

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Інформаційних радіотехнологій та технічного захисту інформації  
(повна назва)

Кафедра Медіаінженерії та інформаційних радіоелектронних систем  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА

### Пояснювальна записка

рівень вищої освіти другий (магістерський)  
(позначення документа)

Комплексна тема. Використання нейромереж у роботі з фото та відео.  
Формування етапів та критеріїв.  
(тема)

Виконав:

студент 2 курсу, групи СТМм-22-1  
Денис БЕЗПАЛЬКО

(прізвище, ініціали)

Спеціальність 171 – Електроніка

(код і повна назва спеціальності)

Тип програми освітньо-професійна  
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи, технології і комп'ютерні засоби мультимедіа  
(повна назва освітньої програми)

Керівник ст. викл. Тетяна ВАСИЛЕНКО  
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри \_\_\_\_\_  
(підпис)

Володимир КАРТАШОВ  
(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет Інформаційних радіотехнологій та технічного захисту інформації

Кафедра Медіаінженерії та інформаційних радіоелектронних систем

Рівень вищої освіти другий (магістерський)

Спеціальність 171 – Електроніка

(код і повна назва)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма "Системи, технології і комп'ютерні засоби мультимедіа "

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_

(підпис)

« \_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ р.

## ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

Студентові Безпалько Денису Віталійовичу

(прізвище, ім'я, по батькові)

1. Тема роботи: Комплексна тема. Використання нейромереж у роботі з фото та відео. Формування етапів та критеріїв.

затверджена наказом по університету від " 20.11.2023" № 1371Ст

2. Термін подання студентом роботи 08.01.2024 р.

3. Вихідні дані до проєкту (роботи) \_\_\_\_\_

1. Огляд та дослідження принципів роботи нейронних мереж у мультимедійній обробці

2. Огляд програмного забезпечення та практичні аспекти

3. Практичне застосування нейромереж у створенні голосу, музики та покращенні якості відео

4. Перелік питань, що потрібно опрацювати в роботі

ВСТУП

1. ПРИНЦИП РОБОТИ НЕЙРОННИХ МЕРЕЖ У МУЛЬТИМЕДІЙНІЙ ОБРОБЦІ

2. НЕЙРОМЕРЕЖІ ДЛЯ СТВОРЕННЯ МУЛЬТИМЕДІЙНОГО КОНТЕНТУ: ОГЛЯД ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ТА ПРАКТИЧНІ АСПЕКТИ

3. ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖ У СТВОРЕННІ ГОЛОСУ, МУЗИКИ ТА ПОКРАЩЕННІ ЯКОСТІ ВІДЕО

ВИСНОВКИ

ПЕРЕЛІК ПОСИЛАННЯ

ДОДАТКИ

5. Перелік графічного матеріалу із зазначенням обов'язкових креслеників, схем, плакатів, комп'ютерних ілюстрацій:

1. Вікно зміни голосу; 2. Аудіозапис внизу інтерфейса; 3. Типи треків; 4. Стиль треків; 5. Вікно налаштування ефектів; 6. Вікно налаштувань чіткості та розширення відео.

## КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Отримання теми дипломної роботи	21.11.23-27.11.23	
2.	Підготовка теоретичної частини	28.11.23-02.12.23	
3.	Робота з технічною частиною	04.12.23-09.12.23	
4.	Оформлення дипломної роботи	11.12.23-20.12.23	
5.	Графічна частина роботи	21.12.23-23.12.23	
6.	Перевірка керівником	24.12.23-26.12.23	
7.	Перевірка на академічний плагіат	27.12.23-28.12.23	
8.	Перевірка завідувачем кафедри, рецензування	29.12.23-02.01.24	

Дата видачі завдання \_\_\_\_\_ 20.11.2023 р. \_\_\_\_\_

Студент \_\_\_\_\_  
  
 (підпис)

Керівник роботи \_\_\_\_\_  
 (підпис)

\_\_\_\_\_ Денис БЕЗПАЛЬКО \_\_\_\_\_

\_\_\_\_\_ ст. викл. Тетяна ВАСИЛЕНКО \_\_\_\_\_  
 (посада, прізвище, ініціали)

## РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи магістра має: 75 сторінок, 22 джерел, 34 рисунка, 2 додатки.

НЕЙРОННІ МЕРЕЖІ, ГЛИБОКЕ НАВЧАННЯ, МУЛЬТИМЕДІЙНИЙ ВМІСТ, ОБРОБКА МУЛЬТИМЕДІА, ГОЛОСОВА ОБРОБКА, МУЗИЧНА ТВОРЧИСТЬ, ВИБІР КАДРІВ, РЕДАГУВАННЯ ВІДЕО.

Об'єкт дослідження – використання нейронних мереж, зокрема глибокого навчання, у обробці та творчому використанні мультимедійного вмісту.

Предмет дослідження – технології глибокого навчання для вдосконалення голосу, створення музики, вибору найкращих кадрів і редагування відео в широкому спектрі завдань у сфері мультимедіа.

Мета роботи – вивчити та застосувати передові методи глибокого навчання для покращення різноманітних аспектів мультимедійного вмісту, а також розробити нові підходи для підвищення якості та швидкості обробки.

Методи дослідження: аналіз принципів нейронних мереж, огляд нейронних мереж для створення мультимедійного контенту та програмного забезпечення, практичне застосування нейронних мереж у генерації голосу, музики та відео.

Кваліфікаційна робота магістра складається з вступу, трьох основних розділів, висновку та списку літератури. У першому розділі розглядаються принципи нейронних мереж у обробці мультимедіа. У другому розділі досліджуються нейронні мережі для створення мультимедійного контенту та надається огляд відповідного програмного забезпечення. У третьому розділі розглядаються практичні застосування нейронних мереж у генерації голосу, створенні музики та покращенні якості відео. У висновку узагальнено результати дослідження та окреслено перспективи нейронних мереж у фото- та відеоманіпуляції.

Кваліфікаційна робота магістра має як теоретичне, так і практичне значення. Його результати можуть бути корисними для подальших досліджень нейронної мережі в мультимедійній сфері та для розробки вдосконалених методологій навчання для підвищення досвіду в цій галузі.

## ABSTRACT

The explanatory note of the master's thesis comprises 75 pages, 22 references, 34 figures, 2 supplements.

NEURAL NETWORKS, DEEP LEARNING, MULTIMEDIA CONTENT, MULTIMEDIA PROCESSING, VOICE PROCESSING, MUSIC CREATION, FRAME SELECTION, VIDEO EDITING.

The object of the research is the utilization of neural networks, particularly deep learning, in the processing and creative utilization of multimedia content.

The subject of the research involves deep learning technologies to enhance voice, create music, select the best frames, and edit videos across a wide range of tasks in the multimedia field.

The aim of the work is to study and apply advanced deep learning methods to improve various aspects of multimedia content and to develop new approaches to enhance processing quality and speed.

Research methods include the analysis of neural network principles, reviewing neural networks for multimedia content creation and corresponding software, as well as practical application of neural networks in voice, music, and video generation.

The master's thesis consists of an introduction, three main sections, a conclusion, and a bibliography. The first section explores the principles of neural networks in multimedia processing. The second section investigates neural networks for multimedia content creation and provides an overview of relevant software. The third section examines practical applications of neural networks in voice generation, music creation, and video enhancement. The conclusion summarizes the research results and outlines the prospects of neural networks in photo and video manipulation.

The master's thesis holds both theoretical and practical significance. Its findings can be valuable for further research on neural networks in the multimedia field and for

the development of advanced teaching methodologies to enhance expertise in this domain.

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

RNN – рекурентні нейронні мережі

GAN – генеративно-супервізійні нейронні мережі

VAE – варіаційні автокодери

RPN – мережа виявлення областей

R-CNN – нейронна мережа на основі регіонів з конволюціями

LSTM – короткострокова пам'ять довгострокової пам'яті

GRU – вентиляційний рекурентний вузол

CNN – конволюційні нейронні мережі

## ЗМІСТ

ВСТУП.....	11
1 ПРИНЦИП РОБОТИ НЕЙРОННИХ МЕРЕЖ У МУЛЬТИМЕДІЙНІЙ ОБРОБЦІ .....	13
1.1 Створення голосу за допомогою нейромереж.....	13
1.1.1 Переробка голосу за допомогою нейромереж .....	20
1.1.2 Створення голосу через нейромережі .....	21
1.1.3 Покращення голосу за допомогою нейромереж .....	23
1.2 Створення музики за допомогою нейромереж .....	24
1.3 Нейромережі в процесі підбору кращих кадрів .....	27
1.4 Монтаж відео з використанням нейромереж .....	29
1.4.1 Розпізнавання об'єктів .....	29
1.4.2 Автоматичне позиціонування елементів .....	30
1.4.3 Колірна та експозиційна корекція .....	31
1.4.4 Автоматизація обрізання та зміни ракурсу .....	32
1.4.5 Автоматичне видалення шуму та відновлення якості.....	34
2 НЕЙРОМЕРЕЖІ ДЛЯ СТВОРЕННЯ МУЛЬТИМЕДІЙНОГО КОНТЕНТУ: ОГЛЯД ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ТА ПРАКТИЧНІ АСПЕКТИ.....	36
2.1 Використання нейромереж у створенні голосу .....	36
2.1.1 Генератор мовлення LOVO (Genny) .....	37
2.1.2 Генератор мовлення Play.ht.....	39
2.1.3 Застосунок Speechify.....	41
2.1.4 Генератор мовлення Voicemaker .....	43

	10
2.2 Створення та обробка музичного контенту з використанням технологій нейромереж .....	45
2.2.1 Музичний генератор Amper Music .....	46
2.2.2 Музичний генератор AIVA .....	48
2.2.3. Музичний генератор Soundful .....	49
2.3 Покращення якості відео з використанням нейромереж та інших технологій .....	51
2.3.1 Інструмент для покращення відео Topaz Video Enhance AI .....	52
2.3.2 Інструмент для покращення відео Dain-App .....	54
2.3.3 Інструмент для покращення відео VEED.IO .....	56
2.3.4 Інструмент для покращення відео Unifab .....	59
3 ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖ У СТВОРЕННІ ГОЛОСУ, МУЗИКИ ТА ПОКРАЩЕННІ ЯКОСТІ ВІДЕО .....	62
3.1 Створення голосу за допомогою Voicemaker .....	62
3.2 Створення музики за допомогою Soundful .....	65
3.3 Покращення відео за допомогою Unifab .....	67
ВИСНОВКИ .....	72
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ .....	74
ДОДАТКИ .....	76
ДОДАТОК А .....	77
ДОДАТОК Б .....	83

## ВСТУП

У центрі сучасної динаміки цифрового світу спостерігається вражаюче зростання мультимедійного контенту, яке змінює способи сприйняття, створення та обміну інформацією. З кожним днем інтернет стає все більш насиченим відео-, фото-, та аудіо-змістом, відображаючи не тільки наші історії, а й культуру, творчість та спосіб вираження. Проте, великий об'єм мультимедійних даних також створює проблеми – від ускладнення обробки до нестачі ефективних засобів аналізу та синтезу. У зв'язку з цим, нові парадигми, зокрема використання нейромереж, набувають великого значення для створення інноваційних рішень у цій сфері.

У контексті сучасні техніки обробки фотографій, відео та звуку потребують суттєвого покращення, яке може бути забезпечене за допомогою нейромереж, особливо глибокого навчання. Нейромережі відкривають нові перспективи в обробці мультимедійного контенту, надаючи здатність не тільки адаптуватися до швидкозмінних потреб, але й самостійно вивчати та вдосконалювати способи обробки даних.

Кваліфікаційна робота магістра фокусується на використанні нейромереж, зокрема глибокого навчання, у обробці та творчому використанні мультимедійного контенту. Охоплюючи широкий спектр завдань, від покращення голосу та створення музики до підбірки найкращих кадрів та монтажу відео, кваліфікаційна робота магістра ставить перед собою мету дослідження та застосування передових технологій у сфері обробки мультимедійного контенту.

Головною метою є не лише огляд та аналіз поточних методів роботи з мультимедійним контентом, але й розробка нових підходів з використанням нейромереж, спрямованих на покращення якості та швидкості обробки. Реалізація цих інноваційних підходів може відкрити нові можливості для професіоналів у сфері мультимедіа, а також знайти широке застосування в різних галузях, від медіа-індустрії до особистого творчого використання.

За допомогою розвитку технологій глибокого навчання та їх впливу на цифрову творчість, кваліфікаційна робота магістра має потенціал перетворити підходи до створення та споживання мультимедійного контенту, відкриваючи нові можливості для його обробки та використання.

Кваліфікаційна робота магістра складається з вступу, трьох розділів, висновку та списку використаних джерел.

У першому розділі розглядається принципи роботи нейронних мереж у мультимедійній обробці. У другому розділі аналізуються нейромережі для створення мультимедійного контенту та проводиться огляд програмного забезпечення. У третьому розділі розглядалось практичне застосування нейромереж у створенні голосу, музики та покращенні якості відео. У висновку підводяться підсумки дослідження та визначаються перспективи розвитку нейромереж у роботі з фото та відео.

Кваліфікаційна робота магістра має теоретико-практичне значення. Її результати можуть бути використані для подальшого дослідження нейронних мереж у сфері мультимедіа, а також для розробки методів навчання та підвищення майстерності.

## 1 ПРИНЦИП РОБОТИ НЕЙРОННИХ МЕРЕЖ У МУЛЬТИМЕДІЙНІЙ ОБРОБЦІ

Можливості нейромереж та їх здатність адаптуватися до складних завдань в сфері мультимедійної обробки сьогодні стає не лише фундаментальною, але й перспективною для створення та покращення мультимедійного контенту. Принципи роботи цих нейронних структур відіграють ключову роль у реалізації творчих та технічних завдань, пов'язаних з обробкою фотографій, відео та аудіоматеріалів.

Мультимедійна обробка, на сьогоднішній день, перетинає межі стандартних технік, і це дозволяє їй відігравати ключову роль у сприйнятті, спілкуванні та творчості. Саме тут виникає необхідність в новаторських підходах до обробки мультимедійних даних, а також виявляється потенціал, що прихований у застосуванні нейромереж для здійснення цих завдань.

Розділ присвячений розгляду принципів функціонування нейромереж у сфері обробки мультимедійного контенту. Він ставить за мету визначення та аналіз способів, за якими нейромережі використовуються у процесах створення, покращення та обробки фото-, відео-, та аудіо-матеріалів. Проведення аналізу цих аспектів дозволить нам отримати глибше розуміння того, як саме нейронні мережі виконують ключові завдання у даній сфері та як їх можна ефективно використовувати для досягнення оптимальних результатів у мультимедійній обробці.

### 1.1 Створення голосу за допомогою нейромереж

Створення голосу за допомогою нейромереж - це процес моделювання або синтезування голосу за допомогою комп'ютерних алгоритмів, які ґрунтуються на нейромережах. Підхід базується на використанні глибокого навчання та штучних нейронних мереж для відтворення голосу, що максимально природно наближений до людського.

Створення голосу через неймережі включає в себе процес аналізу та синтезу звукових характеристик людського голосу. Це може включати в себе розпізнавання особливостей тону, інтонації, артикуляції та інших аспектів голосу людини.

Розробка моделей для створення голосу через неймережі включає в себе різноманітні архітектури та методики, що базуються на глибокому навчанні. Однією з популярних архітектур є генеративно-змагальні мережі (GAN), які використовуються для створення реалістичних медіа-контенту, такого як зображення, відео та звук. У випадку генерації голосу, GAN може використовуватись для створення аудіосигналів, які звучать природніше та мають більшу схожість з людським мовленням.

Генеративно-змагальні мережі (GAN) складаються з двох основних частин: генератора та дискримінатора (рис. 1.1), які працюють у взаємодії для створення реалістичних даних. У контексті синтезу голосу через неймережі, GAN можуть використовуватись для створення аудіозаписів, що максимально наближені до природного звучання людського мовлення.

- Генератор: частина мережі, яка відповідає за створення аудіо-даних. У випадку голосового синтезу, генератор намагається створити звук, який схожий на людський голос. Він працює на основі вивчених патернів з навчальних даних, створюючи нові аудіосигнали;

- Дискримінатор: частина мережі, яка функціонує як критик і намагається розрізнити між синтезованим аудіо та реальними аудіозаписами. Він отримує на вхід як синтезовані, так і реальні аудіодані, і навчається відрізнити одне від іншого.

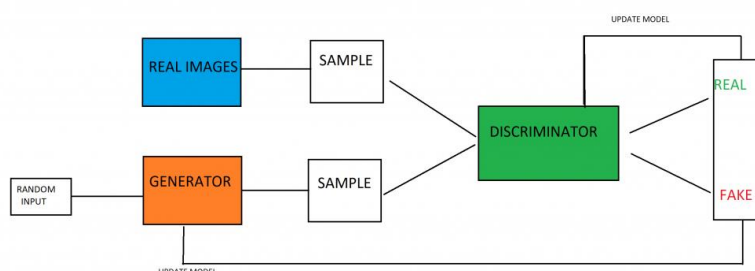


Рисунок 1.1 – Схематичне зображення генеративно-змагальної мережі [16]

Ці дві частини мережі працюють у взаємодії. Генератор старається підвищити якість створених аудіозаписів, відповідаючи на фідбек, який він отримує від дискримінатора про його здатність обдурити дискримінатор. У той же час, дискримінатор намагається поліпшити свою здатність розрізняти справжні аудіодані від синтезованих.

Взаємодія між генератором і дискримінатором стимулює обидві частини мережі до постійного покращення, що в результаті призводить до створення все більш реалістичних та природних аудіозаписів, які можуть бути використані для синтезу голосу.

Ще однією ефективною архітектурою є рекурентні нейронні мережі (RNN) та їх варіації, такі як LSTM (Long Short-Term Memory) або GRU (Gated Recurrent Unit). Вони ефективно використовуються для моделювання послідовностей та дозволяють зберігати контекст при генерації тексту чи аудіо, що робить їх привабливими для синтезу мовлення.

Рекурентні нейронні мережі (RNN) є типом нейронних мереж, призначених для роботи з послідовними даними. Основна сутність полягає в здатності зберігати попередні стани та використовувати їх для обробки нових вхідних даних у вигляді послідовностей.

Фрагмент нейронної мережі  $A$  отримує вхідне значення  $x_t$  і генерує вихідне значення  $h_t$  (рис. 1.2). Існування зворотного зв'язку дозволяє передавати інформацію від одного кроку навчання мережі до наступного. Це забезпечує можливість враховувати попередні результати при роботі з новими даними або кроками.

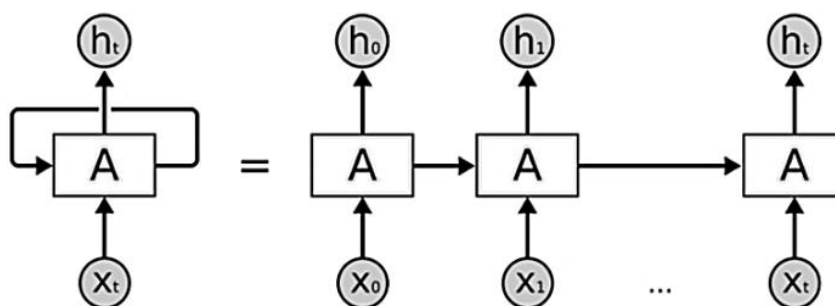


Рисунок 1.2 – Рекурентна мережа у розгортці [17]

У контексті синтезу голосу, RNN можуть бути використані для моделювання залежностей між аудіофрагментами, що служить для створення більш послідовних та природніх звукових послідовностей, які наближені до людського мовлення.

Одними з найпоширеніших варіантів RNN є LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Unit). Вони вирішують проблему втрати інформації, що виникає при тривалій обробці послідовностей, дозволяючи мережам зберігати та використовувати інформацію з попередніх кроків у часі для кращого управління майбутніми даними. LSTM використовує три вентиля: вентиль оновлення, вентиль забуття і вентиль виходу. Ці елементи керують потоком інформації, необхідної для передбачення виходу мережі. Вентиль оновлення визначає, чи потрібно передавати попередній стан наступному елементу рекурентної мережі. Інші вентиля виконують додаткові математичні операції на вхідних даних. Отже, в цілому LSTM використовує дві операції з двома новими наборами ваг, що дозволяє краще моделювати та аналізувати довгострокові залежності в даних. (рис. 1.3).

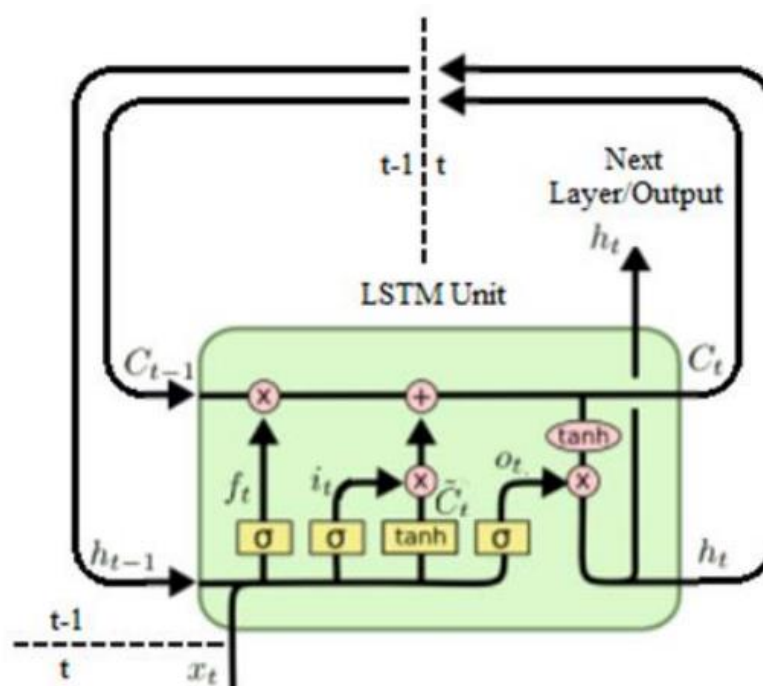


Рисунок 1.3 – . Схема архітектури LSTM [18]

GRU, або вентиляний рекурентний вузол, є варіантом RNN, який подібний до LSTM. Він також призначений для вирішення проблеми короткочасної пам'яті у

моделей RNN. У відмінність від LSTM, GRU використовує лише два вентиля - вентиль оновлення та вентиль забування, замість трьох у LSTM (рис. 1.4).

Вентиль оновлення вирішує, чи нести попередній стан до наступного елементу рекурентної мережі. Це дозволяє контролювати, скільки та яку інформацію зберігати у внутрішньому стані моделі. Вентиль забування використовує додаткові математичні операції з новим набором ваг, щоб визначити, яка інформація повинна бути забута або проігнорована. Ці елементи управляють тим, яка інформація зберігається та оновлюється у внутрішньому стані мережі, що сприяє її здатності пам'ятати та аналізувати довгострокові залежності в даних.

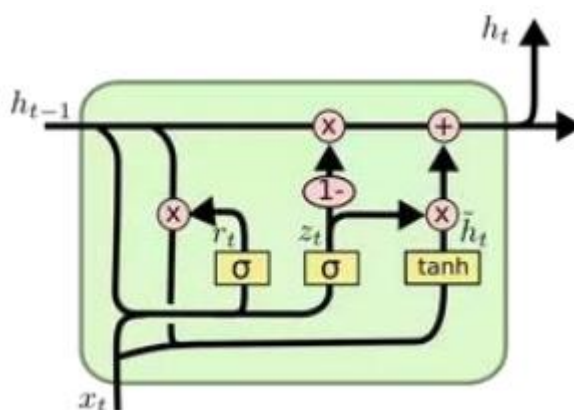


Рисунок 1.4 – Схема архітектури GRU [19]

Застосування RNN у створенні голосу може включати моделювання мелодій, акустичних особливостей та ритму мовлення. Це дозволяє створювати більш природні звукові фрагменти, які мають більшу внутрішню логіку та зв'язок між собою.

Проте, у випадку довгих послідовностей, RNN можуть стикатися з проблемою зниклого градієнту, коли важко зберігати корисну інформацію на довгий період часу через обмежену пам'ять моделі. Це спонукало до розвитку більш складних архітектур, таких як Transformer, які можуть ефективно працювати з послідовностями та зберігати довготривалі залежності.

Загалом, RNN та їхні варіації є потужними інструментами у створенні голосу через нейромережі, здатними враховувати послідовність даних та створювати більш природні та звуково узгоджені аудіозаписи.

Застосування конволюційних нейронних мереж (CNN) також може бути важливим у визначенні структури голосового сигналу та його подальшому створенні. CNN використовуються для роботи з аудіофайлами, аналізу спектрограм та інших характеристик звуку, що допомагає у створенні більш реалістичного звучання.

Конволюційні нейронні мережі (Convolutional Neural Networks, CNN) - це тип нейромережі, який широко використовується для обробки зображень, але також знаходить застосування у сферах обробки аудіо та сигналів, включаючи голосові дані.

Архітектура CNN (рис.1.5) відтворює організацію зв'язків між нейронами у людському мозку, яка була натхненна структурою зорової кори. У зоровій корі окремі нейрони реагують лише на обмежену область в полі зору. Кілька таких областей перекриваються, щоб покрити всю візуальну область. Таке розподілення структур у зоровій корі послужило взірцем для архітектури з конволюційними шарами у нейронних мережах, що спрямовані на обробку візуальних даних.

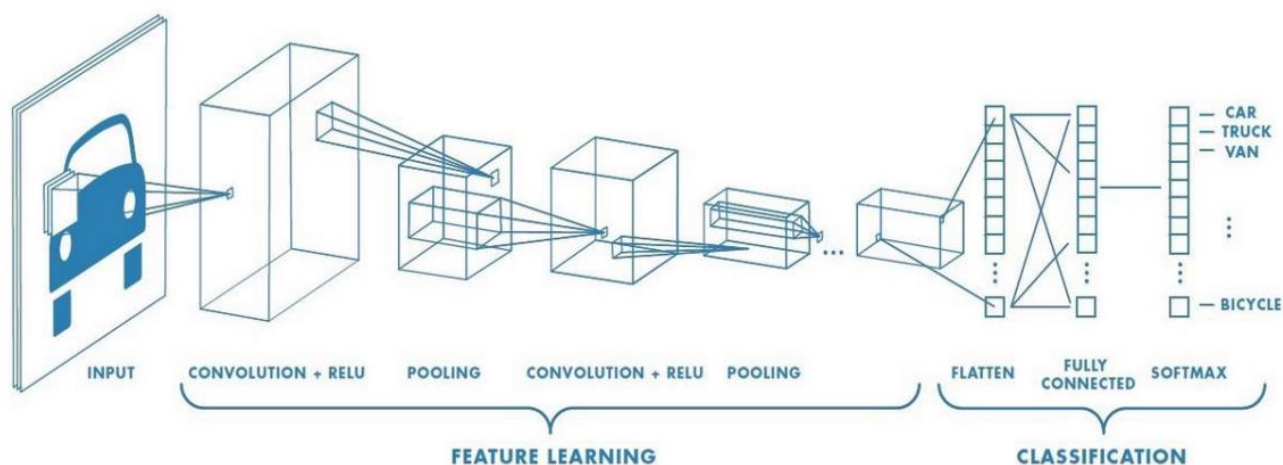


Рисунок 1.5 – Архітектура конволюційних нейронних мереж [20]

У випадку створення голосу за допомогою CNN, ці мережі можуть застосовуватися для аналізу звукових характеристик голосу, таких як частоти, спектральні особливості, ритм, тембр, та інші акустичні параметри. CNN можуть впізнавати шаблони та особливості у звукових сигналах, що дозволяє їм адаптуватися до різних аспектів мовлення та звукової інформації.

Однією з ключових переваг CNN є виявлення локальних особливостей у вхідних даних, що дає змогу виділяти певні аспекти звукового сигналу. Це може бути корисним у визначенні різних аспектів голосу, таких як інтонація, артикуляція, частотні характеристики, та інші особливості мовлення.

Використання CNN для обробки аудіоданих може включати застосування фільтрів, виявлення акустичних шаблонів, а також використання глибокого навчання для автоматизованого виявлення та аналізу акустичних особливостей в голосових сигналах.

Застосування CNN у сфері синтезу голосу дозволяє використовувати їхню потужність у виявленні структурних характеристик аудіо, що сприяє покращенню якості та точності створених звукових даних.

Окрім архітектур, важливим етапом у створенні нейромереж для синтезу голосу є використання навчальних даних. Ці дані є основою для навчання моделі розпізнавати та відтворювати різні аспекти голосу, такі як тон, інтонація, швидкість мовлення та артикуляція.

Навчальні дані для генерації голосу можуть включати аудіозаписи різних людей, що мають різні стилі мовлення, акценти, вікові та гендерні особливості. Це дозволяє моделі засвоювати широкий спектр варіацій у мовленні та робить її більш адаптивною та універсальною.

Ретельний відбір навчальних даних є важливим, оскільки якість та репрезентативність цих даних напряму впливає на якість синтезованого голосу. Чим більше та різноманітніше дані, тим більш точно модель може відтворювати різні особливості мовлення.

Процес підготовки навчальних даних також може включати передобробку аудіозаписів, таку як фільтрація шумів, нормалізація гучності, видалення артефактів чи обрізка сегментів для досягнення більшої чіткості та стабільності в навчальних даних.

Отримані дані використовуються для навчання моделі, де вона вчиться відтворювати голос на основі цих зразків. Це навчання може включати ітеративний

процес, де модель поступово покращується, адаптуючись до нових даних та оптимізуючи свої параметри для досягнення кращої продуктивності.

Використання відповідно підготовлених та репрезентативних навчальних даних є ключовим фактором успішності нейромереж у сфері синтезу голосу, оскільки це визначає їхню здатність точно відтворювати та адаптуватися до різних нюансів мовлення.

Особливо важливим в контексті синтезу голосу є оцінка якості створеного звуку. Метрики, такі як чистота звуку, правдоподібність та природність голосу, є ключовими показниками для оцінки ефективності та реалістичності синтезованого голосу.

Нейромережі навчаються на великій кількості аудіозаписів, а потім використовують ці дані для створення нових голосів, які можуть бути використані у різних цілях, від синтезу мовлення для людей з порушеннями до використання в аудіокнигах, рекламних матеріалах та інших цифрових медіа.

### 1.1.1 Переробка голосу за допомогою нейромереж

Переробка голосу за допомогою нейромереж включає в себе широкий спектр методів та технік, які дозволяють змінювати аудіосигнал у різних аспектах. Основна ідея полягає в тому, щоб застосовувати нейромережі для обробки голосу з метою поліпшення, зміни чи трансформації його характеристик.

- Шумопригнічення. Одним з типових завдань у переробці голосу є видалення шуму або мінімізація його впливу на аудіосигнал. Нейромережі можуть бути навчені розпізнавати шум у вхідних аудіо та відфільтровувати його, поліпшуючи чистоту та якість голосу;

- Еквалізація та покращення якості звуку. Нейромережі можуть застосовуватися для зміни тональності голосу, підвищення чи зниження певних частот або покращення загальної якості аудіосигналу;

- Зміна тембру чи голосу. Іншим аспектом переробки голосу є зміна його характеристик, таких як тембр, стиль мовлення, швидкість мовлення тощо. Це може бути корисним для створення різноманітних голосів чи ефектів у відтворенні аудіо;

- Аудіоаналіз та синтез. Нейромережі можуть бути навчені розпізнавати та аналізувати акустичні особливості голосу для подальшого синтезу нових аудіозаписів або редагування на основі цієї аналітики;

- Підвищення якості мовлення. Технології переробки голосу також можуть використовуватися для покращення якості мовлення, зокрема у випадках, коли людина має порушення мовлення чи інші проблеми з артикуляцією.

Навчання нейромереж для переробки голосу вимагає великої кількості аудіоданих, розглядання різноманітних аспектів голосу та адаптацію моделей для досягнення бажаного результату. Технології переробки голосу стають все більш потужними та використовуються у різних сферах, від розваг та медіа до медичних додатків та розвитку штучного інтелекту.

### 1.1.2 Створення голосу через нейромережі

Створення голосу - це складний процес, що використовується для генерації аудіосигналів, які звучать подібно людському мовленню. Нейромережі, що застосовуються в цій сфері, навчаються аналізувати та моделювати різні аспекти мовлення, враховуючи тон, інтонацію, артикуляцію та інші акустичні характеристики.

Процес потребує великої кількості аудіоданих, з яких нейромережі можуть вивчати патерни та особливості голосу. Вони "вчитимуться" створювати нові голосові зразки, які можуть бути використані у різних цілях, від генерації мовлення для різних відомостей до створення персоналізованих голосових асистентів чи емоційних виразів.

Архітектура нейромережі для створення голосу базується на складних математичних операціях, які враховують акустичні особливості голосу та звукові характеристики мовлення. Процес створення голосу за допомогою нейромереж включає кілька ключових етапів:

- Введення аудіоданих: Нейромережа отримує на вхід велику кількість аудіо-записів, що включають різноманітність мовленнєвих особливостей та характеристик;

- Аналіз акустичних особливостей: Мережа аналізує вхідні дані, розрізняючи різні аспекти мовлення, такі як інтонація, тембр голосу, частотні характеристики та інші акустичні особливості;

- Вивчення патернів: Нейромережа навчається впізнавати та уявляти патерни мовлення, враховуючи особливості голосу з різних аспектів;

- Генерація аудіосигналів: На основі вивчених патернів мережа генерує нові аудіосигнали, що відтворюють природній голос. Процес включає моделювання акустичних властивостей голосу;

- Оцінка та покращення: Створені аудіосигнали оцінюються за різними метриками, такими як якість звуку, природність мовлення, чіткість тощо. Модель може піддаватися покращенню за допомогою зміни параметрів, що дозволяє поліпшувати результати генерації голосу.

Це лише загальний опис процесу. Фактичний робочий механізм нейромережі для створення голосу ґрунтується на використанні різноманітних шарів нейронів, які обробляють вхідні аудіодані. Вхідні дані проходять через мережу, де кожен шар нейронів проводить обчислення з урахуванням попередньо встановлених ваг та параметрів.

Кожен нейрон у мережі приймає вхідні дані, проводить розрахунки на основі вагових коефіцієнтів, застосовує функцію активації і передає свій вихід до наступного шару нейронів. Процес повторюється через кілька шарів, де кожен наступний шар нейронів отримує вже більш абстрактні та складні функції від попередніх.

У кінцевому шарі мережі генерується вихідний сигнал, що є результатом обчислень усіх попередніх шарів. Сигнал може бути проаналізований для оцінки природності та якості мовлення, а сама мережа може піддаватися корекції чи доналаштуванню для поліпшення результатів генерації голосу. Технологія застосовується у багатьох сферах, включаючи розробку аудіокниг, відеоігор, додатків для людей з порушеннями мовлення, аудіореklamних матеріалів та інших мультимедійних

проектів. Її важливість полягає в можливості створення реалістичного та якісного голосу, що максимально наближений до природного мовлення.

### 1.1.3 Покращення голосу за допомогою нейромереж

Покращення голосу за допомогою нейромереж - це процес вдосконалення якості аудіосигналу для досягнення більш природного, чіткого та якісного звучання голосу. Це може включати усунення шуму, підвищення чіткості мовлення, роботу з резонансами або підвищення природності звучання.

Нейромережі використовуються для аналізу аудіосигналів та виявлення акустичних дефектів, таких як шум, ехо чи інші артефакти, що можуть знижувати якість звучання. Вони можуть бути навчені розпізнавати ці артефакти та автоматично коригувати їх, поліпшуючи якість голосу в аудіосигналі.

Процес може також включати апаратні або програмні засоби для обробки голосу, що допомагають покращити якість мовлення. Наприклад, фільтри для підвищення чіткості голосу, методи регулювання гучності або техніки для видалення побічного шуму.

Крім того, покращення голосу може включати і психоакустичні аспекти, оскільки іноді не тільки сам звук важливий, але й сприйняття цього звуку людиною. Такі фактори, як співвідношення сигнал-шум, рівень чіткості голосу та його природність, враховуються під час оптимізації алгоритмів покращення голосу за допомогою нейромереж.

Процес покращення голосу за допомогою нейромереж є важливим у розвитку систем голосового управління, технологій синтезу мови та покращення якості звукового вмісту в різних сферах, від розваг до комунікації та медицини.

## 1.2 Створення музики за допомогою нейромереж

Створення музики за допомогою нейромереж - це захопливий процес, де штучний інтелект використовується для генерації нових музичних композицій або звукових секцій. Нейромережі в цій області використовуються для аналізування, узагальнення та створення нових музичних образів.

Одним з методів є використання глибоких нейронних мереж для навчання на великій кількості існуючих музичних композицій.

Коли нейромережі навчаються на великій кількості музичних композицій, вони аналізують їх структури, мелодії, гармонії та ритми. Глибокі нейронні мережі можуть розрізняти різні елементи музики і вивчати, як вони пов'язані між собою у контексті різних жанрів та виконавців. Це включає в себе розпізнавання основних музичних патернів, взаємозв'язків між нотами, акордами, темпом, структурою композицій та іншими аспектами.

Після вивчення цих елементів нейромережа може використовувати свої знання для створення нових музичних сегментів або композицій, які відповідають зразкам, які вона вивчила під час навчання. Вона може генерувати нову музику, враховуючи стилі, жанри та особливості, які були характерні для вихідних композицій.

Це дозволяє не лише відтворювати існуючі стилі, а й експериментувати з їх поєднанням або створювати нові, унікальні музичні композиції, що базуються на тому, що було вивчено в процесі навчання.

Інші підходи включають у себе генерацію нової музики з нуля на основі певних стилів або відомих музичних шаблонів, що використовують мережі, щоб вони розуміли та відтворювали характеристики цих стилів.

Підхід до створення музики через нейромережі є захоплюючим, оскільки мережі навчаються не тільки відтворювати існуючі музичні шаблони, але й генерувати нову музику з нуля, засновану на вивчених характеристиках стилю.

Мережі можуть бути навчені розуміти особливості різних музичних стилів: вони аналізують ритм, характерні акорди, мелодійні лінії, темп, динаміку та інші

характеристики. Потім вони використовують ці знання для того, щоб створювати нові музичні фрагменти або композиції, які відповідають цим стилям або враховують їхні особливості.

Підхід дозволяє експериментувати зі створенням нових музичних творів, які мають певний стиль, але в той же час є унікальними і відображають творчий потенціал мережі. Вона може генерувати музику, що відповідає певному жанру або стилю, враховуючи властивості цього стилю, але зі своєю власною оригінальністю та неповторністю.

Створення музики за допомогою нейромереж включає в себе використання спеціалізованих моделей глибокого навчання, таких як рекурентні нейронні мережі (RNN), генеративно-супервізійні нейронні мережі (GAN) або варіаційні автокодери (VAE).

Рекурентні нейронні мережі (RNN) - це тип нейромереж, які працюють з послідовностями даних, такими як текст, звук чи музика. Як було сказано вище, RNN використовують свою структуру для зберігання інформації з попередніх кроків у вигляді "пам'яті", що дозволяє їм аналізувати послідовності та створювати прогнози на основі цієї інформації. У музичному контексті вони можуть вивчати музичні структури, ритми та мелодії.

Генеративно-супервізійні нейронні мережі (GAN) - як було сказано вище, це дві нейромережі: генеративна та дискримінативна, які працюють в парі. Генеративна мережа створює нові зразки, у нашому випадку, музичні фрагменти, а дискримінативна оцінює, наскільки вони схожі на реальні. Процес тренування дозволяє генеративній мережі створювати все більш реалістичні результати.

Варіаційні автокодери (VAE) - це тип глибоких нейронних мереж, які використовуються для генерації нових даних на основі вхідного набору даних, при цьому вони стараються зберегти схожість з вихідними даними, але в одночасно створюють нові, варіативні екземпляри.

Основна ідея VAE полягає в тому, щоб навчити нейромережу кодувати вхідні дані в латентний простір, де кожен елемент цього простору представляє собою різні

характеристики даних. Він навчається розуміти розподіл даних у цьому просторі складних параметрів (рис.1.6).

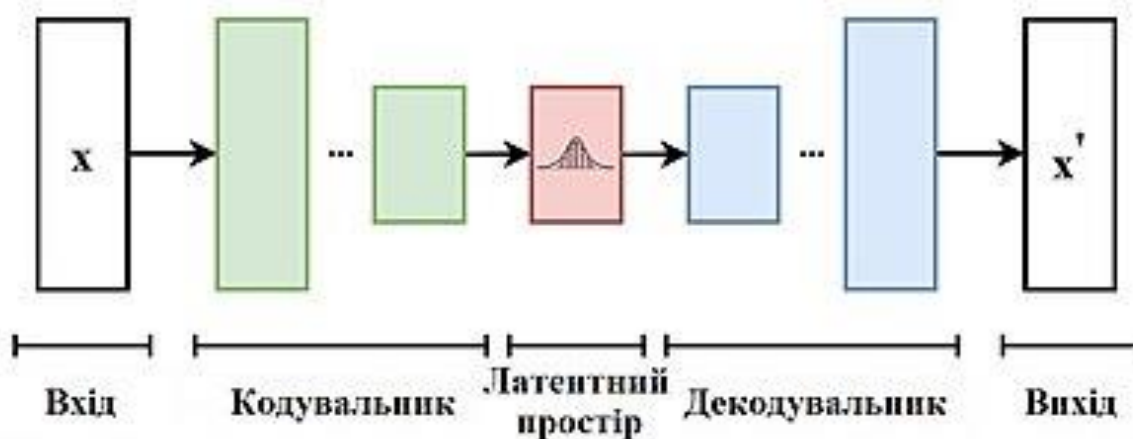


Рисунок 1.6 – Базова схема варіаційного автокодувальника [21]

У випадку музичних даних, VAE може аналізувати музичні патерни, ритми, мелодії та інші характеристики. Після навчання мережі можна використовувати для генерації нових музичних сегментів або композицій, що базуються на вивчених характеристиках.

Однією з особливостей VAE є їх здатність генерувати різні варіанти нових даних з деякою ступенем варіативності. Це означає, що при кожній генерації можуть створюватися нові, але схожі на вихідні дані екземпляри, що робить їх корисними для творчості та експериментів у музичному та інших творчих напрямках.

Всі неймережі, які були приведені вище, мають свої унікальні особливості та застосування у створенні музики, допомагаючи системі розуміти та генерувати музичні шаблони, роблячи їх більш природними та музикально виразними. Вони можуть працювати з різними форматами музичних даних, такими як MIDI, аудіофайли або спектрограми. Неймережі навчаються на великій кількості музичних записів, аналізуючи їхні патерни та характеристики.

Глибокі нейронні мережі можуть вивчати музичні структури, акустичні особливості та гармонійні зв'язки. Зазвичай, вони здатні генерувати нову музику шляхом передбачення наступного музичного фрагмента на основі попередніх даних.

Технічно, під час навчання моделей музичні дані можуть бути перетворені у числові вектори, що дозволяє нейромережам аналізувати та генерувати музику у форматі, який їм зрозумілий та легкий для обробки.

Крім того, деякі алгоритми можуть використовувати спеціалізовані функції втрат (loss functions) для відтворення музичних патернів та гармоній з точністю, наближеною до оригіналу, з метою генерації більш природної та музикальної схожості в нових композиціях, створених нейромережею.

Ці техніки не лише допомагають у створенні нових мелодій, але й можуть сприяти в пошуку нових музичних напрямків або вдосконаленні існуючих, надаючи композиторам та артистам інструменти для творчості та експериментацій.

### 1.3 Нейромережі в процесі підбору кращих кадрів

У процесі підбору кращих кадрів за допомогою нейромереж використовуються різноманітні техніки комп'ютерного зору та машинного навчання. Основна мета полягає в тому, щоб система автоматично вибирала кращі кадри з великої кількості відеоматеріалу.

Нейромережі можуть бути навчені розпізнавати та виділяти візуальні особливості кадрів, такі як яскравість, колір, контрастність, композиція, рух тощо. Вони аналізують кожен кадр з відео та приймають рішення щодо того, які з них найкраще відповідають певним критеріям, наприклад, чіткості, естетичному вигляду чи важливості в контексті всього відео.

Це може включати в себе використання різних моделей нейромережі, таких як сверточні нейронні мережі (CNN), які добре підходять для обробки візуальних даних. Вони аналізують піксельні дані кожного кадру та враховують їх особливості для вибору найбільш якісних та релевантних зображень. CNN ефективно застосовуються для аналізу зображень та відео. Вони використовують концепцію фільтрів,

які просуваються по вхідних даних, щоб виділити важливі ознаки. У відеофайлах, CNN можуть аналізувати різні аспекти кадрів: текстуру, колір, контраст, форми об'єктів та їх розташування.

У процесі підбору кращих кадрів, CNN можуть визначати важливість кожного кадру на основі його візуальних характеристик. Вони можуть враховувати такі параметри, як якість, різноманітність кольорів, наявність об'єктів, їх розмір та рух. Нейромережа аналізує всі ці аспекти для визначення того, які кадри є найбільш релевантними, чіткими або естетично приємними.

Це дозволяє автоматично виділяти ключові моменти, які можуть бути корисними в контексті відеоредакції, скорочення часу перегляду або в інших областях, де потрібно швидко знаходити найкращі кадри у великому обсязі відеоматеріалу.

Крім CNN, існують інші моделі, які можуть застосовуватися для аналізу відео: рекурентні нейронні мережі (RNN) та трансформери.

Рекурентні нейронні мережі (RNN) зазвичай використовуються для роботи з послідовностями, але вони також можуть застосовуватися у відеоаналітиці. RNN можуть допомагати враховувати хронологічний контекст кадрів та визначати їхню важливість залежно від того, як вони поєднуються в послідовності.

Трансформери - це архітектура нейронних мереж, яка набула широкого застосування в обробці послідовних даних, таких як текст або відео. Однією з ключових переваг трансформерів є їхній розподіленість та здатність працювати з довгими послідовностями даних. Вони використовують механізми уваги, що дозволяє їм звертати увагу на різні частини послідовності.

У контексті аналізу відео, трансформери можуть аналізувати великі фрагменти відеоматеріалу, розглядаючи їх в контексті один відносно одного. Це дозволяє їм здійснювати аналіз залежностей та відношень між різними частинами відео та виділяти ключові аспекти в рамках цілої послідовності.

Трансформери відносно нові в галузі нейромереж та мають великий потенціал у багатьох областях, включаючи обробку природної мови, зображень та відео. Вони здатні здійснювати високоякісний аналіз великих обсягів даних, що робить їх цінним інструментом для завдань обробки та аналізу відео.

Ці різні моделі можуть використовуватися разом або окремо для аналізу відео та вибору найкращих кадрів в залежності від потреб користувача або конкретного завдання. Вони сприяють автоматизації процесу обробки великих обсягів відеоданих та швидкої ідентифікації важливих моментів у відеофайлах.

Це стає особливо корисним при автоматичній обробці великої кількості відеоматеріалу, допомагаючи швидше знаходити найкращі моменти, що може використовуватися в редакторських процесах, рекламі, аналізі відео та багатьох інших сферах.

#### 1.4 Монтаж відео з використанням нейромереж

Коли говориться про монтаж відео з використанням нейромереж, це передбачає застосування штучного інтелекту та конкретних алгоритмів для покращення та оптимізації процесу монтажу.

Нейромережі можуть використовуватись для автоматизації певних завдань в процесі монтажу відео, таких як: розпізнавання об'єктів на кадрах, автоматичне позиціонування елементів на відеокадрах, колірна та експозиційна корекція, автоматизація обрізання та зміни ракурсу, видалення шуму та покращення якості відео.

##### 1.4.1 Розпізнавання об'єктів

Розпізнавання об'єктів на кадрах відео є ключовою функцією для полегшення процесу монтажу. Нейромережі, зокрема сверточні нейронні мережі (CNN), здатні автоматично визначати різні об'єкти на відеокадрах.

CNN аналізують кожен кадр у відео та розпізнають на них об'єкти або області інтересу. Це можуть бути обличчя людей, транспортні засоби, будівлі, тварини або будь-які інші речі, що присутні на кадрі. Процес розпізнавання допомагає з автоматичної ідентифікації та маркування об'єктів.

Це важливо для редакторів відео, оскільки вони можуть використовувати цю інформацію для швидкого доступу до необхідних елементів для монтажу. Наприклад, якщо потрібно зібрати послідовність кадрів з певним об'єктом або певними характеристиками, система розпізнавання допоможе знайти відповідні кадри швидше, зменшуючи час, який витрачається на пошук необхідних об'єктів серед великої кількості відеоданих.

#### 1.4.2 Автоматичне позиціонування елементів

Автоматичне позиціонування елементів у відео - це процес, коли нейромережі аналізують кожен кадр та визначають оптимальне місце для розташування певних елементів, таких як текстові блоки або графіка.

Це може включати в себе різні аспекти:

- Розпізнавання контексту: Нейромережі аналізують зміст кадру та його характеристики, щоб визначити, де розмістити певні елементи. Наприклад, вони можуть враховувати композицію кадру, щоб розмістити текст так, щоб він був чітко видимим та не перекривав важливі об'єкти;

- Оптимізація відповідно до контенту: Враховуючи зміст кадру, алгоритми можуть встановити оптимальні параметри для позиціонування тексту чи графіки. Наприклад, вони можуть вибирати кольори тексту, щоб вони контрастували з фоном, що поліпшує читабельність;

- Адаптація до різних форматів відео: Нейромережі можуть також враховувати різні аспекти відеоконтенту, наприклад, відмінності в освітленні, рух камери, що дозволяє автоматично підлаштовувати позицію тексту або графіки для максимальної зручності.

Для перших двох завдань можуть використовуватися CNN або RNN, які навчаються розпізнавати характеристики кадрів та їхній контекст, допомагаючи розмістити елементи відео відповідно до їхнього змісту та композиції. Щодо адаптації

до різних форматів відео: тут CNN часто використовують для аналізу різних аспектів відеоконтенту, таких як освітлення, кут камери тощо. Це дозволяє адаптувати позицію тексту або графіки для оптимального відображення у різних умовах.

Зокрема, для розпізнавання об'єктів, що може бути важливим для розміщення елементів на кадрі, використовуються сверточні нейронні мережі (CNN), тоді як для роботи з послідовністю даних, які можуть впливати на позиціонування (таких як контекст чи композиція), використовуються рекурентні нейронні мережі (RNN).

У кожному конкретному випадку моделі можуть бути адаптовані або комбіновані для досягнення оптимальних результатів залежно від вимог завдання та особливостей відеоданих.

В результаті Процес дозволяє ефективно та автоматично розміщувати елементи на кадрах відео, що зменшує час ручної обробки та поліпшує загальний вигляд та сприйняття відеоматеріалу.

#### 1.4.3 Колірна та експозиційна корекція

Колірна та експозиційна корекція є ключовими аспектами в обробці відео, оскільки вони можуть значно вплинути на його вигляд та сприйняття глядачем. Нейромережі допомагають автоматизувати ці процеси, забезпечуючи кращу якість та візуальний зовнішній вигляд відеоматеріалу.

Для корекції кольорів можуть використовуватися глибокі нейронні мережі, зокрема CNN або GAN. Вони навчаються аналізувати та змінювати колірний баланс, насиченість, тон та інші параметри кольорів для поліпшення загального вигляду відео.

Нейромережі також допомагають у виправленні експозиції, тобто управлінні яскравістю, контрастом та освітленням на кожному кадрі. Це може бути особливо корисним для відео, де освітлення може змінюватися в залежності від умов зйомки.

Для цих завдань можуть використовуватися різні моделі нейромереж, такі як GAN для глибокої корекції кольорів та параметрів освітлення, або CNN для більш точної роботи зі зміною окремих кольорів чи тонів.

Ці підходи дозволяють автоматизувати процес корекції кольорів та експозиції, забезпечуючи відео більш якісним та зручним для подальшого використання чи публікації.

#### 1.4.4 Автоматизація обрізання та зміни ракурсу

Автоматизація обрізання та зміни ракурсу відео - це процес, коли нейронетичні мережі аналізують кожен кадр та вирізають або трансформують його для покращення композиції відео або видалення зайвих частин.

Обрізання кадрів може бути виконана за допомогою сегментації об'єктів та виявленням ключових елементів. Мережі можуть використовувати алгоритми сегментації для виділення конкретних об'єктів або регіонів у кадрі. Після виділення ці області можуть бути вирізані або змінені. U-Net та Mask R-CNN: ці архітектури часто використовуються для сегментації об'єктів у візуальних даних. U-Net добре підходить для вирізання об'єктів на основі їхньої структури, тоді як Mask R-CNN використовується для точного виділення областей з об'єктами у великих зображеннях.

U-Net має архітектуру енкодер-декодер, яка дозволяє ефективно сегментувати об'єкти на зображеннях. Вона складається з двох основних частин: енкодера, який зменшує розмір зображення та виділяє його ключові функції, і декодера, який відновлює зменшене зображення до його початкового розміру, в той же час вирізняючи об'єкти. Архітектура добре працює для завдань сегментації об'єктів, особливо коли необхідно знайти об'єкти на основі їхньої структури та контуру.

Архітектура базується на алгоритмі R-CNN (Region-based Convolutional Neural Network) та дозволяє точно виділяти області з об'єктами на великих зображеннях. Mask R-CNN не тільки виділяє об'єкти, але й надає маски, що точно вказують області, де знаходяться об'єкти. Вона використовує шари Region Proposal Network (RPN) для пропозиції областей, де можуть знаходитися об'єкти, та використовує маски для точного виділення цих областей (рис. 1.7).

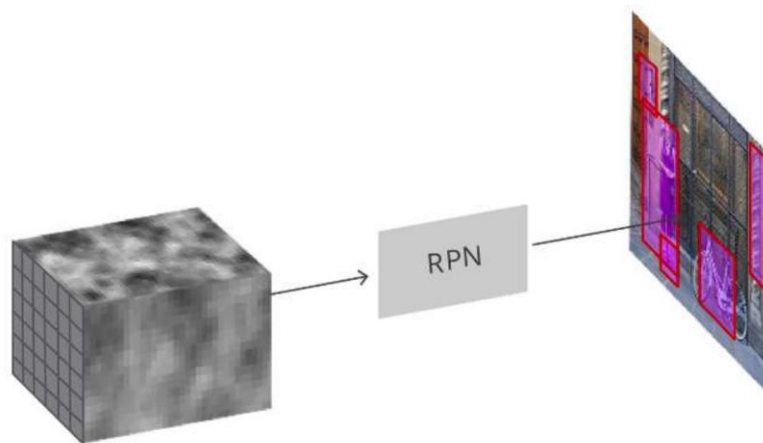


Рисунок 1.7 – RPN використовує згортковий набір ознак і генерує якоря на зображенні [22]

Інші алгоритми можуть аналізувати кадр для виявлення ключових елементів або об'єктів, що несуть значення, та автоматично обрізати кадр навколо них для поліпшення візуального ефекту або збільшення уваги до цих елементів. Алгоритми детекції об'єктів (YOLO, SSD) спрямовані на знаходження та обмеження прямокутниками (bounding boxes) об'єктів у кадрі. Вони можуть використовуватися для виявлення та обрізання кадрів навколо ключових об'єктів, підкреслення їхньої важливості та забезпечення уваги глядача на ці елементи.

Зміна ракурсу створюється за рахунок реконструкції сцени. Це включає в себе алгоритми, які можуть переформатувати ракурс кадрів, створюючи панорамні зображення або змінюючи перспективу, щоб покращити загальний вигляд відео або звернути увагу на певні аспекти сцени. Глибокі архітектури, основані на GAN: Мережі, які використовують глибокі генеративно-протилежні мережі (GAN), можуть перетворювати один ракурс відео на інший. Ці моделі навчаються реконструювати відео з різних ракурсів або перспектив, створюючи нові кадри, які можуть покращити візуальний ефект або змінити перспективу на сцені.

Ці процеси допомагають автоматизувати оптимізацію композиції та забезпечують більш естетичний вигляд відеоматеріалу, спрямований на визначення та видалення зайвих або менш важливих елементів у кадрі, а також на поліпшення сприйняття глядачем.

#### 1.4.5 Автоматичне видалення шуму та відновлення якості

Виправлення акустичних артефактів, таких як ехо та реверберація, важливе для поліпшення якості звуку у відео та аудіоматеріалах. Ехо виникає через відбивання звуку від поверхонь, що може створювати неприємний та затруднюючий сприйняття ефект. Реверберація - це певний тип відлуння, коли звук відображається від різних поверхонь та випромінюється нерівномірно.

Існують різні методи та нейромережеві підходи для усунення ехо та реверберації: акустичне моделювання, автоенкодер, CNN та LSTM мережі

Нейронні мережі можуть використовуватись для акустичного моделювання, що дозволяє передбачити, як звук буде відбиватися від різних поверхонь та середовищ.

Деякі моделі можуть використовувати фізичні принципи, щоб моделювати, як звук відбивається від різних поверхонь та середовищ. Це може включати розрахунки часу затримки, амплітуду відбитого сигналу та інші параметри. Інші моделі можуть використовувати математичні функції передачі для врахування властивостей поверхні та акустичного середовища для передбачення, як звук буде відбиватися від різних структур.

Автоенкодери можуть бути використані для видалення реверберації шляхом відтворення оригінального звуку без непотрібних ефектів, які виникають через реверберацію. Автоенкодери - це нейронні мережі, які навчаються репрезентувати вхідні дані у вигляді більш компактного коду. Для видалення реверберації, автоенкодер може бути навчений реконструювати вхідний звуковий сигнал без цього ефекту, тобто видаляти артефакти та відновлювати оригінальний сигнал.

CNN та LSTM мережі: ці моделі можуть бути використані для розпізнавання та фільтрації акустичних сигналів, спрямованих на видалення артефактів, таких як ехо та реверберація.

Сверточні нейронні мережі (CNN) можуть аналізувати акустичні спектрограми, використовуючи 2D-зображення для розпізнавання та фільтрації акустичних сигналів. Вони можуть виявляти шаблони або особливості, які можуть бути пов'язані з ехо чи реверберацією та піддаються подальшій обробці для їх вилучення.

Long Short-Term Memory (LSTM): ці рекурентні нейронні мережі ефективно працюють з послідовностями даних, такими як аудіо, пам'ятаючи попередні відомості про аудіосигнал. Вони можуть бути використані для аналізу та модифікації аудіосигналу з метою вилучення артефактів, таких як ехо та реверберація.

Всі ці методи допомагають зменшити ефекти ехо та реверберації, забезпечуючи чистіший та більш природний звуковий сигнал. Однак, враховуйте, що це складні завдання, і часом комбінування різних підходів може бути ефективнішим для досягнення високої якості звуку.

Висновок по розділу: Розділ, присвячений принципам функціонування нейромереж у сфері обробки мультимедійного контенту, визначає та аналізує способи, за якими нейромережі використовуються у процесах створення, покращення та обробки фото-, відео-, та аудіоматеріалів. Досліджуючи ці аспекти, отримано глибше розуміння того, як саме нейромережі реалізують важливі завдання у цій сфері та як їх можна застосовувати для досягнення найкращих результатів у мультимедійній обробці. Подальший аналіз вказаних аспектів, таких як створення голосу, музичне творчість, вибір кращих кадрів та монтаж відео з використанням нейромереж, вказує на великий потенціал цих технологій у сучасних мультимедійних застосуваннях. Це свідчить про перспективи та актуальність використання нейромереж для подальшого розвитку та вдосконалення області мультимедійної обробки.

## 2 НЕЙРОМЕРЕЖІ ДЛЯ СТВОРЕННЯ МУЛЬТИМЕДІЙНОГО КОНТЕНТУ: ОГЛЯД ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ТА ПРАКТИЧНІ АСПЕКТИ

У сучасному цифровому світі мультимедійний контент стає все більш важливим та впливовим у візуальному та аудіо сприйнятті інформації. Завдяки стрімкому розвитку технологій нейромереж, нові можливості відкриваються для оптимізації та покращення якості цього контенту. Аналіз та використання нейромереж у голосовій обробці стають ключовими в контексті створення ідеального аудіо. Отримання високоякісної музичної інтерпретації та обробка музичного контенту за допомогою технологій нейромереж посилюють музичний ефект, роблячи його більш інтенсивним та емоційним. Покращення якості відео, зокрема його чіткості та зображення, завдяки впровадженню нейромереж та інших передових технологій, стає необхідним етапом для візуальної глибини та зручності сприйняття інформації. Оцінка ефективності технологій нейромереж у покращенні якості голосу відкриває нові можливості для оптимізації та удосконалення якості звукових записів, забезпечуючи високу відтворюваність та реалістичність голосових даних.

Огляд спрямований на дослідження різноманітних методів використання нейромереж для вищезгаданих цілей, а також на аналіз сучасного програмного забезпечення, що використовується для мультимедійної обробки та творчості.

### 2.1 Використання нейромереж у створенні голосу

У сучасному світі мультимедійний контент стає визначальним у сприйнятті інформації та взаємодії з нею. Завдяки стрімкому розвитку технологій, виникають нові можливості у створенні та оптимізації аудіо- та відеоматеріалів. Одним із ключових етапів цього процесу є обрання підходів до роботи з голосом. Було обране створення голосу з нуля. Це рішення підкреслює необхідність створення унікального звукового вмісту з врахуванням особливостей, властостей та вимог специфічного контенту. Обрання цього підходу дозволяє забезпечити унікальність та високу

ідентифікацію голосового контенту, що є ключовим для підвищення якості та ефективності в мультимедійних проєктах. Вибір також передбачає більший рівень контролю та адаптації параметрів голосу відповідно до потреб в конкретних аудіо- та відеоматеріалах. Такий підхід до створення голосу відкриває нові горизонти для розробки унікальних аудіо та відео зразків з врахуванням потреб користувачів у цифровому просторі.

Таким чином, обране створення голосу з нуля відображає стратегічний вибір, спрямований на створення унікального та персоналізованого контенту, що відповідає вимогам сучасного мультимедійного середовища.

### 2.1.1 Генератор мовлення LOVO (Genny)

LOVO, зокрема інструмент Genny, є платформою для генерації голосу за допомогою штучного інтелекту. Вони спеціалізуються на текст-на-мовлення технологіях, де застосовуються нейромережі та глибоке навчання для створення надзвичайно реалістичних голосів.

LOVO надає можливість користувачам створювати голосовий контент для різних потреб, таких як аудіокниги, рекламні матеріали, аудіо візуалізації та багато іншого. Вони пропонують різні налаштування та варіанти голосів, що дає можливість користувачам вибирати голосові характеристики, які найкраще підходять під їхні потреби (рис. 2.1).

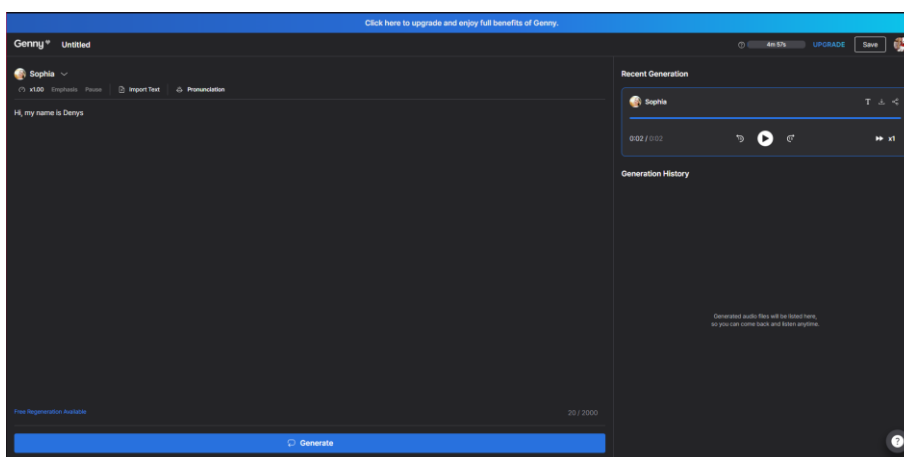


Рисунок 2.1 – Інтерфейс LOVO

LOVO також підтримує багато мов та акцентів, що робить їхню платформу привабливою для глобального ринку. Їхні технології базуються на передових алгоритмах генерації голосу, що робить їхні голоси досить реалістичними та природними у звучанні.

LOVO використовує різні архітектури нейронних мереж для створення своїх голосів. Зазвичай, у їхній технології використовують комбінацію глибоких нейронних мереж для синтезу голосу, таких як рекурентні нейронні мережі (RNN), рекурентні нейронні мережі з довготривалими короткочасними пам'яттями (LSTM), або навіть глибокі згорткові нейронні мережі (CNN) для обробки аудіоданих.

Їхні алгоритми базуються на глибокому навчанні та моделях генерації послідовностей, які навчаються відтворювати природні аспекти мовлення, такі як інтонація, акценти та мелодійність голосу. Однак точні деталі архітектур нейронних мереж, що використовуються в LOVO, можуть бути конфіденційною інформацією компанії.

Переваги використання:

- Реалістичність голосу: Їхня технологія надає надзвичайно реалістичні голоси, які добре підходять для різноманітних аудіо- та відеопроєктів;
- Багатомовність: Підтримка багатьох мов та акцентів, що дозволяє створювати контент для різних аудиторій;
- Налаштування голосу: Є можливість налаштовувати параметри голосу, такі як тон, швидкість, акценти тощо;
- Різноманітність використання: Від аудіокниг до відеопрезентацій, LOVO може бути корисним для різних проєктів.

Недоліки використання:

- Вартість: Зазвичай, подібні платформи вимагають платну підписку або оплату за кожне використання, що може бути витратним. 2-тижнева безкоштовна пробна версія PRO, включаючи їхні бета-версії голосів найвищої якості. Початкова: 4 долари США на місяць за 30 хвилин генерації голосу на місяць Базовий план становить 24 долари США на місяць (при щорічній оплаті) і включає 2 години генерації.

ції голосу на місяць. Професійний план коштує 48 доларів США при оплаті щомісяця або 24 долари США на місяць (зі знижкою 50% при щорічній оплаті) і включає 5 годин генерації голосу разом із бета-версією голосів;

- Неідеальна якість: Хоча голоси дуже реалістичні, можуть бути моменти, коли генерований голос не звучить повністю природно;
- Не всі голоси підтримують налаштування акценту або паузи;
- Клонування голосу підтримує лише англійську мову.

### 2.1.2 Генератор мовлення Play.ht

Play.ht - це сервіс, спрямований на перетворення текстового контенту в аудіоформат. В основному використовується для створення аудіоконтенту для веб-сторінок, блогів, іншого онлайн-контенту (рис. 2.2).

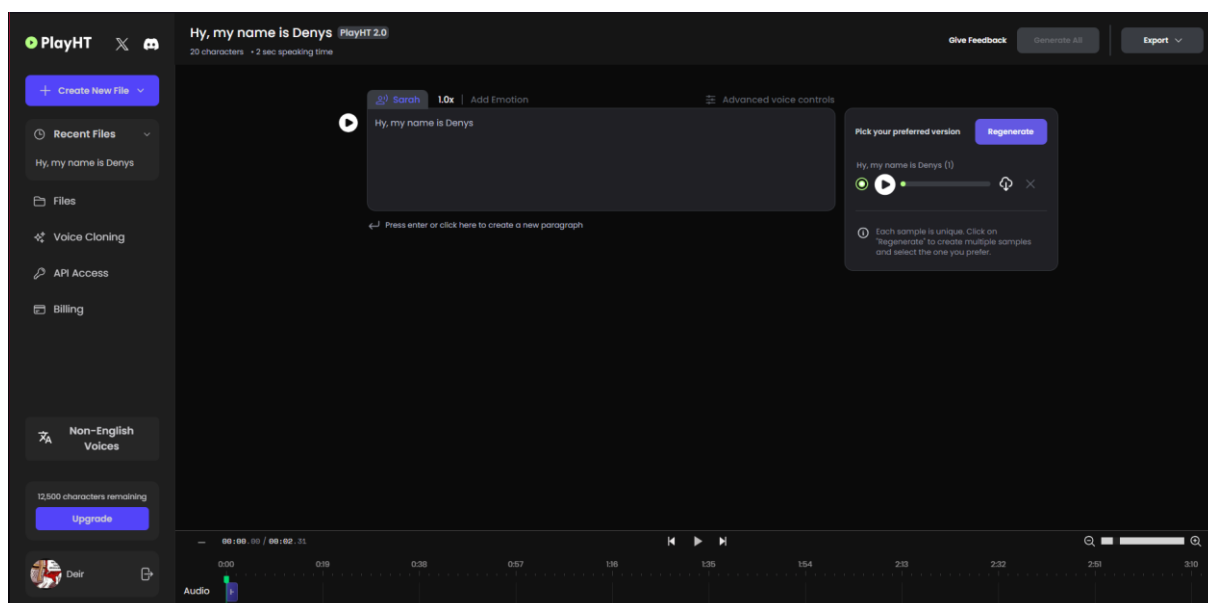


Рисунок 2.2 – Інтерфейс Play.ht

Основні функції Play.ht включають:

- Текст-на-мовлення (TTS): Перетворення тексту у голосовий контент за допомогою технологій синтезу мовлення;

- Інтеграція з веб-сайтами: Можливість інтегрувати аудіофайли безпосередньо на веб-сторінки;

- Настроювання голосу: Вибір різних голосів, тонування та швидкості мовлення для відповідності певним стилям чи потребам контенту;

- Аналітика: Надає звіти та аналітику стосовно використання аудіоконтенту на веб-сторінках.

Сервіс дозволяє створювати аудіоверсії контенту для покращення доступності, аудіотексти для веб-сторінок, а також для розширення аудиторії, яка більше впізнає та засвоює інформацію через аудіоформат.

Play.ht використовує штучний інтелект для синтезу мовлення і ймовірно базується на моделях глибокого навчання, таких як рекурентні нейронні мережі (RNNs) або модифікації, такі як LSTM (Long Short-Term Memory) або GRU (Gated Recurrent Unit), що дозволяє аналізувати та відтворювати мовленнєві послідовності. Однак конкретні деталі про використані моделі нейронних мереж в Play.ht можуть бути конфіденційними та не доступними для публічного огляду.

Отже, ось кілька плюсів та мінусів використання Play.ht для створення голосового контенту:

Переваги використання:

- Легкість використання: Інтуїтивний і простий інтерфейс, що дозволяє швидко створювати аудіоконтент з тексту;

- Широкий вибір голосів: Можливість вибрати із різних голосів та налаштувати їхні параметри (швидкість, тон, акцент тощо) для досягнення бажаного звучання;

- Інтеграція з веб-сайтами: Можливість вбудовувати аудіофайли безпосередньо на веб-сторінки, що робить контент більш доступним для користувачів;

- Аналітика та звіти: Надає звіти та аналітику про використання аудіоконтенту, що може бути корисним для відстеження популярності та ефективності контенту.

Недоліки використання:

- Обмежені можливості налаштування: Може бути обмежена можливість налаштування параметрів голосу порівняно з іншими програмами;
- Обмежена реалістичність голосу: Голосові файли можуть не завжди звучати настільки природно, як хотілося б;
- Враховуючи ці переваги та недоліки, Play.ht може бути корисним інструментом для швидкого створення аудіоконтенту, але його можливості можуть бути обмеженими в порівнянні з іншими програмами з синтезу мовлення.

### 2.1.3 Застосунок Speechify

Speechify - це мобільний застосунок, який дозволяє перетворювати текст у звуковий контент, щоб його можна було прослуховувати на різних пристроях. Основна функція цього застосунку - конвертація текстового контенту у аудіо, щоб забезпечити більший доступ до інформації для користувачів (рис. 2.3).

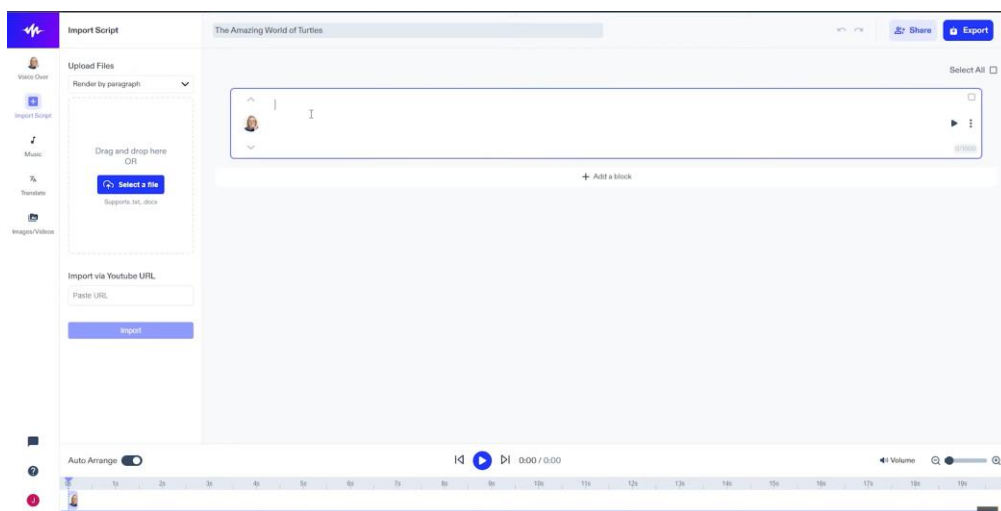


Рисунок 2.3 – Інтерфейс Speechify

Основні можливості Speechify включають:

- Текст-в-мовлення (TTS): Здатність перетворювати текст на звуковий контент, що дозволяє слухати текст у вигляді аудіо;

- Доступність на різних пристроях: Підтримка різних пристроїв, що дозволяє користувачам переключатися між пристроями та продовжувати прослуховування з того місця, де вони зупинилися;

- Синхронізація з іншими платформами: Інтеграція з іншими сервісами або платформами для більшої універсальності та зручності.

Застосунок може бути корисним для людей, які шукають більші можливості доступу до текстового контенту через аудіоформат, щоб прослуховувати текст під час переміщення або в інших ситуаціях, коли читання неможливе або не зручне.

Speechify має свої переваги та недоліки, пов'язані зі створенням голосу:

Переваги використання:

- Доступність: Застосунок є мобільним, що робить його легкодоступним та зручним для використання в рухливих ситуаціях;

- Технологія Текст-в-мовлення (TTS): Speechify використовує технологію TTS, що дозволяє перетворювати текст на звуковий контент;

- Додаткові функції: Деякі версії програми можуть мати функції швидкого збереження та редагування тексту, що робить її більш функціональною.

Недоліки використання:

- Обмежені можливості налаштування голосу: Обмежена можливість налаштування параметрів голосу порівняно з деякими іншими програмами для синтезу мовлення;

- Обмежені функції: У порівнянні з деякими іншими програмами Speechify може мати менше додаткових функцій чи можливостей.

Варто врахувати, що оцінка програми залежить від потреб користувача та конкретного використання. Якщо необхідно просте та зручне перетворення тексту на аудіо, Speechify може відповідати цим потребам.

## 2.1.4 Генератор мовлення Voicemaker

Програма має два двигуни штучного інтелекту: стандартний (якість, що на рівні або навіть краще, ніж у більшості інших програм TTS) і нейронний TTS (ще вища якість). Інтерфейс програми показаний нижче (рис. 2.4).

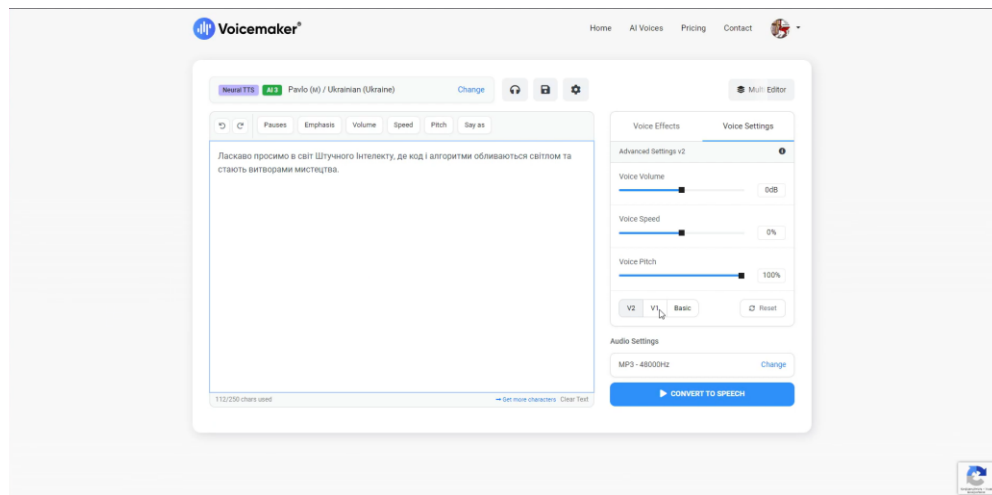


Рисунок 2.4 – Інтерфейс Voicemaker

Двигуни штучного інтелекту відрізняються за своєю архітектурою та методами роботи.

- Стандартний двигун: Використовує класичні методи обробки тексту в мовлення. Він базується на підходах, які використовуються в TTS-системах із загальними правилами та шаблонами для створення звукового відтворення тексту. Результат може бути якісним, але не завжди дуже природним;

- Нейронний двигун: Використовує нейромережі, зокрема глибоке навчання, для генерації мовлення. Нейронні мережі дозволяють системі адаптуватися до контексту та нюансів мовлення, надаючи більш природну та реалістичну звучання. Результатом є мовлення, яке більш природно звучить і має вищу якість, оскільки нейромережі навчаються на великій кількості даних і можуть враховувати більше деталей та контексту.

Також є підтримка SSML (Speech Synthesis Markup Language) - це мова розмітки, яка дозволяє додавати додаткові аудіо ефекти та контролювати параметри

вимови при створенні синтезу мовлення за допомогою програмного забезпечення, такого як текст-в-мовлення (TTS) системи.

Це дає декілька переваг:

- Контроль параметрів вимови: SSML дає можливість точно налаштувати швидкість вимови, паузи, гучність, інтонацію та інші аспекти мовлення;
- Додавання ефектів: Ви можете вставляти ефекти, такі як зміна голосу (чоловічий, жіночий), звукові ефекти (сміх, плач, шепіт), музичні сигнали тощо;
- Контроль нюансів мовлення: SSML дозволяє вам передати паузи для кращого розуміння аудиторії, коригування вимови для складних слів, а також вставляти альтернативний варіант вимови.

Це потужний інструмент для додавання реалізму та природності синтезованому мовленню.

Переваги:

- Можливість налаштовувати паузи, швидкість, гучність;
- Функція вимови дати, часу, дробів тощо (часто неправильно вимовлювана іншими програмами TTS);
- Підтримка SSML;
- 2 двигуни ШІ: стандартний та нейронний;
- Безкоштовний план дає змогу спробувати всі його можливості.

Недоліки:

- Просте інтерфейсне оформлення;
- Ціна.

Voicemaker - це інноваційний інструмент для перетворення тексту в мовлення, що пропонує два типи штучного інтелекту: стандартний та нейронний. Це забезпечує високу якість мовлення, яке звучить природно та чітко. Один з ключових плюсів цієї платформи - широкі налаштування. Ви можете точно контролювати різні параметри вимови, включаючи швидкість, гучність та паузи, для досягнення бажаного звучання мовлення.

Крім того, підтримка SSML робить Voicemaker ще більш привабливим. Мовна розмітка дозволяє додавати ефекти та контролювати нюанси вимови, забезпечуючи більш природне та реалістичне мовлення. Ще однією вагомою перевагою є доступність ціни. Voicemaker пропонує високу якість звуку за доступну ціну, роблячи його привабливим рішенням для широкого кола користувачів.

Отже, у цій кваліфікаційній роботі магістра було обрано Voicemaker, інструмент, що поєднує високу якість мовлення з широким спектром можливостей налаштування та доступною ціною.

## 2.2 Створення та обробка музичного контенту з використанням технологій нейромереж

Штучний інтелект, безсумнівно, відкриває нові можливості для музикантів-любителів, надаючи їм інноваційні методи для покращення творчого процесу. У музичній індустрії, як і в багатьох інших сферах, ШІ використовується як додатковий інструмент, а не заміна людської творчості.

Експерти, дослідники, музиканти та звукорежисери активно розвивають нові шляхи інтеграції штучного інтелекту в музику. Деякі програми розроблені для створення творів у стилі різних композиторів, тоді як інші використовують алгоритми машинного навчання для формування абсолютно нових пісень та звуків.

Ці інструменти також чудові тим, що вони відкриті для широкого загалу. Це означає, що кожен може мати доступ до них та спробувати покращити та розвивати ці технології.

Софти, які створюють музику за допомогою штучного інтелекту, можуть використовувати різні архітектури нейронних мереж. Деякі з них використовують рекурентні нейронні мережі (RNN) або їх варіанти, такі як LSTM (Long Short-Term Memory) або GRU (Gated Recurrent Unit), що дозволяють моделі враховувати контекст та послідовність даних у музичних творах.

Також можуть застосовувати глибокі згорткові нейронні мережі (CNN) для обробки аудіофайлів та витягання характеристик звуків.

Деякі інструменти музичного творчості можуть експериментувати з варіаціями автокодувальних моделей, які вміють відтворювати аудіо чи музичні фрагменти зі своїх внутрішніх репрезентацій.

Конкретний вибір нейронних мереж може залежати від завдань, які вони ставлять перед собою у конкретному софті, а також від унікальних особливостей та власностей, які розробники хочуть вбудувати в свій продукт.

На жаль, детальна інформація про конкретні нейронні мережі, які використовуються в програмах, не є загальнодоступною. Компанії часто не розголошують конкретні технічні деталі своїх нейромереж через комерційні чи конфіденційні причини. Вони можуть використовувати різні архітектури нейронних мереж для генерації музики, проте конкретні дані щодо використання алгоритмів чи моделей можуть бути недоступними для широкої аудиторії.

### 2.2.1 Музичний генератор Amper Music

Amper є одним із найбільш легких у використанні музичних генераторів штучного інтелекту, роблячи його ідеальним вибором для початківців у музичному світі, що бажають працювати з музикою, створеною штучним інтелектом.

Для користування Amper не потрібні глибокі знання музичної теорії чи композиції, оскільки він створює музичні треки з інших зразків. Ці треки можна налаштовувати за бажанням, регулюючи музичні клавіші, темп, окремі інструменти тощо. Наприклад, ви маєте можливість налаштувати весь інструмент відповідно до настрою чи атмосфери, яку хочете створити (рис. 2.5).

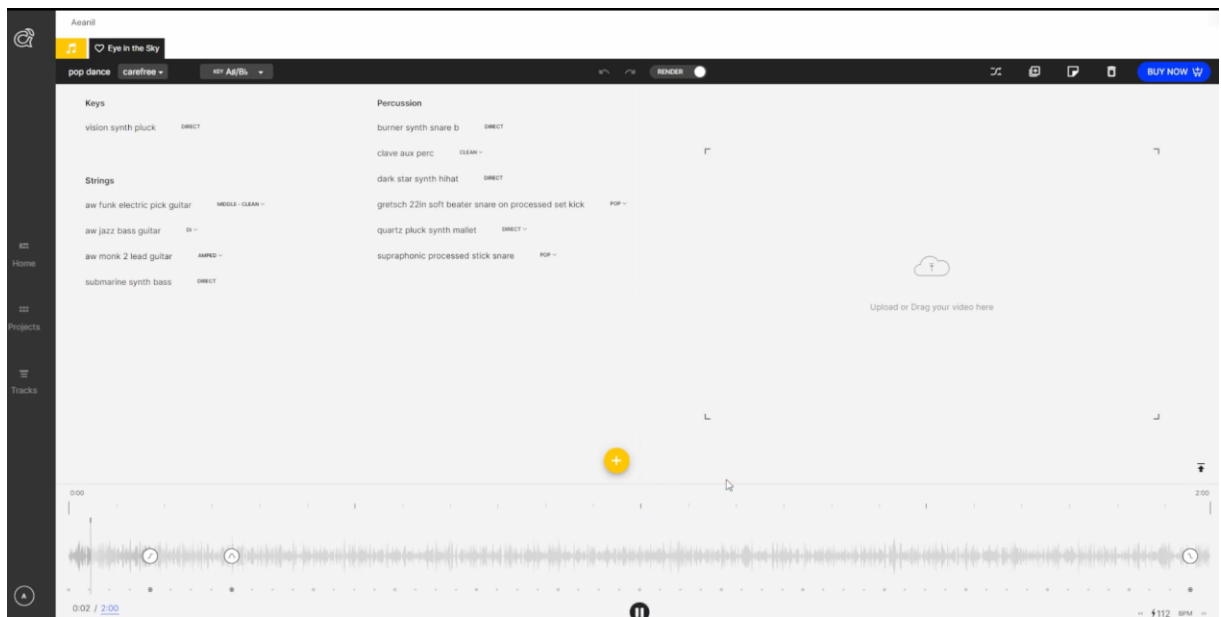


Рисунок 2.5 – Інтерфейс Amper Music

Хмарна платформа є ідеальним вибором для контент-креаторів та індивідуальних осіб, які бажають розробляти саундтреки та звукові ефекти для ігор, фільмів або подкастів. У версії з підпискою на преміум доступ до ще більшого розмаїття можливостей, що доповнюють ваше мистецтво.

Amper Music має кілька ключових переваг:

- Швидке створення музики для різних форматів контенту, таких як подкасти, фільми та відеоігри;
- Розмаїття музичних зразків та інструментів, які створюють безліч можливостей для експериментів;
- Використання спеціальних інструментів для поліпшення творчого процесу у музичному виробництві;
- Хмарна платформа, що дає можливість працювати у віртуальному середовищі з музикою.

## 2.2.2 Музичний генератор AIVA

Один із музичних генераторів, що використовують штучний інтелект, це AIVA (рис. 2.6), випущений у 2016 році. Він неперервно вдосконалюється для створення музичних супроводів для різних цілей, від реклами до відеоігор і фільмів.

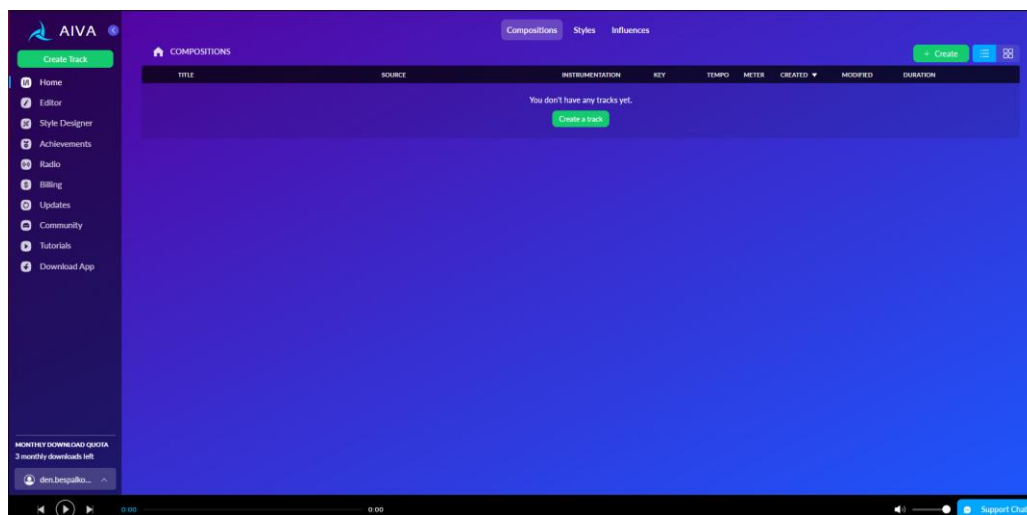


Рисунок 2.6 – Інтерфейс AIVA

Оригінальний альбом, «Opus 1 for Piano Solo», відзначив старт AIVA, представивши його у світ музики. Інструмент дозволяє користувачам створювати музику з самого початку, а також допомагає генерувати варіації існуючих творів, уникнувши проблем з ліцензуванням.

AIVA дозволяє легко створювати музику у різних жанрах та стилях, спочатку вибравши певний стиль за замовчуванням. Він також може внести зміни до поточних музичних творів, розширюючи можливості користувача у творчому процесі.

Ось декілька ключових переваг AIVA:

- Розмаїті пресети і різні музичні формати;
- Функціональна безкоштовна версія, яка дозволяє користувачам випробувати основні функції;
- Можливість редагування звукових доріжок та творчість внесення змін у наявні треки;

- Інтересний результат - музика, створена AIVA, показує потенціал цієї платформи у створенні музичних творів.

### 2.2.3. Музичний генератор Soundful

Сервіс Soundful використовує потужність штучного інтелекту для створення безкоштовної фонові музики одним натисканням кнопки для ваших відео, стрімів, подкастів та інших медійних проєктів. Процес використання платформи інтуїтивно зрозумілий: треба обрати жанр, налаштувати параметри, й отримати унікальний трек.

Унікальність музики гарантується завдяки алгоритмам Soundful, які навчаються нота за нотою разом з провідними продюсерами та звукоінженерами галузі. Використання одноразових семплів у поєднанні з теорією музики гарантує, що Soundful ніколи не відтворить існуючу пісню або свій власний контент. Платформа пропонує користувачам понад 50 шаблонів у різних музичних жанрах для початку.

Soundful - це платформа, яка використовує штучний інтелект для створення унікальної фонові музики для відео, подкастів, стрімів та інших медійних проєктів. І ось як це працює: користувач обирає жанр музики і налаштовує параметри, такі як настрій, темп, інструменти тощо. Потім штучний інтелект Soundful використовує ці дані, щоб згенерувати унікальний трек, який можна використовувати у своїх проєктах (рис. 2.7).

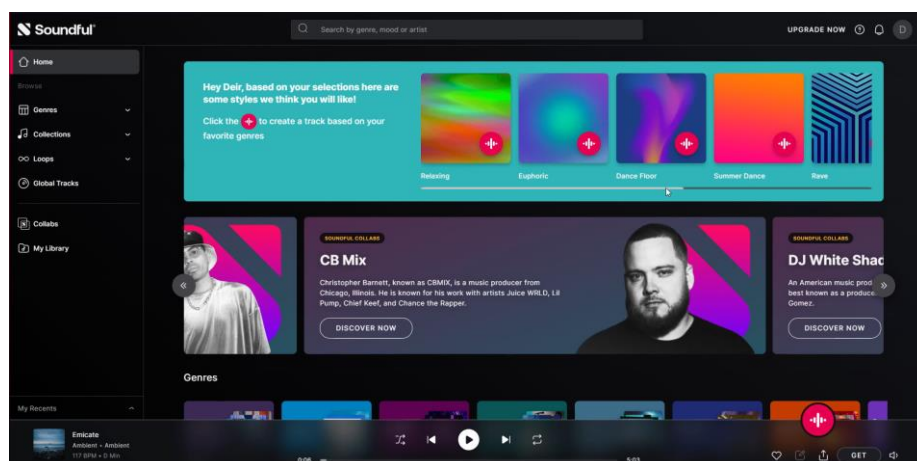


Рисунок 2.7 – Інтерфейс Soundful

Soundful використовує алгоритми, що навчилися музичній теорії по шаблонам, але генерує унікальні композиції, які не повторюють жодну існуючу музику. Тобто, ви завжди отримаєте щось нове й унікальне.

Ще одна цікава річ полягає у процесі навчання алгоритмів. Soundful використовує ноти та музичні семпли, щоб постійно покращувати якість генерованої музики, і це робить його потужним інструментом для творчості.

Але як завжди, є й обмеження. Наприклад, хоча можна отримати унікальну музику, яка не повторює існуючі композиції, вона все одно може не завжди ідеально підходити для конкретного проєкту. Тому експерименти та варіації генерації музики можуть бути потрібні, щоб знайти той самий звук.

Загалом, Soundful відкриває нові можливості для творчості, особливо у випадках, коли потрібно швидко отримати унікальну фонову музику для медійних проєктів, та дозволяє експериментувати зі звуками без спеціальних навичок у музичному творчості.

Платформа Soundful має кілька ключових переваг, які роблять її привабливою для різноманітності користувачів у сфері аудіо-творчості:

- Унікальність музики: Soundful використовує штучний інтелект для генерації унікальних треків, що не мають аналогів на ринку. Це дозволяє користувачам отримувати свіжу та унікальну музику для своїх проєктів;

- Простота використання: Інтерфейс Soundful є інтуїтивно зрозумілим, що робить процес створення музики швидким та легким навіть для тих, хто не має спеціалізованих знань у цій області;

- Гнучкість і кастомізація: Користувачі можуть налаштовувати параметри музики, вибираючи жанр, настрій, темп і багато іншого, щоб створити звук, який ідеально підходить для їхнього контенту;

- Безкоштовний доступ та ліцензування: Платформа надає безкоштовний доступ до створення музики, а також має гнучкі ліцензійні умови, що робить використання треків безпроблемним у різних медійних проєктах;

- Студійна якість звуку: Генерована музика Soundful має високу якість, наближену до студійної, що робить її придатною для різноманітних професійних використаннях.

Вибір Soundful може бути обумовлений комбінацією цих факторів: унікальність звуку, простота використання, гнучкість налаштувань, доступність безкоштовної музики та висока якість аудіо. Ці переваги роблять Soundful привабливою опцією для творчих професіоналів, студентів та брендів, які шукають унікальний та якісний звук для своїх проєктів.

### 2.3 Покращення якості відео з використанням нейромереж та інших технологій

У цифрову епоху відео стало одним із найпопулярніших засобів взаємодії та сприйняття інформації. Покращення якості відео стає важливим завданням для тих, хто прагне зробити свій контент більш привабливим та професійним. Інноваційність нейромереж та розвиток програмного забезпечення відкривають нові можливості для вдосконалення відеоматеріалів. В цій темі розглянуті різноманітні програми та веб-сайти, які ґрунтуються на принципах штучного інтелекту та нейромереж і спрямовані на оптимізацію якості відео, його обробку та покращення зображення.

Є кілька програм та онлайн-сервісів, які використовують нейромережі для покращення якості відео. Деякі з них:

- Topaz Video Enhance AI: Програма використовує штучний інтелект для збільшення роздільної здатності відео, зменшення шуму та покращення деталей;

- Dain-App: Dain-App використовує штучний інтелект для створення більш плавного відео шляхом додавання кадрів;

- VEED.IO: Це онлайн-платформа, яка пропонує інструменти для редагування та покращення відео з використанням нейромереж для фільтрів, кольорової корекції та ін.;

- Unifab: Unifab - це компанія, що працює у сфері обробки відео за допомогою неймереж, зосереджуючись на розвитку алгоритмів для покращення та оптимізації якості відеоконтенту.

Ці інструменти використовують різні алгоритми неймереж для різних аспектів оптимізації відео - від збільшення роздільної здатності до покращення колірної гами та зменшення шуму. Кожен має свої особливості та можливості, тому вибір залежить від конкретних потреб та уподобань.

### 2.3.1 Інструмент для покращення відео Topaz Video Enhance AI

Topaz Video Enhance AI (рис. 2.8) - це програма, яка використовує неймережі для покращення якості відео. Вона працює з роздільною здатністю, шумом та деталями, щоб зробити відео більш чітким і якісним. Основна її функція - збільшення роздільної здатності відео, включаючи збільшення розміру кадру до більш високої роздільної здатності, наприклад, з 1080p до 4K.

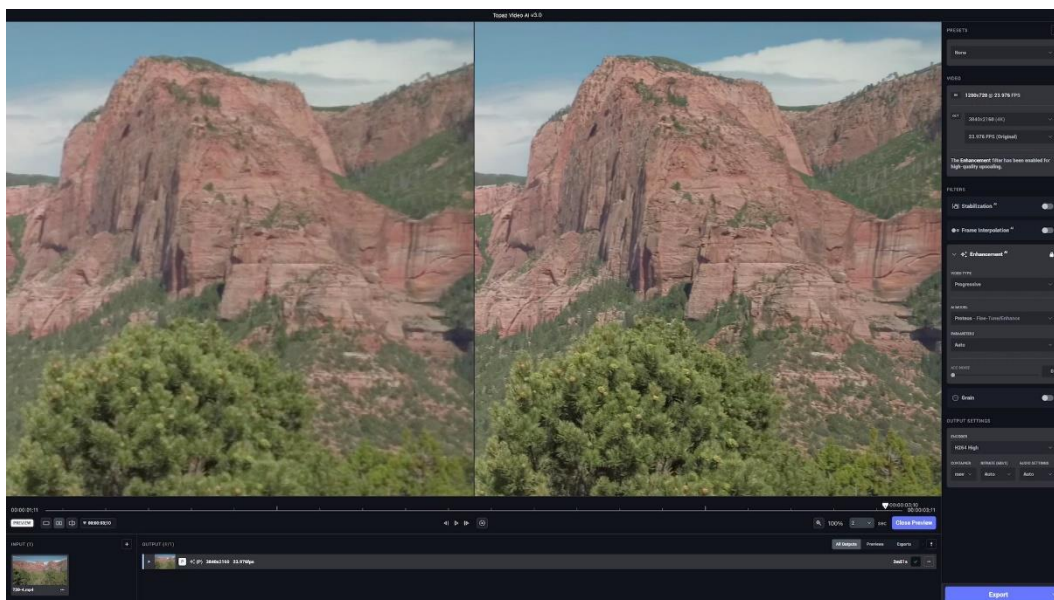


Рисунок 2.8 – Інтерфейс Topaz Video Enhance AI

Це досягається завдяки роботі нейромережі, яка аналізує відео та заповнює деталі, що втрачені при збільшенні розміру, що дозволяє підвищити чіткість та якість кадрів.

Програма також допомагає у зменшенні шуму на відео, що особливо важливо для матеріалу з низькою роздільною здатністю або в умовах поганого освітлення. Вона використовує алгоритми, які дозволяють знизити шум, не втрачаючи при цьому деталей.

Topaz Video Enhance AI використовує комбінацію нейромереж для досягнення своїх результатів. Однією з ключових нейромереж, яку використовує програма, є глибока згорткова нейромережа (CNN). CNN використовується для аналізу кожного кадру відео, визначення його особливостей, структури та деталей. Вона допомагає відтворити відсутні деталі та підвищити роздільну здатність без втрати чіткості.

Також можливо використання глибоких генеративно-зустрічних мереж (GANs) або інших архітектур для генерації відсутніх деталей у відео, оптимізації кольору та підвищення загальної якості.

Ці нейромережі працюють у поєднанні, використовуючи дані з кожного кадру відео для аналізу та підвищення його якості. Це дозволяє програмі ефективно оптимізувати якість відео, зберігаючи деталі та покращуючи вигляд.

Переваги використання Topaz Video Enhance AI:

- Підвищення роздільної здатності: Програма дозволяє збільшувати роздільну здатність відео, що особливо корисно для відео низької якості або з малою роздільною здатністю;

- Зменшення шуму: Вона ефективно працює зі зменшенням шуму на відео, покращуючи загальний вигляд матеріалу;

- Збереження деталей: При підвищенні роздільної здатності чи оптимізації відео, програма намагається зберегти якість і деталість зображення.

Недоліки використання Topaz Video Enhance AI:

- Час обробки: Обробка відео з високим розділенням може займати значний час через складність роботи нейромереж та обчислювальні потужності, що потрібні для цього процесу;

- Вартість: Програма може бути відносно дорогою, особливо якщо використовується регулярно або на комерційних цілях;

- Неідеальність результатів: Хоча програма робить вражаючу роботу, результати можуть не завжди бути ідеальними. Деякі відео можуть мати артефакти або втрату деталей при підвищенні роздільної здатності.

Загалом, Topaz Video Enhance AI є потужним інструментом для тих, хто хоче підвищити якість свого відеоконтенту, зокрема збільшити роздільну здатність, зменшити шум та покращити загальний вигляд відеоматеріалів.

### 2.3.2 Інструмент для покращення відео Dain-App

Dain-App (рис. 2.9) - це програма, яка використовує нейромережі для створення більш плавного відео шляхом додавання кадрів. Це досягається за допомогою техніки, відомої як "інтерполяція кадрів". Основна ідея полягає в тому, щоб заповнити проміжки між існуючими кадрами новими кадрами, що створюють плавніше перехід між ними.

Наприклад, якщо вихідне відео має низьку частоту кадрів, Dain-App може використовувати штучний інтелект для створення додаткових кадрів між існуючими, щоб зробити рух більш плавним та природним. Це особливо корисно для відео з низькою частотою кадрів, таких як анімації, старі відео, чи матеріал знятий на дешевіше камери.

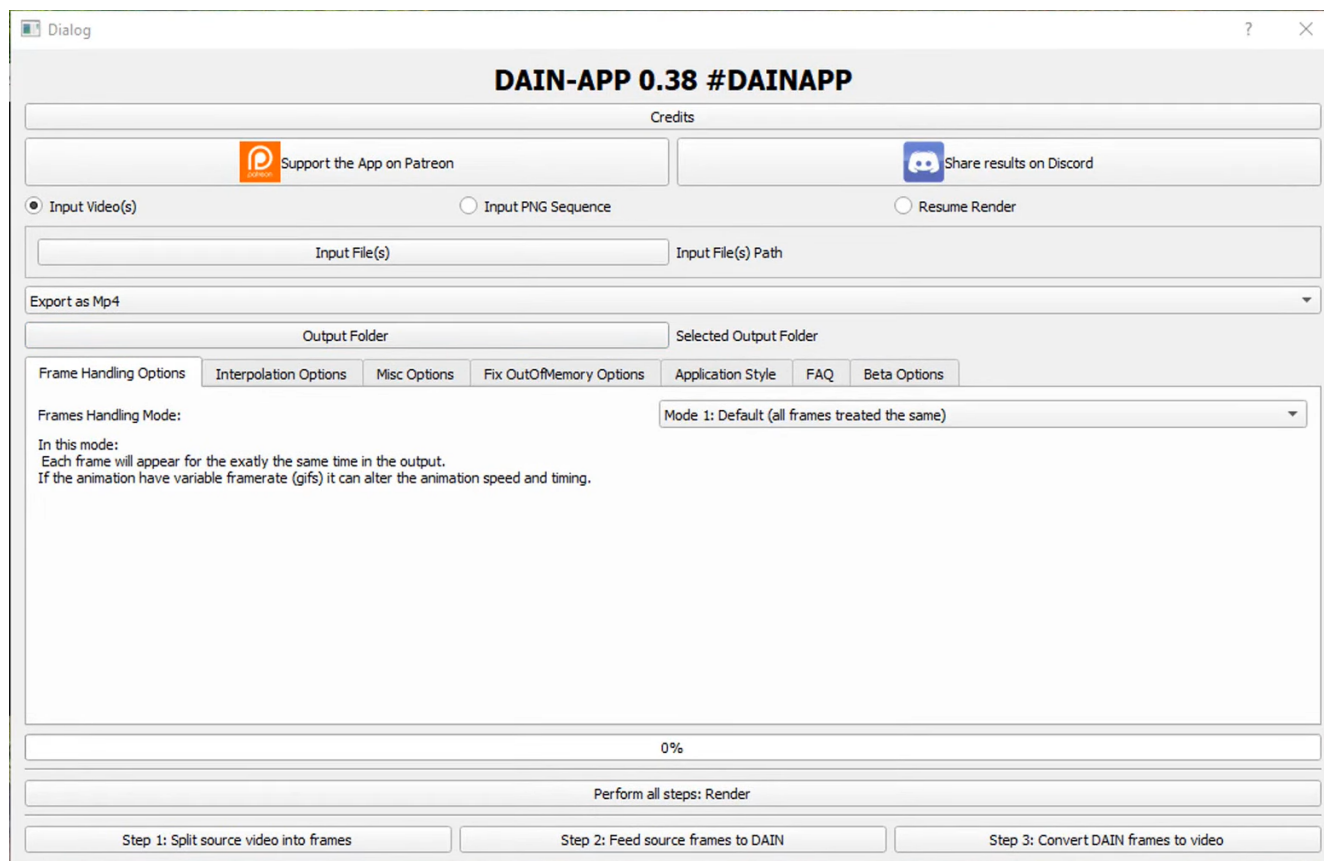


Рисунок 2.9 – Інтерфейс Dain-App

Однак важливо зазначити, що якщо програма використовується неправильно або для відео з високою частотою кадрів, це може призвести до втрати якості або з'явлення артефактів. Тому варто ретельно налаштовувати параметри або експериментувати з різними відео, щоб отримати найкращий результат.

Dain-App використовує глибокі згорткові нейронні мережі (CNN) для досягнення своїх результатів. Вона базується на архітектурі, яка здатна генерувати нові кадри між існуючими відеокадрами.

Програма використовує нейромережі для аналізу попередніх та наступних кадрів у відео, вивчаючи залежності між ними. Потім вона генерує нові кадри, що відповідають цим залежностям, для створення більш плавної анімації чи відео.

Така архітектура нейромережі дозволяє Dain-App розпізнавати та моделювати рух об'єктів у відео, щоб створити додаткові кадри, які плавно переходять між існуючими. Вона намагається зробити перехід між кадрами більш непомітним та природним за рахунок глибокого розуміння динаміки руху у відео.

Переваги використання:

- Поліпшення плавності відео: Програма дозволяє додавати кадри для покращення плавності руху на відео з низькою частотою кадрів;
- Візуальне поліпшення: Це допомагає покращити вигляд анімацій та іншого контенту з недостатньою плавністю;
- Автоматизація процесу: Програма робить процес додавання кадрів більш автоматизованим, вимагаючи менше ручної роботи з боку користувача.

Недоліки використання:

- Втрата якості: При неправильному застосуванні або обробці відео з високою частотою кадрів може виникнути втрата якості або артефакти, такі як розмитість або аномальний рух;
- Час обробки: Додавання нових кадрів може бути високовитратною операцією, особливо для великих файлів або відео з високою роздільною здатністю;
- Потреба в налаштуваннях: Часто для досягнення найкращих результатів потрібно експериментувати з параметрами програми, що може зайняти час та вимагати певного рівня експертизи.

Розуміння цих плюсів та мінусів допоможе краще використовувати Dain-App та отримати найкращі результати при обробці відео.

### 2.3.3 Інструмент для покращення відео VEED.IO

VEED.IO (рис. 2.10) - це онлайн-платформа для редагування відео, яка використовує інструменти на основі штучного інтелекту та нейромереж для різних завдань, включаючи фільтри, кольорову корекцію та багато іншого.

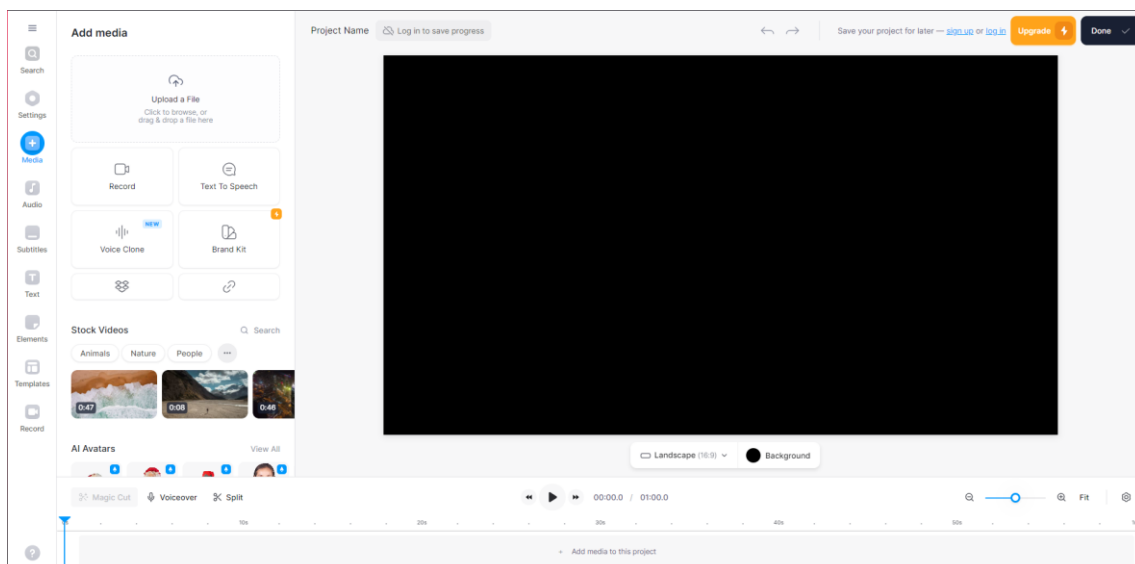


Рисунок 2.10 – Інтерфейс VEED.IO

Ось кілька ключових функцій та можливостей цієї платформи:

- Редагування відео: VEED.IO надає можливість обробляти, обрізати, з'єднувати та редагувати відео в онлайн-режимі без необхідності встановлення спеціального програмного забезпечення;

- Фільтри та ефекти: Платформа пропонує різноманітні фільтри та ефекти для відео, які можна застосовувати для покращення вигляду контенту. Вони можуть бути засновані на нейромережах для покращення кольорів, освітлення та загального вигляду відео;

- Кольорова корекція: Засоби коригування кольору дозволяють покращити кольорову палітру відео, змінюючи контрастність, насиченість та інші параметри, щоб зробити відео більш привабливим та яскравим;

- Автоматичні інструменти: Є інструменти, що використовують штучний інтелект для автоматичної обробки, такі як автоматична підписка, стабілізація відео та інше;

- Відео з підписом та анімації: VEED.IO дозволяє легко додавати підписи, текст та анімацію до відео за допомогою шаблонів та інструментів.

Платформа створена з урахуванням користувача, який шукає простий інструмент для редагування відео, але в той же час хоче мати доступ до різних функцій

покращення та редагування з використанням потужних технологій штучного інтелекту.

VEED.IO використовує широкий спектр інструментів та алгоритмів для обробки та редагування відео, але конкретні нейронні мережі, які використовуються в цій платформі, не завжди є публічно відомими. Такі платформи часто використовують комбінацію глибоких згорткових мереж (CNN), рекурентних нейронних мереж (RNN), глибоких генеративно-зустрічних мереж (GANs), а також інших архітектур для різних завдань обробки відео.

Наприклад, для функцій кольорової корекції, стабілізації відео або покращення якості, можуть використовуватися різні нейромережі для роботи з пікселями, обробки зображень, розпізнавання образів та багато іншого.

Такі системи використовують навчання з учителем на великих наборах даних для покращення якості відео, фільтрації, аналізу та обробки. Вони можуть використовувати попередньо навчені моделі, адаптовані до конкретних завдань обробки відео, або ж навчатися на власних даних, залежно від потреб користувача та функцій, які надає платформа VEED.IO.

Переваги використання VEED.IO:

- Легкість використання: Платформа має інтуїтивний інтерфейс, який дозволяє швидко редагувати відео без спеціалізованих навичок;
- Онлайн доступ: Немає необхідності встановлювати спеціальне програмне забезпечення. Редагування відео відбувається онлайн через веб-браузер;
- Широкий функціонал: Платформа пропонує різноманітні інструменти для редагування, від кольорової корекції та фільтрів до додавання тексту та анімації;
- Використання штучного інтелекту: Вона використовує нейромережі для різних завдань обробки відео, що може покращити якість та результати редагування.

Недоліки використання VEED.IO:

- Обмежені можливості в порівнянні з деякими desktop програмами: У порівнянні з деякими професійними програмами для редагування відео, онлайн-платформи можуть мати обмежені можливості або меншу гнучкість;

- Залежність від інтернет-з'єднання: Редагування відео в онлайн-режимі потребує стабільного інтернет-з'єднання, щоб працювати без перебоїв;
- Обмеженість функцій для безкоштовної версії: Деякі продуктивні функції можуть бути доступні тільки в платних версіях платформи;
- Час обробки: Обробка великих файлів відео або складних ефектів може займати більше часу через обмежену потужність операцій на серверах.

Платформа має свої переваги, але також варто враховувати й можливі обмеження, щоб визначити, чи вона відповідає тим чи іншим потребам у редагуванні відео.

### 2.3.4 Інструмент для покращення відео Unifab

DVDFab, утворена в 2003 році, визнана міжнародним лідером у розробці програмного забезпечення для обробки мультимедійних даних. Вони спеціалізуються на створенні та вдосконаленні інструментів для користувачів у всьому світі.

Unifab (рис. 2.11) є дочірнім проєктом DVDFab та StreamFab. Однією з інновацій Unifab є Video Enlarger AI. Його моделі штучного інтелекту навчені розпізнавати, аналізувати та покращувати відеоконтент за допомогою передових алгоритмів глибокого навчання, що надає більш реалістичний та вражаючий візуальний ефект.

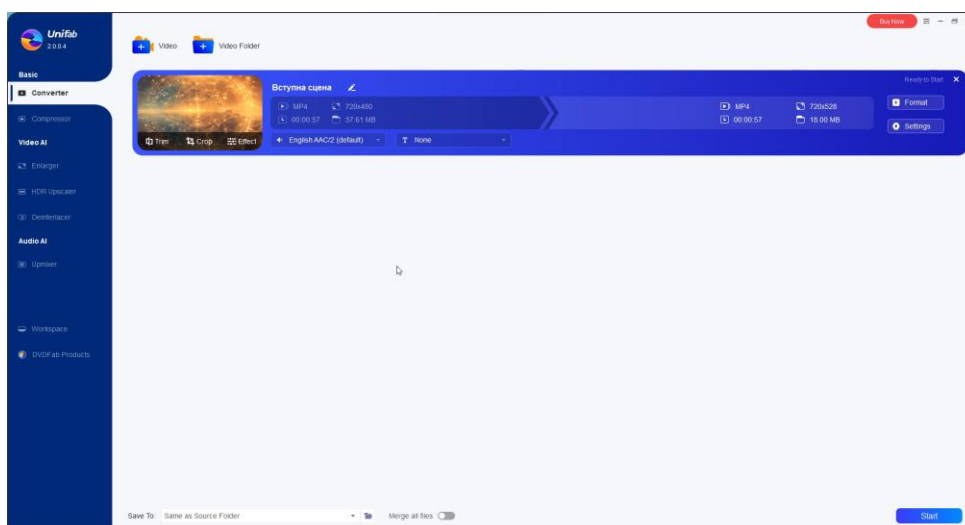


Рисунок 2.11 – Інтерфейс Unifab

Унікальна можливість Unifab Video Enlarger AI - збільшення роздільної здатності відео до 4K. Інструмент не лише здатен підняти роздільну здатність відео до 720P, 1080P та вражаючого 4K Ultra HD, але й подвоїти роздільну здатність відео відносно оригінальної, створюючи вражаючий ефект. Покращуйте якість відео для кожного жанру

UniFab Video Enlarger AI розробляє універсальну модель покращення, що забезпечує вражаючі результати для всіх видів відео - від телешоу з низькою роздільною здатністю до чорно-білих фільмів, домашнього відео та анімації. Це програмне забезпечення для відновлення відео дозволяє досягти результатів без витрати часу на вибір правильної моделі для ваших потреб.

Простий та зрозумілий інтерфейс робить UniFab Video Enlarger AI доступним для початківців. Часто програмне забезпечення для покращення відео вимагає розуміння штучного інтелекту для коректної роботи, але UniFab Video Enlarger AI звільняє від необхідності налаштування складних параметрів. Його штучний інтелект забезпечує автоматизацію процесу, виконуючи роботу за вас.

Ще однією перевагою є збільшена швидкість роботи - до 50 разів швидше, ніж зазвичай. Завдяки передовим технологіям апаратного прискорення графічного процесора, таким як NVIDIA CUDA, AMD і Intel Quick Sync, штучний інтелект UniFab Video Enlarger забезпечує ефективну обробку відео, дозволяючи більш продуктивно виконувати завдання в обмежені строки.

Переваги використання програми:

- Універсальність в роботі з будь-яким типом відео: Здатність працювати з різноманітними форматами відео та вирізнятися у покращенні різних типів контенту - чорно-білі фільми, домашнє відео, анімація тощо;

- Простий інтерфейс: Легкість використання для початківців, відсутність необхідності в налаштуванні складних параметрів;

- Автоматизація: Вбудований штучний інтелект, який забезпечує автоматичне вирішення завдань без потреби в особливих знаннях користувача;

- Швидкість роботи: За допомогою апаратного прискорення графічного процесора, програма забезпечує значне прискорення обробки відео.

Недоліки використання програми:

- Обмеження в роботі: Хоча програма здатна обробляти різні типи відео, її ефективність може бути обмежена у випадку дуже складних сцен або особливо низької якості вихідного матеріалу;

- Недостатня гнучкість: Необхідність використовувати вбудовані моделі, що може обмежувати користувача у творчих підходах до обробки відео та виборі оптимального методу покращення.

Вибір програми UniFab Video Enlarger AI може бути обґрунтований його універсальністю та легкістю використання. Програма надає простір у використанні та автоматизований процес для користувачів будь-якого рівня знань. Крім того, здатність програми працювати з різними типами відео та швидкість обробки завдань можуть бути важливими аспектами при виборі цієї програми. Однак варто врахувати обмеження в гнучкості та можливість обробки складних відеосцен.

Висновок по розділу: В розділі розглядаються різноманітні методи використання нейромереж для створення голосу, музики та покращення відеоряду, а також проводиться аналіз сучасного програмного забезпечення, що використовується для мультимедійної обробки та творчості. У цьому контексті були обрані Voicemaker, Soundful та Unifab як представники передових технологій у створенні голосу та обробці музичного та відеоконтенту.

### 3 ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖ У СТВОРЕННІ ГОЛОСУ, МУЗИКИ ТА ПОКРАЩЕННІ ЯКОСТІ ВІДЕО

Робота над кваліфікаційною роботою магістра демонструє використання комплексного підходу для створення відеоматеріалу з використанням голосу, музики та покращення його якості. Незважаючи на сучасний рівень розвитку нейромереж, їхні можливості в створенні відео поки що обмежені, що змусило використовувати ручний підхід у цьому процесі. Проте, нейромережі знаходять своє застосування у покращенні якості вже готового відеоматеріалу.

Одним із найскладніших етапів було вибір програмного забезпечення та онлайн-ресурсів, що ефективно виконували б покладені функції та були доступними для реалізації. Після аналізу різних варіантів були обрані такі програми: Voicemaker для голосу, Soundful для музичного супроводу та Unifab для покращення візуальної якості відеоряду. Ці інструменти стали ключовими для виконання основної частини завдання, створюючи можливість для подальшого розвитку творчого процесу.

#### 3.1 Створення голосу за допомогою Voicemaker

Процес користування платформою Voicemaker вимагає створення облікового запису для активації пробного періоду. Це можна зробити шляхом авторизації через Google, Facebook або Twitter (рис. 3.1). Після успішної автентифікації користувач отримує доступ до інтерфейсу платформи (рис. 3.2).

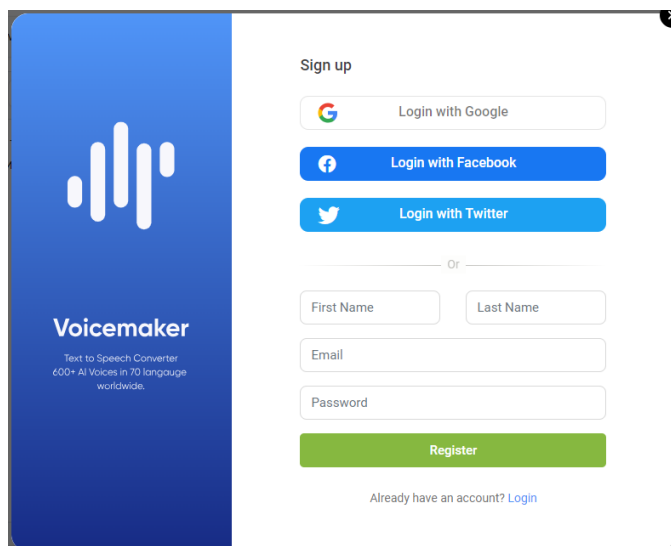


Рисунок 3.1 – Поле авторизації

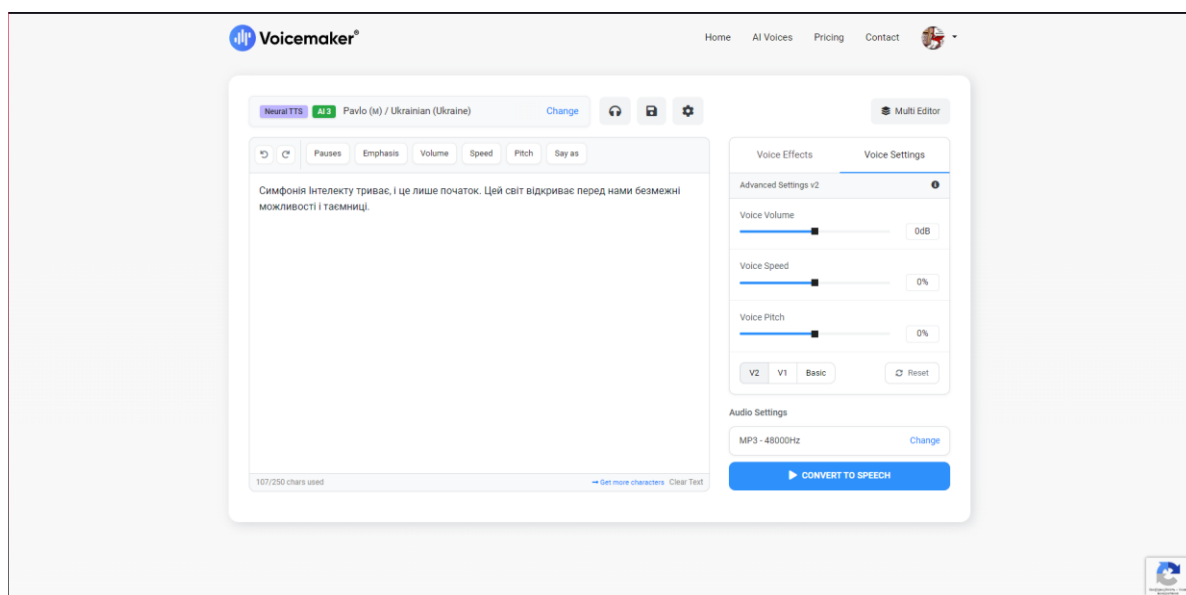


Рисунок 3.2 – Основний інтерфейс

Інтерфейс Voicemaker пропонує можливість змінювати голосові параметри. Для вибору іншого голосу потрібно натиснути кнопку "Change". У відкритому вікні можна обирати не лише конкретний голос та регіон, а й визначати гендерні та нейронні мережеві налаштування (рис. 3.3). Важливо відзначити, що обраний голос повинен відповідати мові тексту, що буде поданий, інакше програма може не коректно працювати.

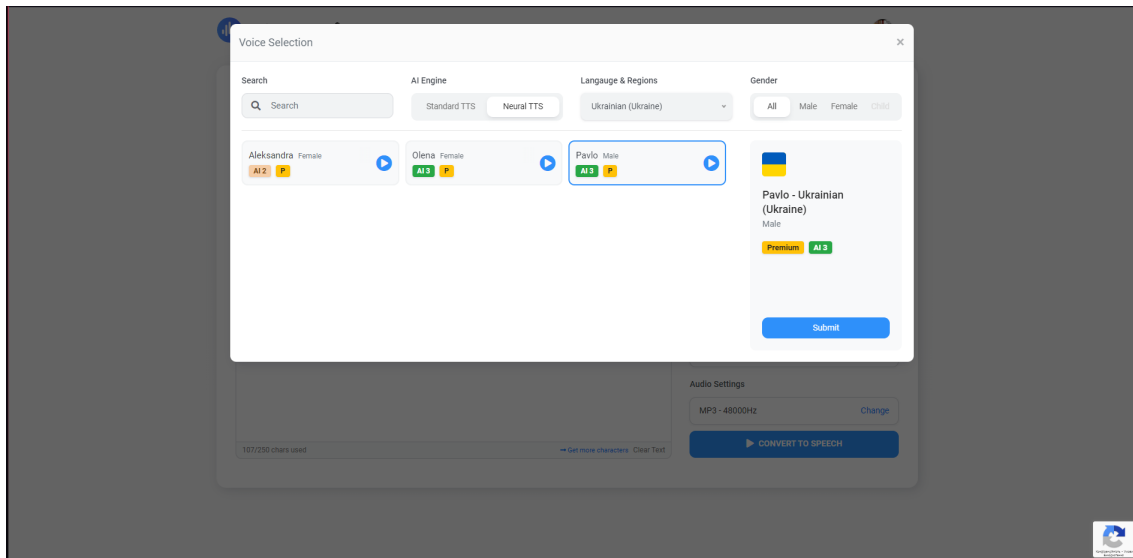


Рисунок 3.3 – Вікно зміни голосу

Після вибору голосу користувач вводить текст у велике поле і після цього активує функцію "Convert to speech" у правому нижньому куті. Перед цим керуючий панеллю можна детально налаштувати гучність, швидкість та інші параметри голосу. Після натискання на кнопку з'являється аудіозапис, який можна прослухати та завантажити, якщо результат відповідає вимогам користувача (рис. 3.4).

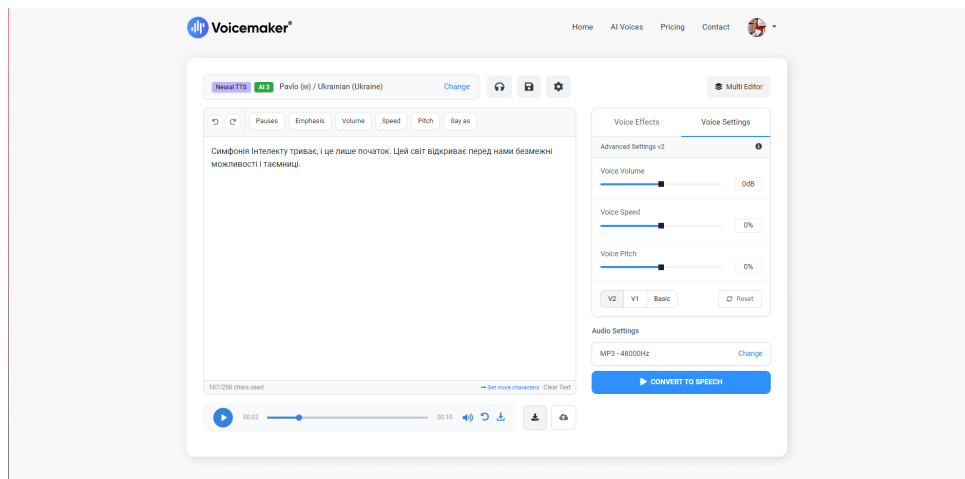


Рисунок 3.4 – Аудіозапис внизу інтерфейса

## 3.2 Створення музики за допомогою Soundful

Процес використання програмного забезпечення Soundful включає авторизацію, яка доступна через обліковий запис Google (рис. 3.5). Після успішної автентифікації відкривається інтерфейс програми, що створює сприятливу основу для роботи з музикою.

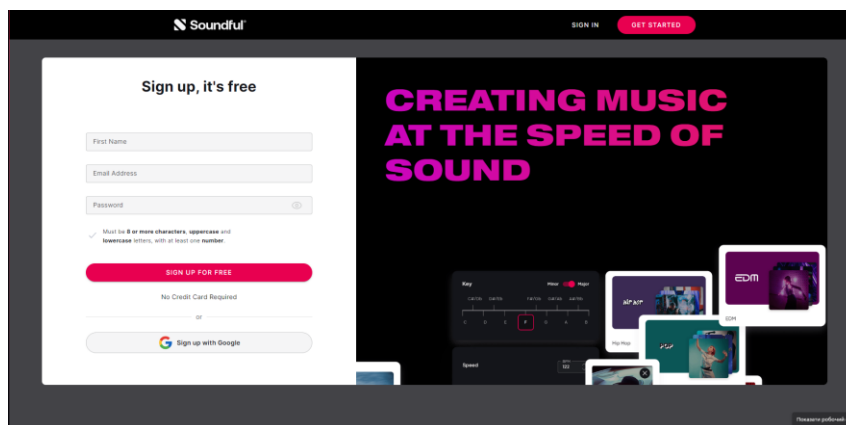


Рисунок 3.5 – Вікно авторизації

Користувач має можливість перейти до процесу створення музики, натиснувши червону кнопку, розташовану у нижньому правому кутку (рис. 3.6), та обравши опцію "Track" (рис. 3.7). Такий крок відкриває доступ до різноманітних типів треків (рис. 3.8), з яких можна обрати потрібний, а потім визначити його стиль (рис. 3.9).

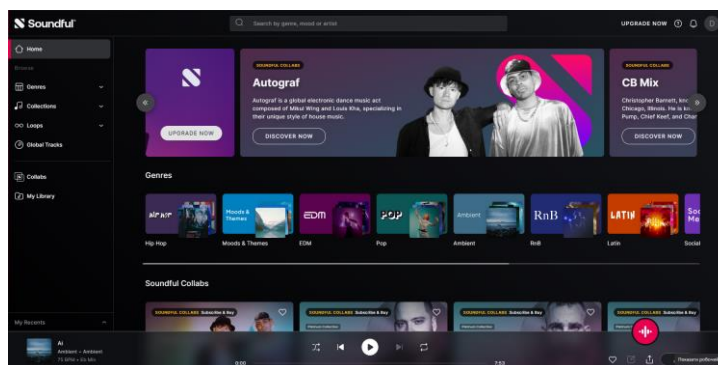


Рисунок 3.6 – Основний інтерфейс

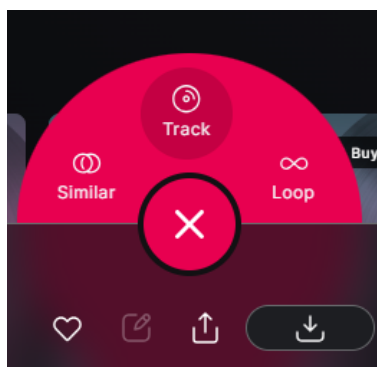


Рисунок 3.7 – Колесо опцій

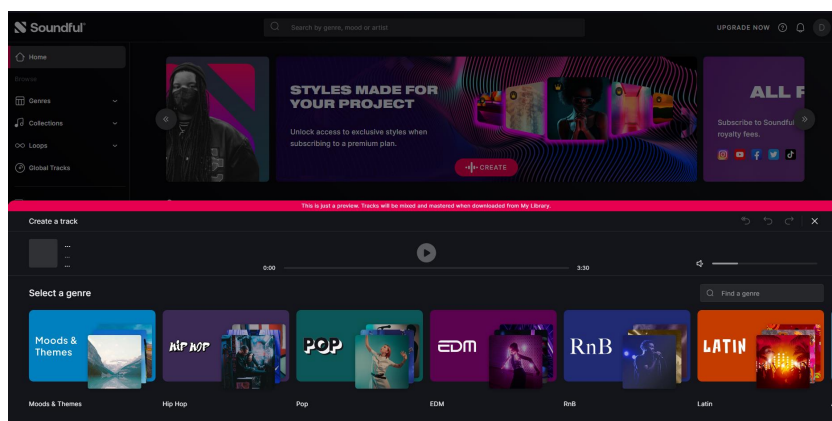


Рисунок 3.8 – Типи треків

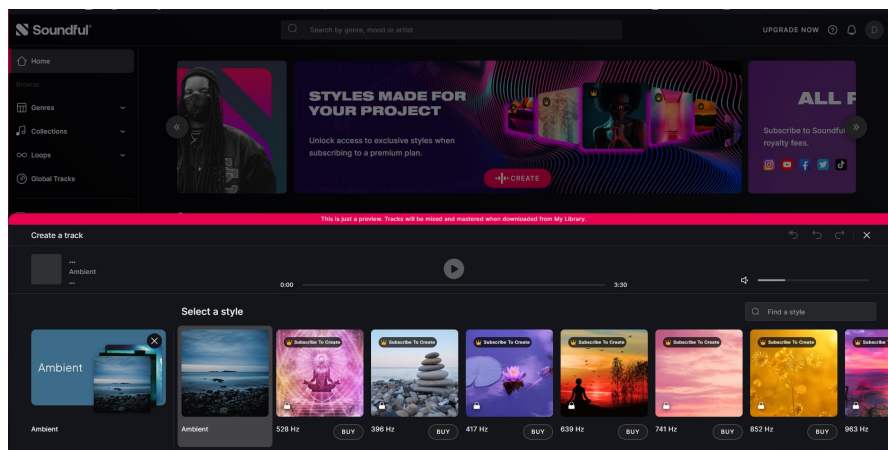


Рисунок 3.9 – Стиль треків

Після обрання стилю музики користувач має можливість налаштувати різні параметри, такі як швидкість треку, музичні ключі та інші характеристики, що впливають на його звучання (рис. 3.10).

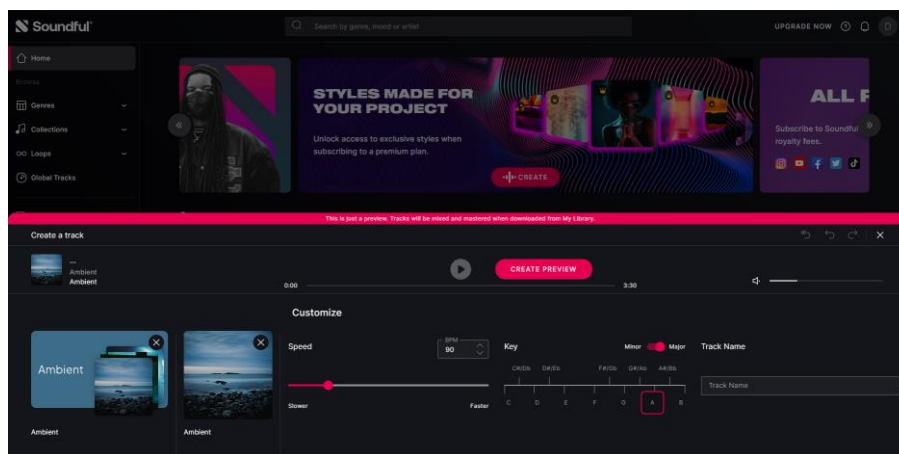


Рисунок 3.10 – Поле налаштування

Після завершення налаштувань користувач може натиснути кнопку "Create preview", щоб прослухати попередньо створену композицію. Якщо результат відповідає очікуванням, можна ввести назву треку та зберегти його, натиснувши кнопку "Save" (рис. 3.11). Створена композиція автоматично заноситься до особистої бібліотеки користувача і стає доступною для завантаження.

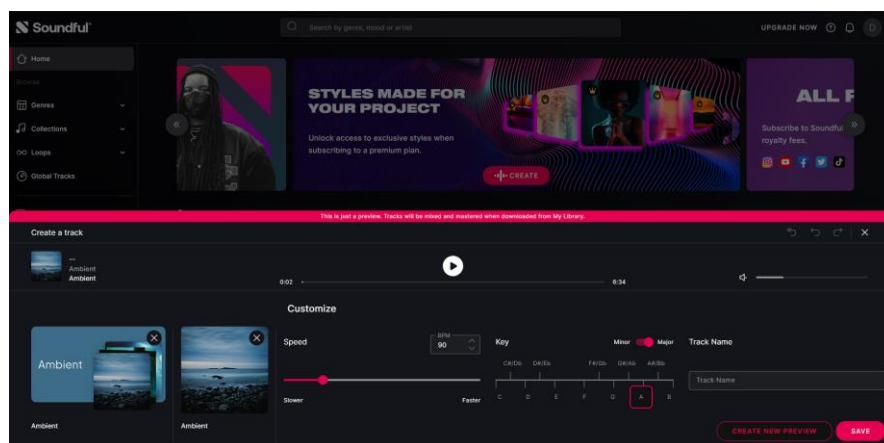


Рисунок 3.11 – Кінцеве поле

### 3.3 Покращення відео за допомогою Unifab

Unifab представляє собою програмне забезпечення, розроблене для операційної системи Windows. Для початку використання програми, відвідувачеві сайту не-

обхідно натиснути кнопку "Free Download" (рис. 3.12). Такий крок відкриває можливість обрати місце збереження установочного файлу, після чого завантажити його на комп'ютер.

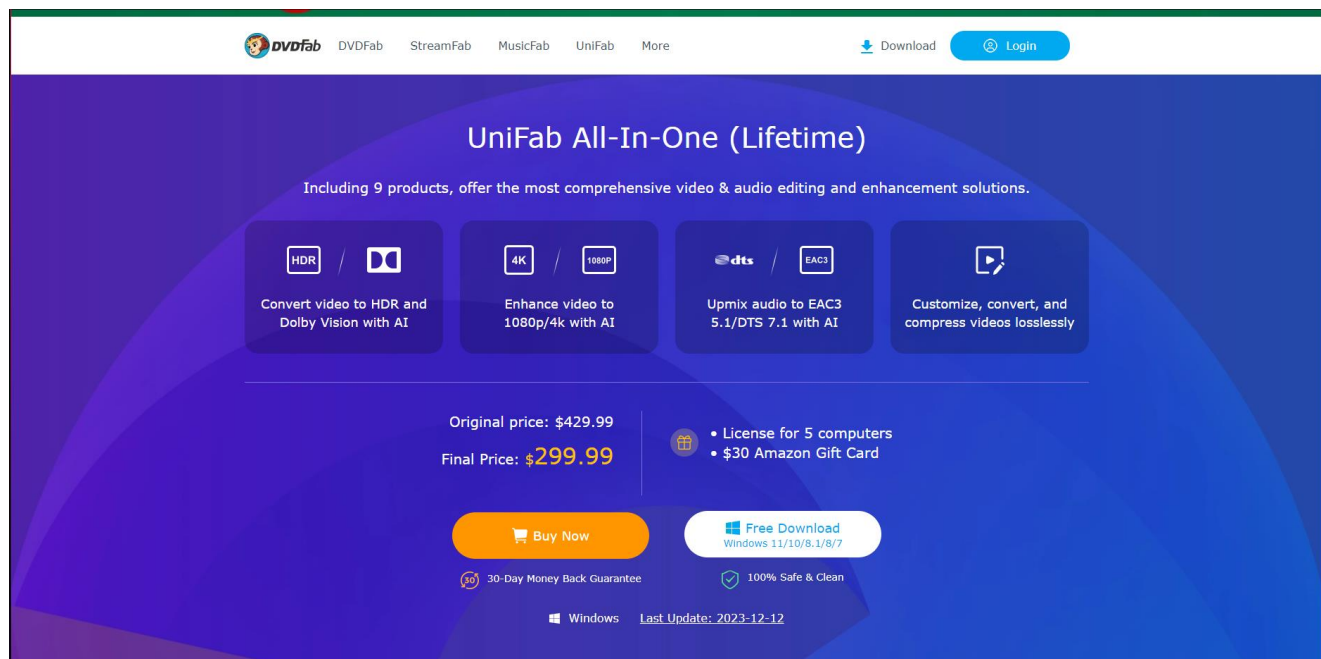


Рисунок 3.12 – Вікно сайту

Після завершення процесу завантаження та успішної установки програми відкривається вікно, де користувач може розпочати використання пробної версії, натиснувши кнопку "Try". Після виконання всіх процедур з реєстрацією відкривається основний інтерфейс (рис. 3.13), де для подальшої роботи з відео необхідно обрати опцію "Video" у лівому верхньому куті, щоб завантажити необхідний відеоряд у програму.

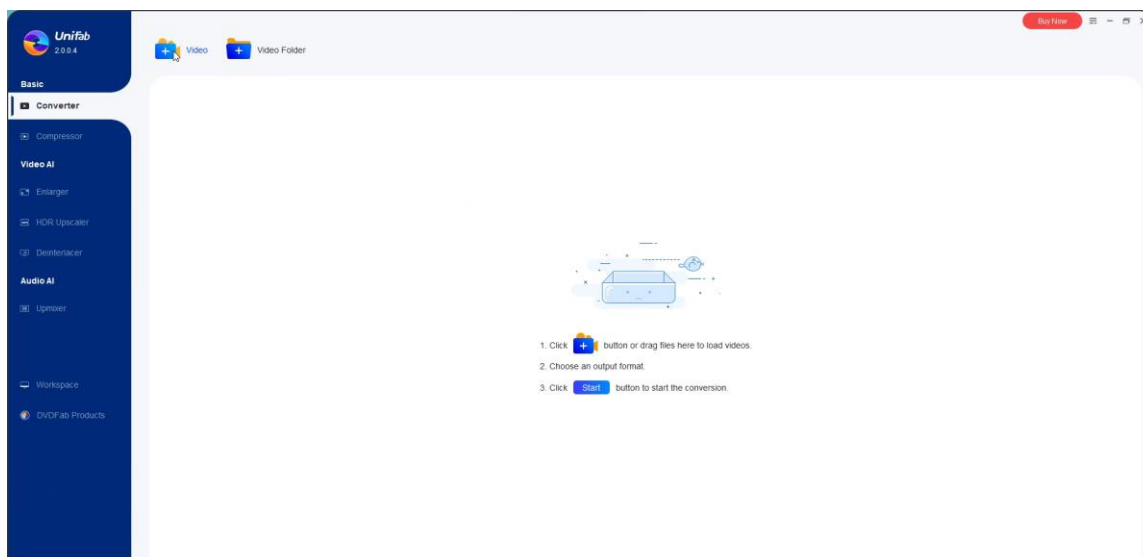


Рисунок 3.13 – Інтерфейс

Після завантаження відео (рис. 3.14) необхідно натиснути кнопку "Effects", після чого відкриється вікно налаштувань (рис. 3.15), де користувач може накласти різноманітні ефекти, регулювати яскравість, контраст та інші параметри.

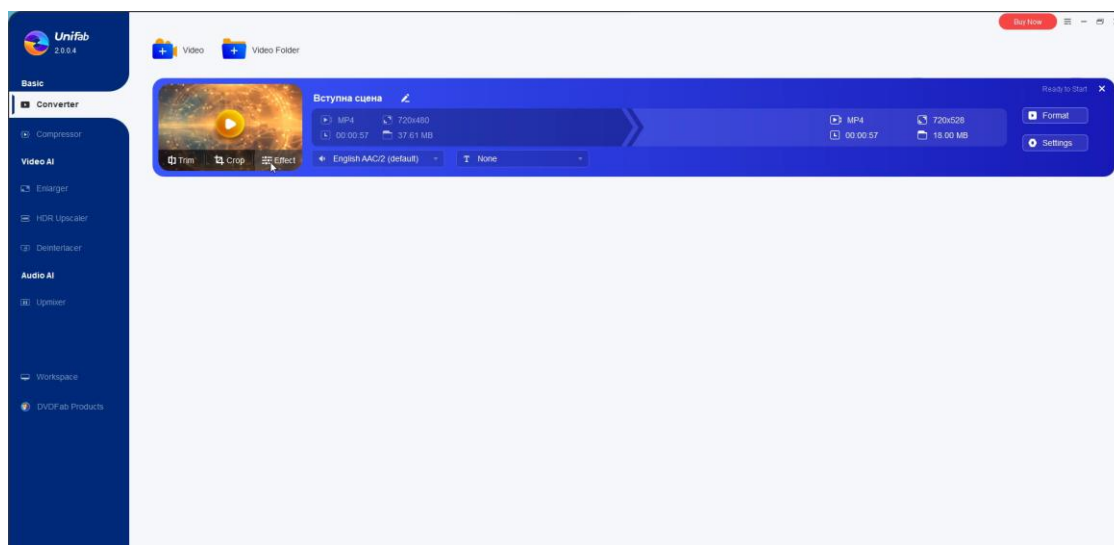


Рисунок 3.14 – Інтерфейс із відео

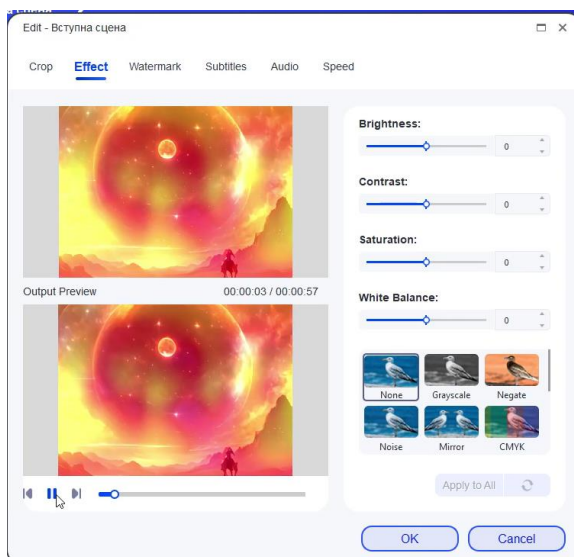


Рисунок 3.15 – Вікно налаштування ефектів

Далі, натиснувши кнопку "Settings" справа від відео, з'являється вікно налаштувань чіткості та розширення відео (рис. 3.16). Користувач може вибрати необхідні параметри, які вплинуть на якість відеоматеріалу.

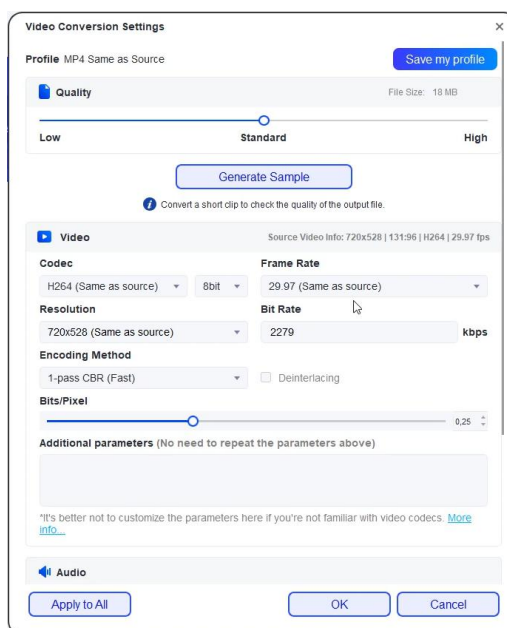


Рисунок 3.16 – Вікно налаштувань чіткості та розширення відео

Після встановлення необхідних параметрів користувач має натиснути кнопку "Generate Sample", після чого програма створить готовий покращений відеоряд з врахуванням зроблених налаштувань.

Висновок по розділу: Після аналізу різних варіантів були обрані такі програми: Voicemaker для створення голосу, Soundful для музичного супроводу та Unifab для покращення візуальної якості відеоряду. Ці інструменти стали ключовими для виконання основної частини завдання, створюючи можливість для подальшого розвитку творчого процесу.

Розділ детально розглядає етапи створення голосу, музики та покращення відео з використанням нейромереж. Програми Voicemaker, Soundful та Unifab були обрані на основі їхньої ефективності та здатності виконати визначені функції, створюючи основу для успішного виконання завдань та відкриваючи перспективи для подальшого розвитку креативного процесу.

Отримані результати підкреслюють важливість інтеграції нейромереж у сучасний процес створення мультимедійного контенту та свідчать про їхній потенціал у покращенні якості та творчому використанні голосу, музики та відео.

## ВИСНОВКИ

Висновок кваліфікаційної роботи магістра на тему "Використання нейромереж у роботі з фото та відео: голос, музика, обробка" полягає у виявленні суттєвого впливу штучного інтелекту на сучасну аудіо- та відеоіндустрію. Дослідження ретельно розглянуло різноманітні інструменти та програми, які використовують нейромережі для генерації звуку та музики з тексту, зокрема LOVO, MURF.ai, AIVA, Voicemaker та інші.

Оцінка цих інструментів врахувала якість їхньої роботи, можливості редагування та налаштувань, що вони пропонують, а також висвітлила їхній інтуїтивний інтерфейс, що робить їх доступними для широкого кола користувачів.

Важливу роль у цьому дослідженні відіграли технічні аспекти, зокрема використання нейронних мереж, аналіз яких дозволив краще розуміти функціонування інструментів у контексті алгоритмів машинного та глибокого навчання.

В ході дослідження функціональний аналіз програмних рішень, зокрема Voicemaker, Soundful та Unifab, дозволив виявити їхні потужності у сфері обробки мультимедійного контенту. Voicemaker надає можливості для створення та налаштування голосу з широким спектром параметрів, дозволяючи адаптувати його для різних потреб. Soundful розширює можливості користувача за допомогою різноманітних музичних стилів та параметрів, сприяючи швидкому створенню музичного супроводу для різних проєктів.

Крім того, Unifab забезпечує обробку відео на платформі Windows та налаштування візуального контенту для покращення якості відеоряду. Ці програмні рішення відкривають можливості для творчого експерименту та підвищення якості мультимедійного матеріалу через налаштування різноманітних ефектів та параметрів.

Використання таких програм дозволяє забезпечити комплексне покращення аудіо та відео контенту, надаючи користувачам широкий спектр інструментів для

творчого вираження та оптимізації мультимедійного матеріалу. Підхід сприяє якісному та ефективному створенню мультимедійних продуктів та проєктів у наукових, творчих чи комерційних цілях.

Кваліфікаційна робота магістра підкреслила потенціал штучного інтелекту для творчого використання у музичній та аудіовізуальній сферах, що вказує на нові можливості та інновації, які він вносить у ці галузі. Результати дослідження відображають не лише сучасний стан розвитку технологій, але й показують перспективи подальшого росту і використання штучного інтелекту в аудіо- та відеоіндустріях.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. 2016, pp. 127–136.
2. Szeliski, R. 2010. Computer Vision: Algorithms and Applications. 2010, pp. 245–254.
3. Foster, D. Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play. 2019, pp. 183–192.
4. Shanmugamani, R. Deep Learning for Computer Vision. 2018, pp. 92–101.
5. Villan, A. F. Mastering OpenCV 4 with Python. 2019, pp. 301–310.
6. Géron, A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. 2019, pp. 408–417.
7. Dadhich, A., Shrimali, V. R. Practical Computer Vision: Extract insightful information from images using TensorFlow and Keras. 2020, pp. 145–154.
8. Forsyth, D., Ponce, J. Computer Vision: A Modern Approach. 2011, pp. 55–64.
9. Iyatomi, H., Gabrys, B. Convolutional Neural Networks in Visual Computing: A Concise Guide. 2016, pp. 76–85.
10. He, K., et al. Mask R-CNN. 2017, pp. 221–230. Режим доступу до ресурсу: <https://arxiv.org/abs/1703.06870> (дата звернення: 15.11.2023).
11. Ronneberger, O., Fischer, P., Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015, pp. 112–121. Режим доступу до ресурсу: <https://arxiv.org/abs/1505.04597> (дата звернення: 12.12.2023).
12. Krizhevsky, A., Sutskever, I., Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. 2012, pp. 32–41. Режим доступу до ресурсу: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> (дата звернення: 05.12.2023).
13. Long, J., Shelhamer, E., Darrell, T. Fully convolutional networks for semantic segmentation. 2015, pp. 88–97. Режим доступу до ресурсу: <https://arxiv.org/abs/1411.4038> (дата звернення: 10.11.2023).

14. NerdyNav - Best AI Voice Generators. Режим доступа до ресурсу: <https://nerdynav.com/best-ai-voice-generators> (дата звернення: 01.12.2023).

15. Unite AI - Best AI Music Generators. Режим доступа до ресурсу: <https://www.unite.ai/best-ai-music-generators> (дата звернення: 20.11.2023)

16. Режим доступа до ресурсу: <https://ihal.it/i-10-migliori-algoritmi-di-apprendimento-automatico> (дата звернення: 05.11.2023)

17. Режим доступа до ресурсу: <https://www.mdpi.com/2076-3417/13/21/11935> (дата звернення: 12.11.2023)

18. Режим доступа до ресурсу: [https://link.springer.com/chapter/10.1007/978-3-030-80285-1\\_13](https://link.springer.com/chapter/10.1007/978-3-030-80285-1_13) (дата звернення: 20.11.2023)

19. Режим доступа до ресурсу: <https://www.intelligenzaartificialeitalia.net/post/deep-learning-definizioni-chiave> (дата звернення: 02.12.2023)

20. Режим доступа до ресурсу: [https://tesi.univpm.it/retrieve/09b43470-4092-4e4f-bab2-b1782ce3be89/Tesi\\_di\\_laurea.pdf](https://tesi.univpm.it/retrieve/09b43470-4092-4e4f-bab2-b1782ce3be89/Tesi_di_laurea.pdf) (дата звернення: 11.12.2023)

21. Режим доступа до ресурсу: [https://uk.wikipedia.org/wiki/Файл:VAE\\_Basic\\_uk.jpg](https://uk.wikipedia.org/wiki/Файл:VAE_Basic_uk.jpg) (дата звернення: 18.12.2023)

22. Режим доступа до ресурсу: <https://tryolabs.com/blog/2018/01/18/faster-r-cnn-down-the-rabbit-hole-of-modern-object-detection> (дата звернення: 08.12.2023)