Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет кібербезпеки
(повна назва)

Кафедра інформаційно-мережної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти *другий(магістерський)*

Дослідження методів відновлення зображень за допомогою ШІ
(тема)

Виконав:
здобувач 2 року навчання,
групи ІМІм-24-2
Максим КАЛІН
(власне ім'я, прізвище)

Спеціальність 172 Електронні комунікації та радіотехніка
(код і повна назва спеціальності)
Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Інформаційно-мережна інженерія
(повна назва освітньої програми)

Керівник доц. Наталія ХАРЧЕНКО
(посада, власне ім'я, прізвище)

Допускається до захисту

Зав. кафедри _____ Микола МОСКАЛЕЦЬ.
(підпис) (власне ім'я, прізвище)

2026 р.

Не містить відомостей, заборонених до відкритого публікування

Здобувач _____ / Максим Калін___/
(підпис) (власне ім'я, прізвище)

Керівник _____ / Наталія Харченко /
(підпис) (власне ім'я, прізвище)

Харківський національний університет радіоелектроніки

Факультет кібербезпеки

Кафедра інформаційно-мережної інженерії

Рівень вищої освіти другий (магістерський)

Спеціальність 172 Електронні комунікації та радіотехніка

Тип програми освітньо-наукова
(код і повна назва)

Освітня програма інформаційно-мережна інженерія
(повна назва)

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

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Каліну Максиму Сергійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів відновлення зображень за допомогою ШІ

затверджено наказом університету від 23 березня 2026р . №232 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 27 травня 2026р .

3. Вихідні дані до роботи
Розглянути основні методи відновлення якості зображень.
Дослідити методи застосування нейронних мереж для відновлення та покращення якості цифрових зображень та фото. Проаналізувати різні моделі ШІ, час їх тренування та скільки ресурсів вони потребують, наскільки ефективно використовувати ту чи іншу модель в залежності від конкретної задачі та виявити їх сильні та слабі сторони.

4. Перелік питань, що потрібно опрацювати в роботі
Вступ
1. Загальні методи відновлення якості
2. Відновлення зображень завдяки ШІ
3. Порівняльний аналіз моделей ШІ
4. Реалізація моделі відновлення зображень за допомогою автоенкодерів
Висновки

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) _____

Слайди у форматі Power Point (назва та мета роботи, класифікація методів відновлення якості зображень, моделі ШІ та їх особливості у задачі покращення якості зображень, порівняльний аналіз моделей ШІ, дослідження кращих випадків для їх використання, дослідження ефективності та часу навчання моделей ШІ, реалізація моделі відновлення зображень за допомогою автоенкодерів, висновки)

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Ознайомлення із завданням. Уточнення ТЗ.	23.03.2026	вик.
2	Підбір літератури за темою роботи	24.03 -31.03.2026	вик.
3	Загальні методи відновлення якості	3.04. -15.04.2026	вик.
4	Відновлення зображень завдяки ШІ	20.04. -2.05.2026	вик.
5	Порівняльний аналіз моделей ШІ	4.05. -09.05.2026	вик.
6	Оформлення презентаційного матеріалу	12.05 - 19.05.2026	вик.
	підготовка до захисту у ЕК		

Дата видачі завдання 23 березня 2026 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Наталія Харченко
(підпис) (посада, власне ім’я, прізвище)

РЕФЕРАТ

Пояснювальна записка: 97 с., 21 рис., 1 табл., 2 додатки, 11 джерел.

Об'єкт роботи – алгоритми обробки зображень.

Предмет дослідження – відновлення зображень штучним інтелектом за допомогою нейронних мереж.

У сучасному світі цифрові зображення використовуються практично всюди: у медицині, системах відеоспостереження, супутникових знімках, комп’ютерному зорі та мультимедійних застосунках. Проте під час отримання або передавання зображень часто виникають різні спотворення. Вони можуть бути спричинені шумами, розмиттям, низькою роздільною здатністю, втратою частини інформації або обмеженнями технічних засобів. Через це проблема відновлення якості зображень є важливою та актуальною. Від того, наскільки добре відновлене зображення, залежить якість подальшої обробки – наприклад, розпізнавання об’єктів, аналіз сцени або класифікація даних.

ЯКІСТЬ ЗОБРАЖЕННЯ, ВТРАТИ, НЕЙРОННА МЕРЕЖА, МАШИННЕ НАВЧАННЯ, АВТОЕНКОДЕРИ, GAN

ABSTRACT

Abstract: 97 pages, 21 figures, 1 table, 2 appendices, 11 references.

Scope of the work – image processing algorithms.

Subject of the study – image restoration using artificial intelligence and neural networks.

In today's world, digital images are used virtually everywhere: in medicine, video surveillance systems, satellite imagery, computer vision, and multimedia applications. However, various distortions often occur during the acquisition or transmission of images. These can be caused by noise, blurring, low resolution, loss of information, or technical limitations. Consequently, the problem of restoring image quality is both important and relevant. The quality of subsequent processing such as object recognition, scene analysis, or data classification depends on how well the image is restored.

IMAGE QUALITY, LOSS, NEURAL NETWORK, MACHINE LEARNING, AUTOENCODERS, GAN.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ

AI (Artificial Intelligence) – штучний інтелект;

CNN (Convolutional Neural Network) – згорткова нейронна мережа;

GAN (Generative Adversarial Network) – генеративно-змагальна мережа;

HR (High Resolution) – зображення високої роздільної здатності;

MSE (Mean Squared Error) – середньоквадратична помилка;

SR (Super Resolution) – надвисока роздільна здатність;

VGG (Visual Geometry Group) – модель згорткової нейронної мережі;

ШІ – штучний інтелект.

ВСТУП

У сучасному світі цифрові зображення використовуються практично всюди: у медицині, системах відеоспостереження, супутникових знімках, комп’ютерному зорі та мультимедійних застосунках. Проте під час отримання або передавання зображень часто виникають різні спотворення. Вони можуть бути спричинені шумами, розмиттям, низькою роздільною здатністю, втратою частини інформації або обмеженнями технічних засобів. Через це проблема відновлення якості зображень є важливою та актуальною. Від того, наскільки добре відновлене зображення, залежить якість подальшої обробки — наприклад, розпізнавання об’єктів, аналіз сцени або класифікація даних. Традиційно для покращення зображень використовують різні методи фільтрації, алгоритми усунення шумів та математичні методи оптимізації. Останніми роками все більшого поширення набувають методи машинного навчання та нейронні мережі, які дозволяють досягати значно кращих результатів у задачах відновлення різкості, видалення шумів і підвищення роздільної здатності. Метою даної дипломної роботи є дослідження методів відновлення якості зображень та реалізація одного з них для покращення візуальних характеристик цифрових зображень. Для досягнення поставленої мети необхідно розглянути основні причини погіршення якості зображень, проаналізувати класичні та сучасні методи їх відновлення, а також оцінити ефективність обраного підходу на практичних прикладах. Об’єктом дослідження є процес відновлення цифрових зображень. Предметом дослідження є методи та алгоритми покращення та відновлення якості зображень. Практичне значення отриманих результатів полягає в можливості використання досліджених методів у реальних прикладних задачах, зокрема в системах комп’ютерного зору та мультимедійних застосунках.

1 ЗАГАЛЬНІ МЕТОДИ ВІДНОВЛЕННЯ ЯКОСТІ

Одним із найпоширеніших методів покращення характеристик зображень є фільтрація. Фільтр Unsharp Mask (нерізка маска) використовується для тонкої кольорової корекції, він є дуже корисним інструментом і призначається для збільшення різкості. Unsharp masking - це технологічний прийом обробки зображення, який дозволяє домогтися ефекту відчуття більшої його різкості за рахунок посилення контрасту тональних переходів. Важливо відзначити, що нерізке маскування не підвищує різкість зображення насправді. Воно не може відновити втрачені деталі на різних етапах обробки зображення. Нерізке маскування лише підсилює локальний контраст зображення на тих ділянках, де спочатку були присутні різкі зміни градацій кольору. Завдяки цьому зображення візуально сприймається як більш різке [1]. Також існують і проблеми нерізкого маскування. При неправильному або надмірному використанні цього фільтру може бути враження «штучності» зображення, що є небажаним ефектом, завдяки якому відновлені краї виглядають з видимими білими відтінками навколо них. Для усунення цього дефекту можна використати модифікований фільтр нерізкої маски, який забезпечує різкість різного цифрового зображення без введення ефекту перевищення. У такому фільтрі зображення згладжується за допомогою традиційного двостороннього фільтра, а потім розмивається за допомогою модифікованого фільтра Баттерворта, замість того, щоб розмити його за допомогою фільтра низьких частот Гауса, як у традиційному фільтрі маски без різкості. Використання цього підходу дозволило б усунути ефект перевищення та відновити якісніші результати. Після застосування фільтра нерізкої маски до краю з великим нахилом (наприклад до темної будівлі на фоні світлого неба) часто виявляють дефект ступінчастості. Для вирішення цього завдання

використовуються фільтри шумозаглушення. Одним із таких фільтрів є згладжування за допомогою гаусіан.

Даний підхід можна описати так: згладжування приглушує шум, дотримуючись вимоги, щоб пікселі були схожі на своїх сусідів. Наступним ефективним методом є метод перетворення гістограм. Гістограма – це діаграма, побудована в стовпчиковій формі, в якої величина показника зображується графічно у вигляді стовпчика. Гістограма наочно характеризує, як величина показника змінюється з часом. Оптимальним, для зорового сприйняття людиною, є зображення, елементи якого мають рівномірний розподіл яркостей. Поліпшення зображень шляхом вирівнювання гістограми - це процес, в якому намагаються досягти рівномірності розподілу яркостей обробленого зображення [1]. Процедура вирівнювання гістограми складається з наступних дій:

- обчислюється гістограма розподілу яркостей елементів зображення H (L);
- будується нормована кумулятивна гістограма CH (L);
- формується нове зображення L '= R * CH (L).

Це перетворення ефективно для поліпшення візуальної якості низько контрастних деталей. Описаний метод перетворення гістограми може бути глобальним, тобто використовувати інформацію про все зображення, та легким, коли для перетворення використовуються локальні області зображення. Глобальні методи гістограмного перетворення є, по суті, табличними методами, основна їхня перевага складається в швидкодії. Для більш детальної обробки зображень використовують ковзаючі методи. Вони забезпечують якісне контрастування дрібних деталей зображення. Разом з тим, вони більш об'ємні по обчислювальним витрат.

Тому при використанні методів класу гістограмних перетворень потрібно шукати компроміс між якістю і швидкодією обробки зображень. Однією з найбільш зручних форм представлення інформації при діагностуванні в медицині

та інших областях є зображення. Це призводить до необхідності розвитку способів діагностики з використанням різноманітних методів [1].

Однак одним з істотних недоліків цих методів є те, що вони переважно забезпечують формування зображень з низькою контрастністю. Тому основна мета методів поліпшення полягає в перетворенні зображень до такого виду, що робить їх більш контрастними і, відповідно, більш інформативнішими. Досить часто на зображенні присутні спотворення лише в певних місцях, які викликані дифракцією світла, недоліками оптичних систем або розфокусуванням. Це призводить до необхідності виконання локальних перетворень на зображенні. Іншими словами, такий адаптивний підхід дає можливість виділити інформативні ділянки на зображенні і відповідним чином їх опрацювати. Викладеним вимогам відповідають методи адаптивного перетворення локального контрасту. Основні кроки даного методу наступні:

- для кожного елемента зображення L (i, j) обчислюють значення локального контрасту C (i, j) в поточній ділянці W з центром в елементі з координатами (i, j);

- обчислюють локальну статистику для поточної ділянки W;

- перетворять (підсилюють) локальний контраст C (i, j), використовуючи для цього нелінійні функції і враховуючи локальну статистику поточної ділянки W;

- відновлюють значення яскравості зображення L '(i, j) з посиленим локальним контрастом. Таким чином, метод посилення контрасту з використанням функції протяжності гістограми ефективно використовується в обробці широкого класу зображень.

Існуючі методи відновлення зображень можна поділити на дві великі категорії: методи, що концентруються на неструктурованих дефектах (шум, розмиття, вицвітання кольору тощо) та методи для структурованих дефектів

(діри, подряпини). Для першої категорії існують як традиційні методи з використанням різних пріорів зображень (нелокальна самоподібність, розрідженість тощо), так і глибокі нейронні мережі.

Для структурованих дефектів складніші методи зазвичай моделюють задачу як «зображення, що фарбує» і використовують потужні можливості семантичного моделювання. Досліджень відновлення від змішаних дефектів (комбінації структурованих і неструктурованих) значно менше. Деякі методи пропонують інструментарій з кількох мереж, кожна для певного типу дефекту. Інші покладаються лише на синтетичні дані. Існуючі дефекти зображення можна умовно розділити на дві групи: неструктуровані дефекти, такі як шум, розмитість, вицвітання кольорів і низька роздільна здатність, і структуровані дефекти, такі як отвори, подряпини і плями. Для перших, неструктурованих, традиційні методи часто накладають різні обмеження на зображення, включаючи нелокальну самоподібність, розрідженість і локальну гладкість. Існує багато методів на основі глибокого навчання для різних видів деградації зображень, таких як згладжування зображень, надвисока роздільна здатність та розмиття. Порівняно з неструктурованою деградацією, структурована деградація є більш складною і часто моделюється як проблема "розфарбовування зображення". Завдяки потужним можливостям семантичного моделювання, більшість існуючих методів розфарбовування засновані на навчанні. Наприклад, автори роботи замаскували дірчасті області в операторі згортки і змусили мережу зосередитися лише на недірчастих особливостях [1]. Щоб отримати кращі результати розфарбовування, багато інших методів враховують як локальну статистику виправлень, так і глобальні структури. Зокрема, в статтях та було запропоновано використовувати шар уваги для використання віддаленого контексту. А потік появи явно оцінюється в роботі Рен та ін., так що текстури в областях дірок можуть бути безпосередньо синтезовані на основі відповідних патчів. Хоча

вищезгадані методи, що базуються на навчанні, можуть досягти чудових результатів незалежно від того, чи йдеться про неструктуровану чи структуровану деградацію, всі вони навчаються на синтетичних даних [1].

Тому їхня ефективність на реальному наборі даних сильно залежить від якості синтетичних даних. Для реальних старих зображень процес деградації, що лежить в їх основі, набагато складніше охарактеризувати, оскільки вони часто серйозно погіршені сумішшю невідомих деформацій. Іншими словами, мережа, навчена лише на синтетичних даних, буде страждати від проблеми розриву домену і погано працюватиме на реальних старих фотографіях. У реальному світі пошкоджене зображення може страждати від складних дефектів у поєднанні з подряпинами, втратою роздільної здатності, вицвітанням кольорів і шумами плівки. Однак значно менш методів вирішують проблему змішаної деградації. У роботі запропоновано інструментарій, який складається з кількох легких мереж, кожна з яких відповідає за певну деградацію. Потім розглядається контролер, який динамічно вибирає оператора з інструментарію. Натхненний роботою, виконує різні згорточні операції паралельно і використовує механізм уваги для вибору найбільш підходящої комбінації операцій. Однак ці методи все ще покладаються на контрольоване навчання на синтетичних даних і, отже, не можуть бути узагальнені на реальні фотографії. Крім того, вони зосереджені лише на неструктурованих дефектах і не підтримують структуровані дефекти, такі як зафарбовування зображень. З іншого боку, Ульянов та ін. виявили, що глибока нейронна мережа за своєю суттю резонує з низькорівневою статистикою зображень і, таким чином, може бути використана як попереднє зображення для сліпого відновлення зображень без зовнішніх навчальних даних. Цей метод має потенціал, хоча і не заявлений в, для відновлення природних зображень, пошкоджених змішаними факторами. Реставрація старих фотографій є класичною проблемою змішаної деградації, але більшість існуючих методів

зосереджені лише на фарбах. Вони дотримуються схожої парадигми, тобто дефекти, такі як подряпини і плями, спочатку ідентифікуються за низькорівневими ознаками, а потім зафарбовуються, запозичуючи текстури з довколишніх ділянок [1].

Однак на моделях ручної роботи та низькорівневих ознаках, які вони використовують, такі дефекти складно виявити та виправити належним чином. Крім того, жоден з цих методів не враховує відновлення деяких неструктурованих дефектів, таких як вицвітання кольору або низька роздільна здатність разом із зафарбовуванням. Таким чином, після реставрації фотографії все ще виглядають старомодними.

На даний момент існує декілька програмних засобів, які пропонують послуги відновлення старих фотографій за допомогою штучного інтелекту. Кожен з цих інструментів має свої переваги та обмеження. Онлайн-сервіси зазвичай є більш доступними, проте їхні можливості можуть бути обмеженими у порівнянні зі складнішими програмами, такими як Photoshop. Ці додатки також можуть відрізнятися за якістю відновлення. Перед використанням будь-якого інструменту варто провести тестове відновлення, оскільки результати можуть варіюватися в залежності від початкового стану фотографії та типу дефектів. Розглянемо деякі найбільш типові з існуючих систем, що використовуються для видалення артефактів зі старих фотографій, зокрема:

- Remini;
- Adobe Photoshop;
- Fotor Photo Restoration;
- LetsEnhance.io;
- Pixlr.

Remini - додаток, який вирізняється серед інших у сфері реставрації старих фотографій. Цей застосунок використовує штучний інтелект та глибоке навчання

для покращення та відновлення старих фотографій, дозволяючи користувачам підвищити їх якість та деталізацію [1].

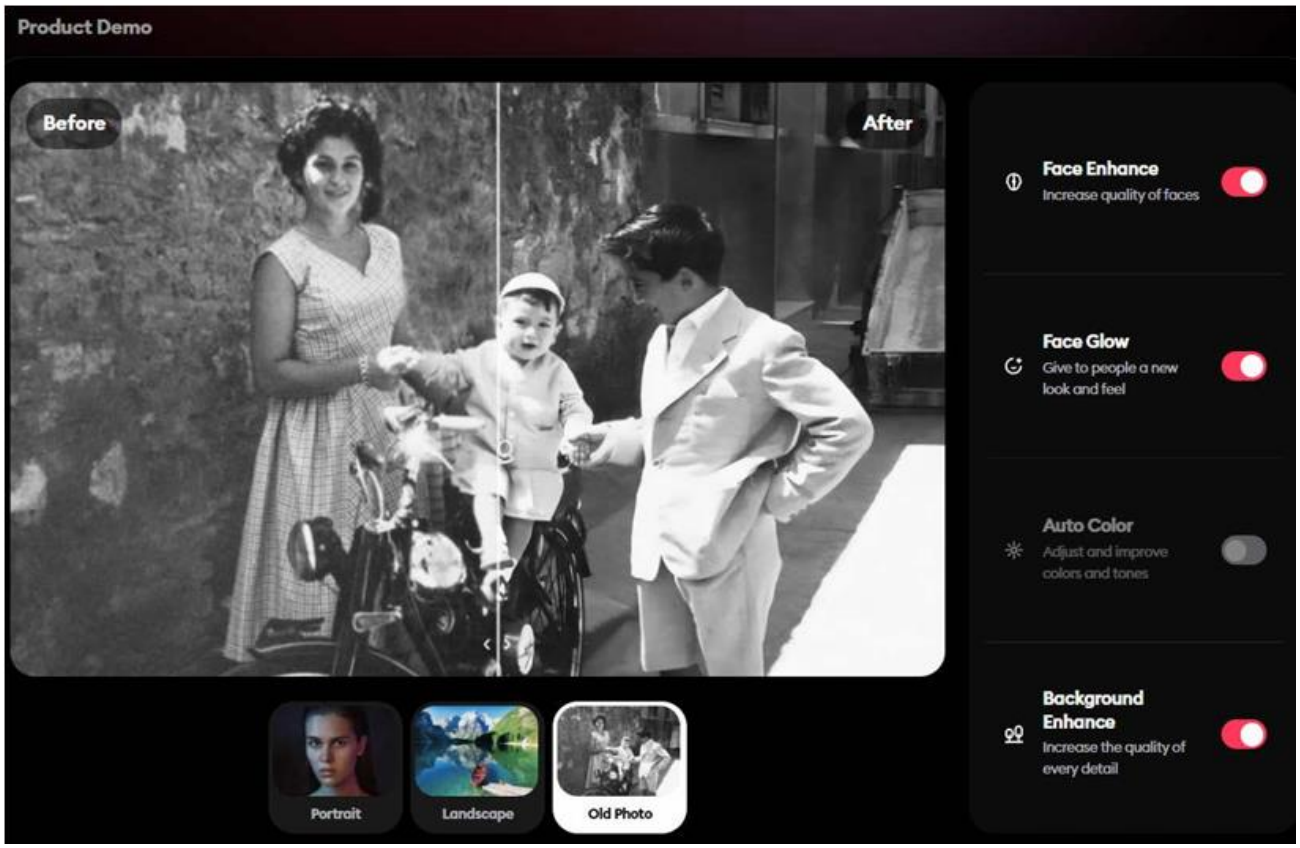

Рисунок 1.1 – Демонстрація Remini

Remini використовує набір алгоритмів машинного навчання, зокрема глибокі нейронні мережі, щоб аналізувати та відновлювати старі фотографії. Цей процес може включати зменшення шумів, відновлення втрачених деталей, поліпшення чіткості та кольорової гами. Однією з головних переваг Remini є його здатність працювати безпосередньо на мобільних пристроях, що дозволяє користувачам легко та швидко відновлювати свої фотографії без необхідності використання складних програм або ресурсів. Додаток пропонує інтуїтивний і простий у використанні інтерфейс, що робить його доступним для широкого кола користувачів будь-якого рівня кваліфікації. Хоча Remini здатний значно покращити якість старих фотографій, варто зауважити, що кінцевий результат

значною мірою залежить від початкової якості зображення та складності дефектів. Надто пошкоджені фотографії можуть потребувати більшого обсягу обробки або не дозволити досягти бажаного результату [1].

Загалом, Remini відомий своєю здатністю покращувати старі фотографії з високою ефективністю та швидкістю, забезпечуючи користувачам можливість відновлення своїх цінних зображень з мінімальними зусиллями.

Переваги:

- зручний інтерфейс користувача. Remini пропонує простий і зрозумілий інтерфейс для користувачів, що дозволяє легко використовувати його функції для поліпшення фотографій;

- автоматизоване відновлення. Додаток автоматично виконує процес відновлення, що дозволяє користувачам отримати поліпшені фотографії без потреби в спеціалізованому знанні;

- висока швидкість обробки. Remini відзначається високою швидкістю обробки фотографій, що дозволяє користувачам швидко отримувати результати. Недоліки:

- залежність від якості вхідних фото: Як і більшість програм для відновлення фотографій, які працюють на основі штучних нейронних мереж, ефективність Remini може залежати від якості вхідного зображення;

- обмеженість налаштувань. Зазвичай такі додатки мають обмежені налаштування, що може обмежити користувачів у виборі певних параметрів чи налаштувань для оптимізації відновлення;

- вимоги до інтернет-з'єднання. Деякі функції можуть вимагати стабільного інтернет-з'єднання, що може бути не завжди доступним.

Adobe Photoshop є одним з найвідоміших та потужних інструментів у сфері реставрації старих фотографій. Цей програмний пакет має широкий спектр інструментів та функцій, що дозволяють не лише редактувати та покращувати

старі фотографії, а й відновлювати їх до високої якості. Adobe Photoshop пропонує безліч інструментів для ретуші та відновлення фотографій [1].

Наприклад, інструменти для видалення дефектів (плям, подряпин, пилу), відновлення колірної гами, збільшення роздільної здатності, усунення шуму та відновлення втрачених деталей.

Рисунок 1.2 – Демонстрація Adobe Photoshop

Основні функції Photoshop, які знаходять застосування у відновленні старих фотографій, включають:

- клонування та виправлення дефектів: Інструменти клонування та виправлення дозволяють видаляти дефекти та нечіткості, реставруючи пошкоджені ділянки зображення;

- відновлення колірної гами: Photoshop пропонує можливості кольорової корекції для відновлення колірної гами старих фотографій та усунення вицвітання;

- фільтри та ефекти: Використання фільтрів і ефектів може допомогти усунути шум, покращити контраст та деталізацію фотографій;

- робота з шарами та історією: З можливістю роботи з шарами, користувачі можуть відстежувати кожен етап відновлення та виправлення. Історія дозволяє повертатися до попередніх змін;

- синхронізація та обробка великої кількості фотографій: Adobe Photoshop Lightroom, який інтегрований з Photoshop, дозволяє робити масову обробку та редакцію фотографій, що включає автоматичні функції відновлення та покращення.

Незважаючи на велику функціональність та можливості Adobe Photoshop, використання цього інструменту вимагає певного рівня експертизи та знань. Однак, знання та вміння користуватися Photoshop може дати надзвичайно потужні та вражаючі результати у відновленні старих фотографій, забезпечуючи великий контроль над процесом та можливості максимального покращення якості зображення.

Переваги:

- різноманітність інструментів. Photoshop має широкий спектр інструментів для виправлення і відновлення фотографій, таких як клонування, реставрація, ретушування і зміна кольорів;

- шари і маски. Функція шарів дозволяє працювати з окремими частинами зображення, що полегшує виправлення дефектів без втрати вихідних даних;

- історія редагування. Можливість відслідковувати всі зміни та кроки редагування дозволяє легко відміняти або повертати зміни в будь який момент;

- зручний інтерфейс користувача. Інтерфейс Photoshop є інтуїтивно зрозумілим і включає в себе багато налаштувань і панелей для редагування [1].

Недоліки:

- складність вивчення і використання. Photoshop є потужним інструментом, але він може бути складним для новачків, і вимагає часу для оволодіння всіма його можливостями;

- вартість. Ліцензійна версія Photoshop може бути дорога для користувачів, особливо для тих, хто використовує його лише для особистого використання;

- вимоги до ресурсів системи. Робота з великими фотографіями та складними процесами може вимагати потужний комп'ютер, що може вплинути на продуктивність;

- часові затрати. Деякі складні процеси редагування можуть займати багато часу, особливо якщо йдеться про великі обсяги фотографій.

Fotor Photo Restoration – це онлайн-сервіс, який пропонує можливості для відновлення та поліпшення старих фотографій. Цей інструмент дозволяє користувачам виконувати видалення плям, відновлення колірних деталей та покращення загального вигляду зображення.

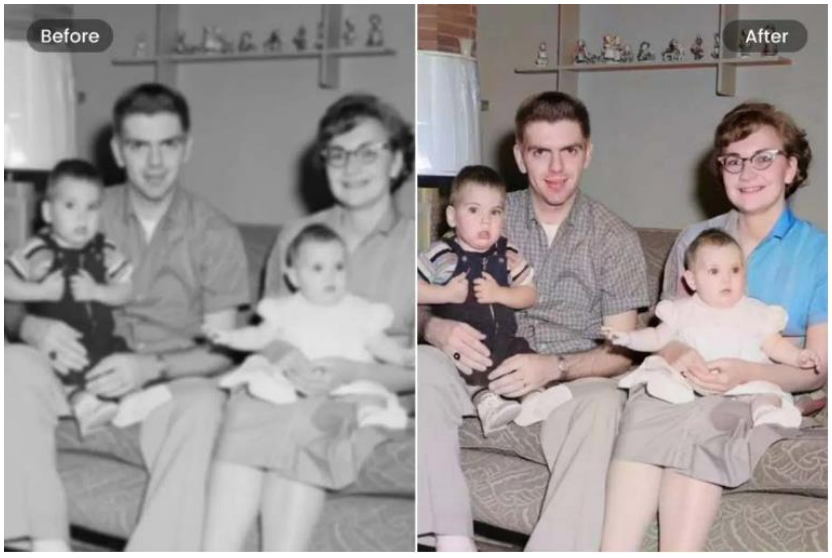

Рисунок 1.3 – Демонстрація Fotor Photo Restoration

Основні функції Fotor Photo Restoration:

- ретуш та реставрація фотографій. Додаток дозволяє виправляти пошкодження на старих фотографіях, такі як подряпини, тріщини, плями або інші дефекти, шляхом замалювання чи видалення цих елементів;

- відновлення кольору та колірний корекції. Fotor пропонує інструменти для відновлення кольору на відбитих фотографіях та можливості налаштування кольорової палітри для поліпшення якості кольорів;

- виправлення контрастності та експозиції. Функції, які дозволяють корегувати контраст, яскравість, експозицію та інші параметри, що можуть покращити загальний вигляд та якість зображення;

- оптичне розпізнавання та автоматична корекція. В деяких випадках, програма може автоматично виявляти пошкодження та намагатися їх відновити або покращити без втручання користувача;

- онлайн та офлайн версії. Додаток пропонує як веб-версію, так і мобільні застосунки, що дає можливість використовувати його на різних платформах.

Fotor Photo Restoration є корисним інструментом для базового відновлення старих фотографій та ретуші, особливо для тих користувачів, яким потрібен простий та доступний інструмент для полегшення процесу реставрації фотографій. Однак для складних відновлень можуть знадобитися більш потужні та розширені програми.

Переваги:

- легкість використання. Програма має простий та дружелюбний інтерфейс, який полегшує користувачам використання;

- онлайн-доступність. Ви можете працювати з фотографіями прямо з веб браузера без необхідності завантажувати та встановлювати програмне забезпечення;

- шаблони та ефекти. Має вбудовані шаблони та ефекти, які можна застосовувати для поліпшення зображень;

- автоматичні інструменти. Можливість використовувати автоматичні інструменти для швидкої корекції фотографій [1].

Недоліки:

- обмежені можливості. Порівняно зі складними програмами для редагування, такими як Photoshop, може бути менше опцій для деталізованого відновлення;

- залежність від мережі. Оскільки це онлайн-сервіс, для роботи потрібне постійне підключення до Інтернету;

- якість результату. Якість відновлення може варіюватися в залежності від фотографій і використовуваних алгоритмів;

- вартість та обмеження безкоштовних версій. Безкоштовні версії сервісу мають певні обмеження. Доступ до повних функцій потребує оплати.

LetsEnhance.io це онлайн-платформа, яка спеціалізується на поліпшенні якості та підвищенні роздільної здатності фотографій за допомогою штучного інтелекту та глибокого навчання. Система пропонує різні інструменти для відновлення старих фотографій, зокрема, усунення шумів, підвищення різкості та покращення деталізації. Сервіс використовує технології штучного інтелекту для автоматичного виявлення дефектів та зменшення втрати деталей під час реставрації. Користувачі можуть завантажити свої фотографії на платформу та використовувати різні інструменти для їх відновлення, покращення різкості та підвищення деталізації.

Переваги:

- швидкість та доступність. Онлайн-платформа, доступна з будь-якого пристрою, що має доступ до мережі Інтернет;

- використання штучного інтелекту. Використання технологій штучного інтелекту для автоматичного виявлення і усунення дефектів;

- простота використання. Інтуїтивний інтерфейс та простий процес реставрації фотографій;

- висока швидкість обробки. Зазвичай обробка відбувається досить швидко.

Недоліки:

- обмежені можливості безкоштовної версії. Певні функції можуть бути обмежені або взагалі недоступні в безкоштовній версії сервісу;

- не завжди точні результати. Ефективність реставрації може варіюватися в залежності від початкового стану фотографій та характеру пошкоджень.

Pixlr – це онлайн-редактор фотографій, який надає користувачам можливість редагувати та покращувати свої зображення.

Переваги:

- зручний інтерфейс. Має інтуїтивно зрозумілий і простий для використання інтерфейс, що дозволяє швидко редагувати фотографії без складних налаштувань;

- багатофункціональність. Має різноманітні інструменти редагування, такі як корекція кольору, зменшення шуму, видалення дефектів тощо, які можна використовувати для реставрації старих зображень;

- доступність. Це онлайн-сервіс, тому використовувати його можна на будь-якому пристрої з підключенням до Інтернету без необхідності встановлення додаткових програм [1].

Недоліки:

- обмежені можливості безкоштовної версії. Деякі додаткові функції або розширені інструменти доступні лише у платних версіях;

- можливість обробки лише онлайн. Всі фотографії потрібно завантажувати для редагування, що може бути незручним для великих обсягів роботи або при обробці конфіденційних зображень [1].

2 ВІДНОВЛЕННЯ ЗОБРАЖЕНЬ ЗАВДЯКИ ШІ

2.1 Роль ШІ у відновленні зображень

Штучний інтелект (ШІ) – це розділ інформатики, що займається створенням комп'ютерних систем, здатних виконувати завдання, які традиційно вимагають людського інтелекту. До таких завдань належать аналіз великих обсягів інформації, розпізнавання образів, мовлення та тексту, прогнозування майбутніх подій і навіть прийняття рішень у складних ситуаціях. Штучний інтелект відкриває нові горизонти для фотозйомки та редагування зображень, використовуючи алгоритми машинного та глибинного навчання. Він радикально змінює підхід до обробки фото, роблячи професійні інструменти доступними навіть для новачків [2]. Від автоматичного покращення якості знімків до видалення небажаних об'єктів і корекції кольору – ШІ розширює можливості фотографів та дизайнерів. Розвиток штучного інтелекту у сфері фотографії розпочався з елементарних інструментів автоматичного покращення зображень, які могли лише незначно коригувати яскравість, контраст або баланс білого. Згодом, із зростанням обчислювальної потужності комп'ютерів та появою нових методів машинного навчання, ці технології еволюціонували в більш складні системи аналізу зображень. Вони не лише автоматично покращували якість фотографій, але й навчилися розпізнавати об'єкти, обличчя та навіть художні стилі [3].

Переломним моментом стало впровадження нейронних мереж, зокрема конволюційних нейронних мереж (CNN), які дозволили аналізувати візуальні дані на глибшому рівні. Завдяки цим алгоритмам стало можливим не лише

коригувати зображення, а й ретельно їх аналізувати, виділяючи ключові особливості та деталі.

Подальший розвиток технологій привів до використання глибинного навчання, зокрема генеративно-змагальних мереж (GAN), що дозволило ШІ створювати абсолютно нові, фотореалістичні зображення. Це відкрило шлях до появи таких інновацій, як діпфейки, автоматична заміна фону, кольоризація старих фотографій та багато інших функцій, які сьогодні активно використовуються як професіоналами, так і аматорами. Таким чином, штучний інтелект перетворився на невід'ємний інструмент у сучасній фотографії, що значно спростив процес зйомки та редагування зображень. Впровадження штучного інтелекту у фоторедагування значно спростило процес створення та відновлення якісних зображень, зробивши його доступним для всіх користувачів. Основний функціонал включає в себе:

- автоматичне покращення зображень: Функції автоматичного налаштування, такі як корекція експозиції, контрастності та насиченості, дозволяють значно спростити редагування фотографій;

- видалення фону: Функція дозволяє легко ізолювати об'єкти на зображенні для подальшого використання в інших проектах або для створення колажів. Цей інструмент незамінний для створення продуктових фото, рекламних матеріалів та професійних портретів;

- портретна ретуш: Інструменти для ретуші портретів дозволяють з легкістю покращувати вигляд обличчя: згладжувати шкіру, підкреслювати очі, додавати природний рум'янець або відбілювати зуби;

- автоматична кольорокорекція: Функції автоматичної корекції кольору аналізують зображення і підлаштовують кольорову палітру для досягнення оптимального результату [3].

2.2 Огляд існуючих підходів для реставрації старих фотографій на основі глибокого навчання

Глибоке навчання відіграє ключову роль у відновленні пошкоджених чи старих фотографій, надаючи можливість відновлювати зображення та покращувати їх якість. На даний момент існує кілька підходів у цьому напрямі.

2.2.1 Генеративно-зорієнтовані моделі

Генеративно-зорієнтовані моделі – це тип нейронних мереж, який складається з двох основних компонентів: генератора і дискримінатора. Цей підхід був представлений у 2014 році та відтоді активно використовується у сфері комп’ютерного зору, обробки зображень і штучного інтелекту [1].

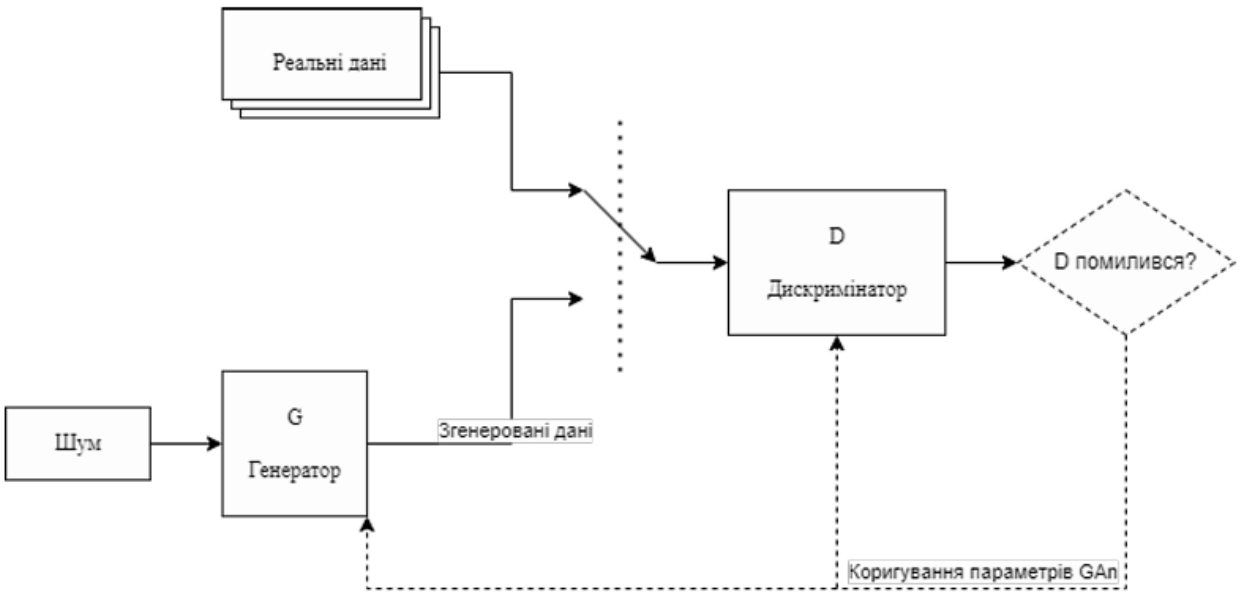

Рисунок 2.1 – Генеративно-зорієнтована модель ШІ

Генеративно-змагальна мережа складається з двох основних компонентів: генератора та дискримінатора. Генератор приймає на вхід зображення низької роздільної здатності та формує його покращену версію. Саме генератор відповідає за реконструкцію деталей, підвищення чіткості контурів, відновлення текстур і створення зображення суперроздільної здатності. Дискримінатор, у свою чергу, виконує роль оцінювача: він отримує як справжні зображення високої роздільної здатності, так і зображення, створені генератором, і намагається визначити, які з них є реальними, а які – згенерованими.

Процес навчання GAN має змагальний характер. Спочатку генератор створює зображення недостатньої якості, оскільки його параметри ще не оптимізовані.

Дискримінатор порівнює такі результати зі справжніми HR-зображеннями та навчається виявляти відмінності між ними. Після цього генератор отримує інформацію про свою помилку і змінює внутрішні параметри так, щоб наступного разу створити більш реалістичне зображення. Таким чином, генератор поступово навчається “обманювати” дискримінатор, тобто створювати такі зображення, які дискримінатору стає складно відрізнити від справжніх.

Важливо, що покращення якості відбувається не лише за рахунок збільшення розміру зображення [1].

Генератор навчається відтворювати високочастотні компоненти, які відповідають за дрібні деталі: волосся, текстуру тканини, траву, шкіру, архітектурні елементи, краї об’єктів тощо. Саме ці елементи часто втрачаються при зменшенні роздільної здатності.

Завдяки змагальному навчанню модель отримує стимул створювати не просто математично близьке до оригіналу зображення, а таке, яке виглядає природно для людського ока. Для цього у GAN використовуються спеціальні функції втрат.

Звичайна попіксельна похибка, наприклад MSE, оцінює різницю між зображеннями на рівні окремих пікселів. Такий підхід добре підвищує числові метрики, але часто призводить до розмиття, оскільки модель намагається усереднити можливі варіанти відновлення. У методах на основі GAN додатково використовуються перцептивні втрати та змагальні втрати. Перцептивні втрати дозволяють порівнювати зображення не лише на рівні пікселів, а й на рівні ознак, які виділяються попередньо навченою нейронною мережею, наприклад VGG. Це дає змогу оцінювати схожість зображень ближче до того, як її сприймає людина.

Змагальна функція втрат змушує генератор створювати зображення, які належать до простору природних зображень. Іншими словами, генератор отримує штраф не лише тоді, коли його результат відрізняється від еталонного HR-зображення, а й тоді, коли дискримінатор визначає його як штучний. Саме ця частина навчання сприяє появі реалістичних текстур і більш чітких деталей. У спрощеному вигляді процес відновлення якості можна описати так:

- на вхід генератора подається зображення низької роздільної здатності;
- генератор створює покращене SR-зображення;
- дискримінатор порівнює згенероване зображення зі справжніми HR-зображеннями;
- обчислюються втрати: попіксельні, перцептивні та змагальні;
- параметри генератора і дискримінатора оновлюються;
- процес повторюється багато разів, доки генератор не починає створювати достатньо реалістичні та деталізовані зображення.

Отже, генеративно-змагальні мережі виконують відновлення якості зображень шляхом навчання на прикладах реальних високоякісних зображень і поступового формування здатності відтворювати правдоподібні деталі. Їхня перевага полягає у тому, що вони орієнтуються не лише на формальну близькість до еталону, а й на візуальну реалістичність результату [1].

Завдяки цьому GAN дозволяють отримувати зображення, які виглядають чіткішими, природнішими та більш деталізованими порівняно з результатами традиційних методів масштабування.

Переваги використання GANs для реставрації старих фотографій включають:

- глибокі та точні результати. GANs можуть генерувати високоякісні зображення з втраченими деталями, які виглядають природні та реалістичні;

- здатність відновлювати відсутні деталі. Вони можуть заповнювати пропущені ділянки або відновлювати втрачені частини фотографій, що допомагає поліпшити зовнішній вигляд фотографій;

- автоматизація процесу реставрації. Після навчання моделі можна використовувати для автоматичного відновлення старих фотографій без значної втрати якості.

До недоліків відносять:

- потребу великої кількості даних. Навчання GANs для реставрації фотографій може вимагати значної кількості фотографій високої якості для досягнення задовільних результатів;

- час навчання. Процес навчання GANs може бути вимогливим за часом та ресурсами, особливо при роботі з великими обсягами даних;

- складність управління. Налаштування параметрів та оптимізація моделі GANs можуть вимагати досить глибоких знань та досвіду в галузі машинного навчання [1].

2.2.2 Автоенкодери (Autoencoders)

Автоенкодери – це тип нейронних мереж, що використовуються для зменшення розмірності даних та відтворення вихідного входу.

Вони складаються з двох основних частин: енкодера та декодера [1].

Енкодер приймає вхідні дані (наприклад, зображення) та перетворює їх у скомпресований представлення, або "латентний простір". Під час навчання енкодер навчається виділяти основні ознаки вхідних даних та створювати універсальне представлення, яке можна використовувати для відтворення вихідного зображення.

Декодер приймає латентний простір (згенерований енкодером) і відновлює вихідні дані, які максимально наближені до вхідного. Декодер намагається відтворити вихідні дані з латентного простору так, щоб вони максимально відповідали початковим даним [1].

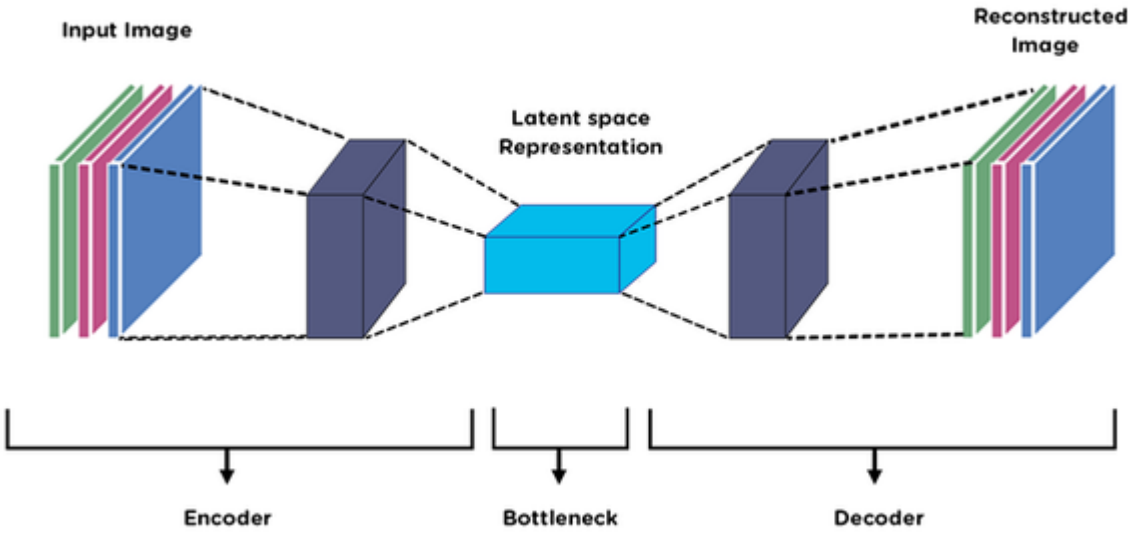

Рисунок 2.2 – Схема автоенкодера

Під час навчання, автоенкодери намагаються зменшити різницю між вхідними та відтвореними даними. Такий процес дозволяє моделі вчитися реконструювати вихідні дані, але також відкидає зайві деталі, що не є важливими для подальшого відтворення. У сфері відновлення старих фотографій,

автоенкодери можуть бути використані для усунення шумів, відтворення втрачених деталей або відновлення пошкоджених частин зображення. Вони можуть навчатися відновлювати пошкоджені частини фотографії, опрацьовуючи цілу серію зображень та навчаючись відтворювати їхні ключові аспекти. Такий підхід допомагає автоенкодерам розпізнавати важливі деталі та реконструювати зображення з них [5].

Переваги:

- здатність відновлювати втрачені деталі. Автоенкодери можуть відновлювати відсутні або пошкоджені ділянки фотографій, допомагаючи відновити вигляд зображення;

- автоматизований підхід до відновлення. Після навчання моделі автоенкодера вона може бути використана для автоматичної реставрації старих фотографій без значної людської участі;

- широкий спектр застосувань. Моделі автоенкодерів можуть бути налаштовані для різних видів реставрації, включаючи заповнення відсутніх ділянок, підвищення роздільної здатності тощо.

Недоліки:

- відсутність контролю за реставрацією. Моделі автоенкодерів можуть відновлювати зображення без чіткого розуміння того, що саме вони відновлюють, що може призвести до непередбачуваних результатів;

- потреба великої кількості даних для навчання. Навчання автоенкодерів може вимагати великої кількості вхідних даних, щоб модель змогла ефективно відтворювати зображення;

- час навчання та складність моделі. Навчання складних моделей автоенкодерів може займати багато часу, особливо при великих розмірах даних.

Процес відновлення якості зображення за допомогою автоенкодера базується на навчанні моделі відтворювати вхідні дані з мінімальною похибкою.

На етапі навчання мережа отримує зображення та поступово налаштовує свої параметри таким чином, щоб різниця між вхідним та відновленим зображенням була мінімальною [5].

Зазвичай для цього використовується функція втрат, наприклад середньоквадратична похибка (MSE), яка оцінює різницю між оригінальним і реконструйованим зображенням . Модель навчається шляхом мінімізації цієї похибки.

Особливість автоенкодера полягає в тому, що він не просто копіює вхідні дані. Завдяки наявності “вузького місця” (bottleneck), мережа змушена навчитися виділяти тільки найважливішу інформацію.

Це дозволяє:

- усувати шум у зображеннях;

- відновлювати відсутні або пошкоджені ділянки;

- покращувати загальну структуру та чіткість зображення.

Відновлення зображень та усунення шуму: Одним із ключових застосувань автоенкодерів є усунення шуму (denoising). У цьому випадку модель навчається на парах даних, де на вхід подається зашумлене зображення, а цільовим результатом є чисте зображення. Таким чином, автоенкодер вчиться ігнорувати шум та відновлювати корисну інформацію. Механізм роботи в цьому випадку можна описати наступним чином:

- вхідне зображення містить шум або спотворення;

- енкодер виділяє основні ознаки, які є стійкими до шуму;

- латентний простір зберігає узагальнене представлення без шумових компонентів;

- декодер реконструює зображення, відновлюючи лише значущі структури;

- у результаті отримується зображення з покращеною якістю, де шум зменшено або повністю усунуто.

Відновлення структури та деталей: Автоенкодери здатні не лише прибирати шум, а й відновлювати загальну структуру зображення. Завдяки навчанню на великій кількості прикладів, модель виявляє приховані закономірності у даних і використовує їх для реконструкції. Це означає, що автоенкодер:

- відновлює контури об’єктів;
- згладжує артефакти;
- зберігає важливі структурні характеристики зображення.

Латентний простір відіграє ключову роль у цьому процесі, оскільки саме в ньому формується узагальнене представлення, що містить найбільш інформативні ознаки зображення.

Тобто, автоенкодери здійснюють відновлення якості зображень шляхом навчання компактного представлення даних і подальшої реконструкції з нього. Завдяки використанню енкодера та декодера модель здатна виділяти суттєві ознаки зображення, усувати шум та відновлювати структуру. На відміну від традиційних методів, автоенкодери не просто обробляють пікселі, а аналізують внутрішню структуру даних, що дозволяє отримувати більш якісні результати [1].

2.2.3 Супервізоване навчання для відновлення зображень

В даному підході модель навчається відтворювати або відновлювати пошкоджені зображення, використовуючи спеціальний навчальний набір, що складається з пар зображень.

Одне зображення в кожній парі є пошкодженим (наприклад, з шумами, артефактами або втратою якості), а інше чистим (тобто оригінальна версія пошкодженного зображення, без дефектів). Ці пари створюють навчальний набір,

який модель використовує для вивчення способів відновлення пошкоджених зображень.

Зазвичай в якості моделей використовуються нейронні мережі, такі як конволюційні мережі (CNN), які навчаються відтворювати вихідні зображення, порівнюючи пошкоджені та чисті версії. Модель намагається максимально точно відтворити чисте зображення на основі пошкодженого.

Тестування моделі для оцінки її здатності відновлювати пошкоджені частини після завершення навчання проводиться на нових зображеннях, які не використовувалися в процесі навчання. Кількісна оцінка зазвичай виконується за допомогою метрик якості, таких як PSNR (Peak Signal-to-Noise Ratio) або SSIM (Structural Similarity Index). Супервізоване навчання застосовується для відновлення старих або пошкоджених фотографій, артефактів у зображеннях, усунення шумів та інших дефектів, що можуть знижувати якість зображень.

Також цей підхід знайшов використання у відеоредакторах, медичних зображеннях та інших областях, де важлива якість та чіткість зображень. Він дозволяє моделі вивчати зв'язки між пошкодженими та чистими зображеннями, що дозволяє їй ефективно відновлювати частини або всі пошкоджені дані [1].

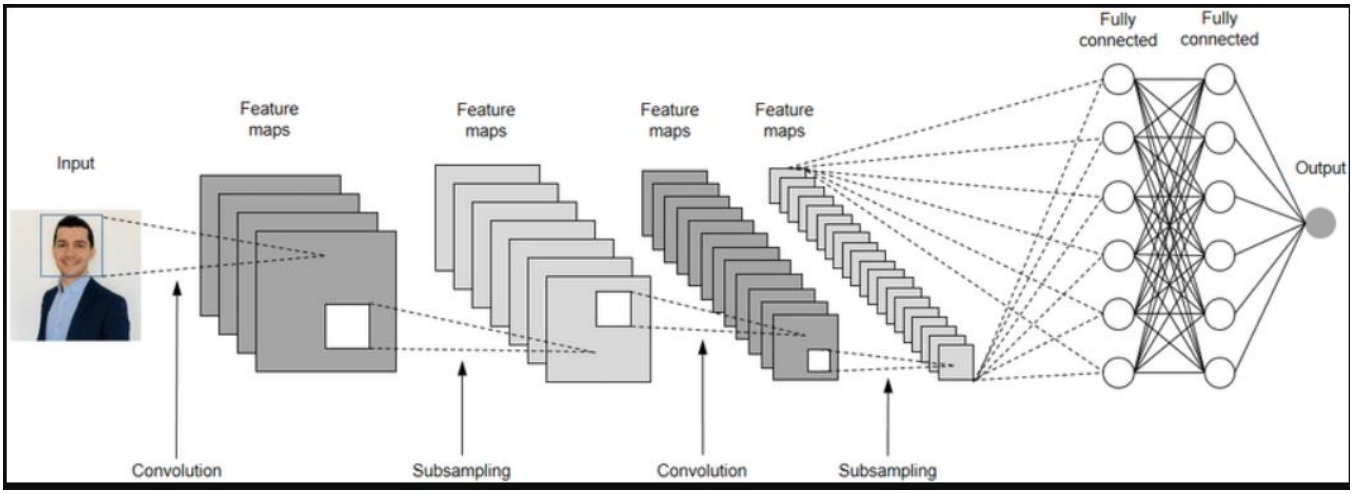

Рисунок 2.3 – Схема роботи CNN

Переваги:

- контроль кінцевого результату. Оскільки використовуються парні зображення, модель може вчитися специфічним шаблонам пошкоджень, що допомагає у кращій відновленні;

- більш ефективна точність відновлення. Завдяки доступу до оригінальних зображень, моделі можуть навчатися точніше відновлювати втрачені деталі та структури;

- можливість контролювати типи пошкоджень. Цей підхід дозволяє створювати датасети з різними типами пошкоджень, що робить модель більш універсальною.

Недоліки:

- залежність від наявності парних зображень. Для супервізованого навчання потрібні парні дані, тобто пошкоджені та непошкоджені версії одних і тих самих зображень. Це може бути складно отримати для старих фотографій;

- обмежений набір пошкоджень для навчання. Модель може бути обмеженою тими типами пошкоджень, які представлені у навчальному наборі. Вона може виявитися менш ефективною при реставрації нових типів пошкоджень;

- вимоги до ресурсів та часу. Навчання моделей для супервізованого відновлення може вимагати значних обчислювальних ресурсів та часу, особливо при великих обсягах даних [1].

2.2.4 Застосування варіаційних автоенкодерів

Застосування варіаційних автоенкодерів у відновленні старих фотографій - це складний, але дуже ефективний процес. VAEs представляють собою клас нейронних мереж, які використовуються для автоматичного відтворення та

покращення якості зображень, особливо тих, що пошкоджені, старі або низької якості. Варіаційні автоенкодери мають дві основні частини - енкодер і декодер. Енкодер перетворює вхідне зображення на вектор у латентному просторі (зазвичай з низьким розміром), що є універсальним представленням зображення. Декодер використовує цей вектор, щоб відновити вихідне зображення. Особливість VAEs полягає у тому, що вони використовують варіаційну інференцію. Замість того, щоб просто вивчати латентний простір, ця модель навчається генерувати параметри розподілу так, що при енкодуванні вхідного зображення отримується статистичний розподіл у латентному просторі, який потім можна використовувати для генерації нових зображень [1].

Після навчання на великому наборі старих зображень VAE може використовуватись для відновлення пошкоджених або зіпсованих фотографій. За його допомогою можна автоматично виявляти та усувати шуми, відновлювати втрачені деталі, покращувати контраст та кольори.мТакож VAE можуть бути використані для реконструкції зображень низької якості. Вони здатні покращити роздільну здатність, зменшуючи розмазання та роблячи зображення більш чіткими та деталізованими.

Деякі варіанти VAEs можуть бути оптимізовані для роботи в реальному часі, що означає, що вони можуть навіть швидко відновлювати старі зображення без значної затримки. VAEs відіграють ключову роль у відновленні старих фотографій, дозволяючи автоматично усувати дефекти, відтворювати втрачені деталі та покращувати загальну якість зображень, що робить їх цінними інструментами для реставрації і реконструкції старих архівних матеріалів [1].

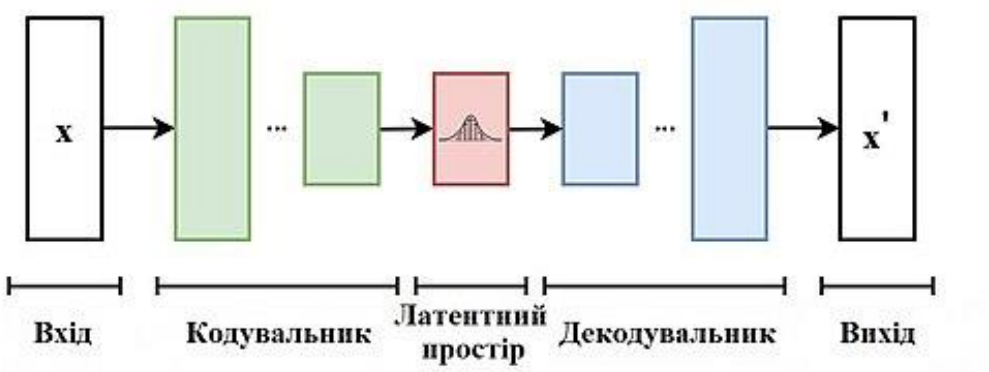

Рисунок 2.4 – Схема варіаційного автоенкодера

Переваги:

- здатність до генеративної моделювання: VAEs можуть генерувати нові зображення на основі вивчених розподілів у векторному просторі. Це може бути корисним для створення нових варіантів старих зображень;

- здатність до репрезентації даних: Вони створюють неперервний простір латентного представлення, що може допомогти у вилученні та репрезентації корисної інформації зі старих фотографій;

- здатність до генерації нових зображень: Можливість генерації нових зображень з декількох векторів в латентному просторі може допомогти в створенні варіацій старих фотографій.

Недоліки:

- обмежена точність відновлення: Оскільки VAEs моделюють генерацію зображень на основі статистичних розподілів, їх точність відновлення деяких деталей може бути обмеженою;

- потреба у великому обсязі даних: Навчання VAEs може вимагати значної кількості даних, особливо для старих фотографій, що може бути складно отримати;

- складність налаштування гіперпараметрів: Підбір гіперпараметрів може бути складним завданням, і невірний вибір може призвести до неадекватних результатів відновлення [1].

2.2.5 Трансформери

Останнім часом трансформери також стали популярними для відновлення зображень. Трансформери - це архітектура глибоких нейронних мереж, спеціалізована на обробці послідовних даних, таких як текст, але також успішно застосовується в обробці зображень.

Їх використовують для вирішення завдань перекладу, генерації тексту, а також для відновлення та покращення зображень. Основна ідея полягає у застосуванні механізму уваги (attention mechanism), який дозволяє мережі звертати увагу на різні частини вхідних даних при обробці. Вони складаються з кількох блоків уваги (attention blocks), кожен з яких має підблоки для квантування уваги на різні частини послідовності. Трансформери успішно використовуються для завдань відновлення та покращення зображень. Вони можуть аналізувати структуру зображення, виявляти дефекти, відновлювати втрачені частини, підсилювати деталі та покращувати загальну якість. У контексті відновлення старих фотографій, трансформери можуть використовувати механізм уваги для виявлення дефектів і артефактів на зображенні [1].

Вони намагаються локалізувати шуми, дефекти та відновлювати втрачені деталі шляхом аналізу різних частин зображення. Також вони можуть використовувати свою здатність до генерації нових частин зображення для реконструкції старих частин, підвищуючи якість та деталізацію відновлених фрагментів.

Трансформери є потужними у роботі з послідовними даними, але використання їх у відновленні зображень може вимагати значних обчислювальних ресурсів та часу для тренування.

Однак вони можуть бути дуже ефективними в покращенні зображень та відновленні їх якості. В цілому, трансформери є потужними інструментами для аналізу та відновлення зображень, вони можуть виявляти та усувати дефекти, відновлювати втрачені деталі та покращувати якість зображення. Тим не менш, їхня ефективність може залежати від обсягу даних та обчислювальних можливостей. Ці підходи використовуються для відновлення старих фотографій, усуваючи шуми, відновлюючи деталі та покращуючи загальну якість зображень, що може бути корисним для відтворення та збереження історичних артефактів [1].

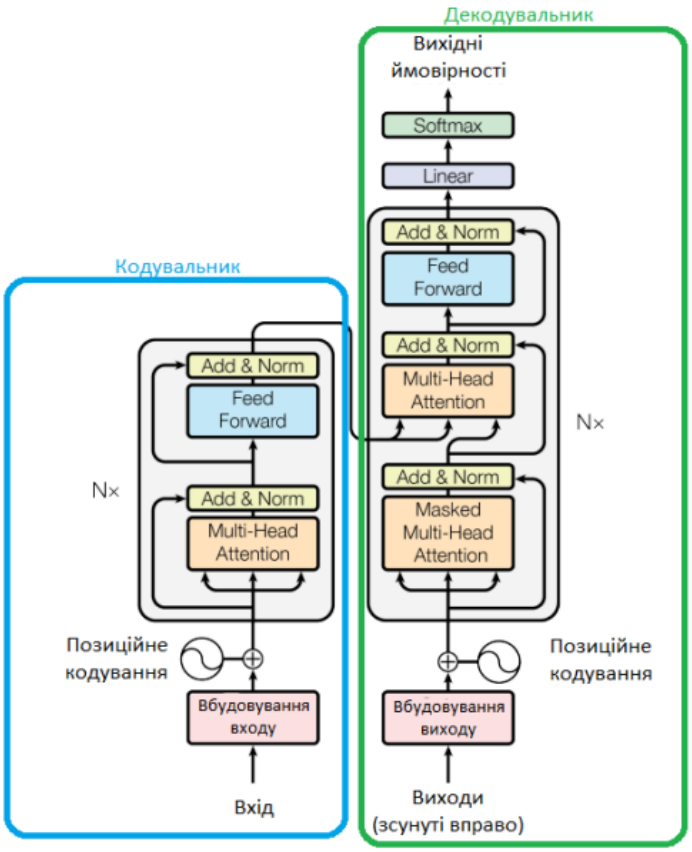

Рисунок 2.5 – Схема Трансформера

Переваги:

- можливість використання контекстуальних залежностей: Трансформери можуть використовувати глибокі контекстуальні залежності для аналізу і відновлення старих зображень, що дозволяє їм краще розуміти шаблони та структури у зображеннях;

- паралельна обробка послідовностей: Вони можуть обробляти послідовності даних паралельно, що дозволяє прискорити процес обробки та використовувати великі обсяги даних без великого збільшення часу навчання;

- здатність до глибокого представлення: Трансформери можуть забезпечити більш глибокі та абстрактні представлення зображень, що допомагає у відновленні структур та деталей.

Недоліки:

- велика обчислювальна складність: Трансформери можуть бути обчислювально витратними, особливо при великому розмірі зображень, що може ускладнити їх застосування для обробки великих даних;

- потреба в великій кількості даних: Як і для більшості глибоких моделей, трансформери можуть вимагати великої кількості даних для ефективного навчання, що може бути важким для забезпечення для старих фотографій;

- складність у використанні: Вони можуть вимагати налагодження багатьох гіперпараметрів та уваги до деталей архітектури для досягнення оптимальних результатів [1].

3 ПОРІВНЯЛЬНИЙ АНАЛІЗ МОДЕЛЕЙ ШІ

Різні ШІ мають свої певні особливості, час потрібний на їх тренування, необхідні їм обчислювальні ресурси, складність використання та реалізації, доступність та різні задачі, в котрих кожна з моделей краща за іншу. Порівняємо ці моделі та визначимо їх корисність в тих чи інших процесах.

3.1. Навчання дискримінатора

Розглянемо навчання моделі генеративно-змагальної мережі Навчання GAN складається з двох основних етапів: навчання дискримінатора та навчання генератора. Зазвичай генератори та дискримінатори GAN навчаються окремо, але можливе й одночасне навчання. Це називається почерговим навчанням. У цьому розділі ми детально розглянемо основні принципи, що лежать в основі кожного з цих процесів.

Навчання дискримінатора є першим етапом навчання GAN. Функціональність моделі генератора залишається незмінною, тоді як функціональність дискримінатора може коливатися для створення «ідеального» дискримінатора. Вхідні дані для навчання дискримінатора походять як з реальних прикладів (тобто вхідних даних, які класифікуються як позитивні приклади), так і з генератора (тобто вхідних даних, які класифікуються як негативні приклади). Обидва типи даних потім класифікуються дискримінатором; їхні прогнози порівнюються з істинним значенням для визначення загальних втрат. За допомогою зворотного поширення дискримінатор отримує штрафні бали за помилки, внаслідок чого змінюються ваги в процесі прийняття рішень. Кожен цикл, під час якого повний навчальний набір даних подається на дискримінатор

для навчання, називається епохою. Протягом процесу навчання відбувається кілька епох, і з кожною епохою ефективність дискримінатора покращується. Навчання зупиняється, коли додаткові епохи мало впливають на ефективність або коли досягається максимально дозволена кількість епох. Розмір партії, критичний параметр, що визначає кількість зразків, які обробляються одночасно, суттєво впливає на ефективність моделі та стабільність навчання [7].

Оптимізація розміру партії є непростим завданням і вимагає емпіричного тестування для збалансування обчислювальної ефективності та точності моделі. Комп'ютерний науковець найкраще підготовлений для точного налаштування цього параметра. Підсумовуючи, навчання дискримінатора передбачає ітеративну оптимізацію з використанням реальних та згенерованих даних для покращення його здатності точно класифікувати вхідні дані, причому продуктивність стабілізується після достатньої кількості епох.

3.2. Навчання генератора

У процесі навчання генеративна модель отримує вхідні дані та намагається створити правдоподібне зображення на виході. Як згадувалося раніше, цей вихідний результат виступає як приклад навчання для моделі дискримінатора, який потім намагається відрізнити синтезований вихідний результат генератора від реального зображення. Дискримінатор накладає штраф на генератор за те, що його розпізнали. Подібний процес для дискримінатора відбувається з генератором: коригування ваг за допомогою зворотного поширення дозволяє вдосконалити генератор, а навчання зупиняється, коли кожна додаткова епоха додає незначну цінність до результату продуктивності або коли досягається максимально дозволена кількість епох.

Зміни в дискримінаторі в процесі навчання ускладнюють досягнення відповідної кінцевої точки для генератора. Тому функціональність моделі дискримінатора залишається незмінною, тоді як функціональність генератора змінюється.

Підсумовуючи, навчання генератора зосереджується на покращенні здатності моделі створювати реалістичні результати шляхом мінімізації втрат через ітеративні оновлення, прагнучи до стабільної збіжності (тобто точки, де функція втрат мінімізована, а параметри навчання досягають стабільного стану, що дозволяє генерувати точні результати) [7].

3.3. Альтернативне навчання

Альтернативний підхід до навчання на сьогодні є стандартною практикою для сучасних GAN і полягає в одночасному навчанні генератора та дискримінатора. У цьому процесі дискримінатор може навчатися протягом заданої кількості епох, а потім той самий процес застосовується для навчання генератора. Цей процес повторюється багато разів, доки не буде досягнуто мінімального покращення функціональності генератора.

Перевага цього методу полягає в тому, що він теоретично покращує функціональність кінцевого генератора, оскільки як дискримінатор, так і генератор стикаються з більш складними завданнями в міру прогресування навчання. Однак це може ускладнити збіжність. Вибір швидкості навчання, ще одного ключового параметра, безпосередньо впливає на швидкість збіжності та стабільність навчання. Висока швидкість навчання може спричинити нестабільність, тоді як низька швидкість може призвести до повільної збіжності. Для навчання GAN перевагу надають графікам швидкості навчання або адаптивним методам, таким як оптимізатори Adam. Визначення оптимальної

конфігурації часто вимагає експертних знань у галузі машинного навчання. У міру зниження продуктивності дискримінатора функція генератора покращується.

Протягом процесу навчання функція генератора продовжує вдосконалюватися, навчаючись на основі зворотного зв'язку, що надається дискримінатором. В результаті генератор може створювати зображення, які дуже точно імітують реальні зображення, роблячи їх майже нерозрізнимі для дискримінатора і, потенційно, для людських спостерігачів. На цьому етапі навіть найідеальніша ефективність дискримінатора не буде кращою за випадковість або 50%.

Якщо мережа продовжує навчатися, то модель генератора адаптується до зворотного зв'язку від дискримінатора, який надаватиме неточний зворотний зв'язок.

Це явище знижує загальну функціональність моделі генератора та ілюструє приклад неможливості збіжності. Хоча цей метод покращує обидві моделі, він може ускладнити збіжність через суперечливу динаміку. Для стабілізації навчання часто використовуються варіанти градієнтного спуску, такі як RMSProp або AdaGrad. Вибір алгоритму оптимізації значною мірою залежить від застосування та характеристик даних.

Це завдання найкраще залишити експертам у галузі машинного навчання. Підсумовуючи, альтернативне навчання GAN покращує ефективність обох моделей завдяки одночасній оптимізації, але суперечлива динаміка може ускладнити збіжність [7].

3.4. Категорії вхідних даних

Вхідні дані слугують відправною точкою для генератора при створенні синтезованих зображень. У перших моделях вхідними даними для GAN були зображення з шумом. У процесі навчання GAN навчаються створювати цілісне зображення на основі цього шуму. Більш практичні застосування вхідних даних GAN зазвичай включають текстові, графічні або відео файли.

Під час навчання ці вхідні дані поєднуються з вихідними цільовими зображеннями, причому метою GAN є генерація цільового зображення. Як і у випадку з усіма програмами машинного навчання, існує значна залежність від вихідних даних.

3.5. Оціночні показники

Для опису «реалістичності» результатів роботи GAN використовуються два терміни: точність та різноманітність. Термін «точність» зазвичай стосується загальної якості отриманого зображення.

Це може залежати від різних характеристик, таких як, зокрема, текстура, структура та деталізація. Різноманітність стосується різноманітності зображень, які можуть генерувати GAN. GAN, що генерують лише один тип зображення або дублікати схожого зображення, мають обмежену практичну цінність і, отже, вважаються такими, що мають низьку різноманітність. Функціональні GAN повинні бути здатні генерувати зображення як з високою точністю, так і з різноманітністю. У задачах дискримінації в машинному навчанні загальна точність, чутливість та специфічність зазвичай використовуються для вимірювання продуктивності нейронної мережі. У генеративних нейронних мережах завдання оцінки є більш складним. Основна проблема полягає в тому,

що «реальне зображення» важко визначити. Тому, на відміну від завдань класифікації, де алгоритм може чітко розрізнити дві категорії, оцінка того, чи є зображення «реальним» чи «нереальним», передбачає суб'єктивне судження (тобто людський вхід) та оцінку багатьох складних характеристик зображення [9].

3.6. Якісні та кількісні методи (піксельна втрата)

Існує дві категорії методів оцінки GAN: якісні методи та кількісні методи. Якісні методи зазвичай передбачають участь оцінювачів-людей і охоплюють, але не обмежуються, методами, що використовують маскованих оцінювачів для оцінки результатів роботи GAN. Поширеною практикою є порівняння вихідних даних GAN з їхніми еталонними зображеннями. Однак такі методи вимагають значних людських ресурсів і схильні до упередженості.

І навпаки, кількісні методи не передбачають людського судження, але можуть бути складними у впровадженні. Кількісні методи охоплюють інструменти, які можуть порівнювати варіації пікселів результатів GAN з їхніми відповідними еталонними аналогами, які повинні бути доступними для порівняння. Ця оцінка відхилення піксель за пікселем називається втратою на рівні пікселів. Втрата на рівні пікселів дозволяє більш об'єктивно оцінити ефективність GAN [8].

Однак цей підхід не є бездоганним показником ефективності GAN, оскільки прості відхилення значень пікселів від еталонних даних не корелюють безпосередньо з реалістичністю. Синтетичне зображення може мати обмежену варіацію пікселів порівняно з його еталонним аналогом, але невідповідність у невеликій області може поставити під сумнів реалістичність усього зображення. Наприклад, якщо GAN генерує зображення людської руки з шістьма пальцями

замість п’яти, реалістичність усього зображення буде поставлена під сумнів. Аналогічно, якщо синтетичне зображення має великі відхилення пікселів, але не містить несумісності, втрата на рівні пікселів недооцінить реалістичність [8].

3.7. Інші об’єктивні методи

До інших об’єктивних показників належать використання SSIM, Inception Score (IS) або Frechet Inception Distance (FID). SSIM порівнює результати GAN з їхніми еталонними аналогами, аналізуючи такі характеристики, як яскравість, контрастність та структура між двома зображеннями. Хоча це і забезпечує більш надійний показник, ніж просте порівняння пікселів, цей метод все ж не вимірює реалістичність зображення. IS використовує модель класифікації, попередньо навчену на великому анотованому наборі даних зображень (наприклад, ImageNET для Inception v3), для оцінки згенерованих результатів [4].

Модель Inception v3 стала популярною, оскільки цей підхід має високу кореляцію з оцінками людей. Однак важливим обмеженням клінічного застосування Inception V3 є висока залежність моделі від бази даних, використаної під час попереднього навчання. Оскільки ImageNET має обмежену кількість медичних зображень, модель, на жаль, іноді має обмежену цінність у медичних застосуваннях.

FID також став популярним показником ефективності для GAN. У цій методології модель Inception v3 використовується для вилучення ознак як із реальних, так і з генерованих зображень, а потім для генерування розподілу. Різниця між цими розподілами відома як FID. Однією з переваг FID є те, що він може виявляти колапс мод, на відміну від IS.

Перенавчання (overfitting) – це небажана поведінка алгоритму, який не здатний узагальнювати та занадто тісно пристосовується до своїх навчальних

даних, що призводить до його нездатності правильно виконувати точні прогнози. Корисною аналогією для розуміння цього поняття є студент-інженер, який запам'ятав усі фізичні вправи у своєму розділі, замість того щоб зрозуміти саму фізику. Наразі не існує простого способу оцінки перенавчання в GAN, а всі метрики мають свої основні недоліки і не мають чіткого золотого стандарту [4].

Ідеальний показник для оцінки функціональності GAN повинен визначатися безліччю факторів, включаючи архітектуру мережі, характер представлених даних та завдання, що виконується. На жаль, конкретні порогові значення для візуалізації сітківки не можуть бути універсально стандартизовані через властиву методам візуалізації мінливість та різноманітний спектр захворювань, що зустрічаються в клінічній практиці. Кожна техніка візуалізації має власний набір параметрів та чутливості, що впливають на визначення порогових значень. Більше того, порогові значення часто потрібно адаптувати до конкретної патології, що досліджується, оскільки різні захворювання сітківки можуть вимагати різних діагностичних критеріїв [9].

З огляду на складність оптимізації цих порогових значень для різних методів та захворювань, внесок експерта з інформатики є необхідним для розробки та налагодження алгоритмів, які можуть динамічно коригувати та застосовувати ці параметри для підвищення точності діагностики [9].

3.8. Функція втрат

Функції втрат оцінюють моделі (генератор або дискримінатор) у процесі навчання. Функції втрат не тільки дозволяють користувачеві оцінити ефективність генератора, але й забезпечують необхідний зворотний зв’язок для моделі генератора з метою її оптимізації за допомогою зворотного поширення та методу градієнтного спуску. Ці функції втрат відображають різницю між

розподілом згенерованих даних та розподілом реальних даних. Мета навчання полягає у створенні дискримінатора, ефективність якого є максимальною і який не здатний відрізнити реальні дані від підроблених, наданих генератором, ефективність якого також є максимальною. Існує три основні функції втрат, які були застосовані і які ми обговоримо в наступних розділах [4].

3.9. Мінімакс-втрата

Перша функція втрат, описана в статті, що представила GAN, автором якої є Ian Гудфеллоу, називається мінімакс-втратою – для її математичного пояснення радимо звернутися до розділу 2.1. Однак ця функція втрат мала недолік: іноді на початку навчання дискримінатор виконував своє завдання розрізнення синтетичних зображень настільки добре, що генератор не встигав за ним (тобто створювати зображення), що призводило до передчасного припинення навчання. Щоб вирішити цю проблему, були розроблені модифікована мінімакс-втрата та функція втрат Вассерштейна [4].

Модифікована мінімакс-втрата та втрата Вассерштейна: модифікована мінімакс-втрата намагається протидіяти передчасному припиненню навчання, визначаючи втрату генератора на основі ймовірності того, що дискримінатор вважає синтетичне зображення реальним, а не порівнюючи прямий розподіл реальних точок даних із синтетичними.

Математично цей підхід визначається як $LG = -Ez{\sim}Pz[\log(D(G(z)))]$. Втрата Вассерштейна, яку часто називають Wasserstein GAN або wGAN, оскільки вона є розширенням класичної GAN, намагається вирішити попередню проблему за допомогою іншого підходу. Вихідні дані дискримінатора Wasserstein GAN відрізняються від вихідних даних класичного дискримінатора GAN. Перший надає оцінку «реальності» (вища оцінка = більш реалістичне зображення), тоді як

класична GAN надає ймовірність, число від 0 до 1, того, що зображення є реальним. Математично втрата Вассерштейна визначається як LG = −Ez~Pz[D(G(z))]. Було продемонстровано, що втрата Вассерштейна зменшує ймовірність дострокового припинення навчання.

Можна створити порогові значення для визначення, чи є зображення реальним чи синтетичним. Метою дискримінатора GAN за Вассерштейном є максимізація точності його оцінки [8].

3.10 Інші функції втрат

Інші варіації функцій втрат включають використання середньоквадратичної втрати. Для стандартних GAN логарифмічна втрата використовується в логарифмічній функції протягом навчання. Однак існують дані, що свідчать про те, що використання середньоквадратичної втрати збільшує ймовірність того, що синтетичні зображення будуть наближатися до своїх реальних аналогів [11].

Усі функції втрат можуть бути корисними для оцінки функціональності GAN; однак основною функцією функцій втрат є керування процесом навчання.

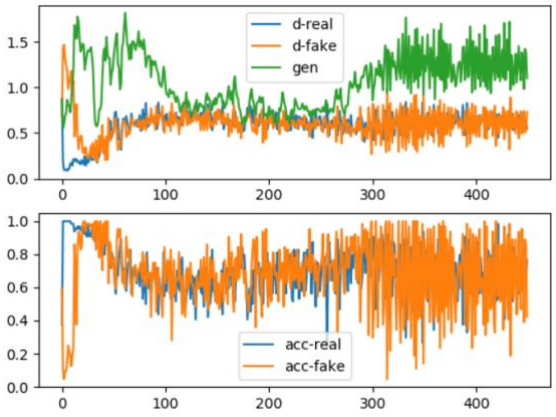

Рисунок 3.1 – Лінійні графіки втрат і точності для стабільної генеративно-змагальної мережі

На рисунку 3.1 зображено два графіки. На верхньому графіку показано лінійні графіки втрат дискримінатора для реальних зображень (синій колір), втрат дискримінатора для згенерованих фейкових зображень (помаранчевий колір) та втрат генератора для згенерованих фейкових зображень (зелений колір).

Ми бачимо, що всі три втрати є дещо нестабільними на початку запуску, перш ніж стабілізуватися приблизно між 100 та 300 епохами. Після цього втрати залишаються стабільними, хоча дисперсія збільшується.

Це приклад нормальної або очікуваної втрати під час навчання. А саме, втрата дискримінатора для реальних і фальшивих зразків приблизно однакова і становить близько 0,5, а втрата генератора дещо вища — від 0,5 до 2,0.

Якщо модель генератора здатна генерувати правдоподібні зображення, то очікується, що ці зображення були б згенеровані між епохами 100 і 300, а також, ймовірно, між 300 і 450.

Нижній підграфік показує лінійний графік точності дискримінатора для реальних (синіх) та фальшивих (помаранчевих) зображень під час навчання. Ми бачимо структуру, схожу на підграфік втрат, а саме: точність спочатку значно відрізняється між двома типами зображень, потім стабілізується між епохами 100–300 на рівні близько 70–80% і залишається стабільною після цього, хоча з підвищеною варіацією.

Часові масштаби (наприклад, кількість ітерацій або епох навчання) для цих закономірностей та абсолютних значень будуть різнитися залежно від задач та типів моделей GAN, хоча графік надає хорошу базову лінію для того, чого слід очікувати під час навчання стабільної моделі GAN [11].

3.11. Відновлення зображень

GAN, призначені для відновлення зображень, використовуються для заповнення зображень, у яких відсутні пікселі або є артефакти (рис. 4).

Крім того, GAN можна застосовувати для генерації проміжних синтетичних зрізів між двома зрізами, отриманими з зображень.

Цю технологію можна використовувати для генерації тонших зрізів з ОКТ на основі протоколу візуалізації з більшою товщиною або для генерації фаз флуоресцентної ангіографії, які не були зафіксовані [7].

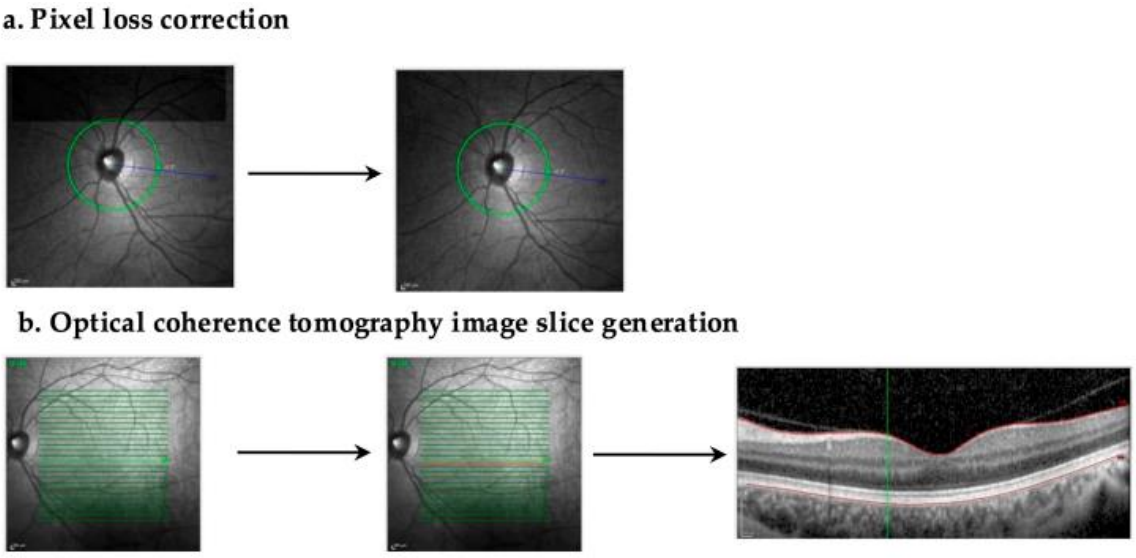

Рисунок 3.2 – Приклади застосування методу відновлення зображень в офтальмології. Метод відновлення зображень може використовуватися для різних завдань в офтальмології, таких як заповнення чорних ділянок на фотографіях (а) та створення проміжних синтезованих зрізів між знімками, отриманими за допомогою оптичної когерентної томографії (б).

3.12. Різновиди GAN

Умовні GAN (cGAN) функціонують аналогічно до стандартних GAN, однак відрізняються тим, що дозволяють користувачеві визначати умовні змінні для зміни процесу навчання. Це означає, що GAN можна навчати, використовуючи не лише надані дані, а й додаткову інформацію.

Теоретично, умова може бути застосована до будь-якої змінної під час навчання; однак, наразі проводяться дослідження, що оцінюють вплив однієї змінної стану, зображень з різних доменів, замаскованих зображень та зображень, керованих тепловою картою. Можливість надавати додаткову інформацію GAN дозволяє краще контролювати процес навчання та бажані характеристики кінцевої моделі [7].

Мультимодальні GAN

У сучасних застосуваннях GAN використовуються мультимодальні вхідні дані. GAN можна застосовувати для перекладу між текстом, зображеннями, мовленням та відео. У медичній галузі GAN типу «мовлення-до-тексту» використовуються для створення клінічних записів шляхом запису аудіофайлів під час клінічних взаємодій. У галузі науки про зір GAN «відео-в-мовлення» використовуються для допомоги пацієнтам із порушеннями зору орієнтуватися у своєму оточенні.

Трансляційні GAN

Трансляційні GAN дозволяють перетворювати один вхідний сигнал на інший вихідний з визначеною властивістю.

Більшість досліджень у сфері медичного застосування GAN зосереджувалася саме на цих типах GAN, з особливим акцентом на моделях «зображення-до-зображення».

GAN «зображення-зображення» використовуються для перетворення ультразвукових зображень та комп'ютерних томограм у МРТ-зображення. У галузі офтальмології дослідження продемонстрували, що фотографії очного дна можна перетворити на відповідні зображення автофлуоресценції флуоресцеїну.

Однією з найбільш вивчених трансляційних GAN є CycleGAN завдяки її простоті та високій ефективності у генеруванні правдоподібних результатів.

CycleGAN функціонує для перетворення зображень з одного розподілу даних у зображення, що належать до іншого розподілу даних з відмінними характеристиками. На відміну від інших GAN, CycleGAN має дві генераторні та дві дискримінаторні функції [8].

У цій конфігурації вихідні дані першого генератора використовуються як вхідні дані для другого генератора. З цієї причини CycleGAN не потребує міток або парної відповідності, оскільки цикл між генераторами дозволяє моделі навчатися характеристикам двох окремих доменів зображень. Це спрощує створення наборів даних для CycleGAN і дозволяє розробляти мережі для шумозаглушення без парних наборів даних. Для трансляційних GAN «зображення-до-зображення» втрату можна оцінити, порівнюючи пікселі синтезованих зображень з їхніми відповідниками в еталонних даних [8].

3.13 Застосування актоенкодерів

Розмірність латентного простору є одним із ключових аспектів роботи з автоенкодерами є правильний вибір розмірності латентного простору, що безпосередньо впливає на обсяг інформації, яку модель може зберегти під час стиснення даних [10].

Невелика кількість вимірів – забезпечує краще стиснення та покращену генералізацію даних, але за рахунок втрати деталей і зниження якості реконструкції.

Висока кількість вимірів – дозволяє зберегти більше інформації та забезпечує точнішу реконструкцію, але знижує рівень стиснення і може призвести до перенавчання моделі.

Вибір кількості вимірів – це компроміс між стисненням та якістю реконструкції. Оптимальне значення залежить від типу даних та мети, яку ми хочемо досягти, і на практиці визначається емпірично.

Автоенкодери використовуються в різних галузях, зокрема:

- відновлення зображень та усунення шуму,
- зменшення розмірності,
- виявлення аномалій у даних,
- стиснення даних та зображень.

3.14 Глибокий конволюційний автоенкодер для відновлення зображень

Запропонована архітектура в основному складається з ланцюжка конволюційних шарів та симетричних деконволюційних шарів.

Як показано на рисунку 3.3, стрибкові зв’язки симетрично з’єднують конволюційні шари з деконволюційними. Метод назвали «RED-Net» - дуже глибокі резидуальні мережі кодера-декодера [10].

3.14.1 Архітектура

Ця архітектура є повністю конволюційною (та деконволюційною. Деконволюція, по суті, є конволюцією з відновленням дискретизації). Після

кожної конволюції та деконволюції додаються випрямні шари. Для завдань відновлення зображень низького рівня не використовується в мережі ані пулінгу, ані розпулінгу, оскільки зазвичай пулінг відкидає корисні деталі зображення, які є необхідними для цих завдань. Варто зазначити, що оскільки конволюційні та деконволюційні шари є симетричними, мережа, по суті, здійснює прогнозування на рівні пікселів, тому розмір вхідного зображення може бути довільним. Вхідні та вихідні дані мережі – це зображення однакового розміру w × h × c, де w, h та c – це ширина, висота та кількість каналів.

Основна ідея полягає в тому, що конволюційні шари діють як екстрактор ознак, який зберігає основні компоненти об'єктів на зображенні, одночасно усуваючи спотворення. Після проходження через конволюційні шари спотворене вхідне зображення перетворюється на «чисте». Під час цього процесу можуть бути втрачені дрібні деталі вмісту зображення.

Потім деконволюційні шари комбінуються для відновлення деталей вмісту зображення. Виходом деконволюційних шарів є відновлена чиста версія вхідного зображення.

Крім того, додаються пропускні з'єднання від конволюційного шару до відповідного дзеркального деконволюційного шару. Передані конволюційні карти ознак сумуються з деконволюційними картами ознак по елементах і передаються до наступного шару після виправлення.

Виходячи з вищезазначеної архітектури, у експериментах використовувались дві мережі, які мають відповідно 20 і 30 шарів, для видалення шуму з зображень, надвисокої роздільної здатності зображень, деблокування JPEG та відновлення зображень.

3.14.2 Деконволюційний декодер

Останнім часом для семантичної сегментації були запропоновані архітектури, що поєднують шари згортки та деконволюції. На відміну від згорткових шарів, у яких кілька вхідних активацій у межах вікна фільтра об’єднуються для отримання єдиної вихідної активації, деконволюційні шари пов’язують одну вхідну активацію з кількома вихідними. Деконволюція зазвичай використовується як навчальні шари з підвищенням частоти дискретизації.

У мережі конволюційні шари послідовно зменшують частоту дискретизації вмісту вхідного зображення до невеликого розміру абстракції. Потім деконволюційні шари підвищують частоту дискретизації абстракції назад до її початкової роздільної здатності.

Окрім використання пропускних з'єднань, основною відмінністю між цією моделлю є те, що мережа є повністю конволюційною та деконволюційною, тобто без пулінгу та депулінгу. Причина полягає в тому, що для відновлення зображень низького рівня метою є усунення пошкоджень низького рівня з одночасним збереженням деталей зображення, а не навчання абстракцій зображення.

На відміну від високорівневих застосувань, таких як сегментація або розпізнавання, пулінг зазвичай усуває багаті деталі зображення і може погіршити ефективність відновлення [10].

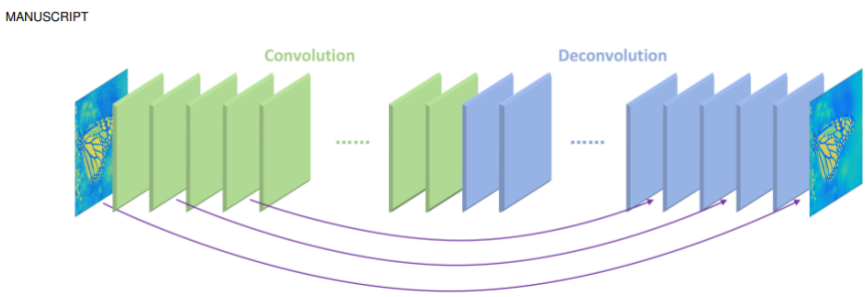

Рисунок 3.3 – Загальна архітектура мережі. Мережа складається з шарів симетричної згортки (кодер) та дезгортки (декодер).

Стрибкові зв’язки підключаються кожні кілька (у наших експериментах - два) шари від згорткових карт ознак до їхніх дзеркальних дезгорткових карт ознак. Відповідь від згорткового шару безпосередньо передається до відповідного дзеркального дезгорткового шару як у прямому, так і в зворотному напрямках.

Можна просто замінити деконволюцію на конволюцію, що дасть архітектуру, дуже схожу на нещодавно запропоновані дуже глибокі повністю конволюційні нейронні мережі. Однак між повністю конволюційною моделлю та нашою моделлю існують істотні відмінності.

Візьмемо для прикладу видалення шуму з зображень. Порівнюємо 5-шарову та 10-шарову повністю конволюційні мережі з цією мережею (що поєднує конволюцію та деконволюцію, але без пропускних з'єднань). Для повністю конволюційних мереж використовуємо заповнення або підвищення частоти дискретизації вхідних даних, щоб вхідні та вихідні дані мали однаковий розмір.

У цій мережі перші 5 шарів є конволюційними, а другі 5 шарів – деконволюційними. Усі інші параметри навчання є ідентичними, тобто навчання проводиться за допомогою SGD зі швидкістю навчання 10−6, рівнем шуму $\sigma = 70$.

Наведено пікове відношення сигнал/шум (PSNR) на валідаційному наборі, яке показує, що використання деконволюції працює краще, ніж повністю конволюційний аналог, як показано на рисунку 3.4. Після досягнення збіжності наша модель демонструє кращі показники PSNR, ніж моделі, з якими її порівнювали.

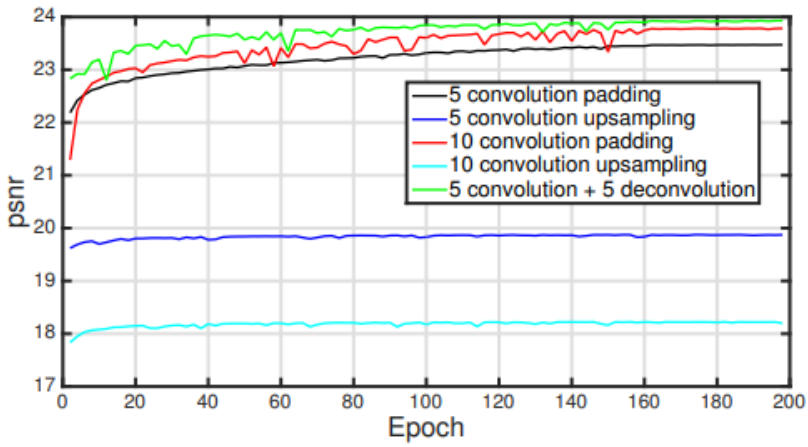

Рисунок 3.4 – Значення PSNR на валідаційному наборі під час навчання.

Крім того, на рисунку 3.5 візуалізуються деякі результати, які є вихідними даними шарів 2, 5, 8 та 10 з 10-шарової повністю конволюційної мережі та нашої мережі. У випадку повністю конволюційної мережі шум усувається поетапно, тобто рівень шуму знижується після кожного шару. Під час цього процесу деталі змісту зображення можуть втрачатися. Проте в цій мережі конволюція зберігає основний зміст зображення. Потім для відновлення деталей використовується деконволюція. Зображення зліва вгорі до правого нижнього кута: чисте зображення, зображення з шумом, вихідні дані conv-2, вихідні дані conv-5, вихідні дані conv-8 та вихідні дані conv-10, де «conv-i» позначає i-й конволюційний шар. Зображення зліва вгорі до правого низу: чисте зображення, зображення з шумом, вихід conv-2, вихід conv-5, вихід deconv-3 та вихід deconv-5, де «deconv-i» означає i-й деконволюційний шар. Частина (а) – візуалізація 10-шарової повністю конволюційної мережі, частина (б) – візуалізація 10-шарової конволюційної та деконволюційної мережі [10].

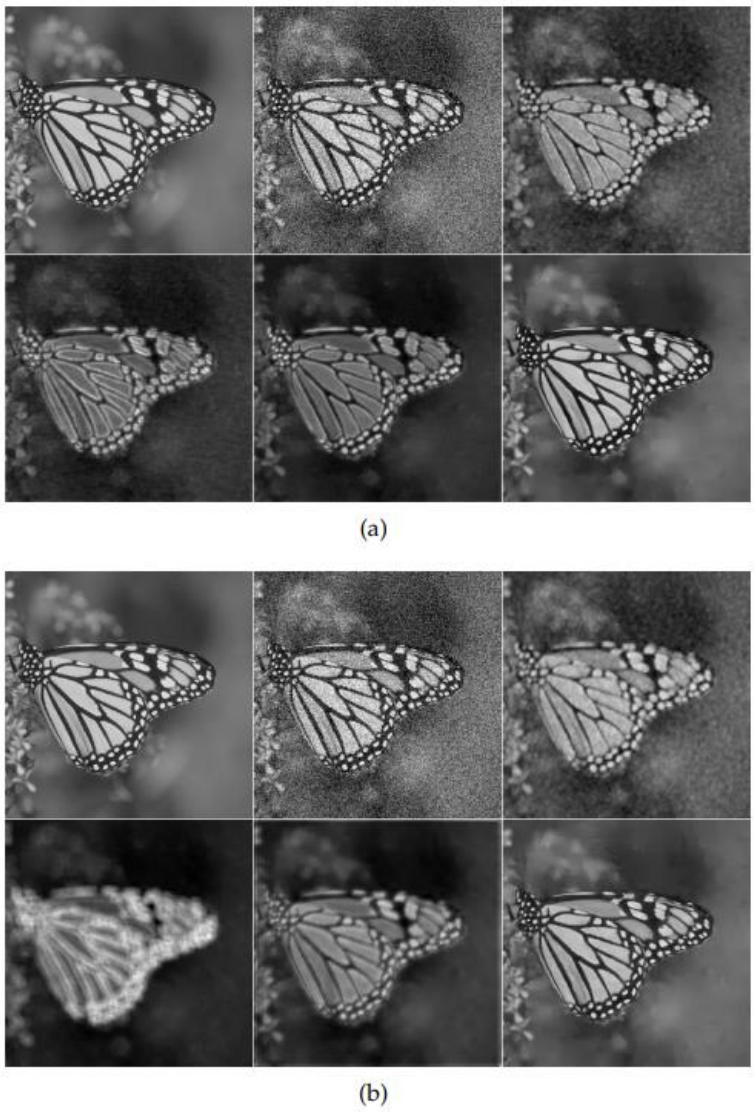

Рисунок 3.5 – (а) Візуалізація 10-шарової повністю конволюційної мережі. (б) Візуалізація 10-шарової конволюційної та деконволюційної мережі.

3.14.3. Пропустити зв’язки

Логічне запитання: чи здатна мережа з деконволюцією відтворити деталі зображення, спираючись лише на його абстракцію? Виявлено, що в неглибоких мережах, які мають лише кілька шарів згортки, деконволюція здатна відтворити деталі. Однак, коли мережа стає глибшою або використовує такі операції, як

максимальне об'єднання, навіть із шарами деконволюції, це не працює так добре, можливо, тому що занадто багато деталей вже втрачено під час згортки та об'єднання.

Друге питання полягає в тому, чи досягає ця мережа приріст продуктивності, коли вона стає глибшою? Спостерігаємо, що глибші мережі в задачах відновлення зображень, як правило, легко страждають від зниження продуктивності. Причина може бути двоякою. По-перше, із збільшенням кількості шарів згортки значна кількість деталей зображення може бути втрачена або спотворена. З огляду лише на абстракцію зображення, відновлення його деталей є недовизначеною проблемою.

По-друге, з точки зору оптимізації, глибокі мережі часто страждають від зникнення градієнтів і стають набагато складнішими для навчання – проблема, яка добре висвітлена в літературі з нейронних мереж. Щоб вирішити ці дві проблеми, натхненні мережами «шосе» та глибокими резидуальними мережами, додаємо пропускні з'єднання між двома відповідними конволюційними та деконволюційними шарами, як показано на рисунку 3.3.

Будівельний блок показано на рисунку 3.6. Є дві причини для використання таких з'єднань. По-перше, коли мережа стає глибшою, як згадувалося вище, деталі зображення можуть втрачатися, що робить деконволюцію менш ефективною у їх відновленні. Однак карти ознак, що передаються через пропускні з'єднання, містять багато деталей зображення, що допомагає деконволюції відновити покращену чисту версію зображення [10].

По-друге, пропускні з'єднання також дають переваги при зворотній передачі градієнта до нижніх шарів, що значно полегшує навчання глибшої мережі.

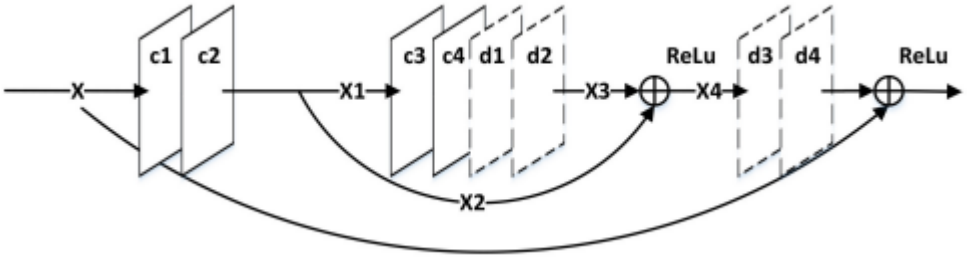

Рисунок 3.6 – Приклад будівельного блоку в запропонованій архітектурі. Прямокутник, обведений суцільною та пунктирною лініями, позначає відповідно згортку та деконволюцію. Символ ⊕ позначає поелементну суму карт ознак.

У цьому випадку прагнемо передати інформацію з конволюційних карт ознак до відповідних деконволюційних шарів. Дуже глибокі мережі типу «highway» по суті є прямими мережами з довготривалою короткочасною пам’яттю (LSTM) із воротами забування, а шари CNN глибокої резидуальної мережі є прямими LSTM без воріт.

Зауважте, що ці мережі загалом не мають формату стандартних прямих LSTM. Замість того, щоб безпосередньо навчати відображення від входу X до виходу Y, прагнемо, щоб мережа підганяла резидуал задачі, який позначається як $F(X) = Y - X$. Така стратегія навчання застосовується до внутрішніх блоків мережі кодування-декодування, щоб зробити навчання більш ефективним. Стрибкові з'єднання передаються кожні два конволюційні шари до їх дзеркальних деконволюційних шарів.

Можливі й інші конфігурації, але ці експерименти показують, що ця конфігурація вже працює дуже добре. Використання таких скорочень полегшує навчання мережі та покращує ефективність відновлення завдяки збільшенню глибини мережі [10].

3.14.4 Тренування

Загалом у цій мережі є три типи шарів: конволюційний, деконволюційний та шар посумкування по елементах. За кожним шаром слідує випрямлений лінійний елемент (ReLU). Нехай X – вхідні дані, тоді конволюційний та деконволюційний шари записуються у вигляді:

$$F\ (X) = \max(0, W_K * X + B_K), \tag{3.1}$$

де W$_k$ і B$_k$ позначають фільтри та зміщення, а $*$ позначає операцію згортки або деконволюції для зручності формулювання. Для шару елементної суми вихідним сигналом є елементна сума двох вхідних сигналів одного розміру з подальшою активацією ReLU:

$$F\ (X_1, X_2) = \max(0, X_1 + X_2). \tag{3.2}$$

Для навчання наскрізного відображення з пошкоджених зображень на чисті зображення необхідно оцінити ваги Θ, що представлені конволюційними та деконволюційними ядрами. Зокрема, маючи набір із N пар навчальних зразків {Xi, Yi}, де Xi – зображення з шумом, а Yi – чиста версія, що є еталонним значенням. Мінімізуємо наступну середньоквадратичну похибку (MSE):

$$L\ (\Theta) = \frac{1}{N} \sum_{i=1}^{N} || \ F\left(X^i; \Theta\right) - Y^i ||_F^2 \ . \tag{3.3}$$

Зазвичай мережа може безпосередньо навчитися відображенню з пошкодженого зображення на чисту версію. Однак ця мережа навчається на адитивному пошкодженні вхідних даних, оскільки між входом і виходом мережі

існує пропускне з'єднання. Ми виявили, що оптимізація для спотворення збігається краще, ніж оптимізація для чистого зображення. У крайньому випадку, якщо вхідними даними є чисте зображення, було б простіше змусити мережу здійснювати нульове відображення (навчання спотворення), ніж підганяти ідентичне відображення (навчання чистого зображення) за допомогою стека нелінійних шарів [10].

Реалізуємо та навчаємо мережу за допомогою Caffe. Емпірично виявили, що використання Adam з базовою швидкістю навчання 10^(−4) для навчання збігається швидше, ніж традиційний стохастичний градієнтний спуск (SGD). Базова швидкість навчання для всіх шарів однакова, на відміну від**,** де для останнього шару встановлюється менша швидкість навчання. Це не є необхідним у цій мережі. Зокрема, градієнти щодо параметрів і-го шару спочатку обчислюються як:

$$g = \nabla_{\Theta_i} L(\Theta_I). \tag{3.4}$$

Потім обчислюються два вектори імпульсу за формулами:

$$m = \beta_1 m + (1 - \beta_1) g, v = \beta_2 v + (1 - \beta_2) g^2. \tag{3.5}$$

Правило оновлення таке:

$$a = a\sqrt{1 - \beta_2^t}\,/(1 - \beta_1^t), \Theta_i = \Theta_i - am/(\sqrt{v} + \epsilon). \tag{3.6}$$

β1, β2 встановлюються відповідно до рекомендованих значень, наведених у. Для формування фрагментів зображень, які слугують навчальним набором для

кожного завдання відновлення зображення, використовуються 300 зображень із набору даних Berkeley Segmentation Dataset (BSD).

3.14.5. Тестування

Незважаючи на те, що мережа навчалася на локальних фрагментах, вона здатна відновлювати зображення довільних розмірів. Отримавши тестове зображення, можна просто пройти по ланцюжку мережі, яка вже зараз демонструє кращі результати, ніж існуючі методи. Щоб досягти ще кращих результатів, пропонуємо обробляти пошкоджене зображення у різних орієнтаціях. На відміну від сегментації, ядра фільтрів у цій мережі усувають лише пошкодження, що зазвичай не залежить від орієнтації вмісту зображення у завданнях відновлення низького рівня. Тому можемо повертати та дзеркально відображати ядра і виконувати проходження кілька разів, а потім усереднювати вихідні дані, щоб отримати сукупність результатів декількох тестів. Бачимо, що це може призвести до дещо кращої продуктивності [10].

3.15 Навчання з використанням симетричних пропускних з’єднань

Метою відновлення є усунення пошкоджень при максимально можливому збереженні деталей зображення.

У попередніх роботах для завдань відновлення зображень на низькому рівні зазвичай використовуються неглибокі мережі. Причиною цього може бути те, що глибші мережі можуть знищити деталі зображення, що є небажаним для піксельної щільної регресії. Ще гірше, використання дуже глибоких мереж може легко призвести до проблем під час навчання, таких як зникнення градієнта.

Використання пропускних з'єднань у дуже глибокій мережі може вирішити обидві вищезазначені проблеми.

По-перше, розробляємо експерименти, щоб показати, що використання пропускних з'єднань є корисним для збереження деталей зображення. Зокрема, дві мережі навчаються для видалення шуму із зображення з рівнем шуму σ = 70.

- У першій мережі використовуємо 5 шарів згортки 3 × 3 з кроком 3. Розмір вхідних даних для навчання становить 243×243, що після 5 шарів згортки дає вектор, який кодує дуже високий рівень абстракції зображення. Потім для відновлення вхідних даних з вектора ознак використовується деконволюція. Результати показано на рисунку 3.7. Можемо спостерігати, що деконволюції складно відновити деталі лише з вектора, що кодує абстракцію вхідних даних. Це явище означає, що просте використання глибоких мереж для відновлення зображень може не привести до задовільних результатів;

- Друга мережа використовує ті самі налаштування, що й перша, але з додаванням пропускних з'єднань. Результати показано на рисунку 3.7. У порівнянні з першою мережею, мережа з пропускними з'єднаннями може відтворити вхідні дані та досягає набагато кращих значень PSNR. Це легко зрозуміти, оскільки карти ознак з великою кількістю деталей на нижніх шарах безпосередньо передаються до верхніх шарів [10].

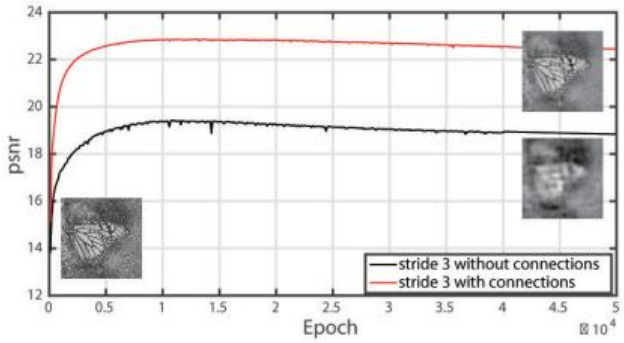

Рисунок 3.7 – Відновлення деталей зображення за допомогою деконволюції та пропускних зв’язків.

Пропускні зв’язки є ефективними для відновлення деталей зображення.

По-друге, навчаємо та порівнюємо п’ять різних мереж, щоб продемонструвати: використання пропускних з’єднань сприяє зворотній передачі градієнта під час навчання, що дозволяє краще відтворити наскрізне відображення, як показано на рисунку 3.8. Ці п’ять мереж – це: мережі з 10, 20 та 30 шарами без пропускних з’єднань, а також мережі з 20 та 30 шарами з пропускними з’єднаннями. Як можна побачити, втрата під час навчання збільшується, коли мережа стає глибшою без пропусків.

На наборі валідації глибші мережі без пропусків досягають нижчого PSNR, і ми навіть спостерігаємо перенавчання для 30-шарової мережі. Ці результати можуть бути пов’язані з проблемою зникнення градієнта. Однак отримуємо менші помилки навчання на навчальному наборі та вищий PSNR і кращу здатність до узагальнення на тестовому наборі при використанні пропускних з'єднань.

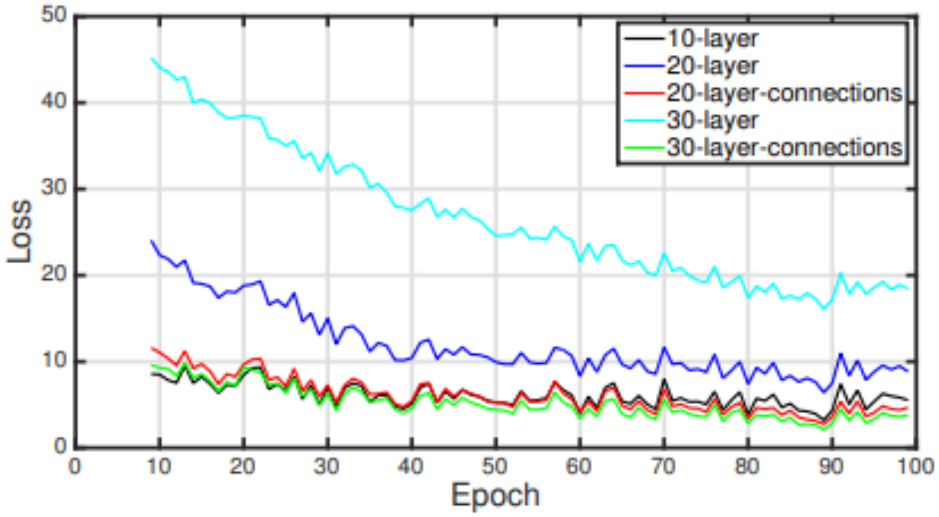

Рисунок 3.8 – Зниження точності на навчальному наборі даних під час навчання.

3.16 Ефективність тестування

Щоб застосовувати моделі глибокого навчання на пристроях з обмеженою обчислювальною потужністю, таких як мобільні телефони, необхідно прискорити етап тестування. Для цієї мережі пропонуємо використовувати зменшення розміру в конволюційних шарах, щоб зменшити розмір карт ознак. Щоб отримати вихідні дані того ж розміру, що й вхідні, використовується деконволюція для збільшення розміру карт ознак у симетричних деконволюційних шарах. Таким чином, ефективність тестування можна значно покращити з майже незначним зниженням продуктивності [10].

Зокрема, використовуємо крок = 2 у конволюційних шарах для зменшення розміру карт ознак. Зменшення розміру в різних конволюційних шарах тестується на очищенні зображень від шуму, як показано на рисунку 3.9.

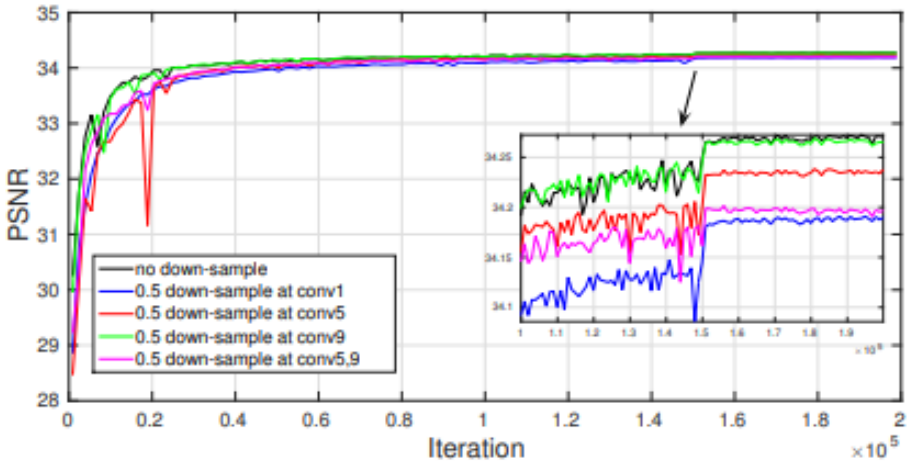

Рисунок 3.9 – Показники PSNR на валідаційному наборі для різних стратегій зменшення роздільної здатності. Позначення «down-sample at conv-i» означає, що зменшення роздільної здатності застосовується в i-му конволюційному шарі, а збільшення роздільної здатності – у його симетричному деконволюційному шарі.

Тестуємо зображення розміром 160×240 на процесорі i7-2600; час тестування для «без зменшення», «зменшення на conv1», «зменшення на conv5», «зменшення на conv9», «зменшення на conv5,9» становить 3,17 с, 0,84 с, 1,43 с, 2,00 с та 1,17 с відповідно. Головне спостереження полягає в тому, що тестові значення PSNR можуть дещо погіршуватися відповідно до зменшення масштабу карти ознак у всій мережі.

Зменшення розміру в першому конволюційному шарі зменшує розмір карт ознак до 1/4, що призводить до майже 4-кратного прискорення тестування, але PSNR погіршується менше ніж на 0,1 порівняно з мережею без зменшення розміру. Зменшення розміру в «conv9» скорочує час тестування на 1/3, але продуктивність майже така ж, як і без зменшення розміру. Як результат, «раніше» зменшення розміру може призвести до дещо гіршої продуктивності, але забезпечує значно вищу ефективність тестування. Це має бути компромісом у різних ситуаціях застосування [10].

3.17 Порівняльний аналіз моделей ШІ

Таблиця 3.1 – Порівняльний аналіз моделей ШІ

Критерій порівняння	**Генеративно-змагальні мережі (GAN)**	**Автоенкодери (Autoencoders)**
Основний принцип роботи	Складаються з двох нейронних мереж — генератора та дискримінатора, які навчаються у змагальному режимі	Складаються з енкодера та декодера, які навчаються стискати та відновлювати зображення
Основна задача	Генерація та відновлення реалістичних високоякісних зображень	Реконструкція вхідного зображення з мінімальною похибкою
Архітектура	Generator + Discriminator	Encoder + Decoder
Принцип навчання	Змагальне навчання між двома мережами	Мінімізація різниці між вхідним та відновленим зображенням
Тип навчання	Adversarial learning (змагальне)	Reconstruction learning (реконструктивне)

Продовження таблиці 3.1

Робота з деталями	Відновлюють дрібні текстури, контури та високочастотні компоненти	Відновлюють основну структуру та важливі ознаки зображення
Якість результату	Дуже висока візуальна реалістичність	Більш згладжений та стабільний результат
Робота із шумом	Добре прибирають шуми та артефакти, але можуть створювати штучні текстури	Ефективно усувають шум за рахунок виділення важливих ознак
Відновлення втрачених ділянок	Можуть генерувати правдоподібні відсутні області	Відновлюють структуру на основі латентного представлення
Основний механізм покращення	Генератор навчається “обманювати” дискримінатор	Мережа навчається компактному представленню даних
Використання латентного простору	Може використовуватись, але не є основним	Є ключовим компонентом моделі
Вузьке місце (bottleneck)	Зазвичай відсутнє	Є основною частиною архітектури
Функції втрат	MSE + perceptual loss + adversarial loss	Найчастіше MSE або reconstruction loss
Робота з текстурами	Дуже добре відновлюють текстури	Текстури можуть бути менш деталізованими
Робота з суперроздільністю	Один із найефективніших підходів	Можуть використовуватись, але поступаються GAN
Реалістичність результату	Висока, орієнтована на людське сприйняття	Вища точність реконструкції, але менша фотореалістичність
Стійкість навчання	Навчання нестабільне та складне	Навчання значно стабільніше
Складність навчання	Висока	Середня
Вимоги до ресурсів	Дуже високі	Нижчі, ніж у GAN
Потреба у великих датасетах	Висока	Висока, але менша ніж у GAN
Швидкість навчання	Повільне та ресурсо-затратне	Швидше та простіше
Ризик появи артефактів	Високий при нестабільному навчанні	Нижчий
Основна перевага	Висока реалістичність і деталізація	Стабільне відновлення структури та усунення шуму
Основний недолік	Складність навчання та нестабільність	Менша деталізація та менш реалістичні результати
Типові задачі	Super-resolution, image restoration, image generation, inpainting	Denoising, compression, reconstruction

Продовження таблиці 3.1

Відновлення старих фотографій	Добре підходять для складної реставрації та генерації деталей	Добре підходять для усунення шуму та реконструкції
Здатність до генерації нових даних	Так	Обмежена
Контроль над результатом	Менш передбачуваний	Більш стабільний і контрольований
Складність реалізації	Висока	Нижча
Використання у сучасних AI-системах	Дуже популярні у суперрезолюції та генерації зображень	Широко використовуються для реконструкції та видалення шумів

Основні відмінності GAN та автоенкодерів

Генеративно-змагальні мережі (GAN) GAN орієнтовані насамперед на:

- створення фотореалістичних результатів;
- генерацію нових деталей;
- покращення візуального сприйняття зображення.

Основні переваги GAN:

- відновлення текстури;
- генерування правдоподібних дрібних деталей;
- створення природного вигляду зображення.

Основні недоліки GAN:

- складні у навчанні;
- потребують великої кількості даних;
- можуть створювати артефакти при нестабільному тренуванні.

Автоенкодери більше орієнтовані на:

- реконструкцію структури зображення;
- усунення шуму;
- стискання інформації;
- стабільне відновлення даних.

Вони простіші у реалізації, стабільніше навчаються, та потребують менше ресурсів.

Основні недоліки автоенкодерів:

- результати можуть бути більш розмитими;

- деталізація нижча, ніж у GAN.

GAN забезпечують високий рівень деталізації та фотореалістичності завдяки змагальному навчанню між генератором і дискримінатором. Водночас автоенкодери орієнтовані на стабільне відновлення структури зображення та ефективне усунення шуму за рахунок використання латентного представлення даних. GAN видають кращі результати у задачах суперроздільності та генерації текстур, однак потребують складнішого навчання та більших обчислювальних ресурсів. Автоенкодери є простішими та стабільнішими моделями, але зазвичай поступаються GAN за рівнем візуальної реалістичності результатів [10].

4 РЕАЛІЗАЦІЯ МОДЕЛІ ВІДНОВЛЕННЯ ЗОБРАЖЕНЬ ЗА ДОПОМОГОЮ АВТОЕНКОДЕРІВ

Розглянемо приклад завдання обробки зображень – відтворенні цифр із набору даних MNIST. Реалізуємо модель на Python і пояснимо, що відбувається на кожному етапі, візуалізуючи процес, щоб краще зрозуміти, як працюють автоенкодери [6].

Імпортуємо необхідні бібліотеки для обробки числових даних, візуалізації, а також побудови та навчання нейронної мережі.

```
import numpy as np
import matplotlib.pyplot as plt
from tensorflow.keras.datasets import mnist
from tensorflow.keras.models import Model
from tensorflow.keras.layers import Input, Dense
from tensorflow.keras.optimizers import Adam
```

Використовуємо набір даних MNIST як вхідні дані, шари Dense для побудови автоенкодера та оптимізатор Adam для навчання ваг моделі. Далі розділимо набір даних на навчальну та тестову сукупності, не беручи до уваги мітки цифр, оскільки автоенкодер навчається відтворювати вхідні дані, а не класифікувати зображення. Це приклад навчання без наставника.

```
(x_train, _), (x_test, _) = mnist.load_data()
```

Виберемо з набору даних одну цифру та відобразимо її.

```
sample = x_test[4]
plt.imshow(sample.reshape(28, 28), cmap="gray")
plt.title("Original - 4")
plt.axis("on")
plt.show()
```

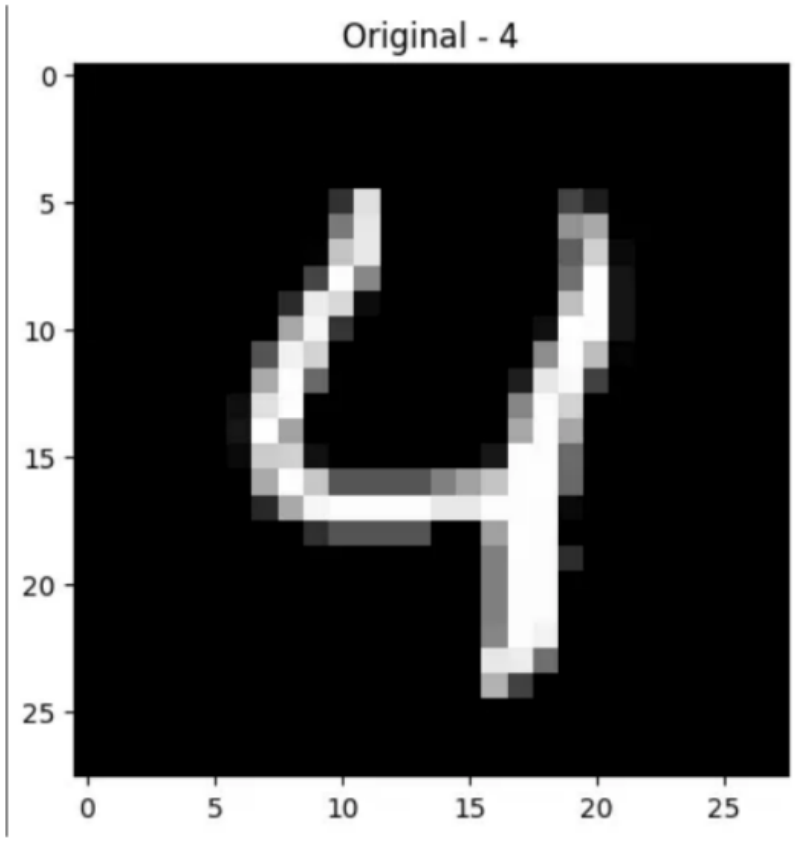

Рисунок 4.1 – Приклад цифри з набору даних MNIST.

Далі перетворимо дані на тип з плаваючою комою та нормалізуємо їх до діапазону від 0 до 1. Нормалізація полегшує та стабілізує процес навчання нейронної мережі, оскільки модель оперує даними з єдиним масштабом.

```
x_train = x_train.astype("float32") / 255.0
x_test = x_test.astype("float32") / 255.0
```

Наступним кроком є перетворення кожного зображення у вектор довжиною 784.

Це необхідно, оскільки шари Dense вимагають вхідних даних у вигляді одновимірних векторів ознак – кожен піксель стає окремою вхідною ознакою, яку потім шари Dense обробляють сукупно [6].

```
x_train = x_train.reshape((len(x_train), 28 * 28))
x_test = x_test.reshape((len(x_test), 28 * 28))
```

Параметри моделі та визначення. На цьому етапі визначаємо архітектуру моделі, а також її основні параметри.

```
input_dim = 784
latent_dim = 32
```

input_dim – визначає розмір вхідних даних (784 пікселі зображення MNIST)

latent_di – визначає кількість вимірів у латентному просторі, що відображає ступінь стиснення даних.

```
input_img = Input(shape=(input_dim,))
```

У функції `input_img` визначаємо вхідні дані для нашого автоенкодера – модель очікує одновимірний вектор із 784 значеннями.

Енкодер:

```
# encoder
encoded = Dense(latent_dim, activation="relu")(input_img)
```

Далі, у блоці `encoded`, створюємо повністю з'єднаний (Dense) шар із 32 нейронами (latent_dim). Також додаємо активаційну функцію ReLU для забезпечення нелінійності.

Декодер:

```
# decoder
decoded = Dense(input_dim, activation="sigmoid")(encoded)
```

Під час декодування виконуємо операцію, зворотну до дії кодера: беремо 32 значення з латентного простору та намагаємося перетворити їх назад у вектор довжиною 784. Активаційна функція сигмоїда обмежує вихідні значення діапазоном від 0 до 1, тобто отримані числа завжди будуть знаходитися в цьому проміжку.

```
# autoencoder
autoencoder = Model(input_img, decoded)
```

Тепер створюємо повну модель автоенкодера, яка пов’язує вхідні дані (input_img) з вихідними (decoded). Модель знає форму даних, які вона отримує, і

може генерувати (відтворювати) вихідні дані. Далі створюємо окрему модель кодера, яка перетворює вхідне зображення на вектор у латентному просторі.

Це дозволяє аналізувати та візуалізувати цифри в цьому просторі. Можна сказати, що він представляє собою своєрідну «суть» кожної цифри, оскільки містить її найважливіші ознаки [6].

encoder = Model(input_img, encoded)

Компіляція та навчання: На цьому етапі компілюємо модель, визначаючи, як вона буде навчатися. Встановлюємо оптимізатор Adam, який відповідає за оновлення ваг моделі, та використовуємо функцію втрат MSE (середньоквадратична похибка), яка вимірює різницю між вихідним зображенням та його реконструкцією піксель за пікселем.

```
# Model compilation
autoencoder.compile(
    optimizer=Adam(),
    loss="mse"
)
```

Розглянемо архітектуру нашої моделі.

```
autoencoder.summary()
```

Layer (type)	Output Shape	Param #
input_layer_1 (InputLayer)	(None, 784)	0
dense_2 (Dense)	(None, 32)	25,120
dense_4 (Dense)	(None, 784)	25,872

Total params: 152,978 (597.57 KB)
Trainable params: 50,992 (199.19 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 101,986 (398.39 KB)

Рисунок 4.2 – Отримані дані

Метод model.summary() надає короткий огляд архітектури автоенкодера, показуючи структуру шарів, розміри їхніх вихідних даних та кількість параметрів, що підлягають навчанню. Це є гарною практикою для візуалізації архітектури моделі.

Як бачимо вище, наш автоенкодер складається з вхідного шару (input_layer_1), шару Dense з 32 нейронами (Encoder) та іншого шару Dense з 784 нейронами (Decoder). Гаразд, давайте тепер навчимо нашу модель. Встановимо 50 епох з партіями по 256 зразків, випадково перемішаємо дані та використаємо тестовий набір для валідації.

```
# Training
history = autoencoder.fit(
    x_train, x_train,
    epochs=50,
    batch_size=256,
    shuffle=True,
    validation_data=(x_test, x_test)
)
```

Візуалізуємо процес навчання на наших даних, побудувавши графік значень втрат як для навчального, так і для валідаційного наборів.

```
loss = history.history["loss"]
val_loss = history.history["val_loss"]
epochs = range(1, len(loss) + 1)
plt.figure(figsize=(8, 5))
plt.plot(epochs, loss, label="Training loss")
plt.plot(epochs, val_loss, label="Validation loss")
plt.xlabel("Epoch")
plt.ylabel("MSE - reconstruction error")
plt.title("Autoencoder training")
plt.legend()
plt.grid(True)
plt.show()
```

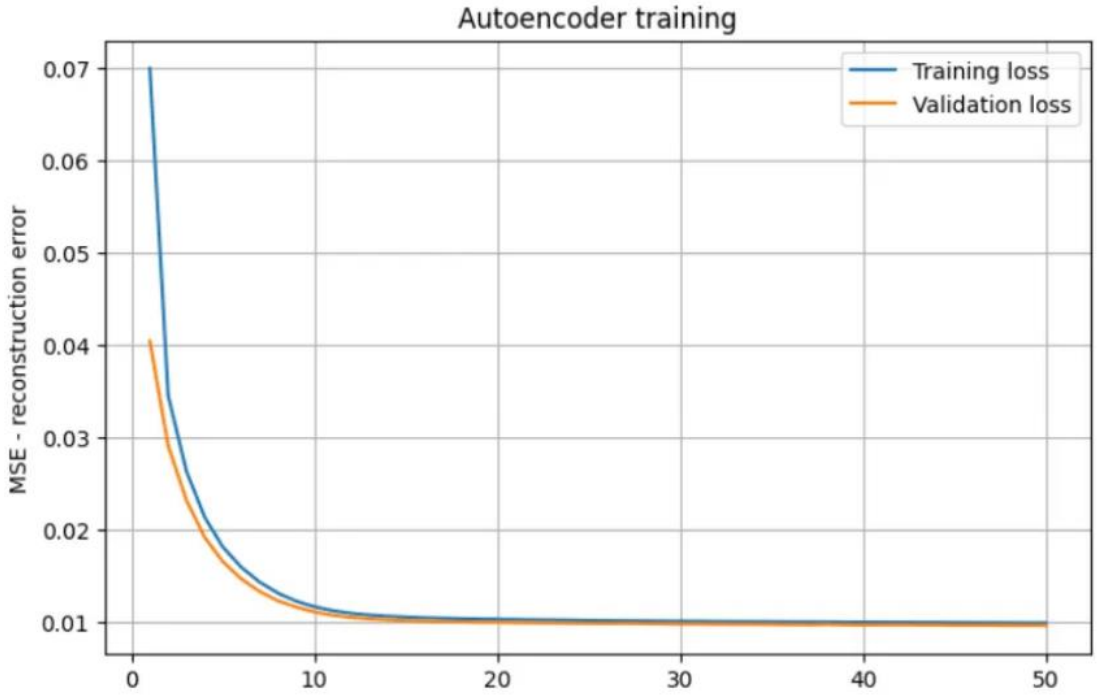

Рисунок 4.3 – Графік залежності

Приблизно через 15 епох наша модель, судячи з усього, досягла стабільних значень втрат навчання та валідації. Тепер маємо навчену модель. Також варто перевірити векторне представлення зразка цифри в латентному просторі – тобто, як саме вона була стиснута [6].

```
latent = encoder.predict(sample.reshape(1, 784))
print("Latent vector:")
print(latent)
```

Представлення однієї цифри у вигляді 32-вимірного вектора:

```
Latent vector:
[ 8.305198   8.910123   5.715289   3.341804   5.220013   4.50231
 12.017851   6.608532   9.079575   9.943344   5.074297   8.533254
  6.790465   4.628155   4.903169   4.27465    3.62276    5.67882
  7.227486   4.593061   3.241135   8.609306   6.993225   8.348385
  9.994142   3.190234   9.053399   3.476817   11.713034   3.839947
 16.681644   6.02702  ]
```

Перевіримо, як модель справляється з відтворенням цифр. Давайте порівняємо вихідні зображення з їхніми відтвореними версіями. Нижче наведено чотири зразки цифр разом із їхніми відтвореними версіями, згенерованими автоенкодером, що дозволяє візуально оцінити якість відтворення.

```
samples = x_test[:4]
reconstructions = autoencoder.predict(samples)
plt.figure(figsize=(10, 4))

for i in range(4):
    # Oryginal
    plt.subplot(2, 4, i + 1)
    plt.imshow(samples[i].reshape(28, 28), cmap="gray")
    plt.title("Oryginal")
    plt.axis("off")
    # Reconstructed
    plt.subplot(2, 4, i + 5)
    plt.imshow(reconstructions[i].reshape(28, 28), cmap="gray")
    plt.title("Reconstruction")
    plt.axis("off")
plt.tight_layout()
```

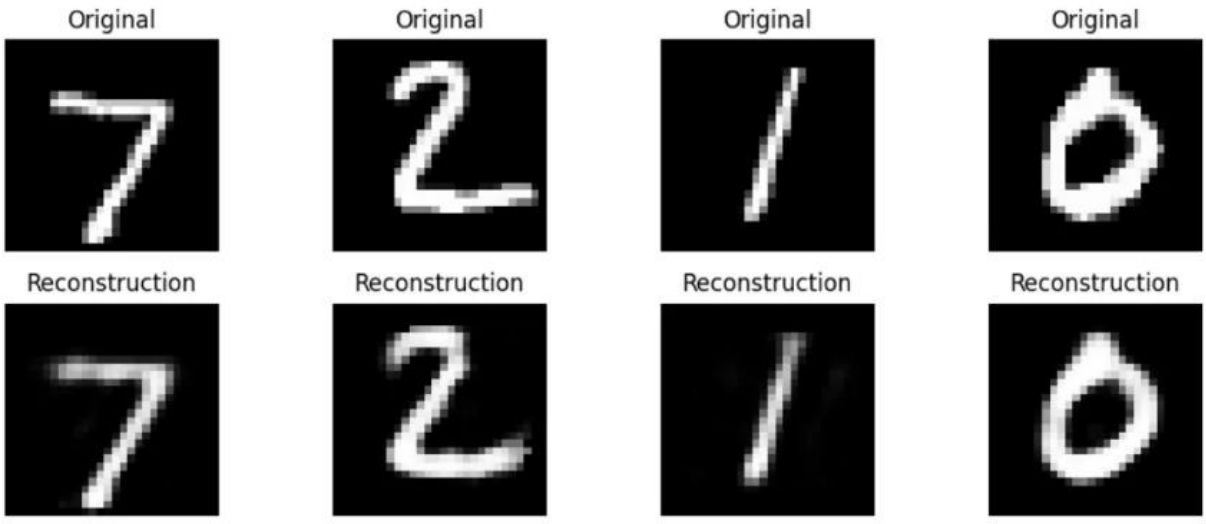

Рисунок 4.4 – Порівняння вихідних даних та їх реконструкції

Реконструкції виглядають досить добре, незважаючи на втрату деяких деталей. Модель правильно відтворила загальні форми та положення цифр. Оскільки наші вхідні дані складаються із зображень, доцільно використовувати конволюційні нейронні мережі (ConvNets) як кодери та декодери. На практиці автоенкодери, що застосовуються до зображень, майже завжди є конволюційними автоенкодерами.

Тобто, автоенкодери є потужним інструментом для навчання стислих представлень даних і становлять основу багатьох сучасних методів обробки зображень. За допомогою латентного простору модель навчається фіксувати найважливіші особливості даних, ігноруючи при цьому непотрібний шум. Хоча у практичних застосуваннях для обробки зображень переважають конволюційні автоенкодери, навіть проста архітектура на основі щільних шарів дає чітке розуміння того, як вони працюють, і ролі прихованого представлення [6].

ВИСНОВКИ

Досліджено сучасні методи відновлення якості цифрових зображень, а також проаналізовано можливості використання штучного інтелекту для покращення фотографій та усунення різних типів дефектів. Розглянуто як класичні методи обробки зображень, зокрема фільтрацію, гістограмні перетворення та адаптивне підсилення контрасту, так і сучасні підходи на основі нейронних мереж.

Особливу увагу було приділено моделям глибокого навчання, таким як генеративно-змагальні мережі (GAN), автоенкодери, варіаційні автоенкодери та трансформери. Було визначено їхні основні переваги, недоліки, складність навчання та ефективність у різних задачах відновлення зображень. Проведений аналіз показав, що сучасні моделі ШІ здатні значно покращувати якість зображень, відновлювати дрібні деталі, усувати шуми, підвищувати різкість та навіть відновлювати пошкоджені частини фотографій.

Розглянуто існуючі програми для відновлення якості фотографій, серед яких Remini, Adobe Photoshop, Fotor, LetsEnhance.io та Pixlr. Кожен із підходів має свої особливості та найкраще підходить для певних типів задач. Простіші онлайн-сервіси забезпечують швидкий результат і зручність використання, тоді як професійні інструменти дають значно більший контроль над процесом обробки.

Досліджено принцип роботи автоенкодерів та їх застосування для відновлення якості зображень. Автоенкодери ефективно працюють із задачами усунення шумів, відновлення структури зображення та покращення його якості. Водночас моделі на основі GAN демонструють кращі результати у відновленні

дрібних деталей та створенні більш реалістичних текстур, хоча потребують значно більших обчислювальних ресурсів і складнішого навчання.

Дослідження підтвердило, що використання штучного інтелекту є одним із самих перспективних напрямів у сфері відновлення та покращення цифрових зображень.

ПЕРЕЛІК ПОСИЛАНЬ

1. Юр’єва К. Г. Математичне та програмне забезпечення системи видалення артефактів з фотографій / НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО». Київ, 2024. 100 с.

2. Штучний інтелект простими словами — вивчаємо та навчаємо. *Claris Verbis*. 25.03.2025. URL: https://www.clarisverbis.com.ua/blogpost/shtuchnyj-intelekt-prostymy-slovamy-vyvchayemo-ta-navchayemo/.

3. Штучний інтелект у фотографії: Як технології змінюють процес зйомки та редагування. *AI 360. Tech-економіка*. 13.02.2025. URL: https://ai360.com.ua/shtuchnyy-intelekt-u-fotohrafii-yak-tekhnolohii-zminiuiut-protses-zyomky-ta-redahuvannia/.

4. Гордун М. В. Методи покращення якості (2Д) зображень генеративно-змагальними мережами / НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО». Київ, 2023. 111 с.

5. Nelson D. What is an Autoencoder?. *Unite.AI*. 20.09.2020. URL: https://www.unite.ai/what-is-an-autoencoder/.

6. Malicki K. Computer Vision: Autoencoders for Image Reconstruction. *Medium*. *How neural networks learn to compress and reconstruct images*. 06.01.2026. URL: https://medium.com/%40malickiart/computer-vision-autoencoders-for-image-reconstruction-55671c478ff9.

7. Raheem Remtulla, Adam Samet, Adam Samet та ін. A Future Picture: A Review of Current Generative Adversarial Neural Networks in Vitreoretinal

Pathologies and Their Future Potentials / Department of Ophthalmology & Visual Sciences, McGill University, Montreal, QC H4A 3SE, Canada. Montreal, 2025. 29 c.

8. Zhou P. Research Developments in Generative Adversarial Networks for Image Restoration and Communication. *Chengdu College of University of Electronic Science and Technology of China, Western High-Tech District, Chengdu 611731, China*. 2025. 23 січ. С. 9.

9. Reemr. Al Maawali, Amnah. Al-Shidi. Optimization Algorithms in Generative AI for EnhancedGAN Stability and Performance. *Applied Computing Journal*. Oman : Faculty of Computing and Information technology, Sohar University, 2025. С. 13.

10. Xiao-Jiao Mao, Chunhua Shen, Yu-Bin Yang. Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections. 2016. 17 с.

11. Brownlee J. How to Identify and Diagnose GAN Failure Modes. *Machine Learning Mastery*. *Generative Adversarial Networks*. 21.01.2021. URL: https://machinelearningmastery.com/practical-guide-to-gan-failure-modes/?utm_source