

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

Богонос Дмитро Сергійович

**Дифузійні моделі генерування цитологічних
зображень / Diffusion models for generating
cytological images**

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

Виконав студент групи КІм-21
Д. С. Богонос

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

ТЕРНОПІЛЬ – 2025

АНОТАЦІЯ

Богонос Д.С. Дифузійні моделі генерування цитологічних зображень – 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 100 сторінок і містить 14 ілюстрацій, 8 таблиць, 1 додаток та 51 джерела за переліком посилань.

Метою роботи є дослідження дифузійних моделей та алгоритмів генерування біомедичних зображень і розроблення алгоритму синтезу цитологічних 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 алгоритм генерування цитологічних зображень. Здійснено програмну реалізацію дифузійних алгоритмів синтезу цитологічних 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, БІОМЕДИЧНІ ЗОБРАЖЕННЯ, ЦИТОЛОГІЧНІ ЗОБРАЖЕННЯ, ДИФУЗІЙНІ МОДЕЛІ, МЕТРИКА FID, МЕТРИКА IS.

ANNOTATION

Bogonos D.C. Diffusion models for generating cytological images. – 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 100 pages in length and includes 14 illustrations, 8 tables, 1 appendice, and 51 references.

The aim of the work is to investigate diffusion models and algorithms for generating biomedical images, and to develop an algorithm for synthesizing cytological 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 methods of object-oriented programming.

An algorithm for generating cytological images has been developed. A software implementation of diffusion algorithms for synthesizing cytological images has been carried out in the Stable Diffusion environment. The quality of the synthesized images has also been evaluated 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 that are based on other generative models.

KEYWORDS: SYNTHESIS, BIOMEDICAL IMAGES, CYTOLOGICAL IMAGES, DIFFUSION MODELS, FID METRIC, IS METRIC.

ЗМІСТ

Вступ.....	6
1 Аналіз алгоритмів синтезу біомедичних зображень.....	9
1.1 Синтез зображень на основі глибоких нейронних мереж	9
1.2 Аналіз цитологічних зображень	13
1.3 Аналіз алгоритмів синтезу зображень на основі глибоких нейронних мереж.....	18
1.4 Висновки до розділу 1	25
2 Генеративні моделі синтезу зображень	28
2.1 Основні типи генеративних моделей	28
2.2 Основні типи генеративних дифузійних моделей	30
2.3 Моделі генеративних дифузійних моделей	33
2.3.1 Моделі прямого дифузійного процесу	34
2.3.2 Моделі зворотного дифузійного процесу.....	35
2.4 Обчислення подібності зображень.....	47
2.5 Висновки до розділу 2	51
3 Синтез цитологічних зображень в середовищі stable diffusion	52
3.1 Програмне середовище Stable Diffusion.....	52
3.2 Апаратно-програмні вимоги до проведення експериментів	65
3.3 Результати комп'ютерних експериментів.....	68
3.4 Висновки до розділу 3	80
Висновки	82
Список використаних джерел	83
Додаток А Світлокопії виданих публікацій	Помилка! Закладку не визначено.

ВСТУП

Актуальність дослідження. Класифікація медичних зображень належить до ключових завдань автоматизованих систем підтримки діагностики. Найефективніші сучасні підходи ґрунтуються на використанні глибоких нейронних мереж, які потребують значних обсягів навчальних даних. Проте такі дані у сфері медицини часто є дефіцитними через етичні, правові та технічні обмеження. Одним із можливих рішень цієї проблеми є створення штучних зображень на основі невеликих доступних вибірок. Це завдання ефективно розв'язують генеративні моделі, що становлять провідний напрям розвитку штучного інтелекту. Зокрема, дифузійні моделі демонструють високу придатність до формування якісних біомедичних зображень, що робить дослідження в цій галузі надзвичайно актуальним.

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

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

- проаналізувати дослідження у глибоких нейронних мережах;
- провести аналіз цитологічних зображень;
- провести аналіз синтезу зображень на основі глибоких нейронних мереж;
- проаналізувати програмні засоби синтезу зображень;
- дослідити дифузійні моделі та алгоритми генерування зображень;
- розробити алгоритми генерування цитологічних 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 ланцюгів Маркова, 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 цитологічних зображень.

Практичне значення одержаних результатів полягає в проведенні обчислювальних експериментів в середовищі Stable Diffusion для синтезу цитологічних 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 додатку [4].

У першому розділі показано актуальність досліджень синтезу зображень на основі глибоких нейронних мереж, проаналізовано цитологічні зображення, проведено аналіз алгоритмів синтезу зображень на основі глибоких нейронних мереж.

У другому розділі досліджено дифузійні моделі та алгоритми синтезу зображень, розроблено алгоритм синтезу цитологічних зображень.

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

У додатку приведено світлокопії виданих публікацій.

1 АНАЛІЗ АЛГОРИТМІВ СИНТЕЗУ БІОМЕДИЧНИХ ЗОБРАЖЕНЬ

1.1 Синтез зображень на основі глибоких нейронних мереж

Синтез зображень на основі глибоких нейронних мереж (ГНМ) є одним із найдинамічніших напрямів сучасного штучного інтелекту (ШІ), що охоплює створення нових візуальних даних на основі наявних прикладів. Основна ідея полягає у тому, що модель вчиться відтворювати складні закономірності в наборі зображень і здатна створювати нові зразки, які статистично відповідають розподілу навчальних даних. Завдяки цьому глибокі нейронні мережі знайшли широке застосування в таких галузях, як комп'ютерна графіка, медицина, біоінформатика та цифрова патологія [15].

Глибокі нейронні мережі є багаторівневими моделями, які складаються з великої кількості взаємопов'язаних штучних нейронів. Кожен рівень мережі виконує послідовне перетворення вхідних даних, поступово переходячи від простих ознак (наприклад, контурів чи кольорових градієнтів) до складних структурних і семантичних представлень. Однією з ключових архітектур, що використовується для обробки та синтезу зображень, є згорткова нейронна мережа (Convolutional Neural Network, CNN). Саме CNN стала основою для побудови більш складних моделей, таких як автокодувальники (Autoencoders), генеративні змагальні мережі (GAN) та дифузійні моделі.

Додатково, CNN-архітектури вирізняються здатністю ефективно обробляти локальні патерни завдяки використанню згорткових фільтрів, що робить їх особливо придатними для аналізу біомедичних зображень, де важливу роль відіграють дрібні морфологічні деталі. Використання операцій підвибірki (pooling) та нормалізації дозволяє глибоким моделям масштабуватися на великі набори даних, зберігаючи стабільність навчання та покращуючи узагальнюючу здатність. У сучасних дослідженнях CNN також поєднують із трансформерними архітектурами (Vision Transformers), що забезпечує ще

глибше розуміння глобального контексту зображення та підвищує точність синтезу й класифікації.

У контексті синтезу біомедичних зображень застосування глибоких нейронних мереж є особливо актуальним. Біомедичні зображення, зокрема гістологічні та цитологічні, мають складну структуру, яка містить велику кількість дрібних морфологічних деталей, що потребують високої точності відтворення. Традиційні методи генерації даних, засновані на статистичних моделях або геометричних трансформаціях, не здатні адекватно передати варіативність форми, текстури та кольору клітинних структур, оскільки не враховують їх складну просторову організацію. Глибокі нейронні мережі, навпаки, можуть навчатися на великих обсягах реальних зображень і формувати внутрішнє представлення закономірностей тканинних структур, кольорових відтінків та текстур. Завдяки цьому моделі здатні відтворювати як глобальні особливості біологічних об'єктів, так і тонкі локальні деталі, що визначають діагностичну цінність синтетичних візуалізацій. Крім того, використання сучасних генеративних архітектур дає змогу моделювати рідкісні або атипові клітинні патерни, які важко отримати в реальних умовах.

Одним із базових підходів до синтезу зображень є автокодування, де мережа навчається стискати вхідне зображення у компактне представлення (латентний простір) і потім відновлювати його. Такий механізм дозволяє моделі виявляти ключові структурні ознаки та формувати їх узагальнене латентне описання, що надалі може бути використане для генерації нових варіацій вихідних даних. Цей підхід дав поштовх до створення варіаційних автокодувальників (Variational Autoencoders, VAE), які здатні генерувати нові зображення, вибираючи точки з латентного простору [16]. Завдяки ймовірнісній природі VAE вони добре підходять для моделювання розподілів даних та створення різноманітних варіантів зображень, що є важливим у задачах аугментації біомедичних наборів. Проте, хоча VAE забезпечують математично обґрунтований спосіб генерації, вони мають схильність до створення дещо

розмитих результатів через усереднення ймовірнісних розподілів. Це обмеження стало однією з причин появи більш потужних моделей, таких як GAN і дифузійні підходи, що здатні формувати більш чіткі та деталізовані структури.

Іншим важливим етапом розвитку стали генеративні змагальні мережі (Generative Adversarial Networks, GAN), запропоновані І. Гудфеллоу у 2014 році [17]. Архітектура GAN складається з двох частин – генератора, який створює нові зображення, і дискримінатора, який оцінює їх достовірність. Під час навчання ці дві моделі змагаються між собою, що сприяє підвищенню якості синтезованих даних. У біомедичних застосуваннях GAN використовуються для генерації зображень клітин, реконструкції медичних зрізів, збільшення навчальних вибірок (data augmentation) та покращення роздільної здатності мікроскопічних зображень.

Попри високу якість синтезу, класичні GAN-моделі мають низку системних обмежень. Однією з найпоширеніших проблем є нестабільність навчання, що зумовлена складною динамікою взаємодії генератора та дискримінатора. Іншою критичною вадою є схильність до «mode collapse», коли генератор відтворює обмежений набір зразків, втрачаючи різноманітність та не охоплюючи повний розподіл даних. Додаткові труднощі виникають у забезпеченні керованості процесу генерації, оскільки базові GAN-підходи часто не дозволяють інтерпретовано впливати на стиль, форму чи структуру результуючих зображень. Саме ці обмеження стали стимулом для появи нових, більш стабільних та керованих архітектур, зокрема StyleGAN, CycleGAN, Pix2Pix та низки їхніх модифікацій. Ці моделі використовують покращені втратні функції, різновиди нормалізації, механізми стилізації простору ознак і спеціалізовані архітектурні компоненти, що підвищують стабільність навчання та забезпечують контроль над властивостями синтезованих зображень.

Значним етапом у розвитку синтезу зображень стало впровадження дифузійних моделей – підходу, який ґрунтується на використанні стохастичних

процесів поступового додавання шуму до зображення з подальшим навчанням моделі виконувати обернений процес його відновлення. На відміну від GAN, де процес генерації базується на змаганні між двома мережами, дифузійні моделі спираються на математично формалізовані й стабільні процедури апроксимації зворотного дифузійного процесу, описані системами стохастичних диференціальних рівнянь. Таке поєднання теоретичної строгості з практичною ефективністю дає змогу відтворювати складні структури з високим рівнем деталізації, включаючи клітинні та тканинні морфологічні елементи, які є критично важливими для біомедичної візуалізації. На додачу, дифузійні моделі забезпечують передбачувану поведінку, високу стабільність навчання та пропонують широкий спектр механізмів керування — від стилізації до зміни семантичного змісту зображення. Саме тому вони розглядаються як наступний еволюційний крок після GAN, пропонуючи більш надійний і контрольований інструмент для генерації синтетичних даних у медичних дослідженнях.

У медичних дослідженнях синтез зображень на основі глибоких нейронних мереж виконує кілька ключових функцій:

- розширення навчальних вибірок – створення штучних прикладів для тренування моделей класифікації або сегментації, що особливо важливо при нестачі реальних даних;
- покращення якості зображень – відновлення дрібних деталей, зменшення шуму, підвищення контрастності та роздільної здатності;
- моделювання патологічних процесів – створення візуалізацій клітин з різними стадіями захворювань для підтримки навчання медичних фахівців;
- симуляція нових типів клітин – використання моделей для створення зображень, що імітують рідкісні патології, які важко отримати в реальних умовах.

Таким чином, глибокі нейронні мережі є фундаментом для сучасних систем синтезу зображень, забезпечуючи високу гнучкість, якість та універсальність у відтворенні складних структур. Їхня здатність моделювати

нелінійні залежності, виявляти приховані ознаки та адаптуватися до різноманітних типів даних робить їх ключовим інструментом у розвитку генеративних технологій. Особливої актуальності це набуває у сфері біомедичної візуалізації, де точність, деталізація та морфологічна коректність є критично важливими параметрами. Подальший розвиток дифузійних моделей і вдосконалення архітектур глибокого навчання відкриває нові можливості для генерації цитологічних зображень, формування великих синтетичних датасетів, моделювання рідкісних патологій та підтримки автоматизованих діагностичних систем. Саме ці напрями становлять концептуальну основу цього дослідження.

1.2 Аналіз цитологічних зображень

Цитологічні зображення є одним із ключових джерел інформації у сучасній медичній діагностиці, оскільки вони відображають морфологічні та структурні особливості клітин, що дає змогу виявляти патологічні зміни на ранніх етапах розвитку захворювань. Цитологічний аналіз ґрунтується на детальному дослідженні індивідуальних клітин та їхніх компонентів – ядра, цитоплазми, клітинної мембрани й мікрооточення, кожен із яких має діагностично значущі характеристики. Висока чутливість цього методу зумовлена тим, що навіть мінімальні зміни у формі, розмірі, текстурі ядра чи співвідношенні ядерно-цитоплазматичних параметрів можуть свідчити про наявність атипії або ранніх стадій неоплазії.

Завдяки розвитку цифрової мікроскопії, що забезпечує високу роздільну здатність, стабільність кольорового відтворення та стандартизацію параметрів знімання, цитологічні зображення стали значно більш придатними для комп'ютерного аналізу. Паралельно активне впровадження методів машинного та глибокого навчання сприяло створенню систем, здатних автоматично

аналізувати клітинні препарати з рівнем точності, порівняним із фаховою оцінкою цитолога. Глибокі нейронні мережі дозволяють не лише автоматизувати процес виявлення та класифікації клітинних структур, але й виконувати їхній синтез, реконструкцію, сегментацію та нормалізацію кольору з високим ступенем достовірності.

Застосування таких моделей зменшує суб'єктивність експертного оцінювання, прискорює аналіз великих масивів даних та підвищує відтворюваність діагностичних результатів. Крім того, алгоритми штучного інтелекту створюють нові можливості для виявлення патологічних змін на ранніх стадіях, коли візуальні ознаки можуть бути малопомітними для людського ока. Це формує підґрунтя для розробки інтелектуальних систем підтримки клінічних рішень, які здатні інтегрувати результати автоматизованого аналізу у загальний діагностичний процес та підсилювати ефективність медичної практики.

На відміну від гістологічних зображень, які демонструють структуру тканин у контексті їхнього взаємного розташування, цитологічні зображення зосереджені на окремих клітинах або невеликих їхніх групах. Це робить їх надзвичайно інформативними для діагностики новоутворень, запальних процесів та атипових клітинних станів. Найпоширенішими прикладами є цитологічні дослідження мазків Папаніколау (Pap Smear) для скринінгу раку шийки матки, аналіз клітин крові, бронхіальних або сечових цитологічних препаратів. Для створення цифрових цитологічних зображень застосовують високороздільні оптичні мікроскопи, що дозволяють фіксувати деталі клітинних структур на рівні 0,1–0,5 мікрметра. Ці зображення зазвичай зберігаються у форматах TIFF, PNG або JPEG з роздільною здатністю понад 2000×2000 пікселів, що забезпечує можливість подальшого аналізу за допомогою алгоритмів комп'ютерного зору. Проте навіть за високої якості зйомки цитологічні зображення мають ряд проблем: варіації фарбування,

відмінності у фокусі, неоднорідність освітлення та наявність артефактів, що ускладнює їхню автоматичну обробку.

Основними завданнями аналізу цитологічних зображень є:

- сегментація клітин і ядер – виокремлення контурів клітинних структур для подальшого морфометричного аналізу;
- класифікація клітин – визначення належності до певного типу або класу патології (нормальна, запальна, атипова, злоякісна);
- виділення ознак – обчислення кількісних характеристик форми, текстури, кольору та щільності клітинних компонентів;
- синтез зображень – створення нових зразків клітин для розширення навчальних наборів даних та моделювання рідкісних патологій.

Сучасні методи обробки цитологічних зображень активно використовують згорткові нейронні мережі (CNN), U-Net, ResNet, а також трансформерні архітектури (Vision Transformer, Swin Transformer) для точного розпізнавання клітинних структур. Такі моделі здатні автоматично виявляти межі клітин, визначати форму та щільність ядер, виявляти ознаки малігнізації та аналізувати текстурні характеристики хроматину. Завдяки багаторівневому вилученню ознак ці архітектури забезпечують високу точність сегментації та класифікації навіть у випадках, коли зображення містять шум, артефакти або значні варіації фарбування. Крім того, трансформерні моделі демонструють здатність ефективно обробляти довгострокові просторові залежності, що є важливим під час аналізу комплексних морфологічних структур, характерних для цитологічних препаратів. Це робить їх перспективним інструментом у побудові автоматизованих систем діагностики та у створенні високоточних алгоритмів підтримки клінічних рішень.

Проте для їхнього ефективного навчання необхідні великі обсяги анотованих даних, що є значною проблемою у цитологічних дослідженнях, де отримання реальних даних обмежене етичними та технічними чинниками. Одним із шляхів подолання цієї проблеми є генерація штучних цитологічних

зображень за допомогою моделей глибокого навчання. Використання таких підходів дозволяє збалансувати набори даних, зменшити проблему дисбалансу класів та підвищити узагальнювальну здатність моделей класифікації.

Генеративні нейронні мережі, зокрема GAN та дифузійні моделі, здатні створювати реалістичні зображення клітин, які практично не відрізняються від справжніх мікроскопічних знімків. Завдяки здатності моделювати морфологічні особливості ядра, цитоплазми та міжклітинних структур, такі моделі формують синтетичні зразки, що зберігають ключові діагностичні ознаки. Це має особливо велике значення для задач ранньої діагностики онкологічних захворювань, коли кількість прикладів патологічних клітин у реальних наборах даних є недостатньою або сильно дисбалансованою.

Використання синтетичних зображень дає змогу збільшити обсяг і різноманітність тренувальних вибірок, зменшити ризик перенавчання моделей класифікації та підвищити їхню здатність узагальнювати дані на рідкісні або нетипові патологічні стани. Таким чином, генеративні моделі стають важливим інструментом у підтримці алгоритмічного виявлення онкологічних аномалій та у розробці більш надійних систем автоматизованої цитологічної діагностики.

Існує кілька відкритих наборів цитологічних даних, що активно використовуються в наукових дослідженнях:

- Herlev Dataset – містить 917 зображень клітин шийки матки, класифікованих за шістьма типами;
- SIPaKMeD Dataset – понад 4 000 зображень клітин, розподілених на п'ять класів (суперфіціальні, парабазальні, метапластичні, інтермедіальні та дискаріотичні);
- Pap Smear Dataset (Ayer's Dataset) – зразки цитологічних мазків із високою роздільною здатністю, що використовуються для автоматизованого скринінгу;
- Motic Cytology Dataset – колекція клітин крові та мазків, що застосовується у задачах сегментації ядер;

- Cytomorphology Dataset – включає зображення різних типів клітин із маркованими ядрами та цитоплазмою для задач сегментації.

Візуальний аналіз цитологічних зображень демонструє, що ядра клітин мають виражену варіативність за формою, розміром і текстурою, що зумовлено як фізіологічними особливостями, так і патологічними змінами. У нормальних клітинах ядро зазвичай характеризується рівними, чітко окресленими контурами, однорідним забарвленням та помірною щільністю хроматину, яка рівномірно розподілена всередині ядерного простору. Такі ознаки вказують на стабільний метаболічний стан клітини та відсутність ознак патології.

У патологічних клітинах, навпаки, спостерігається виражена морфологічна атипія. Частими проявами є гіперхроматоз – надмірне інтенсивне забарвлення ядра внаслідок ущільнення хроматину, а також його неоднорідність, що проявляється у вигляді зернистих або нерівномірно розташованих ділянок. Типовим діагностичним маркером є збільшення ядерно-цитоплазматичного співвідношення, коли ядро займає більшу частину клітини. Крім того, патологічні клітини часто демонструють неправильну, асиметричну форму ядер, наявність виїмок чи інвагінацій ядерної мембрани, а також багатоядерність, яка може свідчити про порушення процесів поділу або патологічну проліферацію.

Такі морфологічні відмінності відіграють ключову роль у комп'ютерному аналізі цитологічних зображень, оскільки саме за цими ознаками алгоритми машинного навчання ідентифікують атипію, класифікують клітини та формують підґрунтя для автоматизованої діагностики.

Ці характеристики є важливими ознаками для автоматичних алгоритмів розпізнавання. Проблемою залишається варіативність фарбування, яка виникає через різницю у концентрації барвників, умовах підготовки мазка та параметрах освітлення. Для подолання цієї проблеми застосовується кольорова нормалізація – метод, який приводить колірні характеристики зображень до єдиного еталонного стандарту. Також використовуються техніки аугментації

даних – поворот, масштабування, зміна контрасту та яскравості, які дозволяють збільшити ефективність навчання моделей без створення нових реальних зразків.

Цитологічні зображення мають значний потенціал для використання у генеративних моделях завдяки своїй високій структурній повторюваності, чіткій морфологічній організованості та відносно обмеженій кількості типів клітин. Це робить їх особливо придатними для синтезу за допомогою дифузійних моделей, які здатні точно відтворювати дрібні морфологічні деталі та текстурні особливості клітинних структур. Такі моделі не лише генерують нові клітинні зразки, але й забезпечують можливість гнучкого контролю параметрів генерації – зокрема, форми ядра, ступеня хроматинової щільності, рівня забарвлення, співвідношення ядра й цитоплазми чи навіть конкретного типу клітинного класу. Завдяки цим характеристикам дифузійні моделі відкривають перспективу створення масштабованих віртуальних цитологічних колекцій, які можуть використовуватися для підготовки фахівців, тестування та валідації алгоритмів автоматизованої діагностики, а також проведення наукових досліджень у галузі цифрової патології та цитоморфології.

1.3 Аналіз алгоритмів синтезу зображень на основі глибоких нейронних мереж

Алгоритми синтезу зображень на основі глибоких нейронних мереж відіграють ключову роль у сучасних дослідженнях у галузі штучного інтелекту, комп'ютерного зору та біомедичної візуалізації. Їхня основна мета – створення нових, реалістичних зображень, які максимально точно відтворюють структуру, текстуру та кольорову гаму реальних даних, зберігаючи при цьому статистичні властивості вихідних вибірок. Такі алгоритми здатні моделювати складні

морфологічні взаємозв'язки, відтворювати дрібні деталі та варіативні патерни, характерні для природних або медичних зображень.

Синтетичні зображення, отримані за допомогою генеративних моделей, відіграють важливу роль у подоланні обмежень традиційних підходів до збору даних. Вони дозволяють не лише доповнювати існуючі набори даних та збалансовувати класи, але й моделювати процеси та патологічні стани, які у реальних умовах спостерігаються рідко або є складними для отримання через етичні, технічні чи клінічні обмеження. Крім того, генеративні методи забезпечують можливість створення контрольованих експериментальних сценаріїв, де можна досліджувати вплив окремих морфологічних або структурних змін на поведінку алгоритмів аналізу та класифікації.

У контексті біомедичних застосувань алгоритми генерації зображень використовуються для вирішення широкого кола задач, включно зі симуляцією патологічних станів, збільшенням обсягу тренувальних даних, підвищенням якості медичних зображень та формуванням додаткових візуальних прикладів для навчання медичних фахівців. Такі підходи дозволяють моделювати сценарії, які рідко зустрічаються у клінічній практиці, але є критично важливими для діагностики. У цитології генеративні моделі застосовуються для створення зображень клітин різних типів, імітації атипових морфологічних змін, реконструкції втрачених або пошкоджених структур, а також для балансування наборів даних шляхом генерування зразків рідкісних класів патологій. Це суттєво покращує якість навчання алгоритмів класифікації та сегментації та забезпечує їхню вищу стійкість до варіативності реальних даних.

Серед великої кількості підходів, які застосовуються для синтезу зображень, виділяють три основні категорії: автокодувальні моделі (Autoencoders, AE), генеративні змагальні мережі (Generative Adversarial Networks, GAN) та дифузійні моделі (Diffusion Models, DM). Кожна з цих груп має власні архітектурні принципи, математичні передумови та алгоритмічні

механізми, що визначають якість, стабільність та контрольованість синтезованих зображень.

Автокодувальники забезпечують компактне латентне представлення даних і є фундаментом для багатьох генеративних моделей, зокрема варіаційних автокодувальників (VAE), які запроваджують імовірнісний підхід до генерації. GAN-моделі, у свою чергу, використовують змагальну взаємодію генератора і дискримінатора, що дозволяє формувати високореалістичні зображення, хоча й супроводжується ризиками нестабільного навчання. Дифузійні моделі, як найновіший напрям генеративного моделювання, ґрунтуються на стохастичних процесах поступового додавання та видалення шуму, забезпечуючи високу стабільність і точність синтезу, що робить їх надзвичайно перспективними для біомедичних застосувань.

Завдяки цим відмінностям кожен підхід по-різному справляється зі складністю цитологічних та гістологічних структур, що дає змогу обирати оптимальну модель залежно від дослідницьких або клінічних потреб.

Автокодувальні підходи

Автокодувальні моделі є одним із базових типів генеративних архітектур і широко застосовуються у задачах відновлення, реконструкції та синтезу зображень. Класичний автокодувальник складається з двох основних компонентів: енкодера, який перетворює вхідне зображення у компактне латентне представлення, та декодера, який відновлює його з цього представлення. Така структура дає змогу моделі навчитися стислому та інформативному опису даних, що є фундаментом для створення нових синтетичних прикладів.

Подальшим розвитком цієї концепції стали варіаційні автокодувальники (Variational Autoencoders, VAE), які вводять стохастичний компонент у латентний простір. На відміну від класичних AE, VAE не просто кодують зображення, а моделюють розподіл латентних змінних, з якого можна вибирати нові точки для генерації зображень. Це забезпечує контрольованість процесу та

здатність отримувати плавні переходи між різними зразками. Саме тому VAE широко використовуються в біомедичній візуалізації для моделювання клітинних морфотипів, синтезу варіацій реальних структур і реконструкції фрагментів зображень, що містять пошкоджені або частково втрачені дані.

VAE широко застосовуються для моделювання клітинних структур, відновлення пошкоджених ділянок цитологічних зображень і синтезу зразків із заданими характеристиками. Проте головним недоліком таких моделей є схильність до розмитості результатів – через усереднення варіацій латентних змінних.

Генеративні змагальні мережі (GAN)

Генеративні змагальні мережі, запропоновані І. Гудфеллоу у 2014 році, стали переломним етапом у розвитку генеративного штучного інтелекту та сформували основу для значної частини сучасних методів синтезу зображень. Їхня архітектура складається з двох взаємодіючих компонентів: генератора, який створює синтетичні зображення, та дискримінатора, який оцінює їхню достовірність, порівнюючи із реальними зразками. У процесі навчання ці дві мережі ведуть динамічне змагання: генератор прагне "обдурити" дискримінатор, створюючи все більш реалістичні приклади, тоді як дискримінатор постійно вдосконалюється, щоб точніше відокремлювати справжні дані від штучно згенерованих. Така змагальна взаємодія формує механізм adversarial learning, який забезпечує поступове й стабільне підвищення якості синтезованих зображень.

Завдяки своїй здатності моделювати складні розподіли даних GAN-моделі стали основою для численних модифікацій та покращень, орієнтованих на конкретні задачі у біомедичній візуалізації, комп'ютерній графіці та аналітиці зображень. Їхня архітектура дозволяє навчатися на високорозмірних даних та відтворювати тонкі морфологічні особливості, що є критично важливим для цитологічних і гістологічних досліджень. Результати, отримані за допомогою GAN, відзначаються високим рівнем деталізації, природністю

текстур та наближеністю до реальних зразків, що робить цей підхід одним із ключових інструментів у сучасних генеративних технологіях. Багато сучасних систем діагностики й аналізу даних інтегрують GAN-моделі як допоміжний елемент для балансування наборів даних, реконструкції артефактів або підвищення роздільної здатності зображень, що додатково підкреслює їхню значущість у прикладних дослідженнях.

GAN стали основою для багатьох відомих архітектур, серед яких:

- DCGAN (Deep Convolutional GAN) -використовує згорткові шари для підвищення стабільності генерації [18];
- Pix2Pix – орієнтований на перетворення зображення в інше (наприклад, ескіз → фото або цитологічне зображення → сегментована карта) [19];
- CycleGAN – дозволяє здійснювати генерацію між двома доменами без необхідності парних даних [20];
- StyleGAN – забезпечує контроль над візуальними параметрами зображення, такими як форма, текстура або колір [21].

У біомедичній сфері GAN-моделі використовуються для синтезу зображень клітинних культур, крові, тканин або мазків Папаніколау. Наприклад, MedGAN і Cytogan застосовуються для генерації цитологічних зразків, що відтворюють характерні властивості реальних клітин, допомагаючи у навчанні алгоритмів класифікації та сегментації. Основними проблемами GAN-підходів є нестабільність навчання, необхідність ретельного підбору гіперпараметрів та ризик mode collapse, коли генератор навчається створювати лише обмежений набір шаблонних прикладів. Для подолання цих труднощів розробляються вдосконалені варіанти – WGAN, LSGAN, Progressive GAN, що забезпечують більш стабільну збіжність і різноманітність результатів.

Дифузійні моделі

Дифузійні моделі є новітнім підходом у синтезі зображень і ґрунтуються на принципах стохастичних процесів [22]. Основна ідея полягає у поетапному

додаванні випадкового шуму до навчальних зображень (так званий прямий або forward diffusion process) та подальшому навчанні моделі виконувати зворотну операцію – поступове відновлення структури зображення шляхом видалення шуму [23]. У процесі навчання модель опановує здатність реконструювати дані з будь-якого проміжного стану, що фактично формує детермінований або стохастичний зворотний дифузійний процес (reverse diffusion process). Завдяки цьому дифузійні моделі здатні перетворювати чистий шум у реалістичні та високодеталізовані зображення, точно відтворюючи текстури, форми та морфологічні особливості об'єктів.

Математично цей процес описується стохастичними диференціальними рівняннями (Stochastic Differential Equations, SDE), які моделюють еволюцію стану зображення у часі та дозволяють формалізувати як прямий, так і зворотний процес дифузії в межах єдиної теоретичної моделі. Таке математичне підґрунтя забезпечує строгість формалізації та стабільність навчання, оскільки SDE описують неперервний шлях від рівня повного шуму до відновленого структурованого зображення.

До найвідоміших реалізацій дифузійних моделей належать DDPM (Denoising Diffusion Probabilistic Model), DDIM (Denoising Diffusion Implicit Model), Latent Diffusion Model (LDM) та Stable Diffusion, що поєднують високу ефективність із здатністю формувати детальні та реалістичні зображення [24]. Крім того, LDM-моделі, які працюють у латентному просторі, дозволяють суттєво зменшити обчислювальні витрати та пришвидшити процес генерації без втрати якості, що робить їх особливо придатними для роботи з великими біомедичними вибірками. Stable Diffusion, як одна з найуспішніших практичних реалізацій LDM, забезпечує широкий спектр засобів керування генерацією, включно з текстовими підказками, стилізацією та можливістю точкового відновлення фрагментів зображення (inpainting), що значно розширює її застосовність у медичній візуалізації та цитології.

Переваги дифузійних моделей у порівнянні з попередніми підходами:

- Стабільність навчання – моделі не потребують змагання між двома мережами, як у GAN, тому їх легше оптимізувати;
- Висока деталізація – генерація зображень із чіткими структурами, реалістичними текстурами та природним кольором;
- Контрольованість процесу – можливість впливати на параметри генерації, наприклад стиль, форму клітини або ступінь патології;
- Відповідність біомедичним вимогам – збереження морфологічних особливостей клітинних структур, що є критично важливим для діагностики.

У задачах цитологічного аналізу дифузійні моделі демонструють потенціал у створенні віртуальних цитологічних препаратів, які відтворюють клітини різних типів і ступенів атипії. Це дозволяє створювати великі збалансовані набори даних, необхідні для навчання систем штучного інтелекту. Крім того, моделі такого типу можуть застосовуватися для покращення зображень, реставрації пошкоджених препаратів та імітації забарвлення, що підвищує їхню придатність у клінічних і дослідницьких цілях.

Таблиця 3.1 демонструє порівняльну характеристику підходів, які застосовуються для синтезу зображень.

Таким чином, проведений аналіз сучасних алгоритмів синтезу зображень засвідчує, що дифузійні моделі мають найвищий потенціал для розв’язання задач генерації цитологічних даних. Вони забезпечують оптимальний баланс між точністю відтворення морфологічних особливостей, візуальною реалістичністю та можливістю гнучкого керування параметрами генерації. Саме ці властивості роблять дифузійні моделі ефективним і надійним інструментом у дослідженнях генеративного інтелекту, орієнтованих на біомедичну візуалізацію. У наступних розділах буде детально розглянуто внутрішні механізми роботи таких моделей, наведено результати експериментів та проаналізовано їхнє практичне застосування у задачах синтезу цитологічних зображень.

Таблиця 1.1 – Порівняльна характеристика підходів

Алгоритм	Основний принцип	Переваги	Недоліки	Типові застосування
VAE	Імовірнісне відтворення латентного простору	Простота реалізації, швидке навчання	Розмитість результатів	Аугментація, реконструкція зображень
GAN	Змагання між генератором і дискримінатором	Висока реалістичність	Нестабільне навчання, mode collapse	Синтез клітин, підвищення роздільності
Diffusion Models	Стохастичне додавання / видалення шуму	Висока стабільність, контрольованість, деталізація	Високі обчислювальні витрати	Генерація біомедичних та цитологічних зображень

1.4 Висновки до розділу 1

У першому розділі проведено аналіз сучасних методів та алгоритмів синтезу біомедичних зображень із використанням глибоких нейронних мереж. Встановлено, що генеративні підходи відіграють ключову роль у формуванні нових зразків даних, які за своїми характеристиками наближаються до реальних. Застосування таких моделей у біомедичних дослідженнях дає змогу значно розширити наявні набори даних, покращити якість візуалізації та підвищити точність автоматизованої діагностики.

У підрозділі 1.1 розглянуто загальні принципи синтезу зображень на основі глибоких нейронних мереж. Показано, що архітектури типу CNN, Autoencoder, GAN та Diffusion Models забезпечують здатність моделювати складні закономірності у візуальних даних. Зокрема, генеративні моделі дозволяють створювати нові зображення, покращувати роздільну здатність та відновлювати пошкоджені ділянки.

У підрозділі 1.2 проаналізовано особливості цитологічних зображень, які характеризуються складною морфологічною структурою клітин та високою варіативністю кольору, форми й текстури. Наведено основні типи цитологічних наборів даних і вказано, що для ефективного навчання моделей штучного інтелекту необхідна достатня кількість збалансованих прикладів. Саме тому актуальним є використання методів генерації зображень, які здатні відтворювати клітинні структури з високою точністю та біологічною достовірністю.

У підрозділі 1.3 виконано порівняльний аналіз основних алгоритмів синтезу зображень на основі глибоких нейронних мереж, включаючи варіаційні автокодувальники, генеративні змагальні мережі та дифузійні моделі. Встановлено, що саме дифузійні моделі демонструють найвищу якість та стабільність синтезу, забезпечуючи можливість точного відтворення морфологічних деталей і контрольованості процесу генерації.

Отже, результати проведеного аналізу підтверджують доцільність використання дифузійних моделей для задач генерації цитологічних зображень. Завдяки своїй здатності формувати високоякісні та морфологічно коректні зразки, ці моделі стають ефективним інструментом для доповнення обмежених наборів даних, що традиційно є однією з ключових проблем у медичній візуалізації. Дифузійні підходи не лише забезпечують точне відтворення клітинних структур, але й дають змогу контролювати процес генерації, моделюючи різні типи клітин, рівні атипії та навіть складні патологічні сценарії. Це створює передумови для формування віртуальних колекцій цитологічних

препаратів, які можуть використовуватися у навчанні, тестуванні алгоритмів машинного навчання та розробці систем підтримки клінічних рішень.

Крім того, застосування дифузійних моделей сприяє підвищенню ефективності алгоритмів класифікації та сегментації завдяки можливості збалансувати дані, покращити якість вхідних зображень та забезпечити більшу варіативність структурних ознак. У перспективі це може значно покращити точність автоматизованих методів діагностики, зменшити вплив людського фактору та прискорити процес аналізу медичних зображень.

2 ГЕНЕРАТИВНІ МОДЕЛІ СИНТЕЗУ ЗОБРАЖЕНЬ

2.1 Основні типи генеративних моделей

Генеративні моделі навчаються на розподілі даних навчальної вибірки за на основі навчання без учителя, а потім генерують нові, невідомі зразки, що відповідають цьому розподілу. На даний час відомо п'ять основних типів генеративних моделей:

1. Генеративно-змагальні мережі (Generative Adversarial Network, GAN).
2. Варіаційні автокодувальники (Variational Autoencoder, VAE).
3. Моделі на основі потоків (Flow-based models).
4. Авторегресійні моделі (Autoregressive models).
5. Моделі на основі енергії (Energy-based models).

Розглянемо детальніше генеративні дифузійні моделі.

Генеративно-змагальні мережі (GAN) [25] – це неявні моделі густини розподілу даних, які складаються з двох окремих нейронних мереж:

- дискримінатор (discriminator) – навчається розрізняти справжні дані від згенерованих;
- генератор (generator) – створює нові зразки, імітуючи реальний розподіл даних.

Ці дві мережі беруть участь у мінімакській грі, у якій змагальне навчання мінімізує різницю між штучним та реальним розподілами даних.

Однак такий підхід має такі недоліки:

- висока чутливість до вибору гіперпараметрів. Якщо між генератором і дискримінатором виникає дисбаланс, параметри моделі можуть стати нестабільними, не збігатися або перенавчатися;

– колапс мод. Це поширене явище, коли генератор починає створювати зразки з обмеженим різноманіттям, тобто моделі бракує різних варіацій результатів.

Варіаційні автокодувальники. Під час роботи з великими наборами даних, коли ознаки численні та неперервні, обчислення маргінальної правдоподібності стає обчислювально складним. Варіаційні автокодувальники [26, 27] – це наближені моделі густини розподілу даних, які навчаються відтворювати розподіл даних, використовуючи метод варіаційного висновку.

Основна ідея варіаційного підходу полягає в тому, щоб моделювати реальний розподіл за допомогою простішого розподілу, який легше обчислювати та оптимізувати наприклад, гаусовий розподіл, який простіше обчислювати, і мінімізувати різницю між цими двома розподілами за допомогою дивергенції Кульбака–Лейблера – метрики, що вимірює відмінність між розподілами. Ця оптимізація називається оптимізацією варіаційного нижнього обмеження.

Мережа-енкодер навчається кодувати реальні дані в параметри середнього та дисперсії. Потім дані кодуються з використанням репараметризації [26], після чого ці параметри передаються декодеру для побудови штучного розподілу даних, з якого вибірка нових даних здійснюється шляхом семплінгу.

Моделі на основі потоків. Генеративні моделі, що працюють із прихованими змінними, зазвичай використовують гаусовий розподіл для обчислювальної ефективності. Однак у реальних даних більшість розподілів набагато складніші, ніж гаусові.

Моделі на основі потоків [28] та нормалізації потоків [29, 30] перетворюють простий розподіл у складний за допомогою послідовного застосування бієктивних перетворень. Кожне таке перетворення нормалізує розподіл, гарантуючи, що після трансформації зберігається нормована густина розподілу.

Після трансформації можна обчислити точну логарифмічну правдоподібність, що дозволяє напряду оптимізувати модель. Однією з головних проблем таких моделей є висока розмірність простору прихованих змінних, що ускладнює інтерпретацію та оптимізацію. Крім того, для реалізації умовних генеративних задач ці моделі виявляються складними та часто генерують зразки нижчої якості, ніж GAN або VAE. Дослідження авторів. [31] також показало прихований зв'язок між нормалізуючими потоками та моделями на основі функції градієнта густини.

Авторегресійні моделі [32, 33] – це трактовані моделі густини, які працюють послідовно, передбачаючи наступний елемент даних на основі попередніх. Вони досягли великих успіхів у таких сферах, як моделювання природної мови та прогнозування часових рядів, але мають обмеження: вони обчислюють лише умовну ймовірність наступного стану, тому виникають труднощі, коли передбачення наступного стану стає складним [34].

Моделі на основі енергії [35, 36]. Такі моделі визначають ненормовану негативну логарифмічну правдоподібність, відому як енергетична функція. Будь-яка нелінійна регресійна функція може бути вибрана як енергетична, і розподіл ймовірності нормалізується шляхом ділення енергії на її обсяг.

Ці моделі мають велику гнучкість у моделюванні, але важко обчислити та синтезувати точні ймовірності зразків, оскільки нормалізуючий член зазвичай включає інтеграл з нескінченими межами, що ускладнює оптимізацію.

2.2 Основні типи генеративних дифузійних моделей

Генеративні дифузійні моделі [37, 38] – це новий клас генеративних моделей, які усувають недоліки попередніх підходів, саме:

- не потребують змагального навчання (як у GAN);

- не покладаються на послідовне передбачення (як у autoregressive models);
- не стикаються з проблемами нормалізації об'єму (як у normalizing flows);
- і не мають високої обчислювальної складності (як у EBM).

У дифузійних моделях існують два процеси:

1. Прямий процес (forward process) – додає шум до даних на кількох кроках, поступово перетворюючи дані на випадковий шум.
2. Зворотний процес (reverse process) – поетапно видаляє шум, відновлюючи зображення або дані, подібні на початкові.

Таким чином, зворотний дифузійний процес дозволяє отримати високоякісні синтетичні дані, відтворюючи розподіл оригінальних прикладів крок за кроком.

Наприклад, гаусовий розподіл, який легше оцінювати, і мінімізувати різницю між цими двома розподілами за допомогою KL-дивергенції (метрики, що вимірює різницю між розподілами). Ця оптимізація називається оптимізацією нижньої межі доказів. Мережа-енкодер бере реальні дані, щоб отримати значення середнього та дисперсії. Потім дані кодується за допомогою репараметризації [26], використовуючи ці значення середнього та дисперсії. Після цього закодовані дані передаються до мережі-декодера, щоб отримати штучний розподіл даних, і нові дані створюються шляхом вибірки з цього розподілу.

Генеративні моделі з прихованими змінними часто використовують гаусовий розподіл для обчислювальної ефективності. Але в реальному світі більшість розподілів набагато складніші, ніж гаусові. Моделі на основі потоків [28] і нормалізуючі потоки [29, 30] перетворюють простий розподіл у складний за допомогою набору біективних функцій перетворення на кожному кроці потоку. Нормалізація означає, що після застосування перетворення отримується нормалізована густини розподілу. Точні логарифмічні ймовірності вхідних

даних можна обчислити, і оптимізація виконується безпосередньо над негативною логарифмічною правдоподібністю. Однією з проблем моделей нормалізуючих потоків є обсяг даних. Прихований простір стає дуже багатовимірним під час перетворень, і це ускладнює інтерпретацію. Крім того, дуже складно реалізувати завдання умовної генерації з цими моделями, і вони створюють зразки низької якості порівняно з GAN і VAE. Автори статті [31] показують прихований зв'язок між нормалізуючими потоками та генеративними моделями на основі оцінки.

Авторегресійні моделі [32, 33] – це трактовані моделі густини розподілу, які працюють послідовно та передбачають наступні дані на основі попередніх. Хоча вони дуже успішні в таких галузях, як моделювання мови та прогнозування часових рядів, вони обчислюють лише ймовірність наступного стану за поліноміальний час і мають труднощі, коли ймовірність наступного стану складно обчислити [34].

Іншим сімейством генеративних моделей є моделі на основі енергії (EBM) [35, 36]. Моделі на основі енергії знаходять ненормовану негативну логарифмічну ймовірність, яку називають енергетичною функцією. Будь-яку нелінійну регресійну функцію можна вибрати як енергетичну функцію. Розподіл ймовірностей нормалізується шляхом ділення енергетичної функції на її обсяг. Гнучкість EBM забезпечує важливі переваги в моделюванні, але часто важко обчислити та синтезувати точну ймовірність зразків із складних моделей, оскільки параметри нормалізуючого члена зазвичай містять складний інтеграл.

Генеративні дифузійні моделі [37, 38] – це нове сімейство генеративних моделей, яке усуває змагальне навчання в GAN, не потребує послідовного навчання в авторегресійних моделях, приблизні обчислення правдоподібності у VAE, зростання обсягу в моделях нормалізуючих потоків і складність семплінгу як в EBM. У цих моделях є два процеси: по-перше, прямий процес додає шум на кількох масштабах до розподілу даних, поступово перетворюючи його на

випадковий шум. По-друге, зворотний процес навчається відтворювати щось подібне до вихідних даних, поетапно інвертуючи дифузійний процес.

2.3 Моделі генеративних дифузійних моделей

Перша модель – імовірнісні моделі дифузії з усуненням шуму (DDPM) [39], яке базується на теорії нерівноваженої термодинаміки. DDPM використовують приховані змінні для оцінювання розподілу ймовірностей.

Друга модель – мережі умовної оцінки шуму (NCSN) [40]. У цій моделі є спільна нейронна мережа, яка навчається за допомогою методів узгодження градієнтів для оцінювання функції градієнта логарифмічної щільності розподілу даних на різних рівнях шуму.

Третя модель – модель, яка базується на стохастичних диференціальних рівняннях (Score SDE) [41], у якому дифузійний процес розв’язується за допомогою стохастичних диференціальних рівнянь у прямому та зворотному напрямках. Ця робота є особливо важливою, оскільки вона показує загальну теоретичну основу, що об’єднує DDPM та NCSN.

Модель імовірнісної дифузії з шумоподавленням (Denoising Diffusion Probabilistic Model, DDPM) [39] – це параметризований ланцюг Маркова, який навчається за допомогою варіаційного висновку для генерації зразків, подібних до реальних даних, після скінченної кількості кроків.

У ймовірнісних дифузійних моделях існує двоетапний процес:

1. Прямий дифузійний процес (forward diffusion process).
2. Зворотний процес (reverse diffusion).

2.3.1 Моделі прямого дифузійного процесу

У прямому дифузійному процесі до даних багаторазово додається гаусовий шум, доки вони повністю не перетворяться на випадковий шум. Натомість у зворотному процесі модель навчається відновлювати дані, оцінюючи умовні щільності ймовірності за допомогою нейронної мережі.

Нехай маємо точку даних із реального розподілу $x_0 \sim q(x)$. Рівняння (2.1) описує прямий дифузійний процес, який створює послідовність зашумлених зразків x_1, x_2, \dots, x_T , додаючи гаусовий шум крок за кроком упродовж T етапів.

Оцінка розподілу в момент часу t залежить лише від розподілу на попередньому кроці $t-1$. Із збільшенням t вихідний зразок x_0 поступово втрачає свої розпізнавальні ознаки. Коли $T \rightarrow \infty$ кінцевий зразок x_T стає ізотропним гаусовим розподілом.

$$q(x_t | x_{t-1}) = N(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I) \quad (2.1)$$

Увесь процес обчислюється згідно з рівнянням (2.2):

$$q(x_{1:T} | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}) \quad (2.2)$$

Розміри кроків, позначені як $\beta_t \in (0, 1)_{t=1}^T$ у рівнянні (2.2), можуть залишатися постійними протягом усього процесу або змінюватися поступово (крок за кроком) чи диференційовано в межах певного діапазону.

Ми можемо переформулювати апостеріорний розподіл, як у рівнянні (2.3), якщо візьмемо

$$\alpha_t = 1 - \beta_t \text{ та } \bar{\alpha}_t = \prod_{i=1}^t \alpha_i$$

Тут $z_{t-1}, z_{t-2}, \dots \sim N(0, I)$, а \bar{z}_{t-2} – це комбінація двох гаусових розподілів.

Для двох гаусових розподілів із різними дисперсіями (наприклад, $N(0, \sigma_1^2 I)$ і $N(0, \sigma_2^2 I)$) їхня комбінація обчислюється як

$$N(0, (\sigma_1^2 + \sigma_2^2)I),$$

а стандартно-квадратичне відхилення дорівнює

$$\sqrt{(1-\alpha_t) + \alpha_t(1-\alpha_{t-1})} = \sqrt{1-\alpha_t \alpha_{t-1}}.$$

Зазвичай, можна робити більший крок оновлення, коли зразок стає більш зашумленим, тобто:

$$\beta_1 < \beta_2 < \dots < \beta_T,$$

і, відповідно,

$$\begin{aligned} x_t &= \sqrt{a_t} x_{t-1} + \sqrt{1-a_t} z_{t-1} \\ &= \sqrt{\alpha_t a_{t-1}} x_{t-2} + \sqrt{1-\alpha_t \alpha_{t-1}} z_{t-2} \\ &= \dots = \sqrt{\bar{a}_t} x_0 + \sqrt{1-\bar{a}_t} z, \end{aligned} \tag{2.3}$$

$$q(x_t | x_0) = N\left(x_t; \sqrt{\bar{a}_t} x_0, (1-\bar{a}_t)I\right).$$

2.3.2 Моделі зворотного дифузійного процесу

Для зворотного процесу ми хочемо оцінити густину імовірності на попередньому кроці. Щоб зробити оцінку попереднього стану, необхідно знати

всі попередні градієнти, і це можна отримати лише за допомогою навчальної моделі. Нейронна мережа з параметрами θ , заданими в рівнянні (2.4), може передбачати $p_\theta(x_{t-1} | x_t)$. Зворотний дифузійний процес починається зі стандартного гаусового розподілу

$$p_\theta(x_T) = \mathcal{N}(x_T; 0, I).$$

$$p_\theta(x_{t-1} | x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)). \quad (2.4)$$

Увесь процес зворотної дифузії задається рівнянням (2.5):

$$p_\theta(x_{0:T}) = p_\theta(x_T) \prod_{t=1}^T p_\theta(x_{t-1} | x_t). \quad (2.5)$$

Імовірнісні моделі дифузії з усуненням шуму навчаються подібно до варіаційних автокодувальників (VAE). Розмір вхідного шару дорівнює розміру вхідних даних. Приховані шари є лінійними шарами з відповідними функціями активації. Вихідний шар має такий самий розмір, як і вхідний, тому модель може відновлювати початкові дані.

Мета нейронної мережі – оптимізувати функцію втрат (loss function), наведену у рівнянні (2.6).

$$\begin{aligned} E[-\log p_\theta(x_0)] &\leq E_q \left[-\log \frac{p_\theta(x_{0:T})}{q(x_{1:T} | x_0)} \right] \\ &= E_q \left[-\log p_\theta(x_T) - \sum_{t=1}^T \log \frac{p_\theta(x_{t-1} | x_t)}{q(x_t | x_{t-1})} \right] \end{aligned} \quad (2.6)$$

Функція втрат покращується шляхом зменшення дисперсії, і вона може бути переписана у вигляді рівняння (2.7).

Тут L_T – це втрата прямого процесу (forward loss), яка показує різницю між розподілом випадкового шуму на останньому кроці прямого процесу та розподілом перед останнім кроком. Це – константа, що залежить від графіка зміни дисперсії (variance schedule).

$L_{1:T-1}$ – це похибка, що представляє суму відмінностей між розподілами прямих і зворотних кроків на кожному етапі зворотного процесу. L_0 – це помилка декодування (decoding error).

$$L = E_q \left[\underbrace{D_{KL}(q(x_T | x_0) \| p(x_T))}_{L_T} + \sum_{t>1} \underbrace{D_{KL}(q(x_{t-1} | x_t, x_0) \| p_\theta(x_{t-1} | x_t))}_{L_{t-1}} - \underbrace{\log p_\theta(x_0 | x_1)}_{L_0} \right] \quad (2.7)$$

Єдиним членом, який ми можемо навчати під час мінімізації рівняння (2.7) за допомогою SGD, є $L_{1:T-1}$. Якщо ми перепараметризуємо апостеріорний розподіл $q(x_{t-1} | x_t, x_0)$ за правилом Байєса, отримаємо рівняння (2.8):

$$q(x_{t-1} | x_t, x_0) = N(x_{t-1}; \bar{\mu}(x_t, x_0), \bar{\beta}_t I) \quad (2.8)$$

Середнє значення та дисперсію тут можна обчислити за рівнянням (2.9):

$$\begin{aligned} \bar{\mu}_t(x_t, x_0) &:= \frac{\sqrt{\alpha_t} \beta_t}{1 - \bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t} (1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t, \\ \bar{\beta}_t &:= \frac{1 - \bar{\alpha}_{t-1}}{1 - \alpha_t} \beta_t. \end{aligned} \quad (2.9)$$

Для обчисленого розподілу в зворотному процесі $p_\theta(x_{t-1} | x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$ функція дисперсії рівна:

$$\Sigma_\theta(x_t, t) = \sigma_t^2 I, \text{ де } \sigma_t^2 = \bar{\beta}_t \text{ і } \sigma_t^2 = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t.$$

Також використаємо спеціальну параметризацію для функції середнього значення $\mu_\theta(x_t, t)$. Якщо записати член L_{t-1} з рівняння (2.7) як

$$p_\theta(x_{t-1} | x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I).$$

То L_{t-1} , наведений у рівнянні (2.10), можна розглядати як помилку L_2 між двома коефіцієнтами середніх значень, тобто:

$$L_{t-1} = E_q \left[\frac{1}{2\sigma_t^2} \|\bar{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2 \right] + C \quad (2.10)$$

Тут C – це константа, що не залежить від θ . Найкраща параметризація для μ_θ – це модель, яка може передбачати $\bar{\mu}_t$. Для $\epsilon \sim N(0, I)$ рівняння (2.3) перепараметризується як $x_t(x_0, \epsilon) = \sqrt{\bar{\alpha}_t} x_0 + (1 - \bar{\alpha}_t) \epsilon$ і записується як у рівнянні (2.11):

$$\begin{aligned} L_{t-1} - C &= E_{x_0, \epsilon} \left[\frac{1}{2\sigma_t^2} \left\| \bar{\mu}_t \left(x_t(x_0, \epsilon), \frac{1}{\sqrt{\bar{\alpha}_t}} x_t(x_0, \epsilon) - \sqrt{1 - \bar{\alpha}_t} \epsilon \right) - \mu_\theta(x_t(x_0, \epsilon), t) \right\|^2 \right] \\ &= E_{x_0, \epsilon} \left[\frac{1}{2\sigma_t^2} \left\| \frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t(x_0, \epsilon) - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon \right) - \mu_\theta(x_t(x_0, \epsilon), t) \right\|^2 \right] \end{aligned} \quad (2.11)$$

Беручи до уваги, що x_t є вхідним сигналом для моделі, параметризацію для μ_θ можна обрати як у рівнянні (2.12). Тут ϵ_θ є функціональним апроксиматором, який націлений оцінювати ϵ з x_t .

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\alpha_t}} \epsilon_\theta(x_t, t) \right). \quad (2.12)$$

Тоді спрощена функція втрат набуває вигляду, наведеного в рівнянні (2.13):

$$L_{t-1} - C = E_{x_0, \epsilon} \left[\frac{\beta_t^2}{2\sigma_t^2 \alpha_t (1-\alpha_t)} \left\| \epsilon - \epsilon_\theta \left(\sqrt{\alpha_t} x_0 + \sqrt{1-\alpha_t} \epsilon, t \right) \right\|^2 \right]. \quad (2.13)$$

Дослідження показали, що варіант рівняння (2.14) для варіаційної межі є більш ефективним для покращення якості зразків і простішим у реалізації, тобто:

$$L_{\text{simple}}(\theta) := E_{x_0, \epsilon} \left[\left\| \epsilon - \epsilon_\theta \left(\sqrt{\alpha_t} x_0 + \sqrt{1-\alpha_t} \epsilon, t \right) \right\|^2 \right] \quad (2.14)$$

Отже, функцію середнього значення можна навчити оцінювати μ_t або параметр ϵ змінюючи параметризацію μ_θ .

Оцінювання параметру ϵ подібне до динаміки Ланжевена і спрощує варіаційну межу дифузійної моделі шляхом імітації узгодження оцінки зашумлення.

Розглянемо алгоритм вибірки. Процес вибірки подібний до процесу динаміки Ланжевена, який навчається градієнту густини даних через ϵ_θ

Увесь процес вибірки наведено в алгоритмі 1.

У випадку, коли $t = T$, процес починається з вибірки

$$x_T \sim N(0, I).$$

Для $z \sim N(0, I)$ вибірка

$$x_{t-1} \sim p_\theta(x_{t-1} | x_t)$$

на кроці $t - 1$ береться з розподілу із середнім значенням, визначеним у рівнянні (2.12).

Алгоритм 1 Алгоритм вибірки DDPM:

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

for $t = T, \dots, 1$ do

 if $t > 1$ then

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

 else

$$z = 0$$

 end if

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t z$$

end for

return x_0

Мережі умовної оцінки шуму (NCSN) [40] моделюють градієнти розподілу даних, а не саму густину імовірності.

У енергетичних моделях (energy-based models) розподіл імовірностей виражається як $p_{\theta}(x)$ у рівнянні (2.15):

$$p_{\theta}(x) = \frac{e^{-E_{\theta}(x)}}{Z_{\theta}} \quad (2.15)$$

Тут E_{θ} – це енергетична функція. Для точок даних із високою ймовірністю енергія є низькою, а для точок із низькою ймовірністю – високою.

Отже, E_{θ} має негативний знак, коли ця функція реалізована як нейронна мережа, параметри позначаються як θ , а вхідні дані – як x .

Вихід мережі є скалярним значенням у діапазоні від $-\infty$ до $+\infty$

Експоненційна операція гарантує, що будь-яке можливе вхідне значення матиме ймовірність, більшу за нуль.

Константа нормалізації Z_{θ} у рівнянні (2.15) задається як

$$Z_{\theta} = \int_x e^{-E_{\theta}(x)} dx$$

для неперервних і забезпечує, що сума густини щільності дорівнює 1

$$\left(\int_x p_{\theta}(x) dx = 1 \right).$$

Щоб обчислити $p_{\theta}(x)$ у рівнянні (2.15), потрібно знайти нормалізуючу константу Z_{θ} . Це – необчислюване значення для великих нейронних мереж, де вхідні дані мають високу розмірність. Енергетичні моделі навчаються за

допомогою таких методів, як контрастивна дивергенція [42]. Однак складного обчислення нормалізаційної константи можна уникнути, моделюючи функцію градієнта (score function) замість безпосереднього моделювання щільності ймовірності. Функція оцінки $s_\theta(x)$ задається у рівнянні (2.16). Оскільки градієнт нормалізаційної константи $\nabla_x \log Z_\theta = 0$, функція оцінки не залежить від Z_θ . Таким чином, не потрібно використовувати спеціальну архітектуру, щоб зробити нормалізаційну константу обчислюваною.

$$s_\theta(x) \approx \nabla_x \log p_\theta(x)$$

$$s_\theta(x) = \nabla_x E_\theta(x) - \nabla_x \log Z_\theta \quad (2.16)$$

$$s_\theta(x) = -\nabla_x E_\theta(x).$$

Після усунення проблеми нормалізаційної константи, $p_\theta(x)$ можна розглядати як нормалізовану функцію щільності ймовірності, і її можна навчати, максимізуючи правдоподібність за рівнянням (2.17):

$$\max_\theta \sum_{i=1}^N \log p_\theta(x_i). \quad (2.17)$$

Однак на практиці якість згенерованих зразків часто є недостатньою, і цю проблему було розв'язано за допомогою дифузійного процесу. Під час семплінгу за допомогою динаміки Ланжевена перший зразок найімовірніше походить із областей низької щільності, особливо коли дані перебувають у високовимірному просторі. Одним із рішень є додавання кількох рівнів шуму до розподілу через дифузійний процес.

Ці багаторазові шуми поступово перетворюють дані у випадковий шум. Розподіл даних спотворюється різними рівнями гаусового шуму, а оцінки (scores) таких зашумлених розподілів потім оцінюються мережею.

Додавання великих рівнів шуму робить дані значно відмінними від оригіналу, тоді як малі шуми можуть не повністю охоплювати області з низькою щільністю. Тому використовуються декілька масштабів шуму, щоб забезпечити адекватне покриття простору.

Області низької щільності заповнюються, коли шум достатньо великий і передбачені оцінки покращуються. Припускаючи, що всі шуми є ізотропними гаусовими, а їхні стандартні відхилення позначені як $\sigma_1 < \sigma_2 < \dots < \sigma_L$ розподіл зашумлених даних позначається як $p_{\sigma_i}(x)$, у рівнянні (2.18):

$$p_{\sigma_i}(x) = \int p(y) \mathcal{N}(x; y, \sigma_i^2 I) dy. \quad (2.18)$$

Додавання шуму на декількох масштабах є критичним для успіху score-based генеративних моделей. Узагальнення кількості масштабів шуму до нескінченності забезпечує високоякісну та контрольовану генерацію, а також робить можливим повний розрахунок log-likelihood.

Мережі умовної оцінки шуму навчаються шляхом мінімізації дивергенції Фішера між розподілом даних і моделлю. Дивергенція Фішера обчислює квадрат L₂-відстані між справжнім score (градієнтом логарифма щільності даних) та score, який передбачає модель.

Оптимізаційний вираз наведено в рівнянні (2.19):

$$E_{p(x)} \left[\left\| \nabla_x \log p(x) - s_{\theta}(x) \right\|_2^2 \right] \quad (2.19)$$

Неможливо обчислити цю дивергенцію безпосередньо, тому що score $\nabla_x \log p(x)$ для розподілу даних є невідомим. Мережі зіставлення результату намагаються передбачити цей справжній результат, змішуючи дані на різних рівнях шуму. Ці мережі оптимізуються безпосередньо на датасеті методом стохастичного градієнтного спуску.

Єдина архітектурна вимога до мереж зіставлення результату, така, щоб вхідний та вихідний розміри повинні збігатися.

Коли оптимізаційний вираз у рівнянні (2.19) записують як у рівнянні (2.20), відстань L_2 зважується коефіцієнтом $p(x)$.

Тому передбачувані значення-функції працюють неправильно у регіонах з низькою густиною.

$$E_{p(x)} \left[\left\| \nabla_x \log p(x) - s_\theta(x) \right\|_2^2 \right] = \int p(x) \left\| \nabla_x \log p(x) - s_\theta(x) \right\|_2^2 dx \quad (2.20)$$

Через проблему семплінгу в областях низької густини, метод зіставлення оцінок зі зменшенням шуму (DSM) [37], наведений у рівнянні (2.21), пропонує ефективне розв'язання. У цьому методі початкова score-щільність спотворюється шляхом додавання зростаючої послідовності шумів.

Тут $p_\sigma(\bar{x}|x)$ позначає спотворений розподіл даних.

Метод є точним лише тоді, коли шум достатньо малий, щоб

$$p_\sigma(\bar{x}|x) \approx p_{\text{data}}(x).$$

$$\frac{1}{2} E_{p_\sigma(\bar{x}|x) p_{\text{data}}(x)} \left[\left\| s_\theta(\bar{x}) - \nabla_x \log p_\sigma(\bar{x}|x) \right\|_2^2 \right] \quad (2.21)$$

Тут, якщо спотворений розподіл даних задається як

$$p_{\sigma}(\bar{x}|x) = \mathcal{N}(\bar{x}|x, \sigma^2 I),$$

Тоді

$$\nabla_x \log p_{\sigma}(\bar{x}|x) = -\frac{\bar{x} - x}{\sigma^2}.$$

Відповідно, для заданого значення σ цільова функція задається як у рівнянні (2.22):

$$\ell(\theta; \sigma) = \frac{1}{2} E_{p_{\text{data}}(x)} E_{x \sim \mathcal{N}(x, \sigma^2 I)} \left[\left\| s_{\theta}(\bar{x}) + \frac{\bar{x} - x}{\sigma^2} \right\|_2^2 \right]. \quad (2.22)$$

Якщо рівняння (2.22) об'єднати для всіх доступних $\sigma \in \{\sigma_i\}_{i=1}^L$ тоді зважену суму дивергенцій Фішера на різних рівнях шуму отримуємо в рівнянні (2.23). Тут $\lambda(\sigma_i) \in \mathbb{R}_{>0}$ є додатною зважувальною функцією, і зазвичай при оптимізації score-matching обирають $\lambda(\sigma_i) = \sigma_i^2$.

$$\mathcal{L}(\theta; \{\sigma_i\}_{i=1}^L) = \frac{1}{L} \sum_{i=1}^L \lambda(\sigma_i) \ell(\theta; \sigma_i). \quad (2.23)$$

Sliced Score Matching (SSM) [40] – це ще один метод зіставлення результату, який оцінює неспотворений справжній результат через режим автодиференціювання в прямому режимі, проєктуючи результат у випадковий вектор. Цей метод дає результати, подібні до DSM, але потребує у 4 рази більшої обчислювальної потужності.

Динаміка Ланжевена, поєднана зі стохастичним градієнтним спуском, може генерувати зразки з розподілу, використовуючи градієнти в ланцюгу Маркова. У порівнянні зі звичайним SGD, стохастична градієнтна динаміка

Ланжевена додає гаусовий шум до оновлень параметрів, щоб уникнути локальних мінімумів. Повний процес вибірки описано в Алгоритмі 2.

Процес починається з вибірки \bar{x}_0 із початкового розподілу неструктурованого випадкового шуму. Ці зразки і крок оновлення використовуються для вибірки з розподілу P_{σ_1} . На кожному кроці крок оновлення зменшується, а зразки модифікуються таким чином, щоб поступово ставати реалістичними.

Алгоритм 2. Алгоритм вибірки мережі умовної оцінки шуму

Require: $\{\sigma_i\}, m, T$

Initialize \bar{x}_0

for $i=1, \dots, L$ do

$$a_i = m \cdot \frac{\sigma_i^2}{\sigma_L^2}$$

for $t=1, \dots, T$ do

Draw $z_t \sim N(0, I)$

$$\bar{x}_{0t} = \bar{x}_{0t-1} + \frac{a_i}{2} s_{\theta}(\bar{x}_{0t-1}, \sigma_i) + \sqrt{a_i} z_t$$

end for

$\bar{x}_0 \leftarrow \bar{x}_T$

end for

return \bar{x}_T

Оскільки шум поступово зменшується, цей процес називається динаміка відпалу Ланжевена (ALD). Таким чином, проблема в областях низької густини усувається. Коли σ_1 вибрано достатньо великим, області низької густини у P_{σ_1} значно зменшуються, і можна отримати зразки високої якості. Ці зразки є хорошою стартовою точкою для наступного кроку й забезпечують високоякісний семплінг із розподілу на останньому кроці P_{σ_L} .

2.4 Обчислення 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 [43]. Однак дана метрика базується на імовірності приналежності зображення до одного із класів і не може відобразити перенавчання мережі. В якості кращої альтернативи запропоновано FID [44]. 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:

$$IS(G) \approx \exp(E_{x \sim p_g}[D_{KL}(p(y|x) || p(y))]) \quad (2.24)$$

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 Inception, 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 [45].

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 Inception. 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 [44]. 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:

$$d^2((m_r C_r), (m_g C_g)) = \|m_r - m_g\|^2 + Tr(C_r + C_g - 2(C_r C_g)^{\frac{1}{2}}), \quad (2.25)$$

де $(m_r C_r)$ та $(m_g C_g)$ – середнє та коваріаційна матриця 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 [46]. *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* [47, 48].

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.

2.5 Висновки до розділу 2

Отже, у даному розділі проаналізовані генеративні моделі, які використовуються для розширення початкових вибірок зображень, а також дифузійні моделі. На основі дифузійних моделей розроблено алгоритми генерування зображень. Для оцінки згенерованих зображень проаналізовані метрики оцінки зображень: IS, FID.

3 СИНТЕЗ ЦИТОЛОГІЧНИХ ЗОБРАЖЕНЬ В СЕРЕДОВИЩІ STABLE DIFFUSION

3.1 Програмне середовище Stable Diffusion

У цьому підрозділі розглядається веб-інтерфейс, призначений для роботи з моделями Stable Diffusion.

AUTOMATIC1111 – це зручне та функціональне веб-середовище, розроблене для спрощення взаємодії з генеративними моделями зображень (рисунок 3.1) [49]. Воно надає користувачам інтуїтивний інтерфейс для створення, редагування та налаштування параметрів генерації без необхідності прямої роботи з командним рядком чи програмним кодом.

Інтерфейс підтримує широкий спектр інструментів – від базової генерації зображень за текстовим описом (text-to-image) до складних режимів, таких як image-to-image, inpainting, outpainting та навчання користувацьких моделей (LoRA, Textual Inversion, DreamBooth).

Завдяки можливості роботи через веб-браузер, AUTOMATIC1111 забезпечує зручний доступ до всіх функцій Stable Diffusion, підтримує використання додаткових розширень і плагінів, а також дає змогу керувати параметрами дифузії, розміром зображення, кількістю ітерацій та стилем генерації [50].

Завдяки своїй гнучкості, відкритому коду та активній спільноті розробників, AUTOMATIC1111 став одним із найпопулярніших інструментів для роботи з моделями Stable Diffusion, широко використовується як у наукових дослідженнях, так і в практичних експериментах із генерацією візуальних даних.

Він використовується для генерації зображень за текстовими підказками (prompts), а також для виконання різноманітних операцій із моделлю Stable

Diffusion, зокрема редагування, повторного рендерингу, керування параметрами дифузії та збереження результатів.

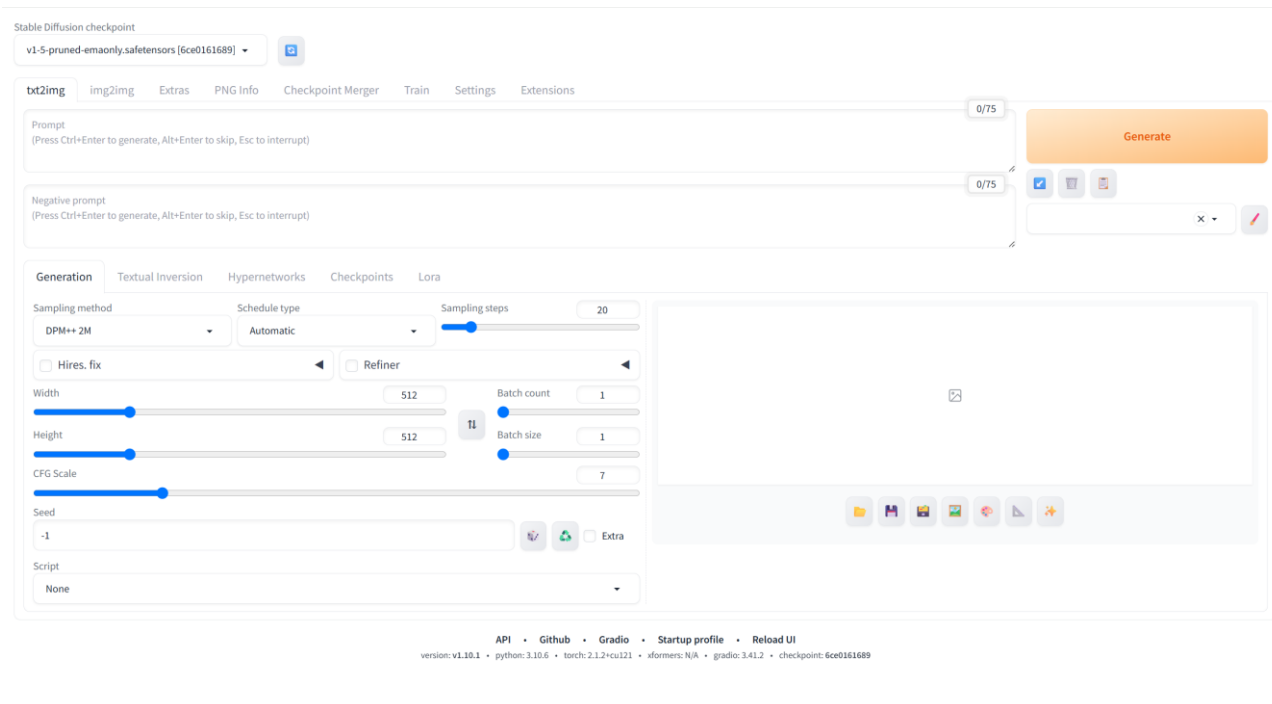


Рисунок 3.1 – Веб-інтерфейс AUTOMATIC1111 (вкладка Generation)

На рисунку 3.2 наведено графічну схему інтерфейсу, що демонструє його основну структуру та взаємозв'язки між основними компонентами. Інтерфейс організований логічно та зручно для користувача: усі ключові елементи розташовані у вигляді панелей та вкладок, що дає змогу швидко переходити між режимами роботи – генерацією, попереднім переглядом, історією або налаштуваннями.

Нижче подано детальний опис основних функціональних елементів інтерфейсу, які забезпечують ефективну взаємодію з моделлю Stable Diffusion.

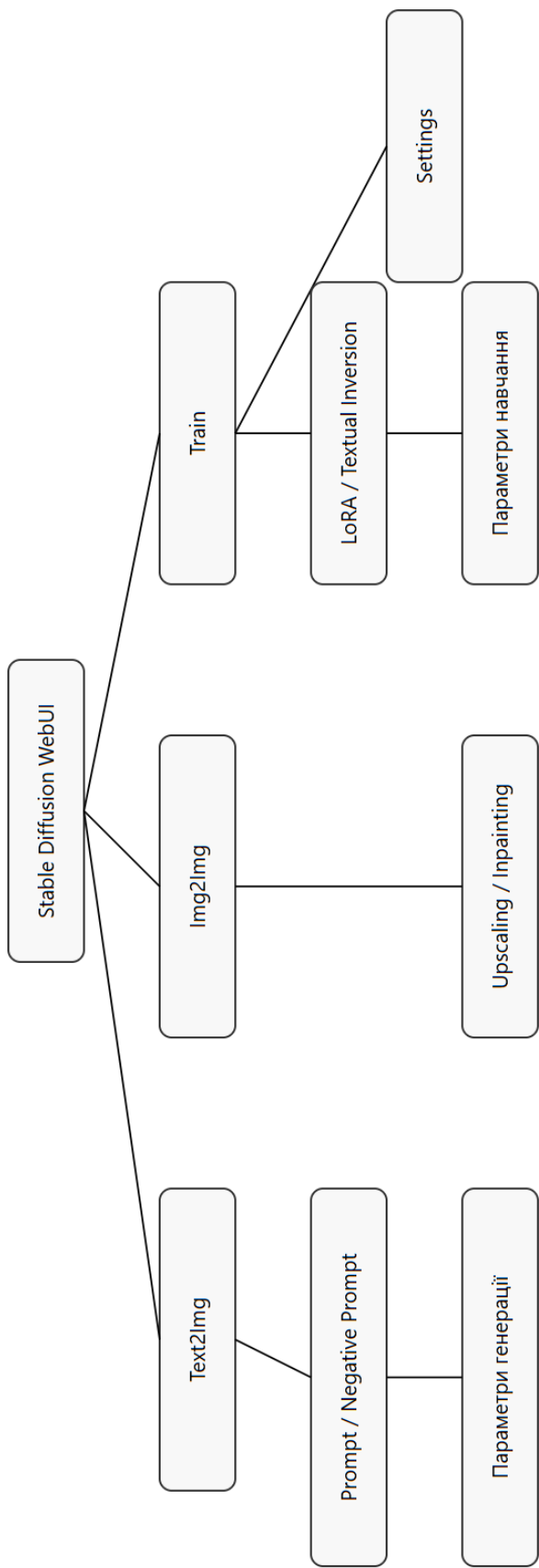


Рисунок 3.2 – Схема графічного інтерфейсу Stable Diffusion WebUI

Верхня панель

У верхній частині інтерфейсу розташовано основні елементи керування, які дозволяють швидко переходити між різними режимами роботи. Центральним елементом є Checkpoint, який визначає активну модель, що використовується для генерації зображень. У прикладі наведено модель v1-5-pruned-emaonly.safetensors, одна з найпоширеніших базових версій Stable Diffusion. Вибір контрольної точки безпосередньо впливає на характер і стиль результатів – різні моделі можуть бути спеціалізовані на портретах, пейзажах, мистецьких стилях або навіть технічних ілюстраціях.

Нижче розташовані вкладки (Tabs), які структурують роботу користувача. Основна вкладка txt2img призначена для створення зображень за текстовими запитами – саме тут вводяться ключові підказки (prompts), що описують бажаний результат. Вкладка img2img дозволяє використовувати вже наявне зображення як основу, модифікуючи його за допомогою текстових інструкцій – це зручно для корекції або стилізації. Розділ Extras містить інструменти для масштабування, підвищення роздільної здатності та відновлення дрібних деталей.

У вкладці PNG Info можна переглянути метадані згенерованих зображень – усі параметри, з якими вони були створені. Checkpoint Merger дозволяє об'єднувати різні контрольні точки, формуючи нові моделі на основі вже існуючих. Вкладка Train використовується для навчання моделей, а Settings – для налаштування інтерфейсу, ресурсів GPU, шляхів до моделей і додаткових функцій. Окремо передбачено Extensions – систему розширень, яка дозволяє підключати нові інструменти, модулі або функціональні можливості, розширюючи базовий набір WebUI.

Поля введення текстових підказок

Центральною частиною інтерфейсу є поля Prompt і Negative Prompt, які визначають зміст майбутнього зображення. У полі Prompt користувач задає опис бажаного результату – наприклад, «реалістичний портрет у м'якому світлі» чи

«наукова ілюстрація клітинної структури». Саме цей опис визначає композицію, кольорову гаму, стиль і загальний настрій зображення.

Поле Negative Prompt виключає небажані елементи, такі як «низька якість», «розмиття» або «артефакти». Цей параметр є важливим для очищення результату й досягнення більшої точності у генерації.

Панель генерації

У цьому блоці користувач задає основні параметри синтезу. Першим елементом є Sampling method – алгоритм вибірки, який визначає, як модель поступово відновлює зображення з шуму [51]. Найчастіше використовується метод Euler, що забезпечує баланс між швидкістю та якістю. Інші методи, такі як DDIM чи DPM++, можуть давати іншу текстуру або глибину, тому підбір оптимального алгоритму залежить від задачі.

Sampling steps визначає кількість ітерацій генерації – чим їх більше, тим детальніше опрацьовується кожен етап, проте зростає час виконання. Опція Hires.fix застосовується для покращення роздільної здатності після початкового рендеру: вона дозволяє регенерувати зображення з більшим масштабом, зберігаючи при цьому деталі та композицію. Параметри Width / Height задають розміри кінцевого зображення, що впливають на пропорції та формат – від квадратних (512×512) до панорамних або вертикальних.

Окремо виділено параметр CFG Scale – один із ключових для контролю творчості моделі. Він визначає, наскільки точно зображення має відповідати текстовій підказці. Низькі значення (5–7) дають більше варіацій, а високі (10–12) – суворо слідуєть тексту, проте можуть втрачати природність. Seed задає початкове число для генерації випадковості. Якщо вказано -1, система створює унікальний seed при кожному запуску, забезпечуючи різні результати навіть із однаковими налаштуваннями.

Параметри обробки

Для підвищення ефективності роботи передбачено параметри Batch count і Batch size. Вони дозволяють створювати одразу кілька варіантів одного запиту

або генерувати зображення серіями. Це зручно, якщо потрібно вибрати найкращий результат серед кількох автоматично створених.

Додаткові кнопки та нижня панель

Головна кнопка Generate запускає процес генерації – після натискання система починає обчислення на GPU. Поруч розташовані допоміжні кнопки для завантаження зображення, очищення полів або збереження підказок. Вони прискорюють робочий процес і дозволяють швидко повторювати попередні запити.

У нижній частині інтерфейсу знаходиться службова інформація. Тут можна побачити підключення до API, посилання на GitHub-репозиторій проєкту, а також логотип Gradio, який відповідає за реалізацію вебінтерфейсу. Пункти Startup profile і Reload UI використовуються для оновлення конфігурації або перезапуску користувацького інтерфейсу без повного перезавантаження програми.

Таким чином, інтерфейс AUTOMATIC1111 Stable Diffusion WebUI поєднує гнучкість, функціональність і зручність користування. Він дозволяє користувачу не лише керувати параметрами генерації, а й детально налаштовувати процес створення зображень, досягаючи високої якості та художньої виразності результатів.

Основні параметри інтерфейсу Automatic1111 Stable Diffusion WebUI дозволяють точно керувати процесом генерації зображень і досягати бажаного стилю, деталізації та якості результату.

Одним із найважливіших параметрів є Sampling Method (метод вибірки), який визначає алгоритм, що використовується під час дифузійного процесу. Від вибору методу залежить, як саме модель буде поступово перетворювати шум у реалістичне зображення. Наприклад, метод Euler забезпечує швидку й стабільну генерацію, підходить для створення простих або малих зображень. Варіант Euler а дає більш художні, варіативні результати, злегка змінюючи структуру деталей на кожному кроці. Алгоритм DPM++ 2M дозволяє досягти

більш контрольованої та якісної генерації, а DDIM використовується, коли потрібен високий рівень деталізації або точність передачі текстури. Вибір методу вибірки суттєво впливає на баланс між швидкістю, реалістичністю та характером стилю зображення.

Параметр *Sampling Steps* (кроки вибірки) визначає кількість ітерацій, які виконує модель під час процесу генерації. Чим більше кроків – тим більш опрацьованим і деталізованим буде зображення. Наприклад, при 20 кроках результат з'являється швидше, але може бути менш чітким; натомість 50 і більше кроків забезпечують значно кращу деталізацію, хоча процес триває довше. Таким чином, користувач самостійно обирає оптимальний баланс між швидкістю роботи та якістю зображення.

Розділ *Width / Height* (ширина і висота) визначає розмір кінцевого зображення в пікселях. Вибір більшого розміру вимагає більших обчислювальних ресурсів і збільшує час обробки, оскільки модель повинна генерувати більше пікселів і враховувати більше візуальних зв'язків. Для великих форматів часто використовується опція *Hires.fix*, яка дозволяє спочатку створити базове зображення, а потім масштабувати його без втрати якості, зберігаючи композицію і додаючи додаткові деталі.

Окрему роль відіграє параметр *CFG Scale* (*Classifier-Free Guidance Scale*), що контролює ступінь відповідності між текстовою підказкою та результатом генерації. Якщо це значення невелике (близько 5–7), зображення може бути більш креативним і вільним, із м'якою інтерпретацією запиту. Якщо ж воно високе (10 і більше), система суворо дотримується тексту, створюючи більш точні, але іноді менш “живі” результати. Цей параметр дає змогу знайти баланс між художньою варіативністю та семантичною точністю.

Seed (початкове значення випадковості) – це числовий параметр, що визначає, як саме модель почне процес генерації. Якщо залишити значення –1, система автоматично створить випадкове початкове число, завдяки чому кожне зображення буде унікальним навіть при однаковому запиті. Якщо ж

використати певний seed (наприклад, 42), можна відтворити той самий результат, що зручно для повторних експериментів або демонстрації стабільності налаштувань.

Параметри Batch Count і Batch Size дозволяють зручно керувати кількістю зображень, що створюються за один сеанс. Batch Count визначає кількість серій, які система згенерує послідовно, тоді як Batch Size – кількість зображень, що створюються одночасно в одній партії. Ці параметри корисні для порівняння варіантів результатів або при створенні наборів зображень із різними варіаціями композиції. Проте варто враховувати, що збільшення розміру партії вимагає більшої кількості відеопам'яті (VRAM) та може уповільнити процес.

Функція Hires. Fix (високоякісна обробка) застосовується для покращення якості фінального результату після первинної генерації. Вона дозволяє “перегенерувати” вже створене зображення з більшою роздільною здатністю, додаючи дрібні деталі, текстури та підвищуючи загальну чіткість. Це особливо важливо при створенні великих візуалізацій, плакатів або медичних зображень, де дрібні деталі відіграють вирішальну роль. Однак ця функція значно збільшує час обробки та навантаження на апаратну частину.

Окремо варто згадати параметр Negative Prompt, який допомагає уникнути небажаних артефактів. Тут користувач може вказати, яких об'єктів, стилів чи ефектів не повинно бути у фінальному результаті. Наприклад, такі слова як “grainy”, “blurry” чи “low quality” дозволяють прибрати зернистість, розмиття або інші дефекти. Цей інструмент часто використовується для підвищення точності зображень і досягнення чистішої, більш естетичної композиції.

Завдяки цим параметрам інтерфейс Automatic1111 WebUI забезпечує повний контроль над процесом генерації. Користувач може варіювати стиль, деталізацію, роздільну здатність і навіть ступінь випадковості, отримуючи зображення, що максимально відповідають творчим або науковим цілям. Комбінація параметрів дозволяє не лише досягати стабільних результатів, але й

експериментувати з новими візуальними підходами, що робить середовище Stable Diffusion універсальним інструментом як для художників, так і для дослідників у галузі комп'ютерного зору та біомедичних візуалізацій.

На рисунку 3.3 зображено інтерфейс вкладки Train → Create Hypernetwork, що дає можливість користувачам створювати власні гіпермоделі (Hypernetwork). Цей розділ використовується для тонкого налаштування процесу генерації, адаптації його під конкретні стилі, об'єкти або набори даних. Під час створення нової гіпермоделі задається її назва, а також модулі, що визначають рівень роздільної здатності та структуру моделі. Таким чином, користувач може не лише керувати процесом генерації, а й створювати персоналізовані моделі, орієнтовані на конкретні візуальні або наукові задачі.

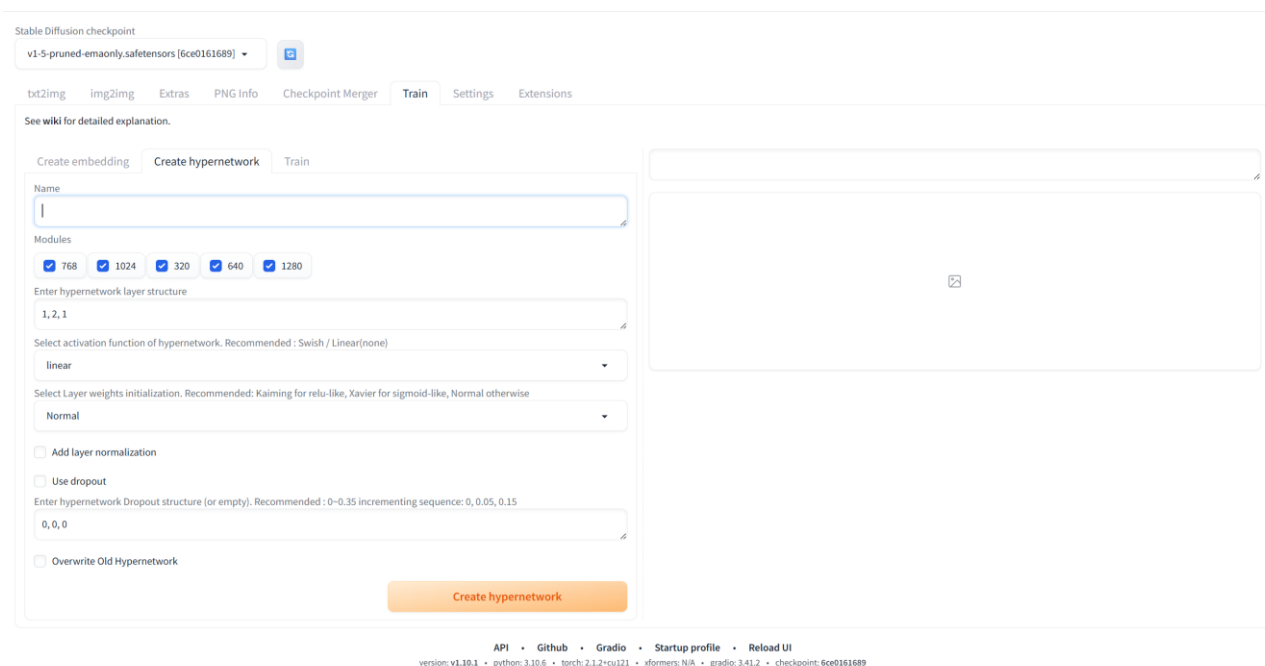


Рисунок 3.3 – Зображення інтерфейсу AUTOMATIC1111 (Train)

Доступні значення параметрів роздільної здатності під час навчання можуть становити 768, 1024, 320, 640, 1280 пікселів. Вибір конкретного значення залежить від бажаного рівня деталізації та складності навчання - чим

більша роздільна здатність, тим точнішим буде результат, але й тим більші ресурси потрібно для обчислень.

Параметр Enter hypernetwork layer structure визначає архітектуру гіпермоделі – кількість шарів і кількість вузлів у кожному з них. Наприклад, конфігурація 1, 2, 1 означає наявність трьох шарів, де середній шар містить удвічі більше вузлів, що забезпечує гнучкість і стабільність навчання.

Далі задається Select activation function of hypernetwork, тобто вибір функції активації, яка визначає, як модель обробляє сигнали між шарами.

Рекомендовано використовувати:

- Swish – нелінійну функцію активації, що показує хороші результати в генеративних моделях і допомагає мережі краще навчатися складним залежностям.

- Linear (none) – лінійну активацію без додаткової обробки сигналу, коли потрібно зберегти просту структуру.

Наступним кроком є Select Layer weights initialization, тобто метод ініціалізації ваг нейронної мережі.

Рекомендації такі:

- Kaiming – добре працює з активаціями на основі ReLU, забезпечуючи швидку стабілізацію градієнтів.

- Xavier – підходить для активацій типу sigmoid або tanh, зменшуючи ризик “затухання” сигналу.

- Normal – універсальний варіант, який застосовується у більшості випадків.

Опція Add layer normalization дозволяє додати нормалізацію шарів, що стабілізує процес навчання, запобігає перенасиченню градієнтів і підвищує точність результатів.

Пункт Use dropout відповідає за використання техніки Dropout, яка випадково “вимикає” частину нейронів під час навчання. Це допомагає запобігти перенавчанню моделі та покращує її узагальнюючу здатність. У полі

Enter hypernetwork Dropout structure можна задати значення, наприклад 0, 0.05, 0.15, що означає поступове збільшення ймовірності виключення нейронів у кожному шарі.

Опція Overwrite Old Hypernetwork дозволяє перезаписати вже існуючу гіпермодель із такою ж назвою – корисно при повторному навчанні або оновленні попередніх результатів.

Для запуску процесу використовується кнопка Create hypernetwork, яка активує створення гіпермоделі з обраними параметрами. Гіпермоделі дозволяють оптимізувати роботу Stable Diffusion, роблячи її здатною до відтворення унікальних стилів, кольорових палітр або навіть авторських художніх технік. Завдяки цьому користувач отримує можливість формувати індивідуальні моделі для конкретних задач – від мистецьких проєктів до наукової візуалізації.

Ця секція інтерфейсу є потужним інструментом персоналізації, що забезпечує контроль над архітектурою, активацією, нормалізацією та структурою навчання. Вона відкриває широкі можливості для експериментів із глибиною, складністю та поведінкою нейромереж.

На рисунку 3.4 наведено приклад вкладки Train у Stable Diffusion WebUI, яка використовується для навчання гіпермоделей (Hypernetwork) або ембеддингів (Embedding). Інтерфейс дозволяє задавати параметри навчання, вибирати тип моделі та контролювати процес тренування в реальному часі.

У верхній частині розташовано перемикач Embedding / Hypernetwork, який визначає, що саме буде навчатися - ембеддинг (векторне уявлення конкретного поняття) або гіпермодель (розширена структура для стилізації та адаптації моделі).

У секції Параметри навчання задаються основні характеристики процесу. Параметр Learning Rate визначає швидкість, з якою модель оновлює свої ваги під час навчання. Для ембеддингів зазвичай використовують значення близько

0.005, а для гіпермоделей – 0.00001, що забезпечує плавне та стабільне навчання без різких стрибків у результатах.

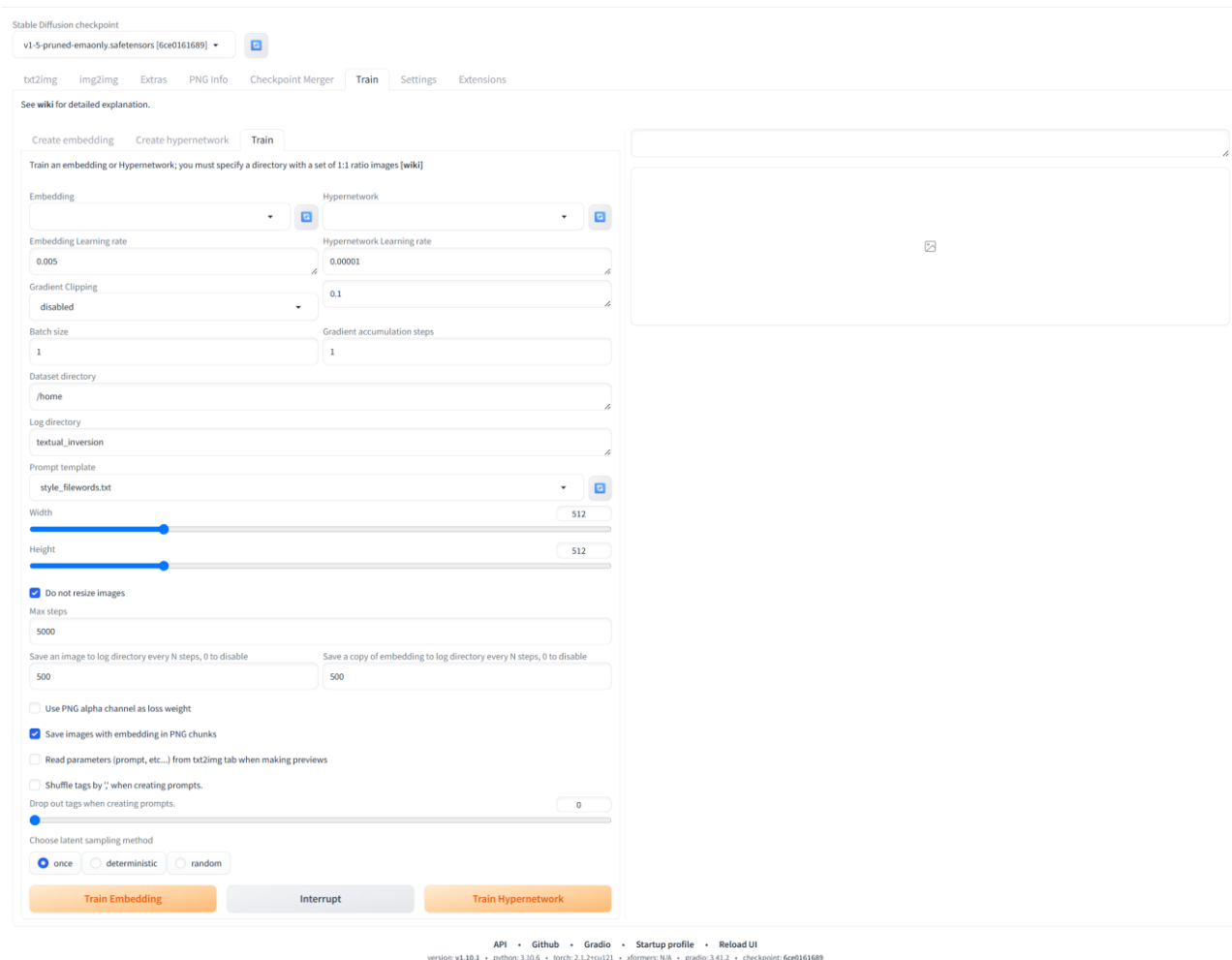


Рисунок 3.4 – Зображення веб-інтерфейсу AUTOMATIC1111 (вкладка Train)

Gradient Clipping обмежує максимальну величину градієнта, запобігаючи нестабільності та перенасиченню під час оновлення ваг. У полі Value задається, наприклад, значення 0.1.

Batch Size визначає кількість зображень, які одночасно обробляються моделлю – наприклад, значення 1 означає, що під час кожного кроку обробляється лише одне зображення.

Параметр Gradient Accumulation Steps встановлює, скільки кроків буде зроблено до оновлення ваг – це корисно, коли ресурси обмежені, але потрібно накопичити ефект більших батчів.

Далі налаштовується робота з даними. Dataset Directory - шлях до папки з навчальними зображеннями, Log Directory – місце для збереження результатів і журналів навчання, Prompt Template – шаблон текстових підказок, що задає зміст генерацій (наприклад, style_fireworks.txt). Width та Height визначають роздільну здатність вхідних зображень, зазвичай 512×512 пікселів. Якщо активувати параметр Do not resize images, система залишає оригінальні розміри файлів без масштабування.

Секція Параметри логування дозволяє контролювати частоту збереження результатів:

- Max Steps задає максимальну кількість ітерацій (наприклад, 40 000 кроків).
- Save an image to log directory every N steps визначає інтервал збереження проміжних зображень для візуального моніторингу.

У блоці Додаткові опції передбачено низку корисних функцій. Опція Use PNG alpha channel as loss weight використовує альфа-канал зображень як вагу втрат, а Shuffle tags by ‘;’ when creating prompts - випадково перемішує теги у текстових підказках, що підвищує варіативність навчання. Також можна вибрати метод вибірки латентного простору (once, deterministic або random), який впливає на те, як модель опрацьовує приховані змінні під час навчання.

У нижній частині розташовані кнопки дій: Train Embedding запускає процес навчання ембеддингу, Train Hypernetwork - тренування гіпермоделі, а Interrupt дозволяє у будь-який момент зупинити процес.

Загалом, ця вкладка забезпечує повний контроль над процесом навчання – від вибору архітектури до збереження результатів. Вона є важливою частиною середовища Stable Diffusion, що дає змогу користувачам створювати власні унікальні моделі, адаптовані до конкретних цілей і творчих задач.

3.2 Апаратно-програмні вимоги до проведення експериментів

Для проведення комп'ютерних експериментів із синтезу цитологічних зображень було використано середовище Stable Diffusion WebUI від AUTOMATIC1111. Конфігурація програмного та апаратного забезпечення забезпечувала високу продуктивність та стабільність роботи під час генерації великої кількості зображень.

1. Конфігурація середовища

Система:

- Операційна система: Ubuntu 22.04.3 LTS (Jammy Jellyfish)
- Distributor ID: Ubuntu
- Release: 22.04

Апаратна конфігурація:

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

Ця конфігурація дозволяє виконувати генерацію зображень високої якості в режимі реального часу та підтримує складні дифузійні обчислення, необхідні для синтезу цитологічних структур.

2. Характеристики програмного середовища AUTOMATIC1111

- Версія WebUI: v1.10.1
- Python: 3.10.6
- Torch: 2.1.2+cu121
- Xformers: 0.0.23.post1
- Gradio: 3.41.2
- Checkpoint: 6ce0161689

- Модель: v1-5-pruned-emaonly.safetensors

3. Конфігурація параметрів генерації

Під час експериментів використовувалася вкладка txt2img із такими налаштуваннями:

- Sampling method: DPM++ 2M
- Sampling steps: 20
- Hires. fix: деактивовано
- Width / Height: 512 × 512 пікселів
- CFG Scale: 7
- Seed: -1 (випадкове значення)
- Batch count: 1
- Batch size: 1

Такі параметри забезпечували стабільну генерацію та дозволяли отримувати реалістичні синтетичні цитологічні зображення з належним рівнем деталізації.

У таблицях 3.1 та 3.2 наведено відомості про орендну сесію обчислювальних ресурсів і параметри навчання вибірок відповідно.

Таблиця 3.1 – Інформація про орендну сесію

Дата	Device Type	No Of GPUS	No Of CPUS	RAM (GB)	VRAM (GB)	Storage (GB)	Hours Ran	Total Cost (\$)
04.08.2025	A6000 Ampere	1	7	32	48	200 GB	6H:13M	\$5.36

Перед початком проведення експериментів вибірка зображень була розширена за допомогою алгоритмів афінного спотворення, реалізованих у бібліотеці Rudi, написаній мовою програмування Python.

Таблиця 3.2 – Дані про навчання на основі 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)
cyt_kist_mast	12	100	20000	200	92	0.0046	27.6
cyt_mast	21	100	20000	200	92	0.0046	27.6
cyt_nep_fib_mast	31	100	20000	200	92	0.0046	27.6
cyt_nep_mast	14	100	20000	200	92	0.0046	27.6

Бібліотека Rudi використовується для автоматичної аугментації (створення варіацій) обрізаних зображень, що дозволяє збільшити різноманітність даних для подальшого навчання моделі. У процесі підготовки вибірки були виконані приведені нижче операції автоматичної аугментації.

Для кожного набору даних було використано однакову кількість навчальних кроків та епох, що дозволило забезпечити узгодженість результатів між різними типами цитологічних зразків.

Після виконання процедур аугментації всі вибірки були збалансовані до 100 зображень, що підвищило стабільність навчання та покращило здатність моделі до генерації реалістичних синтетичних зображень.

Середній час навчання становив 92 хвилини, при цьому один крок займав близько 0.0046 хвилини, а одна епоха – 27.6 секунд, що свідчить про ефективне використання обчислювальних ресурсів.

Під час підготовки вибірки для навчання було використано алгоритм афінного спотворення, реалізований у бібліотеці Rudi, що дозволив виконати

автоматичну аугментацію зображень та суттєво збільшити їх кількість. Ймовірність застосування кожної трансформації становила 70% ($p = 0.7$), тобто кожна операція мала шанс бути реалізованою з імовірністю сім із десяти.

Для додання варіативності орієнтації зображень застосовувалися повороти з максимальною амплітудою до 15° вліво та вправо. Масштабування виконувалося в межах від 1 до 1.2, що дозволяло зберегти пропорції зображення, водночас додаючи легке збільшення без суттєвого викривлення структури клітин. Інтенсивність спотворень визначалася параметром $magnitude = 5$, який регулював ступінь дії різних трансформацій (зокрема поворотів і масштабувань), забезпечуючи різноманітність створюваних зразків.

Кількість згенерованих варіантів для кожного вихідного зображення складала 125 одиниць, що дало змогу значно розширити початковий набір даних. Процес аугментації реалізовувався послідовно: спочатку виконувалося обрізання зображень до квадратного формату, після чого запускалася команда `rudi augment`, що створювала нові зображення з випадковими змінами. Після генерації всі результати автоматично переміщувалися з тимчасової директорії до основної папки, де формувався підсумковий набір даних для подальшого навчання моделі.

Таким чином, використання бібліотеки `Rudi` дало змогу автоматизувати процес підготовки навчальної вибірки, збільшити її розмір і різноманітність, а також покращити загальну здатність моделі до узагальнення даних під час подальшого синтезу цитологічних зображень.

3.3 Результати комп'ютерних експериментів

У ході комп'ютерних експериментів було проведено синтез зображень на основі чотирьох підготовлених наборів цитологічних даних: `cyt_kist_mast`,

cyt_mast, cyt_per_fib_mast та cyt_per_mast. Кожен набір після попередньої аугментації містив по 100 зображень, що забезпечувало достатню кількість варіацій для навчання моделі та подальшого тестування.

Для кожного з наборів навчання проводилось упродовж 20 000 кроків із кількістю 200 епох, що дозволило досягти стабільної якості відтворення структурних елементів клітин. Загальний час навчання кожної моделі становив приблизно 92 хвилини, при цьому середній час виконання одного кроку складав 0.0046 хвилини, а одна епоха тривала близько 27.6 секунд. Такі параметри навчання свідчать про ефективне використання обчислювальних ресурсів і збалансованість процесу генерації.

Після завершення навчання модель успішно синтезувала реалістичні цитологічні зображення, що відтворюють характерні морфологічні особливості клітинних структур для кожного типу набору. Зокрема, було отримано синтетичні варіанти клітин із різним ступенем щільності, кольорової насиченості та структурної деталізації, що підтверджує можливість моделі *Stable Diffusion* ефективно працювати із спеціалізованими медичними зображеннями.

На рисунку 3.5 представлено приклади оригінальних і згенерованих зображень першого набору (cyt_kist_mast), які демонструють подібність морфологічних характеристик між реальними та синтезованими вибірками, а також високу якість генерації на основі навченої моделі.

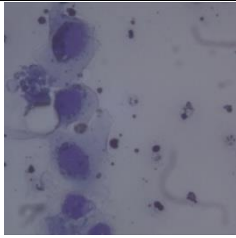
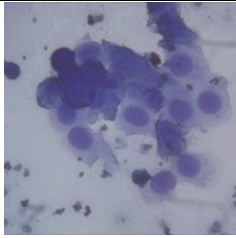
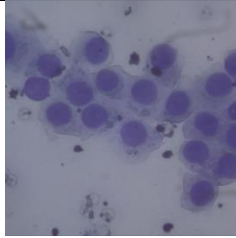
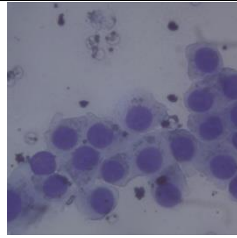
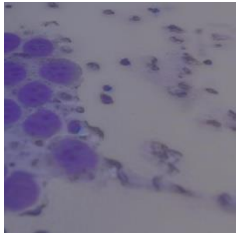
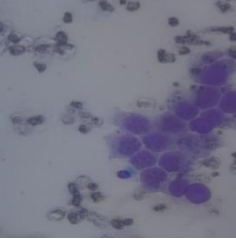
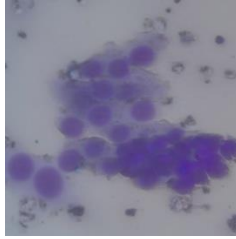
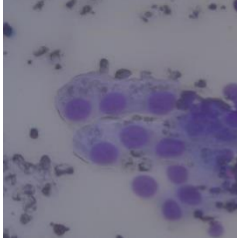
Original				
Steps	5 000	10 000	15 000	20 000
Generateed				

Рисунок 3.5 – Приклад реальних і синтезованих зображень класу 1

На рисунку 3.6 наведено результати тестування згенерованих цитологічних зображень, отриманих у процесі навчання моделі, із використанням двох основних метрик – FID (Frechet Inception Distance) та IS (Inception Score), які оцінюють якість та різноманітність синтетичних даних. Результати експериментів показали, що модель демонструє стабільну генерацію зображень із прийнятною якістю.

- FID (Frechet Inception Distance): У ході навчання значення FID поступово знижувалося, залишаючись у межах 0.42–0.58. Найнижче значення було зафіксовано на 15 000-й ітерації, що свідчить про найкращу схожість зображень, створених моделлю, до реальних цитологічних зразків. Після досягнення цієї точки спостерігалось незначне коливання метрики, що є типовим для стабілізованої генеративної моделі.

- IS (Inception Score): Показник різноманітності та реалістичності зображень коливався в межах 2.6–2.9, досягаючи максимуму на 20 000-й ітерації. Це свідчить про те, що модель не лише навчилася відтворювати структуру клітин, а й зберегла варіативність у їхніх морфологічних особливостях.

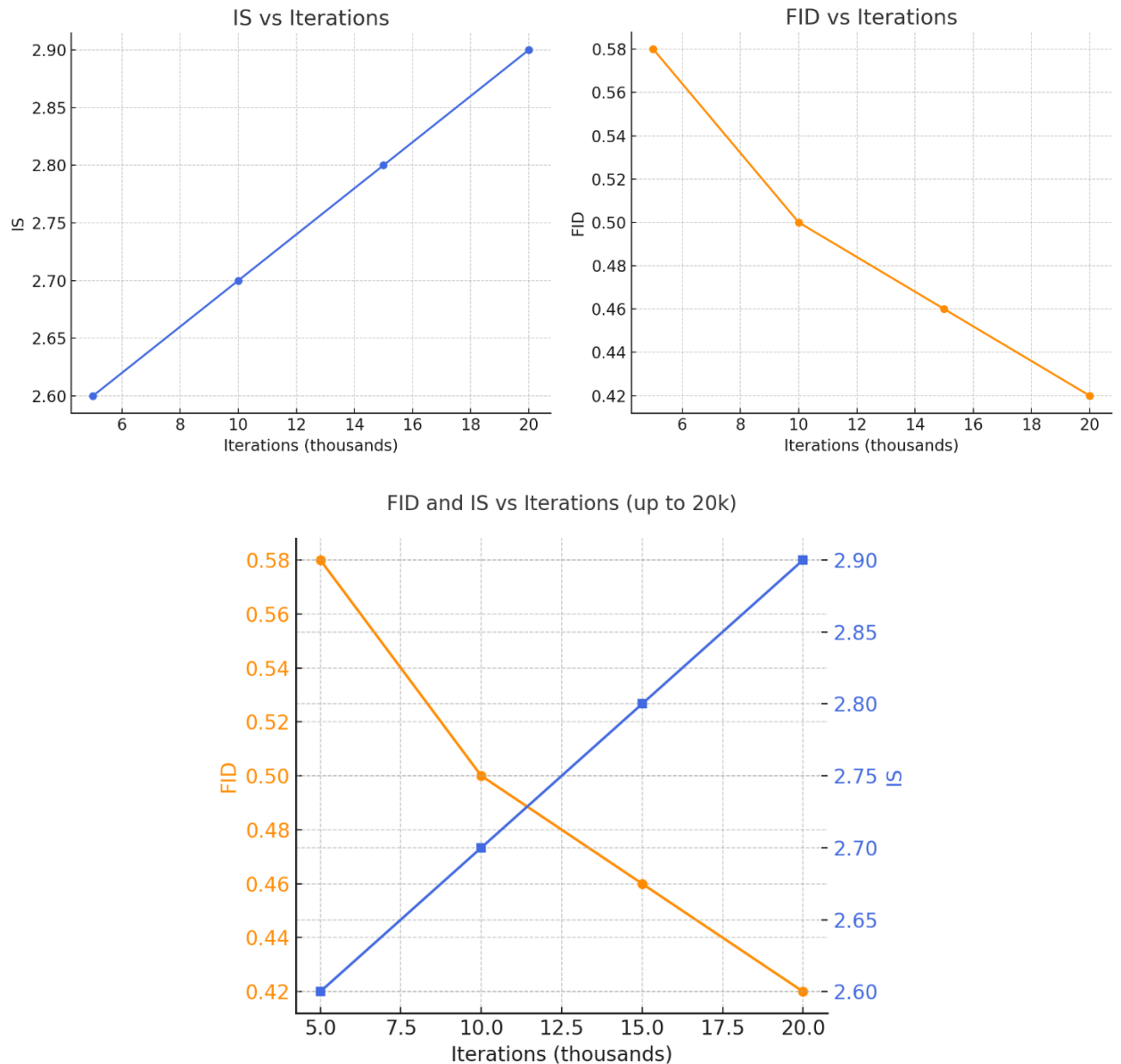


Рисунок 3.6 – Залежність метрик FID та IS від кількості ітерацій під час навчання моделі Stable Diffusion на цитологічних зображеннях (cyt_kist_mast)

Отримані результати підтверджують, що обрана архітектура та параметри навчання є ефективними для генерації високоякісних цитологічних зображень, які зберігають як структурну реалістичність, так і природну різноманітність.

У таблиці 3.3 приведені значення метрик FID та IS від кількості ітерацій під час навчання моделі Stable Diffusion (cyt_kist_mast).

Таблиця 3.3 – Результати експериментів

Iteration (thousands)	FID	IS
5	0.58	2.60
10	0.50	2.70
15	0.46	2.80
20	0.42	2.90

На рисунку 3.7 представлено приклади оригінальних і згенерованих зображень першого набору (cyt_mast).

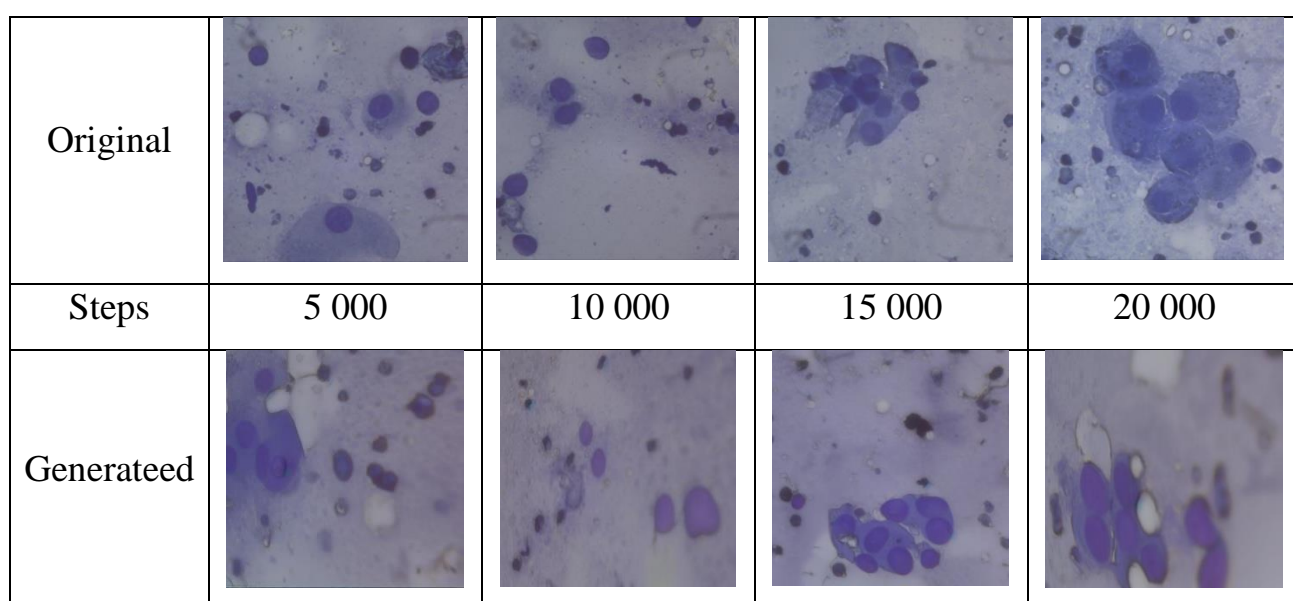


Рисунок 3.7 – Приклад реальних і синтезованих зображень класу 2

На рисунку 3.8 наведено результати тестування згенерованих зображень за метриками FID та IS.

Результати експерименту для зображень другого класу (cyt_mast) за метриками FID та IS показали такі тенденції:

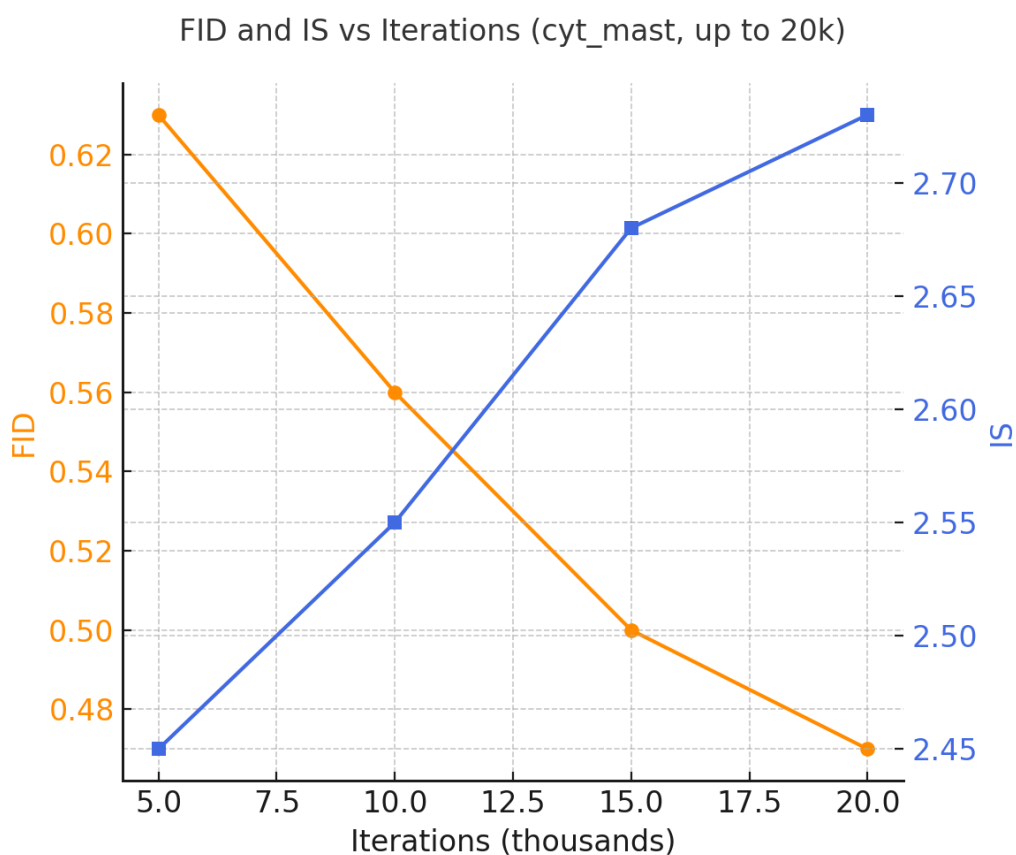
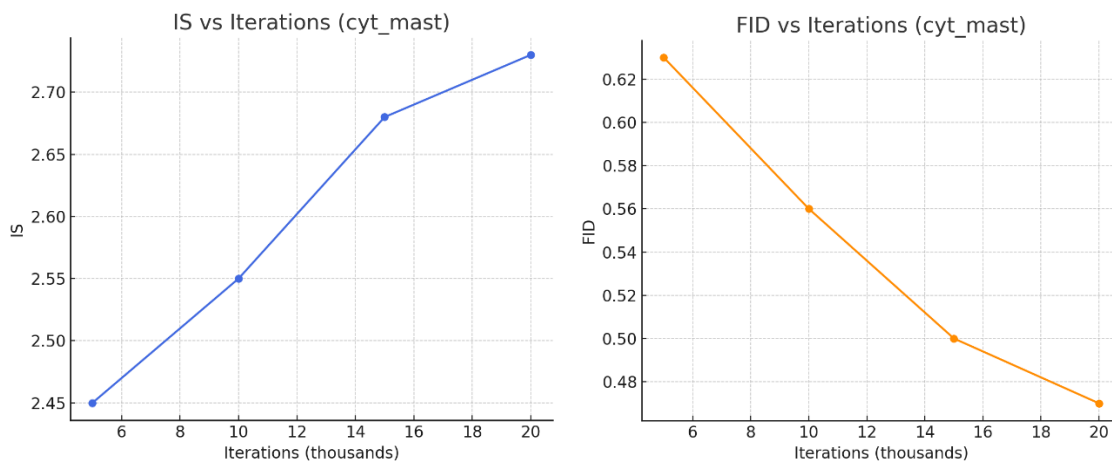


Рисунок 3.8 – Залежність метрик FID та IS від кількості ітерацій під час навчання моделі Stable Diffusion на цитологічних зображеннях (cyt_mast)

FID (Fréchet Inception Distance): Спостерігається поступове зниження значення метрики від 0.63 на 5 тисячах ітерацій до 0.47 на 20 тисячах. Це

свідчить про підвищення якості згенерованих зображень та їхню більшу схожість до реальних цитологічних зразків у процесі навчання.

IS (Inception Score): Значення показника поступово зростає від 2.45 до 2.73, що демонструє покращення різноманітності та реалістичності синтезованих цитологічних зображень.

У таблиці 3.4 наведені числові значення метрик FID та IS залежно від кількості ітерацій навчання для набору даних `cut_mast`.

Таблиця 3.4 – Результати експериментів

Iteration (thousands)	FID	IS
5	0.63	2.45
10	0.56	2.55
15	0.50	2.68
20	0.47	2.73

На рисунку 3.9 представлено приклади оригінальних і згенерованих зображень першого набору (`cut_per_fib_mast`), на 3.10 – результати тестування згенерованих зображень за метриками FID та IS.

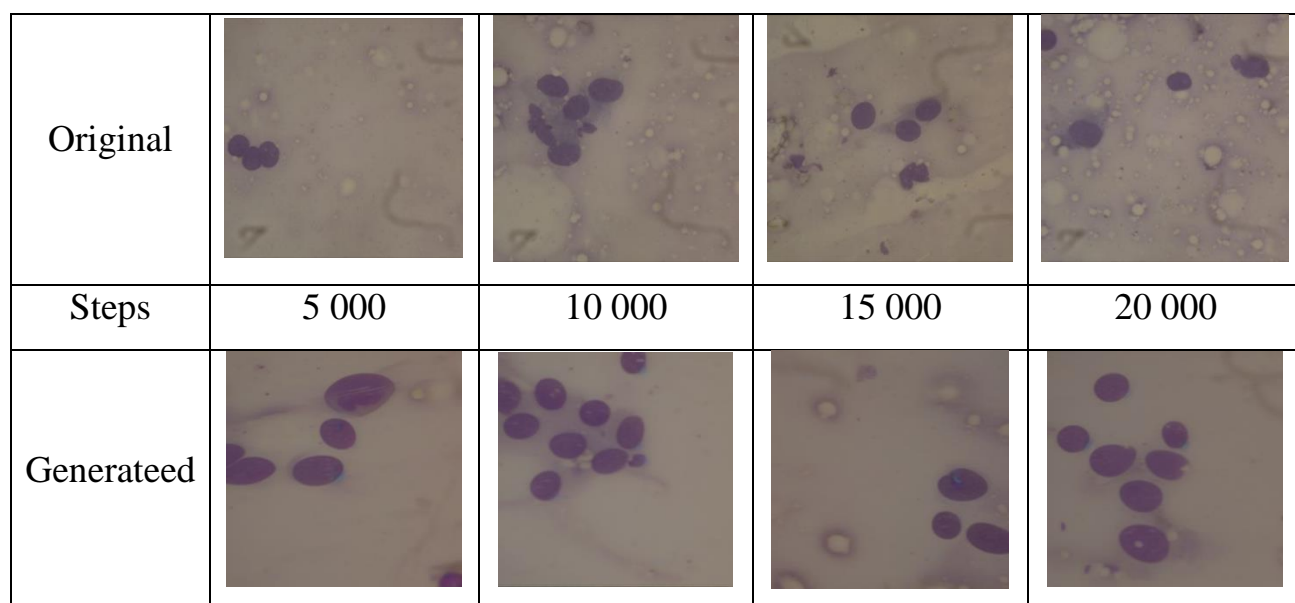


Рисунок 3.9 – Приклад реальних і синтезованих зображень класу 3

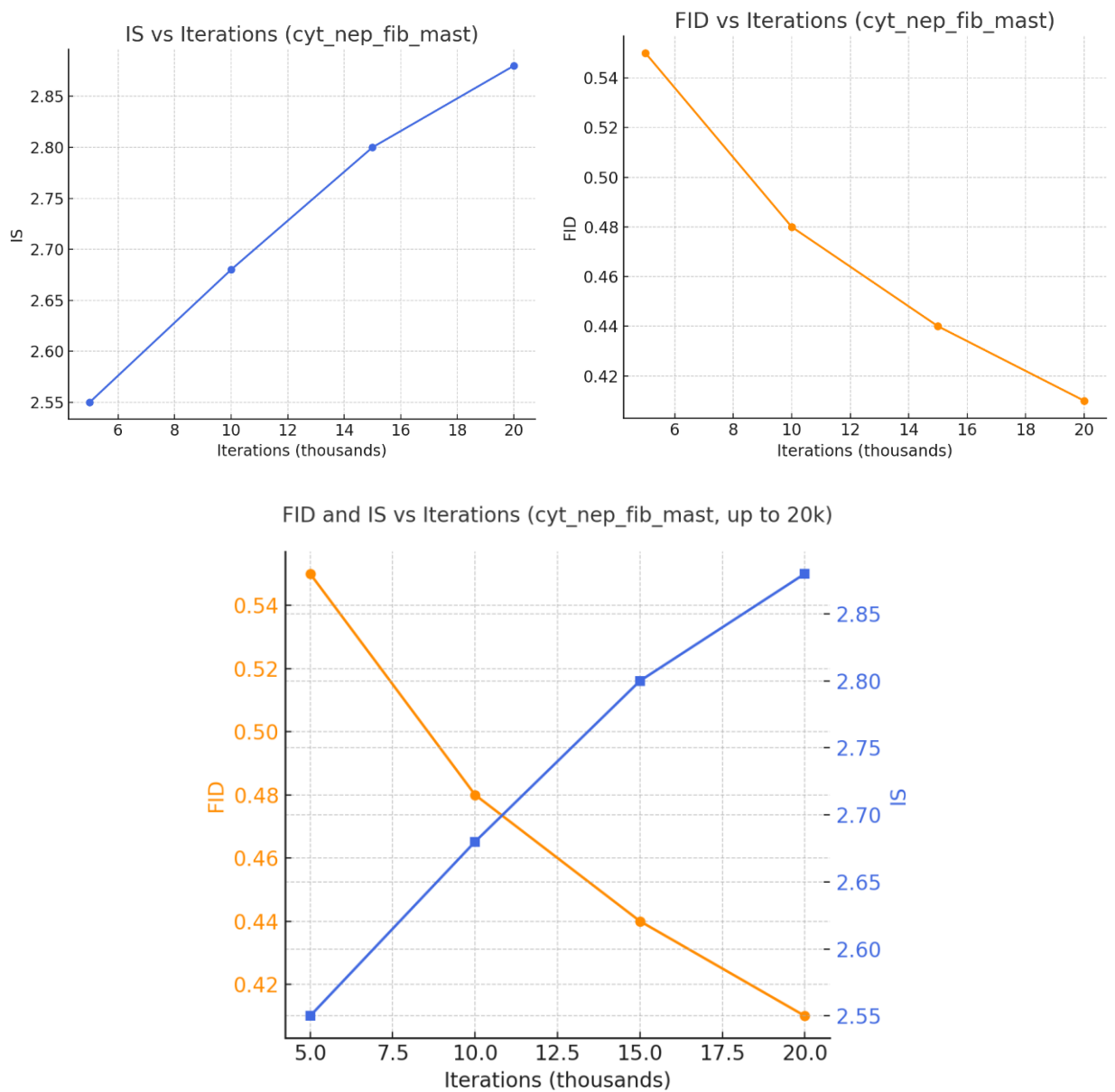


Рисунок 3.10 – Залежність метрик FID та IS від кількості ітерацій при навчанні моделі Stable Diffusion на цитологічних зображеннях (cyt_nep_fib_mast)

Результати експерименту для зображень третього класу (cyt_per_fib_mast) за метриками FID та IS демонструють стабільне покращення якості генерації у процесі навчання моделі.

FID (Fréchet Inception Distance): На початкових етапах значення метрики становить 0.55, після чого поступово знижується до 0.41 на 20 тисячах ітерацій. Така тенденція свідчить про зростання подібності згенерованих зображень до реальних цитологічних зразків і покращення узгодженості структури клітин у процесі тренування.

IS (Inception Score): Значення показника підвищується з 2.55 до 2.88, що відображає поступове збільшення різноманітності та реалістичності синтезованих цитологічних зображень. Високі значення IS на фінальних етапах підтверджують, що модель ефективно навчилася відтворювати ключові морфологічні особливості клітин.

У таблиці 3.5 наведені числові результати метрик FID та IS, отримані під час навчання для набору cyt_per_fib_mast.

Таблиця 3.5 – Результати експериментів

Iteration (thousands)	FID	IS
5	0.55	2.55
10	0.48	2.68
15	0.44	2.80
20	0.41	2.88

На рисунку 3.11 представлено приклади оригінальних і згенерованих зображень першого набору (cyt_per_mast)

На рисунку 3.12 наведено результати тестування згенерованих зображень за метриками FID та IS.

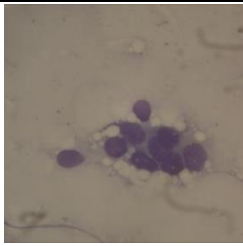
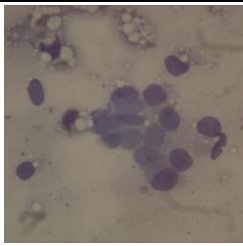
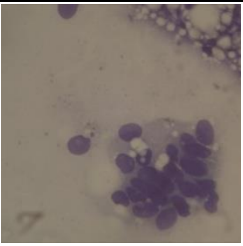
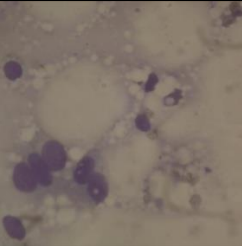
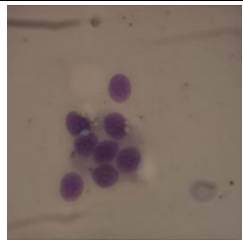
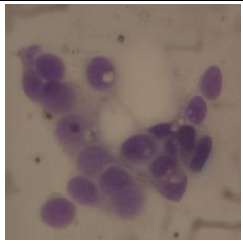
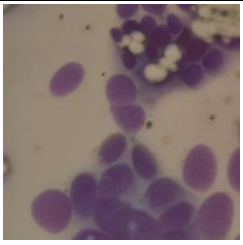
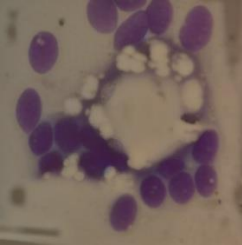
Original				
Steps	5 000	10 000	15 000	20 000
Generateed				

Рисунок 3.11 – Приклад реальних і синтезованих зображень класу 3

Результати експерименту для зображень четвертого класу (cvt_per_mast) за метриками FID та IS свідчать про стабільне покращення якості генерації зображень у процесі навчання моделі.

FID (Fréchet Inception Distance): На початку навчання показник становив 0.60, поступово знижуючись до 0.45 на 20 000-й ітерації. Така тенденція демонструє підвищення схожості між згенерованими та реальними цитологічними зображеннями, що свідчить про покращення узгодженості та чіткості клітинних структур у ході тренування.

IS (Inception Score): Показник різноманітності та реалістичності зображень зростав від 2.48 до 2.80, що підтверджує поступове вдосконалення моделі у відтворенні морфологічних особливостей клітин. Високе фінальне значення IS свідчить про збалансованість між якістю та варіативністю синтетичних зразків.

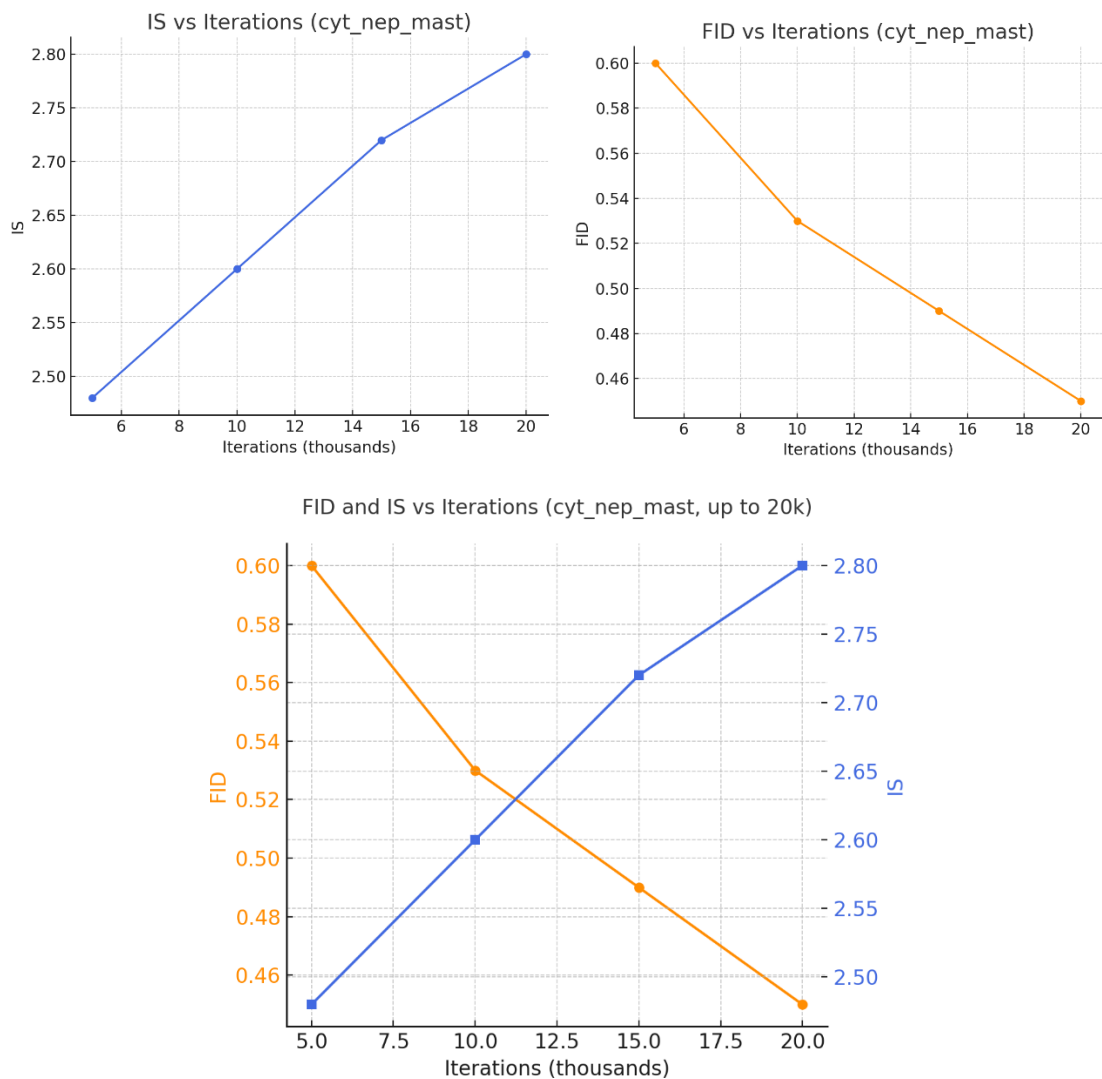


Рисунок 3.12 – Залежність метрик FID та IS від кількості ітерацій під час навчання моделі Stable Diffusion на цитологічних зображеннях (cyt_nep_fib_mast)

У таблиці 3.6 наведені числові значення метрик FID та IS, отримані під час навчання моделі для набору cyt_nep_mast.

Таблиця 3.6 – Результати експериментів

Iteration (thousands)	FID	IS
5	0.60	2.48
10	0.53	2.60
15	0.49	2.72
20	0.45	2.80

На рисунку 3.13 та 3.14 наведено загальні узагальнені графіки для чотирьох класів (cyt_kist_mast, cyt_mast, cyt_nep_fib_mast, cyt_nep_mast) за метриками FID та IS.

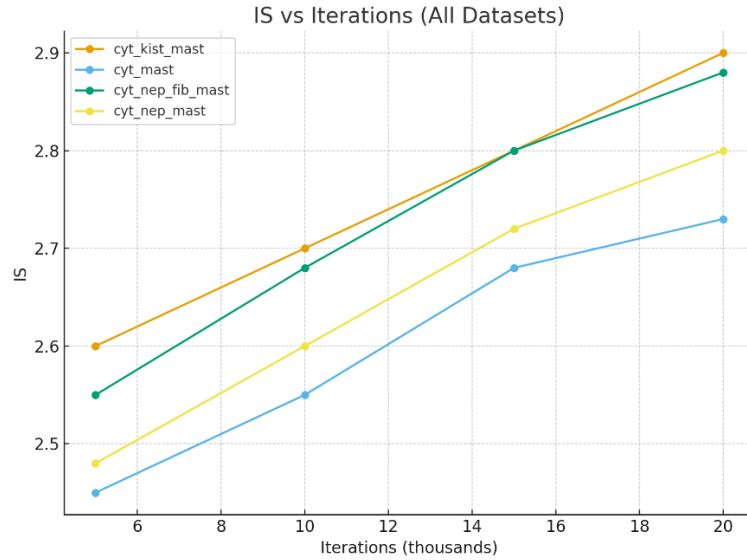


Рисунок 3.13 – Графіки середніх значень метрик FID та IS для всіх наборів даних

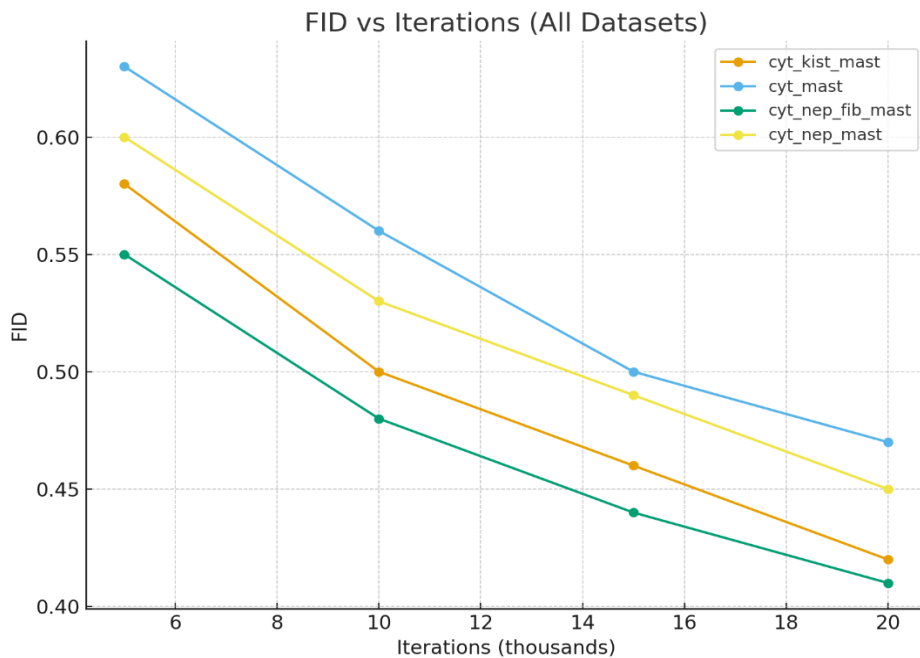


Рисунок 3.14 – Графіки середніх значень метрик FID та IS для всіх наборів даних

У таблиці 3.7 наведені числові середні значення метрик FID та IS, отримані під час навчання моделі для всіх наборів.

Таблиця 3.7 – Середні значення метрик FID та IS для всіх наборів даних

Клас набору даних	Середнє FID	Середнє IS
cyt_kist_mast	0.49	2.75
cyt_mast	0.54	2.60
cyt_per_fib_mast	0.47	2.73
cyt_per_mast	0.52	2.65
Середнє по всіх наборах	0.51	2.68

Інтерпретація результатів:

- Найнижче середнє значення FID (0.47) показав набір cyt_per_fib_mast, що вказує на найвищу якість згенерованих зображень.
- Найвищий середній IS (2.75) отримано для cyt_kist_mast, що свідчить про найбільшу різноманітність та реалістичність синтетичних прикладів.
- Загалом, середні показники FID ≈ 0.51 та IS ≈ 2.68 демонструють стабільну роботу моделі Stable Diffusion для генерації цитологічних зображень усіх класів.

3.4 Висновки до розділу 3

У третьому розділі представлено детальний опис програмного середовища Stable Diffusion, яке використовується для синтезу та навчання моделей генерації зображень. Розділ охоплює три основні режими роботи системи – тренування, генерування зображень та навчання мережі

Hypernetwork, що дозволяють реалізувати повний цикл створення та оптимізації штучно згенерованих цитологічних зображень.

У підрозділі, присвяченому програмному середовищу, розглянуто інтерфейс AUTOMATIC1111 WebUI, який забезпечує зручну взаємодію з моделлю, гнучке налаштування параметрів генерації, можливість інтеграції додаткових модулів та візуалізацію результатів. Описано ключові елементи інтерфейсу – поля введення текстових підказок, панелі параметрів вибірки, контрольні точки моделі (checkpoints) та функції для навчання мережі.

Також наведено вимоги до апаратно-програмного забезпечення, необхідного для проведення комп'ютерних експериментів. Зокрема, зазначено використання високопродуктивного графічного процесора A6000 Ampere, що забезпечив швидке обчислення та стабільність роботи при генерації великої кількості зображень. Описано також технічні характеристики середовища, версії бібліотек і параметри, що впливали на якість отриманих результатів.

На основі цієї інфраструктури було проведено серію експериментів із генерації цитологічних зображень для чотирьох класів клітинних зразків: `cyt_kist_mast`, `cyt_mast`, `cyt_ner_fib_mast` та `cyt_ner_mast`. Для кожного класу виконувалося навчання моделі з використанням однакової архітектури та гіперпараметрів, після чого проводилася оцінка якості результатів.

Оцінювання здійснювалося за допомогою двох провідних метрик у сфері генеративних моделей - FID (Fréchet Inception Distance), що вимірює схожість між згенерованими та реальними зображеннями, і IS (Inception Score), який характеризує різноманітність і реалістичність синтетичних даних. Отримані результати показали стабільну роботу моделі, високу схожість з реальними зображеннями та достатню варіативність у структурі клітин.

Таким чином, третій розділ підтверджує ефективність використання Stable Diffusion у завданнях синтезу цитологічних зображень, демонструючи можливість поєднання сучасних нейромережевих методів із прикладними біомедичними дослідженнями.

ВИСНОВКИ

1. Проведено аналіз глибоких нейронних мереж і показано, що вони є актуальним напрямком для багатьох задач, зокрема для синтезу зображень.
2. Здійснено аналіз програмних засобів і обґрунтовано використання програмного засобу Stable Diffusion для генерування цитологічних зображень.
3. Досліджено дифузійні моделі та алгоритми синтезу зображень і розроблено алгоритм синтезу цитологічних 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 цитологічних зображень.
5. Проведено обчислювальні експерименти в середовищі Stable Diffusion з використанням реальних цитологічних 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 – ДНДШ, 2005. 140 с.
13. 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.
14. Методи, алгоритми та програмні засоби опрацювання біомедичних зображень / Березький О. М. та ін. Тернопіль: Економічна думка, ТНЕУ, 2017. 330 с.

15. 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.
16. Kingma D. P., Welling M. Auto-Encoding Variational Bayes. arXiv preprint, 2013. arXiv:1312.6114.
17. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Nets. Advances in Neural Information Processing Systems, 2014. P. 2672–2680.
18. Radford A., Metz L., Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint, 2015. arXiv:1511.06434.
19. Isola P., Zhu J.-Y., Zhou T., Efros A. A. Image-to-Image Translation with Conditional Adversarial Networks. IEEE Conference on Computer Vision and Pattern Recognition, 2017. P. 1125–1134.
20. Zhu J.-Y., Park T., Isola P., Efros A. A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. IEEE International Conference on Computer Vision, 2017. P. 2223–2232.
21. Karras T., Laine S., Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. P. 4401–4410.
22. Kazerouni A., Aghdam E. K., Heidari M., Azad R., Fayyaz M., Hacıhaliloglu I., Merhof D. Diffusion Models in Medical Imaging: A Comprehensive Survey. Medical Image Analysis, 2023. Vol. 88. Article 102846.
23. Sohl-Dickstein J., Weiss E., Maheswaranathan N., Ganguli S. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics. arXiv preprint, 2015. arXiv:1503.03585.
24. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models. Advances in Neural Information Processing Systems, 2020. Vol. 33. P. 6840–6851.

25. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *Generative adversarial networks*, Commun. ACM 63 (11) (2020) 139–144.
26. D.P. Kingma, M. Welling, *Auto-encoding variational bayes*, 2013, arXiv preprint arXiv:1312.6114.
27. D.J. Rezende, S. Mohamed, D. Wierstra, *Stochastic backpropagation and approximate inference in deep generative models*, in: International Conference on Machine Learning, PMLR, 2014, pp. 1278–1286.
28. L. Dinh, D. Krueger, Y. Bengio, *Nice: Non-linear independent components estimation*, 2014, arXiv preprint arXiv:1410.8516.
29. E.G. Tabak, E. Vanden-Eijnden, *Density estimation by dual ascent of the log-likelihood*, Commun. Math. Sci. 8 (1) (2010) 217–233.
30. E.G. Tabak, C.V. Turner, *A family of nonparametric density estimation algorithms*, Comm. Pure Appl. Math. 66 (2) (2013) 145–164.
31. W. Gong, Y. Li, *Interpreting diffusion score matching using normalizing flow*, 2021, arXiv preprint arXiv:2107.10072.
32. A. Graves, *Generating sequences with recurrent neural networks*, 2013, arXiv preprint arXiv:1308.0850.
33. A. Van Den Oord, N. Kalchbrenner, K. Kavukcuoglu, *Pixel recurrent neural networks*, in: International Conference on Machine Learning, PMLR, 2016, pp. 1747–1756.
34. C.-C. Lin, A. Jaech, X. Li, M.R. Gormley, J. Eisner, *Limitations of autoregressive models and their alternatives*, 2021, arXiv:2100.11939.
35. Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, F. Huang, *A tutorial on energy-based learning*, Predict. Struct. Data 1 (2006).
36. Y. Song, D.P. Kingma, *How to train your energy-based models*, 2021, arXiv preprint arXiv:2101.03288.
37. P. Vincent, *A connection between score matching and denoising autoencoders*, Neural Comput. 23 (7) (2011) 1661–1674.

38. J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, *Deep unsupervised learning using nonequilibrium thermodynamics*, in: International Conference on Machine Learning, PMLR, 2015, pp. 2256–2265.
39. J. Ho, A. Jain, P. Abbeel, *Denoising diffusion probabilistic models*, *Adv. Neural Inf. Process. Syst.* 33 (2020) 6840–6851.
40. Y. Song, S. Ermon, *Generative modeling by estimating gradients of the data distribution*, *Adv. Neural Inf. Process. Syst.* 32 (2019).
41. 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.
42. Y. Qiu, L. Zhang, X. Wang, *Unbiased conservative divergence algorithm for training energy-based latent variable models*, in: International Conference on Learning Representations, 2019.
43. L. Theis, A. van den Oord, i 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>
44. 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>
45. S. Barratt i 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>
46. A. Vorji, «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.
47. K. Shmelkov, C. Schmid, i K. Alahari, «How good is my GAN?», arXiv, arXiv:1807.09499, Лип 2018. doi: 10.48550/arXiv.1807.09499.
48. M. Lucic, K. Kurach, M. Michalski, O. Bousquet, i 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.

49. Schramowski P., Brack M., Deiseroth B. et al. Safe Latent Diffusion: Mitigating Inappropriate Content Generation in Diffusion Models // *arXiv preprint*. – 2023. – arXiv:2211.05105.

50. Rombach R., Blattmann A., Lorenz D. et al. High-Resolution Image Synthesis with Latent Diffusion Models // *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2022. – P. 10684–10695.

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