

ХЕРСОНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
(повне найменування вищого навчального закладу)
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ДИЗАЙНУ
(повне найменування інституту, назва факультету (відділення))
КАФЕДРА ПРОГРАМНИХ ЗАСОБІВ ТА ТЕХНОЛОГІЙ
(повна назва кафедри (предметної, циклової комісії))

Пояснювальна записка

до кваліфікаційної роботи

бакалавра

(освітній рівень)

на тему: «Розробка мистецького програмного додатку на основі
генеративного штучного інтелекту»

Виконав: здобувач освіти 4 року навчання,
групи 4ПР2 спеціальності
121 - «Інженерія програмного забезпечення»
(шифр і назва спеціальності)

Шевченко Максим Андрійович

(прізвище та ініціали)

Керівник д.т.н., проф. Жарікова М.В.

(прізвище та ініціали)

Рецензент к.т.н. доцент Вишемирська С.В.

(прізвище та ініціали)

Хмельницький - 2024

Херсонський національний технічний університет

(повне найменування вищого навчального закладу)

Факультет, відділення Інформаційних технологій та дизайну
 Кафедра Програмних засобів і технологій
 Освітній рівень бакалавр
 Спеціальність 121 – Інженерія програмного забезпечення

(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри

Програмних засобів та технологійк.т.н. доц. О.Є. Огнєва

“ ____ ” _____ 2024 р.

З А В Д А Н Н Я**НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧУ**Шевченко Максиму Андрійовичу

(прізвище, ім'я, по батькові)

Тема роботи «Розробка мистецького програмного додатку на основі генеративного штучного інтелекту»керівник роботи д.т.н. професор Жарікова М.В.,

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджена наказом вищого навчального закладу від 09.02.2024 р. №217-С2. Строк подання здобувачем роботи 15.06.2024 р.3. Вихідні дані до роботи літературні та періодичні джерела, теоретичні та експериментальні дослідження процесу генерації зображень з використанням моделей латентної дифузії

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити):

1) аналіз поточних засобів та інструментів у сфері генеративного мистецтва;2) дослідження моделей латентної дифузії;3) практичне використання моделей латентної дифузії та аналіз отриманих результатів.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

1. Змагальні мережі.2. Моделі дифузії.3. Схема взаємодії VAE, U-Net та CLIP під генерації зображення.4. Схема роботи варіаційного автоенкодера.5. Структура мережі U-Net.

6. Алгоритм навчання моделі CLIP

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів виконання роботи	Термін виконання етапів роботи	Примітки
1.	Отримання завдання	09.02.2024	Виконано
2.	Підбір літератури	09.02.2024-21.02.2024	Виконано
3.	Аналіз предметної області	22.02.2024-28.02.2024	Виконано
4.	Розробка та обґрунтування завдання	01.03.2024-14.03.2024	Виконано
5.	Розробка концептуальної моделі	15.03.2024-21.03.2024	Виконано
6.	Розробка алгоритму	22.03.2024-04.04.2024	Виконано
7.	Проектування програмної системи	05.04.2024-11.04.2024	Виконано
8.	Розробка інтерфейсу програмної системи	12.04.2024-18.03.2024	Виконано
9.	Тестування програмної системи	19.04.2024-25.04.2024	Виконано
10.	Оформлення пояснювальної записки	26.04.2024-09.05.2024	Виконано
11.	Захист кваліфікаційної роботи	15.06.2024	Виконано

Здобувач вищої освіти _____ Шевченко М.А.
(підпис) (прізвище та ініціали)

Керівник роботи _____ Жарікова М.В.

(підпис)

(прізвище та ініціали)

РЕФЕРАТ

Кваліфікаційна робота бакалавра: 152 сторінки, 54 рисунка, 3 таблиці, 1 додаток, 47 джерел.

Мета роботи – підвищення якості вихідних зображень та ефективності підбору параметрів генерації при зменшенні необхідних обчислювальних ресурсів, що задіяні у процесі генерації зображень із використанням моделей латентної дифузії.

Об’єкт дослідження – процес генерації зображень із використанням моделей латентної дифузії.

Предмет дослідження – інформаційна та структурна моделі процесу генерації зображень з використанням моделі латентної дифузії та налаштування параметрів їх використання в практичних завданнях генеративного мистецтва.

Методи дослідження – методи теоретичного моделювання, теоретичні основи процесу генерації зображень із використанням моделей латентної дифузії, метод експертної оцінки отриманих зображень.

Новизна роботи полягає в тому, що застосовано модель латентної дифузії при генерації зображень, що дозволило покращити якість генерованих зображень та пришвидшити робочий процес, прискоривши етап підбору оптимальних параметрів генерації при зменшенні необхідної обчислювальної потужності.

Практична цінність результатів роботи полягає у тому, що сформовані у роботі рекомендації та вказівки дозволяють користувачам покращити якість генерованих зображень та оптимізувати процес підбору параметрів генерації із зменшенням кількості необхідних обчислювальних ресурсів при генерації зображень з використанням моделей латентної дифузії у завданнях генеративного мистецтва.

Ключові слова: *моделі латентної дифузії, машинне навчання, генеративні моделі, генеративне мистецтво, генерація зображень, оптимізація параметрів генерації, Stable Diffusion, ControlNet.*

АНОТАЦІЯ

Кваліфікаційна робота бакалавра складається зі вступу, трьох розділів, висновку, списку використаних джерел та додатку.

Роботу присвячено дослідженню моделей латентної дифузії у контексті генеративного мистецтва із метою глибокого аналізу та розуміння технічних аспектів роботи даного класу генеративних моделей та формування переліку рекомендацій для налаштування параметрів, які використовуються моделлю латентної дифузії під час генерації зображень із метою підвищення їх якості та швидкості генерації.

Результати дослідження подано у вигляді, що дозволяє побачити, оцінити та використати отримані рекомендації у процесі генерації зображень із використанням моделей латентної дифузії.

В результаті роботи сформовано комплекс конкретних рекомендацій для оптимізації процесу налаштування параметрів генерації при роботі із моделями латентної дифузії з метою покращення якості генерованих зображень та ефективності використання обчислювальних ресурсів, задіяних під час генерації. Отримані рекомендації є універсальними для будь-яких систем генерації зображень, які використовують моделі латентної дифузії.

Результати роботи можуть бути практично використані у задачах генеративного мистецтва. Вони дозволяють підвищити якість вихідних зображень, оптимізувати та пришвидшити робочий процес та зменшити кількість необхідних обчислювальних ресурсів при роботі із моделями латентної дифузії у задачах генеративного мистецтва.

Їх впровадження надасть змогу скоротити витрати на необхідні обчислювальні ресурси та збільшити якість генерованих зображень за рахунок оптимізації процесу підбору оптимальних параметрів генерації при роботі із моделями латентної дифузії у задачах генеративного мистецтва.

ABSTRACT

The bachelor's thesis consists of an introduction, three chapters, a conclusion, a list of sources and an appendix.

The thesis is devoted to the study of latent diffusion models in the context of generative art with the aim of in-depth analysis and understanding of the technical aspects of the study of this class of generative models and the formation of a list of recommendations for setting the parameters used by the latent diffusion model during image generation in order to improve their quality and generation speed.

The results of the work are presented in a form that allows you to see, evaluate and use the received recommendations in the process of image generation using latent diffusion models.

As a result of the work, a set of specific recommendations was formed to optimize the process of setting the generation parameters when working with latent diffusion models in order to improve the quality of the generated images and the efficiency of the use of computing resources used during generation. The resulting recommendations are universal for any imaging systems that use latent diffusion models.

The results can be practically used in the tasks of generative art. They allow you to improve the quality of the output images, optimize and speed up the work process, and reduce the amount of necessary computing resources when working with latent diffusion models in generative art tasks.

Their implementation will make it possible to reduce the costs of necessary computing resources and increase the quality of generated images by optimizing the process of selecting optimal generation parameters when working with latent diffusion models in the tasks of generative art.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	9
ВСТУП	10
РОЗДІЛ 1. АНАЛІЗ ПОТОЧНИХ ЗАСОБІВ ТА ІНСТРУМЕНТІВ У СФЕРІ ГЕНЕРАТИВНОГО МИСТЕЦТВА	13
1.1. Основи генерації зображень	13
1.1.1. Введення до генерації зображень	13
1.1.2. Історичний огляд розвитку генерації зображень	14
1.1.3. Аналіз ключових досягнень у галузі генерації зображень	16
1.2. Сучасні підходи до генерації даних	17
1.2.1. Генеративні змагальні мережі (GAN)	17
1.2.2. Моделі, засновані на дифузії	20
1.2.3. Трансформери (Transformers)	23
1.2.4. Варіаційні автоенкодера (VAE)	24
1.3. Огляд популярних систем генерації зображень	26
1.3.1. Stable Diffusion	26
1.3.2. MidJourney	27
1.3.3. DALL-E 2	28
1.3.4. DeepDream	29
1.3.5. Nvidia GauGAN	30
1.4. Проблематика вирішення завдань генеративного мистецтва	32
1.5. Висновок до першого розділу	33
РОЗДІЛ 2. ДОСЛІДЖЕННЯ МОДЕЛЕЙ ЛАТЕНТНОЇ ДИФУЗІЇ	35
2.1. Моделі, засновані на дифузії	35
2.2. Латентні дифузійні моделі	36

	10
2.3. Архітектура моделей латентної дифузії	37
2.3.1. Варіаційні автоенкодери (VAE)	39
2.3.2. Модель U-Net	40
2.3.3. Модель CLIP Text Encoder	42
2.4. Навчання моделей латентної дифузії	44
2.5. Генерація зображень із текстового опису	48
2.6. Основні параметри генерації	51
2.7. Особливості моделей латентної дифузії	61
2.8. Обмеження та вимоги використання моделей латентної дифузії	62
2.9. Функції моделей латентної дифузії та розширення, сумісні з ними	65
2.9.1. Text-To-Image	65
2.9.2. Image-To-Image	66
2.9.3. Inpainting	67
2.9.4. ControlNet	69
2.10. Висновок до другого розділу	71
РОЗДІЛ 3. ПРАКТИЧНЕ ВИКОРИСТАННЯ МОДЕЛЕЙ ЛАТЕНТНОЇ ДИФУЗІЇ ТА АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ	74
3.1. Способи практичного використання дифузійних моделей	74
3.1.1. Google Colab	74
3.1.2. Онлайн-сервіси	82
3.1.3. Stable Diffusion WebUI	82
3.2. Експериментальне дослідження та аналіз результатів	83
3.2.1. Дослідження впливу параметрів генерації	84
3.2.2. Створення зображень на основі текстового опису	100
3.2.3. Створення зображень на основі вхідного зображення	102

	11
3.2.4. Керування відтворенням зображення	104
3.2.5. Пост-обробка згенерованих зображень	108
3.3. Оптимізація параметрів генерації зображень при роботі з моделями латентної дифузії	114
3.3.1. Вибір моделі (Checkpoint)	115
3.3.2. Текстовий опис (Prompt)	117
3.3.3. Запобігання небажаним зображенням (Negative prompt)	118
3.3.4. Керування відповідністю текстовому опису (CFG Scale)	119
3.3.5. Кількість кроків очищення шуму (Sampling steps)	120
3.3.6. Алгоритми очищення шуму (Sampling methods)	121
3.3.7. Розмір зображення (Image size)	123
3.3.8. Значення початкового шуму (Seed)	123
3.3.9. Значення доданого шуму (Denoising strength)	124
3.4. Висновок до третього розділу	125
ВИСНОВКИ	128
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	131
ДОДАТОК А. ЛІСТИНГ ПРОГРАМИ	137

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

МН – машинне навчання

МД – моделі дифузії

МЛД – моделі латентної дифузії

ПЗ – програмне забезпечення

LDM – latent diffusion model

SD – Stable Diffusion

UI – User Interface

GPU - Graphics processing unit

CU – computing units

ВСТУП

Генерація зображень з використанням алгоритмів та моделей машинного навчання – це одна з ключових тем, яка зараз знаходиться у центрі уваги науковців та творців. На даний момент існує множина методів та підходів, які можуть бути використані у задачах генеративного мистецтва, однак, усі вони мають різну якість вихідних результатів та ефективність відносно до задіяних обчислювальних ресурсів.

Тож вибір інструментів та методів вирішення завдань генеративного мистецтва є дуже важливим етапом, на який впливають кілька факторів, такі як якість та швидкість генерації зображень, вимоги до технічних характеристик обчислювальної машини та ефективність їх використання. Вибір генеративної моделі залежить від бюджету, технічних характеристик ЕОМ та вимог до генерованих зображень.

Технічні вимоги до обчислювальної машини, на якій виконується генерація зображень, є ключовим обмежуючим фактором. Абсолютна більшість наявних генеративних моделей потребує для своєї роботи значної обчислювальної потужності, такої як, наприклад, обсяг відеопам'яті, швидкість ядра та кількість одиниць обчислення графічного процесора, які зазвичай наявні тільки на потужних комерційних серверах, що робить їх використанням коштовним, а процес налаштування параметрів генерації менш гнучким та доступним для користувача.

Через наведені фактори, до появи моделей латентної дифузії, масове самостійне дослідження та використання генеративних моделей не було доступним та зручним для звичайного користувача. Моделі латентної дифузії стали інноваційним рішенням у завданнях генеративного мистецтва через ряд таких переваг, як висока якість вихідних результатів, гнучкість налаштування та їх нова архітектура, яка призвела до суттєвого зниження вимог до технічних характеристик ЕОМ та збільшення ефективності їх використання у обчислювальних процесах, задіяних під час генерації.

Метою роботи є підвищення якості вихідних зображень та ефективності підбору параметрів генерації при зменшенні необхідних обчислювальних ресурсів, що задіяні у процесі генерації зображень із використанням моделей латентної дифузії.

Це потребує дослідження ефективності моделей латентної дифузії, як інструменту для створення якісних зображень у завданнях генеративного мистецтва.

Для досягнення мети дослідження необхідно розв'язати наступні задачі:

1. Провести аналіз існуючих технологій, які дозволяють виконувати генерацію зображень (GAN, VAE, дифузійні моделі).
2. Здійснити порівняння моделей латентної дифузії із іншими генеративними моделями та дослідити переваги використання даного класу моделей, відносно інших існуючих варіантів.
3. Теоретично дослідити та проаналізувати принципи роботи моделей латентної дифузії, їх архітектуру, алгоритми навчання та використання, параметри генерації та можливості застосування даного класу генеративних моделей у задачах генеративного мистецтва.
4. Провести експериментальні дослідження процесу генерації зображень із використанням моделей латентної дифузії та проаналізувати отримані під час практичного дослідження результати.
5. Сформулювати перелік рекомендації для оптимального налаштування параметрів під час генерації зображень із використанням моделей латентної дифузії, які дозволять покращити якість вихідних зображень, пришвидшити та оптимізувати робочий процес та зменшити кількість необхідних обчислювальних ресурсів.

Об'єктом дослідження є процес генерації зображень із використанням моделей латентної дифузії.

Предметом дослідження є інформаційна та структурна моделі процесу роботи моделі латентної дифузії, дослідження теоретико-прикладних засад створення та функціонування даних моделей

та основи налаштування їх параметрів генерації для використання в практичних завданнях генеративного мистецтва.

Методи дослідження базуються на основних принципах роботи моделей латентної дифузії. Використано методи теоретичного моделювання, теоретичні основи процесу генерації зображень із використанням моделей латентної дифузії, метод оцінки отриманих зображень людськими експертами та практичне використання моделей латентної дифузії

Новизна роботи полягає в тому, що отримав подальший розвиток напрямок генеративного мистецтва, що базується на застосуванні моделей латентної дифузії при генерації зображень. Запропонований підхід дозволяє покращити якість генерованих зображень та пришвидшити робочий процес, прискоривши етап підбору оптимальних параметрів генерації при зменшенні необхідної кількості обчислювальної потужності.

Практичне значення отриманих в роботі результатів полягає у тому, що сформовані у роботі рекомендації та вказівки дозволяють користувачам покращити якість генерованих зображень та оптимізувати процес підбору параметрів генерації із зменшенням кількості необхідних обчислювальних ресурсів при генерації зображень зі використанням моделей латентної дифузії у завданнях генеративного мистецтва

Структура та обсяг кваліфікаційної роботи. Відповідно до мети, завдань і предмета дослідження, кваліфікаційна робота складається з реферату, вступу, трьох основних розділів і висновків, списку використаних джерел та додатків.