Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Центр післядипломної освіти
(повна назва)
Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Порівняльний аналіз методів NeRF та 3D Gaussian Splatting для задачі 3D-реконструкції обличчя в естетичній медицині
(тема)

Виконав:
здобувач другого року навчання,
групи СШІздм-24-1
Роман ЛЕЩЕНКО
(власне ім'я, прізвище)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)
Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Системи штучного інтелекту
(повна назва освітньої програми)
Керівник доц. Каріна СЕЛІВАНОВА
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ШІ _____ Лариса ЧАЛА
(підпис) (власне ім'я, прізвище)

2026 р.

Харківський національний університет радіоелектроніки

Факультет Центр післядипломної освіти

Кафедра Штучного інтелекту

Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

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

Освітня програма Системи штучного інтелекту
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«_____» _____ 20 ___ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Лещенку Роману Олеговичу
(прізвище, ім'я, по батькові)

1. Тема роботи Порівняльний аналіз методів NeRF та 3D Gaussian Splatting для задачі 3D-реконструкції обличчя в естетичній медицині

затверджена наказом університету від 6 квітня 20 26 р. № 48Стз

2. Термін подання здобувачем роботи до екзаменаційної комісії 26 травня 20 26 р.

3. Вихідні дані до роботи Науково-технічні публікації, дані Інтернет-джерел та відомих наукових проектів щодо реконструкції зображень та 3Д сцен за допомогою NeRF та 3DGS, сучасні архітектури Neural Radiance Fields та 3DGS, тестувальні набори даних та функції розрахунку метрик якості.

4. Перелік питань, що потрібно опрацювати в роботі
1) Аналіз предметної галузі
2) Математичний огляд моделей
3) Розробка та реалізація експериментальних моделей
4) Експериментальні дослідження та аналіз результатів

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / термін виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	06.04.2026	виконано
2	Аналіз предметної галузі	06.04.2026 – 10.04.2026	виконано
4	Вибір моделей для дослідження	11.04.2026 – 15.04.2026	виконано
5	Програмна реалізація моделей та алгоритмів	16.04.2026 – 26.04.2026	виконано
6	Проведення експериментів	27.04.2026 – 03.05.2026	виконано
7	Підготовка пояснювальної записки	04.05.2026 – 17.05.2026	виконано
8	Надання записки на перевірку керівнику	18.05.2026 – 25.05.2026	виконано
9	Захист перед ЕК	26.05.2026	виконано

Дата видачі завдання 6 квітня 20 26 р.

Здобувач _____
(підпис)

Керівник роботи _____ доц. Каріна СЕЛІВАНОВА
(підпис) (посада, власне ім’я, прізвище)

РЕФЕРАТ

Пояснювальна записка: 68 с., 6 рис., 2 табл., 1дод., 62 джерела.

ЕСТЕТИЧНА МЕДИЦИНА, МЕДИЧНА ВІЗУАЛІЗАЦІЯ, НЕЙРОННІ ПОЛЯ ВИПРОМІНЮВАННЯ, ФОТОГРАММЕТРІЯ, ШТУЧНИЙ ІНТЕЛЕКТ, FASTAVATAR, HEADNERF, NERF, 3D GAUSSIAN SPLATTING, 3D-РЕКОНСТРУКЦІЯ.

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

Предмет дослідження – методи нейронних полів випромінювання (NeRF) та тривимірного гаусового розсіювання (3D Gaussian Splatting) для побудови 3D-моделей в естетичній медицині.

Мета роботи – порівняльний аналіз найбільш актуальних методів 3D-реконструкції обличчя у сфері естетичної медицини, а саме нейронних полів випромінювання (NeRF) та 3D Gaussian Splatting (3DGS).

Методи дослідження – методи комп'ютерного зору та машинного навчання, диференційований об'ємний рендеринг (ray marching, splatting), експериментальне моделювання простору, математичний та статистичний аналіз результатів за допомогою геометричних, фотометричних, перцептивних і клінічних метрик.

Виконано огляд сучасних архітектур 3D-реконструкції обличчя. Здійснено програмну реалізацію моделей HeadNeRF та FastAvatar. Проведено експериментальне моделювання та комплексний порівняльний аналіз цих методів за геометричними (RMSE, CD), фотометричними (PSNR, SSIM, LPIPS) та клінічними (LRE, ICC) метриками, а також за швидкістю рендерингу для визначення їхньої придатності у практичній естетичній медицині.

ABSTRACT

Master’s thesis contains: 68 pp., 6 fig., 3 tabl., 1 ann., 62 references.

AESTHETIC MEDICINE, ARTIFICIAL INTELLIGENCE, FASTAVATAR, HEADNERF, MEDICAL IMAGING, NERF, NEURAL RADIANCE FIELDS, PHOTOGRAMMETRY, 3D GAUSSIAN SPLATTING, 3D RECONSTRUCTION.

Object of research – the process of creating hyper-realistic 3D human face models.

Subject of research – Neural Radiance Fields (NeRF) and 3D Gaussian Splatting methods for 3D modeling in aesthetic medicine.

Purpose of the work – comparative analysis of the most relevant 3D face reconstruction methods in the field of aesthetic medicine, namely Neural Radiance Fields (NeRF) and 3D Gaussian Splatting (3DGS).

Research methods – computer vision and machine learning methods, differentiable volume rendering (ray marching, splatting), experimental spatial modeling, mathematical and statistical analysis of results.

A review of modern architectures for 3D face reconstruction was performed. Software implementation of HeadNeRF and FastAvatar models was carried out. Experimental modeling and a comprehensive comparative analysis of these methods were conducted using geometric (RMSE, CD), photometric (PSNR, SSIM, LPIPS), and clinical (LRE, ICC) metrics, as well as rendering time, to determine their practical suitability in aesthetic medicine.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів 7
Вступ 8
1 Аналіз предметної галузі 9
1.1 Область застосування, задачі 3D-представлення та фотограмметрії .. 9
1.2 Методи представлення 3D-простору 11
1.3 Використання 3D-реконструкції в медицині 16
1.4 Нейронні поля випромінювання (NeRF) 18
1.5 3D Gaussian Splatting (3DGS) 20
1.6 Постановка задачі 22
2 Математичний огляд моделей 24
2.1 NERF 24
2.2 Основні моделі вдосконалення NERF 28
2.3 Алгоритм HeadNeRF 29
2.4 3DGS – математичний опис та види реалізації 32
2.5 Метод SuGaR (Surface-Aligned Gaussian Splatting) 37
2.6 Метод FastAvatar 39
3 Розробка та реалізація експериментальних моделей 43
3.1 Огляд датасетів для навчання моделей 43
3.2 Опис середовища розробки та обраних інструментів для реалізації. 47
3.3 Реалізація експеременту порівняня моделей 49
4 Експериментальні дослідження та аналіз результатів 51
4.1 Геометричні метрики 51
4.2 Фотометричні та перцептивні метрики 52
4.3 Клінічні метрики 55
4.4 Аналіз результатів 56
Висновки 59
Перелік джерел посилання 60
Додаток А Відомість кваліфікаційної роботи 68

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

КТ – комп'ютерна томографія;

ШІ – штучний інтелект;

AR – Augmented Reality – доповнена реальність;

CAD – Computer-Aided Design – система автоматизованого проєктування;

CD – Chamfer Distance – відстань Чамфера;

FEA – Finite Element Analysis – метод скінченних елементів;

ICC – Intraclass Correlation Coefficient – внутрішньокласовий коефіцієнт кореляції;

LPIPS – Learned Perceptual Image Patch Similarity – перцептивна схожість ділянок зображення;

LRE – Landmark Error – похибка анатомічних орієнтирів;

MLP – Multi-Layer Perceptron – багатошаровий перцептрон;

MSE – Mean Squared Error – середньоквадратична помилка;

NeRF – Neural Radiance Fields – нейронні поля випромінювання;

PSNR – Peak Signal-to-Noise Ratio – пікове відношення сигнал/шум;

ReLU – Rectified Linear Unit – випрямлений лінійний блок;

RMSE – Root Mean Square Error – середньоквадратична похибка;

SfM – Structure from Motion – структура з руху;

SfS – Shape from Shading – форма із затінення;

SSIM – Structural Similarity Index Measure – індекс структурної подібності;

3DGS – 3D Gaussian Splatting – тривимірне гаусове розсіювання;

3DMM – 3D Morphable Models – тривимірні параметричні моделі.

ВСТУП

Сучасна естетична медицина стрімко переходить від використання плоских 2D-знімків до точних тривимірних цифрових моделей, які необхідні для передопераційного планування, лицьової гармонізації та створення AR-аватарів пацієнтів. Традиційні методи 3D-реконструкції часто виявляються неефективними або потребують іонізуючого випромінювання (КТ), тому перспективним напрямком є використання сучасних алгоритмів штучного інтелекту. Зокрема, методи нейронних полів випромінювання (NeRF) та тривимірного гаусового розсіювання (3D Gaussian Splatting) дозволяють синтезувати фотореалістичні 3D-моделі на основі розріджених наборів 2D-даних, що відкриває нові можливості для віртуального хірургічного планування.

Метою роботи є порівняльний аналіз найбільш актуальних методів 3D-реконструкції обличчя у сфері естетичної медицини – NeRF та 3D Gaussian Splatting. Для досягнення поставленої мети сформульовано такі завдання:

– дослідити моделі та методи побудови NeRF і 3DGS;

– обрати для реалізації моделі, які відповідають вимогам у сфері естетичної медицини та реалізувати обрані алгоритми;

– дослідити метрики, які показують якість 3D-моделей та їх застосовність у клінічній сфері;

– провести порівняння моделей за обраними метриками та проаналізувати результати.

Об’єктом дослідження: процес 3D-реконструкції обличчя для потреб естетичної медицини.

Предмет дослідження: моделі та алгоритми NeRF і 3D Gaussian Splatting для генерації та рендерингу тривимірних сцен.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Область застосування, задачі 3D-представлення та фотограмметрії

Фотограмметрія – це наука про отримання достовірної інформації навколишнього середовища і фізичних об’єктів та за допомогою процесу запису, вимірювання та інтерпретації фотографічних зображень і моделей електромагнітного випромінювання та інших явищ.

3D-представлення є фундаментом для багатьох сучасних технологічних досягнень, забезпечуючи можливість цифрового опису та візуалізації складних структур. Основними сферами застосування є:

– комп'ютерна графіка: створення реалістичних моделей та середовищ для індустрії розваг і симуляцій;

– віртуальна реальність (Virtual Reality, VR): розробка імерсивного досвіду для навчання, ігор та медицини;

– автономна навігація: побудова детальних карт середовища та детекція перешкод для безпечного руху безпілотного транспорту;

– медична візуалізація: детальне відображення анатомічних структур для діагностики та планування хірургічних втручань.

Предметна галузь фотограмметрії та 3D-візуалізації є складним і динамічно розвиненим напрямком, що поєднує методи комп’ютерного зору, комп’ютерної графіки, машинного навчання та інженерії. Ця галузь пройшла шлях від традиційних геометричних підходів до аналізу зображень до сучасних нейромережевих методів, здатних створювати фотореалістичні тривимірні сцени на основі двовимірних даних. Сучасна цифрова фотограмметрія базується на використанні цифрових зображень та обчислювальних систем для формування геометричної моделі об’єкта.

Генерація 3D-зображень на основі 2D-входів належить до класу задач inverse rendering, у межах яких виконується відновлення прихованих

параметрів 3D-сцени (геометрії, матеріалів, освітлення або структури сцени) за її двовимірними проєкціями. До цього класу належать:

– 3D reconstruction – відновлення форми та геометрії сцени;

– Novel view synthesis – синтез нових ракурсів;

– Depth estimation – оцінка карти глибини;

– Pose estimation – визначення положення камери чи об’єкта.

Еволюція методів 3D-представлення характеризується переходом від базових геометричних конструкцій до складних моделей, керованих даними. Перші підходи були орієнтовані на геометричні примітиви та просторові структури даних. Традиційно для розв’язання цієї задачі сформувалися два базові класи математичних моделей: геометричні та фотометричні.

Геометричні підходи базуються на законах проєктивної геометрії та встановленні відповідностей між різними кадрами. Фундаментальним алгоритмом у цьому напрямі є Structure from Motion, який за колекцією зображень дозволяє одночасно вирахувати траєкторію переміщення камери та побудувати розріджену 3D-хмару точок. Після етапу SfM зазвичай залучають метод Multi-View Stereo, що дозволяє відтворити щільну тривимірну геометрію сцени, спираючись на принцип фотометричної узгодженості пікселів.

Фотометричні моделі вивчають характер взаємодії світла з поверхнями об'єктів. Класичним прикладом є задача Shape from Shading, що полягає у реконструкції 3D-форми за одним знімком шляхом аналізу градієнтів затінення (залежності інтенсивності пікселя від кута падіння променів). Щоб уникнути неоднозначностей під час такої реконструкції, застосовують підхід Photometric Stereo, який використовує серію зображень об'єкта, зроблених з однієї позиції камери, але за різних умов освітлення.

Попри міцне математичне підґрунтя, традиційні моделі мають суттєві недоліки: геометричні методи (SfM, MVS) часто дають збій під час обробки дзеркальних або позбавлених вираженої текстури поверхонь, тоді як

фотометричні (SfS, PS) критично залежать від суворо контрольованих умов освітлення. Ці обмеження стали поштовхом до розробки новітніх методів, що спираються на використання нейронних мереж та диференційованого рендерингу.

1.2 Методи представлення 3D-простору

З розвитком обчислювальних потужностей і методів машинного навчання з'явилися більш складні та гнучкі способи представлення. Процес розробки можна представити через такий ієрархічний перехід:

– ранні методи та геометричні моделі: початкові підходи до 3D-представлення ґрунтувалися на явних геометричних описах, таких як полігональні сітки та конструктивна суцільна геометрія (CSG). Ці методи стали основою систем автоматизованого проєктування (CAD) та ранньої комп'ютерної графіки;

– об'ємні представлення: запровадження воксельних сіток дозволило створювати дискретне представлення 3D-простору, придатне для медичної візуалізації та обробки об'ємних даних. Однак воксельні сітки характеризуються високими витратами пам'яті та обмеженнями роздільної здатності;

– хмари точок (Point Cloud): хмари точок, які представляють об'єкти як набір дискретних точок у просторі, набули популярності з появою LiDAR та сенсорів глибини. Хоча вони ефективні для отримання сирих просторових даних, відсутність інформації про зв'язки між точками створює труднощі для подальшої обробки;

– методи на основі сіток (Mesh): полігональні сітки, що складаються з вершин, ребер і граней, забезпечують компактне та універсальне представлення. Вони широко застосовуються у комп'ютерній графіці та анімації. З часом такі методи були адаптовані для підтримки динамічних і деформованих моделей;

– неявні представлення: такі техніки, як Signed Distance Function (SDF), забезпечують безперервний та неявний спосіб опису поверхонь. SDF особливо корисні для виявлення зіткнень та аналізу форм;

– нейронні представлення: сучасні досягнення призвели до появи представлень на основі нейронних мереж, таких як Neural Radiance Fields (NeRF), які кодують 3D-сцени за допомогою нейронних мереж. Ці методи революціонізували задачі 3D-реконструкції та синтезу нових ракурсів;

– гібридні методи: поєднання кількох методів представлення, таких як DMTet, Tri-plane та 3D Gaussian Splatting, дозволяє використовувати переваги кожного підходу для ефективнішої роботи зі складними сценаріями.

Різноманітність сучасних 3D-даних підкреслюється використанням різних форматів, серед яких вокселі (Voxel), полігональні сітки (Mesh), хмари точок (Point cloud), октадерева (Octree), дескриптори (Descriptor), дані RGB-D, графи (Graph), проєкції (Projections). Візуалізація різних форматів представлення 3D-даних наведена на рисунку 1.1.

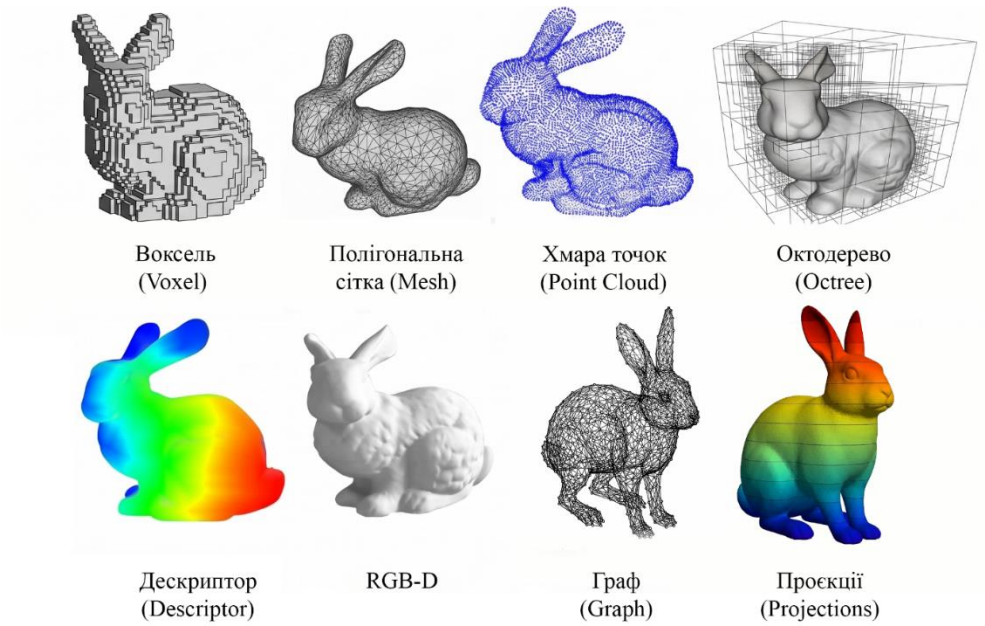


Рисунок 1.1 – Візуалізація різних форматів представлення 3D-даних

Представлення у вигляді хмари точок (Point Cloud) є одним із базових методів моделювання та аналізу тривимірних (3D) структур. Хмара точок являє собою набір точок у просторі, який описує зовнішню поверхню об’єкта або сцени. Зазвичай ці точки визначаються координатами у тривимірному просторі (x, y, z) і можуть додатково містити інформацію про колір, інтенсивність або нормальні вектори. Хмари точок створюються за допомогою технологій 3D-сканування, таких як LiDAR (Light Detection and Ranging), фотограмметрія та камери глибини, які забезпечують високоточне захоплення форми об’єктів і навколишнього середовища.

Однією з головних переваг хмар точок є їхня простота та безпосереднє представлення тривимірного світу. На відміну від інших способів 3D-представлення, таких як полігональні сітки (mesh) або воксельні ґратки (voxel grid), хмари точок не потребують інформації про зв’язки між елементами чи об’ємних даних, що спрощує їх створення та обробку. Водночас така простота створює певні труднощі. Хмари точок часто є розрідженими, неструктурованими та можуть містити шум, через що задачі сегментації, розпізнавання й реконструкції стають значно складнішими.

Представлення у вигляді полігональної сітки (Mesh) є одним із фундаментальних підходів у тривимірній комп’ютерній графіці та обчислювальній геометрії для опису поверхні 3D-об’єкта. Воно складається з вершин, ребер і граней, які формують поліедральну структуру. Вершини є точками у тривимірному просторі, ребра з’єднують пари вершин, а грані (зазвичай трикутники або чотирикутники) утворюються ребрами, що поєднують три або більше вершин.

Полігональні сітки набули широкого застосування завдяки оптимальному поєднанню простоти та виразності. Вони дозволяють із довільною точністю апроксимувати складні геометричні форми шляхом зміни кількості вершин і граней. Крім того, mesh-представлення добре підтримується сучасним графічним обладнанням, що забезпечує ефективність під час рендерингу та моделювання фізичних процесів.

Гнучкість цього підходу також дозволяє ефективно обчислювати характеристики поверхні, зокрема нормалі та кривину, які є важливими для реалістичної візуалізації та фізичних симуляцій.

Представлення за допомогою Signed Distance Function (SDF) широко використовується для опису водонепроникних (water-tight) форм і поверхонь тривимірних об’єктів. SDF кодує геометрію об’єкта шляхом визначення відстані від будь-якої точки простору до найближчої поверхні об’єкта, при цьому знак значення вказує, чи знаходиться точка всередині об’єкта (від’ємне значення), чи поза ним (додатне значення). Такий підхід забезпечує гладкий і неперервний опис поверхонь.

Однією з базових концепцій у цій галузі є використання SDF для моделювання неявних поверхонь, де поверхня визначається нульовим рівнем функції відстані. Цей метод дозволяє природно обробляти топологічні зміни, зокрема об’єднання та розділення поверхонь, що є особливо корисним у симуляціях динаміки рідин і деформівних об’єктів.

Neural Radiance Field (NeRF) став революційним підходом у галузі тривимірного представлення та синтезу нових ракурсів зображення. Цей метод кодує об’ємне представлення сцени у нейронній мережі, здатній генерувати нові види складних сцен на основі обмеженого набору вхідних зображень. Метод NeRF використовує повнозв’язну глибоку нейронну мережу для прогнозування кольору та щільності точки у просторі на основі її просторових координат і напряму спостереження.

Це дозволяє моделі відтворювати фотореалістичні зображення з нових точок огляду шляхом інтегрування передбачених значень уздовж променів, що проходять через сцену. Висока якість згенерованих NeRF-зображень зробила цей підхід одним із базових методів у сфері комп’ютерного зору та тривимірної комп’ютерної графіки.

3D Gaussian Splatting (3DGS) є інноваційним методом представлення для моделювання та рендерингу тривимірних сцен із використанням гаусових розподілів. Основна ідея 3DGS полягає у представленні сцени як

набору тривимірних гаусіан, які дозволяють ефективно апроксимувати геометрію та зовнішній вигляд сцени. Кожна гаусіана містить просторову та кольорову інформацію, забезпечуючи плавне й неперервне відтворення поверхонь і текстур.

Таке представлення має низку переваг порівняно з традиційними моделями на основі NeRF. Насамперед, 3DGS забезпечує можливість рендерингу в реальному часі, що є критично важливим для застосувань у віртуальній та доповненій реальності. Досягнення високої швидкості роботи пояснюється використанням неперервної природи гаусових функцій, які можуть ефективно відтворюватися за допомогою point-based методів і оптимізуватися через сучасні структури прискорення обчислень.

Deep Marching Tetrahedra (DMTet), запропонований Tianye Shen та співавторами, є сучасним методом тривимірного представлення, який розвиває концепцію marching tetrahedra – підходу для виділення ізоповерхонь. На відміну від класичних методів marching cubes або marching tetrahedra, що працюють на фіксованій сітці, DMTet використовує глибоке навчання для підвищення гнучкості та точності представлення.

В основі DMTet лежить нейронна мережа, яка прогнозує заповненість та зовнішній вигляд тривимірного простору. Як правило, мережа отримує на вхід латентний код або неявну функцію, що кодує форму та структуру об’єкта. Простір дискретизується на тетраедри – невеликі об’ємні елементи тетраедральної форми. Для кожної вершини тетраедра обчислюються значення заповненості, які визначають, чи знаходиться точка всередині або поза поверхнею об’єкта. Отримані значення використовуються для виділення поверхні шляхом інтерполяції між вершинами тетраедра, подібно до принципу роботи традиційних marching-методів, але з більш адаптивними та навченими межами поверхні.

Tri-plane є методом представлення тривимірного простору, який умовно поділяє його на три ортогональні площини, зазвичай вирівняні вздовж осей X, Y та Z, що дозволяє ефективно кодувати та відтворювати

3D-інформацію. Кожна площина відображає різні аспекти просторової структури, а разом вони формують цілісне представлення сцени або об’єкта.

Підхід tri-plane базується на припущенні, що складні тривимірні структури можна розкласти на простіші двовимірні проєкції, які згодом об’єднуються для формування повної 3D-моделі. Така декомпозиція спрощує обробку та маніпулювання тривимірними даними, зменшуючи розмірність задачі з трьох вимірів до двох. На практиці інформація про 3D-об’єкт або сцену кодується у трьох окремих двовимірних сітках або текстурах, кожна з яких відповідає одній із ортогональних площин.

Ці сітки можуть зберігати різноманітні атрибути, зокрема колір, глибину чи значення заповненості, які надалі можуть оброблятися стандартними методами двовимірної обробки зображень. Однією з головних переваг tri-plane представлення є здатність поєднувати високу обчислювальну ефективність із якістю 3D-реконструкції. Завдяки використанню внутрішньої структури тривимірних об’єктів і сцен цей метод забезпечує високий рівень деталізації та точності без необхідності складної та ресурсомісткої об’ємної 3D-обробки. Це робить tri-plane особливо придатним для застосувань у режимі реального часу, зокрема у віртуальній та доповненій реальності.

1.3 Використання 3D-реконструкції в медицині

Сучасна медицина перейшла від використання виключно плоских рентгенографічних знімків до точних тривимірних цифрових та фізичних моделей. 3D-реконструкція створює гіперреалістичні анатомічні орієнтири, які використовуються для передопераційного планування, створення індивідуальних імплантатів та візуалізації результатів хірургічних втручань.

У клінічній практиці головна мета 3D-реконструкції – мінімізація хірургічних ризиків, скорочення часу операції та забезпечення ідеального прилягання медичних виробів до анатомії пацієнта.

3D-технології в медицині, зокрема поверхневе сканування обличчя, дозволяють точно оцінювати об’єм і форму щелепно-лицевих структур без використання іонізуючого випромінювання. Вони значно покращують діагностику вроджених і набутих деформацій, а також забезпечують більш обґрунтоване планування хірургічних втручань. Такі підходи дають можливість об’єктивно відстежувати результати лікування у динаміці.

3D-моделювання на основі КТ дозволяє хірургам моделювати кістки лицьового черепа, які є надзвичайно тонкими, а травми часто призводять до багатоуламкових переломів, які загрожують порушенням бінокулярного зору. Лікарі використовують непошкоджену сторону обличчя як дзеркальний шаблон для відновлення симетрії.

В ортопедії 3D-друк та цифрова реконструкція використовуються для лікування складних переломів та створення пацієнт-специфічних імплантатів (PSI). На основі 3D-моделей кісток створюються індивідуальні шаблони для свердління та розпилу.

Віртуальне хірургічне планування (VSP) в ортогнатичній хірургії використовується для прогнозування переміщення щелеп (остеотомія Ле Фор I, BSSO). 3D-моделі піддаються методу скінченних елементів (FEA). Біомедичні інженери імітують фізіологічні навантаження (наприклад, ходьбу або жування), щоб перевірити, чи витримає індивідуальний металевий або полімерний імплантат тиск, і чи не призведе він до руйнування сусідньої кісткової тканини. Цей підхід є основою доказового дизайну імплантатів.

На відміну від загальної хірургії, де фокус іде на міліметрову точність кісткових структур, естетична медицина працює з м'якими тканинами, текстурою шкіри та психологією пацієнта. Тут 3D-реконструкція вирішує проблему узгодження очікувань пацієнта з реальними можливостями хірурга.

Системи з доповненою реальністю, дозволяють створювати 3D-аватар пацієнта. Це дозволяє хірургу у режимі реального часу модифікує

об'єми (наприклад, змінює проєкцію кінчика носа або імітує введення філерів), демонструючи пацієнту анатомічні обмеження, що спрощує планування ринопластики та лицьової гармонізації. Ці цифрові моделі конвертуються в 3D-друковані напрямні для проведення точних розрізів хрящів та кісток під час операції, а також в індивідуальні зовнішні назальні шини для післяопераційної фіксації.

Сучасні алгоритми штучного інтелекту, методи машинного навчання та інноваційні системи рендерингу, такі як нейронні поля випромінювання (NeRF) та тривимірне гаусове розсіювання (3DGS), дозволяють синтезувати фотореалістичні 3D-моделі на основі розріджених наборів 2D-даних, суттєво зменшуючи потребу в інвазивних процедурах та дозах іонізуючого опромінення.

1.4 Нейронні поля випромінювання (NeRF)

Запропонована у 2020 році архітектура нейронних полів випромінювання (NeRF) здійснила фундаментальний зсув у сфері комп'ютерного зору. NeRF принципово відрізняється від класичних методів тим, що не потребує попередньо заданої топології сітки. Натомість NeRF представляє сцену як безперервну (неявну) об'ємну функцію, яка апроксимується за допомогою повністю зв'язаної глибокої нейронної мережі – багатошарового персептрона (MLP). Мережа приймає на вхід безперервну 5D-координату, що складається з просторової локації точки та напрямку огляду камери.

Після процесу позиційного кодування MLP видає прогнозовану густину об'єму та колір для кожної точки простору (рисунок 1.2).

Для генерації пікселя нового зображення NeRF використовує алгоритми об'ємного рендерингу (ray marching), підсумовуючи колір та густину всіх точок вздовж променя зору. Цей математичний підхід дозволяє

створювати деталізовані фотореалістичні 3D-сцени з повним урахуванням оклюзій, складної геометрії та віддзеркалень.

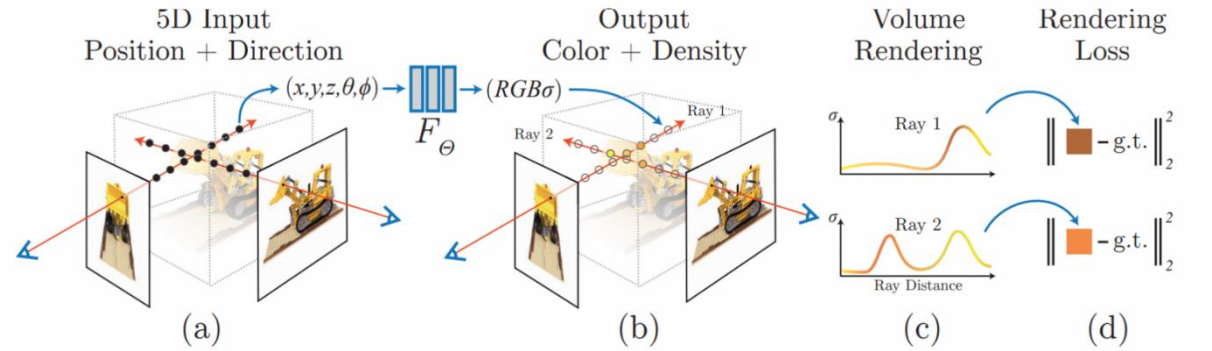


Рисунок 1.2 – Процес синтезу даних за домогою нейронних полів випромінювання

Дослідження Ruthberg та колег (2025) адаптувало NeRF для обробки монокулярного ендоскопічного відео з метою створення 3D-реконструкції операційного поля в режимі реального часу. Це знайшло застосування у ендоскопічній хірургії носових пазух. Для усунення проблеми відносності масштабу застосовувалася попередня калібрувальна матриця 1x1 мм. На кадаверних зразках NeRF-моделі були суміщені з післяопераційними КТ-сканами (які виступали як «ground truth»). Результати показали субміліметрову точність, що підтвердило здатність NeRF відстежувати критичні анатомічні зони (наприклад, поблизу орбіти чи внутрішньої сонної артерії) без використання громіздких стереокамер або зовнішніх систем трекінгу.

Генеративні варіанти NeRF (GNeRF) застосовуються для відновлення об'ємних 3D-моделей кісток або судин на основі розрідженого набору 2D-знімків. Щоб обійти етичні перешкоди зі збором даних, дослідники створюють цифрові реконструкційні рентгенограми зі збережених КТ-сканів, навчаючи NeRF відновлювати об'ємні дані. Це відкриває шлях до отримання КТ-подібних 3D-моделей пацієнта лише за кількома

стандартними рентгенівськими знімками, знижуючи дозу іонізуючого випромінювання.

Незважаючи на фотореалізм, методи на базі NeRF мають критичний недолік для медичного застосування – їхня геометрія є неявною (implicit). Оскільки об'єкт представлений як хмара щільності, з нього вкрай важко витягнути чітку, гладку полігональну сітку. Без полігональної сітки використання стандартних CAD-інструментів, застосовуваних у віртуальному хірургічному плануванні для виконання остеотомій, переміщення кісткових сегментів або розрахунку об'єму введеного філера, стає неможливим. Процес оптимізації та рендерингу вимагає надвисоких обчислювальних потужностей і займає години, що критично для невідкладної хірургії.

1.5 3D Gaussian Splatting (3DGS)

Розуміючи обчислювальні та геометричні обмеження NeRF, у 2023 році наукова спільнота представила новий метод – 3D Gaussian Splatting (3DGS). Метод 3DGS являє собою алгоритмічний підхід до задачі синтезу нових ракурсів та просторової реконструкції, що базується на явному об'ємному поданні. На відміну від неявних методів на базі нейронних полів випромінювання (наприклад, NeRF), 3DGS апроксимує геометрію та фотометрію сцени за допомогою множини анізотропних 3D-гауссіанів (Gaussian splats).

3D-гауссіани зберігаються у компактному форматі, де кожен сплат описується набором параметрів. Зазвичай до них належать:

– позиція: тривимірне положення центру сплату;

– коваріацію: масштаб та просторову орієнтацію;

– непрозорість: ступінь впливу сплату на сформоване зображення;

– колір сплату;

– додаткові характеристики матеріалу, такі як блиск, відбиття та заломлення світла.

Завдяки такому компактному поданню можливо представляти велику кількість сплатів із відносно невеликим обсягом даних. Це робить Gaussian splats особливо придатними для рендерингу складних сцен у високій роздільній здатності. На рисунку 1.3 показано, як тривимірні дані представляються за допомоги цього методу.



Рисунок 1.3 – Представлення тривимірної моделі за допомоги 3D-гауссіан

3DGS застосовує надзвичайно швидкий процес диференційованої растеризації (splatting). Гаусіани проєктуються з 3D-простору на 2D-площину камери та змішуються для формування кінцевого зображення.

Навчання 3D-гауссіанів – це процес оптимізації, де параметри гауссіанів підлаштовуються так, щоб при їх проєктуванні на 2D-площину вони максимально точно відтворювали набір тренувальних фотографій. Навчання зазвичай починається з ініціалізації розрідженої хмари точок (наприклад, отриманої за допомогою алгоритму Structure from Motion, такого як COLMAP). Алгоритм рендерить зображення з певної позиції, обчислює втрати (різницю) порівняно з реальною вхідною фотографією, і на основі цього оновлює параметри 3D-гаусіанів доти, доки синтезовані зображення не стануть ідентичними реальним. Алгоритм роботи представлено на рисунку 1.4.

Здатність методів 3DGS миттєво відмальовувати складні просторові об'єкти ідеально синхронізується з апаратними можливостями AR-гарнітур.

Це дозволяє створювати системи змішаної реальності (MR-DINS у стоматології або ортопедії), де хірург бачить голографічну 3D-модель судин і пухлини, яка з субміліметровою точністю накладена на реальне обличчя пацієнта. Відстеження інструментів у режимі реального часу допомагає уникнути пошкодження лицьового нерва або сонної артерії.

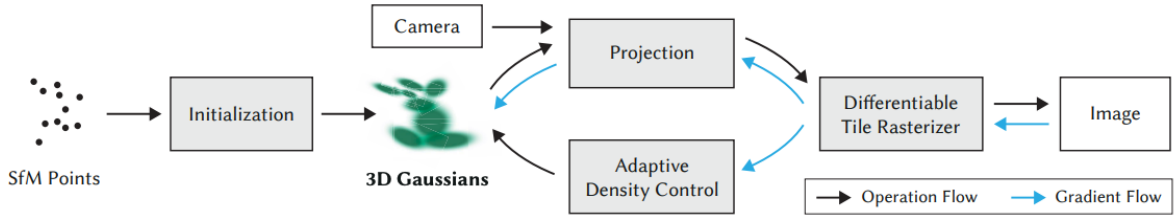


Рисунок 1.4 – Алгоритму роботи 3DGS

Оскільки 3DGS явно зберігає об'єкти як набір часток, хірурги можуть математично «відтинати» або редагувати певні гауссіани (наприклад, віртуально видаляти пухлину або зміщувати кістку), що є неможливим у NeRF.

Вже існують медицинські MedGS, SurgicalGS, UltraGS, фреймворки які використовують переваги 3DGS такі як швидкість відмальовування моделей або гнучкість налаштування гауссіан, що дозволяє використовувати метод з ультразвуковим та рентгенівським випромінбванням.

1.6 Постановка задачі

Метою магістерської роботи є порівняльний аналіз найбільш актуальних методів 3D реконструкції обличчя у сфері естетичної медицини, а саме 3D Gaussian Splatting та нейронні поля випромінювання (NeRF) . Для досягнення поставленої мети необхідно:

– дослідити моделі та методи побудови NeRF;

– дослідити моделі та методи побудови 3DGS;

– обрати для реалізації моделі NeRF та 3DGS які відповідають вимогам у сфері естетичної медицини;

– реалізувати обрані моделі;

– дослідити метрики та показники які показують якість отриманих 3D моделей та їх застосовність у клінічній сфері;

– провести порівняння моделей за обраними метриками;

– проаналізувати отримані результати та зробити висновки відносно моделей.

2 МАТЕМАТИЧНИЙ ОГЛЯД МОДЕЛЕЙ

2.1 NERF

В основі алгоритму NeRF лежить математичне формулювання, яке поєднує методи глибокого навчання з класичними законами фізичної оптики. Сцена не зберігається у вигляді дискретних об'єктів, а моделюється як неперервне поле випромінювання за допомогою багатошарового перцептрона (MLP).

Сцена представляється як неперервна векторнозначна функція (2.1), що відображає 3D-координату та 2D-напрямок огляду у локальну щільність та колір:

$$F_\Theta : (x, d) \longmapsto (c, \sigma), \tag{2.1}$$

де x = (x, y, z)– точка у просторі;

d = (θ, φ) – напрямок огляду;

c = (r , g, b) – колір;

σ – об'ємна щільність;

Θ – параметри мережі.

Щільність σ залежить лише від положення x (фізична властивість матерії), а колір c – додатково від напрямку d, що дозволяє моделювати складні фотометричні ефекти, такі як відблиски та напівпрозорість, де сприйманий колір змінюється при переміщенні камери.

Для синтезу двовимірного зображення на основі навченої безперервної 5D-функції NeRF використовує методи класичного об'ємного рендерингу (Classic Volume Rendering). Процес генерації починається з трасування променів. З оптичного центру віртуальної камери через кожен піксель площини ображення у тривимірний простір випускається промінь параметризований рівнянням прямої:

$$r(t) = o + t \cdot d, \tag{2.2}$$

де o – точка початку променя (позиція камери);

d – одиничний вектор напрямку променя;

t – скалярний параметр відстані.

Очікуваний колір $C(r)$ пікселя обчислюється класичним інтегралом об'ємного рендерингу, (2.3):

$$C(r) = \int_{t_n}^{t_f} T(t)\,\sigma\big(r(t)\big)\,c(r(t),\,d)\,dt, \tag{2.3}$$

де $T(t) = exp\left(-\int_{t_n}^{t} \sigma\big(r(s)\big)\,ds\right)$ функція пропускання виражає собою накопичене пропускання вздовж променя від точки t до точки t_n. З фізичної точки зору, $T(t)$ визначає ймовірність того, що фотон світла пролетить від t до t_n, не зіткнувшись із непрозорою частинкою.

Оскільки аналітичне обчислення цього безперервного інтеграла для функції, представленої складною нейронною мережею, є неможливим, алгоритм використовує чисельну апроксимацію за допомогою стратифікованої вибірки. Промінь розбивається на рівномірно розподілені інтервали. Всередині кожного інтервалу випадковим чином вибирається точкаю Така рандомізація гарантує, що в процесі багаторазових ітерацій навчання мережа буде оцінювати безперервний простір, а не дискретну сітку точок. Чисельне інтегрування перетворюється на формулу альфа-композитингу (alpha compositing), (2.4):

$$\hat{\mathrm{C}}\,(r) = \sum_{i=1}^{N} T_i \left(1 - e^{-\sigma_i \delta_i}\right) c_i, \tag{2.4}$$

де $T_i = exp\big(-\sum_{j=1}^{i-1} \sigma_j \delta_j\big)$;

$\delta_i = t_{i+1} - t_i$ – відстань між зразками, крок дискретизації;

$\alpha_i = 1 - exp(-\sigma_i \delta_i)$ – класична альфа-композиція.

Подача сирих декартових координат безпосередньо в багатошаровий перцептрон призводить до незадовільних результатів. Глибокі мережі зі стандартними функціями активації, такими як ReLU, мають так зване спектральне зміщення: вони прагнуть апроксимувати низькочастотні сигнали і є вкрай неефективними при вивченні високочастотних деталей, таких як складні текстури або різкі геометричні грані.

Для подолання цієї фундаментальної проблеми NeRF використовує метод позиційного кодування (Positional Encoding). Перед подачею в мережу вхідні координати проектуються у простір вищої розмірності за допомогою набору періодичних гармонічних функцій (синусів і косинусів) з експоненціально зростаючими частотами.

$$\gamma(p) = \left(sin(2^0\pi p),\ cos(2^0\pi p),\ \dots,\ sin(2^{L-1}\pi p),\ cos(2^{L-1}\pi p)\right). \quad (2.5)$$

В оригінальній статті NeRF використовують значення L=10 для γ(x) і L=4 для $\gamma(p)$.

Сліпий відбір точок уздовж променя є обчислювально неефективним, оскільки більша частина простору часто є порожньою, а оклюзовані області позаду щільних об'єктів не роблять внеску у фінальний колір пікселя. Оригінальний NeRF вирішує цю проблему за допомогою ієрархічного семплінгу (Hierarchical Volume Sampling), використовуючи дві нейронні мережі: «грубу» і «точну». Спочатку «груба» мережа оцінює набір з стратифікованих точок N_c, очікуваний колір для «грубої» нейромережі:

$$\widehat{C_c}(r) = \sum_{i=1}^{N_c} w_i c_i, \backslash emsp w_i = T_i(1 - exp\ exp\ (-\sigma_i\delta_i)\ ). \quad (2.6)$$

На основі отриманих ваг будується функція щільності ймовірності (PDF) вздовж променя. Нормалізовані ваги $\hat{w}_i = \frac{w_i}{\sum_{j=1}^{N_c} w_j}$, вказують на розподіл матерії. Потім, з використанням методу

вибірки оберненого перетворення (inverse transform sampling), алгоритм вибирає додаткові точок у тих регіонах, де наявність матерії є найбільш імовірною. «Точна» мережа обробляє об'єднаний набір точок для генерації фінального зображення.

Оптимізація обох мереж здійснюється за допомогою мінімізації середньоквадратичної помилки (Mean Squared Error, MSE) між кольорами, згенерованими мережами, та істинними кольорами пікселів з навчальної вибірки для всіх променів r у батчі R, функція втрат:

$$L = \sum_{r \in R} \left[\|\widehat{C_c}(r) - C(r)\|_2^2 + \|\widehat{C_f}(r) - C(r)\|_2^2 \right], \quad (2.7)$$

де $C(r)$ – істинний колір пікселя;

$\widehat{C_c}(r)$, $\widehat{C_f}(r)$ – прогнози грубої і точної мереж відповідно.

Попри високу якість синтезу зображень, базова архітектура алгоритму NeRF характеризувалася низкою фундаментальних обмежень, що перешкоджали її масштабуванню та інтеграції в системи реального часу.

До недоліків належала надмірна обчислювальна ресурсомісткість , а низька швидкість рендерингу на етапі інференсу робила алгоритм непридатним для інтерактивних застосунків. Використання нескінченно тонких променів для просторової вибірки зумовлювало появу виражених артефактів аліасингу за умов зміни масштабу або відстані від камери до об'єкта.

Значним обмеженням класичного підходу була відсутність здатності до генералізації просторових ознак. Реконструкція кожної нової сцени вимагала ініціалізації навчання моделі «з нуля» та наявності десятків прецизійно відкаліброваних ракурсів, що унеможливлювало коректну роботу методів в умовах дефіциту вхідних даних.

Парадигма оригінальної моделі базувалася на строгих припущеннях щодо абсолютної статичності середовища та непрозорості поверхонь. Це створювало суттєвий бар'єр для її імплементації в динамічних та

спеціалізованих доменах, зокрема в медицині, де моделювання деформацій тканин та об'ємна візуалізація внутрішніх анатомічних структур є критично важливими завданнями.

2.2 Основні моделі вдосконалення NERF

Для розв'язання проблеми появи артефактів аліасингу архітектура Mip-NeRF імплементувала концепцію міпмапінгу (mipmapping) – техніки, що традиційно використовується у класичних конвеєрах рендерингу для запобігання згладжуванню текстур. Замість трасування одновимірних променів, Mip-NeRF запровадив трасування конусів (рисунок 2.1). Кожен піксель зображення розглядається як основа конуса, що розширюється в міру просування вглиб сцени. Цей конус перетинає тривимірну геометрію, формуючи об'ємні ділянки у вигляді обмежувальних конічних зрізів.

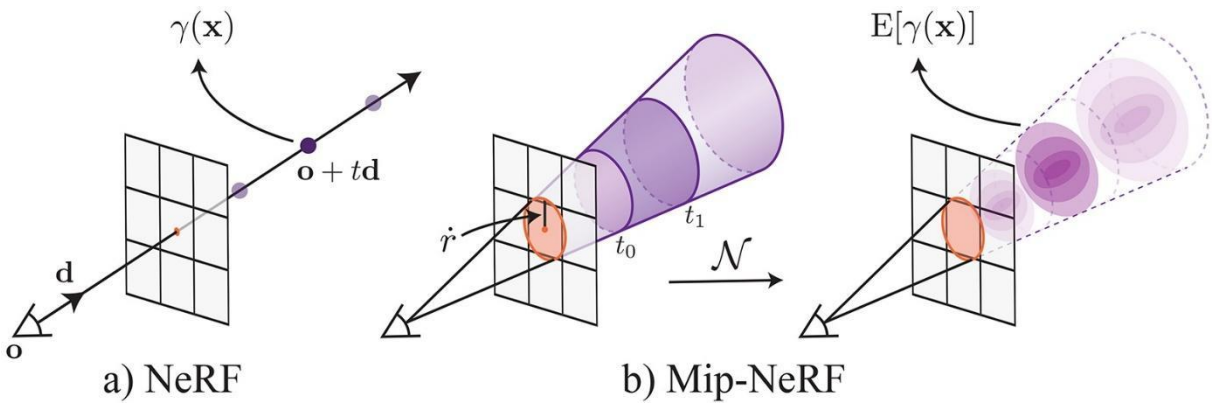


Рисунок 2.1 – Порівняння трасування одновимірних променів та конусів

Для того, щоб нейронна мережа могла обробляти ці об'ємні зрізи замість дискретних точок, розробники Mip-NeRF створили математичний механізм – інтегроване позиційне кодування (Integrated Positional Encoding, IPE). IPE обчислює математичне сподівання (інтеграл) цих функцій у межах усього об'єму конічного зрізу. Аналіз цього підходу виявляє глибинний фізичний зміст: якщо об'єм зрізу великий (камера знаходиться далеко),

позитивні та негативні значення високочастотних синусоїд взаємно компенсуються, і їхній інтеграл наближається до нуля.

Окрім інтеграції IPE, Mip-NeRF також спростив архітектуру мережі. У базовому NeRF на вхід MLP подавалася конкатенація позиційного кодування та самих «сирих» координат. Автори Mip-NeRF виявили, що ця конкатенація не робить значущого внеску у точність і лише ускладнює формалізацію IPE, тому вони повністю її видалили, спираючись виключно на ознаки інтегрованого позиційного кодування.

Модель Instant Neural Graphics Primitives (Instant-NGP) скоротила час навчання та збільшила швидкість рендерингу. Instant-NGP розділяє навчання NeRF на 3 основні складові та пропонує вдосконалення для кожної з них, що дозволяє здійснювати навчання NeRF у реальному часі. Ці 3 складові включають:

– покращений алгоритм навчання та рендерингу завдяки схемі проходження променів, де заміняється фіксоване позиційне кодування на навчану таблицю ознак з мультирівневим хеш-кодуванням;

– меншу, повністю об’єднану нейронну мережу;

– ефективне багаторівневе хеш-кодування на ба зі векторів ознак, яке заміняє позиційне кодування.

2.3 Алгоритм HeadNeRF

У задачах комп’ютерного зору та графіки параметричні моделі голови 3D Morphable Models (3DMM) є критично важливими для низьковимірного представлення морфології та експресії обличчя. Концептуально 3DMM представляє довільне обличчя як лінійну комбінацію базисних компонентів, отриманих методом аналізу головних компонент на наборі сканів, приведених у щільну вершинну відповідність. Модель розділяє обличчя на три ортогональних підпростори: ідентичність, що описує індивідуальну форму; експресія, що моделюється як додатковий

PCA-підпростір або набір FACS-блендшейпів; текстура, що описує колір і відбивні властивості шкіри.

FLAME – це артикульована модель голови, що поєднує статистичні переваги 3DMM з технікою LBS, запозиченою з моделі тіла SMPL. FLAME перевершує класичні 3DMM з кількох причин:

- явна артикуляція щелепи, шиї та очей;
- дезентенглінг ідентичності, пози й експресії;
- покриття повної голови з шиєю та внутрішньою поверхнею рота;
- диференційовність та інтеграція у PyTorch-пайплайни;
- коригувальні блендшейпи пози що усувають артефакти LBS.

Традиційні підходи, засновані на текстурованих мешах, обмежені топологічною жорсткістю та нездатністю моделювати складні структури, тоді як 2D GAN-моделі демонструють незадовільну багатокутну узгодженість через відсутність явного 3D-представлення. Стратегічним рішенням є інтеграція нейронних полів яскравості як 3D-проксі, що дозволяє неявно кодувати геометрію та зовнішній вигляд. Архітектура HeadNeRF спрямована на подолання обчислювальної складності оригінального NeRF, забезпечуючи фотореалістичний рендеринг у реальному часі зі збереженням параметричного та семантичним контролю.

В основі HeadNeRF лежить MLP-функція h_θ, яка інтегрує позиційне кодування координат із набором семантичних векторів. Архітектура розроблена таким чином, щоб забезпечити стабільність геометрії незалежно від умов освітлення.

HeadNeRF визначається як функція R, що генерує зображення I на основі параметрів камери P (включаючи внутрішні та зовнішні матриці) та чотирьох латентних кодів:

$$I = R\left(z_{id}, z_{exp}, z_{alb}, z_{ill}, P\right), \tag{2.8}$$

де z_{id} – ідентичність;

z_{exp} – вираз;

z_{alb} – альбедо;

z_{ill} – освітлення.

Для забезпечення геометричної стабільності h_θ реалізована через внутрішнє розщеплення. Спочатку функція приймає позиційне кодування $\gamma(x)$ та геометричні коди (z_{id}, z_{exp}) для обчислення щільності σ та проміжного вектора ознак. Потім цей вектор конкатенується з кодами зовнішнього вигляду (z_{alb}, z_{ill}) для отримання фінального 256-вимірного вектора ознак *F*:

$$h_\theta: \big(\gamma(x), z_{id}, z_{exp}, z_{alb}, z_{ill}\big) \mapsto (\sigma, F). \tag{2.9}$$

HeadNeRF виконує апсемплінг низькороздільної карти ознак (32x32) до фінального зображення (256x256). На відміну від GIRAFFE, де використовуються згортки 3x3, HeadNeRF застосовує Conv2D 1x1 у фінальних шарах. З погляду комп'ютерного зору, 3x3 згортки вносять просторові залежності між сусідніми пікселями, які не узгоджуються з 3D-геометрією променів. Використання 1x1 згорток діє як per-pixel MLP, зберігаючи незалежність променів та гарантуючи геометричну цілісність при зміні ракурсу.

Навчання здійснюється на 2D-даних за цільовою функцією:

$$L = L_{data} + L_{per} + L_{dis}, \tag{2.10}$$

де L_{data} – Photometric Loss, попіксельна відповідність у зоні маски голови, MSE між рендерингом та істиним зображенням;

L_{per} – Perceptual Loss, базується на активаціях VGG16. Цей компонент є вирішальним для синтезу дрібних деталей: мікрозморшок, текстури бороди та проміжків між зубами;

L_{dis} – Disentangled Loss, регуляризація латентних кодів відносно початкових значень 3DMM.

HeadNeRF дозволяє виконувати перенос експресії шляхом послідовної заміни латентного коду цільового об’єкта кодами, екстрагованими з референсного відео. Завдяки диспозиції ознак, ідентичність та умови освітлення залишаються інваріантними.

Попри високу ефективність, HeadNeRF має певні вразливості. Модель демонструє труднощі з аксесуарами, такими як головні убори, через їхню низьку представленість у навчальних вибірках. Крім того, умови освітлення у FaceSEIP обмежені лише чотирма фіксованими станами, що призводить до дискретності тіней при спробі плавної зміни коду.

2.4 3DGS – математичний опис та види реалізації

Метод 3D Gaussian Splatting (3DGS) запропонував перехід до явного представлення простору за допомогою дискретної множини геометричних примітивів – тривимірних гауссіанів. Ідея 3DGS ґрунтується на концепції «сплеттингу об'єму», вперше запропонованій для рендерингу поверхонь у 2001 році за допомогою еліптичних зважених середніх (Elliptical Weighted Average, EWA).

Замість нейронної мережі сцена описується великою множиною навчуваних тривимірних гаусових примітивів, кожен з яких має фізичний центр, повну матрицю коваріації, непрозорість та залежний від напрямку погляду колір.

Базовою одиницею представлення у 3DGS є тривимірний анізотропний гаусіан. Кожен такий примітив визначається у світовому просторі своїм центром (середнім значенням) та матрицею коваріації. Просторовий вплив гаусіана на довільну точку простору задається експоненційною функцією:

$$G(x) = exp\ exp\left(-\frac{1}{2}(x - \mu)^{\top}\Sigma^{-1}(x - \mu)\right). \quad (2.11)$$

У цьому виразі вектор центра μ позначає положення гаусіана, а матриця коваріації Σ розміром 3×3 визначає його форму, розмір та орієнтацію. Геометрично рівні значення цієї функції утворюють еліпсоїд: ізотропний гаусіан з коваріацією, пропорційною одиничній матриці, має форму кулі, тоді як анізотропний гаусіан – форму витягнутого у певному напрямку еліпсоїда. Саме анізотропність дозволяє небагатьом примітивам ефективно описувати видовжені структури сцени – ребра, тонкі поверхні, складки.

Безпосередня оптимізація матриці коваріації як вільного набору з шести чисел є проблематичною: у процесі градієнтного спуску матриця може втратити властивість додатної напіввизначеності й перестати бути коректною коваріацією. Щоб уникнути цього, у 3DGS коваріація параметризується через масштаб і обертання:

$$\Sigma = R\,S\,S^{\top}\,R^{\top}, \quad (2.12)$$

де S – матриця масштабування;

R – матриця обертання.

У цій декомпозиції діагональна матриця S задає масштабування й зберігається у вигляді тривимірного вектора, а матриця R задає обертання й зберігається як одиничний кватерніон. Така декомпозиція має наочну інтерпретацію: спочатку одинична куля розтягується вздовж осей координат відповідно до масштабів, а потім повертається у потрібну орієнтацію. За будь-яких значень параметрів результуюча матриця залишається коректною коваріацією, що робить оптимізацію стабільною.

Окрім центра, масштабу та обертання, кожен гаусіан зберігає скалярну непрозорість та параметри кольору.

Реальні поверхні виглядають по-різному з різних боків через відблиски, дзеркальні відбиття та підповерхневе розсіювання. Щоб модель могла відтворювати цей залежний від ракурсу вигляд, колір кожного гаусіана задається не сталим значенням RGB, а функцією від напрямку погляду, розкладеною за базисом сферичних гармонік:

$$c(d) = \sum_{l=0}^{L} \sum_{m=-l}^{l} k_l^m Y_l^m(d), \quad (2.13)$$

де k_l^m– навчувані коефіцієнти;

$Y_l^m(d)$ – функції сферичних гармонік;

l – ступень;

m – порядок.

Центр гаусіана переводиться зі світових координат у простір камери за допомоги формули:

$$\mu' = \pi(W\mu + t), \quad (2.14)$$

де π – перспективна проєкція;

W – матриця перегляду;

t – вектора зсуву.

Перетворення коваріації складніше. Перспективна проєкція є нелінійною через ділення на глибину, тому її не можна застосувати до коваріації безпосередньо. Метод спирається на класичну техніку EWA-сплатингу, перспективне перетворення локально наближається лінійним за допомогою матриці Якобі J. Двовимірна коваріація обчислюється так:

$$\Sigma' = J\,W\,\Sigma\,W^{\top}\,J^{\top}. \qquad (2.15)$$

Тут матриця W позначає лінійну частину матриці перегляду (обертання камери), а матриця Якобі J задає афінне наближення перспективної проєкції й залежить від фокусних відстаней та координат (x, y, z) центра гаусіана у просторі камери. Оскільки для формування зображення важлива лише двовимірна площина, з результуючої матриці беруть верхній лівий блок 2×2. Наближення працює тим точніше, чим менші гаусіани й чим більша фокусна відстань камери; для надто великих примітивів воно вносить помітне спотворення, що частково мотивує процедуру розділення гаусіанів, описану нижче.

Після проєкції всіх гаусіанів формується підсумкове зображення. Для кожного пікселя потрібно врахувати внесок усіх гаусіанів, що його перекривають, у порядку від ближнього до дальнього. Колір пікселя обчислюється накопиченням внесків гаусіанів спереду назад за схемою альфа-композитингу:

$$C(x) = \sum_{i \in N} c_i\,\alpha_i'\prod_{j=1}^{i-1}(1-\alpha_j'), \qquad (2.15)$$

де c_i – це RGB-колір i-го гауссіана (визначений через сферичні гармоніки);

α_i' – його адаптована непрозорість;

$\prod_{j=1}^{i-1}(1-\alpha_j')$ – добуток відображає кумулятивний коефіцієнт пропускання світла від попередніх шарів.

Ресурсна ефективність алгоритму досягається тим, що потік зупиняє розрахунок (Early Stop) одразу, як тільки акумульована непрозорість досягає насичення (наближається до 1.0).

Ключова інженерна знахідка 3DGS – тайлова растеризація. Екран розбивається на квадратні плитки розміром 16×16 пікселів. Гаусіани відсікаються за пірамідою видимості камери, після чого для кожної плитки

складається список примітивів, що її перекривають. Гаусіани сортуються за глибиною один раз на плитку, а не для кожного пікселя окремо. Растеризація кожної плитки виконується паралельно окремим блоком потоків графічного процесора: потоки завантажують гаусіани у спільну пам'ять і накопичують колір та непрозорість, проходячи список спереду назад.

Параметри всіх гаусіанів – центри, масштаби, обертання, непрозорості й коефіцієнти кольору – навчаються градієнтним спуском. На кожній ітерації обирається один із навчальних ракурсів, сцена рендериться з відповідної камери, а результат порівнюється з еталонним зображенням. Функція втрат поєднує попіксельну різницю з метрикою структурної подібності. Увесь конвеєр рендерингу диференційовний, тому градієнти функції втрат поширюються назад аж до параметрів кожного гаусіана. Для прискорення автори виводять градієнти явно, не покладаючись на автоматичне диференціювання.

Самого лише оновлення параметрів недостатньо: початкова розріджена хмара точок не може передати всі деталі сцени, а деякі гаусіани виявляються зайвими. Тому 3DGS періодично виконує адаптивний контроль щільності – три операції, керовані величиною градієнтів та непрозорістю.

Клонування застосовується там, де геометрія недостатньо покрита. Якщо невеликий гаусіан отримує великий градієнт позиції, поряд створюється його копія, яка зміщується у напрямку градієнта; це додає примітивів у деталізованих, але порожніх ділянках.

Розділення застосовується там, де один завеликий гаусіан намагається описати складну ділянку. Такий примітив замінюється двома меншими, розташованими в межах вихідного еліпсоїда; це уточнює представлення і зменшує спотворення афінного наближення проєкції.

Відсікання видаляє гаусіани, непрозорість яких опустилася нижче заданого порога, а також надмірно великі примітиви. Це стримує зростання моделі й прибирає примітиви, що не впливають на зображення.

2.5 Метод SuGaR (Surface-Aligned Gaussian Splatting)

У процесі стандартної оптимізації гауссіани прагнуть мінімізувати помилку рендерингу, часто утворюючи товсті «хмари» там, де має бути тонка поверхня. Це робить класичні методи екстракції поверхонь, такі як Marching Cubes, неефективними. Оскільки функція щільності в 3DGS дуже розріджена і зосереджена лише в центрах частинок, Marching Cubes на основі воксельної сітки видає зашумлений результат.

Метод SuGaR пропонує механізм регуляризації, що зближує реальну щільність гауссіан $d(p)$ з ідеальною моделлю поверхні. Ключовим моментом є перехід до SDF (Signed Distance Function) – знакової дистанційної функції.

Для виведення апроксимації SDF ми вводимо припущення про площинність гауссіан ($s_g \approx 0$). У такому випадку щільність у точці визначається переважно внеском найближчої гауссіани g^*. Ми знаходимо наменшу g^* за квадратом відстані Махаланобіса, яка враховує не лише лінійну відстань, але й форму та орієнтацію самої гаусіани:

$$g^* = arg\ arg\ \left\{ \left(p - \mu_g\right)^T \Sigma_g^{-1} \left(-\mu_g\right) \right\}, \tag{2.16}$$

де p – координати заданої точки у тривимірному просторі;

μ_g – вектор координат центру (середнього значення) поточної гаусіани;

Σ_g^{-1} – обернена матриця коваріації гаусіани.

Якщо гауссіани ідеально вирівняні та плоскі, ми можемо визначити «ідеальну» дистанційну функцію , що базується на поточному розподілі щільності:

$$f(p) = \pm s_{g^*}\sqrt{-2\ log\ log\ (d(p))}\,, \qquad (2.17)$$

де s_{g^*} – найменший фактор масштабування найближчої гауссіани g^*.

Для досягнення такого стану SuGaR накладає три умови регуляризації:

– Limited overlap, мінімізація сумарної щільності в точках між центрами гауссіан;

– Flatness, примусове зменшення одного з факторів масштабування до нуля;

– Opacity, використання ентропійних втрат для того, щоб гауссіани були або повністю непрозорими, або видалялися.

SuGaR проводить екстракція мешу за допомогою алгоритму Пуассона. Процес екстракції в SuGaR складається з наступних етапів:

Рендеринг карт глибини: система генерує карти глибини з ракурсів навчальних камер.

Точна вибірка точок: для кожного пікселя на промені зору v вибираються точки в межах довірчого інтервалу $[-3\sigma_g(v), 3\sigma_g(v)]$, де $\sigma_g(v)$ – стандартне відхилення гауссіани в напрямку камери. Це дозволяє точно знайти перетин променя з ізоповерхнею.

Обчислення нормалей: використовується аналітичний градієнт щільності $\nabla d(p)$, який дає точний вектор нормалі до поверхні.

Глобальна реконструкція: алгоритм Пуассона створює замкнений меш, що є стійким до шумів та пропусків у даних.

Це дозволяє отримати деталізовану топологію, яка зберігає дрібні елементи сцени, недоступні для воксельних методів.

Фінальним і найбільш інноваційним етапом SuGaR є Binding – прив’язка нових плоских гауссіан до трикутників отриманого мешу. Це створює гібридну модель, де геометрія представлена полігонами, а візуальна складова – фотореалістичними гауссіанами. При цьому позиція кожної гауссіани жорстко зафіксована відносно вершин трикутника тобто координати барицентричні. Також замість складних кватерніонів для обертання використовується одне комплексне число, що обмежує ротацію гауссіани площиною трикутника. При деформації або анімації мешу масштабування гауссіан автоматично корегується через відношення середньої довжини сторін деформованого трикутника до оригінального. Це запобігає розривам «текстури» при розтягуванні моделі.

2.6 Метод FastAvatar

Метод FastAvatar вирішує дві фундаментальні проблеми, що залишалися: висока обчислювальна вартість пер-суб'єктної оптимізації та залежність якості від ракурсу вхідного зображення. FastAvatar – це пов'язно-незалежний прямопоширювальний (pose-invariant feed-forward) фреймворк, що відновлює повний 3DGS-аватар обличчя з єдиного зображення довільного ракурсу приблизно за секунди.

FastAvatar представляє собою гібридну архітектуру, яка поєднує стабільність апріорних геометричних знань із гнучкістю диференційованого рендерингу, долаючи розрив між швидкістю feed-forward мереж та якістю попіксельної оптимізації. Архітектура складається з трьох логічних компонентів: шаблонна модель 3DGS обличчя, кодувальник-декодувальник та необов'язковий етап test-time refinement.

Шаблонна модель 3DGS обличчя – статистично усереднений «середній» аватар, побудований з тренувального набору суб'єктів. Для кожного з N тренувальних суб'єктів окремо оптимізується власний 3DGS-аватар, що складається з анізотропних гаусіанів. Початкові позиції цих

гаусіанів задаються тривимірними ландмарками FLAME-меша, який підгоняється до зображень суб'єкта; це гарантує, що в усіх суб'єктів k-й гаусіан позиційно відповідає одній і тій самій точці обличчя (наприклад, кутку губ або центру брови).

Сцена або об'єкт моделюється набором з K анізотропних гаусіан. Кожен гаусіан G_k у моделі повністю описується п'ятьма групами параметрів:

– центром $\mu_k \in \mathbb{R}^3$;

– коваріаційною матрицею $\Sigma_k = R_k S_k S_k^T R_k^T$;

– непрозорістю $\alpha_k \in [0,1]$;

– коефіцієнтами сферичних гармонік c_k.

Рендеринг виконується класичним альфа-композитингом упорядкованих за глибиною гаусіанів. Після того як усі N індивідуальних моделей оптимізовано, формується шаблонна модель T: для кожного індексу k параметри статистично усереднюються по всіх N суб'єктах. Отриманий шаблон є геометрично та апієрансно середнім обличчям – сильним пріором, навколо якого мережа надалі вивчає індивідуальні відхилення.

Кодувальник-декодувальник, що отримує одне зображення обличчя на вході й передбачає параметричні зсуви до кожного гаусіана шаблонної моделі. Система використовує MLP-декодер для прогнозування дельт (Δ). Оновлення параметрів кожного Гаусіана k виконується за наступними формулами:

$$\mu_k = \mu_k^T + \Delta\mu_k \,, \tag{2.18}$$

$$s_k = s_k^T \cdot exp\ exp\ (\Delta s_k) \,, \tag{2.19}$$

$$q_k = (q_k^T + \Delta q_k), \tag{2.20}$$

$$\alpha_k = \sigma\left((\alpha_k^T) + \Delta\alpha_k\right), \tag{2.21}$$

$$c_k = c_k^T + \Delta c_k. \tag{2.22}$$

Такий механізм гарантує, що масштаб і прозорість завжди залишатимуться у фізично коректних межах завдяки експоненційному та сигмоїдальному перетворенням.

Для запобігання колапсу моделі до «середнього обличчя» ми застосовуємо стратегію відокремленого навчання:

– навчання декодера на латентних кодах, що дозволяє йому вивчити простір можливих деформацій обличчя;

– навчання енкодера з використанням ознак ArcFace. Ми мінімізуємо Cosine Similarity Loss між передбаченим кодом та еталонним , що змушує енкодер ігнорувати ракурс і фокусуватися виключно на ідентичності особи.

Необов'язковий етап test-time refinement, що додатково оптимізує приблизно 300 ітерацій (≈3 с) для адаптації до тонких особистих рис, окулярів, прикрас тощо. Алгоритм уточнення під час інференсу:

– енкодер генерує початковий латентний код з вхідного фото;

– виконується швидка оптимізація коду та параметрів зовнішнього вигляду.

Оскільки кожен Гаусіан прив’язаний до топології FLAME, ми використовуємо Linear Blend Skinning (LBS). Трансформація Гаусіанів (позиція, ротація, масштаб) розраховується через ваги скіннінгу вершин сітки, що забезпечує плавне відтворення міміки в реальному часі.

До обмежень можна выднести: якість залежить від різноманіття тренувального набору: суб'єкти з рисами, відсутніми у тренувальній вибірці (рідкісні етнічні типи, виражена асиметрія, нестандартні зачіски), будуть гірше представлені у шаблоні. Шаблонна модель базується на FLAME-параметризації, що добре описує саму голову, але обмежено працює з аксесуарами, прикрасами та екстремальними виразами – їх

компенсує лише test-time refinement. Метод реконструює статичний аватар; для анімації потрібно додатково керувати ним через FLAME-параметри, що потенційно зменшує точність динаміки. Нарешті, обчислення вимагає сучасного графічного процесора рівня NVIDIA A100 для повної швидкості 10 мс – на споживчому обладнанні час інференса буде вищим.

Завдякі FastAvatar у контексті естетичної медицини відкривається можливість миттєвого створення фотореалістичної 3D-моделі пацієнта з єдиного знімка, з подальшим редагуванням рис обличчя у латентному просторі для симуляції результатів хірургічних втручань – напрямок, що логічно продовжує дослідження точної геометрії, описані в наступному підрозділі.

3 РОЗРОБКА ТА РЕАЛІЗАЦІЯ ЕКСПЕРИМЕНТАЛЬНИХ МОДЕЛЕЙ

3.1Огляд датасетів для навчання моделей

Вибір датасету – це важливий етап у розробці будь-якого рішення на базі машинного навчання. Архітектура моделі визначає механізм того, як вона вчиться, але саме датасет визначає, чого вона навчиться. Академічна спільнота активно попвнюється новими наборами зображень для 3Д реконструкції облич. Розглянемо основні зних.

Набір даних NeRSemble розроблений як високоточна багатокамерна платформа для побудови динамічних нейронних полів радіації людських голів з урахуванням часу. Архів містить понад 4700 синхронізованих відео послідовностей надвисокої роздільної здатності для більше ніж 220 унікальних осіб. Зйомка проводилася в лабораторній студії на масив із 16 каліброваних камер промислового класу з частотою 73 кадри на секунду. Набір супроводжується тривимірними MVS-реконструкціями, щільними хмарами точок та високоякісними картами глибини. Головною метою авторів було надати дослідникам безпрецедентну базу для вивчення міміки, емоцій та поворотів голови з усіх ракурсів одночасно. Також у датасет інтегровано прецизійний 3D-трекінг у топології FLAME, виконаний за допомогою сучасного алгоритму VHAP.

FaceScape є масштабною базою даних тривимірних моделей обличчя, розробленою Університетом Нанкіну для навчання генеративних алгоритмів та деталізованої 3D-реконструкції. Набір містить 18 760 унікальних тривимірних сканів, отриманих від 938 суб'єктів, де кожен знятий у 20 фіксованих мімічних виразах. До складу датасету входять як сирі хмари точок з камер, так і топологічно уніфіковані (TU) 3D-сітки та прецизійні карти зміщень, що фіксують мікротекстуру пір шкіри. Додатково FaceScape пропонує власну білінійну параметричну модель 3DMM для

спрощення інтеграції геометричних апріорі. Ціль створення проекту полягала у подоланні дефіциту якісних 3D-даних з покроковою мімічною деталізацією для глибокого навчання.

RenderMe-360 одна з найбільших публічних 4D-бібліотек цифрових активів, розробленою для подолання обмежень існуючих баз даних та прискорення досліджень у сфері створення реалістичних 3D-аватарів голови. Датасет містить понад 243 мільйони кадрів та понад 800 тисяч відеопослідовностей, отриманих від 500 унікальних суб'єктів різного віку, статі та етнічного походження. Зйомка проводилася в студійній системі з 60 синхронізованих 2K-камер із частотою 30 кадрів на секунду, що забезпечує повний 360-градусний огляд голови. База даних супроводжується багатим набором анотацій: калібрувальними параметрами камер, масками сегментації фону, 3D-сканами, 2D/3D ключовими точками обличчя та готовим трекінгом параметрів FLAME. Основна ціль проекту полягала у створенні універсального майданчика для тестування п'яти ключових завдань: синтезу нових ракурсів, анімації міміки, рендерингу та редагування волосся, а також генерації розмовних голів.

NoW Challenge (Not-quite-in-the-Wild) розроблений Sanyal, Bolkart, Feng і Black у Max Planck Institute for Intelligent Systems і представлений разом із моделлю RingNet на CVPR 2019. Набір даних був створений з метою встановлення єдиного об'єктивного стандарту для оцінювання точності алгоритмів тривимірної реконструкції обличчя за поодинокими монокулярними зображеннями в неконтрольованих («диких») умовах та усунення методологічних розбіжностей під час порівняння різних 3DMM-алгоритмів за допомогою єдиної системи оцінювання. До його складу входять 2 054 високоякісні 2D-фотографії 100 унікальних суб'єктів (розподілені на 20 осіб для валідації та 80 для тестування). Кожне зображення фіксує людину під різними кутами зйомки, з різною мімікою, в умовах складного природного освітлення та з типовими перешкодами (оклюзіями) на кшталт окулярів чи волосся. Еталонними

даними виступають високоточні тривимірні скани голів, отримані за допомогою професійних 3D-сканерів.

Набір даних Multiface був розроблений дослідницькою лабораторією Reality Labs (Meta AI) з метою подолання дефіциту високоякісних багатокамерних даних у відкритому академічному середовищі та стимулювання досліджень у сфері нейрорендерингу облич та VR-телеприсутності. База даних містить деталізовані записи синхронізованих відеопослідовностей надвисокої роздільної здатності для 13 унікальних суб’єктів, кожен з яких відтворює понад 100 різноманітних мімічних виразів. Для зйомки використовувалася спеціалізована купольна установка Mugsy, обладнана щільним масивом камер (до 150 кольорових камер у версії v2) з роздільною здатністю 4096x2668 пікселів та частотою 30/90 кадрів на секунду. Комплект постачання включає синхронізовані RGB-зображення, калібрувальні параметри камер, аудіодоріжки, а також повністю відстежені динамічні 3D-сітки та розгорнуті карти текстур. Основною метою проекту є надання дослідникам прецизійного інструментарію для розробки реалістичних цифрових аватарів.

ILSH – Imperial Light-Stage Head (Imperial College London, ICCV 2023 Workshop). Набір даних ILSH був розроблений дослідниками Імперського коледжу Лондона спеціально для підтримки академічних змагань та тестування алгоритмів синтезу нових ракурсів голови людини в умовах складного освітлення. До його складу входять дані для 52 унікальних суб'єктів, зафіксованих у професійній установці Light Stage, що складається з геодезичного купола діаметром 2.5 метри. Зйомка здійснювалася на 24 синхронізовані камери промислового класу Balser з роздільною здатністю 4K (4096x3000 пікселів) за одночасної роботи всіх 82 джерел білого світла на базі RGBW-діодів. Датасет надає дослідникам 1 248 деталізованих портретних зображень (RAW-файли), точні калібрувальні матриці камер та маски меж обличчя. Головною метою проекту є надання високоякісних 2D-фотографій з унікальним розподілом тіней та світла для

перевірки здатності нейромереж моделювати складну фотометричну структуру шкіри.

FaceVerse є спеціалізованою базою даних, створеною для побудови та навчання дрібнозернистої 3D-морфованої моделі обличчя з можливістю тонкого керування геометричними деталями. Набір даних має гібридну структуру та складається з двох частин: великомасштабної вибірки з 60 000 злитих RGB-D зображень від споживчих датчиків глибини та високоточної вибірки з 2 000 щільних 3D-сканів голів. Усі дані зібрані виключно серед представників азіатської етнічної групи, охоплюючи широкий спектр вікових категорій, гендерного розподілу та мімічних виразів. Ціль створення FaceVerse полягала у вирішенні проблеми низької узагальнювальної здатності західних 3DMM-моделей на азіатських обличчях та забезпеченні можливості нелінійного редагування мікрозморшок. Проект надає готовий фреймворк на базі диференційованого рендерингу для підгонки параметрів моделі за поодинокими 2D-зображеннями.

Набір даних INSTA був створений з метою розширення можливостей моделювання та реконструкції об'ємних 3D-аватарів голови людини в реальному часі за допомогою технологій нейронних полів та Gaussian Splatting. Він складається з високоякісних монокулярних відеопослідовностей, знятих на мобільні пристрої, для 10 унікальних суб'єктів, які демонструють активну та різноманітну міміку. Головною особливістю датасету є наявність прецизійно сегментованих масок переднього плану, де фон повністю залитий чорним кольором для полегшення тренування нейромереж. Усі кадри супроводжуються попередньо розрахованими параметрами моделі FLAME, що забезпечує стабільну ініціалізацію геометрії. Ціль створення цього ресурсу полягала у спрощенні розробки та тестування алгоритмів відтворення міміки в умовах одного ракурсу камери.

У роботі було використано FaceScape датасет який надає різноматні зображення для опрацювання у тому числі і в різних позах і кутах зйомки.

3.2 Опис середовища розробки та обраних інструментів для реалізації

Python виступає ключовою мовою реалізації проєкту. Python – це високорівнева, інтерпретована мова програмування загального призначення з динамічною суворою типізацією. Розроблена наприкінці 1980-х років Гвідо ван Россумом.

Мова є мультипарадигмальною: вона нативно підтримує об'єктно-орієнтоване, імперативне (процедурне) та меншою мірою функціональне програмування. Головним проектним принципом Python є підвищення читабельності вихідного коду та продуктивності розробника.

Для візуалізації обраного зображення, рендерів та прорахованих метрик обрано десктопний GUI PyQt6. PyQt – це набір Python-прив'язок (bindings) для кросплатформного фреймворку Qt, який оригінально розроблений мовою C++. PyQt є одним із найпопулярніших інструментів для розробки графічних інтерфейсів користувача (GUI) у середовищі Python, оскільки дозволяє поєднати продуктивність і багатство компонентів C++-фреймворку з високорівневим синтаксисом Python.

Для обробки зображень окрім стандартних методів Python використано бібліотеку OpenCV (Open Source Computer Vision Library). Це кросплатформна бібліотека алгоритмів комп'ютерного зору, обробки зображень та чисельного аналізу. Базова кодова база реалізована на C++ з акцентом на оптимізацію продуктивності у реальному часі. Також використовується scikit-image. Це бібліотека алгоритмів обробки зображень з відкритим вихідним кодом для мови програмування Python. Вона є частиною розширеної екосистеми SciPy та проєктувалася з орієнтацією на наукові, інженерні та медичні дослідження. На відміну від OpenCV, яка оптимізована для задач комп'ютерного зору у реальному часі, scikit-image

розробляється з пріоритетом на алгоритмічну прозорість, точність реалізації еталонних математичних методів та чистоту Python-коду.

Для обчислення метрик для порівняння 3D моделей використані бібіліотеки як загального призначення так і спеціалізовані.

SciPy (Scientific Python) – це відкрита бібліотека високорівневих математичних алгоритмів та зручних функцій для наукових та інженерних обчислень. Вона структурно надбудовується над бібліотекою NumPy і розширює її можливості, надаючи повноцінний інструментарій для чисельного інтегрування, оптимізації, інтерполяції, обробки сигналів та розв'язання диференціальних рівнянь.

NumPy (Numerical Python) – це базова бібліотека для наукових обчислень у середовищі Python. Її головним завданням є подолання архітектурних обмежень продуктивності інтерпретатора Python (зокрема, динамічної типізації та накладних витрат під час виконання циклів) при здійсненні інтенсивних математичних операцій. Бібліотека досягає цього шляхом делегування обчислень на рівень скомпільованого коду (написаного на C, C++ та Fortran).

SciPy та NumPy використовуються у програмі для роботи з векторами, прорахунки специфічних ознак у алгоритмі FastAvatar, та інші методи бібліотек.

MediaPipe Face Landmarker – це високорівневий компонент бібліотеки MediaPipe Tasks (розробленої корпорацією Google), призначений для вирішення задач комп'ютерного зору, пов'язаних із локалізацією та аналізом облич на зображеннях або у відеопотоці. Інструмент базується на моделях глибокого навчання і забезпечує інференс у режимі реального часу з оптимізацією для кінцевих пристроїв (edge computing). Використовується для метрики LRE – наскільки лендмарки на згенерованому обличчі збігаються з оригіналом.

PyTorch – це фреймворк для тензорних обчислень та машинного навчання, розроблений лабораторією FAIR. torchvision є офіційним

розширенням PyTorch, яке надає базовий функціонал для обробки візуальних даних. Ці компоненти використовуються для обчислення LPIPS (perceptual distance через AlexNet features).

Для розробки програмного застосунку було обрано середовище PyCharm. PyCharm є інтегрованим середовищем розробки (IDE) для мови програмування Python. Воно було створене компанією JetBrains і є одним з найбільш популярних інструментів для розробки на Python. Середовище має багато функцій, що полегшують роботу з Python, таких як автодоповнення коду, рефакторинг, налагоджування, візуальне моделювання та підтримка віртуальних середовищ. Інтерфейс PyCharm простий і зрозумілий, але в той же час має багато можливостей для налаштування і розширення.

3.3 Реалізація експеременту порівняня моделей

Для реалізації експеременту 3D-реконструкції та моделювання обличчя було обрано алгоритми FastAvatar та HeadNeRF. Зазначені підходи мають фундаментальну архітектурну схожість: обидва алгоритми базуються на парадигмі неявних нейронних репрезентацій (зокрема, Neural Radiance Fields) і спрямовані на створення фотореалістичних, анімованих 3D-моделей людської голови. Спільною рисою є використання параметричних моделей обличчя як апріорної геометричної бази. Це дозволяє алгоритмам декомпозувати процес рендерингу та незалежно керувати параметрами ідентичності, міміки та просторового положення.

З метою мінімізації обчислювальних витрат та прискорення збіжності моделей (model convergence) під час навчання, було вирішено не ініціалізувати мережі випадковими значеннями. Для спрощення імплементації та забезпечення відтворюваності результатів, базові ваги нейронних мереж було інтегровано з офіційних репозиторіїв відповідних проєктів. Такий підхід дозволив зосередити обчислювальні ресурси на етапі тонкого налаштування та адаптації моделей під специфічні умови

поточного дослідження. На рисунку 3.1 представлено приклад рендеру методів від ісходного зображення до рендеру HeadNeRF та FastAvatar.

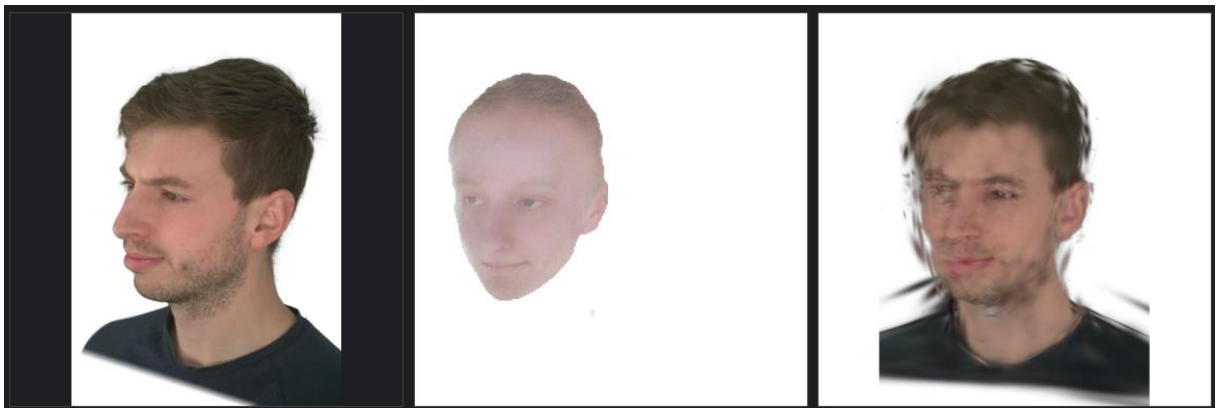


Рисунок 3.1 – Візуалізація роботи алгоритмів

Для проведення повноцінних тестів та порівнянь моделей розроблено окреме ізольоване віртуальне середовище. Воно спирається на стекз підтримкою CUDA. CUDA – це програмно-апаратна архітектура паралельних обчислень, що дозволяє істотно збільшити обчислювальну продуктивність завдяки використанню графічних процесорів NVIDIA. CUDA дозволяє програмістам реалізовувати на спеціальному спрощеному діалекті мови C алгоритми, які використовуються в графічних процесорах NVIDIA, і включати спеціальні функції в текст програми на C. Архітектура CUDA дозволяє розробнику на свій розсуд організовувати доступ до набору інструкцій GPU і керувати його пам'яттю. Ця технологія підтримує кілька мов програмування. Серед них Java, Python і деякі інші.

4 ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

4.1 Геометричні метрики

Геометричні метрики (RMSE, Chamfer Distance) оцінюють точність відновленої тривимірної форми у фізичному просторі – наприклад, у міліметрах. Вони порівнюють реконструйовану поверхню чи хмару точок з еталонною геометрією і є критичними там, де геометрія важливіша за зовнішній вигляд.

RMSE (Root Mean Square Error) вимірює середнє геометричне відхилення реконструйованої моделі від еталонної. Для тривимірної поверхні вона обчислюється як квадратний корінь із середнього квадрата евклідових відстаней між зіставленими точками двох моделей:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \| p_i - q_i \|^2}, \quad (4.1)$$

де p_i – точка реконструйованої моделі;

q_i – найближча відповідна точка еталонної поверхні (зіставлення зазвичай виконується алгоритмом ICP);

N – кількість зіставлених пар точок.

Результат виражається у фізичних одиницях (міліметрах для облич) або у відносних безрозмірних одиницях для зображень. Метрика завжди невід’ємна; ідеальне значення дорівнює нулю, і чим менше значення, тим точніша реконструкція. Перевагою RMSE є інтуїтивна фізична інтерпретація та чутливість до систематичних відхилень, а основними обмеженнями – залежність від попереднього вирівнювання моделей і висока чутливість до викидів, оскільки піднесення до квадрата підсилює великі помилки. Для оцінки NeRF та 3DGS RMSE застосовують після вилучення

поверхні (полігональної сітки) з реконструкції, що для 3DGS є нетривіальною задачею.

Chamfer Distance – базова геометрична метрика для порівняння двох хмар точок, інваріантна до порядку точок. Вона дорівнює сумі двох доданків: для кожної точки першої хмари береться квадрат відстані до найближчої точки другої хмари, і навпаки, після чого обидва доданки усереднюються:

$$CD(A,B) = \frac{1}{|A|}\sum_{a\in A} \min_{b\in B} \| a-b \|^2 + \frac{1}{|B|}\sum_{b\in B} \min_{a\in A} \| b-a \|^2, \quad (4.2)$$

де A – реконструйована хмара точок;

B – еталонна;

|A| та |B| – їхні розміри.

Метрика невід'ємна, найкраще значення – нуль, а її масштаб залежить від нормалізації сцени. Перевагами є просте й швидке обчислення за допомогою просторових індексів та диференційовність, завдяки якій CD використовують і як функцію втрат, і як метрику оцінювання. Головне обмеження – висока чутливість до викидів і «плаваючих» точок: окрема ізольована точка значно погіршує значення метрики. Це робить пряме застосування CD до 3DGS проблематичним, адже гаусіани після оптимізації розташовані неструктуровано й містять артефакти-флоутери; коректне обчислення CD потребує попереднього вилучення впорядкованої поверхні спеціалізованими методами.

4.2 Фотометричні та перцептивні метрики

Метрики на основі зображень оцінюють якість двовимірного рендеру з нового ракурсу порівняно з опорним знімком. Вони не звертаються до геометрії безпосередньо, а порівнюють зображення попіксельно (PSNR), за

структурною подібністю (SSIM) або за перцептивною схожістю у просторі ознак глибокої мережі (LPIPS).

PSNR (Peak Signal-to-Noise Ratio)– найпоширеніша метрика якості рендеру, що логарифмічно виражає попіксельну схожість двох зображень. Вона обчислюється через середньоквадратичну похибку між зображеннями:

$$PSNR = 10 \cdot log_{10}\left(\frac{MAX^2}{MSE}\right),\ MSE = \frac{1}{MN}\sum_{i=1}^{M}\ \sum_{j=1}^{N}\ (I(i,j) - K(i,j))^2, \quad (4.3)$$

де I – опорне зображення;

K – відрендероване;

M×N – їхній розмір;

MAX – максимально можливе значення пікселя (255 для 8-бітних зображень або 1 для нормалізованих).

Результат виражається в децибелах; ідентичні зображення дають нескінченність, а вищі значення відповідають кращій якості. Перевагами PSNR є простота, швидкість обчислення та єдина зрозуміла шкала, що робить її зручним орієнтиром. Проте метрика погано узгоджується з людським сприйняттям: два зображення з однаковим PSNR можуть виглядати помітно по-різному, оскільки метрика не враховує структурних спотворень і семантики.

SSIM (Structural Similarity Index Measure) – індекс структурної подібності SSIM оцінює структурну, а не попіксельну подібність зображень, спираючись на припущення, що зорова система людини чутлива саме до структурної інформації. Метрика поєднує три складники – схожість за яскравістю, контрастом і структурою – і обчислюється у ковзному вікні:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (4.4)$$

де μ_x, μ_y – локальні середні яскравості;

$\sigma_x^2,\ \sigma_y^2$ – локальні дисперсії;

σ_{xy} – коваріація фрагментів x та y;

C_1, C_2 – малі константи стабілізації, що залежать від динамічного діапазону.

Значення SSIM лежать у діапазоні від −1 до 1 (на практиці від 0 до 1); одиниця відповідає ідентичності зображень, а вищі значення – кращій якості. Перевагою SSIM є краща за PSNR кореляція з суб'єктивною оцінкою якості та чутливість до локальних структурних артефактів. Обмеженнями є слабка чутливість до сильних кольорових спотворень і залежність результату від розміру вікна.

LPIPS (Learned Perceptual Image Patch Similarity) вимірює значення перцептивної схожості між зображеннями. В основі LPIPS лежить ідея, що людська зорова система сприймає зображення з точки зору локальних ділянок, а не глобальних особливостей.

Вона обчислює схожість між локальними ділянками двох зображень, порівнюючи представлення їх глибинних ознак. Ці глибинні ознаки отримуються шляхом пропускання ділянок через попередньо навчену глибоку нейронну мережу. LPIPS має кілька переваг над іншими метриками якості зображень.

По-перше, вона базується на глибоких нейронних мережах, які довели свою високу ефективність у відображенні людської інтерпретації. По-друге, має високу точність і добре корелює із людським сприйняттям, що робить його надійною метрикою для оцінки якості зображення.

Нарешті, LPIPS є дуже гнучким і може бути легко адаптуватися до різних застосувань шляхом точного налаштування глибокої нейронної мережі. Для оцінки моделей в цьому досліджені використовувалася метрика LPIPS основана на AlexNet.

4.3 Клінічні метрики

У цьому розділі моделі представлені метрики які виміріють точність за антропараметричними метриками або точність за зразковою моделю. Це критично саме у медицині бо демострують відповідність 3D моделі облиичю пацієнта та показають можливість застосування моделей у практичній медицині.

Похибка анатомічних орієнтирів Landmark Error (LRE) оцінює точність відновлення положень окремих анатомічно значущих точок (орієнтирів) і, на відміну від інтегральних метрик, дозволяє локалізувати помилку в клінічно важливих зонах. Усереднена по всіх орієнтирах похибка обчислюється як середнє евклідових відстаней між відповідними точками реконструкції та еталона:

$$LRE = \frac{1}{K}\sum_{k=1}^{K} \| p_k^{rec} - p_k^{gt} \|_2, \quad (4.5)$$

де p_k^{rec} та p_k^{gt} – координати k-го орієнтира на реконструкції та на еталоні відповідно;

K – кількість орієнтирів.

Часто застосовують нормалізовану версію, де похибку ділять на характерний розмір обличчя (наприклад, міжочну відстань), що робить метрику безрозмірною та порівнянною між об’єктами різного масштабу. Метрика невід’ємна, найкраще значення – нуль; вона безпосередньо пов’язана з антропометричними вимірюваннями – лінійними відстанями та кутами між орієнтирами. Перевагою є клінічна інтерпретованість і можливість регіонального аналізу, а обмеженнями – залежність від точності ідентифікації самих орієнтирів (яка має власну похибку) та обмежена кількість стандартних точок, що не покриває «гладких» ділянок поверхні.

Внутрішньокласовий коефіцієнт кореляції Intraclass Correlation Coefficient (ICC) оцінює не точність реконструкції, а надійність і

відтворюваність вимірювань, виконаних на ній: наскільки узгоджуються повторні вимірювання того самого об’єкта, отримані різними експертами або в різні моменти часу.

Коефіцієнт виражається через компоненти дисперсії за результатами дисперсійного аналізу; одна з найпоширеніших форм – двофакторна модель з випадковими ефектами для абсолютної узгодженості – має вигляд:

$$ICC = \frac{MS_R - MS_E}{MS_R + (k-1)\,MS_E + \frac{k}{n}(MS_C - MS_E)}, \quad (4.6)$$

де MS_R – середній квадрат між об’єктами;

MS_C – середній квадрат між експертами або повтореннями;

MS_E – залишковий середній квадрат;

n – кількість об’єктів;

k – кількість вимірювань на об’єкт.

Значення ICC лежать у діапазоні від 0 до 1: за усталеними рекомендаціями значення нижче 0,5 свідчить про погану надійність, від 0,5 до 0,75 – про помірну, від 0,75 до 0,9 – про хорошу, а понад 0,9 – про відмінну. Перевагою ICC є статус загальноприйнятого стандарту в медицині та психометрії й наявність довірчих інтервалів. Головне обмеження полягає в тому, що існує десять різних форм ICC з відмінними припущеннями, тому в звіті завжди слід явно зазначати використану форму, інакше результат можна хибно інтерпретувати.

4.4 Аналіз результатів

Перші прогони моделей у бенчмарку показали натсупні результати, наведені у таблиці 4.1.

Таблиця 4.1 – Порівняльні метрики на сирих даних

	HeadNeRF	FastAvatar
CD	5.9	52.659
RMSE	58.856	149.474
PSNR	14.291	4.781
SSIM	0.726	0.378
LPIPS	0.313	0.843
LRE	0.029	0.556
ICC	0.999	0.456
time	29.026	3.577

Точність 3D-геометрії (CD, RMSE): показники похибки побудови геометрії у FastAvatar вищі у порядок разів (CD) порівняно з HeadNeRF. Це вказує на серйозні розриви у 3D-сітці або хмарі точок. Модель FastAvatar не змогла коректно відновити форму об'єкта.

Теж саме можно сказати за точність зображення – значення PSNR та SSIM вказують що FastAvatar не зберіг структуру обєкта на відміну від HeadNeRF. Ідентічність та мімка теж не збережені. Хоча HeadNeRF показує не погані результати але не великої якості. FastAvatar – зовсім не справляється з задачею.

Все говорить про те що необхідно доначвати моделі. Після донавчання найкращі результати виглядають більш релевантними для оцінки (таблиця 4.2).

Якість картинки (Візуальне сприйняття) FastAvatar показав кращий PSNR (16.42 проти 14.29), тобто математично попіксельна різниця кольорів у нього краща. HeadNeRF кращий за SSIM (0.918) та LPIPS (0.048). Це означає, що око людини сприйме результат HeadNeRF як набагато чіткіший і реалістичніший.

Таблиця 4.2 – Порівняльні метрики на тренерованих даних

	HeadNeRF	FastAvatar
CD	2.77	6.23
RMSE	13.34	39.11
PSNR	14.291	16.42
SSIM	0.918	0.809
LPIPS	0.048	0.276
LRE	0.013	0.030
ICC	0.999	0.999
time	29.026	3.577

Точність 3D-геометрії (CD, RMSE) Геометрія FastAvatar (RMSE 39.11) вийшла набагато «грубішою» і менш точною. Він гірше вловив глибину обличчя і загальну форму голови.

Міміка та ідентичність (LRE, ICC) Обидва алгоритми впоралися зі збереженням особистості на відмінно – ICC 0.999 означає, що на рендері та сама людина, що й в оригіналі. Однак HeadNeRF точніше передав рухи губ та очей (LRE 0.013 проти 0.030).

Продуктивність (time) Головний вагомий козир FastAvatar. Він відпрацював за 3.577 проти 29.026 у HeadNeRF (приблизно у 8.1 раза швидше).

Дані показують що HeadNeRF видає красиву, чітку картинку з хорошою геометрією та мімікою, але робить це довго. FastAvatar відпрацьовує швидко і зберігає впізнаваність людини, але розплачується за це втратою 3D-деталей та «замилюванням» фінального зображення.

Також слід зауважити що дані отримані від наявного датасету відрізняється від заявлених показників дослідників. Враховуючи покращення результатів після донавчання можна предположити що можна отримати кращі результати більш відповідні заявленним. Але навіть заявленні показують гірші результати у збережені геометрій і структури обличчя ніж HeadNeRF.

ВИСНОВКИ

У результаті виконання магістерської кваліфікаційної роботи було успішно вирішено актуальну задачу – проведено порівняльний аналіз сучасних методів 3D-реконструкції обличчя для потреб естетичної медицини, а саме нейронних полів випромінювання (на прикладі HeadNeRF) та 3D Gaussian Splatting (на прикладі FastAvatar). Завдання дослідження виконано в повному обсязі: проаналізовано математичні моделі, обрано інструментарій, здійснено програмну реалізацію та проведено комплексне тестування розроблених рішень.

Алгоритм HeadNeRF продемонстрував значно вищу точність 3D-геометрії (RMSE 13.34 проти 39.11 у FastAvatar) та кращу якість синтезованого зображення за перцептивними метриками (SSIM 0.918, LPIPS 0.048). Це означає, що HeadNeRF створює деталізованішу картинку, що критично для точної медичної діагностики.

Обидва алгоритми відмінно впоралися зі збереженням ідентичності пацієнта (показник надійності ICC склав 0.999). Проте HeadNeRF виявився точнішим у передачі мікрорухів та орієнтирів (похибка LRE 0.013 проти 0.030).

Головною перевагою методу FastAvatar (3DGS) стала швидкість рендерингу – час його роботи склав 3.577 с, що приблизно у 8,1 раза швидше за показники HeadNeRF (29.026 с). Однак така швидкодія досягається ціною втрати дрібних просторових деталей та часткового «замилювання» фінального результату.

Отримані результати відкривають перспективи для розробки гібридних архітектур, які б поєднували геометричну точність методів NeRF зі швидкістю рендерингу та можливістю редагування 3DGS. Подальші дослідження також мають бути спрямовані на розширення навчальних вибірок (датасетів) для покращення роботи моделей з нестандартними анатомічними особливостями та медичними аксесуарами.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. BARF: bundle-adjusting neural radiance fields / C.-H. Lin et al. *2021 IEEE/CVF international conference on computer vision (ICCV)*, Montreal, QC, Canada, 10–17 October 2021. 2021. URL: https://doi.org/10.1109/iccv48922.2021.00569 (date of access: 20.03.2026).

2. Mip-NeRF: a multiscale representation for anti-aliasing neural radiance fields / J. T. Barron et al. *2021 IEEE/CVF international conference on computer vision (ICCV)*, Montreal, QC, Canada, 10–17 October 2021. 2021. URL: https://doi.org/10.1109/iccv48922.2021.00580 (date of access: 20.03.2026).

3. SurgicalGS: dynamic 3D gaussian splatting for accurate robotic-assisted surgical scene reconstruction / J. Chen et al. *Lecture notes in computer science*. Cham, 2025. P. 572–582. URL: https://doi.org/10.1007/978-3-032-05141-7_55 (date of access: 11.05.2026).

4. Wenzheng C., Jun G. Learning to predict 3D objects with an interpolation-based differentiable renderer. *arXiv.org*. URL: https://arxiv.org/abs/1908.01210 (date of access: 11.05.2026).

5. 3D gaussian splatting vs nerf: the end game of 3D reconstruction?. *PyImageSearch*. URL: https://pyimagesearch.com/2024/12/09/3d-gaussian-splatting-vs-nerf-the-end-game-of-3d-reconstruction/ (date of access: 11.05.2026).

6. Curless B., Levoy M. A volumetric method for building complex models from range images. *The 23rd annual conference*, Not Known. New York, New York, USA, 1996. URL: https://doi.org/10.1145/237170.237269 (date of access: 11.05.2026).

7. Zhekov G. 3D Gaussian Splatting: a new frontier in rendering. *Chaos Blog*. URL: https://blog.chaos.com/3d-gaussian-splatting-new-frontier-in-rendering (date of access: 11.05.2026).

8. Gaussian splatting: 3D reconstruction and novel view synthesis, a review / A. Dalal et al. *IEEE access*. 2024. P. 1. URL: https://doi.org/10.1109/access.2024.3408318 (date of access: 11.05.2026).

9. DeepView: view synthesis with learned gradient descent / J. Flynn et al. *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Long Beach, CA, USA, 15–20 June 2019. 2019. URL: https://doi.org/10.1109/cvpr.2019.00247 (date of access: 11.05.2026).

10. DeepVoxels: learning persistent 3D feature embeddings / V. Sitzmann et al. *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Long Beach, CA, USA, 15–20 June 2019. 2019. URL: https://doi.org/10.1109/cvpr.2019.00254 (date of access: 11.05.2026).

11. Differentiable Monte Carlo ray tracing through edge sampling / T.-M. Li et al. *ACM transactions on graphics*. 2019. Vol. 37, no. 6. P. 1–11. URL: https://doi.org/10.1145/3272127.3275109 (date of access: 12.05.2026).

12. Differentiable volumetric rendering: learning implicit 3D representations without 3D supervision / M. Niemeyer et al. *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Seattle, WA, USA, 13–19 June 2020. 2020. URL: https://doi.org/10.1109/cvpr42600.2020.00356 (date of access: 12.05.2026).

13. Density estimation using Real NVP. *arXiv.org*. URL: https://arxiv.org/abs/1605.08803 (date of access: 12.05.2026).

14. GNeRF: gan-based neural radiance field without posed camera / Q. Meng et al. *2021 IEEE/CVF international conference on computer vision (ICCV)*, Montreal, QC, Canada, 10–17 October 2021. 2021. URL: https://doi.org/10.1109/iccv48922.2021.00629 (date of access: 12.05.2026).

15. Global illumination with radiance regression functions / P. Ren et al. *ACM transactions on graphics*. 2013. Vol. 32, no. 4. P. 1–12. URL: https://doi.org/10.1145/2461912.2462009 (date of access: 12.05.2026).

16. Accurate, large minibatch SGD: training imagenet in 1 hour. *arXiv.org*. URL: https://arxiv.org/abs/1706.02677 (date of access: 12.05.2026).

17. Henzler P., Mitra N. J., Ritschel T. Learning a neural 3D texture space from 2D exemplars. *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Seattle, WA, USA, 13–19 June 2020. 2020. URL: https://doi.org/10.1109/cvpr42600.2020.00838 (date of access: 12.05.2026).

18. IEEE transactions on pattern analysis and machine intelligence. *IEEE transactions on pattern analysis and machine intelligence*. 2004. Vol. 26, no. 4. P. 0_2. URL: https://doi.org/10.1109/tpami.2004.1265858 (date of access: 12.05.2026).

19. Hierarchical representations and explicit memory: learning effective navigation policies on 3D scene graphs using graph neural networks / Z. Ravichandran et al. *2022 IEEE international conference on robotics and automation (ICRA)*, Philadelphia, PA, USA, 23–27 May 2022. 2022. URL: https://doi.org/10.1109/icra46639.2022.9812179 (date of access: 12.05.2026).

20. HeadNeRF: a realtime nerf-based parametric head model / Y. Hong et al. *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, New Orleans, LA, USA, 18–24 June 2022. 2022. URL: https://doi.org/10.1109/cvpr52688.2022.01973 (date of access: 12.05.2026).

21. Image quality assessment: from error visibility to structural similarity / Z. Wang et al. *IEEE transactions on image processing*. 2004. Vol. 13, no. 4. P. 600–612. URL: https://doi.org/10.1109/tip.2003.819861 (date of access: 13.05.2026).

22. The application of generative neural radiance fields in medical imaging: a systematic review / T. Kabika et al. *Medical physics*. 2026. Vol. 53, no. 2. URL: https://doi.org/10.1002/mp.70296 (date of access: 13.05.2026).

23. Kajiya J. T., Von Herzen B. P. Ray tracing volume densities. *ACM SIGGRAPH computer graphics*. 1984. Vol. 18, no. 3. P. 165–174. URL: https://doi.org/10.1145/964965.808594 (date of access: 13.05.2026).

24. Kanatani K., Sugaya Y., Kanazawa Y. Bundle adjustment. *Guide to 3D vision computation*. Cham, 2016. P. 149–161. URL: https://doi.org/10.1007/978-3-319-48493-8_11 (date of access: 13.05.2026).

25. NeRSemble: multi-view radiance field reconstruction of human heads / T. Kirschstein et al. *ACM transactions on graphics*. 2023. Vol. 42, no. 4. P. 1–14. URL: https://doi.org/10.1145/3592455 (date of access: 13.05.2026).

26. Kutulakos K. N., Seitz S. M. A theory of shape by space carving. *Proceedings of the seventh IEEE international conference on computer vision*, Kerkyra, Greece, 20–27 September 1999. 1999. URL: https://doi.org/10.1109/iccv.1999.791235 (date of access: 13.05.2026).

27. A Survey of 3D Reconstruction: The Evolution from Multi-View Geometry to NeRF and 3DGS / S. Liu et al. *Sensors*. 2025. Vol. 25, no. 18. P. 5748. URL: https://doi.org/10.3390/s25185748 (date of access: 13.05.2026).

28. Local deep implicit functions for 3D shape / K. Genova et al. *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Seattle, WA, USA, 13–19 June 2020. 2020. URL: https://doi.org/10.1109/cvpr42600.2020.00491 (date of access: 13.05.2026).

29. Local implicit grid representations for 3D scenes / C. Jiang et al. *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Seattle, WA, USA, 13–19 June 2020. 2020. URL: https://doi.org/10.1109/cvpr42600.2020.00604 (date of access: 13.05.2026).

30. Neural volumes / S. Lombardi et al. *ACM transactions on graphics*. 2019. Vol. 38, no. 4. P. 1–14. URL: https://doi.org/10.1145/3306346.3323020 (date of access: 13.05.2026).

31. MedGS: gaussian splatting for multi-modal 3D medical imaging. *arXiv.org*. URL: https://arxiv.org/abs/2509.16806 (date of access: 14.05.2026).

32. MedGS: learning 3D structures from sequential medical imaging with gaussian splatting. *arXiv.org e-Print archive*. URL: https://arxiv.org/html/2509.16806v4 (date of access: 14.05.2026).

33. NeRF: representing scenes as neural radiance fields for view synthesis. *arXiv.org*. URL: https://arxiv.org/abs/2003.08934 (date of access: 14.05.2026).

34. E-NeRF: neural radiance fields from a moving event camera / S. Klenk et al. *IEEE robotics and automation letters*. 2023. P. 1–8. URL: https://doi.org/10.1109/lra.2023.3240646 (date of access: 14.05.2026).

35. NeRF: representing scenes as neural radiance fields for view synthesis / B. Mildenhall et al. *Computer vision – ECCV 2020*. Cham, 2020. P. 405–421. URL: https://doi.org/10.1007/978-3-030-58452-8_24 (date of access: 14.05.2026).

36. SG-NeRF: sparse-input generalized neural radiance fields for novel view synthesis / K. Xu et al. *Journal of computer science and technology*. 2024. URL: https://doi.org/10.1007/s11390-024-4157-6 (date of access: 14.05.2026).

37. Neural radiance fields in medical imaging: challenges and next steps. *arXiv.org e-Print archive*. URL: https://arxiv.org/html/2402.17797v1 (date of access: 14.05.2026).

38. NoW challenge. *NoW*. URL: https://now.is.tue.mpg.de/ (date of access: 14.05.2026).

39. RenderMe-360: a large digital asset library and benchmarks towards high-fidelity head avatars / W. Cheng et al. *Advances in neural information processing systems 36*, New Orleans, Louisiana, USA, 10–16 December 2023. San Diego, California, USA, 2023. P. 7993–8005. URL: https://doi.org/10.52202/075280-0350 (date of access: 24.05.2026).

40. PixelNeRF: neural radiance fields from one or few images / A. Yu et al. *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Nashville, TN, USA, 20–25 June 2021. 2021. URL: https://doi.org/10.1109/cvpr46437.2021.00455 (date of access: 14.05.2026).

41. Point-NeRF: point-based neural radiance fields / Q. Xu et al. *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, New Orleans, LA, USA, 18–24 June 2022. 2022. URL: https://doi.org/10.1109/cvpr52688.2022.00536 (date of access: 15.05.2026).

42. ACM, Siggraph, Staff A. SIGGRAPH 1998 conference proceedings: computer graphics annual conference series. Pearson Education, 1998. 472 p.

43. Waheed U. B. Kronecker neural networks overcome spectral bias for PINN-based wavefield computation. *IEEE geoscience and remote sensing letters*. 2022. P. 1. URL: https://doi.org/10.1109/lgrs.2022.3209901 (date of access: 15.05.2026).

44. RegNeRF: regularizing neural radiance fields for view synthesis from sparse inputs / M. Niemeyer et al. *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, New Orleans, LA, USA, 18–24 June 2022. 2022. URL: https://doi.org/10.1109/cvpr52688.2022.00540 (date of access: 15.05.2026).

45. Rosinol A., Leonard J. J., Carlone L. NeRF-SLAM: real-time dense monocular SLAM with neural radiance fields. *2023 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Detroit, MI, USA, 1–5 October 2023. 2023. URL: https://doi.org/10.1109/iros55552.2023.10341922 (date of access: 15.05.2026).

46. Neural radiance fields (nerf) for 3D reconstruction of monocular endoscopic video in sinus surgery / J. S. Ruthberg et al. *Otolaryngology–Head*

and neck surgery. 2025. URL: https://doi.org/10.1002/ohn.1105 (date of access: 15.05.2026).

47. Single-image tomography: 3D volumes from 2D cranial x-rays / P. Henzler et al. *Computer graphics forum*. 2018. Vol. 37, no. 2. P. 377–388. URL: https://doi.org/10.1111/cgf.13369 (date of access: 15.05.2026).

48. Stanley K. O. Compositional pattern producing networks: a novel abstraction of development. *Genetic programming and evolvable machines*. 2007. Vol. 8, no. 2. P. 131–162. URL: https://doi.org/10.1007/s10710-007-9028-8 (date of access: 15.05.2026).

49. BA-Net: dense bundle adjustment network. *arXiv.org*. URL: https://arxiv.org/abs/1806.04807 (date of access: 15.05.2026).

50. The lumigraph / S. J. Gortler et al. *The 23rd annual conference*, Not Known. New York, New York, USA, 1996. URL: https://doi.org/10.1145/237170.237200 (date of access: 15.05.2026).

51. The unreasonable effectiveness of deep features as a perceptual metric / R. Zhang et al. *2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Salt Lake City, UT, 18–23 June 2018. 2018. URL: https://doi.org/10.1109/cvpr.2018.00068 (date of access: 16.05.2026).

52. CLPIPS: a personalized metric for ai-generated image similarity. *arXiv.org e-Print archive*. URL: https://arxiv.org/html/2604.01234v1 (date of access: 16.05.2026).

53. Unstructured lumigraph rendering / C. Buehler et al. *Seminal graphics papers: pushing the boundaries, volume 2*. New York, NY, USA, 2023. P. 497–504. URL: https://doi.org/10.1145/3596711.3596764 (date of access: 16.05.2026).

54. Waechter M., Moehrle N., Goesele M. Let there be color! Large-scale texturing of 3D reconstructions. *Computer vision – ECCV 2014*. Cham, 2014. P. 836–850. URL: https://doi.org/10.1007/978-3-319-10602-1_54 (date of access: 16.05.2026).

55. Kanatani K. Representation of 3D rotations. *Group-Theoretical methods in image understanding*. Berlin, Heidelberg, 1990. P. 197–235. URL: https://doi.org/10.1007/978-3-642-61275-6_6 (date of access: 16.05.2026).

56. Multiface: a dataset for neural face rendering. *arXiv.org*. URL: https://arxiv.org/abs/2207.11243 (date of access: 16.05.2026).

57. Xian S., Zhang L. Robustness measurement of comprehensive evaluation model based on the intraclass correlation coefficient. *Mathematics*. 2025. Vol. 13, no. 11. P. 1748. URL: https://doi.org/10.3390/math13111748 (date of access: 16.05.2026).

58. FaceScape: A large-scale high quality 3D face dataset and detailed riggable 3D face prediction / H. Yang et al. *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Seattle, WA, USA, 13–19 June 2020. 2020. URL: https://doi.org/10.1109/cvpr42600.2020.00068 (date of access: 16.05.2026).

59. UltraGS: gaussian splatting for ultrasound novel view synthesis. *arXiv.org e-Print archive*. URL: https://arxiv.org/html/2511.07743v1 (date of access: 16.05.2026).

60. HRAvatar: high-quality and relightable gaussian head avatar. *arXiv.org*. URL: https://arxiv.org/abs/2503.08224 (date of access: 16.05.2026).

61. ILSH: the imperial light-stage head dataset for human head view synthesis / J. Zheng et al. *2023 IEEE/CVF international conference on computer vision workshops (ICCVW)*, Paris, France, 2–6 October 2023. 2023. URL: https://doi.org/10.1109/iccvw60793.2023.00119 (date of access: 17.05.2026).

62. Zielonka W., Bolkart T., Thies J. Instant volumetric head avatars. *2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Vancouver, BC, Canada, 17–24 June 2023. 2023. URL: https://doi.org/10.1109/cvpr52729.2023.00444 (date of access: 17.05.2026).