

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра комп'ютерної інженерії

Павловський Тарас Мирославович

**Генерування гістологічних зображень на основі
стохастичних диференціальних рівнянь / Generation
histological images based on stochastic differential
equations**

спеціальність: 123 - Комп'ютерна інженерія
освітньо-професійна програма - Комп'ютерна інженерія
Кваліфікаційна робота

Виконав студент групи КІм-21
Т. М. Павловський

Науковий керівник:
д.т.н., проф. О.М. Березький

ТЕРНОПІЛЬ – 2025

АНОТАЦІЯ

Павловський Т.М. Генерування гістологічних зображень на основі стохастичних диференціальних рівнянь. – Рукопис. Кваліфікаційна робота на здобуття освітнього ступеня «магістр» за спеціальністю 123 «Комп'ютерна інженерія», освітньо-професійна програма. Західноукраїнський національний університет, Тернопіль, 2025.

Робота написана обсягом 95 сторінок і містить 12 ілюстрацій, 7 таблиць, 1 додаток та 49 джерел за переліком посилань.

Метою роботи є дослідження дифузійних моделей та алгоритмів генерування гістологічних зображень і розроблення алгоритмів синтезу гістологічних зображень для підвищення якості штучних зображень.

Методи досліджень базуються на використанні математичного аналізу, теорії імовірностей та математичній статистиці, теорії нейронних мереж, теорії алгоритмів та методах об'єктно-орієнтованого програмування.

Розроблено алгоритми генерування гістологічних зображень. Здійснено програмну реалізацію дифузійних алгоритмів синтезу гістологічних зображень в середовищі Stable Diffusion. Також зроблено оцінку якості синтезованих зображень на основі метрик IS, FID.

Результати роботи можуть бути використані для задач класифікації біомедичних зображень у системах автоматичного діагностування.

Можливими напрямками подальших досліджень є дослідження алгоритмів генерування зображень, які базуються на інших моделях генерування.

КЛЮЧОВІ СЛОВА: СИНТЕЗ, ГІСТОЛОГІЧНІ ЗОБРАЖЕННЯ, ДИФУЗІЙНІ МОДЕЛІ, МЕТРИКА FID, МЕТРИКА IS.

ANNOTATION

Pavlovskyi M.O. Generation histological images based on stochastic differential equations. – Manuscript. A qualification thesis for obtaining the Master’s degree in the specialty 123 “Computer Engineering,” educational-professional program. West Ukrainian National University, Ternopil, 2025.

The thesis is 95 pages in length and includes 12 illustrations, 7 tables, 1 appendice, and 49 references.

The aim of the work is to study diffusion models and algorithms for generating histological images, as well as to develop algorithms for the synthesis of histological images in order to improve the quality of artificial images.

The research methods are based on the use of mathematical analysis, probability theory and mathematical statistics, neural network theory, algorithm theory, and object-oriented programming methods.

Algorithms for generating histological images have been developed. A software implementation of diffusion algorithms for the synthesis of histological images has been carried out in the Stable Diffusion environment. An assessment of the quality of the synthesized images has also been performed based on the IS and FID metrics.

The results of the work can be used for biomedical image classification tasks in automatic diagnostic systems.

Possible directions for further research include the study of image generation algorithms based on other generative models.

KEYWORDS: SYNTHESIS, HISTOLOGICAL IMAGES, DIFFUSION MODELS, FID METRIC, IS METRIC.

ЗМІСТ

ВСТУП	5
1 АНАЛІЗ ЗАСОБІВ СИНТЕЗУ БІОМЕДИЧНИХ ЗОБРАЖЕНЬ.....	8
1.1 Синтез зображень на основі стохастичних диференціальних рівнянь .	8
1.2 Аналіз гістологічних зображень	13
1.3 Аналіз програмних засобів синтезу зображень.....	18
1.4 Висновки до розділу 1.....	24
2 МОДЕЛІ ТА АЛГОРИТМИ ГЕНЕРУВАННЯ ЗОБРАЖЕНЬ НА ОСНОВІ СТОХАСТИЧНИХ ДИФЕРЕНЦІАЛЬНИХ РІВНЯНЬ.....	26
2.1 Моделі прямого генерування.....	26
2.2 Моделі зворотного генерування.....	27
2.3 Алгоритми прямого генерування.....	33
2.4 Алгоритми зворотного генерування	36
2.5 Метрики для оцінки подібності зображень	38
2.6 Дифузійна модель синтезу зображень на основі стохастичних диференціальних рівнянь	43
3 СИНТЕЗ ГІСТОЛОГІЧНИХ ЗОБРАЖЕНЬ ЗАСОБАМИ ГЕНЕРАТИВНОГО ІНТЕЛЕКТУ	45
3.1 Програмне середовище Stable Diffusion	45
3.2 Апаратно-програмне забезпечення для проведення комп'ютерних експериментів	59
3.3 Результати комп'ютерних експериментів	63
3.4 Висновки до розділу 3.....	74
ВИСНОВКИ.....	76
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	77
Додаток А	Помилка! Закладку не визначено.
Світлокопії виданих публікацій	Помилка! Закладку не визначено.

ВСТУП

Актуальність дослідження. Проблема точного розпізнавання та класифікації біомедичних зображень залишається однією з центральних для сучасних інтелектуальних діагностичних систем. Методи глибинного навчання, які нині визначають рівень розвитку комп'ютерного аналізу зображень, потребують великих і різноманітних навчальних вибірок. У медичній сфері отримати такі дані складно через обмежений доступ, конфіденційність та високу вартість їх збирання. Це зумовлює потребу в генерації синтетичних зображень, які могли б компенсувати нестачу реальних даних. Генеративні моделі, зокрема дифузійні підходи, відкривають нові перспективи у створенні якісних медичних зображень. Відтак дослідження в напрямі генерації біомедичних даних із використанням дифузійних моделей є надзвичайно важливим і своєчасним.

Мета і завдання дослідження. Метою кваліфікаційної роботи є дослідження дифузійних моделей та алгоритмів генерування гістологічних зображень і розроблення алгоритмів синтезу гістологічних зображень для підвищення якості штучних зображень.

У відповідності із поставленою метою у кваліфікаційній роботі вирішити такі задачі:

- проаналізувати використання дифузійних моделей для синтезу зображень;
- провести аналіз гістологічних зображень;
- провести аналіз підходів до синтезу зображень на основі стохастичних диференціальних рівнянь;
- проаналізувати програмні засоби синтезу зображень;
- дослідити дифузійні моделі та алгоритми генерування зображень на основі стохастичних диференціальних рівнянь;

- розробити алгоритми генерування гістологічних зображень;
- провести обчислювальні експерименти для синтезу гістологічних зображень і оцінити якість синтезованих зображень на основі метрик IS, FID.

Об’єкт дослідження – процес синтезу біомедичних зображень.

Предмет дослідження – дифузійні моделі та алгоритми синтезу зображень.

Методи досліджень. У кваліфікаційній роботі використано математичний аналіз, теорію імовірностей та математичну статистику, теорію нейронних мереж, теорію алгоритмів та методи об’єктно-орієнтованого програмування.

Наукова новизна одержаних результатів полягає в дослідженні дифузійних моделей та алгоритмів генерування зображень, розроблені алгоритму генерування гістологічних зображень.

Практичне значення одержаних результатів полягає в проведенні обчислювальних експериментів в середовищі Stable Diffusion для синтезу гістологічних зображень заданої якості.

Апробація роботи. За результатами досліджень опубліковані двоє тезів доповідей на III Всеукраїнській науково-практичній конференції молодих вчених і студентів «Інтелектуальні комп’ютерні системи та мережі» (25 листопада 2025 р., м. Тернопіль, Західноукраїнський національний університет) [1, 2]:

1. Павловський Т.М., Богонос Д.С. Дифузійні моделі для синтезу цитологічних зображень. *Інтелектуальні комп’ютерні системи та мережі* : тези доп. III Всеукр. наук.-практ. конф. студентів, аспірантів, молодих вчених (25 листопада 2025 р.). Тернопіль : ЗУНУ, 2025. С. 154-155.

2. Павловський Т.М., Богонос Д.С. Синтез гістологічних зображень на основі дифузійних моделей. *Інтелектуальні комп’ютерні системи та мережі* : тези доп. III Всеукр. наук.-практ. конф. студентів, аспірантів, молодих вчених (25 листопада 2025 р.). Тернопіль : ЗУНУ, 2025. С. 156-157.

Кваліфікаційна робота написана згідно методичних рекомендацій [3, 4].
Магістерська робота написана в рамках науковому напрямку дослідницької групи «комп'ютерні системи штучного інтелекту» під керівництвом професора Березького О.М. [5-14].

Кваліфікаційна робота складається із трьох розділів, висновків, списку використаної літератури та додатків [4].

У першому розділі показано актуальність синтезу зображень для розширення навчальних вибірок, проаналізовані програмні засоби для генерування зображень, зроблено аналіз гістологічних зображень.

В другому розділі досліджено дифузійні моделі та алгоритми синтезу зображень на основі стохастичних диференціальних рівнянь, розроблено алгоритми синтезу гістологічних зображень.

В третьому розділі описано програмне середовище Stable Diffusion, описано обчислювальні експерименти для синтезу гістологічних зображень і проведено їх оцінку на основі метрик IS, FID.

У додатку приведено світлокопії виданих публікацій.

1 АНАЛІЗ ЗАСОБІВ СИНТЕЗУ БІОМЕДИЧНИХ ЗОБРАЖЕНЬ

1.1 Синтез зображень на основі стохастичних диференціальних рівнянь

Стохастичні диференціальні рівняння (СДР) є фундаментальним математичним інструментом для опису процесів, у яких присутній випадковий компонент або шум. У контексті генеративного штучного інтелекту ці рівняння стали основою для побудови сучасних дифузійних моделей, що відтворюють складні структури, зокрема біомедичні та гістологічні зображення [15]. Дифузійні моделі імітують природні процеси розсіювання та зворотної дифузії, що дозволяє поступово перетворювати випадковий шум у впорядковане, реалістичне зображення з високим рівнем деталізації [16]. Завдяки використанню стохастичних процесів такі моделі здатні точно відтворювати природну варіативність біологічних тканин, що робить їх особливо ефективними для роботи зі складними мікроструктурами, характерними для гістологічних даних. Більш того, математичний опис дифузії через СДР забезпечує контроль над швидкістю додавання та усунення шуму, дозволяючи управляти як глобальною структурою зображення, так і дрібними локальними деталями.

Основна ідея синтезу зображень на основі стохастичних процесів полягає у тому, що генерація здійснюється шляхом поступового моделювання розподілу даних. На початковому етапі модель додає до вихідних зображень шум, який розсіює інформацію, перетворюючи їх у випадковий стан. Далі, у процесі навчання, модель опановує зворотню задачу – як відновити оригінальне зображення із зашумленого. У межах цього процесу нейронна мережа фактично вивчає градієнти щільності розподілу даних, що дозволяє їй точно реконструювати морфологічні закономірності тканин і повторювати характерні патерни забарвлення. Такий підхід дозволяє створювати нові

зображення, які за статистичними характеристиками подібні до реальних, навіть якщо вони ніколи не з'являлися у вихідному наборі даних. Крім того, стохастична природа дифузійного процесу забезпечує високу різноманітність результатів, що є критично важливим для формування великих аугментованих наборів гістологічних даних.

Математично стохастичні диференціальні рівняння описують зміну стану системи під впливом як детермінованих, так і випадкових факторів. Загальна форма СДР має вигляд:

$$dx = f(x,t)dt + g(x,t)dW_t,$$

де x – вектор стану, $f(x,t)$ – детермінована частина, що описує закономірну динаміку системи, $g(x,t)$ – функція інтенсивності шуму, а W_t – вінерівський процес, який моделює випадкові флуктуації. У випадку дифузійних моделей ці рівняння визначають, як інформація «розсіюється» у просторі даних і як вона може бути відновлена шляхом зворотного інтегрування рівняння у часі.

Процес генерації зображень складається з двох етапів:

1. Прямий процес (forward diffusion) – до зображення послідовно додається шум на кожному кроці, доки воно не перетворюється у чистий гаусівський шум. Цей процес визначається як випадковий рух частинок у латентному просторі, що підпорядковується СДР.

2. Зворотний процес (reverse diffusion) – модель навчається відновлювати дані, поступово усуваючи шум. На кожному кроці вона використовує апроксимацію градієнта щільності розподілу даних, щоб наблизитися до реального зображення.

Стохастичний характер процесу дозволяє моделі створювати необмежену кількість варіацій зображень, які залишаються статистично узгодженими з вихідним розподілом. Саме тому дифузійні моделі набули широкого застосування для синтезу складних біомедичних структур, де навіть

невеликі варіації в текстурі або морфології клітин мають значення для досліджень і навчання [17].

Одним із ключових підходів, який базується на СДР, є дифузійне моделювання з використанням імовірнісних методів (Diffusion Probabilistic Models). У таких моделях випадковий процес дифузії задається як ланцюг Маркова, де кожен крок зменшує кількість інформації у даних, а потім за допомогою нейронної мережі реалізується зворотний процес. Це дозволяє отримати розподіл даних, максимально наближений до реального, без необхідності змагального навчання, як у генеративних змагальних мережах (GAN) [18].

Додатковою перевагою такого підходу є його висока стабільність: процес навчання дифузійних моделей не страждає від проблеми колапсу моделі, характерної для генеративно-змагальних мереж (GAN), а також не демонструє різких коливань або «вибуху» градієнтів під час оптимізації. Це обумовлено тим, що процес навчання базується на добре сформульованій і математично строго визначеній задачі відновлення шуму, а не на змаганні двох моделей із протилежними цілями. Завдяки цьому дифузійні моделі навчаються більш передбачувано, легко масштабуються на великі набори даних і рідше потребують спеціальних технік стабілізації.

Строга математична формалізація дифузійних процесів дозволяє контролювати кожен етап генерації – від інтенсивності шумових перетворень до швидкості зворотної дифузії. Це забезпечує не лише плавний перебіг синтезу, але й можливість точного відтворення структури даних із контрольованим рівнем деталізації. Користувач або дослідник може впливати на форму, контури, текстуру та морфологічні властивості синтезованих елементів, коригуючи параметри шуму, кількість ітерацій або режими вибірки.

У контексті біомедичної візуалізації така керованість має особливе значення. На відміну від художніх або стилізованих зображень, гістологічні та цитологічні структури містять дрібні деталі, що мають критичне діагностичне навантаження. Навіть мінімальні деформації, неточності контурів клітинних

ядер, змінена форма тканинних структур або порушення природних співвідношень можуть призвести до хибної інтерпретації. Дифузійні моделі, завдяки своїй стабільності та контрольованості, здатні забезпечувати точне відтворення таких особливостей, зберігаючи морфологічну достовірність і природну текстурність тканин.

Крім того, вони дозволяють досягати високої відтворюваності результатів: задаючи однакові параметри та початкові умови (seed), дослідники отримують ідентичні зображення, що критично важливо для валідації методів, порівняльних експериментів і клінічних досліджень. У свою чергу, зміна окремих компонентів процесу дає змогу генерувати широкий спектр реалістичних варіацій тканинних структур, не втрачаючи статистичної відповідності вихідному домену.

Подальший розвиток цієї ідеї призвів до появи моделей, заснованих на стохастичних диференціальних рівняннях (SDE-based models), таких як Score-based Generative Models (SGM) та SDEdit [19]. Вони описують процес дифузії у вигляді безперервного стохастичного потоку, що дозволяє керувати швидкістю і напрямком «руху» точок у просторі даних. Такі підходи дають змогу гнучко контролювати якість, стиль і деталізацію зображення, що є особливо важливим при моделюванні цитологічних зразків, де навіть мікроскопічні зміни у клітинній структурі можуть мати діагностичне значення.

Додатково, моделі на основі SDE забезпечують можливість часткового редагування зображень, коли до початкового зразка додається керований рівень шуму, після чого виконується його реконструкція із збереженням морфологічної структури. Це робить такі методи надзвичайно корисними для створення варіативних навчальних наборів, дослідження рідкісних патологій або формування контрольованих експериментальних сценаріїв, де важливо моделювати зміну окремих клітинних компонентів без втрати загальної анатомічної цілісності.

Для моделювання біомедичних і цитологічних даних стохастичні моделі мають низку переваг:

1. Висока реалістичність синтезованих зображень. Розв'язання СДР дозволяє враховувати природну варіативність біологічних структур.

2. Стійкість до шумів і артефактів. Оскільки модель навчається на зашумлених даних, вона ефективно відновлює деталі навіть у складних випадках.

3. Гнучкість у навчанні. Параметри рівнянь можна адаптувати під різні типи тканин або клітин.

4. Можливість керованого синтезу. Модель може генерувати зображення із заданими характеристиками (наприклад, певна морфологія клітин або рівень патологічних змін).

Використання стохастичних диференціальних рівнянь у генеративному штучному інтелекті забезпечує математично обґрунтований, формально описаний та відтворюваний підхід до створення нових зображень. На відміну від традиційних методів генерації, де процес навчання часто залежав від емпіричних прийомів або нестабільних оптимізаційних стратегій, дифузійні моделі покладаються на чітко визначені стохастичні процеси, що гарантують контрольованість та прогнозованість результатів. Завдяки цьому синтез зображень ґрунтується не лише на емпіричній інтерпретації даних, а й на строгій математичній структурі процесу додавання та усунення шуму.

У порівнянні з попередніми методами, такими як варіаційні автокодувальники чи генеративно-змагальні мережі, дифузійні моделі демонструють значно вищу стабільність навчання: вони не схильні до колапсу моделі, не вимагають точного балансування компонентів, а їхнє навчання менш чутливе до гіперпараметрів. Крім того, вони забезпечують суттєво вищу якість реконструкції складних текстур, коректніше відтворення структури об'єктів та кращу здатність до узагальнення. Це дозволяє створювати не лише візуально реалістичні, а й морфологічно достовірні зразки, що особливо важливо для біомедичних зображень, де більше значення мають точність мікроструктур і збереження діагностично значущих патернів, ніж художня виразність.

1.2 Аналіз гістологічних зображень

Гістологічні зображення є ключовим інструментом у медичній діагностиці, наукових дослідженнях і навчальному процесі, оскільки вони відображають мікроструктуру тканин, дозволяючи виявляти патологічні зміни, оцінювати морфологічні особливості клітин та міжклітинного матриксу, а також визначати стадії розвитку захворювань. Як розділ морфології, гістологія досліджує будову та організацію тканин, їхню архітектуру, клітинний склад і характер взаємодії між структурними елементами, що робить аналіз гістологічних зрізів фундаментальним джерелом медичної інформації. Саме тому точність і чіткість візуалізації є критично важливими для коректної інтерпретації біологічних процесів, диференціації нормальних та патологічних структур і формування обґрунтованих клінічних висновків.

Зазвичай гістологічні зображення отримують за допомогою світлової, електронної або флуоресцентної мікроскопії після попереднього фарбування зразків. Найпоширенішим методом є забарвлення гематоксилином та еозином (H&E), яке забезпечує чітке розмежування ядер (сині або фіолетові) та цитоплазми (рожева), дозволяючи легко ідентифікувати клітинні структури та основні морфологічні патерни. Окрім цього, застосовуються спеціальні методики, такі як Masson Trichrome, PAS, Immunohistochemistry (ІНС), які використовуються для візуалізації сполучної тканини, полісахаридів, білкових маркерів або специфічних клітинних компонентів. Таке фарбування суттєво підвищує контрастність зображення і полегшує подальший аналіз як для лікаря-патологоанатома, так і для алгоритмів машинного навчання.

Проте навіть за наявності стандартизованих методів фарбування гістологічні зображення часто залишаються неоднорідними через варіації освітлення, відмінності у протоколах підготовки зразків, різну товщину зрізів, нерівномірне фокусування або появу оптичних артефактів. Такі фактори ускладнюють автоматизовану обробку, призводять до нестабільності

результатів і потребують додаткових методів нормалізації, стандартизації кольору та фільтрації шумів перед застосуванням алгоритмів комп'ютерного аналізу.

Основною метою обробки гістологічних зображень є сегментація тканинних компонентів, визначення морфометричних характеристик клітин і структур (зокрема, площі ядер, співвідношення паренхіми до стромы, щільності судинної сітки), а також виявлення патологічних змін, таких як некрози, фіброз, проліферативні утворення чи атипія клітин. Для досягнення цих завдань застосовуються як класичні методи комп'ютерного зору (фільтри текстур, порогова сегментація, кластеризація, морфологічні операції), так і сучасні глибокі нейронні мережі, здатні аналізувати високорівневі та приховані патерни в будові тканин. Особливо ефективними є моделі на основі згорткових нейронних мереж (CNN), трансформерні архітектури та сегментаційні моделі типу U-Net, які демонструють високу точність у класифікації клітинних структур, виявленні патологічних ділянок та морфологічній стратифікації зразків. Поєднання цих методів створює основу для побудови автоматизованих систем діагностики, які підвищують точність аналізу та зменшують суб'єктивний вплив людського фактора.

Під час роботи з гістологічними зображеннями виникає низка типових проблем:

по-перше, велика морфологічна варіативність тканин, коли структура клітин значно відрізняється навіть у межах одного органа або типу патології;

по-друге, наявність артефактів, таких як складки зрізу, пил, повітряні бульбашки чи пошкодження скла;

по-третє, оптична неоднорідність, спричинена нерівномірним освітленням або варіаціями кольорового тону під час сканування; і, нарешті, обмежена кількість якісно розмічених даних, оскільки створення таких наборів вимагає участі досвідчених гістопатологів і є надзвичайно трудомістким процесом.

Типовий набір гістологічних зображень містить цифрові мікрофотографії тканинних зрізів у форматах JPG, PNG, TIFF або SVS із роздільною здатністю від 512×512 до 8192×8192 пікселів. Ці дані забезпечують достатню деталізацію для розпізнавання клітинних ядер, стромальних елементів, протоків і судинних структур. Часто одне зображення охоплює кілька тисяч клітин, що потребує ефективних алгоритмів сегментації та класифікації тканинних компонентів. Крім того, великі гістологічні препарати часто мають формат Whole Slide Images (WSI), які можуть досягати десятків гігапікселів та зберігаються у багаторівневій структурі з різними ступенями масштабування. Це вимагає використання спеціалізованих інструментів для обробки, таких як OpenSlide або ASAP, а також застосування патч-базованих методів, коли великі зображення поділяються на фрагменти для локального аналізу. Така стратегія дозволяє ефективно обробляти гігантські обсяги даних, зберігаючи при цьому точність виявлення дрібних клітинних деталей і локальних патологічних змін.

Залежно від морфологічних характеристик гістологічні зразки поділяють на:

- нормальні тканини, які мають впорядковану архітектуру, чіткі клітинні межі, однорідну стромальну структуру;
- доброякісні зміни, що характеризуються локальними порушеннями структури, збільшенням розміру ядер або нерівномірністю забарвлення без ознак злоякісності;
- патологічні (злоякісні) утворення, які відзначаються високою клітинною щільністю, поліморфізмом ядер, гіперхромією та деструкцією тканинної архітектури.

Для підвищення точності аналізу гістологічних даних проводиться попередня обробка (preprocessing), що включає нормалізацію кольору, компенсацію нерівномірного освітлення, фільтрацію шумів, корекцію артефактів сканування та масштабування зображень до єдиного простору. Нормалізація кольору є особливо важливою, оскільки різні лабораторії та протоколи фарбування (зокрема H&E) можуть суттєво відрізнятися, що

ускладнює подальший автоматизований аналіз. Після етапу вирівнювання даних виконується сегментація клітин і тканинних компонентів, яка може здійснюватися за допомогою класичних алгоритмів порогової фільтрації, методів кластеризації, таких як K-means або Gaussian Mixture Models, а також сучасних глибоких архітектур, зокрема U-Net, ResNet та Mask R-CNN [20]. Сегментація дозволяє виділити ядра, цитоплазму, стромальні структури, міжклітинний матрикс або конкретні патологічні утворення, формуючи структуровані дані для подальшого аналізу. На наступному етапі здійснюється класифікація тканинних ділянок, метою якої є розмежування нормальних, проліферативних, запальних, фіброзних та інших типів тканин. Отримані ознаки можуть бути використані не лише для діагностики, але й для побудови прогнозних моделей перебігу захворювання, кількісного морфометричного аналізу та створення великих маркованих датасетів для навчання систем машинного зору.

Сучасні дослідження демонструють високу ефективність використання генеративних моделей у гістології, оскільки вони здатні відтворювати складні морфологічні патерни та мікроструктури тканин із точністю, достатньою для діагностичних і навчальних завдань. Такі моделі можуть синтезувати нові зразки тканин, які зберігають текстурні, колірні та структурні властивості реальних гістологічних зображень, зокрема характер розташування ядер, щільність клітинного складу та особливості стромального матриксу. Це відкриває нові можливості для створення великих навчальних наборів даних для штучних нейронних мереж; балансування класів у медичних датасетах (особливо при рідкісних патологіях, де реальні дані представлені в недостатньому обсязі); формування віртуальних навчальних колекцій для підготовки майбутніх лікарів і патоморфологів; а також тестування та валідації алгоритмів автоматизованої діагностики без залучення реальних пацієнтів. Використання синтетичних гістологічних зображень також дозволяє мінімізувати етичні ризики та зменшити залежність від дорогих лабораторних

процедур, забезпечуючи дослідникам більшу свободу у побудові експериментів і моделюванні рідкісних патологічних станів.

Особливо перспективними виявилися дифузійні моделі, засновані на стохастичних диференціальних рівняннях, які дозволяють генерувати реалістичні гістологічні зображення з високою структурною достовірністю [21]. Завдяки покроковому моделюванню процесу усунення шуму ці моделі здатні відтворювати дрібні морфологічні деталі, включно з формою клітинних ядер, щільністю стромальних елементів та текстурою міжклітинного матриксу. Вони забезпечують контроль над рівнем шуму, контрастом, текстурою й деталізацією клітинних структур, що робить їх цінним інструментом для розширення та покращення медичних баз даних. Крім того, можливість адаптації таких моделей до вузькоспеціалізованих наборів даних дозволяє створювати синтетичні зразки, максимально наближені до конкретних патологічних сценаріїв, що суттєво підвищує ефективність навчання діагностичних алгоритмів та покращує їх узагальнюючу здатність.

Таким чином, аналіз гістологічних зображень є важливою складовою біомедичних досліджень і систем підтримки клінічних рішень. Використання методів комп'ютерного зору, глибокого навчання та генеративного штучного інтелекту забезпечує новий рівень автоматизації, точності та надійності аналізу тканин. Дифузійні моделі відкривають можливості для створення реалістичних синтетичних гістологічних зразків, які можуть стати основою для навчання, тестування та розвитку інтелектуальних діагностичних систем у медицині. Крім того, сучасні дослідження демонструють, що поєднання дифузійних моделей із методами доменної адаптації та самонавчання дозволяє досягати ще більшої відповідності синтетичних даних реальним зразкам, що підсилює їх практичну цінність для клінічних застосувань.

1.3 Аналіз програмних засобів синтезу зображень

У сучасних дослідженнях генеративного штучного інтелекту важливу роль відіграють програмні засоби, що реалізують процес синтезу зображень на основі дифузійних та нейронних моделей. Такі системи дозволяють не лише створювати реалістичні візуальні об'єкти, але й моделювати складні структури, зокрема цитологічні або гістологічні зразки. Для вибору оптимального інструмента необхідно враховувати особливості архітектури моделі, можливість навчання на власних даних, вимоги до апаратного забезпечення та ступінь відкритості програмного коду. Особливо важливим є також рівень керованості генеративного процесу, підтримка розширень та сумісність з науковими фреймворками, адже від цього залежить точність синтезу, здатність моделі відтворювати дрібномасштабні морфологічні деталі та можливість інтеграції в дослідницькі пайплайни.

До найпоширеніших програмних засобів, що використовуються для генерації біомедичних зображень, належать Stable Diffusion, ComfyUI, DALL-E та MidJourney. Кожен із них має свої унікальні переваги та обмеження залежно від цілей дослідження, типу даних і доступних обчислювальних ресурсів. Stable Diffusion та його інтерфейси на кшталт ComfyUI вирізняються відкритістю коду, можливістю донавчання на спеціалізованих медичних наборах даних та високим рівнем контролю над параметрами дифузійного процесу, що робить їх придатними для наукового застосування. Натомість DALL-E та MidJourney орієнтовані переважно на художню та креативну генерацію, мають закриту архітектуру та обмежені можливості адаптації до біомедичних задач, проте демонструють високу візуальну якість і узгодженість між текстовими описами та сформованими зображеннями.

Stable Diffusion – це модель глибокого навчання, що базується на дифузійних процесах і реалізує генерацію зображень із текстових описів або заданих шаблонів. Архітектура моделі складається з варіаційного

автокодувальника (VAE), блоку U-Net та текстового енкодера CLIP. VAE стискає вхідне зображення в латентний простір, де здійснюється моделювання шуму, а U-Net виконує зворотну дифузію, поступово відновлюючи структуру зображення. Механізм перехресної уваги (cross-attention) забезпечує зв'язок між текстом і візуальними характеристиками.

Завдяки роботі в латентному просторі (latent space) Stable Diffusion істотно знижує обчислювальні витрати порівняно з моделями, що виконують дифузію у піксельному просторі. Це дозволяє генерувати зображення високої роздільної здатності на звичайних GPU із 6–8 GB відеопам'яті. Крім того, архітектура підтримує механізми донавчання (LoRA, DreamBooth, Hypernetwork), що дає змогу адаптувати модель до спеціалізованих задач, зокрема медичних, де важливо зберегти морфологічну точність структури та унікальні текстурні патерни. Таким чином, Stable Diffusion є одним із найбільш універсальних і гнучких генеративних інструментів, який поєднує високу якість синтезу, продуктивність та масштабованість для наукових і прикладних досліджень.

Stable Diffusion відзначається такими характеристиками:

- відкритий код і доступність ваг моделі для користувачів;
- можливість навчання на власних наборах даних, у тому числі медичних;
- висока деталізація зображень (до 1024×1024 пікселів і більше);
- підтримка локального запуску без необхідності використання хмарних сервісів;
- розширюваність за допомогою плагінів (DreamBooth, LoRA, Textual Inversion) для адаптації до спеціалізованих задач.

Завдяки відкритій архітектурі Stable Diffusion активно використовується в наукових і медичних дослідженнях, зокрема для генерації цитологічних зображень, моделювання клітинних структур, а також створення аугментованих наборів даних для навчання систем діагностики.

Comfy UI - це графічний інтерфейс, побудований на основі Stable Diffusion, який дозволяє користувачам візуально налаштовувати процес генерації. Система використовує модульний підхід із побудовою робочих процесів у форматі node-based графів, де кожен блок (node) відповідає за окрему операцію: завантаження моделі, генерацію шуму, виконання дифузійних кроків, обробку текстових підказок, постобробку зображення тощо. Така структура дає змогу гнучко конструювати власні пайплайни з урахуванням специфіки дослідницької задачі, варіювати алгоритми вибірки, роздільну здатність, параметри шуму й методи донавчання моделей.

На відміну від класичного WebUI, Comfy UI дозволяє візуалізувати повний ланцюг обчислень і експериментувати з різними архітектурними конфігураціями без необхідності написання коду. Це робить платформу особливо зручною для наукових експериментів у галузі біомедичної візуалізації, де потрібно тестувати різні варіанти дифузійних процесів, комбінувати моделі або застосовувати кастомні алгоритми обробки гістологічних та цитологічних структур. Завдяки підтримці великих моделей (SDXL, SD3, Flux) і можливості інтеграції зовнішніх скриптів Comfy UI є потужним середовищем для створення високоточних та контрольованих генеративних рішень.

Система працює за принципом блокового моделювання (node-based), що дає змогу створювати складні пайплайни генерації без написання коду.

Основні переваги Comfy UI:

- гнучка структура для побудови власних сценаріїв синтезу;
- можливість візуальної інтеграції етапів дифузії, шуму, обробки й стилізації;
- підтримка великих моделей (SD XL, SD 3.0, Flux) і зовнішніх розширень;
- зручність для навчальних цілей і демонстрацій генеративних процесів.

Comfy UI особливо ефективна для експериментів із гістологічними даними, коли потрібно дослідити вплив різних параметрів дифузії (кількість кроків, рівень шуму, масштаб контенту) на форму клітинних структур чи текстуру тканини.

DALL-E - це пропрієтарна система, розроблена компанією OpenAI, що використовує трансформерну архітектуру для генерації зображень із текстових підказок. Модель функціонує на основі багатомодальної інтеграції тексту та візуальної інформації, завдяки чому здатна створювати складні композиції, поєднувати абстрактні концепції та формувати семантично узгоджені результати високої якості. На відміну від відкритих рішень, DALL-E не підтримує локальне навчання на власних наборах медичних даних, що обмежує її застосування у вузькоспеціалізованих наукових задачах, але гарантує стабільність, передбачуваність та високу якість генерації у загальних сценаріях.

DALL-E застосовується переважно у творчих ілюстраціях, прототипуванні дизайнів та генерації унікальних художніх стилізацій, однак може бути корисною і в біомедичній сфері – передусім для створення узагальнених або схематичних гістологічних та цитологічних зображень, які використовуються у презентаціях, навчальних посібниках або демонстраційних матеріалах. Завдяки здатності точно інтерпретувати текстові описи модель дає змогу формувати пояснювальні схеми, стилізовані репрезентації клітинних структур та інші допоміжні візуалізації, що сприяють підвищенню якості подачі інформації в освітніх та науково-популярних матеріалах.

Особливості:

- закритий код і робота виключно через API;
- обмежені можливості кастомізації;
- висока фотореалістичність при генерації загальних об'єктів, але слабша точність у наукових структурах [22].

MidJourney – онлайн-платформа для генерації художніх зображень, орієнтована передусім на креативні та дизайнерські завдання. Модель використовує пропрієтарні алгоритми глибинного навчання, які забезпечують високу художню експресивність, складні стилістичні композиції та характерну естетику, що й сприяло її широкій популярності серед дизайнерів, ілюстраторів та дослідників візуальних форм.

Система пропонує велику кількість художніх стилів, точні механізми роботи з текстовими підказками та можливість швидкої генерації якісних зображень, проте її функціональність принципово орієнтована на творчі завдання, а не на наукове моделювання. У контексті біомедичних задач MidJourney може бути корисною для створення стилізованих концептів клітинних структур, ілюстративних матеріалів або узагальнених візуальних схем, що застосовуються у презентаціях, освітніх курсах та комунікаційних матеріалах.

Однак відсутність можливості навчання моделі на власних медичних даних, обмежений контроль над внутрішніми параметрами генерації та закритість архітектури значно звужують її застосування в наукових дослідженнях. Через це MidJourney не може забезпечити точного відтворення гістологічних чи цитологічних структур і є менш придатною для задач моделювання, аугментації або аналізу даних у біомедичній галузі порівняно з відкритими дифузійними моделями. До того ж результати, отримані за допомогою MidJourney, орієнтовані переважно на художній стиль, а не на структурну відповідність реальним біологічним об'єктам. Це призводить до появи артефактів, стилізованих деталей або спрощень, які неприпустимі для медичних задач, де навіть мінімальні відхилення у формі ядра чи текстурі тканини можуть суттєво вплинути на діагностичну інтерпретацію. Унаслідок цього MidJourney розглядається радше як інструмент ілюстративної або концептуальної візуалізації.

В таблиці 1.1 приведено порівняльну характеристику програмних засобів.

Таблиця 1.1 – Порівняльна характеристика програмних засобів

Критерій	Stable Diffusion	Comfy UI	DALL-E	MidJourney
Доступність	Відкритий код	Відкритий код (надбудова)	Закритий, API	Пропріетарний
Можливість навчання	Так (власні дані, LoRA, DreamBooth)	Так (через Stable Diffusion)	Ні	Ні
Графічний інтерфейс	Базовий (UI через AUTOMATIC1111)	Розширений блоковий інтерфейс	Інтернет-інтерфейс API	Веб-інтерфейс (Discord)
Якість зображень	Висока деталізація	Висока, залежно від моделі	Висока фотореалістичність	Художній стиль
Використання в біомедицині	Так, ефективно для синтезу та аугментації цитологічних даних	Так, зручне експериментальне середовище	Обмежено	Обмежено
Обчислювальні ресурси	GPU \geq 6 GB VRAM	GPU \geq 8 GB VRAM	Хмарні	Хмарні
Тип ліцензії	Creative ML OpenRAIL-M	Відкрита (на базі SD)	Комерційна	Комерційна

1.4 Висновки до розділу 1

У цьому розділі було проведено системний аналіз сучасних методів, математичних моделей та програмних засобів, що використовуються для синтезу біомедичних і цитологічних зображень. Особливу увагу приділено дифузійним підходам, які базуються на стохастичних диференціальних рівняннях і становлять основу новітніх генеративних технологій у галузі штучного інтелекту. Дослідження показало, що стохастичні диференціальні рівняння забезпечують математично строгий опис процесу генерації зображень через послідовне додавання та усунення шуму. Такий підхід дозволяє моделі формувати латентний простір із високим ступенем структурної організованості, у якому відбувається плавне відновлення даних, що в підсумку забезпечує реалістичність, відтворюваність та контрольованість синтезованих візуальних структур.

Завдяки цьому підходу моделі здатні відновлювати складні структури та текстури з високим рівнем деталізації. Використання стохастичних процесів дає змогу імітувати природну варіативність біологічних об'єктів, що робить їх надзвичайно ефективними для генерації медичних зображень, зокрема цитологічних препаратів. Крім того, дифузійні моделі дозволяють контролювати рівень шуму, ступінь згладженості чи деталізації, що дає змогу адаптувати процес синтезу під конкретні діагностичні чи дослідницькі потреби. Завдяки такій гнучкості штучно згенеровані зразки можуть відтворювати як типові морфологічні патерни, так і рідкісні варіації, які важко отримати в реальних умовах, що значно розширює можливості моделювання патологічних процесів і підвищує ефективність навчання алгоритмів медичної діагностики.

У результаті аналізу цитологічних зображень було встановлено, що вони містять високу морфологічну складність, значну варіативність клітинних структур і чутливі до змін у методах фарбування, освітлення та фокусування

[23]. Ці фактори ускладнюють автоматизовану обробку й потребують великих кількостей якісно розмічених даних для навчання моделей глибокого навчання. Використання синтетичних зображень, отриманих за допомогою дифузійних моделей, дозволяє компенсувати нестачу даних, збалансувати вибірки та покращити якість класифікації патологічних клітин. Проведений огляд програмних засобів синтезу зображень засвідчив, що найбільш придатними для наукових та медичних завдань є відкриті системи, такі як Stable Diffusion та Comfy UI. Вони забезпечують гнучке налаштування параметрів генерації, можливість додаткового навчання на власних наборах даних, а також високу якість синтезованих результатів [24]. У свою чергу, пропрієтарні рішення DALL-E та MidJourney, хоч і демонструють вражаючу візуальну якість, мають обмежене застосування в біомедичних дослідженнях через закритість і відсутність адаптації до медичних даних.

Таким чином, можна зробити висновок, що дифузійні моделі на основі стохастичних диференціальних рівнянь є перспективним напрямом у синтезі гістологічних зображень. Вони забезпечують високу реалістичність, контрольованість і відтворюваність результатів, що робить їх ефективним інструментом для створення навчальних вибірок, тестування алгоритмів сегментації, а також розробки інтелектуальних систем підтримки медичних рішень у гістологічній діагностиці.

2 МОДЕЛІ ТА АЛГОРИТМИ ГЕНЕРУВАННЯ ЗОБРАЖЕНЬ НА ОСНОВІ СТОХАСТИЧНИХ ДИФЕРЕНЦІАЛЬНИХ РІВНЯНЬ

2.1 Моделі прямого генерування

Зашумлення даних на багатьох масштабах шуму є основною ідеєю для методів генерування зображень. Узагальнимо цю ідею до нескінченного числа масштабів шуму, так щоб зашумлені розподіли даних еволюціонували відповідно до стохастичного диференціального рівняння (SDE) у міру посилення шуму.

Наша мета – побудувати дифузійний процес $\{\mathbf{x}(t)\}_{t=0}^T$, індексований неперервною часовою змінною $t \in [0, T]$, такий що $\mathbf{x}(0) \sim p_0$, для якого ми маємо вибірку незалежних ідентично розподілених даних, а також $\mathbf{x}(T) \sim p_T$, для якого існує зручна форма для ефективною генерації вибірок. Іншими словами, p_0 – це розподіл даних, а p_T – це апіорний розподіл. Цей дифузійний процес можна моделювати як розв’язок стохастичного диференціального рівняння СДР:

$$d\mathbf{x} = f(\mathbf{x}, t)dt + g(t)d\mathbf{w}. \quad (2.1)$$

де \mathbf{w} – це стандартний процес Вінера (тобто броунівський рух);

$f(\cdot, t): R^d \rightarrow R^d$ – векторнозначна функція, що називається коефіцієнтом дрейфу процесу $\mathbf{x}(t)$;

$g(\cdot): R \rightarrow R$ – скалярна функція, відома як коефіцієнт дифузії процесу $\mathbf{x}(t)$.

Для спрощення викладу ми припускаємо, що коефіцієнт дифузії є скаляром (а не матрицею розміру $d \times d$) і не залежить від \mathbf{x} . СДР має єдиний сильний розв’язок за умови, що коефіцієнти є глобально ліпшицевими як щодо стану, так і щодо часу [25].

Далі ми позначаємо через $p_t(\mathbf{x})$ густину ймовірності процесу $\mathbf{x}(t)$, а через $p_{st}(\mathbf{x}(t)|\mathbf{x}(s))$ – ядро переходу від $\mathbf{x}(s)$ до $\mathbf{x}(t)$, де $0 \leq s < t \leq T$.

Зазвичай p_T – це неструктурований апіорний розподіл, який не містить інформації про p_0 , наприклад, гаусівський розподіл із фіксованими середнім та дисперсією. Існують різні способи побудови SDE у рівнянні (2.1) таким чином, щоб воно розмивало розподіл даних у фіксований апіорний розподіл.

2.2 Моделі зворотного генерування

Почавши з вибірок $\mathbf{x}(T) \sim p_T$ та виконавши зворотний хід процесу, ми можемо отримати вибірки $\mathbf{x}(0) \sim p_0$. Результат Андерсона [26] стверджує, що обернений (у часі) дифузійний процес також є дифузійним процесом, який "працює назад у часі" і задається зворотним СДР:

$$d\mathbf{x} = [f(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})] dt + g(t) d\bar{\mathbf{w}}, \quad (2.2)$$

де $\bar{\mathbf{w}}$ – стандартний вінерівський процес у випадку, коли час тече назад від T до 0,

dt – нескінченно малий від'ємний крок часу.

Коли скор-функція (градієнт лог-щільності) для кожного маргінального розподілу $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ відома для всіх t , ми можемо побудувати зворотній дифузійний процес за рівнянням (2.2) і симулювати його, щоб отримувати вибірки з p_0 .

Оцінку градієнт лог-щільності розподілу можна визначити шляхом навчання бально-орієнтованої моделі на вибірках за допомогою бального зіставлення [27, 28]. Щоб оцінити $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$, ми можемо тренувати залежну

від часу скор-модель $s_\theta(\mathbf{x}, t)$ використовуючи неперервне узагальнення рівнянь [29]:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_t \left\{ \lambda(t) \mathbb{E}_{\mathbf{x}(0)} \mathbb{E}_{\mathbf{x}(t)|\mathbf{x}(0)} \left[\left\| s_\theta(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log p_{0t}(\mathbf{x}(t)|\mathbf{x}(0)) \right\|_2^2 \right] \right\}. \quad (2.3)$$

Тут $\lambda: [0, T] \rightarrow \mathbb{R}_{>0}$ – додатна вагова функція, t рівномірно вибирається на $[0, T]$, $\mathbf{x}(0) \sim p_0(\mathbf{x})$, а $\mathbf{x}(t) \sim p_{0t}(\mathbf{x}(t)|\mathbf{x}(0))$.

За достатньої кількості даних і моделювальної здатності бальне зіставлення гарантує, що оптимальний розв’язок рівняння (2.3), позначений $s_{\theta^*}(\mathbf{x}, t)$, збігається з $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ для майже всіх \mathbf{x} і всіх t .

Як і у SMLD та DDPM, зазвичай обирають

$$\lambda \propto \frac{1}{\mathbb{E} \left[\left\| \nabla_{\mathbf{x}(t)} \log p_{0t}(\mathbf{x}(t)|\mathbf{x}(0)) \right\|_2^2 \right]}.$$

Зауважимо, що рівняння (2.3) використовує, але можуть застосовуватися й інші варіанти зіставлення балів – зокрема зіставлення зрізаних балів [28] та зіставлення балів методом скінченних різниць [30].

Зазвичай нам потрібно знати перехідне ядро $p_{0t}(\mathbf{x}(t)|\mathbf{x}(0))$, щоб ефективно розв’язати рівняння (2.3).

Коли функція дрейфу $f(\cdot, t)$ є афінною, перехідне ядро завжди має вигляд гаусівського розподілу, середнє та дисперсію якого часто можна виписати в замкненому вигляді та отримати стандартними методами [31].

Для більш загальних SDE ми можемо розв’язувати рівняння Колмогорова (пряме рівняння Фоккера–Планка) [25], щоб отримати $p_{0t}(\mathbf{x}(t)|\mathbf{x}(0))$.

Альтернативно, можна симулювати СДР, щоб отримувати вибірки з $p_{0t}(\mathbf{x}(t)|\mathbf{x}(0))$, і замінити зіставлення оцінок знешумлення у рівнянні (2.3) на зіставлення зрізаних балів під час навчання моделі – це дозволяє уникнути обчислення $\nabla_{\mathbf{x}(t)} \log p_{0t}(\mathbf{x}(t)|\mathbf{x}(0))$.

Шумові збурення, які використовуються в SMLD та DDPM, можна розглядати як дискретизації двох різних СДР. При використанні N рівнів шуму кожне ядро збурення $p_{\sigma_i}(\mathbf{x}|\mathbf{x}_0)$ у SMLD відповідає розподілу \mathbf{x}_i у наступному ланцюгу Маркова:

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \sqrt{\sigma_i^2 - \sigma_{i-1}^2} \mathbf{z}_{i-1}, \quad i = 1, \dots, N, \quad (2.4)$$

де $\mathbf{z}_{i-1} \sim \mathbf{N}(0, \mathbf{I})$, а $\sigma_0 = 0$ введено для спрощення запису.

У межі при $N \rightarrow \infty$ множина $\{\sigma_i\}_{i=1}^N$ переходить у неперервну функцію $\sigma(t)$, \mathbf{z}_i переходить у $\mathbf{z}(t)$, а дискретний ланцюг Маркова $\{\mathbf{x}_i\}_{i=1}^N$ переходить у неперервний стохастичний процес $\{\mathbf{x}(t)\}_{t=0}^1$, де ми використовуємо неперервну часову змінну $t \in [0, 1]$ замість індексу i . Процес описується наступним СДР:

$$d\mathbf{x} = \sqrt{\frac{d[\sigma^2(t)]}{dt}} d\mathbf{w}. \quad (2.5)$$

Аналогічно, для ядер збурення $\{p_{\alpha_i}(\mathbf{x}_i|\mathbf{x}_0)\}_{i=1}^N$ у DDPM дискретний ланцюг Маркова має вигляд:

$$\mathbf{x}_i = \sqrt{1 - \beta_i} \mathbf{x}_{i-1} + \sqrt{\beta_i} \mathbf{z}_{i-1}, \quad i = 1, \dots, N \quad (2.6)$$

Коли $N \rightarrow \infty$, рівняння (2.6) переходить у наступне SDE:

$$d\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}dt + \sqrt{\beta(t)}d\mathbf{w}. \quad (2.7)$$

Отже, шумові збурення, використані в SMLD та DDPM, відповідають дискретизаціям СДР з рівнянь (2.5) та (2.7).

Цікаво, що СДР з рівняння (2.5) завжди породжує процес із зростаючою дисперсією при $t \rightarrow \infty$, тоді як СДР з рівняння (2.7) дає процес із фіксованою дисперсією, рівною одиниці, якщо початковий розподіл має одиничну дисперсію. Через цю відмінність ми називаємо рівняння (2.5) Variance Exploding (VE) SDE, а рівняння (11) – Variance Preserving (VP) SDE.

Розглянемо підхід на основі рівняння:

$$d\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}dt + \sqrt{\beta(t)(1 - e^{-2\int_0^t \beta(s)ds})}d\mathbf{w}. \quad (2.8)$$

Використовуючи ту саму функцію $\beta(t)$ та однаковий початковий розподіл, дисперсія стохастичного процесу, породженого рівнянням (2.8), на кожному проміжному кроці часу завжди залишається обмеженою зверху дисперсією VP SDE. З цієї причини ми називаємо рівняння (2.8) sub-VP SDE.

Оскільки VE, VP та sub-VP SDE мають афінні (лінійно-афінні) коефіцієнти дрейфу, їхні перехідні ядра $p_{0t}(\mathbf{x}(t)|\mathbf{x}(0))$ є гаусівськими і можуть бути обчислені в замкненому вигляді. Це робить навчання за рівнянням (2.3) особливо ефективним.

Бально-орієнтовані моделі дають змогу застосувати ще один чисельний метод для розв'язання СДР у зворотному часі. Для будь-якого дифузійного процесу існує відповідний детермінований процес, траєкторії якого мають ті самі маргінальні густини ймовірності $\{p_t(\mathbf{x})\}_{t=0}^T$, що й СДР. Цей детермінований процес задовольняє загальному диференціальному рівнянню:

$$d\mathbf{x} = \left[f(\mathbf{x}, t) - \frac{1}{2} g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) \right] dt . \quad (2.9)$$

Отже, рівняння (2.9) є ODE потоку ймовірності (probability flow ODE). Коли бальна-функцію апроксимують часовою бально-орієнтованою моделлю (зазвичай нейронною мережею), такий процес є прикладом нейронної ODE [32].

Завдяки зв'язку з нейронними СДР ми можемо обчислити густину, визначену рівнянням (2.9), використовуючи формулу миттєвої зміни змінних [8]. Це дозволяє нам обчислювати точну правдоподібність для будь-яких вхідних даних .

Ми обчислюємо log-likelihood для рівномірно деквантизованих даних і порівнюємо лише з моделями, оціненими таким самим способом (не включаючи моделі з варіаційною деквантизацією [33]).

Інтегруючи рівняння (2.9), ми можемо закодувати будь-яку точку даних $\mathbf{x}(0)$ у латентний простір $\mathbf{x}(T)$. Декодування можна виконати, інтегруючи відповідне ОДЕ для SDE у зворотному часі. Так само, як і в інших оборотних моделях – таких як нейронних диференціальних рівняннях [34, 35] – ми можемо маніпулювати латентним представленням для редагування зображень: виконувати інтерполяцію, зміну "температури".

На відміну від більшості сучасних оборотних моделей, наше кодування є однозначно ідентифікованим. Це означає, що за достатніх обсягів даних, моделювальної здатності та точності оптимізації кодування для вхідного прикладу однозначно визначається розподілом даних [36].

Подібно до нейронними СДР, ми можемо генерувати зразки $\mathbf{x}(0) \sim p_0$, розв'язуючи рівняння (2.9) з різних фінальних умов $\mathbf{x}(T) \sim p_T$. Використовуючи фіксовану стратегію дискретизації, ми отримуємо конкурентні зразки, особливо у поєднанні з коректорами Застосування black-box ODE-розв'язувача [37] дозволяє:

- отримувати вибірки високої якості;
- гнучко балансувати між точністю та швидкістю.

Неперервна структура нашої моделі дозволяє нам генерувати не лише зразки з p_0 , але й з умовного розподілу $p_0(\mathbf{x}(0)|\mathbf{y})$, якщо відомий розподіл $p_t(\mathbf{y}|\mathbf{x}(t))$.

Маючи forward SDE із рівняння (2.1), ми можемо отримувати вибірки з $p_t(\mathbf{x}(t)|\mathbf{y})$, починаючи з розподілу $p_T(\mathbf{x}(T)|\mathbf{y})$ та розв'язуючи умовне SDE у зворотному часі:

$$d\mathbf{x} = \left\{ f(\mathbf{x}, t) - g(t)^2 (\nabla_{\mathbf{x}} \log p_t(\mathbf{x}) + \nabla_{\mathbf{x}} \log p_t(\mathbf{y}|\mathbf{x})) \right\} dt + g(t) d\bar{\mathbf{w}}. \quad (2.10)$$

Загалом рівняння (2.10) дозволяє розв'язувати широкий клас обернених задач (inverse problems) за допомогою score-based генеративних моделей, якщо відома або наближена градієнтна інформація для forward-процесу: $\nabla_{\mathbf{x}} \log p_t(\mathbf{y}|\mathbf{x}(t))$.

У деяких випадках можна навчити окрему модель, щоб апроксимувати forward-процес $\log p_t(\mathbf{y}|\mathbf{x}(t))$ та обчислити його градієнт. В інших випадках градієнт можна оцінити за допомогою евристик або фахової інформації про домен.

Імпутація є окремим випадком умовної генерації. Припустимо, що ми маємо неповний об'єкт даних \mathbf{y} , у якому відомо лише певну підмножину ознак, $\Omega(\mathbf{y})$.

Імпутація зводиться до генерування вибірок із розподілу $p(\mathbf{x}(0)|\Omega(\mathbf{y}))$, що можна виконати, використовуючи безумовну модель.

Колоризація є спеціальним випадком імпутації, за винятком того, що відомі компоненти даних між собою пов'язані.

2.3 Алгоритми прямого генерування

У статті [29] автори узагальнили процес зашумлення до нескінченної кількості кроків у моделях DDPM та NCSN і визначили його як стохастичний процес у неперервному часі.

Стохастичні процеси є розв'язками стохастичних диференціальних рівнянь (СДР). СДР задаються як у рівнянні (2.11):

$$dx = f(x,t)dt + g(t)dw, \quad (2.11)$$

де $f(.,t):R^d \rightarrow R^d$ – векторна функція, що називається дрейфовим коефіцієнтом;

$g(t) \in R$ – скалярна функція, що називається коефіцієнтом дифузії;

w – стандартний броунівський рух;

dw – нескінченно малий білий шум.

Вибір СДР не завжди однаковий, оскільки існують сотні способів спотворити розподіл даних. Процеси зашумлення в імовірнісній моделі дифузії з усуненням шуму (DDPM) та мережі умовної оцінки шуму (NCSN) є різними версіями СДР. Відповідні СДР наведені в рівнянні (2.12) для DDPM і рівнянні (2.13) для NCSN.

Тут $\beta\left(\frac{t}{T}\right) = T\beta_t$ і $\sigma\left(\frac{t}{T}\right) = \sigma_t$ коли T прямує до нескінченності.

$$dx = -\frac{1}{2}\beta(t)x dt + \sqrt{\beta(t)} dw, \quad (2.12)$$

$$dx = \sqrt{\frac{d[\sigma(t)^2]}{dt}} dw. \quad (2.13)$$

Розв'язки СДР у зворотному часі є процесами зворотної дифузії, які поступово перетворюють шум у дані. Для кожного СДР існує зворотне-СДР, як наведено в рівнянні (2.14). Тут розподіл вибірки в будь-який момент часу t позначається як $p_t(x)$.

Рівняння (2.17) використовує метод оцінки зіставлення знешумлення, щоб знайти зважену комбінацію дивергенцій Фішера у зворотному часі. Хоча цей метод є обчислювально більш ефективним, його можна оптимізувати з такою ж ефективністю, як методи, такі як зіставлення зрізаних оцінок [37]. dt позначає нескінченно малі кроки часу в зворотному напрямку, оскільки SDE має бути розв'язане у зворотному часі – від $t = T$ до $t = 0$

$$dx = \left[f(x, t) - g^2(t) \nabla_x \log p_t(x) \right] dt + g(t) dw. \quad (2.14)$$

Для всіх стохастичних дифузійних процесів існує детермінований процес, що відповідає звичайному диференціальному рівнянню (ODE). Потік ймовірностей ODE, наведене в рівнянні (2.15), є звичайним диференціальним рівнянням, чий траєкторії мають ту саму маргінальну розподільну функцію, що й СДР у зворотному часі. Як SDE у зворотному часі, так і потоці ймовірностей ODE дозволяють виконувати вибірку з того самого розподілу даних.

Рівняння (2.15) є прикладом нейронною ODE, оскільки оцінка-функція передбачається часовою оціночною моделлю (нейронною мережею).

$$dx = \left[f(x, t) - \frac{1}{2} g^2(t) \nabla_x \log p_t(x) \right] dt. \quad (2.15)$$

Стохастичні диференціальні рівняння розв'язуються з набором неперервних випадкових змінних $\{x(t)\}_{t \in [0, T]}$. Ці випадкові змінні слідують стохастичним траєкторіям, коли індекс часу t рухається від початкового часу 0 до кінцевого часу T .

Нехай маргінальна густина ймовірності $x(t)$ позначається як $p_t(x)$. Тут $t \in [0, T]$ подібний до $i = 1, 2, \dots, L$ коли ми маємо скінченну кількість шумів, а $p_t(x)$ тут подібний до $p_{\sigma_i}(x)$. Розподіл до застосування шуму в момент $t = 0$ - це реальний розподіл даних $p_0(x) = p(x)$. Додавання шуму до $p(x)$ стохастичним процесом протягом достатньо великого часу T перетворює його на пояснюваний шумовий розподіл $\pi(x)$ який називається попередній розподіл.

Коли ми додаємо останній і найбільший шум σ_L то отримуємо розподіл:

$$p_{\sigma_L}(x).$$

СДР, наведене в рівнянні (2.11), визначається вручну, так само як $\sigma_1 < \sigma_2 < \dots < \sigma_L$, коли ми маємо шум з обмеженою шкалою. Залежна від часу оціночна модель $s_\theta(x, t) \approx \nabla_x \log p_t(x)$ тренується для розв'язання зворотного-СДР за допомогою зіставлення оцінок. Ця модель подібна до $s_\theta(x, \sigma_i)$ коли ми маємо шум з кінцевою шкалою. Ціль тренування, наведена в рівнянні (2.16), є неперервною та зваженою комбінацією дивергенцій Фішера між оцінкою реального розподілу даних і моделлю. Тут $U(0, T)$ позначає рівномірний розподіл на часовому інтервалі $[0, T]$, а $\lambda: R \rightarrow R_{>0}$ - це додатна зважувальна функція.

Зазвичай вона рівна:

$$\lambda(t) \propto \frac{1}{E[\|\nabla_x \log p(x(t) | x(0))\|_2^2]}.$$

Щоб збалансувати величину різних зіставлених оціночних втрат протягом часу.

$$E_{t \in U(0,T)} \left[E_{p_t(x)} \lambda(t) \|\nabla_x \log p_t(\mathbf{x}) - s_\theta(x,t)\|_2^2 \right] \quad (2.16)$$

$$\theta^e = \arg_\theta \min E_t \left\{ \lambda(t) E_{x(0)} E_{x(t)|x(0)} \left[\|s_\theta(x(t),t) - \nabla_x \log p_{0t}(x(t)|x(0))\|_2^2 \right] \right\} \quad (2.17)$$

Необхідно знати перехідне ядро $p_{0t}(x(t)|x(0))$. щоб розв'язати рівняння (2.17).

Перехідне ядро є гаусовим розподілом, коли $f(.,t)$ є афінним, а середнє та дисперсія можуть зазвичай бути отримані в закритій формі стандартними методами. Процеси шуму, які використовуються в DDPMs і NCSNs, відповідають різним версіям СДР. СДР у рівнянні (2.13) завжди є процесом вибухаючої дисперсії, коли $t \rightarrow \infty$ Рівняння (2.12) має фіксовану дисперсію, коли $t \rightarrow \infty$ і початковий розподіл має одиничну дисперсію. Через цю різницю рівняння (2.13) називають дисперсно вибуховим СДР, а рівняння (2.12) – захищеним від варіацій СДР. Крім того, існує ще одне СДР, яке добре працює для likelihood, і воно називається sub-VP СДР.

Оскільки VE, VP та sub-VP СДР мають афінний дрейфовий коефіцієнт, перехідні ядра $p_{0t}(x(t)|x(0))$ є гаусовими і можуть бути обчислені в закритій формі.

2.4 Алгоритми зворотного генерування

Після тренування залежної від часу оціночної моделі s_θ вибірка з p_0 вимагає розв'язання SDE у зворотному часі за допомогою числових методів.

Існує багато універсальних числових розв'язувачів, які розв'язують СДР. Наприклад, метод Ейлера–Маруяма та стохастичні методи Рунге–Кутта

можуть бути застосовані до СДР у зворотному часі для генерації вибірок. Метод вибірки в імовірнісній моделі дифузії з усуненням шуму (DDPM) відповідає спеціальній дискретизації VP-SDE у зворотному часі. Також зворотні дифузійні семплери [37] працюють трохи краще, ніж класичні семплери, які дискретизують СДР у зворотному часі так само, як і пряму дискретизацію, і тому вибірка є простою, коли задана пряма дискретизація.

У семплерах Predictor–Corrector (PC) спочатку числовий розв’язувач СДР, який називається predictor, робить передбачення для вибірки на наступному кроці часу на кожному часовому кроці. Потім оціночно-базований підхід MCMC, який називається corrector, виправляє маргінальний розподіл вибірки. PC-семплери узагальнюють оригінальні методи вибірки NCSNs та DDPMs. NCSN використовує функцію тотожності як предиктор і ланжевенівську динаміку як коректор, тоді як DDPM використовує числові розв’язувачі як predictor та функцію тотожності як corrector.

Алгоритми для Predictor–Corrector семплерів для VE-SDE та VP-SDE наведені в [29].

Algorithm 1: VE-SDE predictor-corrector sampling algorithm

```

 $x_N \sim N(0, \sigma_{\max}^2 I)$ 
for  $i = N - 1, \dots, 0$  do
     $x'_i \leftarrow x_{i+1} + (\sigma_{i+1}^2 - \sigma_i^2) s_{\theta}^e(x_{i+1}, \sigma_{i+1})$ 
     $z \sim N(0, I)$ 
     $x_i \leftarrow x'_i + \sqrt{\sigma_{i+1}^2 - \sigma_i^2} z$ 
    for  $j = 1, \dots, M$  do
         $z \sim N(0, I)$ 
         $x_i \leftarrow x_i + \eta s_{\theta}^e(x_i, \sigma_i) + \sqrt{2\eta} z$ 
    end for

```

```

end for
return  $x_0$ 

```

Algorithm 2: VP-SDE predictor-corrector sampling algorithm

```

 $x_N \sim N(0, I)$ 

```

```

for  $i = N-1, \dots, 0$  do

```

$$x'_i \leftarrow (2 - \sqrt{1 - \beta_{i+1}})x_{i+1} + \beta_{i+1} s_\theta^e(x_{i+1}, i+1)$$

$$z \sim N(0, I)$$

$$x_i \leftarrow x'_i + \sqrt{\beta_{i+1}} z$$

```

for  $j = 1, \dots, M$  do

```

$$z \sim N(0, I)$$

$$x_i \leftarrow x_i + \eta s_\theta^e(x_i, i) + \sqrt{2\eta} z$$

```

end for

```

```

end for

```

```

return  $x_0$ 

```

Тут N задає число дискретизації для СДР у зворотному часі, а M задає кількість кроків corrector у алгоритмах. Розв'язував зворотного СДР використовується як метод predictor, а відпал динаміки Ланжевена— як метод corrector. Розмір кроку динаміки Ланжевена позначається як $\{\eta\}_{i=0}^{N-1}$.

2.5 Метрики для оцінки подібності зображень

Мета оцінки подібності зображень полягає в тому, щоб виміряти відстань між синтетичними та реальними зображеннями. Більшість існуючих

методів використовують початкову модель Inception для відображення зображень у просторі меншої розмірності. Найпоширенішою метрикою на даний час є IS, яка вимірює відстань за допомогою дивергенції Кульбака-Лейблера [38]. Однак дана метрика базується на імовірності приналежності зображення до одного із класів і не може відобразити перенавчання мережі. В якості кращої альтернативи запропоновано FID [39]. Ця метрика безпосередньо вимірює відстань Фреше на просторі ознак за допомогою апроксимації нормального розподілу.

Оскільки дані метрики базуються на попередньо підготовленій моделі Inception (яка навчалася на датасеті ImageNet), то їх значення можуть погіршуватися при застосуванні до інших датасетів. Серед всіх метрик найбільш популярними та релевантними метриками для оцінки якості синтезованих за допомогою ГЗМ зображень є IS та FID. Дані метрики досить добре себе зарекомендували у багатьох дослідженнях та показали хорошу кореляцію із оцінками експертів.

IS. Дана метрика базується на моделі нейронної мережі для класифікації зображень Google Inception V3. Ця модель призначена для класифікації кольорових зображень. В якості навчального набору даних використано набір даних ImageNet, який включає близько 1,2 мільйона RGB зображень, що поділені на 1000 класів.

Дана метрика показала хорошу кореляцію із оцінками, що зроблені експертами на датасеті CIFAR-10.

$$IS(G) \approx \exp(E_{x \sim p_g}[D_{KL}(p(y|x) || p(y))]), \quad (2.18)$$

де E – математичне сподівання;

$x \sim p_g$ показує, що x є зображенням, що синтезоване із розподілу p_g (розподіл генератора);

D_{KL} є відстанню Кульбака-Лейблера між розподілом умовної імовірності $p(y|x)$ та маргінальним розподілом $p(y)$.

Передбачається, що умовний розподіл даних, який містить значущі об'єкти, повинен мати низьку ентропію, а маргінальний розподіл (синтезовані зображення є різноманітними) – високу.

IS працює наступним чином. Для прикладу візьмемо 5 тисяч синтетичних зображень. Для того, щоб отримати умовний розподіл класів, потрібно класифікувати дані зображення мережею Insertion, яка поверне вектор імовірностей $p(y|x)$. Для отримання маргінального розподілу потрібно просумувати умовний розподіл для кожного зображення $p(y) = \frac{1}{5000} \sum_{i=1}^{5000} p(y|x_i)$.

Далі потрібно обчислити відстань Кульбака-Лейблера між умовним розподілом кожного синтезованого зображення та загальним маргінальним розподілом. Середнє значення даних відстаней і буде значенням метрики IS [40].

Отже, IS вимірює середню відстань Кульбака-Лейблера між умовним розподілом $p(y|x)$ та маргінальним розподілом класів $p(y)$. Тобто дана метрика взагалі не розглядає розподіл оригінальної вибірки, а тому не може оцінити наскільки добре синтезовані генератором зображення подібні до оригінальної вибірки. Дана метрика оцінює лише їх різноманітність. Мінусами даної метрики є чутливість до роздільної здатності самих зображень та до змін в самій мережі, яка використовується для класифікації.

Мінімальним значенням даної метрики є 1, а максимальним – кількість класів, яку може класифікувати мережа Insertion. В даному випадку – 1000.

Для того, щоб отримати високе значення IS, потрібно щоб синтетичні зображення містили чіткі об'єкти та щоб генератор синтезував різноманітні зображення із всіх класів [39]. Відповідно, якщо хоча б одна із цих умов незадовільна – оцінка буде низькою.

FID порівнює розподіли оригінальних та синтетичних даних. Для того, щоб обчислити *FID* між реальними та синтезованими зображеннями, дані перетворюються в простір ознак з використанням конкретного шару моделі *Inception*, а саме – *pool3 layer*. Простір ознак використовується для того, щоб представити зображення у просторі меншої розмірності, де схожі зображення представлені у відносно тих самих регіонах. На виході ми отримуємо карти активацій (карти ознак). Далі ці карти ознак апроксимуються із використанням двох нормальних розподілів. Тоді відстань між ними обчислюється наступним чином:

$$d^2((m_r C_r), (m_g C_g)) = \|m_r - m_g\|^2 + Tr(C_r + C_g - 2(C_r C_g)^{\frac{1}{2}}), \quad (2.19)$$

де $(m_r C_r)$ та $(m_g C_g)$ – середнє та *covariance* реального та синтезованого розподілу даних відповідно,

Tr – сума діагональних елементів матриці.

Чим нижче значення метрики, тим менша відстань між розподілами, відповідно розподіли більш подібні між собою [41]. *FID* метрика досить чутлива до спотворень на зображеннях (зсув, шум і т.д.). Чим більше спотворень, тим більшим буде значення метрики.

Низьке значення *FID* свідчить про те, що розподіли реальних та синтетичних зображень подібні між собою. Проте на практиці, якщо модель має низьке значення *FID*, то це свідчить про те, що зображення мають високу якість або різноманітність, або й те й інше. Така поведінка може значно ускладнити діагностику моделі.

Автори також показують, що дана метрика більше збігається із людськими оцінками та більш стійка до шуму, ніж *IS* [42-43].

Дані метрики є досить популярними в області синтезу зображень за допомогою *GAN* мереж. Але вони мають свої недоліки.

Метрика *IS* має такі обмеження:

- 1) значення метрики сильно залежить від того, що може класифікувати модель Inception;
- 2) синтез зображень іншого набору класів, яких немає в оригінальному датасеті ImageNet, може спричинити низьке значення IS;
- 3) якщо класифікатор не може визначити ознаки, які відносяться до навчального датасету, то зображення низької якості можуть отримувати високі оцінки.

Мережа Inception натренована на датасеті ImageNet. Якщо обчислювати IS для іншого датасету, то класифікатор може досить погано визначити деякі ознаки, відповідно зображення низької якості будуть отримувати високі оцінки.

FID також базується на моделі Google Inception. Але на відміну від IS, дана метрика може визначити залежності між класами. Тобто якщо модель генерує тільки одне зображення на кожен клас, то IS може бути досить високим, проте FID буде низьким. Також метрика FID погіршується при додаванні різних артефактів до зображення.

Метрика IS дійсно показує кореляцію з якістю та різноманітністю створених зображень, що пояснює широке використання на практиці. Однак дана метрика оцінює лише розподіл синтезованих зображень, проте не бере до уваги те, наскільки синтезовані та оригінальні зображення є подібними. Як наслідок, це може спонукати моделі просто вивчати чіткі та різноманітні зображення (або навіть деякі шуми) замість розподілу оригінальних даних. Ця метрика обмежена тільки вимірюванням того, наскільки синтезовані зображення є різноманітними, а FID вимірює відстань між розподілом синтезованих та реальних даних.

2.6 Дифузійна модель синтезу зображень на основі стохастичних диференціальних рівнянь

Генерування зображень на основі дифузійної моделі відбувається в програмному середовищі Stable Diffusion. Базова модель Stable Diffusion навчається на великій вибірці зображень [44-46]. Навчання на основі своєї вибірки відбувається в середовищі нейромережі Hypernetwork. Ця мережа корегує ваги базової моделі. Алгоритм генерування зображень на основі дифузійної моделі складається з таких кроків:

1. Навчання на основі свого dataset зображень в середовищі Hypernetwork.
2. Процес зашумлення початкового dataset I_C .
3. Процес знешумлення.

Деталізуємо ці кроки. Початкова вибірка перетворюється в латентний простір: $I_C \rightarrow Z_{0C}$. На основі Z_{0C} обчислюємо значення зашумлення на кожному кроці t так:

$$Z_t = \sqrt{\alpha_t} Z_{0C} + \sqrt{1 - \alpha_t} \varepsilon_t,$$

де α_t – коефіцієнт, який визначає швидкість зашумлення на кроці t .

Значення кроку t вибираємо із діапазону:

$$t \in [0, T],$$

де T – кількість кроків;

ε_t – значення випадкового гаусівського шуму на кроці t .

Значення ε_t обчислюється згідно виразу:

$$\varepsilon_t \sim N(E, D),$$

де N – нормальний закон розподілу із математичним сподіванням $E=0$ і дисперсією $D=1$.

Значення знешумлення обчислюємо згідно виразу:

$$Z_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(Z_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \hat{\varepsilon}_t \right),$$

де $\hat{\varepsilon}_t$ – оціночне значення шуму на кроці t ;

$\bar{\alpha}_t$ – коефіцієнт, який визначає рівень шуму на попередньому кроці t ;

β_t – коефіцієнт, що контролює рівень зменшення шуму.

Після виконання процесу знешумлення (після проходження $t=T$ кроків) утворюється вектор Z_{1C} у латентному просторі. Потім енкодер перетворює Z_{1C} у множину зображень I_{CD} , причому $I_{CD} \gg I_C$. Якість згенерованих зображень перевіряємо метриками IS і FID .

3 СИНТЕЗ ГІСТОЛОГІЧНИХ ЗОБРАЖЕНЬ ЗАСОБАМИ ГЕНЕРАТИВНОГО ІНТЕЛЕКТУ

3.1 Програмне середовище Stable Diffusion

У цьому підрозділі розглянуто веб-інтерфейс, призначений для взаємодії з моделями Stable Diffusion. Одним із найпоширеніших середовищ для роботи з цими моделями є AUTOMATIC1111, що поєднує зручність використання, високу функціональність і гнучкість налаштування [47]. Це веб-орієнтоване рішення створене для спрощення процесу генерації та редагування зображень без необхідності роботи з командним рядком або написання програмного коду (рисунок 3.1). Інтерфейс AUTOMATIC1111 підтримує широкий спектр інструментів – від базової генерації зображень на основі текстових описів (text-to-image) до більш складних режимів, таких як image-to-image, inpainting, outpainting, а також навчання користувацьких моделей за допомогою методів LoRA, Textual Inversion та DreamBooth.

Користувач має можливість регулювати параметри процесу дифузії, зокрема кількість ітерацій, силу шуму, масштаб зображення, стиль генерації та вибір моделі чи контрольних чекпоінтів. Завдяки роботі через веб-браузер, AUTOMATIC1111 забезпечує зручний доступ до всіх функцій Stable Diffusion на локальному комп'ютері або сервері. Програмне середовище підтримує підключення додаткових розширень і плагінів, що розширюють можливості системи, зокрема для автоматичної обробки серій зображень, керування масками чи створення композицій із кількох джерел.

Відкритий вихідний код і активна спільнота розробників сприяють постійному вдосконаленню інтерфейсу. Саме тому AUTOMATIC1111 став одним із найпопулярніших і найефективніших інструментів для взаємодії з моделями Stable Diffusion. Він широко застосовується у наукових і навчальних дослідженнях, зокрема для синтезу гістологічних зображень, створення

аугментованих наборів даних і тестування нових підходів до генерації біомедичного контенту.

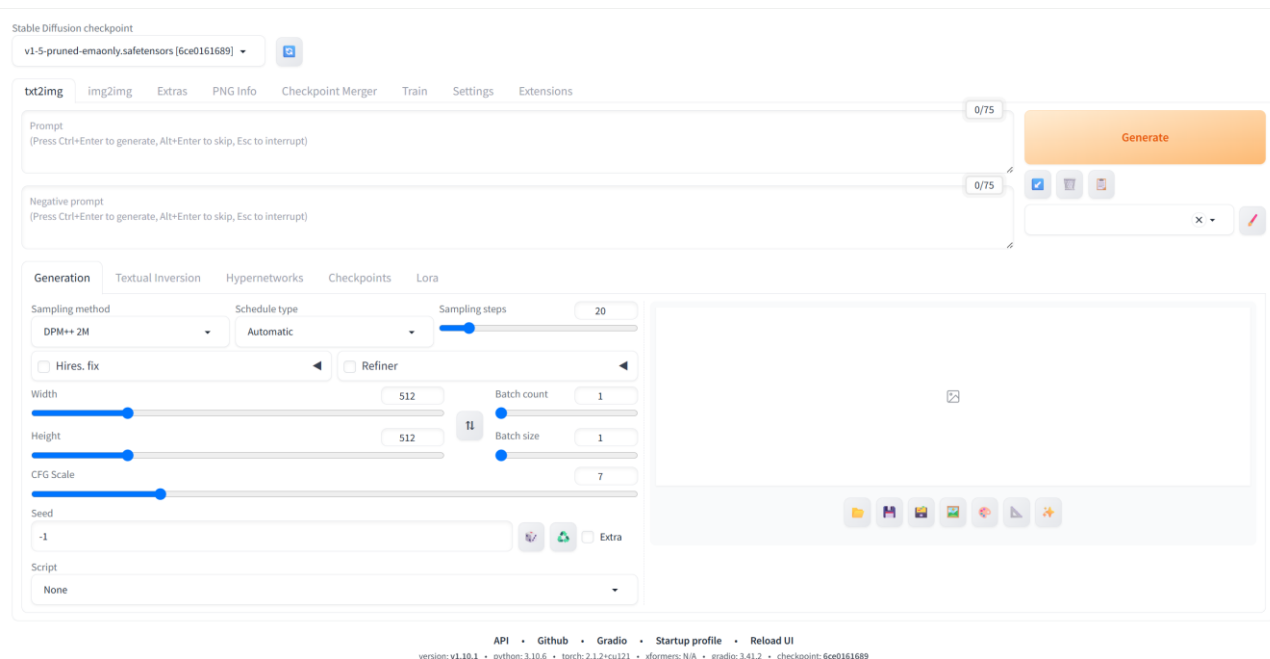


Рисунок 3.1 – Головна вкладка “Generation” у середовищі AUTOMATIC111

Він застосовується для генерації зображень на основі текстових підказок (prompts), а також для виконання різноманітних операцій із моделлю Stable Diffusion - зокрема редагування, повторної генерації, налаштування параметрів дифузійного процесу та збереження отриманих результатів.

На рисунку 3.2 подано структурну схему інтерфейсу, що ілюструє його основну архітектуру та взаємозв'язки між ключовими компонентами.

Інтерфейс побудований логічно й орієнтований на зручність користувача: усі основні елементи згруповано у вигляді панелей та вкладок, що дозволяє швидко перемикатися між режимами роботи – генерацією, попереднім переглядом, історією або налаштуваннями системи.

Нижче наведено детальний опис основних функціональних елементів інтерфейсу, які забезпечують повноцінну взаємодію з моделлю Stable Diffusion та ефективно керування процесом синтезу зображень.

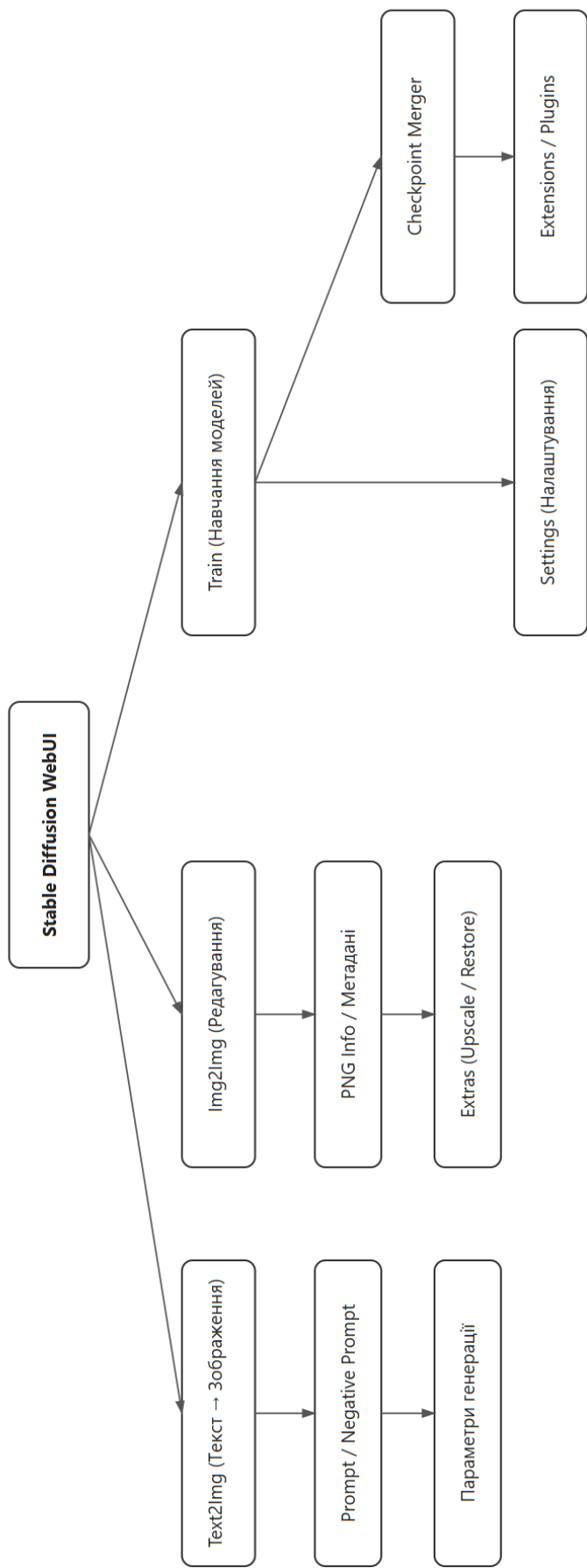


Рисунок 3.2 – Логічна схема основних компонентів інтерфейсу Stable Diffusion WebUI

Верхня панель

У верхній частині інтерфейсу розташовані основні елементи керування, що забезпечують швидкий доступ до різних режимів роботи. Центральним компонентом є Checkpoint – параметр, який визначає активну модель, що використовується для генерації зображень.

У наведеному прикладі використано модель v1-5-pruned-emaonly.safetensors, яка є однією з найпоширеніших базових версій Stable Diffusion. Вибір контрольної точки має суттєвий вплив на характер і стиль отриманих результатів, оскільки різні моделі спеціалізуються на певних типах зображень - наприклад, портретах, пейзажах, художніх композиціях або технічних ілюстраціях.

Безпосередньо під головним меню розташовані вкладки (Tabs), які структурують роботу користувача та розподіляють функціональні можливості інтерфейсу за напрямками.

- txt2img – основна вкладка, призначена для генерації зображень на основі текстових описів (*prompts*). Саме тут користувач формує текстову підказку, що визначає зміст і стиль майбутнього зображення.

- img2img – режим, який дозволяє використовувати вже наявне зображення як основу для подальшої генерації. Завдяки цьому можна виконувати редагування, стилізацію або часткову заміну елементів за допомогою текстових інструкцій [48].

- Extras – набір допоміжних інструментів для масштабування, підвищення роздільної здатності та покращення дрібних деталей зображення.

- PNG Info – вкладка для перегляду метаданих, що містять усі параметри, використані під час генерації конкретного зображення.

- Checkpoint Merger – модуль, який надає змогу об'єднувати кілька контрольних точок, формуючи нові моделі на основі існуючих, що корисно для створення спеціалізованих варіантів Stable Diffusion.

- Train – розділ для навчання моделей, у тому числі адаптованих під певний стиль або тематику.

- Settings – панель налаштувань, де користувач може змінювати параметри інтерфейсу, розподіл ресурсів GPU, шляхи до моделей та інші технічні опції.

- Extensions – система розширень, яка дозволяє підключати додаткові модулі, інструменти або функції, розширюючи базові можливості Stable Diffusion WebUI.

Завдяки такій організації елементів користувач має швидкий доступ до всіх ключових функцій середовища, що забезпечує ефективну роботу як у базовому режимі генерації, так і під час виконання складних експериментів із дифузійними моделями.

Поля введення текстових підказок

Центральне місце в інтерфейсі займають поля Prompt і Negative Prompt, які визначають зміст та якість майбутнього зображення. Поле Prompt використовується для формування текстового опису бажаного результату. Саме тут користувач задає основну ідею зображення, вказуючи його стиль, композицію, кольорову гаму або настрій. Наприклад: *«реалістичний портрет у м'якому освітленні»* чи *«детальна наукова ілюстрація клітинної структури»*. Введений опис безпосередньо впливає на те, яким буде згенерований візуальний результат.

Поле Negative Prompt виконує протилежну функцію - воно дає змогу виключити з генерації небажані характеристики або об'єкти. Сюди зазвичай додають фрази на кшталт *«низька якість»*, *«розмиття»*, *«деформовані деталі»* чи *«артефакти»*. Використання негативних підказок дозволяє суттєво підвищити якість кінцевого зображення, усуваючи типові похибки моделі та роблячи результат більш точним і візуально привабливим.

Панель генерації

У цьому блоці користувач задає основні параметри, що визначають процес синтезу зображення.

Одним із ключових налаштувань є Sampling method – алгоритм вибірки, який визначає спосіб поступового відновлення зображення з шуму.

Найчастіше використовується метод Euler, що забезпечує оптимальний баланс між швидкістю роботи та якістю результату. Однак існують й інші алгоритми, наприклад DDIM, DPM++ чи Neun, які можуть формувати різну глибину, текстуру або м'якість контурів, тому вибір конкретного методу залежить від поставленої задачі та бажаного стилю зображення.

Параметр Sampling steps визначає кількість ітерацій, протягом яких модель виконує зворотний процес дифузії. Збільшення цього показника підвищує деталізацію та точність зображення, проте потребує більше часу для генерації.

Функція Hires.fix використовується для підвищення роздільної здатності зображення після первинного рендеру. Вона виконує повторну генерацію з більшим масштабом, зберігаючи композицію, кольори та ключові деталі. Цей режим особливо ефективний для створення великих ілюстрацій або макрозображень гістологічних зразків, де важлива точність дрібних структур. Параметри Width і Height задають розміри кінцевого зображення. Вони впливають на співвідношення сторін і формат - від квадратного (512×512 пікселів) до панорамного чи вертикального. Вибір цих значень залежить від цілей генерації та вимог до подальшої обробки.

Особливе значення має параметр CFG Scale (Classifier-Free Guidance) - він контролює ступінь відповідності між текстовою підказкою та результатом генерації. Невеликі значення (5–7) забезпечують більшу варіативність і творчість моделі, тоді як високі (10–12) змушують її суворо дотримуватись опису, іноді зменшуючи природність зображення.

Параметр Seed визначає початкове значення випадкової послідовності, яка впливає на структуру шуму та варіації результатів. Якщо встановлено значення -1, система автоматично генерує новий seed під час кожного запуску, що дозволяє отримувати унікальні варіанти зображень навіть за незмінних параметрів генерації.

Параметри обробки

Для підвищення продуктивності та зручності роботи в інтерфейсі передбачені параметри Batch count і Batch size, які керують кількістю одночасно створюваних зображень.

Параметр Batch count визначає, скільки серій генерації буде виконано послідовно, тоді як Batch size задає кількість зображень, що створюються у межах однієї серії. Завдяки цьому користувач може отримати кілька варіантів результату за один запуск, що значно прискорює процес вибору найкращого зображення. Такий підхід особливо корисний під час експериментів із параметрами дифузії або текстовими підказками - він дозволяє порівнювати результати візуально та оперативно визначати оптимальні налаштування для подальшої генерації.

Додаткові кнопки та нижня панель

Головним елементом керування у нижній частині інтерфейсу є кнопка Generate, яка запускає процес генерації зображення. Після її натискання система ініціює обчислення на графічному процесорі (GPU), використовуючи задані параметри моделі та текстові підказки. Поруч розташовані допоміжні елементи - кнопки для завантаження вихідного зображення, очищення полів введення або збереження підказок. Вони забезпечують швидкий доступ до основних дій, полегшують повторення попередніх запитів і прискорюють робочий процес.

У нижній частині інтерфейсу розміщується службова інформаційна панель, де відображаються технічні дані - статус підключення до API, посилання на GitHub-репозиторій проєкту, а також логотип Gradio, на базі якого реалізовано вебінтерфейс. Окремі пункти, такі як Startup profile та Reload UI, використовуються для оновлення конфігурації системи або перезапуску інтерфейсу без необхідності повного перезавантаження програми. Таким чином, інтерфейс AUTOMATIC1111 Stable Diffusion WebUI поєднує гнучкість, функціональність і зручність користування. Він дозволяє не лише керувати параметрами генерації, а й детально налаштовувати процес створення

зображень, досягаючи високої якості, реалістичності та художньої виразності результатів.

Основні параметри генерації

Панель параметрів Stable Diffusion WebUI забезпечує точне керування процесом синтезу зображень і дає змогу досягати потрібного стилю, деталізації та рівня якості.

Одним із найважливіших параметрів є Sampling Method (*метод вибірки*), який визначає алгоритм дифузійного процесу - тобто спосіб, за допомогою якого модель поступово перетворює шум у структуроване зображення.

- Метод Euler забезпечує швидку й стабільну генерацію, підходить для створення простих або невеликих зображень.
- Варіант Euler a надає результатам більш художнього, варіативного характеру, змінюючи структуру деталей на кожному кроці.
- Алгоритм DPM++ 2M забезпечує контрольовану генерацію з високою якістю, зберігаючи плавні переходи та реалістичні текстури.
- Метод DDIM часто використовується, коли необхідна підвищена деталізація або точна передача мікроструктур.

Вибір алгоритму суттєво впливає на баланс між швидкістю, реалістичністю та стилем результату.

Параметр Sampling Steps визначає кількість ітерацій, які модель виконує під час процесу відновлення зображення. Менша кількість кроків (приблизно 20) забезпечує швидке отримання результату, однак із дещо нижчою деталізацією. Зі збільшенням кількості кроків (до 50 і більше) зростає якість і глибина опрацювання, проте відповідно збільшується час генерації. Отже, користувач має змогу самостійно обрати оптимальний компроміс між швидкістю виконання й якістю кінцевого результату.

Розділ Width / Height, CFG Scale, Seed, Batch та Hires.fix

Параметри Width і Height визначають розмір кінцевого зображення в пікселях і безпосередньо впливають на пропорції, формат та рівень деталізації результату. Зі збільшенням роздільної здатності зростають обчислювальні

витрати, оскільки модель повинна генерувати більше пікселів і враховувати складніші візуальні взаємозв'язки між ними. Для створення великих форматів часто застосовується функція Hires.fix, яка дозволяє спочатку згенерувати базове зображення в нижчій роздільній здатності, а потім масштабувати його без втрати якості, зберігаючи композицію та додаючи додаткові деталі. Такий підхід забезпечує високу чіткість і зменшує ризик появи артефактів під час збільшення масштабу.

Параметр CFG Scale (Classifier-Free Guidance Scale) визначає ступінь відповідності між текстовою підказкою та згенерованим результатом. Низькі значення (5–7) забезпечують більшу свободу творчої інтерпретації, дозволяючи моделі відходити від точного змісту запиту та створювати більш художні, варіативні результати. Високі значення (10 і більше) змушують систему суворо дотримуватися текстової інструкції, формуючи точні, але іноді менш «живі» зображення. Правильний вибір CFG Scale дозволяє знайти баланс між креативністю та семантичною точністю.

Параметр Seed (початкове значення випадковості) визначає, з якої точки модель розпочинає процес генерації. Якщо встановити -1, система автоматично генерує випадкове початкове число, завдяки чому кожен запуск створює унікальне зображення, навіть за однакових налаштувань. Якщо ж вказати певне значення (наприклад, 42), можна відтворити той самий результат, що є зручним для повторних експериментів, демонстрації стабільності або порівняльного аналізу.

Параметри Batch Count і Batch Size визначають кількість зображень, що створюються за один сеанс генерації.

- Batch Count задає кількість послідовних серій, які система генерує одна за одною.

- Batch Size визначає кількість зображень, що створюються одночасно в межах однієї серії. Такі налаштування зручні для отримання кількох варіантів результату або формування наборів із різними композиційними варіаціями. Водночас збільшення Batch Size потребує більшого обсягу відеопам'яті

(VRAM) і може підвищити час обробки, особливо під час роботи з моделями великої роздільної здатності.

Функція Hires.fix (High-Resolution Fix) виконує вторинну генерацію зображення з метою покращення його якості після первинного рендеру. Вона дозволяє підвищити роздільну здатність, деталізацію та чіткість контурів, додаючи дрібні текстури та мікродеталі. Ця функція є особливо корисною при створенні великоформатних візуалізацій, плакатів або медичних і гістологічних зображень, де навіть незначні структурні деталі мають суттєве значення для аналізу. Проте слід враховувати, що використання Hires.fix збільшує час генерації та навантаження на обчислювальні ресурси системи.

Окремої уваги заслуговує параметр Negative Prompt, який використовується для усунення небажаних елементів або артефактів під час генерації. У цьому полі користувач може вказати, які об'єкти, стилі чи ефекти не повинні бути присутні у фінальному результаті. Наприклад, такі терміни, як “grainy”, “blurry” чи “low quality” допомагають уникнути зернистості, розмиття або зниження якості зображення. Використання негативних підказок дозволяє зробити результат точнішим, естетично чистішим і більш професійним, особливо у випадках, коли необхідна чітка передача дрібних структур - наприклад, під час синтезу гістологічних зображень.

Завдяки сукупності параметрів інтерфейс AUTOMATIC1111 WebUI забезпечує користувачеві повний контроль над усіма аспектами генерації. Можна варіювати стиль, ступінь деталізації, роздільну здатність і навіть рівень випадковості, отримуючи зображення, які максимально відповідають поставленим творчим або науковим завданням. Гнучка комбінація налаштувань дозволяє як досягати стабільних, відтворюваних результатів, так і експериментувати з новими візуальними підходами. Це робить середовище Stable Diffusion універсальним інструментом для художників, дизайнерів, дослідників та фахівців із біомедичної візуалізації.

На рисунку 3.3 представлено інтерфейс вкладки Train → Create Hypernetwork, яка призначена для створення власних гіпермоделей

(Hypernetwork). Цей розділ дає змогу користувачу виконувати тонке налаштування процесу генерації, адаптуючи його до конкретних стилів, об'єктів або тематичних наборів даних. Під час створення нової гіпермоделі задається її назва, а також конфігураційні параметри, що визначають рівень роздільної здатності та архітектуру мережі.

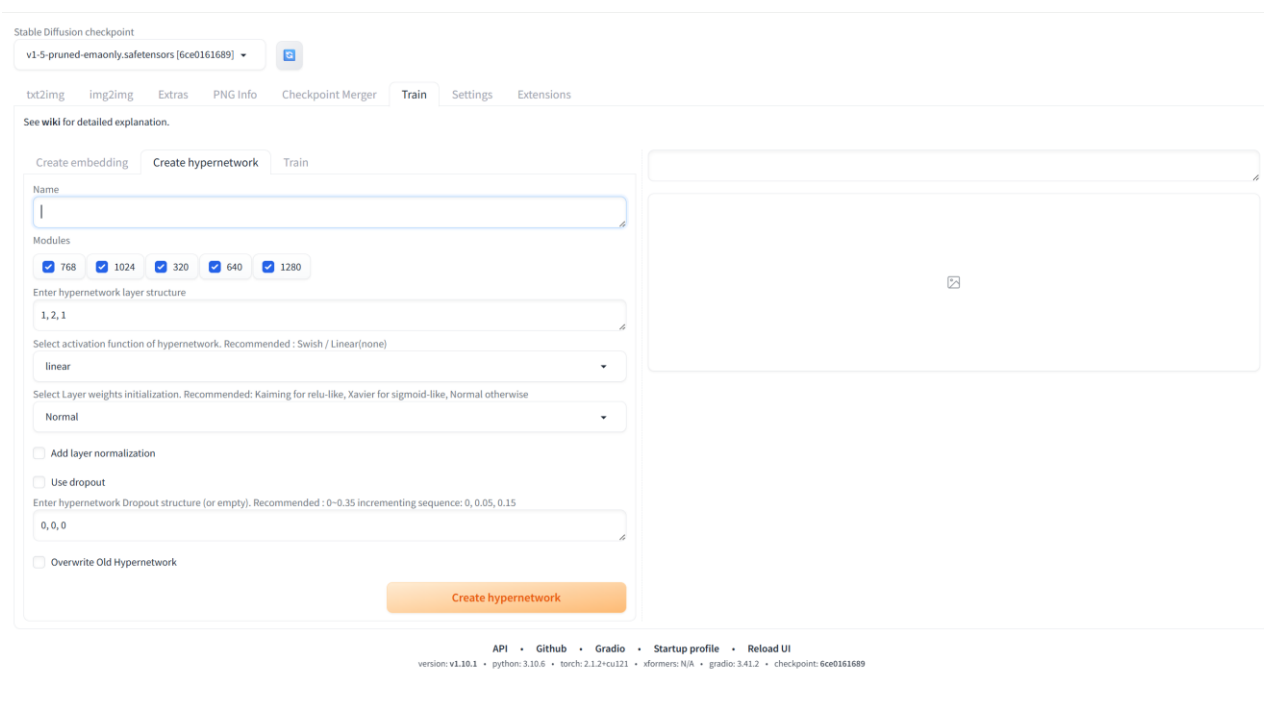


Рисунок 3.3 – Зображення веб-інтерфейсу AUTOMATIC1111 (Train)

Під час створення або навчання гіпермоделей у середовищі Stable Diffusion WebUI користувач має можливість задавати низку параметрів, які впливають на архітектуру, поведінку та якість результатів моделі. Зокрема, параметри роздільної здатності під час навчання можуть становити 320, 640, 768, 1024 або 1280 пікселів. Вибір конкретного значення залежить від бажаного рівня деталізації: чим більша роздільна здатність, тим точніше відтворюються деталі зображення, проте тим вищі обчислювальні витрати та час обробки. Параметр Enter hypernetwork layer structure визначає архітектуру гіпермоделі – кількість шарів та кількість вузлів у кожному з них. Наприклад, конфігурація 1, 2, 1 означає, що модель має три шари, де середній шар містить

удвічі більше вузлів, ніж інші. Така структура сприяє підвищенню гнучкості та стабільності навчання, дозволяючи моделі краще адаптуватися до різних типів даних.

Наступним етапом є Select activation function of hypernetwork, який визначає функцію активації – тобто спосіб, у який модель обробляє сигнали між шарами. Зазвичай використовують Swish, нелінійну функцію, що показує високу ефективність у генеративних моделях, або Linear (none) - варіант без додаткової активації, який забезпечує простішу структуру мережі.

Параметр Select Layer weights initialization задає метод ініціалізації ваг нейронної мережі. Найчастіше застосовуються такі підходи: Kaiming - ефективний для активацій на основі ReLU, Xavier - добре працює з функціями sigmoid і tanh, а Normal є універсальним варіантом, що підходить для більшості сценаріїв навчання. Опція Add layer normalization дозволяє активувати нормалізацію шарів, яка стабілізує процес навчання, запобігає надмірному росту градієнтів і покращує точність результатів. Додатково параметр Use dropout відповідає за застосування техніки Dropout - випадкового “вимикання” частини нейронів у процесі навчання для запобігання перенавчанню. У полі Enter hypernetwork Dropout structure задаються числові значення, наприклад 0, 0.05, 0.15, що означає поступове збільшення ймовірності вимкнення нейронів у кожному наступному шарі.

Під час оновлення або повторного навчання моделі можна скористатися параметром Overwrite Old Hypernetwork, який дозволяє перезаписати вже існуючу гіпермодель із тією ж назвою. Це зручно для оновлення попередніх результатів без створення нових копій. Завершальним кроком є натискання кнопки Create hypernetwork, що запускає процес створення гіпермоделі з вибраними параметрами. Гіпермоделі дають змогу розширити можливості Stable Diffusion, оптимізуючи процес генерації та роблячи модель здатною відтворювати унікальні стилі, кольорові палітри й авторські техніки. Завдяки цьому користувач може створювати персоналізовані моделі, орієнтовані на

конкретні задачі – від художніх експериментів до наукової візуалізації гістологічних або цитологічних структур.

Отже, дана секція інтерфейсу є потужним інструментом персоналізації, який надає контроль над архітектурою, активацією, нормалізацією та параметрами навчання. Вона відкриває широкі можливості для експериментів із глибиною, складністю й поведінкою нейронних мереж, забезпечуючи високу гнучкість і точність у процесі створення генеративних моделей.

На рисунку 3.4 представлено приклад вкладки Train у середовищі Stable Diffusion WebUI, яка призначена для навчання гіпермоделей (Hypernetwork) або ембеддингів (Embedding). Цей інтерфейс надає користувачу можливість гнучко задавати параметри навчання, обирати тип моделі, відстежувати прогрес і контролювати процес тренування в режимі реального часу.

У верхній частині вкладки розміщено перемикач Embedding / Hypernetwork, який визначає тип об'єкта для навчання: ембеддинг (векторне представлення певного поняття, стилю чи об'єкта) або гіпермодель (багаторівнева структура, що дозволяє адаптувати модель до конкретного стилю, типу даних чи завдань).

У секції параметрів навчання задаються основні характеристики процесу. Параметр Learning Rate визначає швидкість оновлення ваг нейронної мережі. Для ембеддингів зазвичай застосовують значення близько 0.005, тоді як для гіпермоделей – 0.00001, що забезпечує стабільне, поступове навчання без різких коливань результатів. Параметр Gradient Clipping обмежує максимальну величину градієнта, запобігаючи перенасиченню та нестабільності – у полі Value зазвичай вказується 0.1. Значення Batch Size задає кількість зображень, що обробляються за один крок тренування (наприклад, 1 означає, що модель обробляє одне зображення на ітерацію). Параметр Gradient Accumulation Steps дозволяє накопичувати градієнти протягом кількох кроків перед оновленням ваг, що корисно при обмежених ресурсах, імітуючи ефект більших пакетів даних.

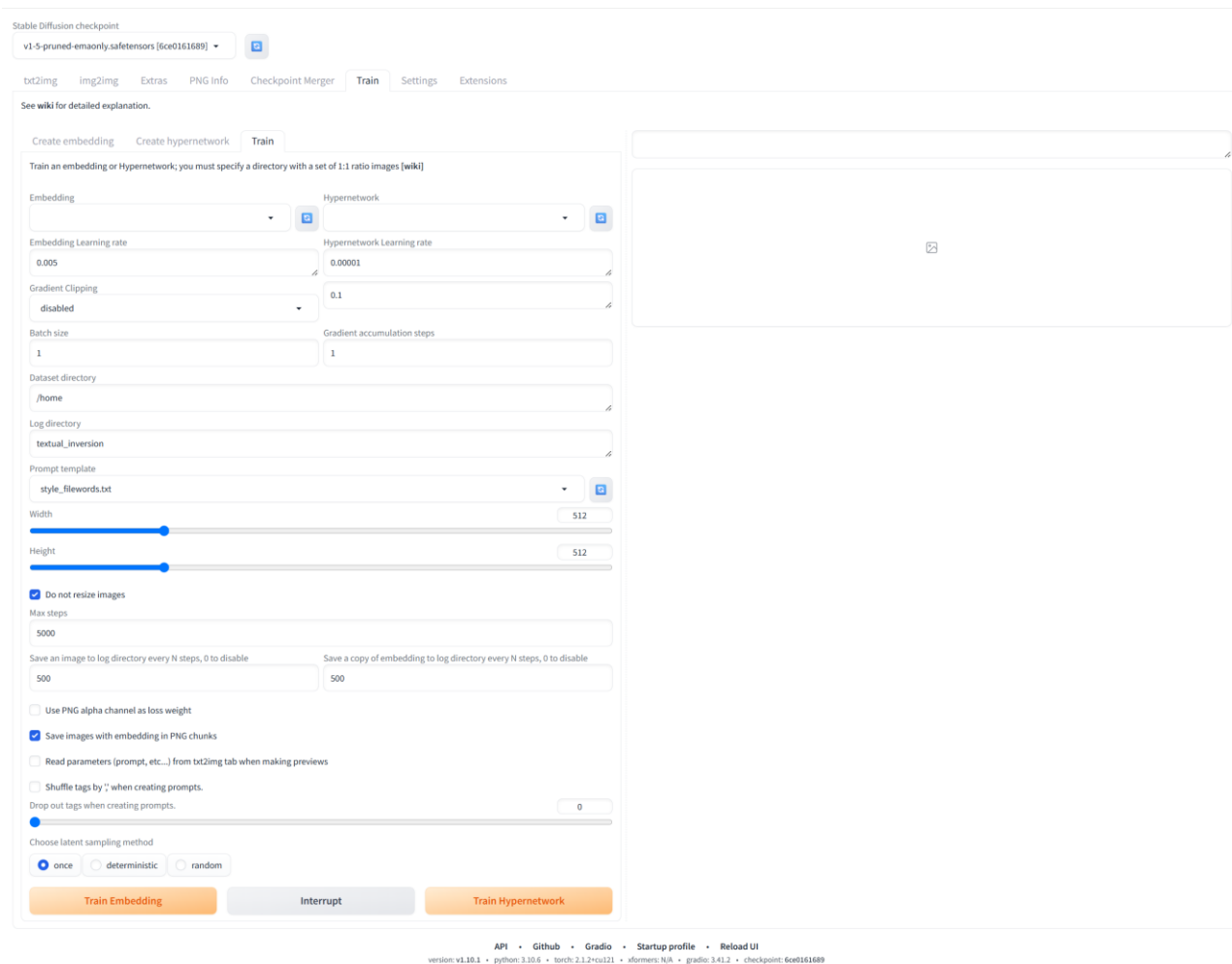


Рисунок 3.4 – Зображення веб-інтерфейсу AUTOMATIC111

Наступна секція стосується роботи з навчальними даними. Поле Dataset Directory вказує шлях до папки, що містить зображення для навчання, Log Directory – місце збереження результатів, журналів і проміжних артефактів. Параметр Prompt Template визначає шаблон текстових підказок, який використовується для генерації (наприклад, style_fireworks.txt). Параметри Width та Height задають роздільну здатність вхідних зображень - найчастіше 512×512 пікселів. Якщо активовано пункт Do not resize images, система зберігає оригінальні розміри файлів без масштабування.

Секція логування дозволяє контролювати частоту збереження проміжних результатів. Поле Max Steps задає загальну кількість ітерацій навчання (наприклад, 40 000 кроків), а параметр Save an image to log directory every N steps визначає інтервал збереження зображень для візуального

моніторингу прогресу. Це дозволяє оцінювати зміни в якості згенерованих результатів на різних етапах навчання.

У блоці додаткових опцій передбачено низку розширених налаштувань. Опція Use PNG alpha channel as loss weight дозволяє використовувати альфа-канал зображення як ваговий коефіцієнт втрат, що підвищує точність обробки складних зразків. Параметр Shuffle tags by ‘;’ when creating prompts перемішує теги у текстових підказках, збільшуючи варіативність навчальних прикладів. Також користувач може обрати метод вибірки латентного простору (once, deterministic або random), який визначає, як модель оброблятиме приховані змінні під час тренування.

У нижній частині вкладки розташовані основні кнопки керування процесом: Train Embedding запускає навчання ембеддингу, Train Hypernetwork – тренування гіпермоделі, а Interrupt дозволяє зупинити процес у будь-який момент без втрати поточного прогресу.

Загалом, вкладка Train у Stable Diffusion WebUI забезпечує повний контроль над навчанням користувацьких моделей - від вибору архітектури й швидкості навчання до параметрів збереження результатів. Її функціональність дозволяє створювати індивідуальні моделі, адаптовані до конкретних завдань - як творчих (генерація зображень у певному стилі), так і наукових, наприклад, моделювання та синтез гістологічних зображень з високою точністю структурних особливостей.

3.2 Апаратно-програмне забезпечення для проведення комп’ютерних експериментів

Для проведення комп’ютерних експериментів використовувалася хмарна конфігурація Stable Diffusion WebUI від Jarvis Labs, що забезпечує високу

продуктивність для генерації зображень за допомогою моделей дифузійного типу.

1. Програмне середовище:

- Операційна система: Ubuntu 22.04.3 LTS (jammy)
- Версія WebUI: AUTOMATIC1111 v1.10.1
- Python: 3.10.6
- Torch: 2.1.2+cu121
- Xformers: 0.0.24+post1
- Gradio: 3.41.2
- Checkpoint: v1-5-pruned-emaonly.safetensors (6ec0616689)

2. Обчислювальні ресурси:

- Графічний процесор: 1 × NVIDIA RTX A6000 (Ampere, CUDA 12.3)
- Центральний процесор: 7 CPU
- Оперативна пам'ять: 32 GB RAM
- Відеопам'ять (VRAM): 48 GB

Ця апаратна конфігурація забезпечує достатній обчислювальний ресурс для стабільної роботи моделей Stable Diffusion 1.5, дозволяючи виконувати генерацію зображень високої якості з підтримкою прискорення на GPU.

Параметри середовища Stable Diffusion WebUI

Основні налаштування під час експериментів мали такі значення:

- Sampling method: DPM++ 2M
- Sampling steps: 20
- Resolution: Width = 512 px, Height = 512 px
- CFG Scale: 7
- Seed: -1 (випадкове значення для кожного запуску)
- Batch count: 1
- Batch size: 1
- Schedule type: Automatic
- Script: None

Таке поєднання параметрів дозволяє досягти балансу між швидкістю рендерингу та якістю зображення. Метод DPM++ 2M забезпечує плавну реконструкцію деталей, а параметр CFG Scale = 7 оптимально поєднує точність виконання текстової підказки з художньою варіативністю результату.

У таблицях 3.1 та 3.2 наведено відповідно параметри орендної сесії обчислювальних ресурсів та основні характеристики процесу навчання вибірок.

Таблиця 3.1 – Інформація про оренду

Дата	Тип пристрою	Кількість GPUS	Кількість CPUS	RAM (GB)	VRAM (GB)	Обсяг сховища (GB)	Hours Ran	Total Cost (\$)
15.07.2025	A6000 Ampere	1	7	32	48	190	6H	5.08

Таблиця 3.2 – Інформація про навчання на основі різних вибірок

Набір даних	Кількість зображень до розширення	Кількість зображень після розширення	Кількість кроків	Кількість епох	Загальний час навчання у (хв)	Час на один крок (хв)	Час на одну епоху (сек)
HISTO_01	21	150	20000	133	120	0.36	54.1
HISTO_02	29	150	20000	133	120	0.36	54.1
HISTO_03	37	150	20000	133	120	0.36	54.1

Перед проведенням експериментів початкова вибірка зображень була доповнена за допомогою алгоритмів афінних перетворень, реалізованих у бібліотеці Rudi, написаній мовою програмування Python.

Під час підготовки навчальної вибірки було виконано автоматичну аугментацію гістологічних зображень із використанням бібліотеки *Rudi*, розробленої мовою програмування Python. Цей підхід дозволив не лише збільшити обсяг даних, але й підвищити стійкість моделі до варіацій у зображеннях, які часто зустрічаються у реальних гістологічних дослідженнях.

У процесі обробки до кожного зображення з імовірністю близько 70% застосовувалися різні афінні перетворення, що забезпечувало створення нових візуально достовірних варіацій. До найпоширеніших трансформацій належали повороти зображень у межах $\pm 15^\circ$, що імітують зміни орієнтації гістологічних зрізів, та масштабування у межах коефіцієнтів від 1.0 до 1.25, яке дозволяло моделювати незначні зміни масштабу або збільшення під час мікроскопії. Також використовувалися локальні спотворення, що змінювали форму клітинних структур без втрати їхньої цілісності. Інтенсивність цих змін регулювалася спеціальним параметром, який визначав силу геометричних і кольорових модифікацій, зберігаючи при цьому реалістичність тканинної структури.

Для кожного початкового зображення було створено приблизно 120–150 нових зразків, що суттєво розширило навчальну базу. Такий підхід дозволив зменшити проблему дисбалансу між класами, особливо для рідкісних типів тканин, і забезпечив більшу репрезентативність вибірки.

Алгоритм аугментації складався з кількох послідовних етапів. Спочатку виконувалося автоматичне обрізання зображень до квадратного формату, що спрощувало подальше завантаження в модель. Потім викликала команда *rudi augment*, яка застосовувала визначені параметри випадковим чином, створюючи численні унікальні копії. Отримані результати автоматично зберігалися у цільову директорію для аугментованих даних, після чого інтегрувалися з початковими зображеннями для формування підсумкової вибірки.

Завдяки такому підходу бібліотека *Rudi* забезпечила повну автоматизацію процесу підготовки даних і дала змогу підвищити варіативність

та узагальнюючу здатність моделі. Це, у свою чергу, зменшило ризик перенавчання та дозволило моделі ефективніше працювати з новими, раніше невідомими гістологічними зразками. Крім того, створена аугментована база даних стала основою для подальшого навчання та тестування дифузійної моделі, сприяючи підвищенню її точності, стабільності та здатності відтворювати реалістичні структурні особливості тканин.

3.3 Результати комп'ютерних експериментів

У ході проведення експериментів було використано три розширені набори гістологічних даних, що відповідали трьом основним класам тканин: HISTO_01, HISTO_02 та HISTO_03. Кожен із цих наборів був попередньо підготовлений шляхом аугментації, у результаті чого кількість зображень у кожній вибірці становила приблизно 150. Такий підхід дав змогу збалансувати обсяг даних між класами та забезпечити достатню кількість прикладів для ефективного навчання моделі.

Навчання моделі проводилося протягом 20 000 кроків, що дозволило досягти високого рівня деталізації зображень і стабільної збіжності результатів. Кількість епох для кожного набору становила 133, що забезпечило повне охоплення навчальних даних. Загальний час навчання для кожної вибірки складав близько 120 хвилин, що свідчить про ефективність використаної конфігурації апаратно-програмного середовища.

Час на один крок залишався сталим і становив 0,36 хвилини, що є оптимальним показником для моделі дифузійного типу з урахуванням обчислювальних можливостей системи. Час на одну епоху складав у середньому 54,1 секунди, забезпечуючи стабільну швидкість обробки даних при збереженні високої якості результатів.

У процесі експериментів було помічено, що модель добре узагальнює навчальні дані для всіх трьох класів. Найбільш реалістичні та структурно точні зображення спостерігалися для наборів HISTO_01 та HISTO_03, де збереження клітинних структур, ядер і тканинної архітекτονіки досягало високого рівня візуальної відповідності реальним гістологічним зрізам. У наборі HISTO_02 результати також продемонстрували високу якість, хоча іноді відзначалися незначні розбіжності у текстурі стромальних ділянок.

На рисунку 3.5 наведено приклад порівняння оригінальних та згенерованих зображень для класу HISTO_01, де чітко видно, що синтезовані зразки зберігають основні морфологічні риси тканини, зокрема структуру клітинних кластерів, форму ядер та характерне забарвлення цитоплазми.

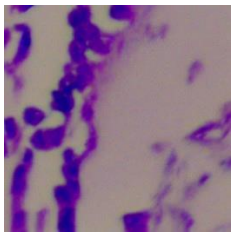
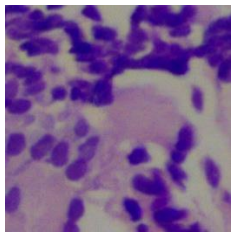
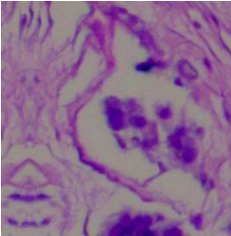
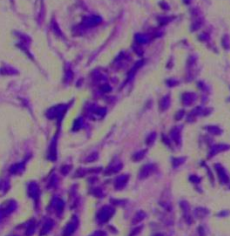
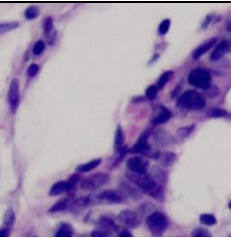
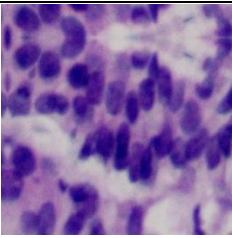
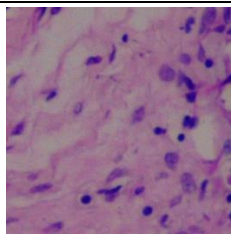
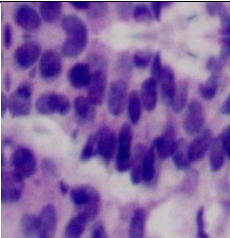
Оригінал				
Кроки	5 000	10 000	15 000	20 000
Генеровані				

Рисунок 3.5 – Приклад порівняння реальних та згенерованих зображень для класу «HISTO_01»

На рисунку 3.6 наведено результати тестування згенерованих гістологічних зображень на основі метрик FID (Fréchet Inception Distance) та IS (Inception Score), які використовувалися для оцінки якості та різноманітності результатів генерації. Під час експериментів спостерігалось поступове покращення показників якості на проміжних етапах навчання.

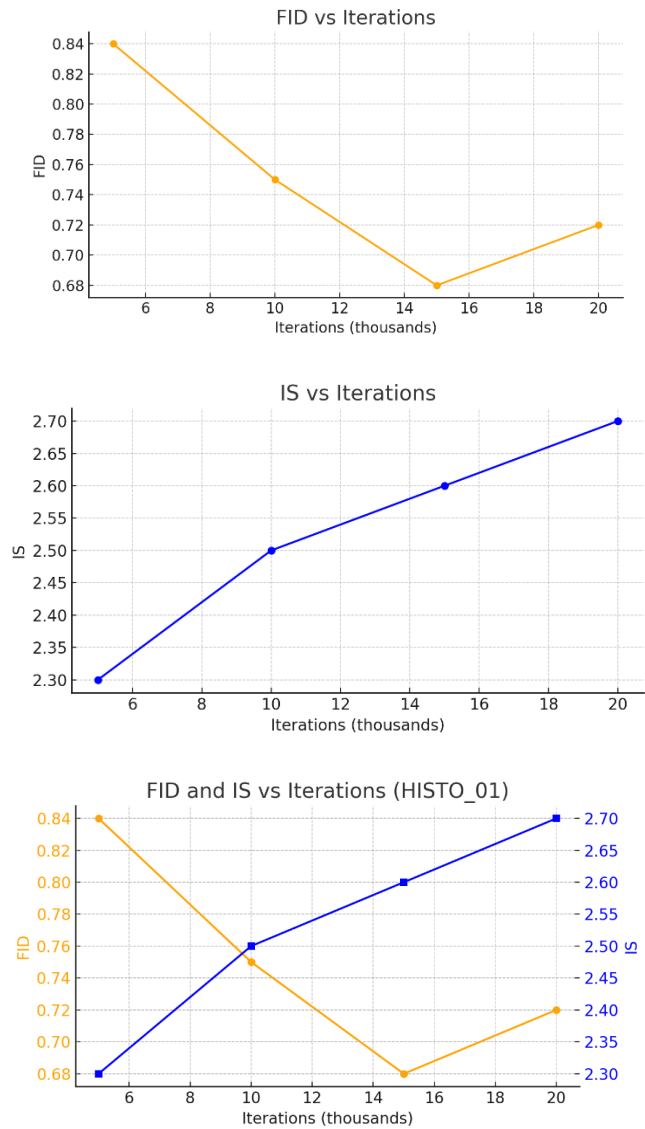


Рисунок 3.6 – Динаміка зміни метрик FID та IS залежно від кількості ітерацій під час навчання моделі Stable Diffusion на гістологічних зображеннях (HISTO_01)

Метрика FID, що відображає ступінь подібності між згенерованими та реальними зображеннями, коливалася в межах 0.68–0.84, поступово зменшуючись із ростом кількості ітерацій. Найкраще значення було зафіксовано приблизно на 15 000 ітераціях, що свідчить про оптимальне співвідношення між деталізацією та узагальненням моделі.

Показник IS, який характеризує різноманітність і реалістичність зображень, знаходився в діапазоні 2.3–2.7. Найвищий результат спостерігався на 20 000 ітераціях, що свідчить про підвищення якості генерації та покращення візуальної варіативності створених зображень.

Отримані значення метрик вказують на те, що модель успішно навчилася відтворювати основні морфологічні особливості гістологічних зразків, забезпечуючи задовільний баланс між точністю, реалістичністю та різноманітністю згенерованих структур. Це підтверджує ефективність застосованої архітектури та параметрів навчання для поставленої задачі синтезу гістологічних зображень.

У таблиці 3.3 наведено експериментальні значення метрик FID та IS, отримані на різних етапах навчання моделі Stable Diffusion для набору гістологічних зображень (HISTO_01).

На рисунку 3.7 наведено приклади реальних та згенерованих зображень із другого набору даних (HISTO_02).

Таблиця 3.3 – Результати експериментів

Кількість ітерацій (×1000)	FID	IS
5	0.84	2.30
10	0.75	2.50
15	0.68	2.60
20	0.72	2.70

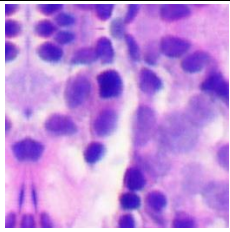
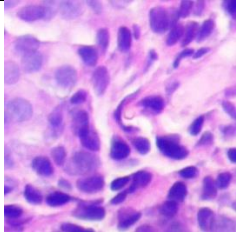
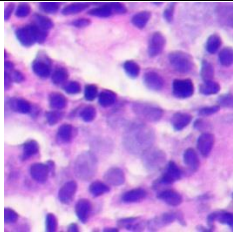
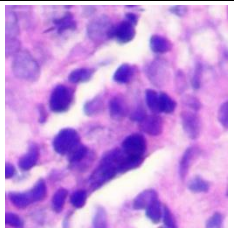
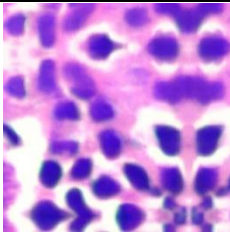
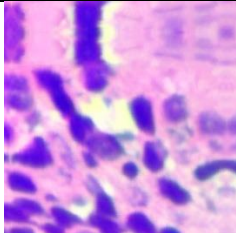
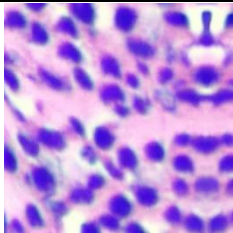
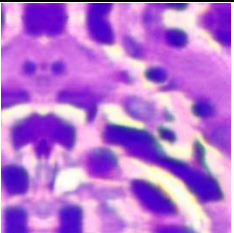
Оригінал				
Кроки	5 000	10 000	15 000	20 000
Генеровані				

Рисунок 3.7 – Приклад порівняння реальних та згенерованих зображень для класу «HISTO_02»

На рисунку 3.8 наведено результати тестування згенерованих гістологічних зображень на основі метрик FID (Frechet Inception Distance) та IS (Inception Score).

Результати експерименту для другого набору гістологічних зображень (HISTO_02) показали стабільну динаміку зміни показників якості генерації.

Метрика FID (Frechet Inception Distance) демонструє тенденцію до поступового зниження - від 0.90 на 5 тисячах ітерацій до 0.70 на 15 тисячах, після чого незначно підвищується до 0.73 на 20 тисячах. Така поведінка свідчить про покращення схожості між згенерованими та реальними зображеннями на проміжних етапах і часткове перенавчання моделі наприкінці процесу.

Показник IS (Inception Score), який відображає рівень різноманітності та реалістичності створених зображень, навпаки, демонструє зростаючий тренд - від 2.20 на початку навчання до 2.62 на 20 тисячах ітерацій. Це свідчить про поступове підвищення якості генерації та покращення структури синтезованих гістологічних зразків.

У таблиці 3.4 наведено значення метрик FID та IS, отримані під час навчання моделі Stable Diffusion для другого набору даних (HISTO_02).

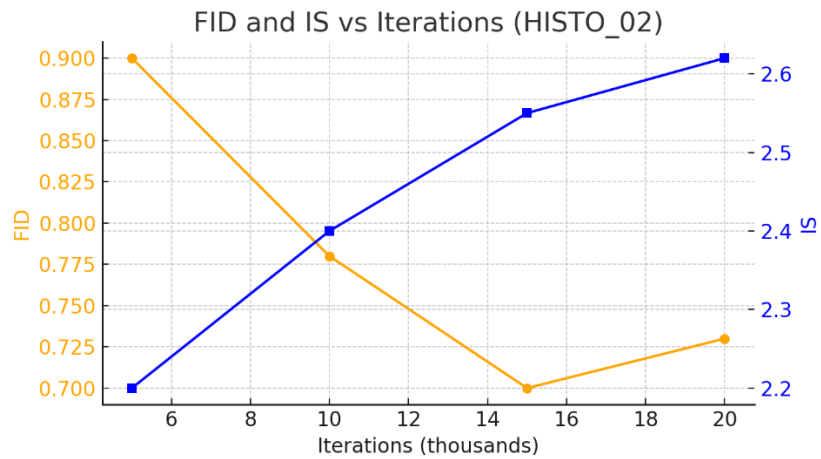
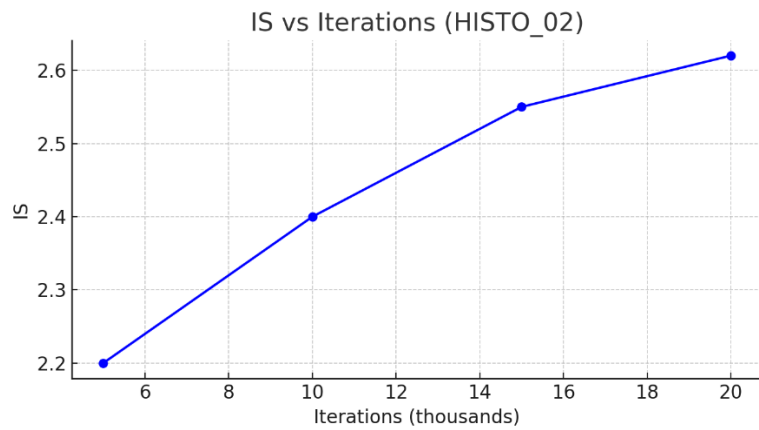
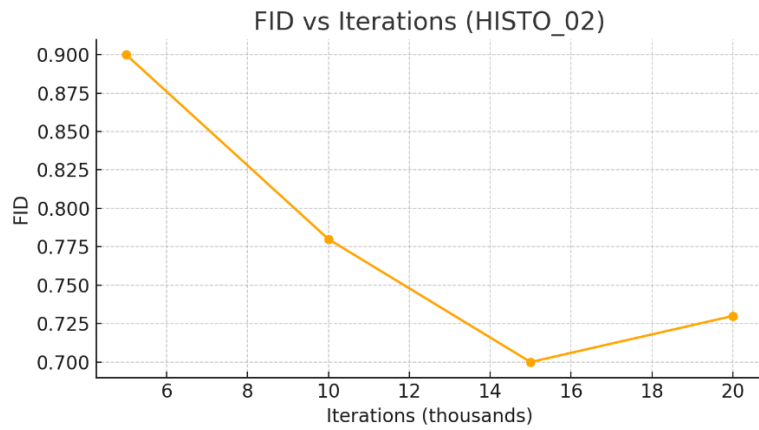


Рисунок 3.8 – Динаміка зміни метрик FID та IS залежно від кількості ітерацій під час навчання моделі Stable Diffusion на гістологічних зображеннях (HISTO_02)

Таблиця 3.4 – Результати експериментів

Кількість ітерацій (×1000)	FID	IS
5	0.90	2.20
10	0.78	2.40
15	0.70	2.55
20	0.73	2.62

На рисунку 3.9 наведено результати тестування згенерованих гістологічних зображень на основі метрик FID (Frechet Inception Distance) та IS (Inception Score).

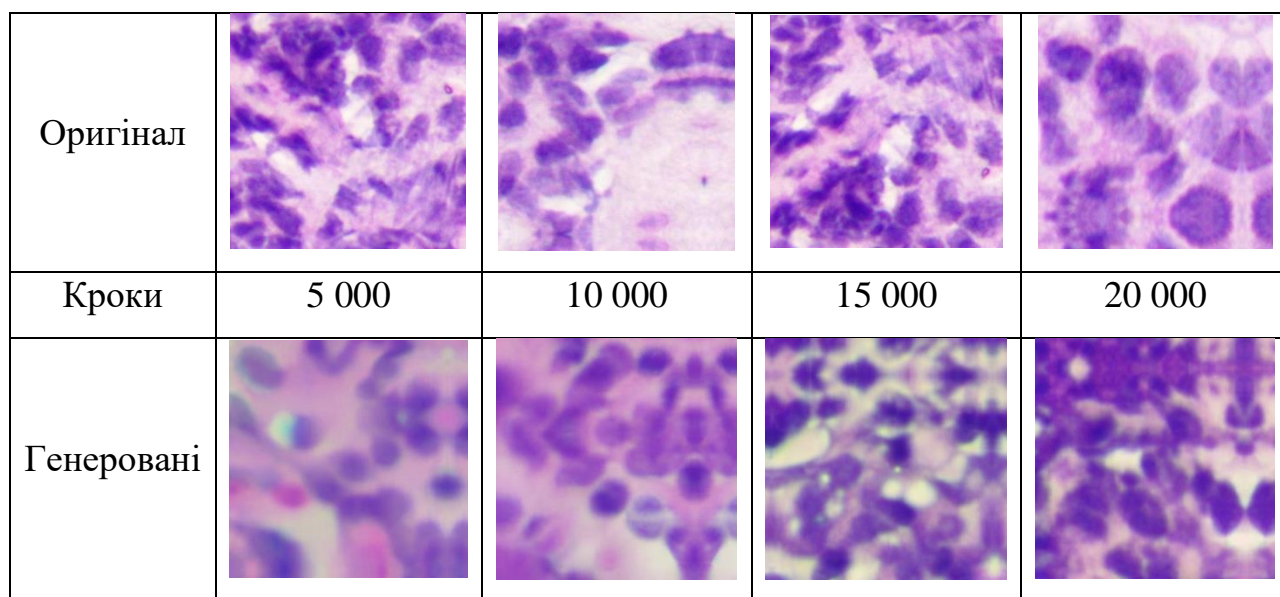


Рисунок 3.9 – Приклад порівняння реальних та згенерованих зображень для класу «HISTO_03»

На рисунку 3.10 наведено результати тестування згенерованих гістологічних зображень на основі метрик FID (Frechet Inception Distance) та IS (Inception Score).

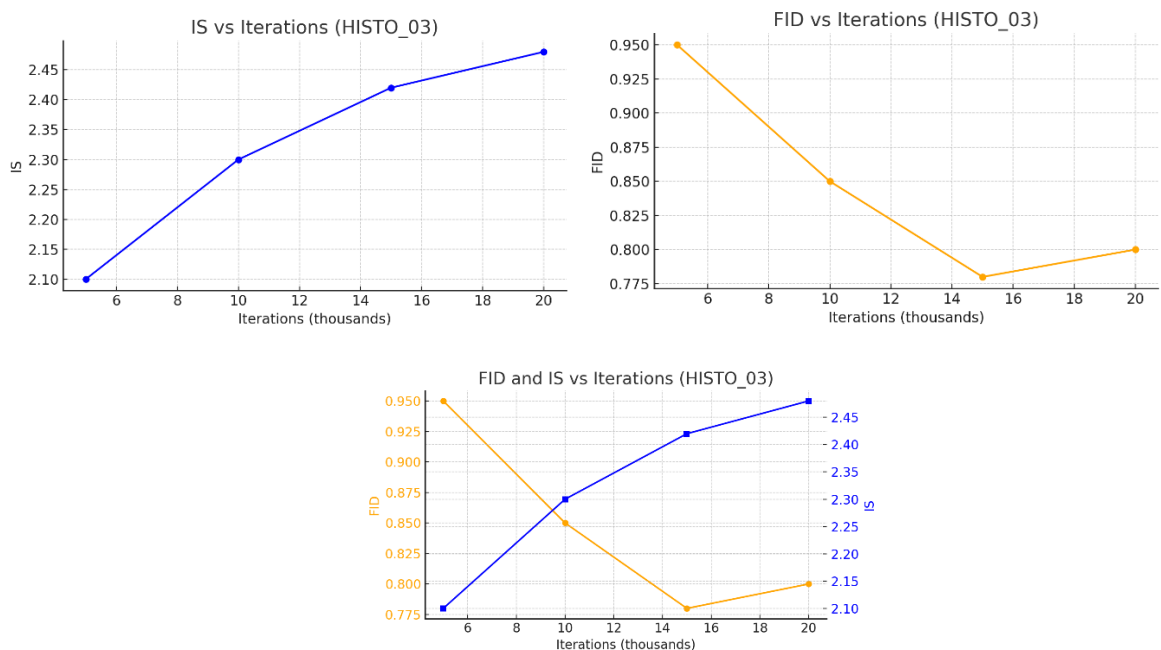


Рисунок 3.10 – Динаміка зміни метрик FID та IS залежно від кількості ітерацій під час навчання моделі Stable Diffusion на гістологічних зображеннях (HISTO_03)

Результати експерименту для третього набору гістологічних зображень (HISTO_03) свідчать про поступове покращення якості генерації впродовж процесу навчання.

Метрика FID (Fréchet Inception Distance) демонструє стабільну тенденцію до зниження - від 0.95 на 5 тисячах ітерацій до 0.78 на 15 тисячах, після чого незначно зростає до 0.80 на 20 тисячах. Така динаміка свідчить про досягнення оптимальної якості зображень на середніх етапах навчання та часткове перенавчання наприкінці.

Показник IS (Inception Score) демонструє зростаючий тренд: від 2.10 на початковому етапі до 2.48 на 20 тисячах ітерацій. Це свідчить про поступове підвищення різноманітності та реалістичності згенерованих гістологічних зразків.

У таблиці 3.5 наведено значення метрик FID та IS, отримані під час навчання моделі Stable Diffusion для третього набору даних (HISTO_03).

Таблиця 3.5 – Результати експериментів

Кількість ітерацій (×1000)	FID	IS
5	0.95	2.10
10	0.85	2.30
15	0.78	2.42
20	0.80	2.48

На рисунках 3.11 та 3.12 наведено узагальнені результати для трьох наборів гістологічних зображень (HISTO_01, HISTO_02, HISTO_03) за метриками FID та IS, які відображають загальну якість та різноманітність синтезованих зображень, отриманих у процесі навчання моделі Stable Diffusion.

В таблиці 3.6 приведені середні значення метрик FID та IS для всіх наборів даних.

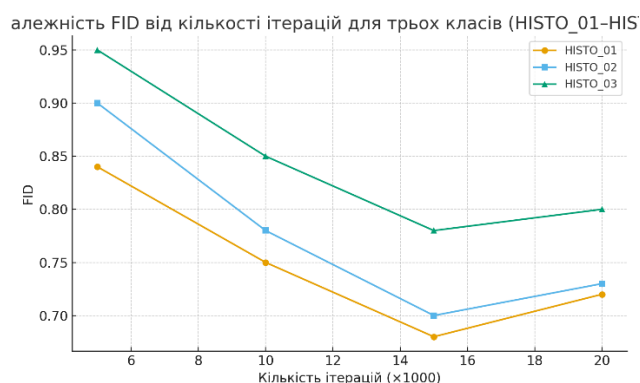


Рисунок 3.11 – Узагальнені графіки середніх значень метрик FID та IS для трьох наборів гістологічних зображень (HISTO_01–HISTO_03)

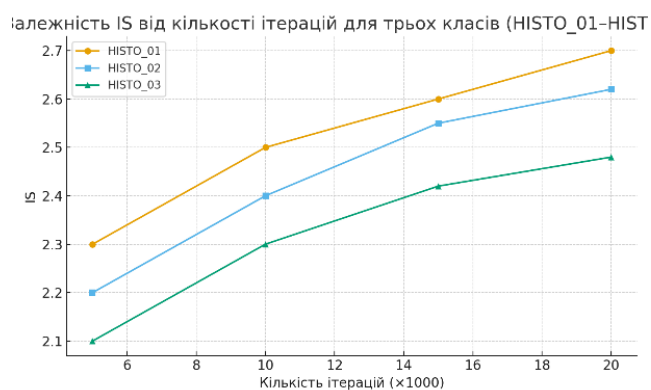


Рисунок 3.12 – Узагальнені графіки середніх значень метрик FID та IS для трьох наборів гістологічних зображень (HISTO_01–HISTO_03)

Таблиця 3.6 – Середні значення метрик FID та IS для всіх наборів даних

Набір даних	Середнє FID	Середнє IS
HISTO_01	0.747	2.525
HISTO_02	0.778	2.443
HISTO_03	0.845	2.325
Середнє по всіх наборах	0.790	2.43

Метрики FID (Fréchet Inception Distance) та IS (Inception Score) є стандартними показниками для оцінки якості згенерованих зображень у генеративних моделях.

FID оцінює ступінь подібності між реальними та синтезованими зображеннями, порівнюючи статистичні характеристики їхніх ознак у просторах ознак нейронних мереж. Чим нижче значення FID, тим ближче згенеровані зображення до реальних, а отже, модель краще відтворює текстуру, кольорову гамму та морфологічну структуру гістологічних зрізів. У контексті даного експерименту це означає, що модель Stable Diffusion успішно навчилася відтворювати характерні риси клітинних утворень, таких як межі клітин, ядра, тканинні структури тощо.

IS, навпаки, вимірює різноманітність і реалістичність згенерованих результатів. Високе значення IS свідчить про те, що модель створює зображення з широким спектром візуальних особливостей, не обмежуючись повторенням одних і тих самих структур. Це особливо важливо для гістологічних даних, де навіть у межах одного класу можуть бути суттєві морфологічні варіації.

Згідно з таблицею 3.6, найкращі показники якості генерації продемонстрував набір HISTO_01, що має найнижчий середній FID (0.747) та найвищий IS (2.525).

Це свідчить про те, що модель добре адаптувалася до морфологічних особливостей цього типу зразків, відтворюючи як текстурну глибину, так і природну варіативність структури.

Набір HISTO_02 показав стабільні результати, що підтверджує його збалансовану складність - модель змогла досягти компромісу між точністю відтворення (FID = 0.778) та різноманітністю (IS = 2.443).

Для набору HISTO_03 спостерігається вищий показник FID (0.845), що вказує на дещо нижчу подібність до реальних даних. Це можна пояснити складнішою морфологічною структурою гістологічних зразків цього типу - можливо, наявністю більш виражених неоднорідностей, переплетених тканин або мікроструктур, які складніше точно відтворити навіть для дифузійних моделей.

У середньому ж по всіх наборах значення метрик FID ≈ 0.79 та IS ≈ 2.43 свідчать про високу якість генерації та стабільність навчання моделі, що підтверджує ефективність використання Stable Diffusion для синтезу реалістичних гістологічних зображень у межах поставленого експерименту.

3.4 Висновки до розділу 3

У третьому розділі було проведено комплексний аналіз процесу генерації та навчання моделей на основі архітектури Stable Diffusion, із докладним описом її інтерфейсу, налаштувань і можливостей. Детально розглянуто роботу середовища AUTOMATIC1111 WebUI, яке забезпечує інтуїтивне керування параметрами генерації, навчання гіпермоделей (Hypernetwork) та підготовки вибірок для експериментів [49].

Наведено технічні характеристики апаратно-програмної конфігурації, що використовувалась під час досліджень – зокрема, систему на базі GPU NVIDIA A6000 (48 GB VRAM), 32 GB оперативної пам'яті та 7 CPU, що забезпечило високу продуктивність при виконанні обчислювально складних операцій. Така конфігурація дозволила досягти стабільної роботи дифузійної моделі та ефективного використання ресурсів під час навчання й генерації зображень.

Для моделювання було створено три експериментальні набори гістологічних зображень (HISTO_01, HISTO_02, HISTO_03), кожен з яких пройшов процес розширення (аугментації) за допомогою бібліотеки Rudi. Застосування афінних перетворень (повороти, масштабування, зміна орієнтації тощо) дозволило збільшити обсяг навчальних даних і підвищити узагальнюючу здатність моделі.

Процес навчання проводився протягом 20 000 ітерацій для кожного набору зображень, що забезпечило достатню глибину оптимізації параметрів. У результаті було отримано згенеровані гістологічні зразки, які зберігають ключові морфологічні характеристики клітинних структур, такі як форма, текстура тканин і кольорова гармонія.

Для оцінки якості синтезованих зображень застосовувалися дві основні метрики - FID (Fréchet Inception Distance) та IS (Inception Score). Їхні значення засвідчили високу подібність згенерованих зображень до реальних

гістологічних даних, а також належний рівень різноманітності результатів. Середні показники метрик ($FID \approx 0.79$, $IS \approx 2.43$) демонструють стабільну генеративну здатність моделі та збалансованість між реалістичністю й варіативністю структури.

Таким чином, проведені експерименти підтвердили ефективність використання моделі Stable Diffusion для задач генерації гістологічних зображень. Отримані результати свідчать про перспективність застосування дифузійних моделей у медичній візуалізації - як для розширення навчальних датасетів, так і для створення синтетичних зразків, придатних для діагностичних і дослідницьких цілей.

ВИСНОВКИ

1. Проведено аналіз гістологічних зображень і показано, що задача розширення цих зображень є актуальною.
2. Здійснено аналіз програмних засобів і обґрунтовано використання програмного засобу Stable Diffusion для генерування гістологічних зображень.
3. Досліджено дифузійні моделі та алгоритми синтезу зображень на основі стохастичних диференціальних рівнянь і розроблено алгоритми синтезу гістологічних зображень, що дало можливість провести обчислювальні експерименти в середовищі Stable Diffusion.
4. Описано інтерфейс програмного середовища Stable Diffusion, завдяки чому було проведено обчислювальні експерименти для генерування гістологічних зображень.
5. Проведено обчислювальні експерименти в середовищі Stable Diffusion з використанням реальних гістологічних зображень і отримано розширений штучний набір даних цих зображень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Павловський Т.М., Богонос Д.С. Дифузійні моделі для синтезу цитологічних зображень. *Інтелектуальні комп'ютерні системи та мережі* : тези доп. III Всеукр. наук.-практ. конф. студентів, аспірантів, молодих вчених (25 листопада 2025 р.). Тернопіль : ЗУНУ, 2025. С. 154-155.

2. Павловський Т.М., Богонос Д.С. Синтез гістологічних зображень на основі дифузійних моделей. *Інтелектуальні комп'ютерні системи та мережі*: тези доп. III Всеукр. наук.-практ. конф. студентів, аспірантів, молодих вчених (25 листопада 2025 р.). Тернопіль : ЗУНУ, 2025. С. 156-157.

3. Березький О.М., Мельник Г.М. Методичні рекомендації до виконання кваліфікаційної роботи з освітнього ступеня "Магістр". Спеціальність: 123 – Комп'ютерна інженерія. Магістерська програма – Комп'ютерна інженерія". Тернопіль: ЗУНУ, 2024. 32 с.

4. Методичні вказівки до оформлення курсових проектів, звітів про проходження практики, випускних кваліфікаційних робіт для студентів спеціальності «Комп'ютерна інженерія» / І.В. Гураль, Л.О. Дубчак / Під ред. О.М. Березького. Тернопіль: ТНЕУ, 2019. 33 с.

5. Березький О.М., Мельник Г.М., Батько Ю.М. Текстурна сегментація біомедичних зображень на основі просторових моментів. Матеріали 4-ї міжнародної науково-технічної конференції "Комп'ютерні науки та інформаційні технології". 15-17 жовтня, 2009, Львів, Україна. Львів: ПП "Вежа і Ко", 2009 С.42–45.

6. Березький О. М., Березька К. М., Попіна С. Ю. Статистичне оброблення цитологічних зображень. Вісник Хмельницького національного університету: зб. наук.-техн. праць. Сер.: Технічні науки. 2012. № 5. С. 161–164.

7. Березький О. М., Березька К. М., Батько Ю. М., Мельник Г. М. Синтез альтернативних рішень при структурному проектуванні систем

автоматизованої мікроскопії. Науковий вісник НЛТУ України: зб. наук.-техн. праць. Львів: РВВ НЛТУ України. 2009. Вип. 19.5. С. 258-268.

8. Березький О. М., Мельник Г.М., Батько Ю.М., Дацко Т. В. Інтелектуальна система для діагностування різних форм раку молочної залози на основі аналізу гістологічних і цитологічних зображень. Науковий вісник НЛТУ України: зб. наук.-техн. праць. Львів: РВВ НЛТУ України. 2013. Вип. 23.13. С. 357-367.

9. Березький О. М., Батько Ю.М., Мельник Г.М. Інформаційно-аналітична система дослідження та діагностування пухлинних клітин на основі аналізу їх зображень. Вісник Хмельницького національного університету. Технічні науки. 2008. №4. С.33-41.

10. Березький О. М., Батько Ю.М., Мельник Г.М. Комп'ютерна система аналізу біомедичних зображень. Вісник Національного університету «Львівська політехніка». Комп'ютерні науки та інформаційні технології. 2009. № 570. С. 84-89.

11. Березький О. М. Методи та алгоритми перетворення контурів зображень в афінному просторі. Вісник Національного університету «Львівська політехніка». Комп'ютерні науки та інформаційні технології. 2009. № 638. С. 185-189.

12. Грицик В.В., Березька К.М., Березький О. М. Моделювання та синтез складних зображень симетричної структури. Львів: УАД – ДНДП, 2005. 140 с.

13. Berezsky O., Melnyk G., Batko Yu., Pitsun O. Regions matching algorithms analysis to quantify the image segmentation results. Sensors & Transducers. 2017. Vol. 208, Iss. 1. P. 44-50.

14. Березький О. М., Мельник Г. М., Березька К. М. Нечітка база знань інтелектуальної системи діагностування видів раку молочної залози. Вісник Хмельницького національного університету. Технічні науки. 2013. №6. С.284-291.

15. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models. Advances in Neural Information Processing Systems, 2020. Vol. 33. P. 6840–6851.

16. Song Y., Ermon S. Score-Based Generative Modeling through Stochastic Differential Equations. International Conference on Learning Representations, 2021. arXiv:2011.13456.
17. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-Resolution Image Synthesis with Latent Diffusion Models. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. P. 10684–10695.
18. Dhariwal P., Nichol A. Diffusion Models Beat GANs on Image Synthesis. Advances in Neural Information Processing Systems, 2021. arXiv:2105.05233.
19. Karras T., Aittala M., Aila T. Elucidating the Design Space of Diffusion-Based Generative Models. Advances in Neural Information Processing Systems, 2022. arXiv:2206.00364.
20. Pinaya W., Tudosiu P. D., et al. Brain Imaging Generation with Latent Diffusion Models. Medical Image Analysis, 2023. Vol. 90. Article 102948.
21. Tudosiu P. D., Dobson J., Kainz B., Rueckert D. Histopathology Image Generation via Diffusion Models. MICCAI Workshop on Medical Image Generation, 2023. arXiv:2303.09546.
22. Pozzi M., Noei S., Robbi E., Cima L., Moroni M., Munari E., Torresani E., Jurman G. Generating Synthetic Data in Digital Pathology Through Diffusion Models: A Multifaceted Approach to Evaluation. medRxiv preprint, 2023. DOI:10.1101/2023.11.21.23298808.
23. Montoya-del-Angel R., Sam-Millan K., Vilanova J. C., Martí R. MAM-E: Mammographic Synthetic Image Generation with Diffusion Models. Sensors, 2024. Vol. 24(7), Article 2076.
24. Kazerouni A., Aghdam E. K., Heidari M., et al. Diffusion Models in Medical Imaging: A Comprehensive Survey. Medical Image Analysis, 2023.
25. Bernt Øksendal. Stochastic differential equations. In Stochastic differential equations, pp. 65–84. Springer, 2003
26. Brian D O Anderson. Reverse-time diffusion equation models. Stochastic Process. Appl., 12(3): 313–326, May 1982.

27. Aapo Hyvarinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(Apr):695–709, 2005.
28. Yang Song, Sahaj Garg, Jiaxin Shi, and Stefano Ermon. Sliced score matching: A scalable approach to density and score estimation. In *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, July 22-25, 2019*, pp. 204, 2019a.
29. Y. Song, J. Sohl-Dickstein, D.P. Kingma, A. Kumar, S. Ermon, B. Poole, *Score-based generative modeling through stochastic differential equations*, 2020, arXiv preprint arXiv:2011.13456.
30. Tianyu Pang, Kun Xu, Chongxuan Li, Yang Song, Stefano Ermon, and Jun Zhu. Efficient learning of generative models via finite-difference score matching. arXiv preprint arXiv:2007.03317, 2020.
31. Simo Sarkk ¨ a and Arno Solin. ¨ Applied stochastic differential equations, volume 10. Cambridge University Press, 2019.
32. Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in neural information processing systems*, pp. 6571–6583, 2018
33. Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-based generative models with variational dequantization and architecture design. In *International Conference on Machine Learning*, pp. 2722–2730, 2019
34. Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. arXiv preprint arXiv:1605.08803, 2016.
35. Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, pp. 10215–10224, 2018.
36. Geoffrey Roeder, Luke Metz, and Diederik P Kingma. On linear identifiability of learned representations. arXiv preprint arXiv:2007.00810, 2020.
37. John R Dormand and Peter J Prince. A family of embedded runge-kutta formulae. *Journal of computational and applied mathematics*, 6(1):19–26, 1980.

38. L. Theis, A. van den Oord, і M. Bethge, «A note on the evaluation of generative models», *ArXiv151101844 Cs Stat*, Квіт 2016, Дата звернення: 14, Листопад 2021. [Online]. Доступний у: <http://arxiv.org/abs/1511.01844>
39. Q. Xu *et al.*, «An empirical study on evaluation metrics of generative adversarial networks», *ArXiv180607755 Cs Stat*, Сер 2018, Дата звернення: 14, Листопад 2021. [Online]. Доступний у: <http://arxiv.org/abs/1806.07755>
40. S. Barratt і R. Sharma, «A Note on the Inception Score». arXiv, 21, Червень 2018. Дата звернення: 17, Травень 2022. [Online]. Доступний у: <http://arxiv.org/abs/1801.01973>
41. A. Borji, «Pros and cons of GAN evaluation measures», *Comput. Vis. Image Underst.*, вип. 179, с. 41–65, Лют 2019, doi: 10.1016/j.cviu.2018.10.009.
42. K. Shmelkov, C. Schmid, і K. Alahari, «How good is my GAN?», arXiv, arXiv:1807.09499, Лип 2018. doi: 10.48550/arXiv.1807.09499.
43. M. Lucic, K. Kurach, M. Michalski, O. Bousquet, і S. Gelly, «Are GANs created equal? a large-scale study», в *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, в NIPS'18. Red Hook, NY, USA: Curran Associates Inc., Груд 2018, с. 698–707.
44. Peebles W., Xie S. Scalable Diffusion Models with Transformers // arXiv. – 2023. – arXiv:2212.09748. DOI: <https://doi.org/10.48550/arXiv.2212.09748>.
45. Liu Z., Luo D., Xu Y., Jaakkola T., Tegmark M. GenPhys: From Physical Processes to Generative Models // arXiv. – 2023. – arXiv:2304.02637. DOI: <https://doi.org/10.48550/arXiv.2304.02637>.
46. Gulrajani I., Hashimoto T.B. Likelihood-Based Diffusion Language Models // arXiv. – 2023. – arXiv:2305.18619. DOI: <https://doi.org/10.48550/arXiv.2305.18619>.
47. Schramowski P., Brack M., Deiseroth B. et al. Safe Latent Diffusion: Mitigating Inappropriate Content Generation in Diffusion Models. arXiv preprint, 2023. arXiv:2211.05105.

48. Kang M., Choi Y. Noise Map Inversion for Editable Image Generation: Encoding Target Images into Noise for High-Precision Image Manipulation. arXiv preprint, 2025. arXiv:2509.25776.

49. Meng C., He Y., Song J. et al. SDEdit: Image Synthesis and Editing with Stochastic Differential Equations. International Conference on Learning Representations, 2022.