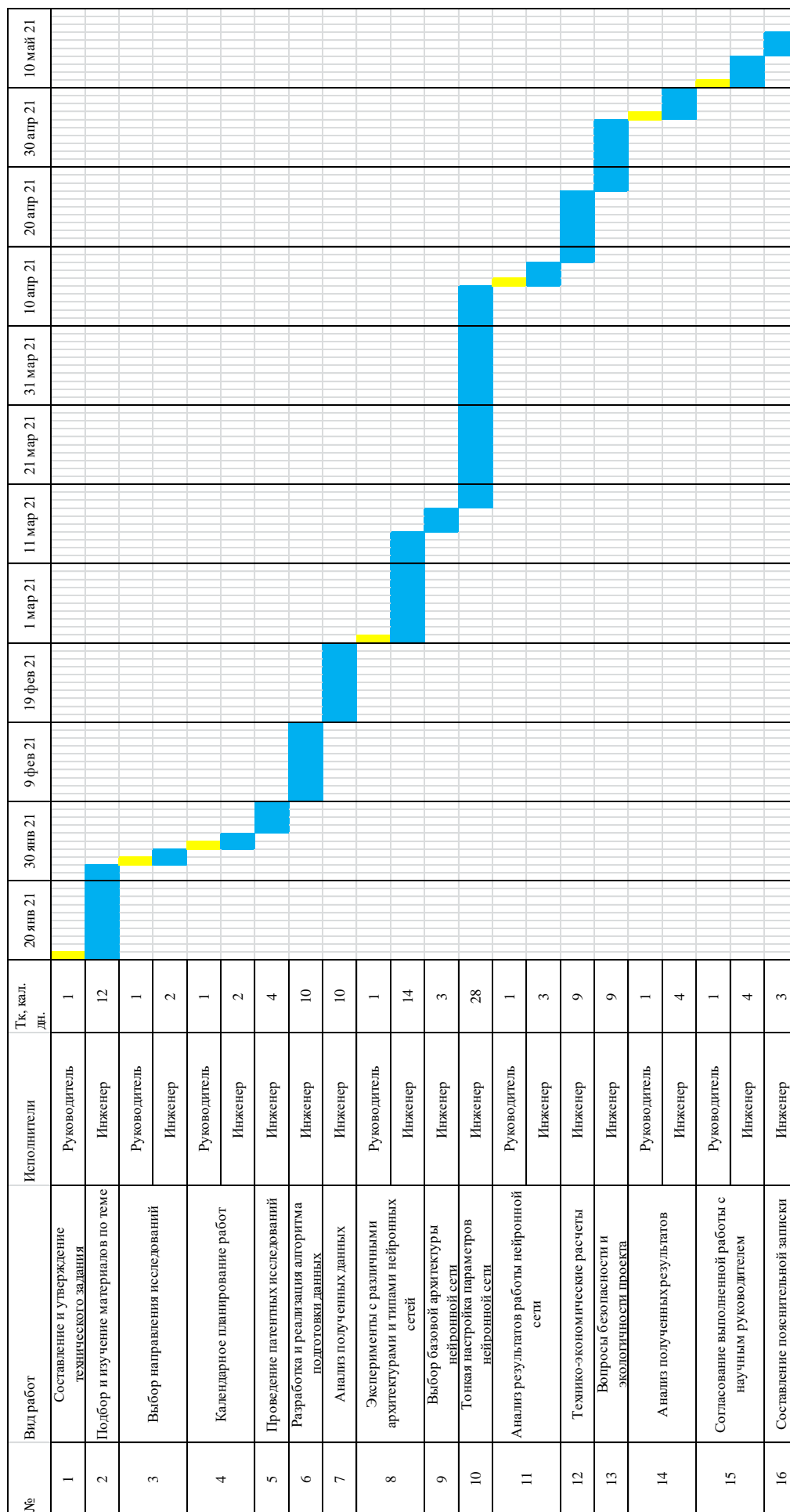


Рисунок 25 – Календарный план-график проекта

5.5 Бюджет проекта

Для полноты и достоверности учета всех расходов сгруппируем все затраты по следующим статьям:

- затраты на материалы;
- затраты на амортизацию;
- основная заработная плата исполнителей;
- дополнительная заработная плата исполнителей темы;
- отчисления во внебюджетные фонды (страховые отчисления);
- накладные расходы.

5.5.1 Материальные затраты

В расчет взяты только затраты на канцелярские товары в размере 1000 рублей.

5.5.2 Основная заработная плата исполнителей темы

Данная статья расходов включает заработную плату научного руководителя и инженера, в его роли выступает исполнитель проекта. Расчет основной заработной платы сводится в таблице 4.

Статья включает основную заработную плату работников, непосредственно занятых выполнением НИТ, и дополнительную заработную плату:

$$Z_{зп} = Z_{осн} + Z_{доп}, \quad (5.7)$$

где $Z_{осн}$ – основная заработная плата;

$Z_{доп}$ – дополнительная заработная плата (12-20 % от $Z_{осн}$).

Основная заработная плата руководителя (лаборанта, студента) от предприятия рассчитывается по следующей формуле:

$$Z_{осн} = Z_{дн} \cdot T_p, \quad (5.8)$$

где $Z_{осн}$ – основная заработная плата одного работника;

T_p – продолжительность работ, выполняемых научно-техническим работником, раб. дн. (таблица 2);

$Z_{дн}$ – среднедневная заработная плата работника, руб.

Среднедневная заработная плата рассчитывается по формуле:

$$Z_{\text{дн}} = \frac{Z_{\text{м}} \cdot M}{F_{\text{д}}}, \quad (5.9)$$

где $Z_{\text{м}}$ – месячный должностной оклад работника, руб.;

M – количество месяцев работы без отпуска в течение года:

при отпуске в 24 раб. дня $M = 11,2$ месяца, 5-дневная неделя;

при отпуске в 56 раб. дней $M = 10,4$ месяца, 6-дневная неделя;

$F_{\text{д}}$ – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн.

Таблица 4 – Баланс рабочего времени

Показатели рабочего времени	Научный руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней	66	118
Потери рабочего времени на отпуск	56	24
Действительный годовой фонд рабочего времени	243	223

Месячный должностной оклад работника:

$$Z_{\text{м}} = Z_{\text{окл}} \cdot k_{\text{р}}, \quad (5.10)$$

где $Z_{\text{окл}}$ – оклад, руб.;

$k_{\text{р}}$ – районный коэффициент, равный 1,3 (для Томска).

Научный руководитель имеет должность профессора и степень доктора технических наук оклад составлял 47104 руб., затем был проиндексирован на 4,3% и составил 49129,5 руб.

Оклад инженера составил 21760 руб., затем осенью был проиндексирован на 4,3% и составил 22695,68 руб.

Таблица 5 – Расчёт основной заработной платы

Исполнители	Разряд	$k_{\text{т}}$	$Z_{\text{окл}}$, руб.	$k_{\text{р}}$	$Z_{\text{м}}$, руб.	$Z_{\text{дн}}$, руб.	$T_{\text{р}}$, раб. дн.	$Z_{\text{осн}}$, руб.
Научный руководитель	–	–	49129,5	1,3	63868,4	2733,46	5	13667,3
Инженер	–	–	22695,68		29504,5	1481,84	78	115583,52
Итого $Z_{\text{осн}}$								129250,82

5.5.3 Дополнительная заработная плата исполнителей темы

Затраты по дополнительной заработной плате исполнителей темы учитывают величину предусмотренных Трудовым кодексом РФ доплат за отклонение от нормальных условий труда, а также выплат, связанных с обеспечением гарантий и компенсаций (при исполнении государственных и общественных обязанностей, при совмещении работы с обучением, при предоставлении ежегодного оплачиваемого отпуска и т.д.). Расчет дополнительной заработной платы ведется по следующей формуле:

$$Z_{\text{доп}} = k_{\text{доп}} \cdot Z_{\text{осн}}, \quad (5.11)$$

где $k_{\text{доп}}$ – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,12 – 0,15).

Дополнительная заработная плата представлена в таблице 6.

Таблица 6 – Расчёт дополнительной заработной платы

Исполнитель	$k_{\text{доп}}$	$Z_{\text{осн}}$	$Z_{\text{доп}}$
Научный руководитель	0,12	13667,3	1640,08
Инженер		115583,52	13870,02
Итого			15510,1

5.5.4 Отчисления во внебюджетные фонды (страховые отчисления)

В данной статье расходов отражаются обязательные отчисления по установленным законодательством Российской Федерации нормам органам государственного социального страхования (ФСС), пенсионного фонда (ПФ) и медицинского страхования (ФФОМС) от затрат на оплату труда работников. Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$Z_{\text{внеб}} = k_{\text{внеб}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}), \quad (5.12)$$

где $k_{\text{внеб}}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

Для учреждений, осуществляющих образовательную и научную деятельность в 2014 году водится пониженная ставка – 27,1%.

Таблица 7 – Отчисления во внебюджетные фонды

Исполнитель	Руководитель	Инженер
Основная заработная плата, руб.	13667,3	115583,52
Дополнительная заработная плата, руб.	1640,08	13870,02
Коэффициент отчислений во внебюджетные фонды	0,271	
Сумма отчислений	4148,3	35081,64
Итого	39229,94	

5.5.5 Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов исследования, оплата услуг связи, электроэнергии, почтовые и телеграфные расходы, размножение материалов и т.д. Их величина определяется по следующей формуле:

$$Z_{\text{накл}} = k_{\text{накл}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}) \quad (5.12)$$

где $k_{\text{накл}}$ – коэффициент, учитывающий накладные расходы.

Величину коэффициента накладных расходов можно взять в размере 16%.

$$Z_{\text{накл}} = 0,16 \cdot 144760,92 = 23161,75 \text{ рубля} \quad (5.13)$$

На основании полученных данных по отдельным статьям затрат составляется калькуляция плановой себестоимости НИ по форме, приведенной в таблице 8.

Таблица 8 – Статьи затрат

Наименование статьи	Сумма, руб.
1. Материальные затраты НТИ	1000
2. Затраты по основной заработной плате исполнителей темы	129250,82
3. Затраты по дополнительной заработной плате исполнителей темы	15510,1
4. Отчисления во внебюджетные фонды	39229,94
5. Накладные расходы	23161,75
6. Бюджет затрат НТИ	208152,61

Таким образом, затраты на разработку составили $C = 208152,61$ руб.

5.5.6 Расчет прибыли

Прибыль от реализации проекта в зависимости от конкретной ситуации (масштаб и характер получаемого результата, степень его определенности и коммерциализации, специфика целевого сегмента рынка и т.д.) может определяться различными способами.

Прибыль составляет 41630,52 руб. (20 %) от расходов на разработку проекта.

5.5.7 Расчет НДС

НДС составляет 20% от суммы затрат на разработку и прибыли. В нашем случае это $(208152,61 + 41630,52) * 0,2 = 49956,63$ руб.

5.5.8 Цена разработки НИР

Цена равна сумме полной себестоимости, прибыли и НДС, в нашем случае $Ц_{НИР}(КР) = 208152,61 + 41630,52 + 49956,63 = 299739,76$ руб.

5.6 Оценка экономической эффективности проекта

Выполнение научно-исследовательских работ оценивается уровнями достижения экономического, научного, научно-технического и социального эффектов. Научный эффект характеризует получение новых научных знаний и отображает прирост информации, предназначенной для внутри научного потребления.

Научно-технический эффект характеризует возможность использования результатов в других проектах и обеспечивает получение информации, необходимой для создания новой техники.

Экономический эффект характеризуется, в частности, выраженной в стоимостных показателях экономией живого общественного труда, а также в возможности применения полученных знаний для создания новых разработок. Социальный эффект проявляется в улучшении условий труда.

Экономическая эффективность проекта обусловлена возрастающей необходимостью в системах распознавания дорожных знаков и в целом автопилотировании, что является востребованным и актуальным на сегодняшний день.

На данном этапе разработки нет возможности количественно оценить экономический эффект. Данная разработка предназначена для использования во всех сферах, связанных с распознаванием и детектированием дорожных

знаков, и для оценки экономического эффекта требуется более глубокое системное исследование.

Для решения этой задачи в дальнейшем необходимо рассчитывать данный показатель, исходя из конкретных условий и масштабы использования, в зависимости от выбранной сферы применения.

5.7 Определение ресурсной эффективности исследования

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования можно определить следующим образом:

$$I_{pi} = \sum a_i b_i, \quad (5.14)$$

где I_{pi} – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки;

a_i – весовой коэффициент i -го варианта исполнения разработки;

b_i – балльная оценка i -го варианта исполнения разработки, устанавливается экспертным путем по выбранной шкале оценивания;

n – число параметров сравнения.

Результат расчета интегрального показателя ресурсоэффективности представлен в таблице 9.

Таблица 9 – Сравнительная оценка характеристик вариантов исполнения проекта

Критерии оценки	Весовой коэффициент параметра	Баллы			
		Б _Р	Б _{WaveGAN}	Б _{MuseGAN}	Б _{GANSynth}
1. Качество звучания	0,3	3	4	2	5
2. Скорость генерации	0,05	3	4	3	5
3. Простота обучения	0,15	4	3	2	3
4. Разнообразие звучания	0,2	5	4	5	3
5. Обучение на неподготовленных данных	0,3	5	1	2	3
ИТОГО	1	20	16	14	19
Значение интегрального показателя	-	4,15	2,95	2,65	3,7

$$I_{p-разработки} = 3 * 0,3 + 3 * 0,05 + 4 * 0,15 + 5 * 0,2 + 6 * 0,3 = 4,15 \quad (5.15)$$

$$I_{p-WaveGAN} = 4 * 0,3 + 4 * 0,05 + 3 * 0,15 + 4 * 0,2 + 1 * 0,3 = 2,95 \quad (5.16)$$

$$I_{p-MuseGAN} = 2 * 0,3 + 3 * 0,05 + 2 * 0,15 + 5 * 0,2 + 2 * 0,3 = 2,65 \quad (5.17)$$

$$I_{p-GANSynth} = 5 * 0,3 + 5 * 0,05 + 3 * 0,15 + 3 * 0,2 + 3 * 0,3 = 3,7 \quad (5.18)$$

Из значений интегрального показателя ресурсоэффективности разработки и ее существующих аналогов можно сделать вывод о ее высокой конкурентоспособности.

5.8 Заключение по разделу

Общая длительность исследования составляет 117 календарных дней (период с 20.01.2019 по 17.05.2019), при этом цена разработки составила 299739,76 рублей. Сравнение значений интегральных показателей эффективности вариантов исполнения разработки показало, что более эффективным вариантом решения технической задачи, поставленной в работе с позиции ресурсной эффективности, является собственная разработка.

Результаты исследования могут быть использованы в как медиа сфере, позволяя создавать новые музыкальные произведения, так и в качестве базы для развития в самостоятельное музыкальное произведение с помощью привлечения профессиональных музыкантов.

6. Социальная ответственность

Целью выпускной квалификационной работы является исследование возможностей генеративно-состязательных сетей в задаче генерации сложных музыкальных произведений, а также методов, повышающих качество звука.

В данном разделе ВКР исследованы вредные и опасные факторы среды, рассмотрены меры по защите работника от возможного негативного воздействия среды, а также возможные чрезвычайные ситуации и необходимые действия, которые разработчик должен выполнить в случае возникновения чрезвычайной ситуации.

Результаты исследований могут быть применены в задачах автоматической генерации музыкальных произведений, используемых как в качестве сопровождения к видео, так и как основа для музыкального творчества.

Были выделены и рассмотрены следующие факторы, воздействующие на разработчика программы:

- недостаточная освещенность рабочей зоны;
- отклонение параметров микроклимата;
- превышения уровня шума;
- нервно-психические перегрузки.

К опасным факторам при работе с персональным компьютером относятся возможность короткого замыкания, влекущего за собой опасность поражения разработчика электрическим током и статическое электричество. Рассмотрены вопросы правового регулирования трудовых отношений, связанных с использованием разработанной системы.

6.1 Правовые и организационные вопросы обеспечения безопасности

Работа в офисе относится ко второй категории тяжести труда – работы выполняются при оптимальных условиях внешней производственной среды и при оптимальной величине физической, умственной и нервно-эмоциональной нагрузки. Согласно трудовому кодексу РФ продолжительность рабочего дня работников не должна превышать 40 часов в неделю. Возможно, сокращение рабочего времени. Для работников, возраст которых меньше 16 лет – не более 24 часа в неделю, от 16 до 18 лет – не более 35 часов, как и для инвалидов I и II группы [37]. Во время регламентированных перерывов целесообразно выполнять комплексы упражнений и осуществлять проветривание помещения.

Конструкция рабочего места и взаимное расположение всех его элементов должны соответствовать антропометрическим, физиологическим и психологическим требованиям, а также характеру работы. Используемые средства отображения информации, требующие точного и быстрого считывания показаний, следует располагать в вертикальной плоскости под углом $\pm 15^\circ$ от нормальной линии взгляда и в горизонтальной плоскости под углом $\pm 15^\circ$ от сагиттальной плоскости. Клавиатура должна располагаться на поверхности стола на расстоянии 100-300 мм от края, обращенного к пользователю (рисунок 23) [38].

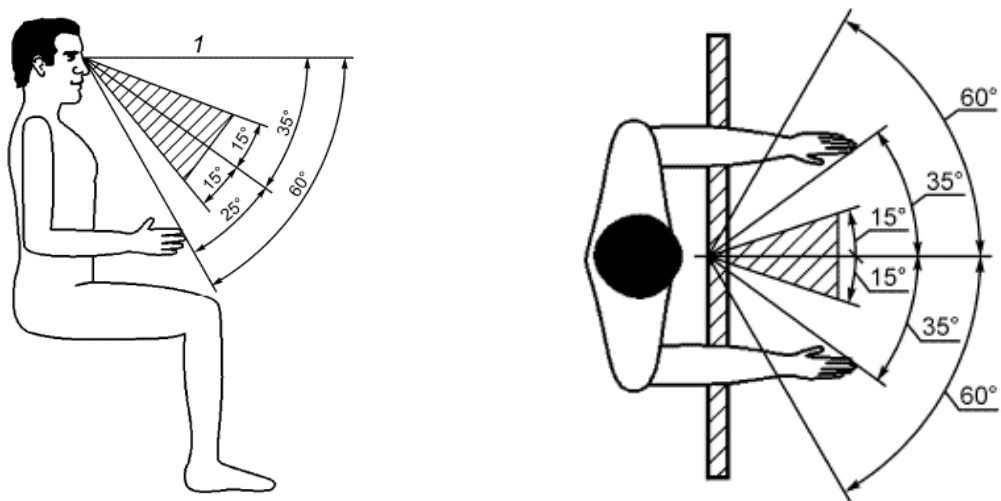


Рисунок 23 – Зоны зрительного наблюдения в горизонтальной плоскости и в вертикальной плоскости [39].

В соответствии с ГОСТ 12.2.032-78. «Система стандартов безопасности труда». Рабочее место, при выполнении работ сидя» рабочий стол может быть любой конструкции, отвечающей современным требованиям эргономики и позволяющей удобно разместить на рабочей поверхности оборудование с учетом его количества, размеров и характера выполняемой работы [38].

Выполнение требований на данном рабочем месте отражено ниже таблице 10, согласно СанПиН 1.2.3685-21 [39] и ГОСТ 12.2.032-78.

Таблица 10 – Требования к организации рабочего места при работе с ПЭВМ

Требование	Требуемое значение	Значение параметра помещения
Высота рабочей поверхности стола	Регулируемая высота (680-800мм) Нерегулируемая высота (725мм)	Нерегулируемая высота (740 мм)
Рабочий стул	Подъемно-поворотный, регулируемый по высоте и углу наклона спинки	Подъемно-поворотный, регулируемый по высоте
Расположение монитора от глаз пользователя	600-700 мм	Соответствует

6.2 Производственная безопасность

Вредные и опасные факторы, воздействующие на сотрудника, устанавливаются согласно ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация» [40]. Могут иметь место следующие факторы, представленные в таблице 11.

Таблица 11 – Возможные опасные и вредные факторы

Факторы	Разработка	Нормативные документы
Отклонение показателей микроклимата	+	СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений
Недостаточная освещенность рабочей зоны	+	СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95
Психофизиологические факторы	+	Трудовой кодекс Российской Федерации от 30.12.2001 №197-ФЗ (ред. от 27.12.2018)
Зрительное напряжение	+	СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания.
Превышение уровня шума	+	СН 2.2.4/2.1.8.562-96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки
Поражение электрическим током	+	ГОСТ 12.1.038-82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов
Статическое электричество	+	ГОСТ 12.1.006-84 ССБТ. «Электромагнитные поля радиочастот. Допустимые уровни на рабочих местах и требования к проведению контроля»
Короткое замыкание	+	ГОСТ 12.1.030-81 Система стандартов безопасности труда (ССБТ). Электробезопасность. Защитное заземление. Зануление. ГОСТ 26522-85 Короткие замыкания в электроустановках.

6.2.1 Отклонение показателей микроклимата

В производственных помещениях, в которых работа с использованием ПЭВМ является основной (диспетчерские, операторские, расчетные, кабины и посты управления, залы вычислительной техники и др.) и связана с нервно-эмоциональным напряжением, должны обеспечиваться оптимальные параметры микроклимата для категории работ 1а и 1б в соответствии с действующими санитарно-эпидемиологическими нормативами микроклимата производственных помещений.

Содержание вредных химических веществ в производственных помещениях, в которых работа с использованием ПЭВМ является основной (диспетчерские, операторские, расчетные, кабины и посты управления, залы вычислительной техники и др.), не должно превышать предельно допустимых концентраций загрязняющих веществ в атмосферном воздухе населенных мест в соответствии с действующими гигиеническими нормативами.

Также указывается, что в помещениях с ПЭВМ должна ежедневно проводиться влажная уборка.

Уставленные гигиенические нормативы для помещений с ВДТ и ПЭВМ для категории работы 1б приведены в таблице 12:

Таблица 12 – Оптимальные и допустимые величины показателей микроклимата на рабочих местах производственных помещений (СанПиН 2.2.4.548-96) [41]

Оптимальные значения характеристик микроклимата				
Период года	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	22 – 24	21 – 25	40 – 60	0,1
Теплый	23 – 25	22 – 26	40 – 60	0,1
Допустимые значения характеристик микроклимата				
Период года	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	20 – 25	19 – 26	15 – 75	0,1
Теплый	21 – 28	20 – 29	15 – 75	0,1 – 0,2

6.2.2 Недостаточная освещенность рабочей зоны

Недостаточная освещенность рабочей зоны помещения, оборудованной ПК, также является одной из причин нарушения зрительной функции, а также влияет на общее самочувствие и эффективность труда.

Искусственное освещение в помещениях для эксплуатации ПК должно осуществляться системой общего равномерного освещения. В случаях преимущественной работы с документами, следует применять системы комбинированного освещения (к общему освещению дополнительно устанавливаются светильники местного освещения, предназначенные для освещения зоны расположения документов). Окна в помещениях, где эксплуатируется вычислительная техника, преимущественно должны быть

ориентированы на север и северо-восток. Нормируемые показатели естественного, искусственного и совмещенного освещения в соответствии с СП 52.13330.2016 указаны в таблице 13 [42].

Таблица 13 – Нормируемые показатели естественного, искусственного и совмещенного освещения

Помещения	Рабочая поверхность и плоскость нормирования КЕО и освещенности (Г-горизонтальная, В-вертикальная) и высота плоскости над полом, м	Естественное освещение		Совмещенное освещение	
		КЕО e_n , %		КЕО e_n , %	
		При верхнем или комбинированном освещении	При боковом освещении	При верхнем или комбинированном освещении	При боковом освещении
Кабинеты, рабочие комнаты, офисы, представительства	$\Gamma - 0,8$	3,0	1,0	1,8	0,6
Помещения	Искусственное освещение				
	Освещенность, лк				
	При комбинированном освещении		При общем освещении	Показатель дискомфорта, М, не более	Коэффициент пульсации освещенности, Кп, %, не более
	Всего	От общего			
Кабинеты, рабочие комнаты, офисы, представительства	400	200	300	40	15

Расчет уровня освещенности рабочего места в помещении:

Длина помещения: $a=6$ м, ширина: $b=3$ м, высота: $H=2,4$ м. Высота рабочей поверхности $h_{рп} = 0,74$ м. Требуемая освещенность $E = 300$ лк.

Коэффициент отражения стен $R_c = 30\%$, потолка $R_n = 50\%$. Коэффициент запаса $k = 1,5$, коэффициент неравномерности $Z = 1,1$. Рассчитываем систему общего освещения лампой накаливания.

Светильник типа ОД, $\lambda = 1,4$

Приняв $h_c = 0,5$ м, определяем расчетную высоту.

$$h = H - h_c - h_{рп} = 2,4 - 0,5 - 0,74 = 1,16 \text{ м.} \quad (6.1)$$

Находим индекс помещения:

$$i = \frac{s}{(a+b)h} = \frac{18}{9 \cdot 1,16} = 1,72 \quad (6.2)$$

Определяем коэффициент использования светового потока:

$$\eta = 0,55 \quad (6.3)$$

Расстояние между светильниками L определяется как:

$$L = \lambda \cdot h = 1,4 \cdot 1,16 = 1,624 \text{ м} \quad (6.4)$$

Количество рядов светильников с люминесцентными лампами определяется по формуле:

$$n_{\text{ряд}} = \frac{B - \frac{2}{3}L}{L} + 1 = \frac{3 - \frac{2}{3} \cdot 1,624}{3} + 1 = 1,64 \approx 2 \quad (6.5)$$

Однако ширина помещения, с учетом расстояния между светильниками не позволяет разметить более одного ряда, поэтому $n_{\text{ряд}} = 1$.

Количество светильников с люминесцентными лампами определяется по формуле:

$$n_{\text{св}} = \frac{A - \frac{2}{3}L}{l_{\text{св}} + 0,5} = \frac{5 - \frac{2}{3} \cdot 1,624}{1,23 + 0,5} = 2,26 \approx 2 \quad (6.6)$$

Размещаем светильники в один ряд. В ряду можно установить 2 светильника типа ОД мощностью 40 Вт. Учитывая, что в каждом светильнике установлено две лампы, общее число ламп в помещении $N = 4$.

Определяем потребный световой поток лампы:

$$\Phi = \frac{E \cdot S \cdot k \cdot Z}{N \cdot \eta} = \frac{300 \cdot 18 \cdot 1,5 \cdot 1,1}{4 \cdot 0,55} = 5400,75 \text{ лм} \quad (6.7)$$

Выбираем ближайшую стандартную лампу: ЛТБ 80 Вт с потоком 5200 лм. Делаем проверку выполнения условий:

$$-10\% \leq \frac{\Phi_{\text{ст}} - \Phi_{\text{расч}}}{\Phi_{\text{ст}}} \cdot 100\% \leq +20\% \quad (6.8)$$

Получаем: $-10\% \leq -3,7\% \leq +20\%$

Определяем электрическую мощность осветительной установки:

$$P = 4 \cdot 80 \text{ Вт} = 320 \text{ Вт} \quad (6.9)$$

6.2.3 Зрительное напряжение

Работа на ПК сопровождается постоянным и значительным напряжением функций зрительного анализатора. Одной из основных

особенностей является иной принцип чтения информации, чем при обычном чтении. Чтобы снизить зрительное напряжение нужно соблюдать визуальные параметры экрана (таблица 14).

Таблица 14 – Допустимые визуальные параметры устройств отображения информации [43]

Параметры	Допустимые значения
Яркость белого поля	Не менее 35 кд/кв.м
Неравномерность яркости рабочего поля	Не более $\pm 20\%$
Контрастность (для монохромного режима)	3:1
Пространственная нестабильность изображения (непреднамеренное изменение положения фрагментов изображения экрана)	Не более $2 \cdot 10L - 4L$, где L – расстояние наблюдения

6.2.4 Психофизические факторы

Работа с ПК сопряжена с воздействием вредных психофизиологических факторов, в частности, нервно-психических перегрузок. Нервно-психические перегрузки – совокупность таких сдвигов в психофизиологическом состоянии организма человека, которые развиваются после совершения работы и приводят к временному снижению эффективности труда. Состояние утомления (усталость) характеризуется определенными объективными показателями и субъективными ощущениями.

Нервно-психические перегрузки подразделяются на следующие:

- умственное перенапряжение;
- перенапряжение анализаторов;
- монотонность труда;
- эмоциональные перегрузки.

При первых симптомах психического перенапряжения необходимо:

- дать нервной системе расслабиться;
- рационально чередовать периоды отдыха и работы;
- начать заниматься спортом;
- ложиться спать в одно и то же время;
- в тяжелых случаях обратиться к врачу.

Естественно, что полностью исключить провоцирующие факторы из жизни вряд ли удастся, но можно уменьшить их негативное воздействие, давая нервной системе необходимый отдых. Суммарное время регламентированных перерывов при работе с ПК зависит от категории трудовой деятельности и уровня нагрузки за рабочую смену. В таблице 15 приведено суммарное время отдыха для каждой категории работ.

Таблица 15. Суммарное время перерывов в зависимости от категории работы и нагрузки

Категория работы с ПЭВМ	Уровень нагрузки за рабочую смену при видах работ с ПЭВМ			Суммарное время регламентированных перерывов при 8-часовой смене, мин
	Группа А, количество знаков	Группа Б, количество знаков	Группа В, часов	
I	До 20000	До 15000	До 2	50
II	До 40000	До 30000	До 4	70
III	До 60000	До 40000	До 6	90

При проведении исследований уровень нагрузки относился к группе В, категория работы III. Согласно таблице, суммарное время перерывов необходимо установить не менее 90 минут. По типовой инструкции по охране труда при работе на персональном компьютере ТОО Р-45-084-01 для данной категории работ требуется установить перерывы по 15 минут каждый трудовой час. [43]

6.2.5 Превышение уровня шума

В производственных помещениях при выполнении основных или вспомогательных работ с использованием ПЭВМ уровни шума на рабочих местах не должны превышать предельно допустимых значений, установленных для данных видов работ в соответствии с действующими санитарно-эпидемиологическими нормативами СН 2.2.4/2.1.8.562-96 [44] установлены допустимые значения уровней звукового давления, создаваемого ПЭВМ (таблица 16).

Таблица 16 – Допустимые значения уровней звукового давления в октавных полосах частот и уровня звука, создаваемого ПЭВМ

Уровни звукового давления, дБ в октавных полосах со среднегеометрическими частотами, Гц.									Уровни звука в дБА
31,5 Гц	63 Гц	125 Гц	250 Гц	500 Гц	1000 Гц	2000 Гц	4000 Гц	8000 Гц	
86	71	61	54	49	45	42	40	38	50

6.2.6 Повышенное значение напряжения в электрической цепи, замыкание которой может произойти через тело человека

При разработке широко используется электричество для питания компьютерной техники, которая может являться источником опасности. Несоблюдение правил ГОСТ 12.1.038-82 ССБТ. «Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов» может привести к опасным последствиям [45]. Поражение электрическим током может произойти при прикосновении к токоведущим частям, находящимся под напряжением, на которых остался заряд или появилось напряжение.

Электрический ток оказывает на человека термическое, электролитическое, биологическое и механическое воздействие. Действие электрического тока на человека приводит к травмам или гибели людей. Для переменного тока частотой 50 Гц допустимое значение напряжения прикосновения составляет 2 В, а силы тока – 0,3 мА, для тока частотой 400 Гц, соответственно – 2 В и 0,4 мА, для постоянного тока – 8 В и 1 мА.

Мерами защиты от воздействия электрического тока являются оградительные устройства, устройства автоматического контроля и сигнализации, изолирующие устройства и покрытия, устройства защитного заземления, устройства автоматического отключения, предохранительные устройства.

6.2.7 Статическое электричество

В помещениях, оборудованных ПЭВМ, токи статического электричества чаще всего возникают при прикосновении персонала к любому из элементов

ПЭВМ. Такие разряды опасности для человека не представляют, однако кроме неприятных ощущений могут привести к выходу оборудования из строя.

Для предотвращения образования и защиты от статического электричества в помещении используются нейтрализаторы и увлажнители, а полы имеют антистатическое покрытие в виде поливинилхлоридного антистатического линолеума.

В ГОСТ 12.1.006-84 ССБТ. «Электромагнитные поля радиочастот. Допустимые уровни на рабочих местах и требования к проведению контроля» [46] установлен максимальный допустимый электростатический потенциал экрана видеомонитора – 500 В.

В качестве мер уменьшения влияния вредных факторов на пользователя используются защитные фильтры для мониторов, увлажнители воздуха. Должны использоваться розетки с заземлением. Требуется проводить регулярную влажную уборку.

6.3 Экологическая безопасность

На данном рабочем месте выявлен предполагаемый источник загрязнения окружающей среды, а именно воздействие на литосферу в результате образования отходов при поломке предметов вычислительной техники и оргтехники.

В ходе разработке возможно повышение температуры ПК, при этом он является источником электромагнитного и ионизирующего излучения, а также шума. Для защиты здоровья сотрудников, работающих с программным обеспечением, рекомендуется соблюдать необходимую дистанцию при работе с компьютером (от 0.5 до 1 м), а также использовать эффективные охлаждающие системы и наиболее современное и эффективное оборудование.

Федеральный закон № 89 от 1998г. «Об отходах производства и потребления» [47] запрещает юридическим лицам самовольно избавляться от опасных отходов. Этим видом деятельности, согласно постановлению Правительства РФ № 340 от 2002 г., могут заниматься только

специализированные структуры. В их число входят и фирмы, которые занимаются утилизацией электронных отходов.

Обращение с отходами регламентируется ГОСТ Р 53692-2009 «Ресурсосбережение. Обращение с отходами».[48]

Поэтому, при необходимости утилизировать вышедшую из употребления электронику наиболее безопасным для окружающей среды способом необходимо обращаться в специализированную компанию по утилизации. Такие компании действуют на всей территории Российской Федерации, в том числе и в Томской области.

6.4 Безопасность в чрезвычайных ситуациях

Наиболее вероятной чрезвычайной ситуацией при разработке алгоритма является пожар на рабочем месте. В качестве противопожарных мероприятий должны быть применены следующие меры:

1. В помещении должны находиться средства тушения пожара, средства связи;
2. Электрическая проводка электрооборудования и осветительных приборов должна быть исправна;
3. Все сотрудники должны знать место нахождения средств пожаротушения и уметь ими воспользоваться, средств связи и номера экстренных служб.

Согласно ГОСТ 12.1.004-91 «Пожарная безопасность. Общие требования», при работе с компьютером необходимо соблюдать следующие нормы пожарной безопасности [49]:

- для предохранения сети от перегрузок запрещается одновременно подключать к сети количество потребителей, превышающих допустимую нагрузку;
- работы за компьютером проводить только при исправном состоянии оборудования, электропроводки;
- иметь средства для тушения пожара (огнетушитель);
- установить количество, размеры и соответствующее конструктивное исполнение эвакуационных путей и выходов;

– обеспечить возможность беспрепятственного движения людей по эвакуационным путям.

Связи с возможностью возникновения пожара разработан следующий план действий:

1. В случае возникновения пожара сообщить о нем руководителю, постараться устранить очаг возгорания имеющимися силами при помощи первичных средств пожаротушения (огнетушитель порошковый, углекислотный О-1ПО (з)-АВСЕ);
2. Привести в действие ручной пожарный извещатель, если очаг возгорания потушить не удастся;
3. Сообщить о возгорании в службу пожарной охраны по телефону 01 или 101, сообщить адрес, место и причину возникновения пожара;
4. Принять меры по эвакуации людей и материальных ценностей;
5. Встретить пожарную охрану, при необходимости сообщить всю необходимую информацию и оказать помощь при выборе наилучшего подхода к очагу возгорания.
6. Рабочее помещение оборудовано в соответствии с требованиями пожарной безопасности. Имеется порошковый огнетушитель, а также пожарная сигнализация и средства связи.

В случае возникновения пожара в здании необходимо немедленно эвакуироваться из здания и направиться на выход в соответствии с планом эвакуации при пожарах и других ЧС (рисунок 24).



Рисунок 24 – План эвакуации при пожарах и других ЧС

6.5 Заключение по разделу

В ходе выполнения работы над разделом «Социальная ответственность» были выявлены опасные и вредные факторы, воздействию которых может подвергнуться человек, проводящий исследование возможностей генеративно-состязательных сетей в задаче генерации сложных музыкальных произведений, а также проведен анализ нормативной документации.

Установлено, что рабочее место удовлетворяет требованиям безопасности. Выполняемая работа не сопряжена с высоким риском травматизма. Освещение на рабочем месте соответствует нормам. Уровни шума находятся в допустимых пределах – источником шума при эксплуатации ПК являются системы охлаждения. Микроклиматические условия соблюдаются за счет использования систем отопления и кондиционирования. В виду того, что корпус ПК является токонепроводящим, необходимость в мерах против статического электричества отсутствует. Во время работы осуществляются перерывы для снижения нагрузки и предотвращения нервно-психических перегрузок. Помещение оборудовано согласно требованиям электробезопасности. Рабочее помещение оборудовано в соответствии с требованиями пожарной безопасности.

Заключение

В ходе проделанной работы были проведены эксперименты с генеративно-состязательными сетями в задаче генерации сложных музыкальных произведений. Для этого был разработан ряд скриптов, позволяющих автоматизировать процесс получения и подготовки данных. Проведены эксперименты с автоэнкодерами и вариационными автоэнкодерами, которые показали, что автоэнкодеры достаточно хорошо могут реконструировать спектрограмм вокальной партии, однако даже с применением вариационных автоэнкодеров не удается создать качественное непрерывное распределение, из которого можно было бы генерировать новые композиции или использовать в качестве скрытого представления. Вокальная партия оказалась слишком сложной для обычной архитектуры генеративно-состязательной сети, которая не смогла сойтись и обучение пришлось остановить. В противовес этому генеративно-состязательная сеть с использованием функции потерь Вассерштейна и штрафом градиента смогла справиться с генерацией вокальной и инструментальных партий, пусть и добавив к ним заметное количество шума. Слишком большое разнообразие партий остальных инструментов не позволило сети изучить их распределение достаточно хорошо, чтобы она могла сгенерировать хотя бы одну партию, похожую на реальную.

В дальнейшем предполагается провести ряд экспериментов, которые приводятся в конце раздела 4. Предполагается, что их комбинация может значительно повысить генеративные возможности сети и позволить сделать генерацию управляемой и более качественной.

СПИСОК ИСТОЧНИКОВ

1. Generative Adversarial Networks // arxiv.org URL: <https://arxiv.org/abs/1406.2661> [Электронный ресурс] (дата обращения: 10.02.21).
2. Шолле Ф. Глубокое обучение на python. - СПб.: Питер, 2018. - 400 с.
3. Andriy Burkov The hundred-page machine learning book. - Andriy Burkov, 2019. - 141 с.
4. Фостер Д. Генеративное глубокое обучение. - СПб: Питер, 2020. - 336 с.
5. MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment // arxiv.org URL: <https://arxiv.org/abs/1709.06298> [Электронный ресурс] (дата обращения: 15.02.21).
6. Описание интерфейса MIDI // midi.ru URL: <http://www.midi.ru/doc/35.htm> [Электронный ресурс] (дата обращения: 20.02.21).
7. Adversarial Audio Synthesis // arxiv.org URL: <https://arxiv.org/abs/1802.04208> [Электронный ресурс] (дата обращения: 25.02.21).
8. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks // arxiv.org URL: <https://arxiv.org/abs/1511.06434> [Электронный ресурс] (дата обращения: 07.03.21).
9. Wasserstein GAN // arxiv.org URL: <https://arxiv.org/pdf/1701.07875v3.pdf> [Электронный ресурс] (дата обращения: 14.03.21).
10. GANSynth: Adversarial Neural Audio Synthesis // arxiv.org URL: <https://arxiv.org/abs/1902.08710> [Электронный ресурс] (дата обращения: 19.03.21).
11. NSynth: Neural Audio Synthesis // magenta URL: <https://magenta.tensorflow.org/nsynth> (дата обращения: 18.03.21).
12. youtube // youtube URL: <https://www.youtube.com/> [Электронный ресурс] (дата обращения: 24.03.21).
13. youtube-dl // github. URL: <https://github.com/ytdl-org/youtube-dl> [Электронный ресурс] (дата обращения: 29.03.21).
14. Мельница // youtube URL: <https://www.youtube.com/c/HeLavisA/playlists> [Электронный ресурс] (дата обращения: 03.04.21).
15. Диана Арбенина и Ночные Снайперы // youtube URL: <https://www.youtube.com/c/DianaArbenina> [Электронный ресурс] (дата обращения: 03.04.21).
16. Город312 // youtube URL: <https://www.youtube.com/user/gorod312band/about> [Электронный ресурс] (дата обращения: 03.04.21).

17. zemfira // youtube URL: https://www.youtube.com/channel/UCaYQXoII8_plzI_nDrmlZJQ [Электронный ресурс] (дата обращения: 03.04.21).
18. FFmpeg // ffmpeg.org URL: <https://ffmpeg.org/> [Электронный ресурс] (дата обращения: 08.04.21).
19. Spleeter by deezer // github URL: <https://github.com/deezer/spleeter> [Электронный ресурс] (дата обращения: 15.04.21).
20. denoiser // github URL: <https://github.com/facebookresearch/denoiser/> [Электронный ресурс] (дата обращения: 17.04.21).
21. Real Time Speech Enhancement in the Waveform Domain // arxiv.org URL: <https://arxiv.org/abs/2006.12847> [Электронный ресурс] (дата обращения: 20.04.21).
22. LSTM: A Search Space Odyssey // arxiv.org URL: <https://arxiv.org/pdf/1503.04069.pdf> [Электронный ресурс] (дата обращения: 22.04.21).
23. librosa // librosa.org URL: <https://librosa.org/doc/latest/index.html> [Электронный ресурс] (дата обращения: 25.04.21).
24. Understanding the Mel Spectrogram // medium.com URL: <https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53> [Электронный ресурс] (дата обращения: 01.05.21).
25. A Scale for the Measurement of the Psychological Magnitude Pitch // scitation URL: <https://asa.scitation.org/doi/10.1121/1.1915893> [Электронный ресурс] (дата обращения: 01.05.21).
26. NumPy // numpy.org URL: <https://numpy.org/> [Электронный ресурс] (дата обращения: 01.05.21).
27. Scipy // scipy.org URL: <https://www.scipy.org/> [Электронный ресурс] (дата обращения: 01.05.21).
28. Tutorial on Variational Autoencoders // arxiv.org URL: <https://arxiv.org/abs/1606.05908> [Электронный ресурс] (дата обращения: 08.05.21).
29. Tensorflow // tensorflow.org URL: <https://www.tensorflow.org/> [Электронный ресурс] (дата обращения: 13.05.21).
30. Keras: the Python deep learning API // keras.io URL: <https://keras.io/> [Электронный ресурс] (дата обращения: 15).
31. Auto-Encoding Variational Bayes // arxiv.org URL: <https://arxiv.org/abs/1312.6114> [Электронный ресурс] (дата обращения: 13.05.21).
32. On Information and Sufficiency // projecteuclid URL: <https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-22/issue-1/On-Information-and-Sufficiency/10.1214/aoms/1177729694.full> [Электронный ресурс] (дата обращения: 15.05.21).
33. GAN — Wasserstein GAN & WGAN-GP // medium URL: <https://jonathan-hui.medium.com/gan-wasserstein-gan-wgan-gp-ba1a2aa1b490> [Электронный ресурс] (дата обращения: 15.05.21).

34. Wasserstein GAN // arxiv.org URL: <https://arxiv.org/abs/1701.07875> [Электронный ресурс] (дата обращения: 15.05.21).
35. Improved Training of Wasserstein GANs // arxiv.org URL: <https://arxiv.org/pdf/1704.00028.pdf> [Электронный ресурс] (дата обращения: 16.05.21).
36. Autoencoding beyond pixels using a learned similarity metric // arxiv.org URL: <https://arxiv.org/abs/1512.09300> [Электронный ресурс] (дата обращения: 16.05.21).
37. Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 27.12.2018) // Гарант URL: <https://base.garant.ru/12125268> [Электронный ресурс] (дата обращения 25.04.2021)
38. ГОСТ 12.2.032-78 Система стандартов безопасности труда (ССБТ). Рабочее место при выполнении работ сидя. Общие эргономические требования // Гарант URL: <https://base.garant.ru/3924216> [Электронный ресурс] (дата обращения: 01.05.2021).
39. СанПиН 1.2.3685-21 "Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания". // Гарант URL: <https://base.garant.ru/400274954/> [Электронный ресурс] (дата обращения: 01.05.2021).
40. ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация» // Электронный фонд URL: <https://docs.cntd.ru/document/120013607> [Электронный ресурс] (дата обращения: 01.05.2021).
41. СанПиН 2.2.4.548–96. Гигиенические требования к микроклимату производственных помещений. // Гарант URL: <https://base.garant.ru/4173106/> [Электронный ресурс] (дата обращения: 01.05.2021).
42. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95. // КонсультантПлюс URL: http://www.consultant.ru/document/cons_doc_LAW_237487/ [Электронный ресурс] (дата обращения: 01.05.2021).
43. ТОИ Р-45-084-01 “Типовая инструкция по охране труда при работе на персональном компьютере” // Электронный фонд URL: <https://docs.cntd.ru/document/1200030047> [Электронный ресурс] (дата обращения: 01.05.2021).
44. СН 2.2.4/2.1.8.562-96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки. // Гарант URL: <https://base.garant.ru/4174553/> [Электронный ресурс] (дата обращения: 01.05.2021).
45. ГОСТ 12.1.038-82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов. // Гарант URL: <https://base.garant.ru/3924202/> [Электронный ресурс] (дата обращения: 01.05.2021).

46. ГОСТ 12.1.006-84 ССБТ. «Электромагнитные поля радиочастот. Допустимые уровни на рабочих местах и требования к проведению контроля» // Гарант URL: <https://base.garant.ru/6305025/> [Электронный ресурс] (дата обращения: 01.05.2021).
47. Федеральный закон № 89 от 1998г. «Об отходах производства и потребления» // КонсультантПлюс URL: http://www.consultant.ru/document/cons_doc_LAW_19109/ [Электронный ресурс] (дата обращения: 01.05.2021).
48. ГОСТ Р 53692-2009 Ресурсосбережение. Обращение с отходами. Этапы технологического цикла отходов. // Гарант URL: <https://base.garant.ru/2175160/> [Электронный ресурс] (дата обращения: 01.05.2021).
49. ГОСТ 12.1.004-91 «Пожарная безопасность. Общие требования». // Гарант URL: <https://base.garant.ru/3922226/> [Электронный ресурс] (дата обращения: 01.05.2021).

Приложение А

Раздел №2 Downloading and preparing data for training «Application of generative adversarial networks in the problem of generating complex musical compositions»

Студент:

Группа	ФИО	Подпись	Дата
8ВМ93	Колобов Ростислав Сергеевич		

Руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ ИШИТР	Спицын Владимир Григорьевич	Д.Т.Н.		

Консультант-лингвист отделения иностранных языков ШБИП:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель ОИЯ	Розанова Яна Викторовна			

Introduction

Sound synthesis for a variety of applications has many uses in creative sound design for music and film. Musicians and artists browse large databases of sound effects. This strategy is painstaking and may not work if a suitable sound effect is not available in the library.

Despite the huge, constantly growing amount of music content on the Internet, the creation of new musical works is still as relevant as the creation of works of other types of art. New genres, performers and musicians appear and disappear. At the same time, there is a growing interest in the automatic generation of musical works. However, at the moment, the results in most cases remain monophonic and exclusively instrumental, which in turn leaves a lot of scope for research. Until now, the question of the most convenient presentation of audio data, which could use the best results in such a task, remains open.

This paper discusses the use of generative adversarial networks in the task of generating music. A set of scripts is also implemented to automate the creation of training data. In addition, a number of experiments are being carried out with autoencoders and variational autoencoders, the purpose of which is to study the possibility of extracting informative features from data, which can significantly reduce the data dimension and simplify the training of a generative adversarial network.

2. Downloading and preparing data for training

In order to collect the training dataset, a number of scripts have been developed to automate the downloading and preparation of data. In the work, audio tracks obtained from the open video hosting Youtube are used as data for training. [1].

2.1 Downloading unprepared data from Youtube

To download unprepared data, two scripts have been developed that based on youtube_dl [13] library. The main goals of the scripts are:

- getting a list of URLs for each video;
- direct download of data, based on the list obtained at the previous stage.

One can get a list of video URLs:

- using the URL of a specified Youtube channel, then the URL of all videos related to this channel will be obtained;
- by specifying the keyword, based on which the search will be made among the videos posted on Youtube.

Further, using the received URLs, the audio tracks are directly downloaded. Next, only the audio components of each video are extracted and saved in webm or mp4 formats. However, with a large number of requests, video hosting can limit the number of requests per day, up to limiting access to the service from a specific IP address. Therefore, the data downloading script also provides the ability to specify a list of proxy addresses, the use of which allows one to download audio tracks from multiple streams using multiple proxies at the same time.

In this thesis, downloading by specifying the URLs of the following Youtube channels was performed:

- folk-rock group "Melnitsa" [14];
- Diana Arbenina and «Nochnyye Snaypery» [15];
- «Gorod 312» [16];
- «Zemfira» [17].

To avoid getting audio from a promo video or interview into the training set, additional filtering is performed based on the duration of the video. Only tracks that are at least two and a half minutes in length and no more than five minutes in length are saved.

2.2 Data preparation

To facilitate the learning process, it is necessary to simplify the training data and bring it to a general form.

2.2.1 Convert audio to the same format

It is necessary to convert the data obtained in the previous step into audio format, and also change the sampling rate. Conversion is performed using the ffmpeg [6] command line utility with the following parameters:

- sampling rate: 16 KHz;
- 16 bit;
- mono;
- wav format.

2.2.2 Splitting audio into musical components

In order to simplify the training of the network, the audio tracks are divided into musical parts:

- vocal;
- drums;
- the rest of the instruments sounds.

It is assumed that such a division makes it possible to simplify the study of complex compositional distributions.

Splitting audio into musical parts is done using the open source library Spleeter [19]. This library provides the ability to use pre-trained neural network models that allow one to divide the original audio into two, four and five musical parts. One of the advantages of this library is its high speed; splitting audio into parts using a GPU can take up to 0.01 of the original audio duration. Since the library does not provide the ability to divide only into the three parts described above, the division

is carried out using a four-part model, after which the parts that are not related to vocals and drums are combined using the ffmpeg utility.

2.2.3 Removing Noise from a Vocal Part

When examining parted audio data, it was found that vocals often contain artifacts similar to echoes from other parts. Since such artifacts are not related to vocals, it becomes necessary to exclude them.

To remove artifacts from vocals, Facebook Denoiser [20] is used - this is a development that is based on a neural network approach, which is an implementation of the article [21]. The model used in the development is based on the LSTM encoder-decoder architecture [22]. The model has been trained specifically to remove noise in audio containing voice, allowing one to remove various types of background noise, including stationary and non-stationary noise, and room reverberation. Noise removal is used exclusively for vocals.

2.2.4 Dividing audio into sections containing vocals

Another simplification is training not on all audio, but on some part of it; only audio segments are used in training, containing speech at least 50% of its duration. In order to split audio into sections containing vocals, the librosa [23] library is used, and in particular its split method, which splits audio into intervals that do not contain silence. For each audio, an additional array is created, of the same size, containing units in those intervals where vocals are present in the audio. Further, on the basis of the obtained array, the selection of sections containing speech of at least 50% of its duration is performed using the specified window size and its offset step. In this work, we used a window with a size of 330000 with a step of 32000, which corresponds to approximately 20 seconds of audio with a step size of two seconds. If the duration of speech in the window is at least 10 seconds, then the windows with the corresponding coordinates are extracted from the audio files of the vocals, drums and other instruments and are transferred to the next stage of preparation, otherwise the window is discarded. Thus, the result of this process is a tensor of the form $N \times 330000 \times 3$, where N is the number of satisfying windows extracted from all audio.

2.2.5 Extraction of acoustic characteristics. Spectrograms.

The data obtained in the previous step is still presented as a sequence of instantaneous values, read at regular intervals (Figure 1). It is rather complicated and computationally expensive to train on such data; therefore, it makes sense to extract frequency features of a lower dimension, but which retain their information content well enough.

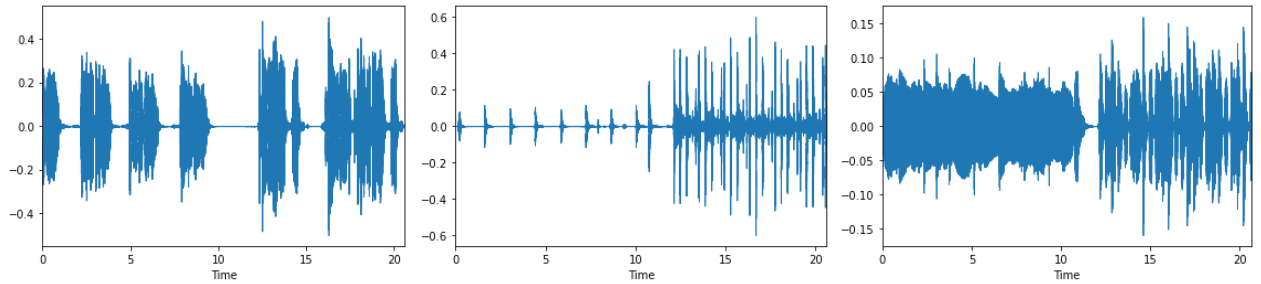


Fig. A1 Data in the form of instant values (vocals, drums, rest)

Spectrograms are the result of a Fast Fourier Transform computation applied over overlapping window segments. This view allows one to visually represent the amplitude of a signal that changes over time at different frequencies. In this case, the ordinate axis is converted to a logarithmic scale, the value at a specific point is converted to decibels. Figure 2 shows an example of the histograms.

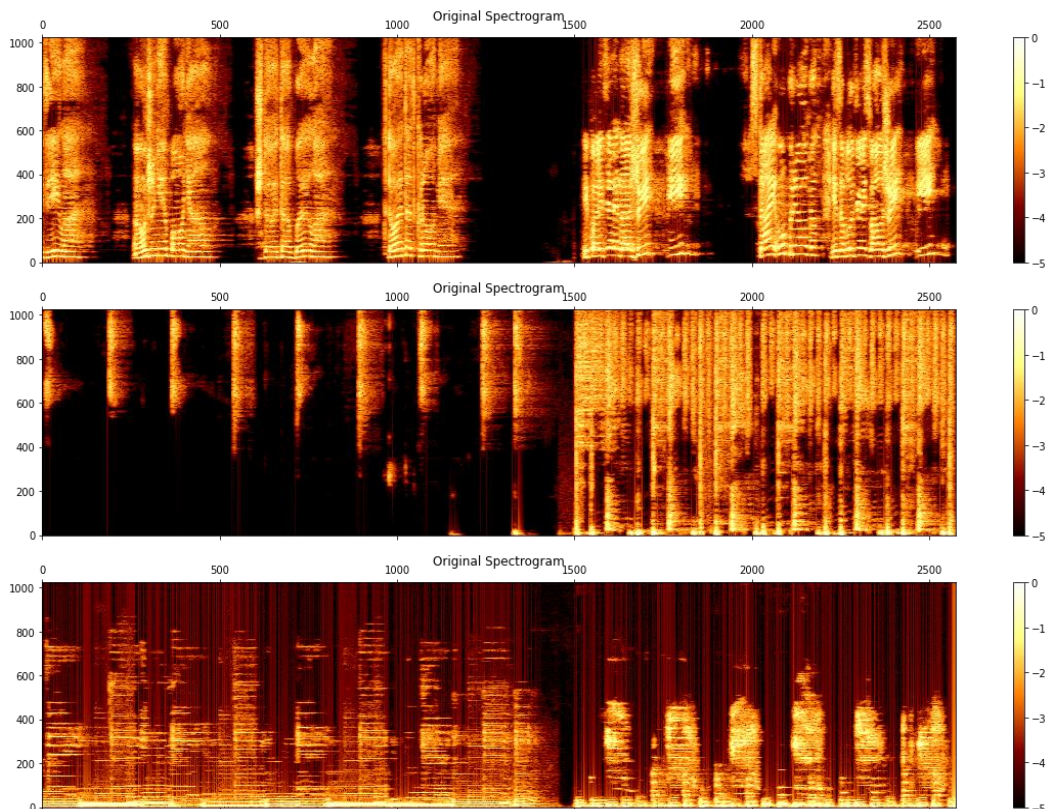


Fig. A2 An example of the spectrograms (vocals, drums, the rest)

However, humans do not perceive frequencies on a linear scale and are better at detecting differences at lower frequencies than at higher frequencies. Also, using histograms as training data is still computationally expensive, and the presence of small details will greatly complicate training. Therefore, it was decided to simplify the data, at the cost of quality loss, and convert the frequency values of the spectrograms to a Mel-scale [24].

A Mel-spectrogram is a spectrogram in which the frequency scale is converted to the Mel-scale - the psychophysical unit of pitch [25].

The spectrograms converted to Mel-scale are shown in Fig. 3

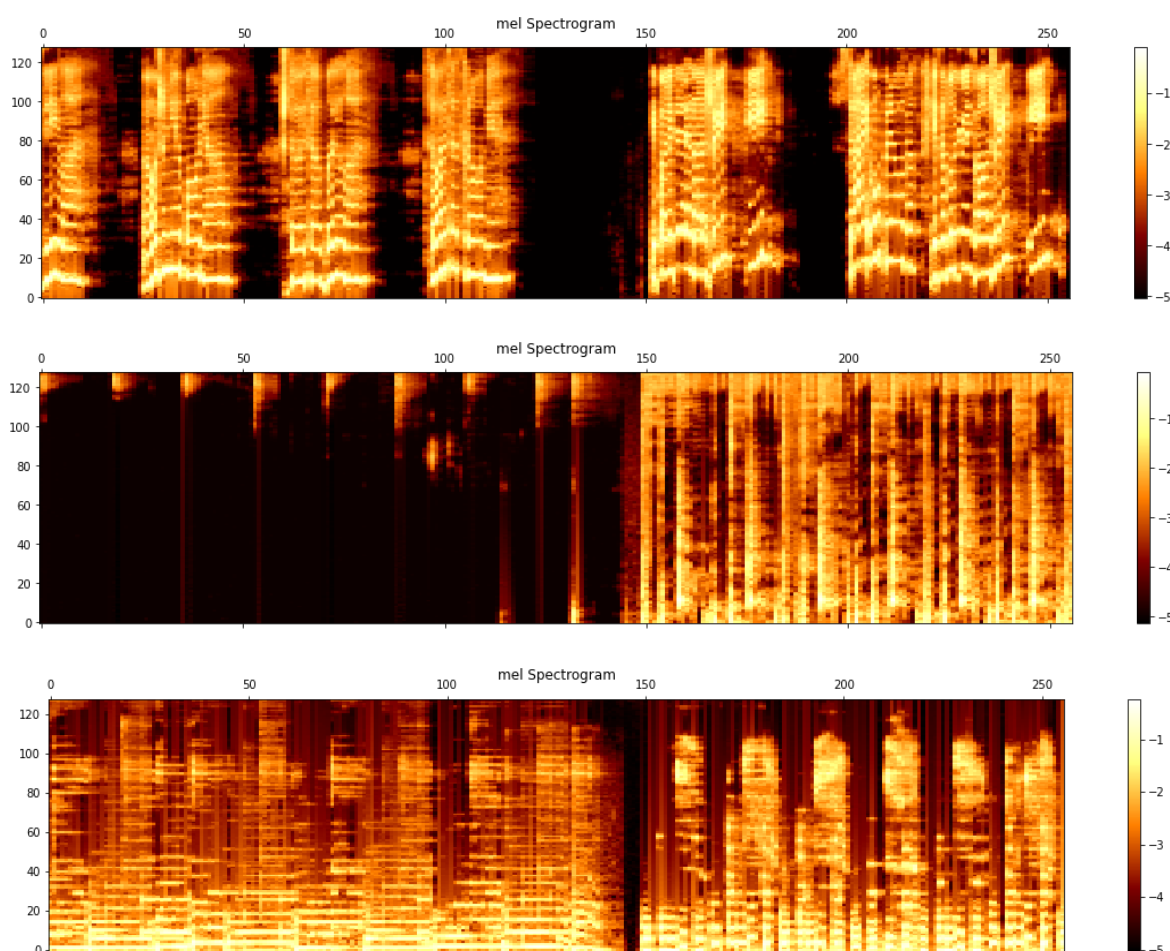


Fig. A3 An example of Mel-spectrograms (vocals, drums, the rest)

Since the transformation significantly compresses the data, some information component is lost, however, the audio data recovered from the Mel-spectrograms still reflects the original data quite well (Figure 4).

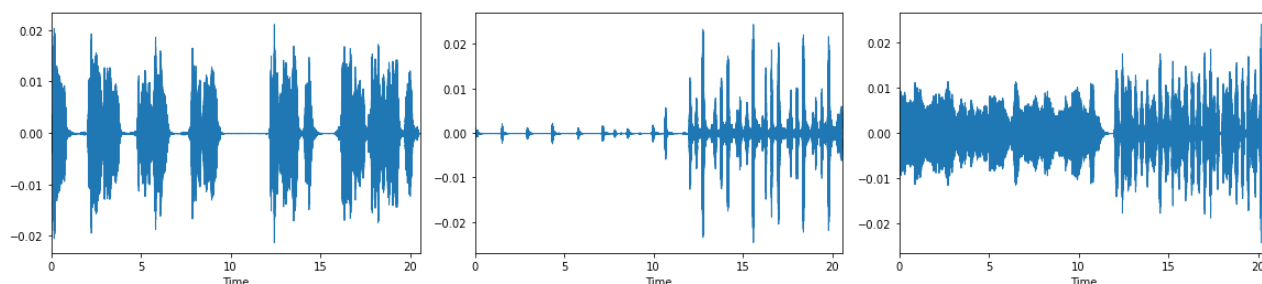


Fig. A4 Recovered audio (vocals, drums, rest)

The transformations were carried out with the following set of parameters:

- window size for fast Fourier transform: 2048;
- window step size: 128;
- number of Mel-frequency components: 128;
- shorten factor: 10;
- minimum frequency: 300;
- maximum frequency: 16000.

The entire conversion process is based on the use of the numpy [26], scipy [27] and librosa libraries.

Thus, the training set is represented by a tensor of the form $N \times 128 \times 256 \times 3$.

3. Initial experiments

To explore the possibility of further reducing the size of the training data, it makes sense to train autoencoder neural networks and extract the resulting latent representation.

3.1 Learning the autoencoder network

First of all, it is necessary to study the ability of a neural network to create a hidden view that allows one to reconstruct data.

Autoencoders [28] refer to feedforward neural networks that allow reconstruction of the input signal at the output. Inside the neural network, there is a hidden layer that describes the data in general. This type of neural networks is

designed primarily in such a way that the neural network does not have the opportunity to ideally copy the input signal to the output; for this, the size of the latent representation must be less than the dimension of the input data. However, due to this limitation, the input signal is restored with errors due to the loss of information during encoding, which forces the neural network to select the most informative features. [4].

An autoencoder neural network consists of two parts: an encoder and a decoder. The task of the encoder is to compress the input data to the dimension of the latent representation, and the decoder is to restore the original signal from the received latent representation.

At this stage, a convolutional auto-encoder is used, the architecture of which is shown in Figure 5. In this case, only a part of the training set is used, namely only the vocal component, limited to 64 values along the time axis and 64 values along the frequency axis.

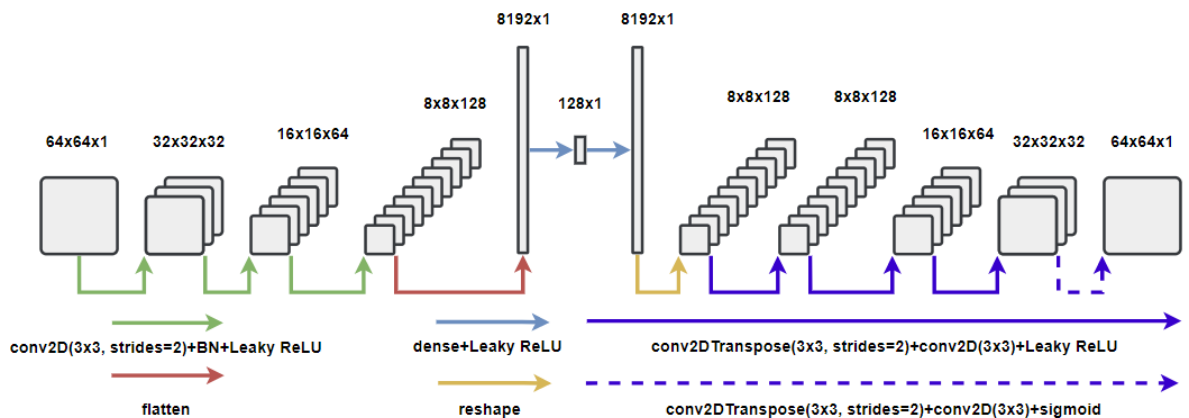


Fig. A5 Autoencoder architecture

This architecture was implemented and trained with TensorFlow [29] implementation using Keras high-level API [30]. Training was performed using binary cross-entropy as a loss function, and adam as an optimizer with a learning rate of 0.001. The training was carried out for 64 epochs, but it was stopped at the 50th epoch due to the absence of a decrease in the loss function, reaching a value of 0.5595 on the training set and 0.5471 on the validation set.

The results of the reconstruction of the input data are shown in Figure 6.

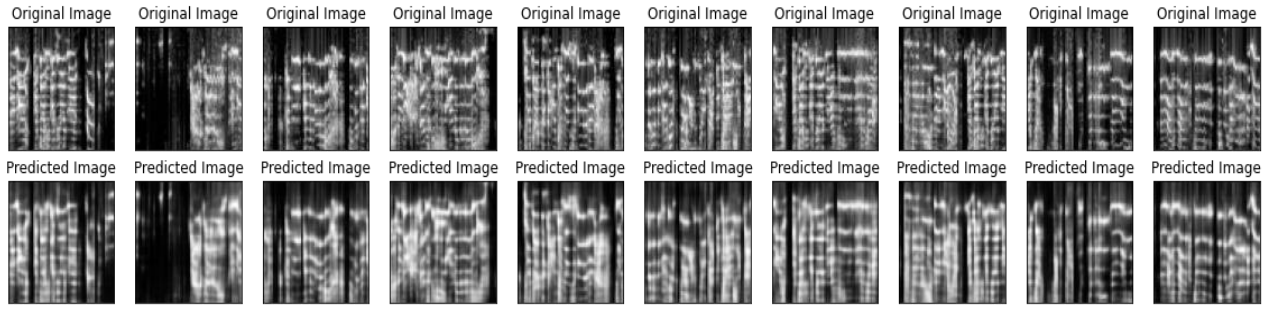


Fig. A6 Results of input data reconstruction

The figure shows that the network did a fairly good job of reconstructing the output. However, since the hidden space of autoencoder networks is not continuous, it cannot be used as a full-fledged representation of data that it has not seen, as well as for generating new ones. Therefore, it makes sense to move on to the next stage, namely, training the neural network of the variational autoencoder, the continuous hidden representation of which corrects the shortcomings of the current neural network.

3.2 Training the variational autoencoder network

Autoencoder neural networks map each image to a point in a hidden space that is not continuous. In a variational autoencoder, each image is mapped to a multivariate normal distribution around a point in hidden space. The continuity of the hidden space is achieved by using two hidden vectors: the vector of means and the vector of standard deviations, which together define the multivariate normal distribution in the hidden space. The mapping to the logarithm of variance is often also chosen, because it can take any real values in the range $(-\infty, \infty)$, which corresponds to the natural range of the output values of the neural network node, while the variance values are always positive. Eventually, the encoder converts each input image into two vectors, μ – for means and $\log(\sigma^2)$ for the log variance, which together define a multivariate normal distribution in hidden space. It is possible to transform an image to a specific z-point in hidden space by sampling from this distribution using the following equation [4]:

$$z = \mu + \sigma * \varepsilon, \quad (3.2.1)$$

where $\sigma = \exp(\log(\sigma^2)/2)$, and ε – a point selected from the standard normal distribution.

Since a random value is selected from the area around μ , then all neighboring points must be decoded into similar images. In this case, reconstruction losses should remain small. This operation is called sampling, and its result is the output of the encoder.

To generate new data, it is not enough just to compare vectors with a distribution. Since there is no restriction on what values the mean and standard deviation vectors take, the encoder can return distributions with different means for different classes or clusters with little deviation from the mean so that encodings do not differ much from the same sample. As a result, the coded distributions are far apart. Therefore, the Kullback-Leibler distance (KL) [32] is introduced into the auto-encoder loss function, which measures how much one probability distribution differs from another. In this case, we need to know how much the resulting distribution differs from the normal one, the loss function using the distance KL is as follows:

$$KL_{loss} = -0.5 * \sum(1 + \log(\sigma^2) - \mu^2 - \exp(\log(\sigma^2))) \quad (3.2.2)$$

This loss function forces the encoder to distribute values close to the standard normal distribution (mean 0 and standard deviation 1). In this case, it is ensured that the distributions are close and actually overlap. This increases the chance that the latent view is truly continuous. However, by optimizing only the KL divergence, the encoder will randomly plot points close to the center. It needs to be more organized by class, so reconstruction loss needs to be used as well. The sum of these two terms is the loss function of the variational autoencoders.

The variational autoencoder network used in the experiment has the architecture shown in Figure 7. The loss function includes not only the reconstruction error, but also the Kullback-Leibner distance.

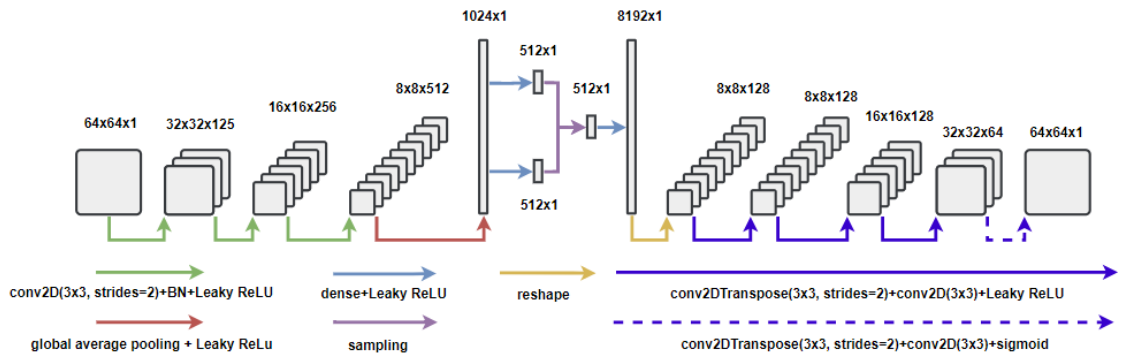


Fig. A7 The architecture of the variational autoencoder

Learning was stopped at the 150th epoch due to the absence of the decrease in the loss function, while with a decrease in the value of losses during reconstruction, an increase in the value of losses KL occurred, which indicates that the network cannot cope with the task of generating a continuous representation.

The results of reconstruction after training the model are shown in the figure 8.

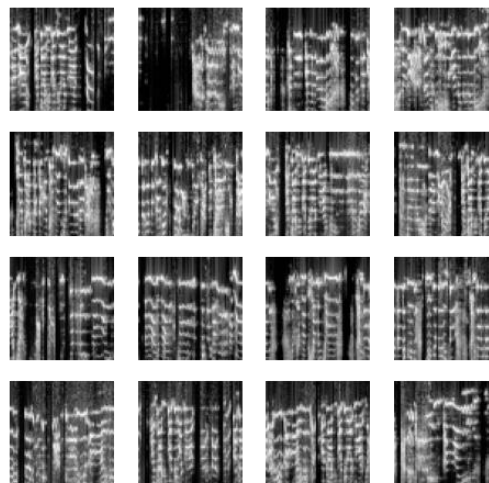


Fig. A8 Results of reconstruction by a variational autoencoder

Figure 8 shows that the results of reconstruction are similar to the real images of spectrograms. However, blurring of the images, which is a serious drawback of variational autoencoders for the problem stated in the thesis, can be seen. This blurring introduces additional noise into the reconstructed audio. This fact can significantly complicate the training of generative models, which, together with a high KL loss, does not allow using the latent representation to generate new data. Therefore, it was decided to continue experiments with generative adversarial networks using spectrogram data.

(Справочное)

Процесс получения и подготовки данных

