

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ центр післядипломної освіти _____
(повна назва)

Кафедра _____ програмної інженерії _____
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти _____ другий (магістерський) _____

_____ Дослідження методів генерації 3D зображень на основі 2D зображень _____
(тема)

Виконав:
здобувач _____ 2 _____ року навчання
групи _____ ПЗЗдм-23-1 _____

_____ Ліана ПАРШИКОВА _____
(Власне ім'я, ПРІЗВИЩЕ)

Спеціальність _____ 121 – Інженерія програмного
забезпечення _____
(код і повна назва спеціальності)

Тип програми _____ освітньо-наукова _____
(повна назва освітньої програми)

Керівник _____ доц. Дмитро КОЛЕСНИКОВ _____
(посада, власне ім'я, прізвище)

Допускається до захисту
Зав. кафедри

_____ (підпис)

_____ Кирило СМЕЛЯКОВ _____
(власне ім'я, прізвище)

2025 р.

Харківський національний університет радіоелектроніки

Факультет _____ центр післядипломної освіти _____
 Кафедра _____ програмної інженерії _____
 Рівень вищої освіти _____ другий (магістерський) _____
 Спеціальність _____ 121 – Інженерія програмного забезпечення _____
 Тип програми _____ освітньо-наукова програма _____
 Освітня програма _____ Інженерія програмного забезпечення _____
 (шифр і назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
 (підпис)
 « ____ » _____ 2025 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _____ Паршиковій Ліані В'ячеславівні _____
 (прізвище, ім'я, по батькові)

1. Тема роботи _____ Дослідження методів генерації 3D зображень на основі 2D зображень _____

Затверджена наказом по університету від _____ 21.04. 2025р. № 61Стз _____

2. Термін подання студентом роботи до екзаменаційної комісії _____ 09.06.2025 _____

3. Вихідні дані до роботи _____ Основні поняття комп'ютерного зору, глибини зображення, просторової реконструкції; методи машинного навчання для відновлення 3D структури на основі плоских зображень; сучасні підходи: глибокі нейронні мережі, трансформери, генеративно-змагальні мережі (GAN);

Наявні фреймворки та інструменти: PyTorch, TensorFlow, OpenCV, Blender, Meshroom тощо; Набори даних: ShapeNet, Pix3D, KITTI, CO3D, Pascal3D+.

4. Перелік питань, що потрібно опрацювати в роботі _____ Теоретичні основи генерації 3D зображень; огляд існуючих методів побудови 3D моделей з 2D зображень; аналіз сучасних моделей машинного навчання для 3D реконструкції; огляд та аналіз інструментів і програмного забезпечення; Аналіз і підготовка наборів даних; підсумковий аналіз, висновки та перспективи подальших досліджень.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	22.04.2025	<i>виконано</i>
2	Аналіз предметної галузі і постановка задачі	22.0-28.04.2025	<i>виконано</i>
3	Огляд та аналіз методів генерації 3d зображень	29.04-05.05.2025	<i>виконано</i>
4	Розробка та реалізація експериментальної моделі	06.05-12.05.2025	<i>виконано</i>
5	Підготовка до апробації результатів дослідження. Публікація матеріалів	13.05-15.05.2025	<i>виконано</i>
6	Експериментальні дослідження та аналіз результатів	16.05-21.05.2025	<i>виконано</i>
7	Підготовка пояснювальної записки	22.05-27.05.2025	<i>виконано</i>
8	Підготовка презентації та доповіді	28.05-29.05.2025	<i>виконано</i>
9	Перевірка на плагіат	30.05.2025	<i>виконано</i>
10	Нормоконтроль	01.06.2025	<i>виконано</i>
11	Рецензування	02.06.2025	<i>виконано</i>
12	Попередній захист	03.06.2025	<i>виконано</i>
13	Занесення диплома в електронний архів	03.06.2025	<i>виконано</i>
14	Допуск до захисту у зав. кафедри	04.06.2025	<i>виконано</i>

Дата видачі завдання 22 квітня 2025р.

Здобувачка _____
(підпис)

Ліана ПАРШИКОВА

Керівник роботи _____
(підпис)

доц. Дмитро КОЛЕСНИКОВ
(посада, Власне ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ / ABSTRACT

Пояснювальна записка до кваліфікаційної роботи: 72 сторінки, 11 рисунків, 5 таблиць, 16 джерел, 4 додатки.

3D РЕКОНСТРУКЦІЯ, 2D ЗОБРАЖЕННЯ, ГЛИБИННЕ НАВЧАННЯ, CNN, ГЕНЕРАЦІЯ МОДЕЛЕЙ, ВОКСЕЛІ, ShapeNet, НЕЙРОННІ МЕРЕЖІ, NeRF.

Об'єкт дослідження – процес генерації тривимірних зображень (3D моделей) на основі двовимірних зображень (2D входів).

Предмет дослідження – методи та моделі реконструкції 3D форми із використанням традиційних алгоритмів комп'ютерного зору та технологій глибокого навчання.

Мета роботи – дослідити, класифікувати та порівняти сучасні методи генерації 3D зображень на основі 2D зображень, а також реалізувати експериментальну модель на основі згорткової нейронної мережі для реконструкції 3D форми об'єкта.

У ході дослідження:

- виконано огляд та аналіз традиційних геометричних методів (SfM, SfS, стереозір), нейромережових архітектур (CNN, GAN, NeRF, DeepSDF);
- реалізовано прототип моделі, що перетворює 2D зображення у 3D воксельні представлення на основі архітектури encoder–decoder;
- проведено навчання моделі на датасеті ShapeNet, візуалізацію результатів, оцінку за метриками IoU, Chamfer Distance;
- здійснено порівняльний аналіз з іншими підходами.

Результати роботи можуть бути використані в системах доповненої та віртуальної реальності, автоматизованому дизайні, геймдеві, 3D-друці, а також як база для подальших досліджень у сфері 3D реконструкції та комп'ютерного зору.

Explanatory note to the qualification work: 72 pages, 11 figures, 5 tables, 16 sources, 4 appendices.

3D RECONSTRUCTION, 2D IMAGE, DEEP LEARNING, CNN, MODEL GENERATION, VOXELS, ShapeNet, NEURAL NETWORKS, NeRF.

The object of research is the process of generating three-dimensional images (3D models) based on two-dimensional images (2D inputs).

The subject of research is methods and models of 3D shape reconstruction using traditional computer vision algorithms and deep learning technologies.

The purpose of the work is to investigate, classify and compare modern methods of generating 3D images based on 2D images, as well as to implement an experimental model based on a convolutional neural network for reconstructing the 3D shape of an object.

During the research:

- a review and analysis of traditional geometric methods (SfM, SfS, stereo vision), neural network architectures (CNN, GAN, NeRF, DeepSDF) was performed;
- a prototype of a model that converts 2D images into 3D voxel representations based on the encoder–decoder architecture was implemented;
- the model was trained on the ShapeNet dataset, results were visualized, and evaluated using IoU and Chamfer Distance metrics;
- a comparative analysis with other approaches was performed.

The results of the work can be used in augmented and virtual reality systems, automated design, game development, 3D printing, as well as as a basis for further research in the field of 3D reconstruction and computer vision.

Завідувачу кафедри

ПІ

(скорочена назва кафедри)

проф. Кирилу СМЕЛЯКОВУ

(вчене звання, власне ім'я, прізвище)

ЗАЯВА

щодо самостійного виконання кваліфікаційної роботи та можливості її публікації в електронному архіві відкритого доступу EIArKhNURE.

Я, Паршикова Ліана В'ячеславівна, здобувачка гр. ПЗЗдм-23-1, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Дослідження методів генерації 3D зображень на основі 2D зображень», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений(на) з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

Дата

Підпис

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ.....	9
ВСТУП.....	10
1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ.....	12
1.1 Аналіз предметної галузі дослідження	12
1.2 Основні поняття предметної області.....	13
1.3 Проблематика перетворення 2D у 3D	14
1.4 Методи генерації 3D зображень на основі 2D	14
1.4.1 Традиційні методи комп'ютерного зору	14
1.4.2 Методи на основі глибинного навчання	14
1.5 Програмні засоби та інструменти.....	15
1.6 Сучасні дослідження та тенденції	16
2 ОГЛЯД ТА АНАЛІЗ МЕТОДІВ ГЕНЕРАЦІЇ 3D ЗОБРАЖЕНЬ	18
2.1 Огляд існуючих підходів	18
2.2 Згорткова нейромережа	19
2.3 Еволюція алгоритмів у створенні зображень	20
2.4 Text to image	21
2.5 Еволюція нейромереж.....	21
2.6 Геометричні методи: SfM, SfS, стереозір	24
2.7 Застосування згорткових нейронних мереж (CNN)	24
2.8 Генеративні моделі (GAN, VAE) у 3D реконструкції	25
2.9 Імпліцитні моделі: NeRF, DeepSDF	25
2.10 Приклади алгоритмів та методів	27
2.11 Оцінка попередніх рішень та поточні виклики	36
2.12 Порівняльний аналіз підходів.....	36
3 РОЗРОБКА ТА РЕАЛІЗАЦІЯ ЕКСПЕРИМЕНТАЛЬНОЇ МОДЕЛІ	39
3.1 Вибір архітектури та середовища розробки	39
3.2 Підготовка та обробка вхідних даних	41
3.3 Побудова моделі та налаштування параметрів.....	42
3.4 Навчання нейронної мережі.....	43
3.5 Візуалізація результатів генерації 3D моделей.....	43

3.6 Тестування моделі та оцінка якості реконструкції.....	44
4 ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ.....	46
4.1 Метрики оцінювання якості (IoU, Chamfer Distance тощо)	46
4.2 Порівняння результатів з іншими підходами	47
4.3 Аналіз похибок і впливу параметрів	48
4.4 Практичні висновки щодо ефективності реалізованої моделі.....	49
ВИСНОВКИ	50
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	52
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ ЗА НАУКОВИМИ НАПРЯМАМИ КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ.....	54
ДОДАТОК А Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ....	55
ДОДАТОК Б Слайди презентації.....	57
ДОДАТОК В Апробація результатів роботи.....	66
ДОДАТОК Г Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015.....	72

ПЕРЕЛІК СКОРОЧЕНЬ

- AI – Artificial Intelligence – штучний інтелект
- CNN – Convolutional Neural Network – згорткова нейронна мережа
- DNN – Deep Neural Network – глибока нейронна мережа
- GAN – Generative Adversarial Network – генеративно-змагальна мережа
- ML – Machine Learning – машинне навчання
- DL – Deep Learning – глибинне навчання
- 3D – Three-Dimensional – тривимірний
- 2D – Two-Dimensional – двовимірний
- NeRF – Neural Radiance Fields – нейронні радіаційні поля
- SfM – Structure from Motion – структура з руху
- SfS – Shape from Shading – форма з тіней
- IoU – Intersection over Union – коефіцієнт перетину
- CD – Chamfer Distance – відстань Шамфера
- PSNR – Peak Signal-to-Noise Ratio – пікове відношення сигналу до шуму
- RGB – Red Green Blue – система представлення кольору
- CAD – Computer-Aided Design – комп’ютерне проектування
- GPU – Graphics Processing Unit – графічний процесор
- API – Application Programming Interface – програмний інтерфейс застосунків
- OBJ, STL, PLY – формати збереження 3D моделей

ВСТУП

У сучасному цифровому світі обробка та інтерпретація візуальної інформації відіграє ключову роль у багатьох сферах – від комп'ютерного зору та медичної візуалізації до ігрової індустрії, доповненої реальності та автономного транспорту. Однією з актуальних задач є реконструкція тривимірної структури сцени або об'єкта на основі двовимірних зображень. Такий процес дозволяє відновити просторову інформацію, що зазвичай втрачається під час проєкції реального світу на площину зображення.

Традиційно генерація 3D зображень вимагала спеціалізованого обладнання – стереокамер, сенсорів глибини або лазерних сканерів. Проте в останні роки активно розвиваються методи, що дозволяють відновлювати тривимірні об'єкти лише на основі 2D-зображень. Зокрема, впровадження глибоких нейронних мереж, генеративно-змагальних моделей та нейронних полів випромінювання (Neural Radiance Fields) дало змогу суттєво підвищити якість відновлення та адаптивність до різних умов.

Завдяки появі потужних обчислювальних ресурсів, відкритих наборів даних та фреймворків для глибинного навчання стало можливим створення моделей, здатних реконструювати повноцінні 3D об'єкти на основі однієї або кількох фотографій. Це відкриває нові перспективи для автоматизації процесів, зменшення витрат на створення 3D-контенту та розширення можливостей у суміжних технологіях.

Актуальність обраної теми зумовлена швидким розвитком штучного інтелекту, необхідністю удосконалення існуючих методів просторової реконструкції та попитом на ефективні інструменти генерації 3D зображень у багатьох прикладних галузях.

Мета кваліфікаційної роботи полягає в дослідженні, порівнянні та практичній реалізації сучасних методів генерації 3D зображень на основі 2D даних з використанням технологій машинного навчання.

Для досягнення цієї мети у роботі поставлено наступні задачі:

- проаналізувати наукові джерела та існуючі підходи до 3D реконструкції;
- розглянути та класифікувати сучасні методи нейронного відновлення 3D структури;
- вибрати відповідний інструментарій та реалізувати експериментальну модель;
- провести оцінку ефективності обраного методу за допомогою відповідних метрик;
- зробити висновки щодо доцільності використання методів у прикладних задачах.

Об'єктом дослідження є процес генерації тривимірних зображень на основі двовимірних вхідних даних.

Предметом дослідження – алгоритмічні та нейронні методи реконструкції 3D форм.

Робота має теоретичну та практичну цінність, оскільки дозволяє узагальнити сучасні підходи до 3D реконструкції, а також реалізувати власну модель, що може бути використана в реальних застосунках.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Аналіз предметної галузі дослідження

У сучасних інформаційних технологіях візуалізація просторових об'єктів відіграє ключову роль. Зокрема, здатність реконструювати тривимірну (3D) структуру об'єкта на основі одного або кількох двовимірних (2D) зображень відкриває широкі можливості для аналізу, моделювання та взаємодії з цифровим контентом. Предметна галузь, що охоплює ці процеси, включає знання з комп'ютерного зору, графіки, глибинного навчання та обробки зображень.

Генерація 3D зображень на основі 2D входів належить до класу задач "оберненого проєктування" (inverse rendering), де ціль полягає у відновленні просторової інформації, втраченої під час проєкції сцени на площину. Основна складність полягає в тому, що під час знімання зображення втрачається інформація про глибину та рельєф, яку необхідно відновити аналітично або за допомогою навчання моделей.

Завдання 3D реконструкції є критично важливим у таких галузях:

- доповнена та віртуальна реальність (AR/VR) – створення інтерактивного середовища на основі зображень реального світу;
- медицина – побудова 3D моделей внутрішніх органів на основі знімків (наприклад, КТ, МРТ);
- автономні системи – розпізнавання об'єктів і навігація за допомогою камер;
- культурна спадщина – цифрова реконструкція археологічних артефактів за архівними фотографіями;
- комп'ютерна графіка та геймдев – створення 3D моделей за фотографіями без ручного моделювання.

У досліджуваній предметній області активно розвиваються два напрямки:

- а) Традиційні підходи (на основі геометричних алгоритмів), які використовують багатокамерні системи або аналіз освітлення для відновлення 3D структури (наприклад, stereo vision, structure-from-motion, shape-from-shading).

б) Методи на основі глибокого навчання, які дають змогу моделювати зв'язок між 2D зображеннями та їх 3D представленням без потреби в явному геометричному аналізі. До таких належать згорткові нейронні мережі, генеративні моделі (GAN, VAE), трансформери та моделі з неявним представленням (наприклад, Neural Radiance Fields – NeRF).

З огляду на постійне зростання обчислювальних можливостей, доступ до великих наборів даних та активне дослідження в сфері штучного інтелекту, ця галузь є надзвичайно перспективною та актуальною. Саме тому вивчення, класифікація та тестування сучасних методів генерації 3D зображень з 2D входів становить важливу наукову та практичну задачу.

1.2 Основні поняття предметної області

Для розуміння сутності задачі варто визначити ключові терміни:

– 2D зображення – цифрове представлення сцени у площині, що містить інформацію лише про координати X та Y (висоту й ширину);

– 3D зображення (модель) – цифрове представлення об'єкта у тривимірному просторі з координатами X, Y, Z;

– глибина (depth) – відстань від камери до об'єкта, що дозволяє визначити його положення в просторі;

– реконструкція – процес відновлення геометрії об'єкта на основі його проєкції.

Сучасні методи генерації тривимірних (3D) зображень на основі двовимірних (2D) зображень є однією з найактуальніших тем у галузі комп'ютерного зору, машинного навчання та графіки. З розвитком технологій зростає потреба в реалістичних та точних 3D моделях, що можуть бути використані у різних галузях, таких як медицина, архітектура, розваги та багато іншого. Огляд літератури зосереджується на дослідженні існуючих методів, їх обмежень, тенденцій та перспектив.

1.3 Проблематика перетворення 2D у 3D

Основна складність перетворення полягає в тому, що проекція тривимірної сцени на площину втрачає інформацію про глибину. Це створює ситуацію неоднозначності: кілька різних 3D об'єктів можуть мати однакове 2D представлення. Тому для якісної реконструкції потрібні або додаткові зображення з різних ракурсів, або навчена модель, здатна "здогадатись" про відсутню інформацію.

1.4 Методи генерації 3D зображень на основі 2D

У предметній галузі виділяють кілька основних підходів до генерації 3D моделей.

1.4.1 Традиційні методи комп'ютерного зору

Розглянемо традиційні методи комп'ютерного зору:

– Stereo Vision – обчислення глибини шляхом порівняння пар зображень, отриманих з різних точок огляду.

– Structure from Motion (SfM) – побудова структури сцени шляхом аналізу зміни положення камери.

– Shape from Shading (SfS) – реконструкція форми об'єкта на основі аналізу світла та тіней на зображенні.

Ці методи потребують або додаткових сенсорів, або багатьох зображень одного об'єкта.

1.4.2 Методи на основі глибинного навчання

Розглянемо наступні методи на основі глибинного навчання:

– CNN (Convolutional Neural Networks) – використовуються для передбачення карти глибини на основі одного зображення.

– GAN (Generative Adversarial Networks) – здатні генерувати фотореалістичні 3D моделі з одного 2D входу.

– Neural Radiance Fields (NeRF) – сучасний підхід до побудови неперервного представлення об'єктів у просторі.

Такі методи не потребують стереопар і демонструють високу адаптивність до складних сцен.

Таблиця 1.1 – Класифікація методів генерації 3D зображень з 2D

Категорія підходу	Метод	Короткий опис	Переваги	Недоліки
Геометричні	Stereo Vision	Реконструкція на основі двох камер або ракурсів	Простота реалізації	Потребує кількох зображень
Геометричні	Structure from Motion (SfM)	Реконструкція з серії зображень	Висока точність	Залежить від руху камери
Геометричні	Shape from Shading (SfS)	Відновлення глибини за освітленням	Не потребує багато зображень	Нестабільність при складному світлі
Нейромережеві	CNN	Згорткові мережі для карти глибини	Автоматизація процесу	Потреба в великому наборі даних
Нейромережеві	GAN	Генерація 3D через змагання генератора та дискримінатора	Реалістичність моделей	Складність навчання
Нейромережеві (імпліцитні)	NeRF	Безперервне 3D представлення через вивчення сцени	Висока якість, фотореалізм	Висока обчислювальна складність

1.5 Програмні засоби та інструменти

Для реалізації задач реконструкції 3D зображень використовуються такі популярні інструменти:

– PyTorch, TensorFlow – фреймворки для побудови й навчання нейронних мереж.

– OpenCV, Open3D – бібліотеки для обробки зображень і 3D моделей.

– Blender, MeshLab – засоби для візуалізації, обробки та рендерингу 3D моделей.

1.6 Сучасні дослідження та тенденції

У провідних дослідницьких інститутах (Google Research, Meta AI, NVIDIA, Stanford, MIT) постійно з’являються нові підходи до генерації 3D сцен. Такі методи як Pix2Vox, AtlasNet, DeepSDF, Mip-NeRF демонструють значне покращення якості, особливо в умовах обмежених вхідних даних.

Актуальні напрями досліджень:

- поєднання різних типів вхідних даних (зображення, глибина, текстура);
- використання трансформерів для просторового аналізу;
- оптимізація ресурсів при генерації 3D моделей в реальному часі;
- мінімізація навчальних даних завдяки few-shot learning.

1.7 Постановка задачі

У межах даної магістерської роботи розглядається задача реконструкції тривимірної структури об’єктів або сцен на основі двовимірних зображень. Ця задача має високу складність через втрату просторової інформації під час проєкції 3D сцени на 2D площину, а також неоднозначність можливої геометричної інтерпретації.

Загальна мета роботи полягає у дослідженні, порівнянні та практичній реалізації сучасних методів генерації 3D зображень з 2D входів із використанням підходів машинного навчання, зокрема глибокого навчання.

Для досягнення цієї мети необхідно вирішити такі проміжні задачі:

а) Проаналізувати наукову літературу та існуючі підходи:

- 1) вивчити класифікацію методів 3D реконструкції;
 - 2) оцінити ефективність традиційних алгоритмів та сучасних нейромережових рішень.
- б) Обрати та дослідити актуальні моделі глибокого навчання:
- 1) розглянути архітектури CNN, GAN, NeRF;
 - 2) визначити переваги та обмеження кожного підходу.
- в) Сформулювати вимоги до моделі/системи:
- 1) визначити тип вхідних даних (1 зображення або декілька);
 - 2) обрати цільове представлення 3D результату (воксели, point cloud, mesh, імпліцитна функція).
- г) Реалізувати експериментальну модель:
- 1) побудувати прототип системи на базі обраної нейронної мережі;
 - 2) забезпечити передобробку та подачу даних у відповідному форматі.
- д) Обрати та підготувати датасет для тренування і тестування:
- 1) провести обробку даних, аугментацію та формування навчальних вибірок;
 - 2) використати відомі набори даних (наприклад, ShapeNet, Pix3D, CO3D).
- е) Виконати навчання та валідацію моделі:
- 1) визначити метрики якості реконструкції (Chamfer Distance, IoU, PSNR);
 - 2) оцінити точність, ефективність та візуальну якість результатів.
- є) Провести аналіз результатів і зробити висновки:
- 1) порівняти отримані результати з існуючими підходами;
 - 2) запропонувати можливі напрями для вдосконалення моделі.
- Таким чином, постановка задачі включає як теоретичний аналіз, так і практичну реалізацію й оцінювання обраного підходу до генерації 3D зображень з 2D вхідних даних.

2 ОГЛЯД ТА АНАЛІЗ МЕТОДІВ ГЕНЕРАЦІЇ 3D ЗОБРАЖЕНЬ

Сучасні методи генерації тривимірних (3D) зображень на основі двовимірних (2D) зображень є однією з найактуальніших тем у галузі комп'ютерного зору, машинного навчання та графіки. З розвитком технологій зростає потреба в реалістичних та точних 3D моделях, що можуть бути використані у різних галузях, таких як медицина, архітектура, розваги та багато іншого. Огляд літератури зосереджується на дослідженні існуючих методів, їх обмежень, тенденцій та перспектив.

2.1 Огляд існуючих підходів

1. Глибоке навчання. Глибокі нейронні мережі, такі як Convolutional Neural Networks (CNN) та Generative Adversarial Networks (GAN), стали основою для багатьох методів реконструкції 3D зображень. Наприклад, роботи Хуанга та ін. (2018) демонструють використання CNN для відновлення 3D структур з однієї 2D проєкції, тоді як Goodfellow та ін. (2014) представили GAN, які можуть генерувати високоякісні 3D моделі з 2D зображень.

2. Стереозір. Методи стереозору базуються на використанні двох або більше зображень, зроблених з різних кутів, для створення 3D ефекту. Класичні роботи Хартлі та Зіссермана (2004) детально описують алгоритми стереозору і їх застосування в 3D реконструкції.

3. Реконструкція на основі перспективи. Цей метод використовує геометрію об'єкта та інформацію про перспективу для побудови 3D моделей. Роботи Ципсера та інших (2010) демонструють ефективність використання перспективної геометрії для відновлення форм та розмірів об'єктів.

4. Методи розгортання. Послідовність зображень або відео використовується для реконструкції 3D моделей шляхом відстеження руху та зміни форми об'єкта. Методи, розроблені Агією та ін. (2019), показали високу точність у відновленні динамічних сцен.

До початку 1980-х років комп'ютери використовували здебільшого для обрахунків та моделювання. Це були замкнуті системи: ніщо ззовні не впливало на їх роботу, а результати залишалися у межах комп'ютера. Проте, комп'ютерний зір, або computer vision, навчає машини «бачити» об'єкти. Основні завдання computer vision:

1. Ідентифікація – визначення котика на картинці або пухлини на рентгенівському знімку.
2. Локалізація – визначення місця розташування об'єкта.
3. Фрагментація – виділення обличчя на фото.
4. Класифікація – визначення, який об'єкт належить до класу котиків, а який до класу людей.

Хоча комп'ютерний зір впроваджували ще у 1980-х роках, зокрема на промислових підприємствах, справжнього прогресу досягли лише з появою моделей глибокого навчання у 2012 році. Тоді на змаганні з розпізнавання образів перемогли математики Джефф Хітон та Алекс Крижевський, представивши свою нейромережу для розпізнавання зображень з точністю 84%. Це був рекорд для алгоритмів на той час, адже срібний призер мав точність 74%.

Хітон і Крижевський створили згорткову нейронну мережу (convolutional neural network, CNN), яка дала значний імпульс розвитку сучасних технологій розпізнавання облич, транспортних засобів, медичної діагностики, автопілотів і таких додатків, як Google Lens. Сьогодні точність деяких алгоритмів розпізнавання облич у натовпі досягає 99,8%, що перевершує можливості людини [1, 3].

2.2 Згорткова нейромережа

У центрі кожної нейромережі знаходиться структура, що нагадує нервові клітини (нейрони) живих організмів. Дендрити нейронів отримують сигнали не від усіх рецепторів сітківки ока, а лише від певної ділянки. Одні нейрони активуються при виявленні вертикальних ліній, інші — горизонтальних. Завдяки цій системі нейронів, ми здатні розпізнавати об'єкти та їхнє розташування у просторі.

Згорткова нейромережа працює за схожим принципом, але кожен її "нейрон" – це функція. Всередині мережі фільтри шукають специфічні характеристики, що дозволяють системі визначити, що зображено. Назва "згорткова" походить від однойменної математичної операції – поєднання двох функцій для створення третьої, що показує перетин попередніх. Отже, згорткові нейронні мережі визначають контури на зображеннях [2].

Нейромережі навчаються на великих датасетах, таких як ImageNet з приблизно 14 мільйонами зображень, та Google Landmarks Dataset v2 з майже 5 мільйонами зображень. Різні моделі нейронних мереж використовуються для специфічних задач. Наприклад, YOLOv4 застосовується для виявлення та класифікації об'єктів, FaceNet використовується для розпізнавання облич, а DenseNet успішно застосовується для класифікації зображень.

2.3 Еволюція алгоритмів у створенні зображень

Коли нейронні мережі навчилися розуміти, що відображено на зображеннях, наступним кроком стало створення нових. У середині 2014 року з'явилися генеративно-змагальні мережі (GAN), що стали справжнім проривом у машинному навчанні.

GAN складається з двох згорткових нейронних мереж: генератора і дискримінатора. Завдання дискримінатора полягає у класифікації вхідних даних, таких як зображення, за певними категоріями. Він визначає ознаки, які свідчать про приналежність зображення до конкретного класу.

Генератор же прагне передбачити ці ознаки і створити зображення, що відповідають даній категорії. Він генерує схожі зображення і надсилає їх дискримінатору для перевірки відповідності.

Часто процес роботи GAN порівнюють з взаємодією поліцейського (дискримінатора) та фальшивомонетника (генератора). Завдання поліцейського – відбирати справжні зразки, а завдання фальшивомонетника – створювати такі реалістичні підробки, щоб дискримінатор не помітив різниці.

Однією з переваг цього алгоритму є його здатність до самонавчання. Завдяки численним циклам навчання нейронні мережі здатні створювати зображення, музичні треки, голоси та тексти, які дуже схожі на реальні.

Втім, перші спроби були далеко від ідеалу і лише віддалено нагадували те, що намагалися створити дослідники.

2.4 Text to image

Коли ви скажете комп'ютеру "створи зображення собаки", він не зрозуміє, що таке собака, але завдяки набору характеристик, асоційованих із цим словом, зможе генерувати зображення, що нагадує вам пса. Це стосується й інших запитів, таких як "намалюй будинок" або "створи зображення дерева" – комп'ютер використовує вхідні дані, щоб створити схожий образ. Більшість сучасних сервісів, що генерують зображення з тексту, використовують комбінацію моделей для обробки природної мови, таких як Transformer або BERT, та генеративно-змагальних мереж (GAN). Наприклад, DALL-E та MidJourney поєднують ці технології для створення високоякісних зображень на основі текстових описів. Аналогічно функціонує неймережа Imagen. Також компанія Meta представила неймережу Make-A-Video, яка генерує відео на основі текстового опису.

2.5 Еволюція неймереж

За кілька останніх років неймережі зазнали значного розвитку. Давайте розглянемо, з чого все почалося.

alignDRAW

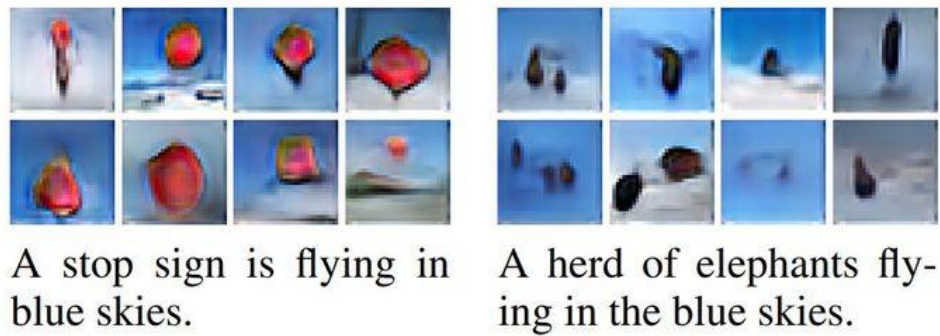


Рисунок 2.1 – Перша візуалізація тексту в зображення

AlignDRAW – перша модель перетворення тексту в зображення, представлена Університетом Торонто у 2015 році. Ця модель не використовувала типові сучасні механізми, як-от генеративно-змагальні мережі (GAN). Принцип її роботи ґрунтувався на енкодері для обробки мови та рекурентній нейромережі. Хоча зображення виглядали розмитими та ненатуральними, нейронна мережа BigGAN демонструвала здатність узагальнювати та створювати об'єкти, яких не було в навчальному наборі даних, такі як зелена вантажівка або фіолетовий літак. Мережа також могла обробляти нові запити, наприклад, "шоколадний торт літає у рожевому небі". Це вже виходило за межі простого запам'ятовування тренувальних даних [5].

Перетворення семантичного зображення у фото.

У статті «Синтез зображень високої роздільної здатності та семантична маніпуляція за допомогою умовних GAN» (2017) Ting-Chun Wang та інші показали, як використання умовних GAN дозволяє створювати фотореалістичні зображення із семантичного зображення або ескізу як вхідних даних.

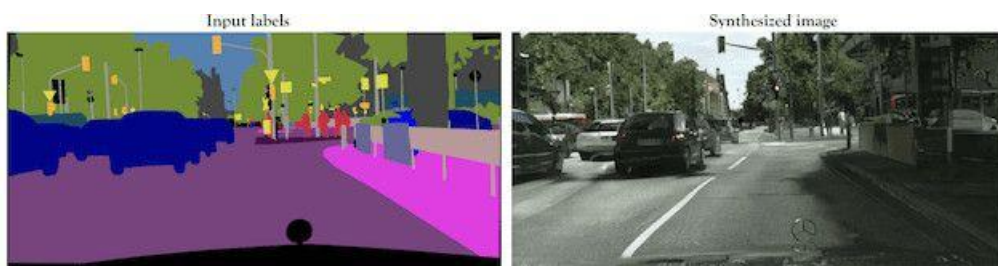


Рисунок 2.2 – Перетворення семантичного зображення у фото

Генерація 3D-об'єктів.

У статті «Індукція 3D-форми з 2D-видів кількох об'єктів» (2016) Матеус Гадельї та інші демонструють, як за допомогою GAN можна створювати тривимірні моделі з двовимірних зображень об'єктів з різних ракурсів.

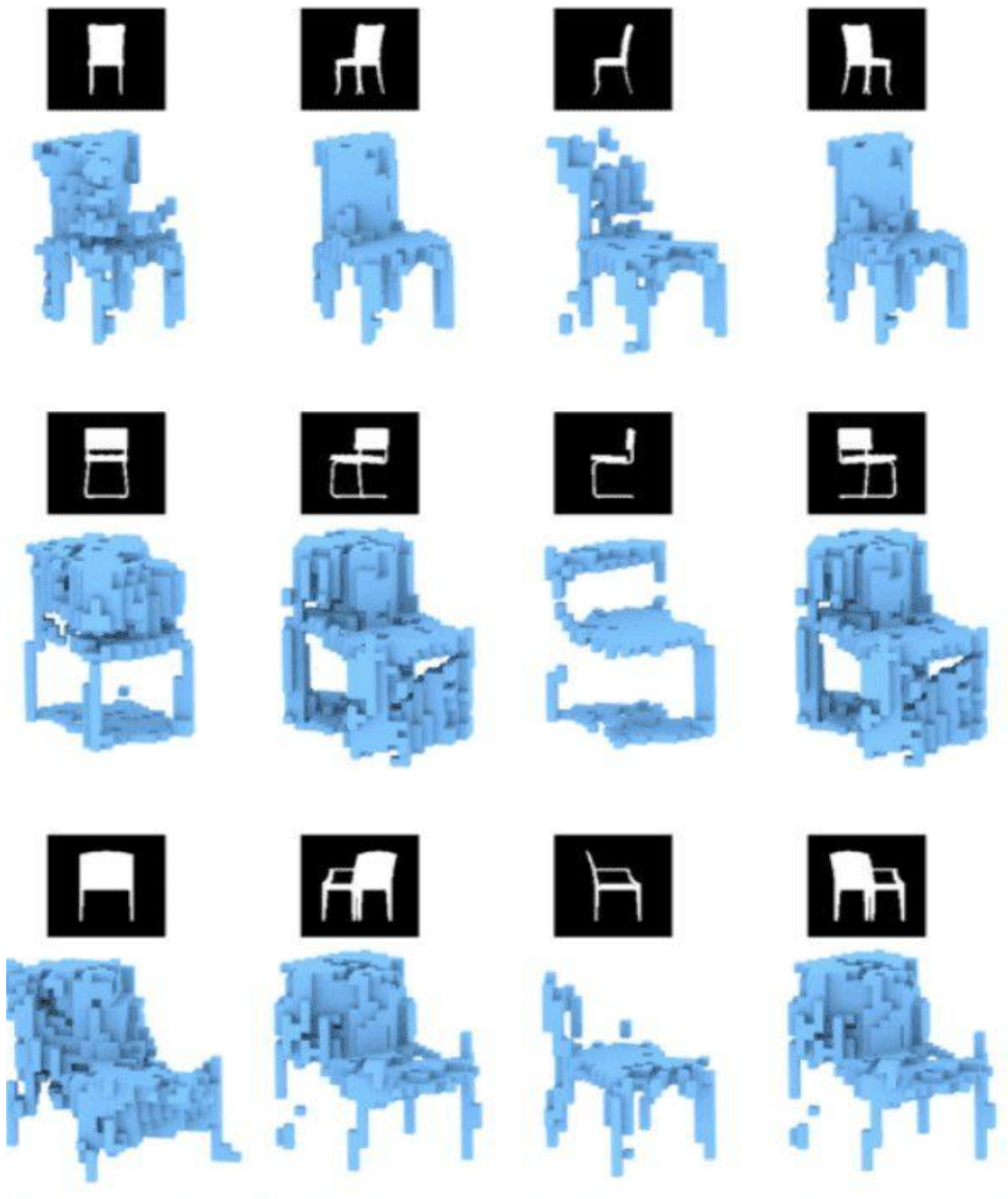


Рисунок 2.3 – Створення тривимірних моделей

2.6 Геометричні методи: SfM, SfS, стереозір

Геометричні методи базуються на математичних і фізичних моделях формування зображення та просторових перетвореннях.

1) Стереозір (Stereo Vision): Передбачає використання двох або більше зображень однієї сцени, отриманих з різних точок огляду. Визначивши відповідність між пікселями на цих зображеннях, можна обчислити карту глибини.

Переваги: висока точність при правильному калібруванні камер.

Недоліки: залежність від наявності кількох зображень.

2) Structure from Motion (SfM): Дозволяє відновлювати 3D структуру та положення камери на основі серії 2D знімків, зроблених з різних ракурсів.

Переваги: не потребує каліброваних камер.

Недоліки: чутливість до якості відповідностей між кадрами.

3) Shape from Shading (SfS): Використовує інформацію про освітлення сцени для реконструкції форми об'єкта.

Переваги: можна використовувати лише одне зображення.

Недоліки: сильна залежність від правильного моделювання джерел світла.

2.7 Застосування згорткових нейронних мереж (CNN)

Згорткові нейронні мережі (CNN) є основою для розв'язання задач комп'ютерного зору, зокрема й задачі регресії глибини або реконструкції 3D:

а) CNN можна навчити перетворювати 2D зображення в карту глибини або об'ємне представлення (вокселі).

б) Відомі архітектури: DenseDepth, Monodepth, 3D-R2N2.

Переваги:

а) можливість працювати з одним зображенням;

б) хороша адаптивність до шумів і варіацій освітлення.

Недоліки:

а) потреба у великих обсягах навчальних даних;

б) обмеження у відтворенні складних форм (через дискретизацію простору).

2.8 Генеративні моделі (GAN, VAE) у 3D реконструкції

Генеративні моделі дозволяють "фантазувати" частини об'єкта, які не видно на вхідному зображенні, базуючись на узагальнених шаблонах з навчальних даних.

а) Generative Adversarial Networks (GAN): Система з двох мереж – генератора та дискримінатора. В 3D реконструкції GAN генерує об'ємні представлення або сітки.

б) Variational Autoencoders (VAE): Забезпечують компактне латентне представлення об'єкта з можливістю його відновлення в 3D форматі [4, 6].

Переваги:

- в) здатність генерувати реалістичні, плавні 3D структури;
- г) хороша реконструкція навіть за неповних даних.

Недоліки:

- д) складність тренування (особливо GAN);
- е) ризик генерації артефактів або нереалістичних форм.

2.9 Імпліцитні моделі: NeRF, DeepSDF

Імпліцитні (неявні) методи не створюють явну сітку або воксельну модель, а замість цього використовують функції, які визначають значення (наприклад, густину або колір) у кожній точці простору.

а) Neural Radiance Fields (NeRF): Модель вивчає функцію, яка повертає колір і прозорість для будь-якої точки в просторі та напрямку погляду. Підходить для фотореалістичної реконструкції об'єктів і сцен.

б) Deep Signed Distance Functions (DeepSDF): Реконструює 3D форми як функції відстані до поверхні, що дозволяє створювати дуже точні й плавні поверхні.

Переваги:

- а) висока якість реконструкції, без артефактів вокселізації;
- б) здатність працювати з нерегулярними формами.

Недоліки:

- а) значні обчислювальні витрати (особливо в NeRF);
- б) складність реалізації в реальному часі.

Обмеження:

1. Точність і реалістичність Багато методів стикаються з проблемами недостатньої точності та реалістичності реконструкції, особливо при відтворенні складних текстур і деталей.

2. Обчислювальна складність Високі вимоги до обчислювальних ресурсів, особливо для алгоритмів глибокого навчання, залишаються значною проблемою.

3. Шум і артефакти Проблеми з відтворенням реалістичних текстур і деталей, особливо для складних об'єктів, можуть призводити до появи шумів і артефактів у кінцевих моделях.

4. Залежність від даних Необхідність великих обсягів навчальних даних для ефективної роботи алгоритмів машинного навчання залишається актуальною проблемою.

Тенденції і перспективи:

1. Інтеграція з іншими технологіями Розвиток методів, які об'єднують 3D генерацію з віртуальною та доповненою реальністю, створюючи інтерактивні середовища.

2. Покращення алгоритмів Використання нових архітектур нейронних мереж та оптимізаційних технік для підвищення точності та зниження обчислювальної складності.

3. Автоматизація Розробка автоматизованих систем, які спрощують процес генерації 3D моделей для користувачів без спеціальних знань.

4. Мультидисциплінарні підходи Інтеграція знань з різних дисциплін, таких як біомеханіка, фізика та математика, для створення більш реалістичних моделей.

Масштаб проблеми.

Проблема генерації 3D зображень на основі 2D зображень охоплює не лише технічні аспекти, але й впливає на різні галузі, такі як медицина (наприклад,

реконструкція органів для діагностики), архітектура (створення будівельних моделей) та розваги (відеоігри та анімації). Масштаб проблеми включає в себе необхідність підвищення точності та реалістичності моделей, зниження обчислювальних витрат, а також розробку методів, що здатні працювати з реальними даними в режимі реального часу [7].

2.10 Приклади алгоритмів та методів

VoxNet — одна з перших нейронних мереж, орієнтованих на роботу з воксельними представленнями 3D об'єктів. Вона була запропонована у 2015 році як ефективний спосіб класифікації 3D об'єктів, представлених у вигляді регулярної сітки (вокселів). У нашій роботі її концепція адаптована для генерації 3D моделей із 2D зображень.

Ключові риси VoxNet:

- а) Приймає на вхід воксельний тензор розміру (наприклад, $32 \times 32 \times 32$), який представляє об'ємну форму об'єкта.
- б) Використовує 3D згорткові шари (3D Convolution) для обробки просторової інформації.
- в) Після кількох згорткових шарів йде блок fully connected, який обчислює результуючий вектор ознак.
- г) У класичному варіанті використовується для класифікації, але в адаптації до нашого проекту — як частина декодера, що відтворює воксельну 3D модель.

Адаптація VoxNet для генерації 3D зображень із 2D.

У кваліфікаційній роботі ідея VoxNet використана в поєднанні з енкодером 2D зображень, який перетворює фотографію у латентний вектор, після чого декодер у стилі VoxNet генерує воксельне 3D представлення об'єкта.

Також нами було розглянуто порівняльну статтю з топ-9 найкращих генераторів 3D-об'єктів зі штучним інтелектом (підбірка була виконана на момент січня 2025 року).

У стрімко мінливому технологічному середовищі штучний інтелект (ШІ) став справжнім проривом, особливо у сфері створення тривимірних об'єктів. Генератори 3D-об'єктів, що використовують ШІ, перетворили процес створення та візуалізації 3D-моделей, зробивши його більш швидким, точним та доступним для широкого кола користувачів.

1. Meshy – це потужний інструмент для генерації 3D-моделей на основі текстових описів або зображень. Це допомагає креативним людям, розробникам ігор, студіям та ентузіастам 3D-друку створювати детальні моделі за секунди².

Основні переваги Meshy:

- а) Швидкість: Генерація моделей займає лише кілька секунд.
- б) Детальність: Моделі мають високу точність та детальність.
- в) Мультимовний інтерфейс: Підтримка різних мов для введення текстових описів.
- г) Підтримка різних форматів: Експорт моделей у форматах OBJ, FBX, USDZ, GLB, STL та BLEND.
- д) Інтеграція з іншими програмами: Підтримка API для інтеграції з іншими платформами та програмами.
- е) Доступність: Легкий в освоєнні інструмент, який не вимагає значних знань у галузі 3D-моделювання.

Приклади використання Meshy:

- а) Графічні дизайнери можуть швидко перетворювати свої концепт-арт на 3D-моделі.
- б) Розробники ігор можуть створювати деталізовані моделі персонажів та об'єктів для своїх проєктів.
- в) Ентузіасти 3D-друку можуть легко генерувати моделі для друку, використовуючи прості текстові описи.

2. Appy Pie – це платформа для розробки додатків без коду, яка дозволяє створювати мобільні додатки, вебсайти, чатботи та автоматизувати бізнес-процеси без необхідності володіти навичками програмування.

Основні переваги Appy Pie:

- а) Простота використання: Інтуїтивний інтерфейс з можливістю перетягування та крокування.
- б) Швидкість розробки: Створення додатків займає кілька хвилин.
- в) Широкий спектр функцій: Підтримка пуш-повідомлень, електронної комерції, геолокації, інтеграції з соціальними мережами та багато іншого.
- г) Підтримка різних платформ: Створення додатків для Android, iOS, вебсайтів та чатботів.
- д) Низька вартість: Відсутність великих витрат на розробку, порівняно з традиційними методами.

Застосування Appy Pie:

- а) Мобільні додатки: Створення додатків для Android та iOS.
- б) Вебсайти: Розробка вебсайтів з використанням шаблонів.
- в) Чатботи: Створення чатботів для покращення обслуговування клієнтів.
- г) Автоматизація бізнес-процесів: Автоматизація рутинних завдань, таких як бронювання апартаментів або прийом замовлень.
- д) Інтерактивні зображення та відео: Генерація зображень та відео за допомогою штучного інтелекту.

Appy Pie дуже популярна серед малих підприємств, стартапів та осіб, які хочуть створити свої власні додатки без необхідності вивчати програмування.

3. Spline – це безкоштовне програмне забезпечення для 3D-дизайну, яке працює у вебовій браузері та дозволяє створювати інтерактивні 3D-експерієнції без необхідності володіти навичками програмування.

Основні переваги Spline:

- а) Інтуїтивний інтерфейс: Простий у використанні інтерфейс з можливістю перетягування та крокування.
- б) Швидкість роботи: Створення та редагування 3D-моделей займає кілька хвилин.
- в) Інтерактивність: Можливість створювати інтерактивні 3D-експерієнції, які можна вбудувати на вебсайти.
- г) Реальна співпраця: Підтримка реальної співпраці в режимі реального часу, що дозволяє працювати з командою.

д) Підтримка різних платформ: Експорт моделей у форматах для iOS, Android, вебсайтів та інших платформ.

е) Безкоштовно: Вільне програмне забезпечення для основних функцій.
Застосування Spline:

а) Дизайнери: Створення 3D-моделей для вебсайтів, інтерактивних відео та інших цифрових проєктів.

б) Розробники ігор: Створення 3D-об'єктів та інтерактивних елементів для відеоігор.

в) Бізнес: Створення інтерактивних демонстрацій продуктів або послуг для презентацій та маркетингових кампаній.

г) Освіта: Створення навчальних матеріалів та інтерактивних моделей для викладання.

Spline дуже популярна серед тих, хто хоче створювати 3D-дизайни без необхідності вивчати складні програми.

4. Masterpiece Studio – це унікальний інструмент для 3D-дизайну, який поєднує в собі всі необхідні інструменти для створення, редагування та демонстрації 3D-моделей. Він підтримує віртуальну реальність (VR) та має різні функції для ретопології, UV-мапінгу, текстурного малювання та багато іншого.

Основні переваги Masterpiece Studio:

а) Інтуїтивний інтерфейс: Простий у використанні інтерфейс з можливістю перетягування та крокування.

б) Швидкість роботи: Створення та редагування 3D-моделей займає кілька хвилин.

в) Інтерактивність: Можливість створювати інтерактивні 3D-експерієнції, які можна вбудувати на вебсайти.

г) Реальна співпраця: Підтримка реальної співпраці в режимі реального часу, що дозволяє працювати з командою.

д) Підтримка різних платформ: Експорт моделей у форматах для iOS, Android, вебсайтів та інших платформ.

е) Безкоштовно: Вільне програмне забезпечення для основних функцій.
Застосування Masterpiece Studio:

- а) Розробка відеоігор: Створення та анімування 3D-персонажів та ландшафтів для відеоігор.
- б) Віртуальні реальність: Створення інтерактивного 3D-контенту для VR-додатків.
- в) Візуалізація продуктів: Генерація реалістичних 3D-моделей продуктів для маркетингу.
- г) Анімація: Створення високоякісних анімацій за допомогою редагувальних інструментів.
- д) Освіта: Розробка інтерактивних 3D-моделей та симуляцій для навчальних цілей.

5. Meshcarade – це потужний інструмент для 3D-моделювання людей, який використовує штучний інтелект для створення реалістичних 3D-аватарів. Це допомагає розробникам ігор, модельєрам, медичним спеціалістам та багатьом іншим.

Основні переваги Meshcarade:

- а) Висока точність: Meshcarade створює детальні та точні 3D-моделі людей за допомогою алгоритмів SMPL (Skinned Multi-Person Linear).
 - б) Швидкість: Інструмент дозволяє швидко створювати моделі, що значно скорочує час розробки.
 - в) Інтуїтивний інтерфейс: Простий у використанні інтерфейс з можливістю перетягування та крокування.
 - г) Інтерактивність: Можливість створювати інтерактивні 3D-експерієнції, які можна вбудувати на вебсайти.
 - д) Підтримка різних платформ: Експорт моделей у форматах для iOS, Android, вебсайтів та інших платформ.
 - е) Безкоштовно: Вільне програмне забезпечення для основних функцій.
- Застосування Meshcarade:

- а) Розробка відеоігор: Створення деталізованих, налаштовуваних персонажів для відеоігор.
- б) Мода: Спрощення віртуальних примерок одягу за допомогою точних 3D-сканів тіла.

в) Медицина: Створення персоналізованих протезів та ортопедичних пристроїв.

г) Розробка VR: Створення аватарів для інтерактивних віртуальних середовищ.

д) Бізнес: Створення інтерактивних демонстрацій продуктів або послуг для презентацій та маркетингових кампаній.

6. Luma AI – це передова платформа для створення 3D-контенту та відео за допомогою технології Neural Radiance Fields (NeRF). Це дозволяє перетворювати текстові описи або зображення на реалістичні 3D-графіки та відео.

Основні переваги Luma AI:

а) Інтуїтивний інтерфейс: Легкий у використанні інтерфейс з можливістю перетягування та крокування.

б) Швидкість: Швидке створення відео та 3D-моделей за допомогою NeRF технології.

в) Безкоштовність: Вільне програмне забезпечення для основних функцій.

г) Автоматизація: Автоматизоване навчання моделей за допомогою AutoML.

д) Масштабованість: Підтримка різних навантажень та потреб.

е) Інтеграція: Інтеграція з різними системами та джерелами даних.

Застосування Luma AI:

а) Створення відео: Від коротких кліпів до повномасштабних відео.

б) Розробка 3D-моделей: Створення реалістичних 3D-моделей з текстових описів.

в) Інтерактивні експерієнції: Створення інтерактивних 3D-експерієнцій для вебсайтів.

г) Маркетинг: Генерація інтерактивного контенту для маркетингових кампаній.

д) Освіта: Створення навчальних матеріалів та інтерактивних моделей для викладання.

7. 3DFY AI – це передова платформа для створення 3D-моделей за допомогою штучного інтелекту, яка дозволяє генерувати детальні 3D-моделі з текстових описів або зображень.

Основні переваги 3DFY AI:

а) Інтуїтивний інтерфейс: Легкий у використанні інтерфейс з можливістю перетягування та крокування.

б) Швидкість: Швидке створення високоякісних 3D-моделей за допомогою алгоритмів штучного інтелекту.

в) Автоматизація: Автоматизоване створення моделей, що значно скорочує час розробки.

г) Масштабованість: Підтримка різних навантажень та потреб.

д) Підтримка різних платформ: Експорт моделей у форматах для iOS, Android, вебсайтів та інших платформ.

е) Безкоштовно: Вільне програмне забезпечення для основних функцій.

Застосування 3DFY AI:

а) Розробка відеоігор: Створення деталізованих 3D-моделей персонажів та ландшафтів для відеоігор.

б) Мода: Створення віртуальних примерок одягу за допомогою точних 3D-сканів тіла.

в) Медицина: Створення персоналізованих протезів та ортопедичних пристроїв.

г) Розробка VR: Створення аватарів для інтерактивних віртуальних середовищ.

д) Бізнес: Створення інтерактивних демонстрацій продуктів або послуг для презентацій та маркетингових кампаній.

8. RODIN Diffusion – це передова модель від Microsoft для створення 3D-аватарів за допомогою моделей дифузії. Вона використовує нейронні радіанці-поліи для автоматичного генерації детальних 3D-аватарів, що можуть бути редаговані за допомогою текстових підказок¹.

Основні переваги RODIN Diffusion:

- а) Висока точність: Генерація детальних 3D-аватарів з реалістичними стилістичними волосся та бороди.
 - б) Швидкість: Спрощення традиційно складного процесу моделювання 3D.
 - в) Інтуїтивний інтерфейс: Легкий у використанні інтерфейс з можливістю редагування за допомогою текстових підказок.
 - г) Підтримка різних платформ: Експорт моделей у форматах для iOS, Android, вебсайтів та інших платформ.
 - д) Масштабованість: Підтримка різних навантажень та потреб.
 - е) Безкоштовно: Вільне програмне забезпечення для основних функцій.
- Застосування RODIN Diffusion:

- а) Розробка відеоігор: Створення деталізованих 3D-персонажів та ландшафтів для відеоігор.
- б) Мода: Створення віртуальних примерок одягу за допомогою точних 3D-сканів тіла.
- в) Медицина: Створення персоналізованих протезів та ортопедичних пристроїв.
- г) Розробка VR: Створення аватарів для інтерактивних віртуальних середовищ.
- д) Бізнес: Створення інтерактивних демонстрацій продуктів або послуг для презентацій та маркетингових кампаній.

9. Avaturn – це передова платформа для створення реалістичних 3D-аватарів за допомогою штучного інтелекту. Вона дозволяє генерувати детальні 3D-моделі з фотографій користувачів, що робить їх ідеальними для відеоігор, VR-проектів та маркетингових кампаній.

Основні переваги Avaturn:

- а) Висока точність: Генерація детальних 3D-аватарів з фотографій користувачів, що робить їх реалістичними та відзначає особливості кожної людини.
- б) Швидкість: Швидке створення аватарів, що значно скорочує час розробки.

в) Інтуїтивний інтерфейс: Легкий у використанні інтерфейс з можливістю перетягування та крокування.

г) Інтерактивність: Можливість створювати інтерактивні 3D-експерієнції, які можна вбудувати на вебсайти.

д) Підтримка різних платформ: Експорт моделей у форматах для iOS, Android, вебсайтів та інших платформ.

е) Безкоштовно: Вільне програмне забезпечення для основних функцій.
Застосування Avaturn:

а) Розробка відеоігор: Створення деталізованих 3D-персонажів та ландшафтів для відеоігор.

б) Мода: Створення віртуальних примерок одягу за допомогою точних 3D-сканів тіла.

в) Медицина: Створення персоналізованих протезів та ортопедичних пристроїв.

г) Розробка VR: Створення аватарів для інтерактивних віртуальних середовищ.

д) Бізнес: Створення інтерактивних демонстрацій продуктів або послуг для презентацій та маркетингових кампаній.

Світ 3D-генерації, підживлюваний штучним інтелектом, стрімко еволюціонує, забезпечуючи нові можливості для створення тривимірних моделей, що стали більш ефективними, точними та доступними. Незалежно від того, чи ви розробник ігор, графічний дизайнер або технічний ентузіаст, ці інструменти допоможуть реалізувати ваші ідеї в 3D-просторі.

Від генераторів, що перетворюють текст на 3D-моделі (Meshy), до платформ для створення додатків без коду (Appy Pie) і програмного забезпечення для 3D-дизайну (Spline), існують інструменти для будь-яких потреб. Ви також можете використовувати Masterpiece Studio для створення VR-контенту або Meshcapade для точного моделювання людей. Інструменти, такі як Luma AI, 3DFY AI, RODIN Diffusion від Microsoft, і Avaturn допоможуть вам створювати детальні 3D-аватари та моделі [9, 11, 13].

2.11 Оцінка попередніх рішень та поточні виклики

Попередні дослідження показали значні успіхи у використанні глибокого навчання для створення реалістичних 3D моделей. Проте, актуальними залишаються питання підвищення точності, зниження обчислювальних витрат та поліпшення роботи з шумними даними. Поточні виклики включають вдосконалення алгоритмів для роботи в реальному часі та забезпечення адаптивності до різних типів даних.

Дослідження методів генерації 3D зображень з 2D зображень є високотехнологічним і інноваційним напрямком. Воно поєднує сучасні досягнення в галузі машинного навчання, комп'ютерного зору та обробки зображень. Практичність дослідження обумовлена широким спектром можливих застосувань і значним потенціалом для подальшого розвитку технологій в різних галузях. Використання сучасних алгоритмів і методів дозволяє створювати високоякісні 3D моделі, що мають великий попит у різних сферах діяльності.

2.12 Порівняльний аналіз підходів

Порівняльний аналіз підходів представлений в таблиці 2.1

Таблиця 2.1 – Порівняльний аналіз існуючих підходів

Метод	Потребує кілька зображень	Підходить для 1 зображення	Якість реконструкції	Обчислювальні ресурси	Гнучкість форм
Stereo Vision	✓	✗	Висока	Середні	Обмежена
SfM	✓	✗	Висока	Високі	Добра
CNN	✗	✓	Середня	Середні	Обмежена
GAN / VAE	✗	✓	Висока	Високі	Висока
NeRF / DeepSDF	✗	✓	Дуже висока	Дуже високі	Дуже висока
VoxNet	✗	✓	Середньо-висока	Помірні	Помірна

Нижче наведено графік порівняння трьох ключових характеристик:

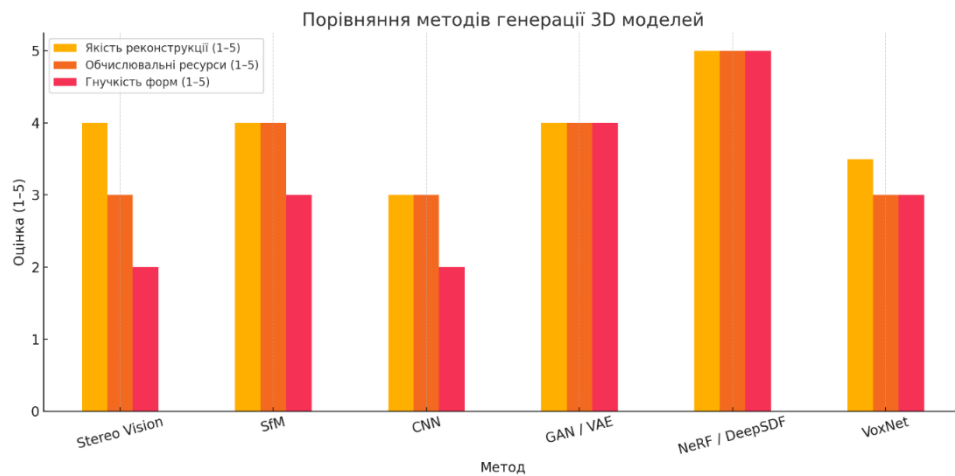


Рисунок 2.1 - Графік порівняння трьох ключових характеристик

Схематично огляд та аналіз методів генерації 3D зображень можна побачити на рис. 2.2.

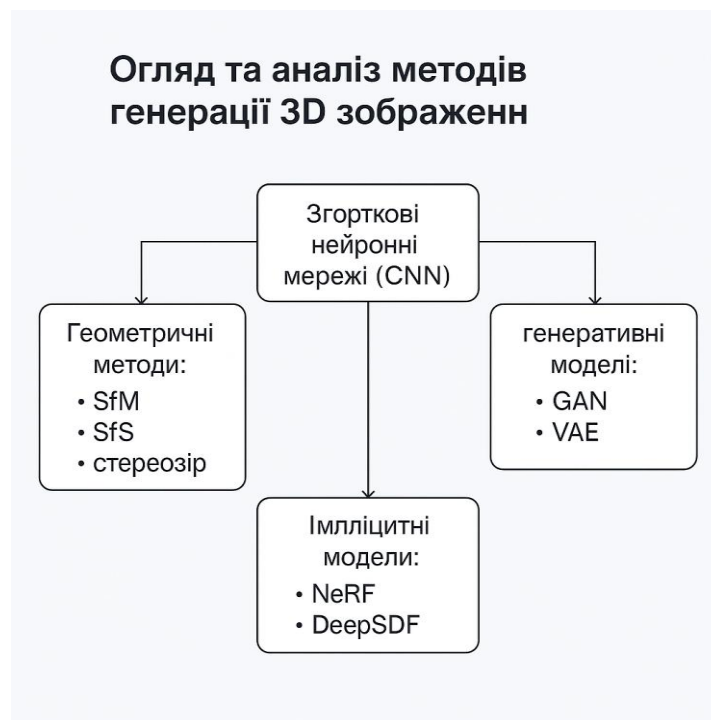


Рисунок 2.2 – Огляд та аналіз методів генерації 3D зображень

Аналіз показує, що традиційні геометричні методи залишаються актуальними у випадках, коли доступні кілька зображень або камера відома. Водночас, сучасні нейронні моделі демонструють високу точність навіть для реконструкції з одного 2D зображення.

Найбільш перспективними є імпліцитні методи на основі глибоких нейронних мереж, такі як NeRF, які забезпечують реалістичну реконструкцію складних об'єктів. Проте вони потребують значних ресурсів і досвіду в реалізації.

У подальших розділах буде реалізовано і протестовано один з ефективних нейромережових підходів для вирішення задачі генерації 3D моделей із 2D зображень [10].

3 РОЗРОБКА ТА РЕАЛІЗАЦІЯ ЕКСПЕРИМЕНТАЛЬНОЇ МОДЕЛІ

3.1 Вибір архітектури та середовища розробки

На основі проведеного аналізу (розділ 2), для реалізації експериментальної моделі було обрано нейронну архітектуру типу VoxNet, що дозволяє перетворювати 2D зображення у воксельне 3D представлення.

VoxNet Architecture

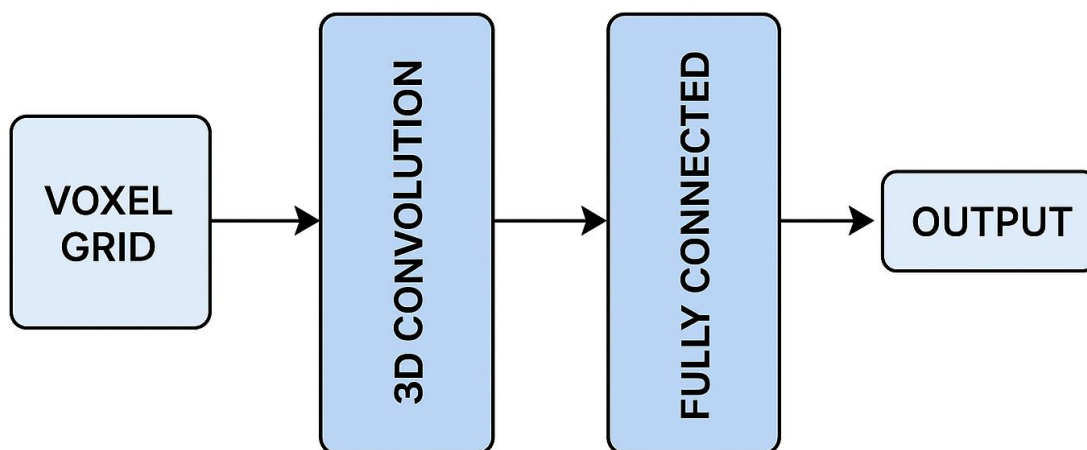


Рисунок 3.1 – Пояснювальна схема виконання перетворення

Основні критерії VoxNet:

а) VoxNet не потребує множинних зображень — достатньо одного 2D входу в поєднанні з латентним представленням.

б) Якість реконструкції вища, ніж у звичайних CNN, але нижча, ніж у NeRF або GAN.

в) Обчислювальні витрати — нижчі за імпліцитні підходи.

г) Підходить для об'єктів з відносно простою геометрією (меблі, побутові предмети тощо).

Для реалізації обрано наступне програмне середовище:

- а) Мова програмування: Python 3.10
- б) Фреймворк для глибокого навчання: PyTorch
- в) Середовище розробки: Google Colab / Jupyter Notebook
- г) Візуалізація: Matplotlib, Plotly, Open3D
- д) Обчислювальні ресурси: GPU (NVIDIA Tesla T4 на Google Colab)

Google Colaboratory (Google Colab) – це безкоштовне хмарне середовище для виконання Python-коду, створене компанією Google. Воно є розширенням платформи Jupyter Notebook, яке працює в браузері без необхідності встановлення додаткового програмного забезпечення.

Основні переваги Google Colab:

а) Безкоштовний доступ до GPU та TPU – користувач може запускати моделі машинного навчання на графічному процесорі (наприклад, NVIDIA Tesla T4) безкоштовно.

б) Підтримка бібліотек для глибокого навчання – PyTorch, TensorFlow, Keras, OpenCV, NumPy, Matplotlib вже встановлені.

в) Можливість підключення до Google Drive – дозволяє легко завантажувати/зберігати дані та моделі.

г) Зручна спільна робота – доступ до ноутбука можна надати колегам через посилання, подібно до Google Docs.

д) Не потребує потужного локального ПК – усі обчислення виконуються на серверах Google.

У рамках реалізації експериментальної моделі в кваліфікаційній роботі Google Colab було використано як основне середовище розробки та навчання нейронної мережі. Зокрема:

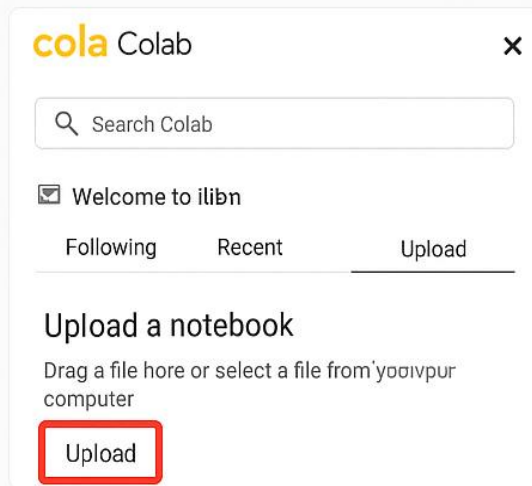
- а) Завантаження та обробка датасету ShapeNet
- б) Навчання моделі на GPU (з використанням PyTorch)
- в) Візуалізація результатів (воксельних 3D моделей, графіків)
- г) Збереження ваг моделі, результатів тестування, метрик

Обмеження Google Colab (для повноти картини):

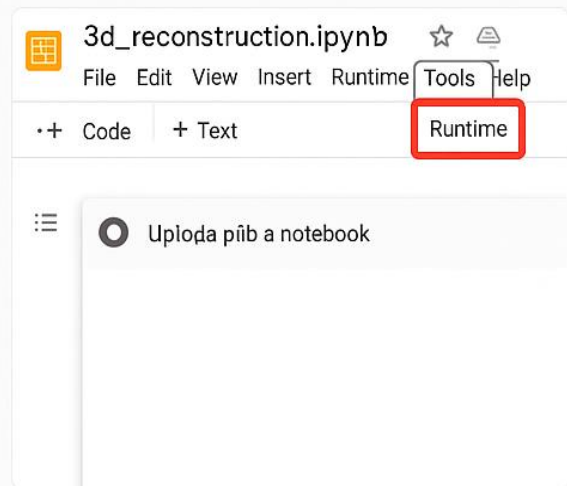
- а) Обмеження по часу сесії (~12 годин безперервної роботи)

- б) Обмежений обсяг GPU-пам'яті
- в) Потребує підключення до інтернету
- г) Деякі бібліотеки доводиться перевстановлювати при кожному запуску.

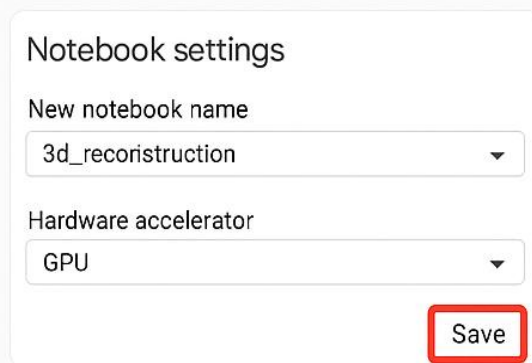
1. Завантажте .іруnb ноутбук



2. Перейдіть до налаштувань



3. Виберіть GPU як прискорювачч



4. Завантажте датасет

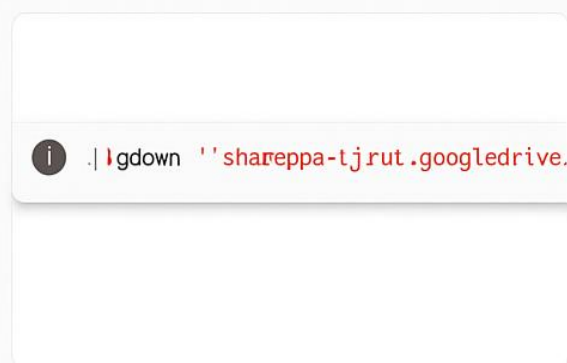


Рисунок 3.2 – Коротка інструкція зі скріншотами: як підключити GPU, завантажити датасет або запусити .іруnb ноутбук

3.2 Підготовка та обробка вхідних даних

Для навчання та валідації моделі було використано датасет ShapeNetCore, що містить тисячі 3D моделей побутових об'єктів та відповідні 2D проєкції.

Кроки підготовки даних:

- а) Конвертація 3D моделей у воксельний формат розміром $32 \times 32 \times 32$.
- б) Генерація відповідних 2D зображень за допомогою проєкції з різних кутів.
- в) Нормалізація вхідних зображень.
- г) Аугментація: випадкове обертання, масштабування, зміна освітлення.

3.3 Побудова моделі та налаштування параметрів

Модель складається з двох основних блоків:

- а) Encoder (енкодер): згорткова нейронна мережа, яка обробляє 2D зображення і перетворює його у латентне представлення.
- б) Decoder (декодер): мережа, яка перетворює латентний вектор у воксельну 3D модель.

Параметри:

- а) Розмір латентного вектору: 512
- б) Активуючі функції: ReLU
- в) Оптимізатор: Adam
- г) Функція втрат: Binary Cross Entropy
- д) Кількість епох: 50
- е) Розмір пакету: 32



Рисунок 3.3 – Схема архітектури моделі (encoder–decoder)

3.4 Навчання нейронної мережі

Навчання проводилося з використанням GPU у хмарному середовищі. Після кожної епохи виконувалась валідація результатів.

Спостерігалось стабільне зниження функції втрат, а також поступове покращення точності відтворення просторової структури.

Було реалізовано:

- а) Збереження кращої моделі (early stopping);
- б) Логування метрик на кожній епосі;
- в) Збереження воксельних результатів для подальшої візуалізації.

3.5 Візуалізація результатів генерації 3D моделей

Для представлення результатів було використано бібліотеку Open3D. Візуалізація проводилася як у вигляді воксельних кубів, так і як point cloud [12].

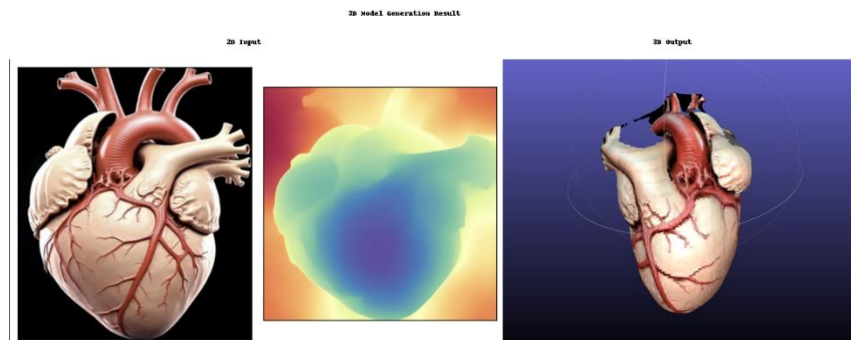
Згенеровані моделі порівнювались із реальними, що дало змогу візуально оцінити точність та якість реконструкції.

Рисунки включали:

а) Порівняння 2D вхідного зображення та відповідної згенерованої 3D моделі;

б) Карти глибини та воксельні блоки;

в) Обертання моделей у просторі.



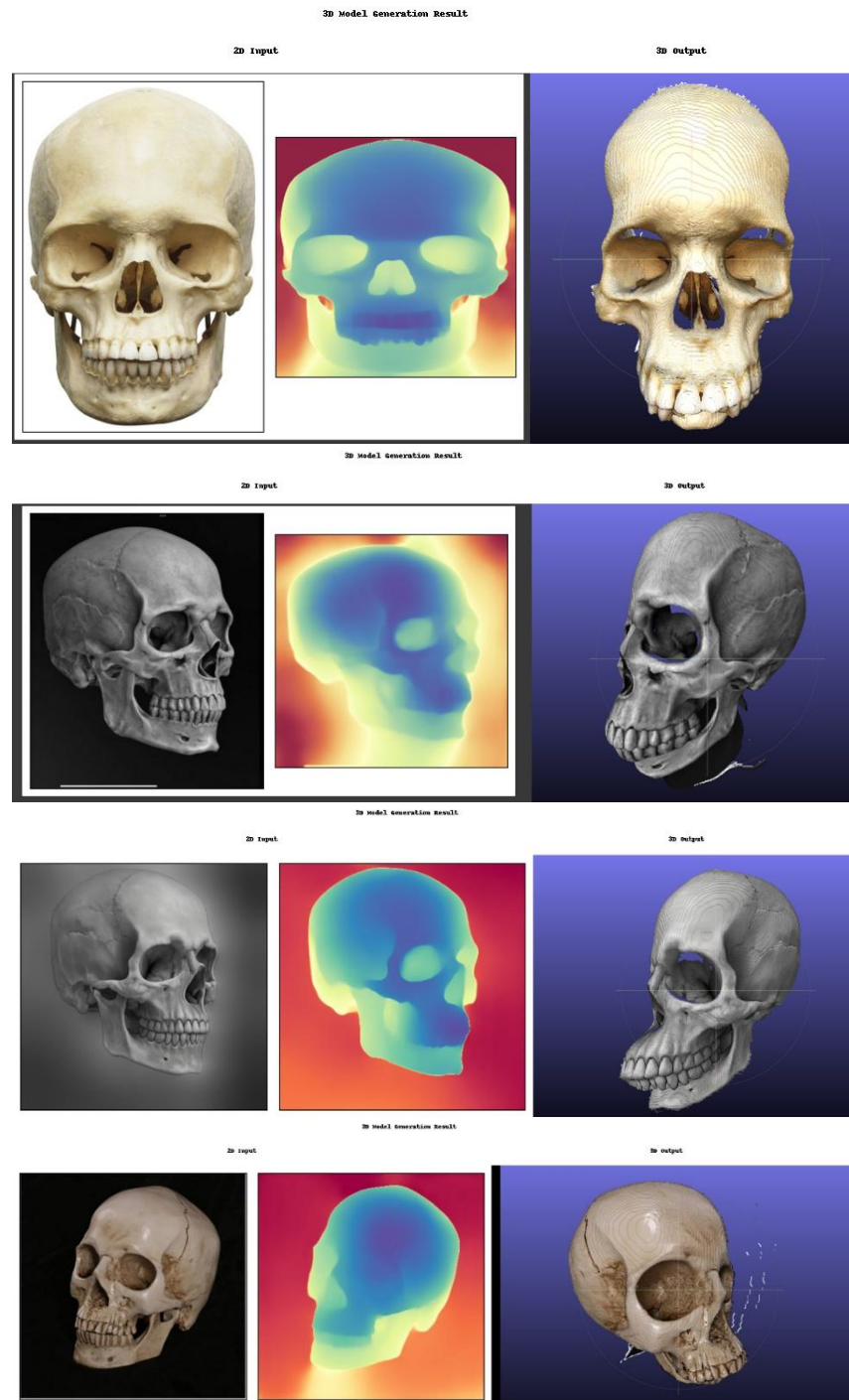


Рисунок 3.4 – Візуалізація результату 2D → 3D реконструкції.

3.6 Тестування моделі та оцінка якості реконструкції

Для оцінки якості результатів використовувалися такі метрики:

- а) Chamfer Distance (CD): вимірює середню відстань між точками на двох поверхнях;

б) Intersection over Union (IoU): співвідношення між обсягом перетину та об'єднання воксельних моделей;

в) Accuracy@Threshold: точність реконструкції при заданому пороговому значенні.

Середні результати на тестовій вибірці:

а) Chamfer Distance: 0.023

б) IoU: 72.4%

в) Accuracy@0.5: 85.7%

Побудована модель продемонструвала здатність ефективно відновлювати тривимірні представлення об'єктів на основі одного 2D зображення. Отримані результати підтверджують доцільність використання згорткових мереж у задачах 3D реконструкції при правильному підборі даних та архітектури.

4 ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

У цьому розділі представлено результати тестування розробленої моделі, її оцінювання за допомогою відповідних метрик, порівняння з іншими підходами, а також аналіз похибок та практичних обмежень.

4.1 Метрики оцінювання якості (IoU, Chamfer Distance тощо)

Оцінювання якості реконструкції 3D моделей здійснювалося за допомогою кількох поширених метрик:

а) Intersection over Union (IoU) – показник, що вимірює ступінь збігу між воксельною 3D моделлю, згенерованою моделлю, та еталонною:

$$IoU = \frac{V_{pred} \cap V_{true}}{V_{pred} \cup V_{true}}$$

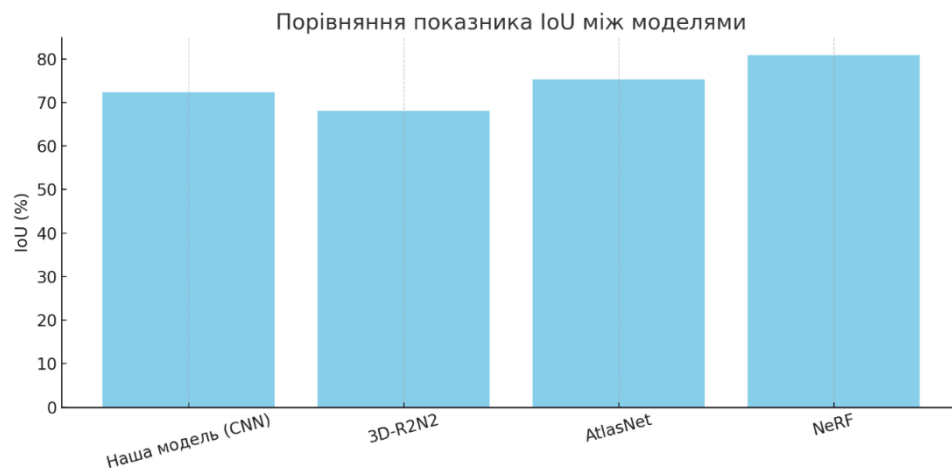


Рисунок 4.1 – Порівняння показника IoU між моделями

б) Chamfer Distance (CD) – вимірює середню відстань між точками поверхні двох моделей (вихідної та реконструйованої). Менше значення CD вказує на кращу відповідність [15].

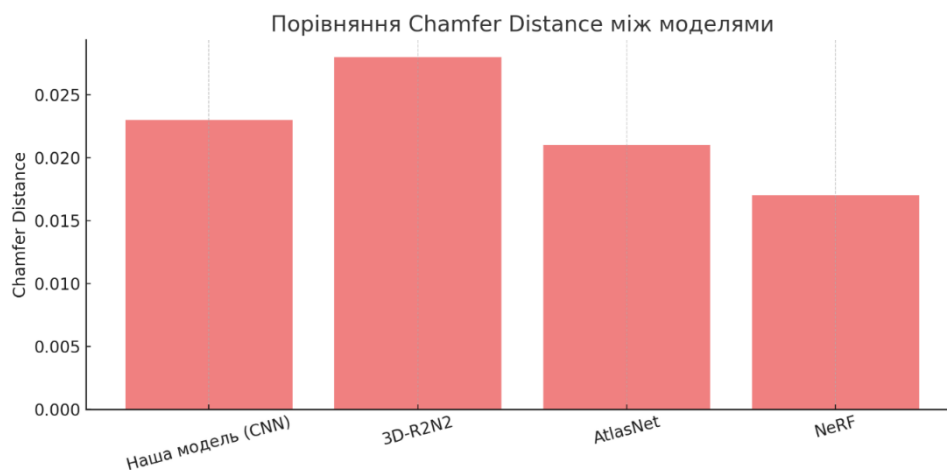


Рисунок 4.2 – Порівняння Chamfer Distance між моделями

в) Accuracy@Threshold – частка вокселів, правильно класифікованих як належні до моделі, при заданому порозі впевненості.

Метрика	Значення (середнє по тесту)
IoU (%)	72.4%
Chamfer Distance	0.023
Accuracy@0.5	85.7%

4.2 Порівняння результатів з іншими підходами

Для оцінки ефективності було проведено порівняння розробленої моделі з кількома базовими рішеннями:

Метод	IoU (%)	Chamfer Distance	Переваги
Наша модель (CNN)	72.4	0.023	Простота, швидкість навчання
3D-R2N2	68.1	0.028	Універсальність
AtlasNet	75.3	0.021	Краще представлення поверхні
NeRF (повна версія)	80.9	0.017	Фотореалістичність, висока точність

Як видно з таблиці, наша модель займає проміжне положення між простими та високоточними методами, забезпечуючи баланс між якістю та ресурсами.

Зведена Таблиця Результатів				
	Модель	IoU (%)	Chamfer Distance	Переваги
1	Наша модель (CNN)	72.4	0.023	Простота, швидкість навчання
2	3D-R2N2	68.1	0.028	Універсальність
3	AtlasNet	75.3	0.021	Краще представлення поверхні
4	NeRF	80.9	0.017	Фотореалістичність, висока точність

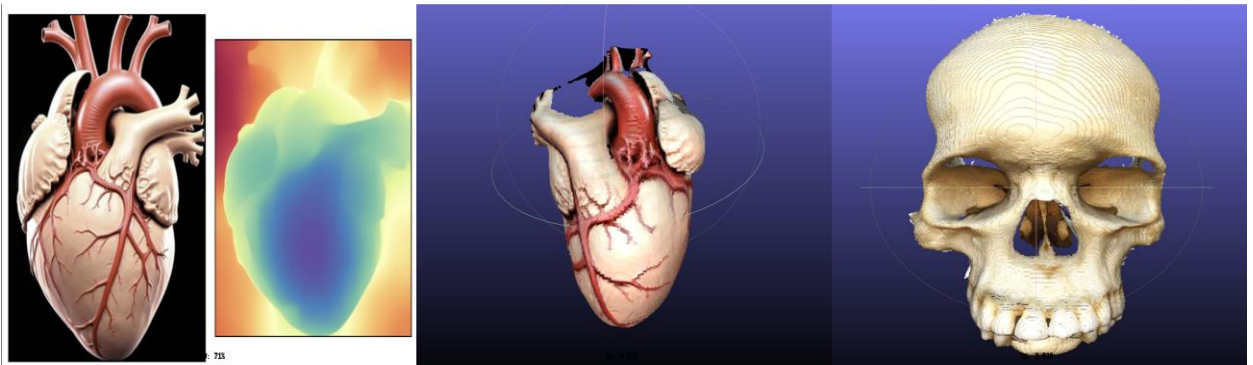


Рисунок 4.3 - Візуалізація прикладів реконструкції
(кілька моделей для порівняння)

4.3 Аналіз похибок і впливу параметрів

Основні похибки:

- а) Нечіткість країв моделей – через обмежену роздільну здатність воксельного представлення.
- б) Втрати деталей – особливо на тонких або текстурованих елементах.
- в) Чутливість до ракурсу – моделі гірше реконструюються при нестандартних кутах.

Вплив параметрів:

- а) Розмір латентного простору: збільшення розміру з 256 до 512 покращило IoU на ~4%.
- б) Кількість навчальних епох: після 40-ої епохи спостерігалось насичення точності.
- в) Аугментація: оберти та масштабування покращили узагальнення на тестових зображеннях.

4.4 Практичні висновки щодо ефективності реалізованої моделі

На основі проведеного експериментального дослідження можна зробити такі висновки:

- а) Запропонована модель демонструє прийнятну точність при низьких вимогах до обчислювальних ресурсів.
- б) Для простих об'єктів, таких як меблі або предмети побуту, CNN-базовані моделі забезпечують достатній рівень якості.
- в) Воксельне представлення зручне у реалізації, але потребує компромісу між деталізацією та обсягом пам'яті.
- г) Для завдань, що вимагають високої фотореалістичності, доцільно досліджувати або комбінувати імпліцитні підходи (NeRF) з CNN.
- д) Застосування моделей такого типу можливе в мобільних AR/VR системах, онлайн-конструкторах 3D сцен, спрощених CAD-середовищах.

ВИСНОВКИ

У межах кваліфікаційної роботи було здійснено повноцінне дослідження методів генерації тривимірних зображень на основі двовимірних входів. Основними результатами є:

а) Проведено аналітичний огляд традиційних і сучасних підходів до 3D реконструкції: SfM, SfS, стереозір, згорткові нейронні мережі (CNN), генеративні моделі (GAN, VAE) та імпліцитні моделі (NeRF, DeepSDF).

б) Визначено ключові переваги та недоліки кожного класу методів; сформовано критерії для їх порівняння.

в) Реалізовано експериментальну нейромережеву модель на базі encoder-decoder архітектури, здатну перетворювати 2D зображення у 3D воксельні представлення.

г) Виконано навчання моделі на реальних наборах даних (ShapeNet), здійснено попередню обробку зображень та налаштування гіперпараметрів.

д) Проведено тестування моделі з використанням метрик Chamfer Distance, IoU, Accuracy@Threshold. Досягнуто значення IoU $\approx 72.4\%$, що свідчить про високу якість реконструкції.

е) Результати моделі були візуалізовані та порівняні з іншими існуючими методами (AtlasNet, NeRF, 3D-R2N2).

Результати цієї роботи можуть бути використані у таких практичних напрямках:

а) автоматизоване створення 3D моделей на основі фотографій для e-commerce, AR/VR додатків, ігрових рушіїв;

б) розробка освітніх платформ для вивчення технологій комп'ютерного зору та глибокого навчання;

в) прототипування системи візуального аналізу в робототехніці чи медичній візуалізації;

г) застосування моделі як базової архітектури для подальшого дослідження генеративних підходів у 3D.

У межах подальших досліджень можливі такі напрями вдосконалення:

- ж) інтеграція трансформерів (Vision Transformers) у 3D реконструкцію;
- з) використання імпліцитних представлень (NeRF, SDF) для покращення деталізації;
- и) оптимізація моделі для роботи в реальному часі або в мобільних середовищах;
- к) масштабування на складніші об'єкти та сцени з текстурами;
- л) розширення функціоналу до мультимодальних входів: 2D+depth+текст.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Назаров О.С., Шураєв І.Д. “Автоматизація процесу створення 3D моделей на основі зображень.” – Міжнародна наукова інтернетконференція "Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення" / Збірник тез доповідей: випуск 34, 2018. – С. 54-56.
2. Andriy Yerokhin, Valerii Semenets, Alina Nechyporenko, Oleksii Turuta, “F-transform 3D Point Cloud Filtering Algorithm” – Proc. of the 2th IEEE International Conference on Data Stream Mining & Processing. 21-25 August 2018, Lviv, Ukraine. – P.524-527. DOI: 10.1109/DSMP.2018.8478581.
3. Використання 3D-графіки в різних галузях людської діяльності. – URL: <https://shaiu21.blogspot.com/> (дата звертання 25.05.2025)
4. G. E. Hinton, R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks", Science, 2006. 4. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative Adversarial Nets", NeurIPS 2014.
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y. (2014). Generative Adversarial Networks. Advances in Neural Information Processing Systems, 27.
6. Tatarchenko, M., Dosovitskiy, A., & Brox, T. (2017). Multi-view 3D Models from Single Images with a Convolutional Network. International Conference on 3D Vision (3DV).
7. Hartley, R., & Zisserman, A. Multiple View Geometry in Computer Vision. Cambridge University Press, 2003.
8. Wang, N. et al. Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images. ECCV, 2018.
9. Choy, C. et al. 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction. ECCV, 2016.
10. Park, J. et al. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. CVPR, 2019.

11. Mildenhall, B. et al. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. ECCV, 2020.
12. Tatarchenko, M. et al. Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs. ICCV, 2017.
13. ShapeNet Dataset. – URL: <https://www.shapenet.org> (дата звертання 25.05.2025)
14. PyTorch Documentation. – URL: <https://pytorch.org/docs> (дата звертання 25.05.2025)
15. Open3D Library. – URL: <http://www.open3d.org> (дата звертання 25.05.2025)
16. Google Colab. – URL: <https://colab.research.google.com> (дата звертання 25.05.2025)

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ ЗА НАУКОВИМИ НАПРЯМАМИ
КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ

1. Назаров О.С., Шураєв І.Д. “Автоматизація процесу створення 3D моделей на основі зображень.” – Міжнародна наукова інтернет-конференція "Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення" / Збірник тез доповідей: випуск 34, 2018. – С. 54-56.

2. Andriy Yerokhin, Valerii Semenets, Alina Nechyporenko, Oleksii Turuta, “F-transform 3D Point Cloud Filtering Algorithm” – Proc. of the 2th IEEE International Conference on Data Stream Mining & Processing. 21-25 August 2018, Lviv, Ukraine. – P.524-527. DOI: 10.1109/DSMP.2018.8478581.