

Міністерство освіти і науки України  
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії  
(повна назва факультету)

Кафедра комп'ютерних наук  
(повна назва кафедри)

## КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

бакалавр

(назва освітнього ступеня)

на тему: Використання генеративної моделі stable diffusion для  
для аугментації даних і поліпшення якості сегментації зображень

Виконав: студент IV курсу, групи СН-41  
спеціальності 122 Комп'ютерні науки

(шифр і назва спеціальності)

(підпис)

Ляпандра М.С.

(прізвище та ініціали)

Керівник

(підпис)

Карнаухов А.К.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Шимчук Г.В.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Боднарчук І.О.

(прізвище та ініціали)

Рецензент

(підпис)

Тиш Є.В.

(прізвище та ініціали)

Тернопіль - 2026

Міністерство освіти і науки України  
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії  
(повна назва факультету)

Кафедра комп'ютерних наук  
(повна назва кафедри)

ЗАТВЕРДЖУЮ  
Завідувач кафедри  
Боднарчук І.О.  
(підпис) (прізвище та ініціали)  
«\_\_» \_\_\_\_\_ 2026 р.

## ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня Бакалавр  
(назва освітнього ступеня)

за спеціальністю 122 Комп'ютерні науки  
(шифр і назва спеціальності)

Студенту Ляпандрі Максиму Степановичу  
(прізвище, ім'я, по батькові)

1. Тема роботи Використання генеративної моделі Stable Diffusion для  
аугментації даних і поліпшення якості сегментації зображень

Керівник роботи Карнаухов Андрій Костянтинович  
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «14» 05 2026 року № 4/9-239

2. Термін подання студентом завершеної роботи 21.06.2026 р.

3. Вихідні дані до роботи наукові літературні джерела

4. Зміст роботи (перелік питань, які потрібно розробити)

Вступ

1. Огляд предметної області.

2. Розробка структури експериментів, підготовка моделі сегментації та набору даних

3. Проведення експериментів для оцінки ефективності генеративної аугментації даних

4. Безпека життєдіяльності, основи охорони праці

Висновки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1. Титулка. 2. Актуальність роботи, мета, задачі дослідження. 3. Поняття сегментації.

4. Аугментація зображень методами глибокого навчання. 5. Основні компоненти Stable Diffusion. 6. Схема експериментів, котрі проводилися. 7. Опис стеку ключових технологій.

8. Опис алгоритму генерації зображень. 9. Приклади зображень із змінним фоном та пташкою. 10. Значення метрик ефективності моделей на тестових даних.

11. Залежність метрики IoU від кількості аугментованих даних для обмеженої тренувальної вибірки. 12. Висновки. Основні результати проведеного дослідження.



## АНОТАЦІЯ

Використання генеративної моделі *stable diffusion* для аугментації даних і поліпшення якості сегментації зображень // Ляпандра Максим Степанович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем та програмної інженерії, кафедра комп'ютерних наук, група СН-41 // Тернопіль, 2026 // С. – 61, рис. – 39, табл. – 1, слайдів – 12, бібліогр. – 42.

**Ключові слова:** аугментація даних, дифузійна модель, згорткові нейронні мережі, комп'ютерний зір, семантична сегментація

У першому розділі розглянуті основні методи сегментації зображень, а також різні підходи до аугментації зображень. Висвітлена дифузійна модель генерації зображень *Stable Diffusion*, показаний принцип її роботи та основні компоненти. описані її переваги в порівнянні з іншими моделями.

У другому розділі наведена структура експериментів, описана підготовка моделі для завдання семантичної сегментації. Для вирішення задачі семантичної сегментації обрано повнозгорткову нейронну мережу *U-Net*. Оцінка ефективності моделей здійснюється за допомогою метрик *IoU* та *Accurasy*. Розглянуто та реалізовано спосіб розширення обмеженої тренувальної вибірки із використанням функції «*Inpaint*», яка дозволяє виділяти маскою окремі частини зображення та змінювати їх з урахуванням текстового запиту. За допомогою даного методу обмежена тренувальна вибірка була збільшена зі 100 зображень до 1100.

У третьому розділі проведені експерименти з навчанням моделі *U-Net* на різних тренувальних вибірках та проведено порівняння результатів, а також надано рекомендації використанню *Stable Diffusion* та змальовано можливості покращення експериментів.

У четвертому розділі описані важливі питання безпеки життєдіяльності та основ охорони праці.

## ANNOTATION

Using the Stable Diffusion Generative Model for Data Augmentation and Improving Image Segmentation Quality // Liapandra Maksym // Ternopil Ivan Pul'uj National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science // Ternopil, 2026 // P. - 61, Fig. - 39, Table - 1, Slide - 12, References - 42.

**Keywords:** data augmentation, diffusion model, convolutional neural networks, computer vision, semantic segmentation

The first section discusses the main methods of image segmentation, as well as various approaches to image augmentation. The diffusion model of image generation Stable Diffusion is highlighted, the principle of its operation and main components are shown. Its advantages compared to other models are described.

The second section presents the structure of the experiments, describes the preparation of the model for the semantic segmentation task. The fully convolutional neural network U-Net was chosen to solve the semantic segmentation task. The effectiveness of the models is assessed using the IoU and Accuracy metrics. A method for expanding the limited training sample using the “Inpaint” function is considered and implemented, which allows you to select individual parts of the image with a mask and change them taking into account the text query. Using this method, the limited training sample was increased from 100 images to 1100.

In the third section, experiments were conducted with training the U-Net model on different training samples and a comparison of the results was made, as well as recommendations for using Stable Diffusion and possibilities for improving the experiments were outlined.

In chapter 4, important issues of life safety and the basics of labor protection are described.

## **ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ СКОРОЧЕНЬ І ТЕРМІНІВ**

GAN (Generative adversarial networks) –генеративно-змагальні мережі.

Stable Diffusion – модель глибокого навчання для перетворення тексту в зображення.

Аугментація даних – процес штучного збільшення розміру та різноманітності вашого набору даних (датасету) шляхом модифікації наявних або генерації нових синтетичних даних. Вона є ключовим інструментом для покращення якості роботи моделей штучного інтелекту та машинного навчання.

## ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1 ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ .....	10
1.1 Підходи до розв'язання задач сегментації зображень .....	10
1.1.1 Поняття сегментації.....	10
1.1.2 Традиційні методи сегментації зображень.....	11
1.1.3 Методи сегментації зображень, що ґрунтуються на глибокому навчанні.....	12
1.2 Огляд методів аугментації зображень .....	16
1.2.1 Поняття аугментації.....	16
1.2.2 Класичні методи аугментації зображень .....	17
1.2.3 Аугментація зображень методами глибокого навчання .....	17
1.2.4 Фреймворки для аугментації даних .....	20
1.3 Опис, переваги та компоненти Stable Diffusion.....	21
РОЗДІЛ 2. РОЗРОБКА СТРУКТУРИ ЕКСПЕРИМЕНТІВ, ПІДГОТОВКА МОДЕЛІ СЕГМЕНТАЦІЇ ТА НАБОРУ ДАНИХ .....	23
2.1 Структура експериментів, що проводяться.....	23
2.2 Підготовка моделі семантичної сегментації .....	24
2.3 Встановлення Stable Diffusion Web UI та доступ до моделі.....	27
2.4 Опис та підготовка набору даних.....	27
2.4.1 Опис набору даних.....	27
2.4.2 Підготовка набору даних.....	29
2.5 Опис стеку ключових технологій, що використовуються.....	29
2.6 Опис та реалізація алгоритму генерації зображень.....	30
РОЗДІЛ 3. ПРОВЕДЕННЯ ЕКСПЕРИМЕНТІВ ДЛЯ ОЦІНКИ ЕФЕКТИВНОСТІ ГЕНЕРАТИВНОЇ АУГМЕНТАЦІЇ ДАНИХ .....	34
3.1 Навчання моделі на повному наборі даних.....	34
3.2 Навчання моделі на обмеженому наборі даних.....	35
3.3 Навчання моделі на обмеженому наборі даних, об'єднаному із синтетичними даними.....	36

3.4 Порівняння результатів .....	41
3.5 Рекомендації щодо використання Stable Diffusion та можливості покращення експериментів .....	47
РОЗДІЛ 4. БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ОХОРОНИ ПРАЦІ .....	49
4.1 Класифікація шкідливих та небезпечних виробничих факторів.....	49
4.2 Вплив вібрації на людину .....	51
ВИСНОВКИ.....	55
ПЕРЕЛІК ДЖЕРЕЛ .....	57

## ВСТУП

**Актуальність теми.** Зараз машинне навчання перебуває на етапі жвавого розвитку і широко застосовується в різних сферах - від персоналізованої реклами та рекомендаційних систем до обробки зображень у медицині, автоматизації промислового виробництва та управління безпілотними апаратами. Частина цих складних завдань вирішується за допомогою комп'ютерного зору - однієї з областей машинного навчання, яка займається розробкою алгоритмів і моделей для обробки та аналізу зображень.

Незважаючи на те, що за останнє десятиліття глибокі нейронні мережі створили переворот у комп'ютерному зорі і показали високі результати в різних галузях, важливо розуміти, що для досягнення належного рівня ефективності та хорошої узагальнюючої здатності такі мережі вимагають значних обсягів попередньо розмічених даних, так і тривалого часу навчання. Однак кількість розмічених даних для реальних завдань часто обмежена. У таких випадках використовують аугментацію (доповнення) даних, тобто різні методики створення додаткових даних шляхом модифікації вихідних даних з допомогою різних трансформацій. Стандартні підходи до аугментації даних містять досить прості трансформації, наприклад такі як повороти, відображення, зрушення масштабування і т.п., і не можуть створювати додаткові дані, які істотно відрізнялися б від вихідних у плані семантики, текстур і геометричних особливостей.

Для розширення набору даних останніми роками дедалі більшого розповсюдження набуває генеративна аугментація зображень, заснована на використанні GAN та дифузійних моделей. У запропонованому дослідженні розглядається використання дифузійних моделей для аугментації даних та покращення якості сегментації зображень на прикладі моделі дифузії Stable Diffusion.

Однак слід зазначити, що навіть використання підходів із застосуванням нейронних мереж не є вичерпним у вирішенні практичних завдань, а лише

наближає результати до бажаного рівня точності та ефективності. Тому вдосконалення результатів роботи наявних методів та пошук нових рішень у галузі обробки та аналізу зображень є актуальним завданням.

**Мета роботи** – дослідження ефективності генеративної аугментації зображень, виконаної за допомогою дифузійної моделі Stable Diffusion на прикладі задачі семантичної сегментації.

**Для досягнення мети виділено ряд завдань:**

- вивчити сучасні моделі сегментації зображень;
- проаналізувати існуючі методи аугментації зображень;
- розробити структуру експериментів з метою оцінки ефективності генеративної аугментації даних з допомогою дифузійної моделі;
- реалізувати алгоритм генерації зображень за допомогою Stable Diffusion;
- провести експерименти та порівняти результати роботи моделі сегментації на даних з генеративною аугментацією та без.

**Практичне значення одержаних результатів.** Результати проведеного дослідження можуть бути успішно використані для медичної діагностики, автономних транспортних засобів, супутникового зондування, робототехніки та ін.

# РОЗДІЛ 1. ОГЛЯД ПРЕМЕТНОЇ ОБЛАСТІ

## 1.1 Підходи до розв'язання задач сегментації зображень

### 1.1.1 Поняття сегментації

Сегментація зображень є важливим підрозділом комп'ютерного зору, що використовується для виділення об'єктів та меж на зображеннях. Найбільш поширене її застосування у таких сферах як медична діагностика, виділення об'єктів на супутникових знімках, автоматичне розпізнавання образів, керування дорожнім рухом, робототехніка та ін.

Сегментація зображення - це метод розбиття цифрового зображення різні групи, звані сегментами, з метою спрощення його аналізу та обробки. При цьому кожному пікселю на зображенні надається певна мітка, що показує, до якого сегменту він відноситься. Сегментація зображень є більш точним і деталізованим підходом порівняно з класифікацією зображень, оскільки, крім визначення класу об'єкта, також виконує його локалізацію на зображенні. Сегментація розмежовує зображення на складові його області чи об'єкти, при чому ступінь деталізації, до якої власне і зводиться таке розмежування, чітко залежить від задачі.

Зважаючи на мету дослідження та завдання сегментація може бути розділена на семантичну сегментацію і сегментацію екземпляра. Перша призначена для класифікації пікселів зображення за відповідними семантичними класами. У другій сегментації відбувається класифікація пікселів на основі екземплярів, а не класів. На рис. 1.1 наведені види сегментації.

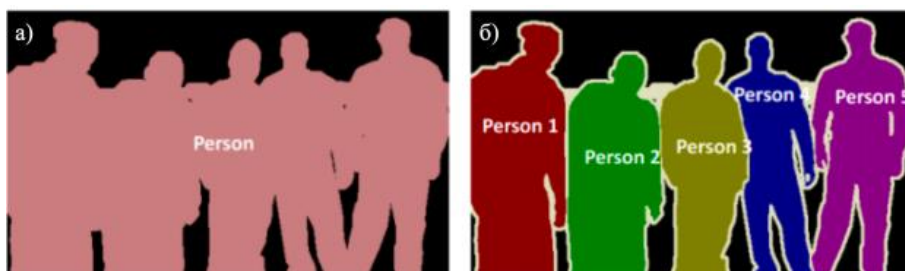


Рисунок 1.1 – Семантична сегментація (а) та сегментація екземплярів об'єктів (б)

Розглянуте у цій роботі завдання є завданням семантичної сегментації - необхідно визначити приналежність пікселів зображень або до призначеного класу, або до фону.

### **1.1.2 Традиційні методи сегментації зображень**

Сегментація зображень почала свій розвиток із примітивних алгоритмів цифрової обробки зображень у поєднанні із алгоритмами оптимізації, зокрема вирощування областей і алгоритм змії [1]. Ці методи були орієнтовані на локальне подання зображення, фокусуючись на відмінностях та градієнтах у пікселях. Згодом з'явилися алгоритми, що використовують глобальне подання вхідного зображення, такі як адаптивна порогова обробка, алгоритм Оцу та алгоритми кластеризації, котрі дозволили отримувати точніші результати [1].

Розглянемо докладніше деякі з методів.

Сегментація зображень методом порогу - це один із найбільш простих методів, за якого визначається певна величина порогу для поділу пікселів на два класи. Властиво коли значення пікселя є більшим, ніж порогове, він належить до одного класу, інакше - до іншого. Порогове значення є особливо ефективним, коли різниця між значеннями пікселів двох класів дуже велика і поріг можна вибрати як середнє значення між двома класами.

Алгоритми сегментації по регіонах засновані на пошуку подібності між сусідніми пікселями та їх групування до загального класу: спочатку деякі пікселі вибираються як вихідні, а потім алгоритм виявляє межі вихідних пікселів і класифікує їх як схожі або несхожі. Далі алгоритм повторюється для безпосередніх сусідів до тих пір, поки все зображення буде сегментовано.

Сегментація країв - це ще один традиційний метод сегментації зображень, який є процесом виявлення меж об'єктів на зображенні. Разом із тим відбувається класифікація кожного пікселя зображення на крайовий і не крайовий, і виділення крайових пікселів в окремий клас.

Сегментація на основі кластеризації використовує алгоритми кластеризації для групування пікселів на зображенні із загальними атрибутами в один сегмент.

Алгоритми кластеризації, як то k-середніх, можуть забезпечувати швидке і якісне виділення сегментів без потреби завдання початкової точки сегментації.

### 1.1.3 Методи сегментації зображень, що ґрунтуються на глибокому навчанні

Класичні повнозв'язні нейронні мережі не зовсім підходять для обробки зображень через розмір самих зображень і кількість пікселів. Щоб обробити таку кількість інформації, потрібно багато нейронів та зв'язків, що робить мережу громіздкою та складнонавченою. Крім того, класичні нейронні мережі не враховують двовимірну структуру зображень і не можуть коректно запам'ятати просторовий зв'язок між об'єктами та їх складовими частинами [2].

Однак для обробки зображень добре підходять згорткові нейронні мережі, що з'явилися в 1988 році. Вони мають двовимірну структуру, яка аналогічна до структури зображень, і аналізують зображення не повністю, а частинами. Це досягається завдяки використанню спеціальних згорткових шарів, які зменшують розмір вхідного сигналу від шару до шару. Операція згортки містить переміщення ядра згортки (матриці-фільтра) по вхідній матриці, при цьому проводиться матричне множення ділянок вхідної матриці на ядро згортки, результат якого зберігається в нову матрицю, яка також називається картою ознак, рис. 1.1 [2].

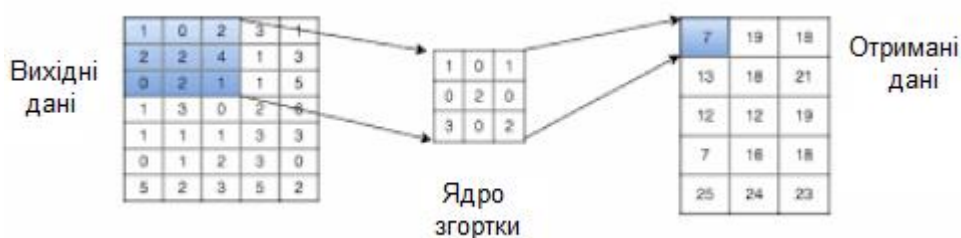


Рисунок 1.2 – Процес згортки

Метод ковзного вікна дозволяє вирішувати задачу сегментації, використовуючи згорткові нейронні мережі. При використанні даного методу ми витягуємо фрагменти з вихідного зображення і передаємо їх до класифікуючої

згорткової нейронної мережі, що дозволяє прогнозувати клас центрального пікселя кожного фрагмента. Однак такий підхід вимагає великих обчислювальних витрат на етапах навчання та тестування, тому що на кожне зображення необхідно передати як мінімум  $N$  фрагментів, де  $N$  - це кількість пікселів у зображенні. Для класифікації пікселів можна використовувати такі попередньо навчені мережі як AlexNet, VGG19, InceptionV3 та ін. Вихідний шар мережі при цьому видає одну з міток, що відповідають класам, на які сегментуємо наше зображення. Однак використання класичної згорткової мережі для класифікації фрагментів зображення потребує значних обчислювальних ресурсів, так як кожен фрагмент має бути оброблений окремо. Це неефективно ще й тому, що суміжні фрагменти зображення можуть перекриватися і повторно оброблятися, що створює додаткове навантаження. Для вирішення цієї проблеми можна скористатися повнозгортковими мережами [3].

Повнозгорткова мережа (FCN - Fully Convolutional Network) є нейронною мережею, що складається з послідовності згорткових шарів, без будь-яких повнозв'язних шарів. Повнозгорткова мережа для семантичної сегментації була представлена в 2015 році командою University of California, Berkeley та описана у статті [4].

Згортки налаштовуються таким чином, щоб розміри зображення (ширина та висота) залишалися незмінними упродовж проходження через мережу. На відміну від підходу на основі ковзного вікна, де кожен витягнутий фрагмент класифікується незалежно, FCN прогнозує категорії всіх пікселів одночасно. Вихідний шар мережі складається з  $C$  карт ознак, де  $C$  - це кількість категорій, включаючи фон, якими класифікується кожен піксель [3]. При цьому на виході мережі для цільової ознаки вийде  $C$  сегментованих зображень, що відповідають  $C$  класам.

У повнозгортковій мережі всі згорткові шари зберігають розміри вихідного зображення. Це означає, що при обробці зображень, котрі мають високою роздільну здатність, виникне потреба у великому обсязі обчислювальних ресурсів, особливо при великій кількості карт ознак або каналів у кожному згортковому шарі (архітектура повнозгорткової мережі наведена на

рис. 1.3). З метою вирішення такої проблеми використовується інший підхід повнозгорткових мереж, а саме повнозгорткову мережу з понижувальною та підвищуючою дискретизацією.

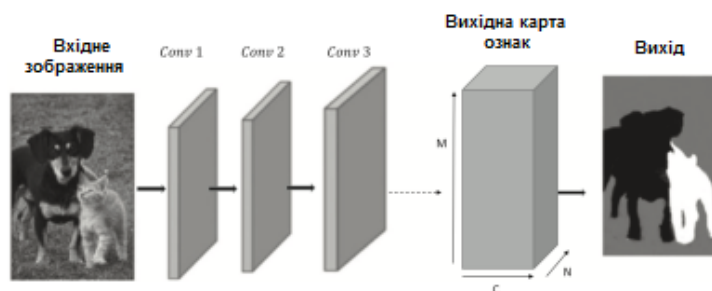


Рисунок 1.3 – Архітектура повнозгорткової нейронної мережі

Повнозгорткова мережа з понижувальною та підвищуючою дискретизацією використовує комбінацію згортки, в якій зображення піддається знижувальній дискретизації (виконуються крокові згортки та/або операції пулінгу) у першій половині мережі, а потім – підвищуючій дискретизації (unpooling, upsampling) в останніх шарах для відновлення просторових розмірів вихідного зображення. Архітектура такої нейронної мережі показана на рис. 1.4.

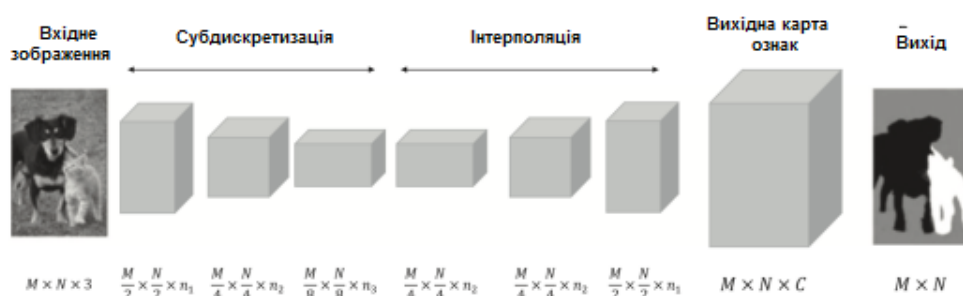


Рисунок 1.4 – Повнозгорткова мережа з понижувальною та підвищуючою дискретизацією

Важливою особливістю таких мереж є використання транспонованої згортки при виконанні операції інтерполяції (upsampling) на етапі підвищення дискретизації. Застосування такої транспонованої згортки забезпечує навчання мережі [3].

Поліпшенням та розвитком FCN стала архітектура U-Net. Вона була створена у 2015 році і представлена у [5]. Топологія мережі має U- подібну форму у напрямку від вхідних шарів до вихідних, звідси і назва – U-Net. Для розв’язання питання втрати інформації, котре існує у шарах зниження дискретизації, застосовуються пропускні з’єднання - skip connections, котрі йдуть від кодера напряду до декодера, обходячи вузьке місце, таким чином зменшують втрату даних. На рис. 1.5 продемонстровано архітектуру U-Net.

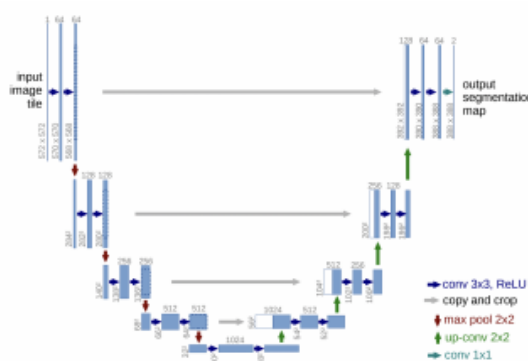


Рисунок 1.5 – Архітектура U-Net

У ту ж саму пору як і U-Net з'явилася інша сегментуюча нейронна мережа – SegNet [6], що також є архітектурою повнозгорткової нейронної мережі для семантичної сегментації зображень за пікселями. Архітектура SegNet проілюстрована на рис. 1.6.

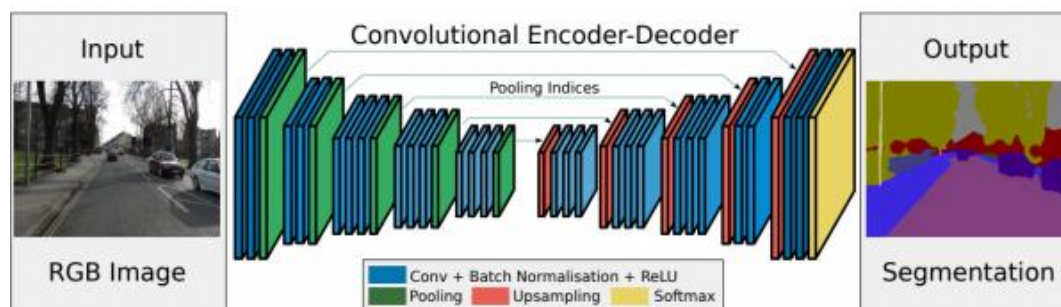


Рисунок 1.6 – Архітектура SegNet

DeerLab – це ще одна архітектура нейронної мережі для вирішення задач у галузі семантичної сегментації. На відміну від попередніх архітектур, таких як

U-Net, DeepLab використовує складніші операції згортки, які дозволяють уникнути втрат інформації при зменшенні розміру зображення. Крім того, в архітектурі реалізовано багатомасштабний підхід до отримання ознак за допомогою семантичного шару (Atrous Spatial Pyramid Pooling), що дозволяє виявляти об'єкти будь-якого розміру на зображенні [1].

## **1.2 Огляд методів аугментації зображень**

### **1.2.1 Поняття аугментації**

Для створення добре працюючих глибоких нейронних мереж необхідно мати достатню кількість різноманітних даних, щоб уникнути перенавчання та їхньої прихильності до конкретних навчальних прикладів. Однак підготовка та розмітка якісних даних найчастіше займає багато часу і вимагає вкладення значних коштів. Крім того, деякі класи об'єктів можуть бути недостатньо представлені в навчальному наборі, що може бути предвісником незбалансованості даних. Для подолання цих труднощів добре використовувати методи аугментації даних.

Аугментація даних є процесом штучного збільшення обсягу тренувальних даних шляхом створення змінених копій вихідних даних. Мета аугментації - збільшити різноманітність даних, щоб модель могла навчатися більш різноманітних зразках і краще узагальнювати знання нових даних.

Методи аугментації зображень можна розділити на кілька видів залежно від типу змін, які вони вносять до набору даних. Існують методи геометричних перетворень (як то афінні перетворення), які змінюють форму та положення об'єктів на зображенні. Також існують методи зміни пікселів, які вносять зміни на рівні окремих пікселів (наприклад, зміна яскравості, контрастності та насиченості). Крім того, для створення додаткових даних використовуються методи створення штучних даних за допомогою GAN та дифузійних нейромереж, які створюють штучні дані на основі навчального набору.

Зазначимо, що перетворення, що застосовуються до навчальної та тестової вибірок, часто можуть відрізнятися. Крім того, залежно від частоти застосування

перетворень, методи можна розділити на постійні та непостійні. Постійні методи застосовуються до всіх зображень у навчальному наборі, а непостійні застосовуються до зображень з деякою ймовірністю або випадковим чином із заданої множини перетворень.

### **1.2.2 Класичні методи аугментації зображень**

Такі методи, звані іноді також базовими методами аугментації, зводяться до геометричних перетворень та фотометричного зсуву зображень. При геометричному перетворенні виконується перевертання, обертання, обрізання, кадрування, перенесення картинки. У фотометричному перетворенні використовуються такі примітивні методи зміни кольору, як зміна колірного простору, вживання різних фільтрів та додавання шуму [7, 9].

### **1.2.3 Аугментація зображень методами глибокого навчання**

За останнє десятиліття глибоке навчання досягло значних результатів і показало свою ефективність у комп'ютерному зорі, класифікації зображень, детектуванні об'єктів, сегментації зображень тощо. У сфері аугментації даних глибоке навчання також знайшло застосування.

У дослідженні, що вийшло в 2021 році, присвяченому огляду останніх тенденцій у глибокому навчанні для аугментації цифрових зображень, автори виділили три основні категорії методів глибокого навчання, що застосовуються для доповнення даних: GAN, нейронне перенесення стилю (Neural Style Transfer) і навчання (Meta metric learning) [7]. При цьому третя категорія складається з трьох моделей – нейронна аугментація (Neural Augmentation), автоматична аугментація (Auto Augmentation) та інтелектуальна аугментація (Smart Augmentation).

Генеративні моделі є предметом зростаючого інтересу та швидкого розвитку. Можна виділити чотири основні методи генерації зображень, заснованих на глибокому навчанні: варіаційні автоенкодері (VAE), потокові моделі (Flow-based models), GAN та дифузійні моделі (Diffusion models). На рис. 1.7 продемонстровано механізми чотирьох вказаних вище методів [11].

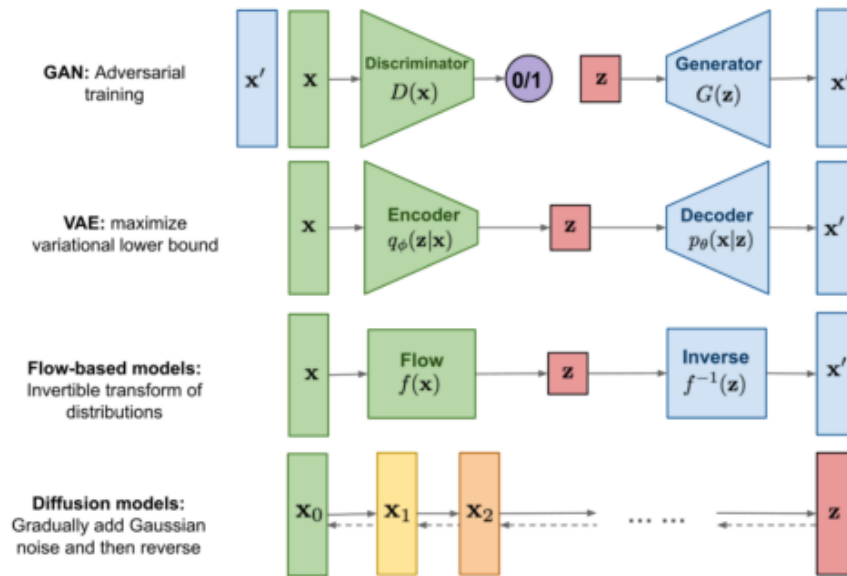


Рисунок 1.7 – Механізми генеративних методів

Розглянемо докладніше GAN, які є одним із найпопулярніших методів машинного навчання для створення синтетичних даних. Вони складаються з двох мереж - генератора та дискримінатора, де генератор приймає на вхід випадковий вектор і створює зразки, які мають виглядати як тренувальні приклади, а дискримінатор класифікує вхідні дані як реальні або згенеровані. Обидві мережі навчаються одночасно, завдяки чому в процесі конкуренції генератор навчається краще створювати реалістичні приклади, а дискримінатор - розрізняти реальні та згенеровані дані. На рис. 1.8 показана загальна схема GAN.

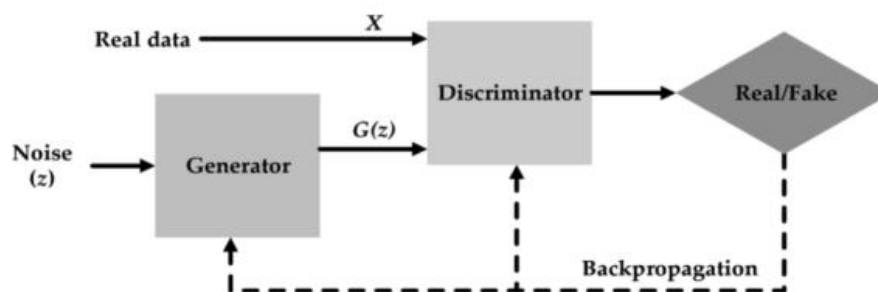


Рисунок 1.8 – Схема GAN

Генеративні моделі дають змогу будувати високоточні фотореалістичні зображення, наближені до реальних даних, оскільки вони навчалися на реальних даних. Крім того, вони дають можливість генерувати потенційно необмежену

кількість синтетичних даних. За останні роки в кількох роботах були спроби дослідити синтетичні дані, отримані за допомогою GAN [12, 13].

Іншим класом перспективних та ефективних генеративних моделей є дифузійні моделі, які вперше були представлені Й. Соул-Дікштейном та його колегами у 2015 році [15]. Також значне впливання на розвиток цього напрямку справила робота, випущена Й. Хо та ін. [16]. Генерування синтетичних даних за допомогою дифузійної моделі, формування з них додаткового набору даних для навчання нейронних мереж було розглянуто у статті Р. Хе та ін. [17]. Команда Б. Трабукко розробила гнучку стратегію аугментації даних, яка генерує варіації реальних зображень, використовуючи моделі дифузії text-to-image, і запропонувала фреймворк DA-Fusion [18].

Дифузійні моделі є класом імовірнісних генеративних моделей, які перетворюють шум на репрезентативну вибірку даних. Навчання складається з двох процесів:

1. Фіксований (або зумовлений) процес прямої дифузії  $q$ , який поступово додає гаусівський шум до зображення доти, доки не вийде чистий шум. На цьому етапі вихідне зображення ( $x_0$ ) повільно ітеративно спотворюється (ланцюжок Маркова) шляхом додавання (масштабованого гаусового) шуму. Цей процес виконується для деяких тимчасових кроків  $T$ , тобто.  $x_T$ . Наприкінці етапу через ітеративне додавання шуму залишається зашумлене зображення, що представляє «ізотропний гаусівський» шум. Таким чином, вихідний розподіл даних перетворюється на гауссовий розподіл.

2. Навчальний процес зворотної дифузії  $p_\theta$ , де нейронна мережа навчається поступово прибирати шум із зображення, починаючи з чистого шуму і до тих пір, поки не буде отримано реальне зображення. Тобто на цьому етапі завдання полягає в тому, щоб ітеративним способом (ланцюжок Маркова) видалити шум, доданий у прямому процесі. Це робиться з використанням моделі нейронної мережі, яка з урахуванням часового інтервалу  $t$  та зашумленого зображення  $x_t$  має спрогнозувати шум, доданий до зображення на кроці  $t-1$  [11].

Ілюстрація процесу прямої та зворотної дифузії показана на рис. 1.9.

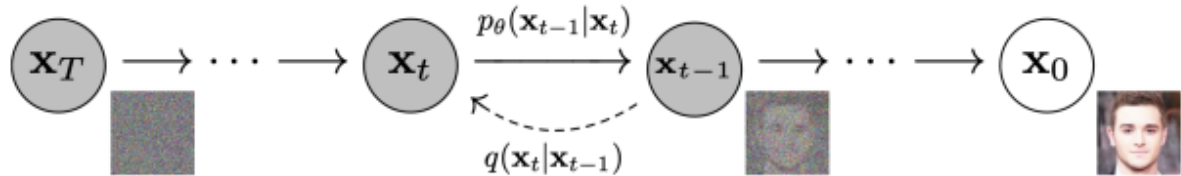


Рисунок 1.9 – Ілюстрація процесу прямої ( $q$ ) і зворотної ( $p_\theta$ ) дифузії

Загальну концепцію генерації зображень із використанням дифузійних моделей можливо описати наступним чином: створюється чистий Гаусівський шум і подається навченій моделі шумозаглушення, щоб отримати абсолютно нове зображення [19].

Існує кілька моделей генерації зображень з урахуванням дифузії. Найбільш відомими є Dall-E 2 від OpenAI, Imagen від Google, Stable Diffusion від StabilityAI, Midjourney. У цій роботі використовується модель Stable Diffusion для генерування синтетичних даних та формування з них додаткового датасету для навчання нейронної мережі.

#### 1.2.4 Фреймворки для аугментації даних

Існує достатньо велика кількість різних фреймворків для аугментації даних, які надають різні методи та функції для створення нових зразків даних, як це наведено в [20]. Перелічимо деякі з них.

- Albumentations - швидкий та гнучкий фреймворк для аугментації зображень та масок. Він підтримує багато різних методів аугментації, таких як зміна розміру, поворот, відображення, зміна кольору і т.д. [21];
- Imgaug - потужний і гнучкий фреймворк для аугментації зображень, який дозволяє створювати складні ланцюжки аугментації, що включають різні методи аугментації;
- Torchvision - фреймворк для роботи з даними в PyTorch, який включає різноманітні методи для аугментації зображень;
- OpenCV (Open Source Computer Vision) – одна з найвідоміших бібліотек з відкритим вихідним кодом для завдань комп'ютерного зору. Вона орієнтована на обробку зображень, захоплення відео та аналіз (включаючи такі

функції, як розпізнавання облич та об'єктів), а також допомагає забезпечити загальну інфраструктуру для додатків комп'ютерного зору;

- Augmentor - бібліотека аугментації зображень у Python, дозволяє застосовувати найбільш актуальні та реальні методи аугментації незалежно від використовуваної платформи або фреймворку для машинного навчання.

Окремо виділимо також бібліотеку Diffusers. Вона не є універсальними фреймворками для аугментації даних, однак за її допомогою можна генерувати синтетичні дані для розширення вихідних датасетів. Diffusers – це бібліотека для сучасних попередньо навчених дифузійних моделей для генерації зображень, аудіо та навіть тривимірних структур молекул.

### **1.3 Опис, переваги та компоненти Stable Diffusion**

Як уже зазначалося, зараз існують різні моделі генерації зображень з використанням дифузії, такі як Dall-E 2 від OpenAI, Imagen від Google, Stable Diffusion від StabilityAI і Midjourney.

Особливу цікавою для нас є модель Stable Diffusion, створена дослідниками та інженерами з CompVis, Stability AI та LAION. Вона навчена на зображеннях 512x512 з підмножини бази даних LAION-5B, найбільшої мультимодальної бази даних, що у вільному доступі [22]. Stable Diffusion має такі переваги в порівнянні з іншими моделями:

- відкритий вихідний код (Stable Diffusion є проектом з відкритим вихідним кодом, що дозволяє дослідникам та розробникам використовувати, покращувати та адаптувати його під свої потреби);

- генерація високоякісних зображень (Stable Diffusion може генерувати зображення високої чіткості з високою роздільністю та чіткими деталями);

- керована генерація зображень (Stable Diffusion дозволяє керувати процесом генерації, змінюючи параметри);

- ефективність (Stable Diffusion має високу швидкість навчання та генерації зображень, що робить його ефективним інструментом для створення великих датасетів);

– гнучкість (Stable Diffusion може використовуватися множиною різних способів - text2img, img2img та ін.).

З цих причин саме Stable Diffusion була обрана для створення синтетичних даних на основі текстового опису, які будуть використовуватися для створення додаткового датасету для навчання нейронної мережі.

Stable Diffusion – це ціла система, що складається з кількох компонентів та моделей, а не одна монолітна модель. Можна виділити три основні компоненти, у кожному з яких є своя нейронна мережа:

1. Кодувальник тексту - спеціальна мовна модель Transformer (текстовий кодувальник моделі ClipText): отримує на вхід текст і дає вихід 77 векторів ембеддингів токенів, кожен у 768 вимірах.

2. Створення інформації зображення за допомогою нейромережі U-Net та алгоритму планування (Scheduler). На цьому етапі відбувається покрокова обробка інформації, що призводить до створення високоякісного зображення. На вхід подаються ембедінг тексту і вихідний багатовимірний масив, що складається з шуму; на виході виходить масив обробленої інформації.

3. Декодер автокодувальника, який малює готове зображення за допомогою масиву обробленої інформації. На вхід приймає масив обробленої інформації розміром (4,64,64), але в виході видає кінцеве зображення розміром (3, 512, 512) [23].

На рис. 1.10 представлені три основні компоненти Stable Diffusion.



Рисунок 1.10 – Три основні компоненти Stable Diffusion

## РОЗДІЛ 2. РОЗРОБКА СТРУКТУРИ ЕКСПЕРИМЕНТІВ, ПІДГОТОВКА МОДЕЛІ СЕГМЕНТАЦІЇ ТА НАБОРУ ДАНИХ

### 2.1 Структура експериментів, що проводяться

У цій роботі проводиться дослідження ефективності генеративної аугментації зображень, виконаної з допомогою дифузійної моделі Stable Diffusion на прикладі задачі семантичної сегментації.

Вихідна ідея експерименту полягає в тому, що, маючи обмежену кількість даних, ми зможемо за допомогою дифузійної моделі згенерувати нові зображення на основі наявних. Тим самим ми зможемо збільшити різноманітність зображень у тренувальній вибірці, підвищити узагальнюючу здатність моделі, що навчається на цих даних, і отримати більш оптимальні результати роботи моделі на тестових даних.

Пропонується наступна схема експериментів, що проводяться:

1. Навчити на повній навчальній вибірці з набору даних нейронну мережу для сегментації зображень, яка використовуватиметься як базовий варіант (baseline) для подальшого порівняння ефективності навчання моделей.

2. Навчити на обмеженій навчальній вибірці з набору даних нейронну мережу для сегментації зображень з тими самими параметрами нейронної мережі, що й у базового варіанта.

3. За допомогою дифузійної моделі Stable Diffusion згенерувати кілька наборів даних із синтетичними зображеннями (з різною кількістю синтетичних зображень на одне вихідне зображення), які будуть використовуватися для розширення обмеженої навчальної вибірки набору даних.

4. Навчити на варіаціях обмеженої навчальної вибірки, об'єднаної з синтетичними даними, нейронну мережу для сегментації зображень з тими ж параметрами нейронної мережі, що й у базового варіанта і вибрати найбільш оптимальну кількість синтетичних даних, що генеруються одне вихідне зображення.

5. Порівняти результати роботи моделей сегментації, навчених на трьох

наборах даних: повному, обмеженому та обмеженому, об'єднаному із синтетичним.

Для аналізу якості роботи нейронних мереж під час навчання будуть аналізуватись такі метрики - IoU (Intersection over Union) та Accuracy.

На рис. 2.1. зображена схема експериментів, що проводяться.

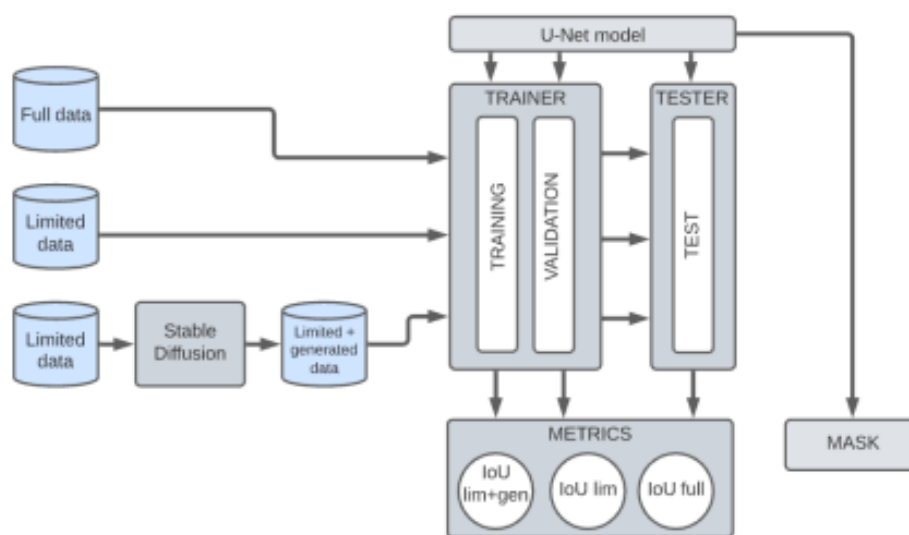


Рисунок 2.1 – Схема експериментів, що проводяться

## 2.2 Підготовка моделі семантичної сегментації

Для вирішення поставленого в роботі завдання було обрано повнозгорткову нейронну мережу для семантичної сегментації зображень U-Net. Як кодувальник використовувалася модель класифікації Resnet34 без останнього «щільного» (dense) шару.

Ініціалізація ваг виконана відповідно до методу, описаного у статті К. Хе у 2015 р., з використанням нормального розподілу [24], оскільки нейронні мережі з ініціалізацією ваг за даним методом показують більш високі результати, ніж з ініціалізацією зі випадковими вагами.

Як функція втрат для навчання та оцінки якості моделі використана функція втрат Тверського (Tversky loss) з базового класу функцій відношення перетину до об'єднання (Intersection over Union) [25]. Ця функція втрат часто

використовується у завданнях сегментації зображень, особливо під час роботи з незбалансованими даними. В її основі лежить індекс Тверської (Tversky index), який вимірює схожість між двома множинами, допомагає вирішити проблему дисбалансу даних і досягти більшого компромісу між точністю (precision) та повнотою (recall) при навчанні моделі [25]. Формула індексу Тверський (2.1):

$$TverskyIndex(A, B) = \frac{|A \cap B|}{|A \cap B| + \alpha |A \setminus B| + \beta |B \setminus A|}, \quad (2.1)$$

де  $A$  і  $B$  - дві порівнювані множини, а  $\alpha$  і  $\beta$  є параметрами, які дозволяють зміщувати акценти у бік точності або повноти і керують величиною штрафів за хибнопозитивні (false positive) і хибнонегативні (false negative) помилки.

Формула функції втрат Тверської (Tversky Loss) (2.2) для завдання семантичної сегментації зображень виглядає наступним чином:

$$TverskyLoss(pp, pt) = 1 - \frac{\Sigma(pp * pt)}{(\Sigma(pp * pt) + \alpha * \Sigma(pp * (1 - pt))) + \beta * \Sigma((1 - pp) * pt)}, \quad (2.2)$$

де:

- $pt$  - справжня бінарна маска,
- $pp$  - передбачена мережею маска,
- $\alpha$  і  $\beta$  - коефіцієнти, які дають змогу налаштувати ваги помилок різних

типів.

Таким чином функція втрат Тверської стимулює модель до мінімізації хибнопозитивних і хибнонегативних помилок.

Методом оптимізації обрано AdamW - метод стохастичної оптимізації, що є модифікованою реалізацією методу Adam, запропонованого в [26].

Для оцінки якості роботи нейронних мереж під час навчання оцінювалися наступні метрики – IoU (Intersection over Union) та Accuracy.

Метрика IoU є популярною метрикою для оцінки якості семантичної сегментації зображень. IoU вимірює ступінь перекриття між передбаченою та

істинною масками та обчислюється шляхом поділу перетину передбаченої та істинної масок на їх об'єднання. Чим вище значення IoU, тим більший збіг між пророкованою та істинною масками. Значення метрики лежить у діапазоні від 0 до 1, де 0 вказує на повну відсутність перекриття між масками, а 1 на повний збіг.

Формула для обчислення IoU (2.3) виглядає так:

$$IoU = \frac{TP}{TP+FP+FN}, \quad (2.3)$$

де:

- TP - кількість чітко передбачених пікселів (true positive);
- FP - кількість хибно-позитивних пікселів (false positive);
- FN - кількість помилково- негативних пікселів (false negative).

Інша метрика для оцінки якості моделі семантичної сегментації – це загальна точність (Accuracy). Вона вимірює частку правильно класифікованих пікселів по відношенню до загального числа пікселів у зображенні. Формула для обчислення Accuracy (2.4):

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}, \quad (2.4)$$

де TN - кількість чітко передбачених пікселів, які не належать до обраного класу (true negative). Решту позначень аналогічні до формули (2.3).

Однак варто врахувати, що Accuracy може бути недостатньо інформативною метрикою у тих випадках, коли класи незбалансовані або є велика кількість пікселів, що належать до фону. У таких випадках, навіть якщо модель передбачає всі пікселі як фон, точність може бути високою через домінування пікселів цього класу.

## **2.3 Встановлення Stable Diffusion Web UI та доступ до моделі**

У цій роботі використовується монтування Stable Diffusion від Automatic1111 - Stable Diffusion Web UI. Ця версія має браузерний інтерфейс, заснований на бібліотеці Gradio, доступний для розгортання на локальній машині з підтримкою GPU шляхом клонування репозиторію з GitHub [27]. WebUI створює віртуальне оточення Python, встановлює всі необхідні залежності та запускає локальний веб-сервер.

Stable Diffusion Web UI має простий та інтуїтивно зрозумілий інтерфейс, що дозволяє швидко та легко генерувати нові або змінювати наявні зображення. Крім того, вона часто оновлюється та розширюється, а доступ до неї можна отримати через API [28]. Для проведення експериментів та генерування нових даних на основі наявних ми будемо звертатися через API до веб-сервера, попередньо розгорнутого на локальній машині. Для роботи зі Stable Diffusion можна також альтернативно скористатися бібліотекою Diffusers [29], вона дозволяє створити та налаштувати пайплайн для генерації зображень.

Для Stable Diffusion існує велика кількість передвибраних моделей для генерації зображень з текстового опису, вільно доступних для скачування та додавання до Stable Diffusion Web UI – контрольні точки (checkpoint). Відповідну контрольну точку можна вибрати на <https://huggingface.co/models>. У цій роботі під час проведення експериментів використовуються моделі Stable-Diffusion-v1-5 [30] і Stable-Diffusion-Inpainting [31].

У Stable Diffusion існує два основних підходи до генерації зображень - "txt2img" та "img2img". Підхід "txt2img" дозволяє перетворювати текстовий запит (prompt) на картинку, тоді за допомогою "img2img" можна генерувати зображення на основі існуючих або змінювати їх.

## **2.4 Опис та підготовка набору даних**

### **2.4.1 Опис набору даних**

Як набір даних для проведення експериментів та оцінки якості роботи

моделей був обраний набір даних Caltech-UCSD Birds-200-2011 [32], доступний для завантаження у вільному доступі з посиланням на авторів [33].

Набір даних містить 11788 зображень у форматі RGB двохсот видів птахів. Кожен вид пов'язаний із статтею у Вікіпедії та організований відповідно до наукової класифікації – порядок, сімейство, рід, вид. Зображення були зібрані за допомогою пошуку зображень на фотохостингу Flickr, а потім відфільтровані шляхом показу кожного зображення кільком користувачам інтернет-майданчика Mechanical Turk. Кожне зображення анотовано із зазначенням рамки, що обмежує, розташування частин птиці (наприклад, дзьоб, хвіст тощо) і мітками ознак (атрибутів). Оригінальний датасет був зібраний в 2011 році для вирішення задач класифікації, розпізнавання категорій підпорядкованості, багатокласового виявлення об'єктів, розпізнавання на основі ознак і являє собою розширену версію набору даних Caltech-UCSD Birds 200. Маски для вирішення задачі семантичної сегментації були додані пізніше додатково. пікселі, що стосуються фону. Приклади маски та зображень з датасету наведено на рис. 2.2.

Ми будемо використовувати цей датасет для завдання сегментації зображень з птахами та відноситимемо пікселі в одному з двох класів: «птах» або «фон».



Рисунок 2.2 – Приклад зображень та масок з датасету

Окремо відзначимо, що зображення в цьому наборі даних перетинаються із зображеннями, що входять до ImageNet. Тому слід бути обережним при використанні мереж, попередньо навчених на ImageNet (або будь-яких мереж, попередньо навчених на зображеннях з набору Flickr).

## **2.4.2 Підготовка набору даних**

У ході експериментів ми досліджуємо можливість покращення роботи моделі за допомогою розширення обмеженого датасету. Оскільки додаткові дані генеруються дифузійною моделлю на основі текстового опису, ми можемо використовувати цю перевагу та штучно обмежити представленість класів птахів у тренувальній, валідаційній та тестовій вибірках. Це дозволить наблизити експеримент до завдання з реального життя, коли модель, навчена на обмежених за розмаїттям даних (у нашому випадку - на одних видах птахів), повинна також добре відпрацьовувати і у випадках, коли на вхід приходять різноманітні дані.

У датасеті представлено 200 видів птахів. До тренувальної вибірки увійшли зображення птахів 150 видів, обраних випадковим чином. Для обмеженої тренувальної вибірки також випадково були відібрані 100 зображень. Валідаційна і тестова вибірки представлені п'ятдесятьма видами птахів, розподілених рандомно, по 25 видів у кожній.

## **2.5 Опис стеку ключових технологій, що використовуються**

PyTorch - open source фреймворк машинного навчання, забезпечує гнучку та ефективну платформу для розроблення та навчання глибоких нейронних мереж. Серед його особливостей можна відзначити динамічний обчислювальний граф, широкий вибір функцій та алгоритмів, підтримку GPU та розподіленого навчання [35].

Segmentation Models PyTorch – бібліотека, заснована на фреймворку PyTorch та призначена для спрощення процесу створення, навчання та використання моделей глибокого навчання для сегментації об'єктів на зображеннях [36].

TensorBoard - інструмент візуалізації та налагодження, розроблений для фреймворку машинного навчання TensorFlow, але також сумісний з фреймворком PyTorch. TensorBoard пропонує зручний спосіб відстеження та аналізу процесу навчання нейронних мереж, візуалізації метрик, графів

обчислень, параметрів моделі, а також візуального відображення зображень, тексту та інших даних [37].

## 2.6 Опис та реалізація алгоритму генерації зображень

Перш ніж перейти до реалізації експериментів, необхідно згенерувати синтетичні дані за допомогою Stable Diffusion. Для створення синтетичних даних та розширення набору даних Caltech-UCSD Birds-200-2011 ми скористаємося підходом «img2img», при якому нейромережа ґрунтується на існуючій картинці та підмішує до неї інформацію з текстового опису.

Для навчання моделі сегментації зображень нам потрібні як самі зображення, і відповідні їм маски. У підході «img2img» є важлива для нас функція – «Inpaint» («вмальовування»), вона дозволяє виділяти маскою окремі частини зображення та змінювати їх з урахуванням текстового запиту. Для "Inpaint" генерації можна завантажити маску як окремий файл. Існує модель, спеціально натренована на роботу з «Inpaint» генераціями: Stable-Diffusion-Inpainting, її ми використовуємо для генерації двох типів нових зображень:

- зображення зі зміненим тлом;
- зображення зі зміненим птахом.

Створення нового зображення складається з наступних кроків:

1. Вибирається зображення та відповідна йому маска, далі задається кількість нових зображень, які необхідно згенерувати на основі одного вихідного.

2. Задається режим генерування: «Inpaint» дозволяє вибрати, яку частину зображення змінювати - яка закрита маскою або яка відкрита. Для генерування нового фону вибирається незакрита маскою частина зображення, а зміни птаці - закрита.

Складається текстовий опис (prompt) змін, які потрібно внести до зображення. Оскільки на одне вихідне зображення генерується кілька нових, то для більшого розмаїття даних тестовий опис вибирається випадковим чином із заздалегідь підготовленого списку описів. Список описів для зміни птахів

задається з назв видів птахів (класів), що увійшли до валідаційної та тестової вибірки. Описи для фону вказують на різні можливі місцезнаходження птахів – «птах на даху», «птах у лісі» тощо.

3. Забороняємо моделі генерувати на нових зображеннях певні категорії, поставивши їх у текстовому описі для `negative prompt`. Наприклад, при генеруванні нового фону в опис необхідно знести категорію птахів, тому що птахів на зображеннях повинні збігатися з розміченими масками і додаткові птахи на фоні нам не потрібні.

4. Підбираємо та задаємо параметри функції «Inpaint», які необхідні для досягнення бажаного результату синтезу нових зображень. Вкажемо деякі:

- «Denoising strength» - числове значення від 0 до 1, де 0 - відсутність змін, а 1 - повністю нова картинка, тобто чим вище це значення, тим сильніше підсумкова картинка буде відрізнятися від вихідної. Шляхом тестування різних варіантів було вирішено встановити цей параметр на значенні 0,75 для генерації птахів та 0,8 - для фону.

- "Sampling method" - метод дискретизації. Для генерації птахів використовувався метод "Euler a", а для фону - "LMS".

- "Seed" - стартова точка, з якої нейромережа потім формує зображення. Оскільки ми генеруємо кілька зображень з одного вихідного і потрібно досягти різних результатів при тому самому запиті, необхідно вказати «випадковий» параметр (`seed = -1`).

5. Через API передаємо певні параметри та зображення з маскою в модель, модель генерує зображення, і ми додаємо їх в наш датасет.

Під час проведення експериментів тестувалися дві моделі, попередньо завантажені в Stable Diffusion Web UI - Stable-Diffusion-v1-5 та Stable-Diffusion-Inpainting. Друга модель показала кращі результати, тому зображення експериментів генерувалися через неї.

На рис. 2.3. продемонстровано приклади штучно згенерованих зображень із зміненним тлом.

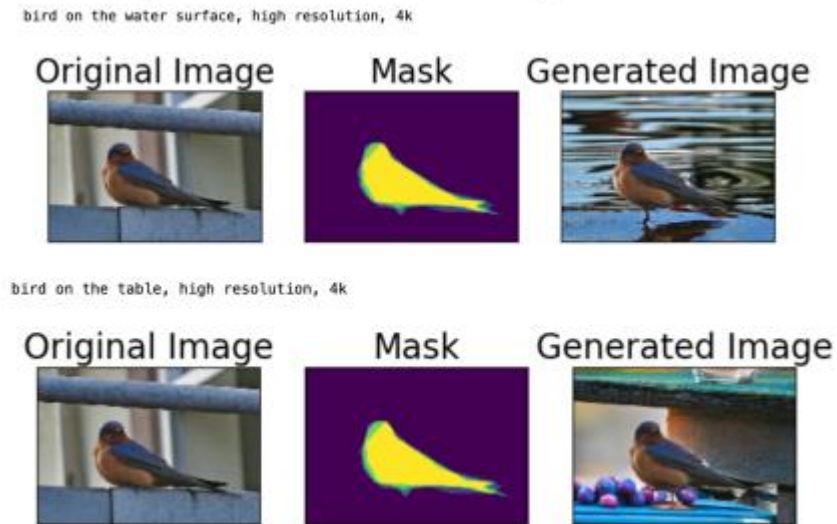


Рисунок 2.3 – Приклади зображень із зміненим тлом

На рис. 2.4. наведено приклади штучно згенерованих зображень із зміненим птахом.

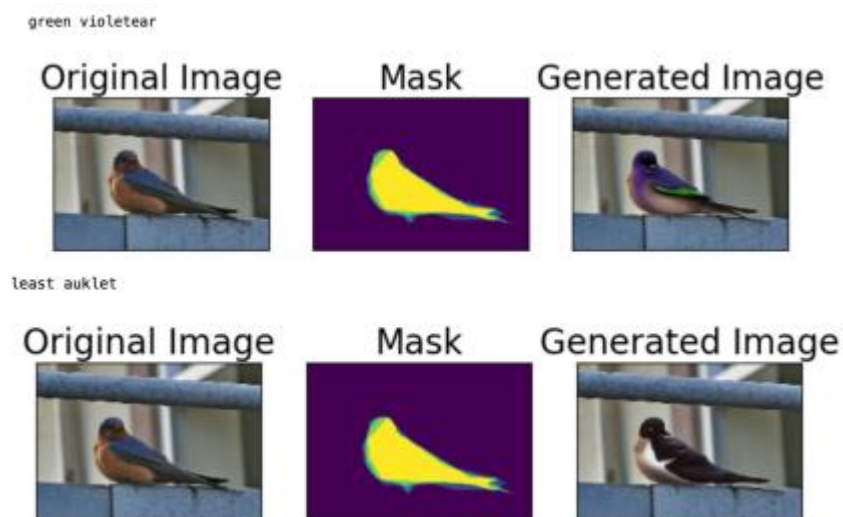


Рисунок 2.4 – Приклади зображень із зміненою пташкою

Параметри, що передаються модель Stable Diffusion, були обрані дослідним шляхом за рахунок багаторазового тестування різних комбінацій.

Зазначимо, що генерації не завжди виявлялися вдалими, тому при доборі параметрів доводилося керуватися не лише ступенем варіативності генерацій, а й стабільністю генерації зображень без мутацій або надмірного домальовування. На рис. 2.5 наведено приклад невдалої генерації.



Рисунок 2.5 – Приклади невдалої генерації зображення (модель домалювала птахові крило, якого немає на розмітці)

Для проведення експериментів було взято 100 зображень обмеженої тренувальної вибірки. Для кожного з вихідних зображень було згенеровано кілька варіантів зі зміненим фоном і зі зміненою пташкою. У результаті було отримано 5 нових наборів даних з різною кількістю нових синтетичних зображень, що припадають на вихідне:

1. 100 вихідних, 100 синтетичних зображень зі зміненим птахом, 100 синтетичних зображень зі зміненим фоном (разом у наборі 300 зображень).
2. 100 вихідних, 200 синтетичних зображень із зміненим птахом, 200 синтетичних зображень зі зміненим фоном (разом у наборі 500 зображень).
3. 100 вихідних, 400 синтетичних зображень із зміненим птахом, 400 синтетичних зображень із зміненим фоном (разом у наборі 900 зображень).
4. 100 вихідних, 500 синтетичних зображень із зміненим птахом, 500 синтетичних зображень зі зміненим фоном (разом у наборі 1100 зображень).
5. 100 вихідних, 700 синтетичних зображень із зміненим птахом, 700 синтетичних зображень зі зміненим фоном (разом у наборі 1500 зображень).

Також було сформовано два додаткові набори даних з метою дізнатися, чи має якийсь один із видів генерованих зображень (генерування нового фону та генерування нового птаха) більший вплив на показники навчання моделей:

1. 100 вихідних, 500 синтетичних зображень зі зміненим птахом (разом у наборі 600 зображень).
2. 100 вихідних, 500 синтетичних зображень зі зміненим фоном (в наборі 600 зображень).

## РОЗДІЛ 3. ПРОВЕДЕННЯ ЕКСПЕРИМЕНТІВ ДЛЯ ОЦІНКИ ЕФЕКТИВНОСТІ ГЕНЕРАТИВНОЇ АУГМЕНТАЦІЇ ДАНИХ

### 3.1 Навчання моделі на повному наборі даних

Перша частина експерименту передбачає навчання нейронної мережі для сегментації зображень на повній навчальній вибірці з набору даних (8990 зображень). На рис. 3.1 представлений графік навчання моделі. Модель навчалася упродовж 598 епох, значення функції втрат на навчальній та валідаційній вибірках зменшуються у міру навчання моделі. Виходячи з того, що значення функції втрат на тренувальній і валідаційній вибірках не мають значних розбіжностей і значення валідаційної наближається до тренувальної, можна зробити висновок, що налаштування ваг нейронної мережі виконано оптимально і наша модель має досить високу узагальнюючу здатність.

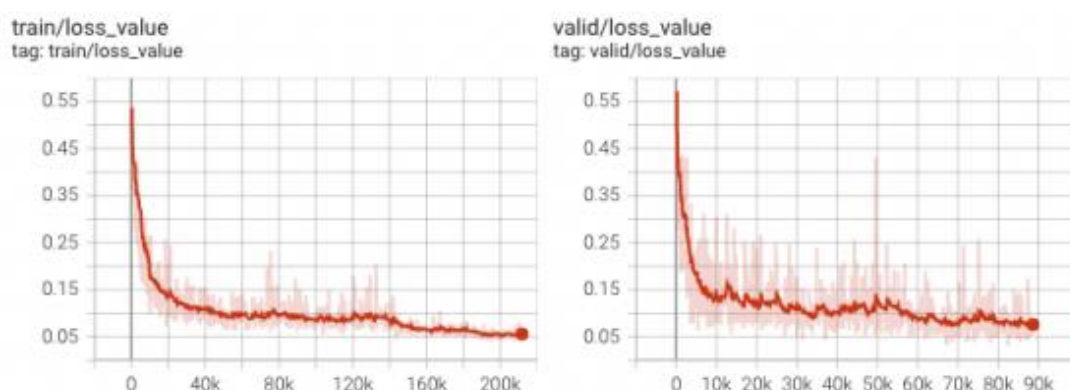


Рисунок 3.1 – Функція втрат

На рис. 3.2 показані значення метрики IoU на тренувальній та валідаційній вибірках у процесі навчання моделі. Модель не перенавчається та показує високі значення метрики на валідаційній вибірці.

Модель, навчену на повному наборі тренувальних даних, вважаємо базовим варіантом і саме з нею порівнюватимемо результати навчання наступних двох моделей нашого експерименту. Також зазначимо, що валідаційна вибірка не бере участі у навчанні моделі і тому може

використовуватись як показник якості роботи моделі.

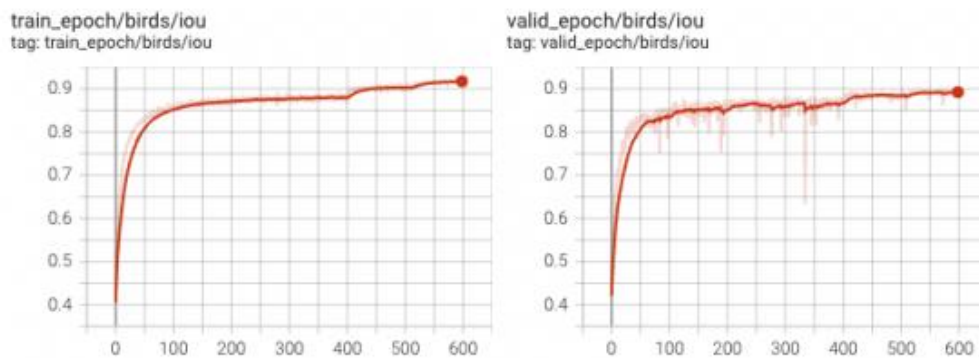


Рисунок 3.2 – Метрика IoU

### 3.2 Навчання моделі на обмеженому наборі даних

Другою частиною експерименту є навчання нейронної мережі для сегментації зображень з тими самими параметрами нейронної мережі, що й у базового варіанту, але вже на обмеженій навчальній вибірці набору даних (100 зображень). На рис. 3.3 та 3.4 показані графіки навчання моделі. Модель навчалася 505 епох. Значення функції втрат і метрики IoU на тренувальній та валідаційній вибірках тепер значно відрізняються, що говорить про перенавчання моделі на тренувальних даних та її досить слабку узагальнюючу здатність.

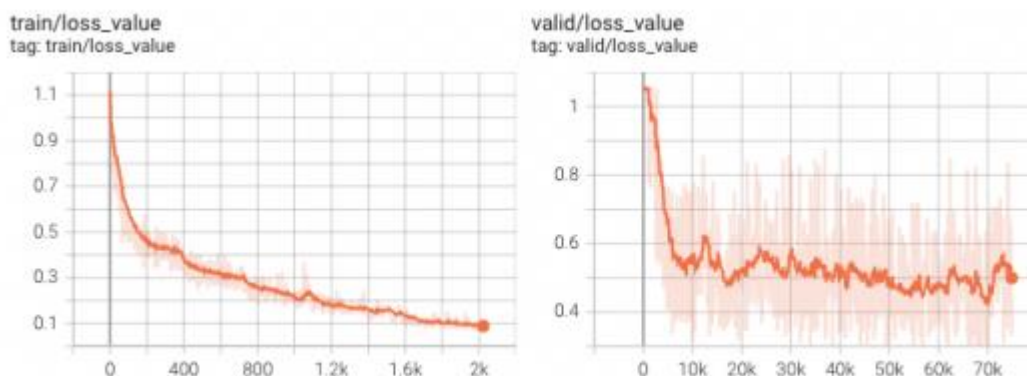


Рисунок 3.3 – Функція втрат

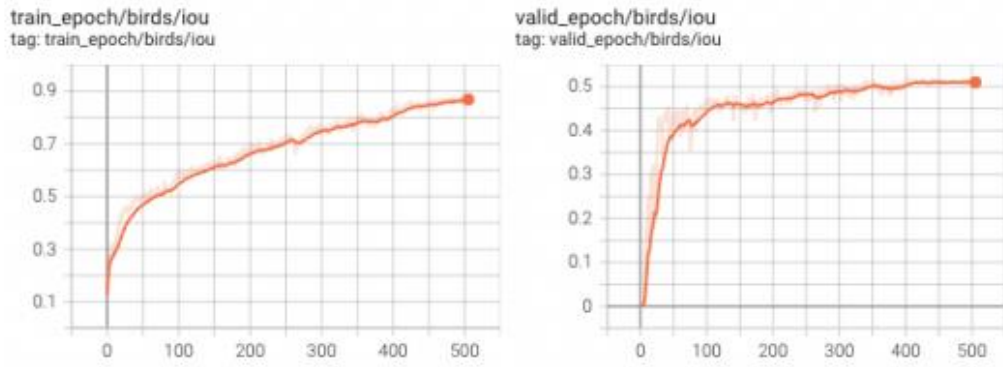


Рисунок 3.4 – Метрика IoU

### 3.3 Навчання моделі на обмеженому наборі даних, об'єднаному із синтетичними даними

У підрозділі 2.6. цієї роботи за допомогою дифузійної моделі Stable Diffusion ми згенерували кілька наборів нових зображень на основі 100 зображень з обмеженого набору даних.

Завданням третьої частини експерименту є навчання нейронних мереж для сегментації зображень з тими ж параметрами нейронної мережі, що й у базового варіанта, на п'яти наборах синтетичних даних, а також виявлення оптимальної кількості зображень, що генеруються.

Перший набір даних - 100 вихідних зображень, 100 синтетичних зображень із зміненим птахом, 100 синтетичних зображень із зміненим фоном (загалом 300 зображень). Процес навчання та оцінки візуалізований на рис. 3.5 та 3.6. Модель навчалася 913 епох.

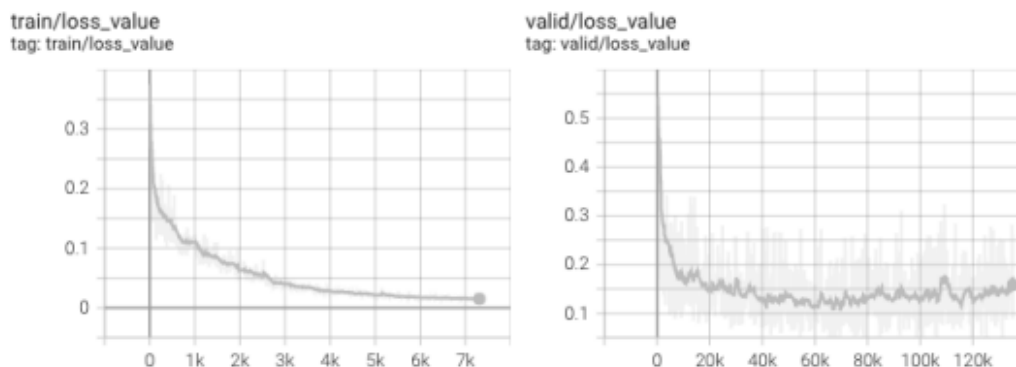


Рисунок 3.5 – Функція втрат (набір 100+200 зображень)

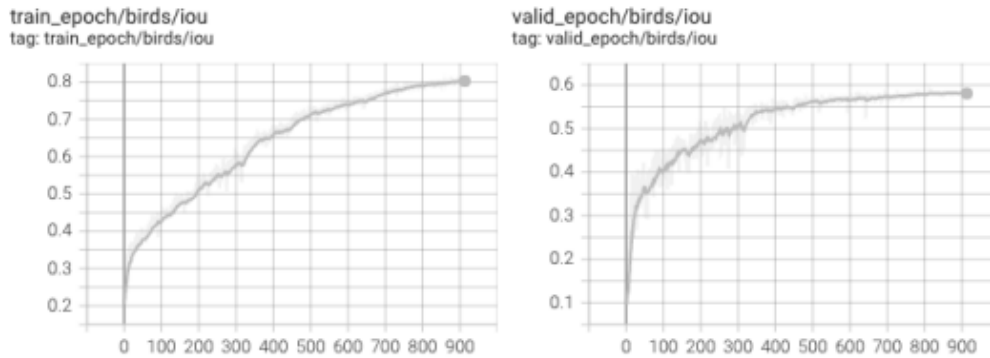


Рисунок 3.6 – Метрика IoU (набір 100 + 200 зображень)

Другий набір даних – 100 вихідних зображень, 200 нових синтетичних зображень із зміненим птахом та 200 нових синтетичних зображень із зміненим фоном (загалом 500 зображень). На рис. 3.7 та 3.8 показаний процес навчання та оцінки моделі. Модель навчалася протягом 557 епох.

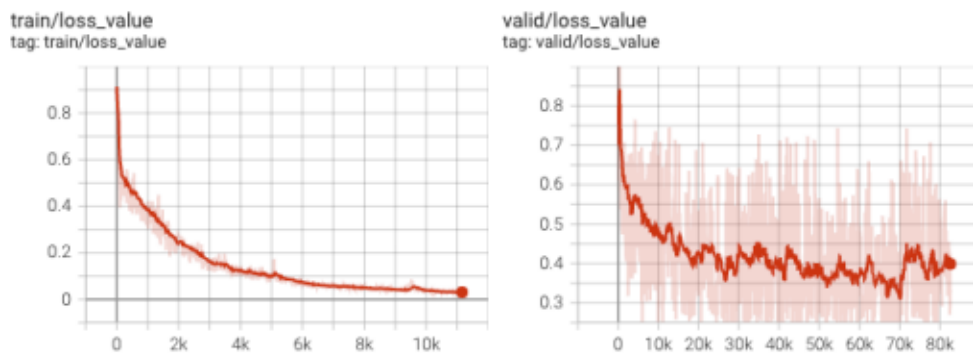


Рисунок 3.7 – Функція втрат (набір 100+400 зображень)

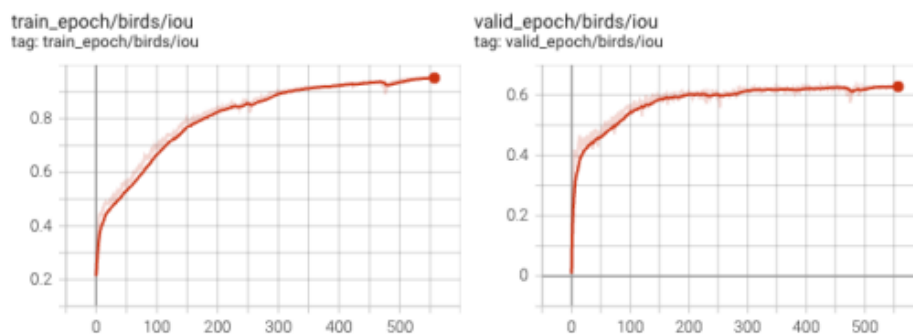


Рисунок 3.8 – Метрика IoU (набір 100 + 400 зображень)

Третій набір даних - 100 вихідних зображень, на основі яких було згенеровано 400 нових зображень зі зміненим птахом та 400 нових зображень зі

зміненим фоном (всього 900 зображень). Процес навчання та оцінки продемонстровано на рис. 3.9 та 3.10. Тривалість навчання – 565 епох.

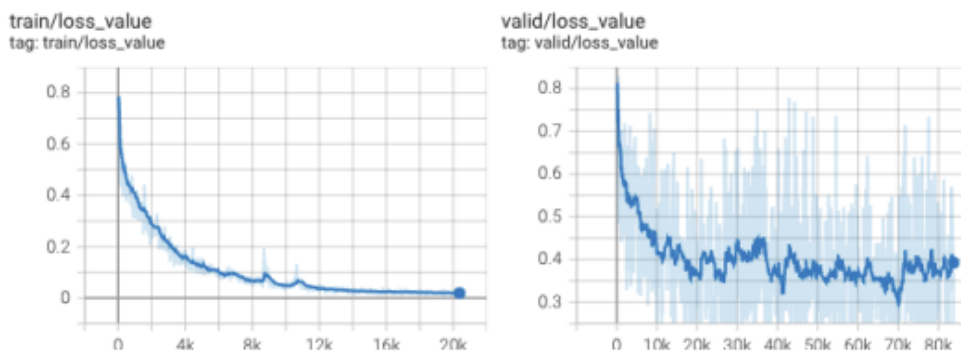


Рисунок 3.9 – Функція втрат (набір 100+800 зображень)

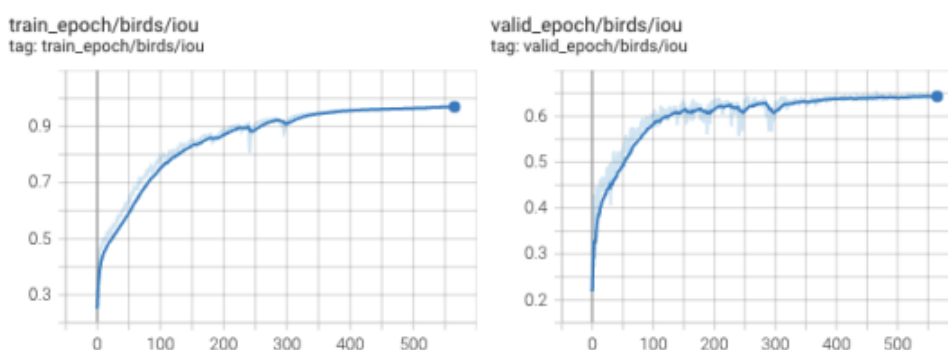


Рисунок 3.10 – Метрика IoU (набір 100 + 800 зображень)

Четвертий набір даних складається з 100 вихідних зображень, 500 зображень зі зміненим птахом і 500 зображень зі зміненим фоном (всього 1100 зображень). Процес навчання та оцінки візуалізований на рис. 3.11 та 3.12. Модель навчалася 342 епохи.

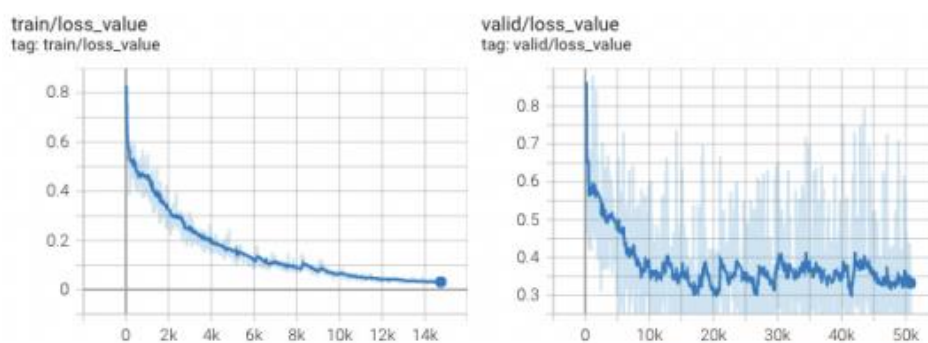


Рисунок 3.11 – Функція втрат (набір 100+1000 зображень)

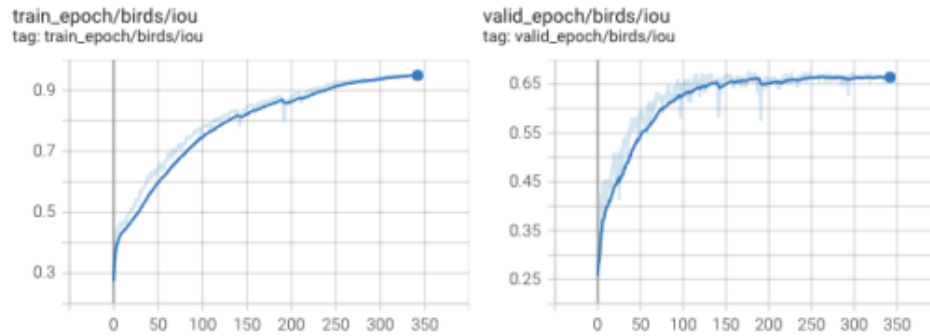


Рисунок 3.12 – Метрика IoU (набір 100 + 1000 зображень)

П'ятий набір даних - 100 вихідних зображень, 700 синтетичних зображень із зміненим птахом, 700 синтетичних зображень із зміненим фоном (всього 1500 зображень). Процес навчання та оцінки продемонстровано на рис. 3.13 та 3.14. Навчання моделі склало 676 епох.

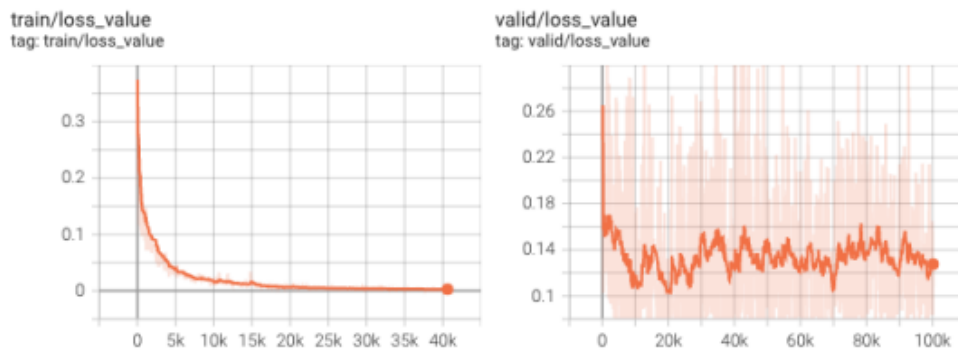


Рисунок 3.13 – Функція втрат (набір 100+1400 зображень)

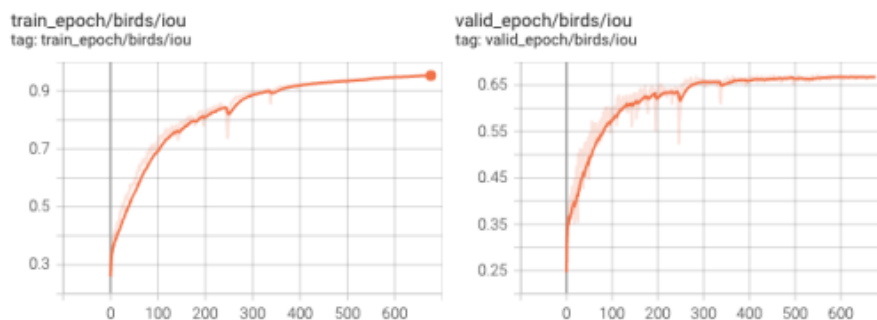


Рисунок 3.14 – Метрика IoU (набір 100 + 1400 зображень)

Також навчимо два додаткові набори даних, які були сформовані з метою дізнатися, чи має якийсь один із видів зображень, що генеруються, більший

вплив на показники навчання моделей:

Перший додатковий набір складається з 100 вихідних і 500 синтетичних зображень тільки зі зміненим птахом (в наборі 600 зображень). Модель навчалася 821 епоху. Результати навчання наведені на рис. 3.15 і 3.16.

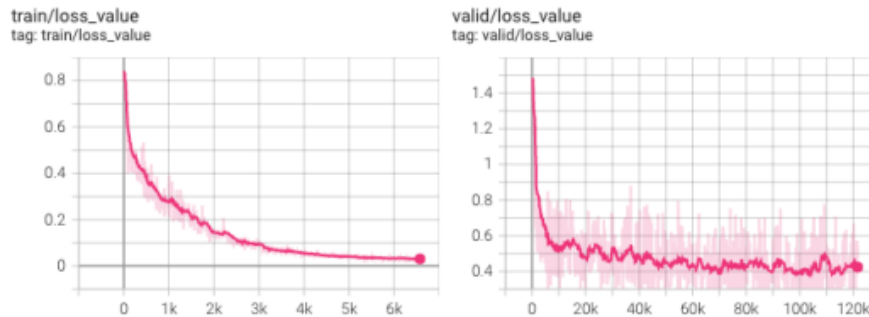


Рисунок 3.15 – Функція втрат (набір 100+500 зображень тільки з птахом)

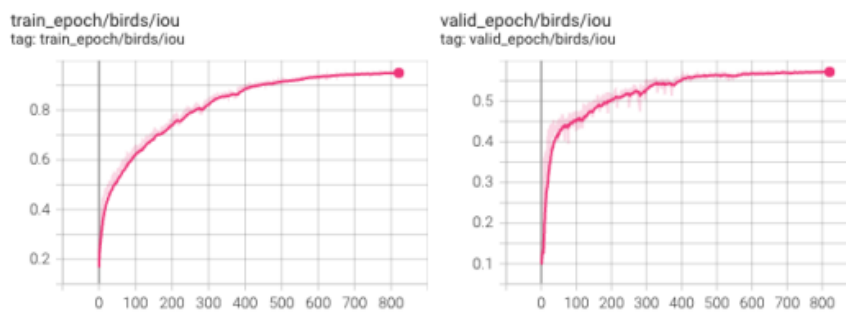


Рисунок 3.16 – Метрика IoU (набір 100 + 500 зображень тільки з птахом)

Другий додатковий набір був сформований зі 100 вихідних, 500 синтетичних зображень тільки зі зміненим фоном (в наборі 600 зображень). Навчання моделі становило 552 епохи. Процес навчання та оцінки наведені на рис. 3.17 та 3.18.

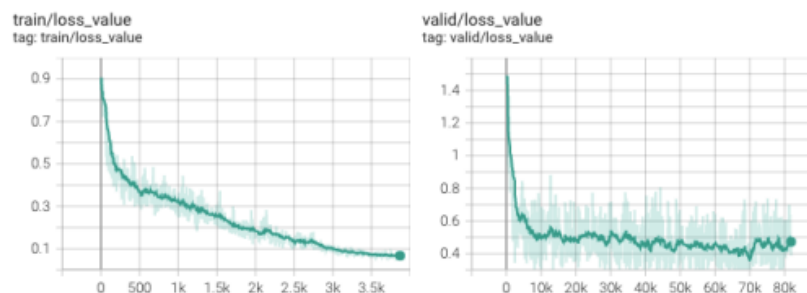


Рисунок 3.17 – Функція втрат (набір 100+500 зображень тільки з фоном)

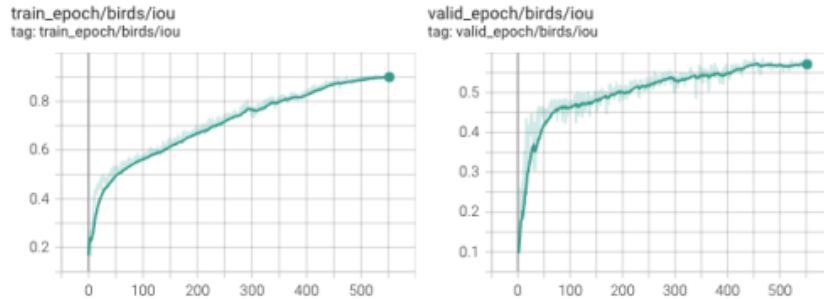


Рисунок 3.18 – Метрика IoU (набір 100 + 500 зображень тільки з фоном)

Аналіз графіків дозволяє стверджувати, що є досить значні відмінності у значеннях функції втрат на тренувальній та валідаційній вибірках, є ознаки перенавчання моделей на тренувальних даних. Це також стосується і значень метрики IoU.

### 3.4 Порівняння результатів

Вище вже було розглянуто окремо кожен частину експерименту. Тепер порівняємо результати. Найкращі результати очікувано показала модель, навчена на повному наборі тренувальних даних. У неї немає ознак перенавчання, значення функції втрат і метрики IoU на тренувальних даних та валідаційних даних знаходяться досить близько один до одного, що говорить про високу узагальнюючу здатність на моделі.

Що стосується моделі, навченої на обмежених даних, то ми бачимо перенавчання моделі на тренувальних даних. При високих значеннях IoU на тренувальній вибірці, значення IoU на валідаційних даних значно поступаються показникам моделі, навченої на повних даних.

Моделі, навчені на обмеженому наборі даних, об'єднаному із синтетичними, показали результати, що лежать між показниками попередніх двох моделей. Незважаючи на перенавчання на тренувальних даних моделі мають більш високу узагальнюючу здатність і більше мають більш високі результати на валідаційних даних, ніж модель, навчена тільки на обмеженому датасеті.

На рис. 3.19 можна простежити залежність якості роботи моделей від кількості даних, що згенерували, на основі вихідних зображень. Тут помаранчевим позначений набір з 200 синтетичними зображеннями, рожевим - із 400 зображеннями, зеленим - із 800 зображеннями, блакитним - із 1000 зображень, синім - із 1400 зображеннями.

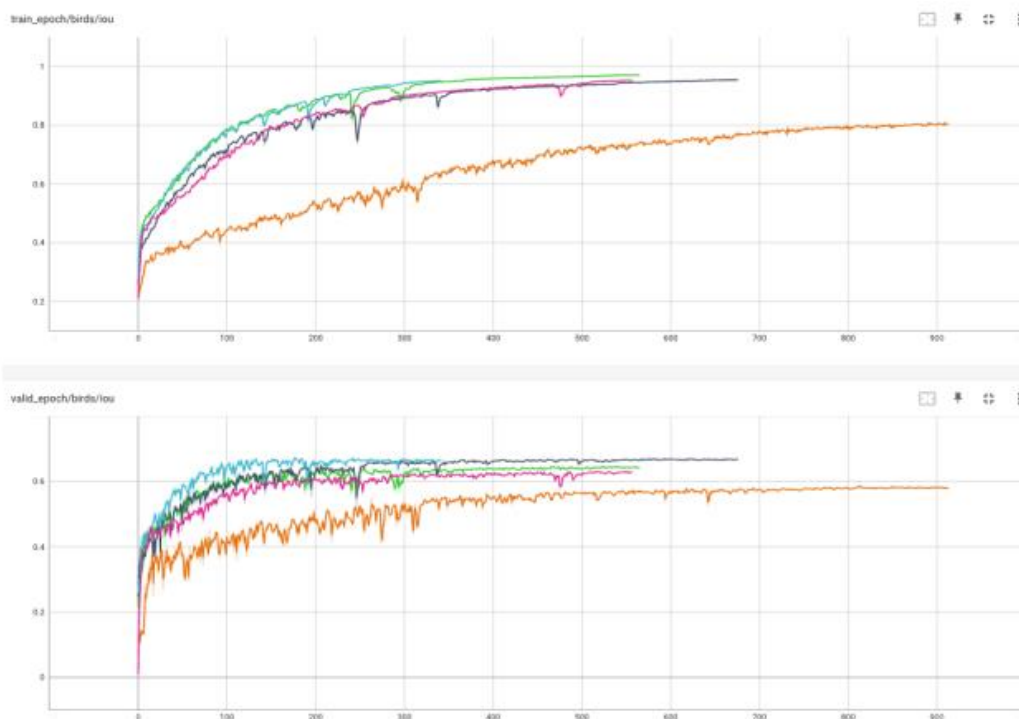


Рисунок 3.19 – Метрика IoU - порівняння синтетичних наборів

Аналіз графіків демонструє, що збільшення кількості даних, що згенерували, до певного моменту призводить до відчутного поліпшення якості роботи моделі, але потім уповільнюється. Різниця в показниках моделей, навчених на четвертому (100 вихідних та 1000 синтетичних) та п'ятому (100 вихідних та 1400 синтетичних) наборах даних, не стільки значна, але поліпшення все одно є. Звідси вважатимемо п'ятий набір даних найоптимальнішим за співвідношенням кількості генерованих даних одне вихідне зображення.

Порівняємо також два додаткові набори даних, ґрунтуючись на графіках, наведених на рис. 3.20. та 3.21. Дані набори були сформовані та навчені для аналізу впливу кожного з видів генерованих зображень (генерування нового фону та генерування нового птаха) на показники якості навчених моделей. Тут

зеленим позначений набір даних із зміненим фоном, рожевим - набір даних із зміненими птахами.

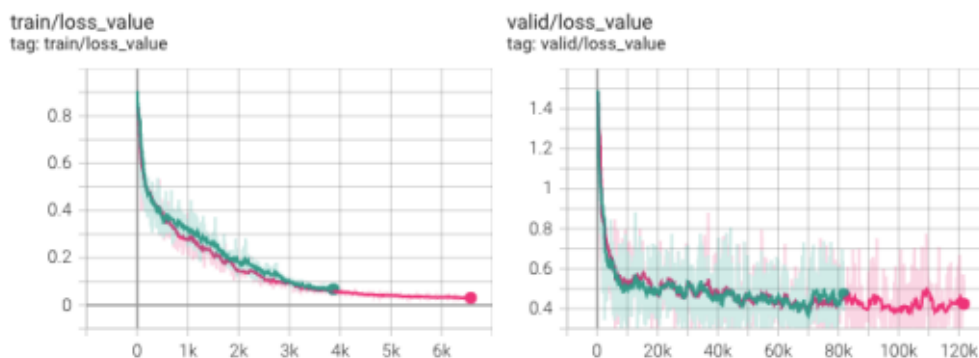


Рисунок 3.20 – Функція втрат - додаткові набори даних

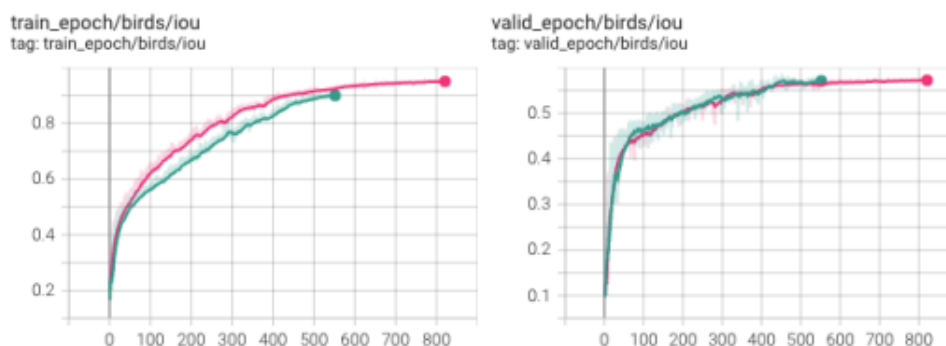


Рисунок 3.21 – Метрика IoU - додаткові набори даних

Аналіз показав, що жоден з видів зображень, що генеруються, не має переважачого впливу на якість навчання моделей. Зображення з новим фоном та зображення зі зміненими птахами демонструють результат приблизно на одному рівні.

Тепер перейдемо до порівняння трьох моделей – моделі, навченої на повних даних, моделі, навченої на обмежених даних та моделі, навченої на обмежених даних, об'єднаних із синтетичними. Останньою виберемо модель, що показала найкращі результати серед розглянутих вище п'яти моделей із синтетичними даними – а саме п'яту модель (100 вихідних зображень та 1400 згенерованих).

Порівняння моделей показано на рис. 3.22 та 3.23. Тут червоним

позначений повний тренувальний набір даних, зеленим - обмежений, синім - обмежений, поєднаний із синтезованим.

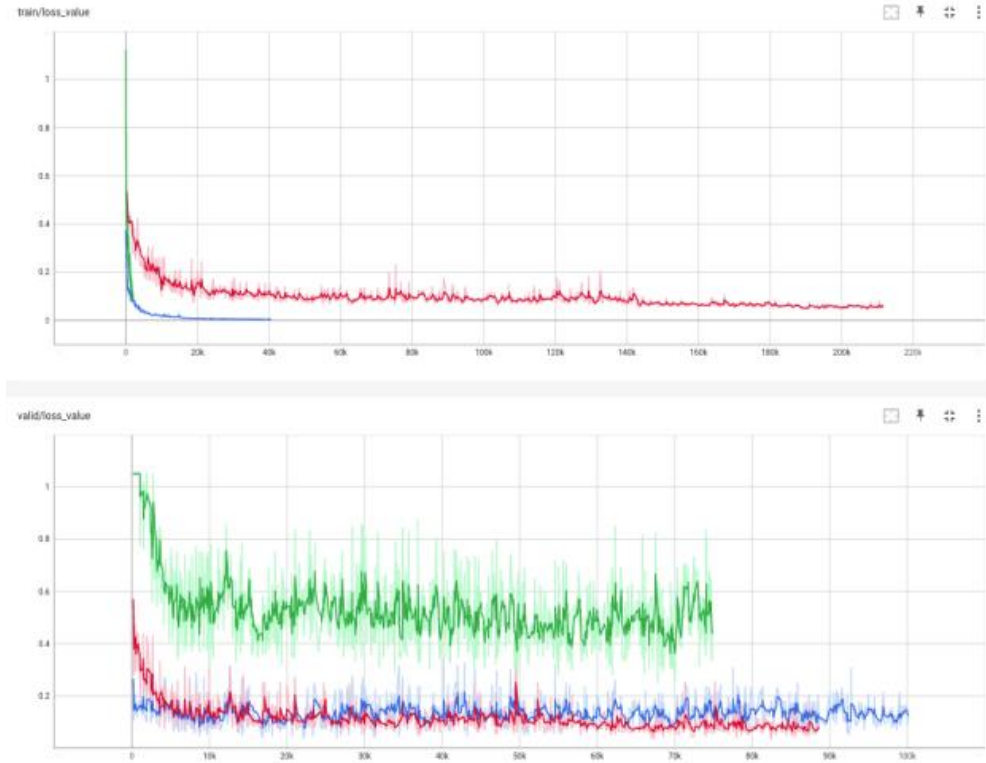


Рисунок 3.22 – Функція втрат

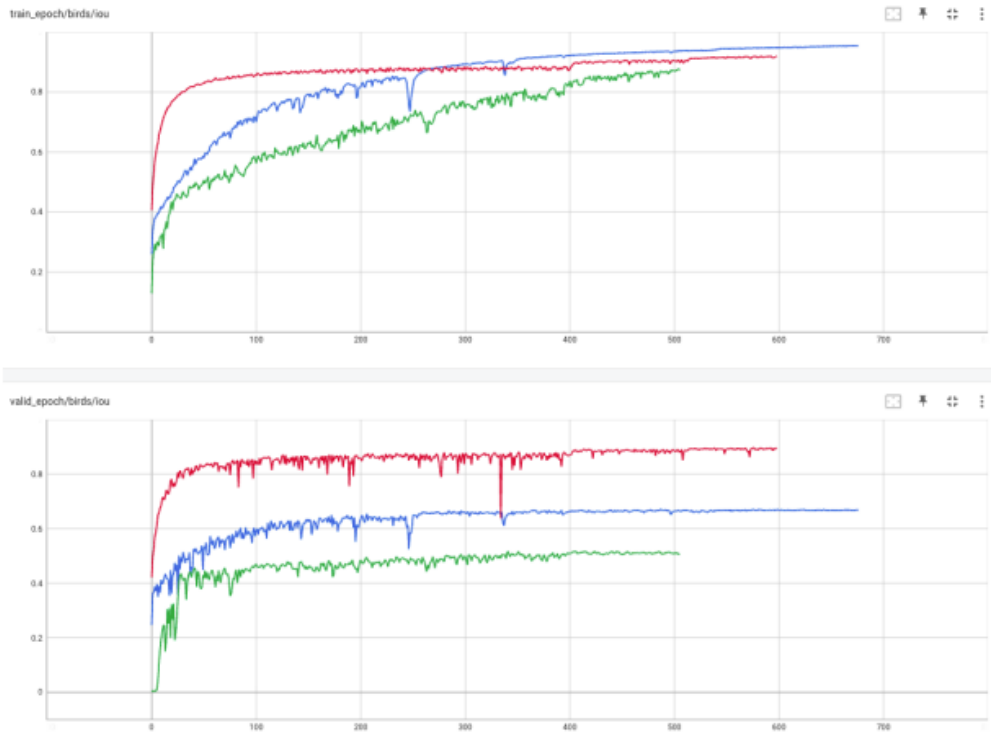


Рисунок 3.23 – Метрика IoU

Тепер порівняємо результати роботи отриманих моделей на тестових даних і переконаємося, що підтверджується вихідна ідея експерименту в тому, що маючи обмежену кількість даних, ми зможемо за допомогою дифузійної моделі збільшити різноманітність зображень у тренувальній вибірці, підвищити узагальнюючу здатність моделі, що навчається на цих даних, та отримати більш оптимальні результати роботи моделі.

Результати оцінки роботи моделей, навчених на повній навчальній вибірці, на обмеженій вибірці та наборах даних з різною кількістю синтезованих зображень, наведені в табл. 3.1.

Таблиця 3.1 – Значення метрик ефективності моделей на тестових даних

<b>Набір даних</b>	<b>Accuracy</b>	<b>IoU</b>
Повна навчальна вибірка (8 990 зображень)	0,9831	0,8847
Обмежена навчальна вибірка (100 зображень)	0,9381	0,508
Обмежена вибірка та синтетичні дані (100 + 200 зображень)	0,9477	0,5689
Обмежена вибірка та синтетичні дані (100 + 400 зображень)	0,9542	0,6272
Обмежена вибірка та синтетичні дані (100 + 800 зображень)	0,9572	0,6319
Обмежена вибірка та синтетичні дані (100 + 1000 зображень)	0,963	0,6523
Обмежена вибірка та синтетичні дані (100 + 1400 зображень)	0,9636	0,6565

На рис. 3.24 графічно відображена залежність цільової метрики IoU від кількості аугментованих даних для обмеженої тренувальної вибірки



Рисунок 3.24 – Залежність метрики IoU від кількості аугментованих даних для обмеженої тренувальної вибірки

Аналіз графіку на рис. 3.24 каже про те, що збільшення обсягу синтезованих даних на основі вихідних даних є доцільним лише до певної межі. Ми бачимо значне покращення значень цільової метрики IoU при додаванні 400% аугментованих даних по відношенню до вихідної тренувальної вибірки. При збільшенні обсягу синтезованих даних до 1000% зростання значення метрики IoU продовжується, але вже не так значне. Після 1000% відбувається насичення даними і додаткове збільшення обсягу даних не призводить до зростання значення метрики. Таким чином, аугментація, що породжує, здатна створювати нові варіації і до певної міри покращувати якість навчання моделі, але не повністю замінити оригінальні дані.

Дані, наведені в табл. 3.1 та на рис. 3.24 дозволяють зробити висновок, що вихідна ідея експерименту підтверджується. Показовими для нас значення метрики IoU. Ми бачимо, що завдяки розширенню обмеженого датасету за рахунок синтетичних даних нам вдалося покращити значення метрики IoU з 0,508 на обмежених даних до 0,6565 на обмеженій вибірці, розширеній за рахунок 1400 синтетичних зображень. Тим самим вдалося наблизитися до значення IoU у базовому варіанті моделі, а покращення склало 14,85%.

### **3.5 Рекомендації щодо використання Stable Diffusion та можливості покращення експериментів**

Досвід генерації нових зображень за допомогою Stable Diffusion показав, що для досягнення кращих результатів необхідно врахувати деякі особливості. Виділимо основні:

- підбір текстового запиту для генерації зображень («prompt» та «negative prompt»): важливо протестувати різні варіанти та підібрати оптимальний текстовий опис запиту для генерації зображення, а також прописати категорії, які моделі буде заборонено генерувати;

- налаштування параметрів «Inpaint»: поряд із підбором текстового запиту цей крок є одним із ключових. Тому необхідно провести експерименти та дійти найбільш оптимальних значень. Іноді краще відмовитися від великої варіативності генерації на користь її більшої стабільності, щоб уникнути генерацій з мутаціями або надмірного малювання;

- вибір відповідної моделі для Stable Diffusion: оскільки існує досить велика кількість продуманих моделей для генерації зображень, важливо протестувати та вибрати відповідну;

- розмір зображень: оскільки Stable Diffusion навчена на зображеннях розміром 512x512 пікселів, то й найвдаліші синтетичні зображення виходять саме у цьому розмірі. При необхідності розмір зображення після створення можна відкоригувати;

Також можна намітити подальші можливості для покращення експериментів та покращення якості роботи моделі сегментації, зокрема:

- експерименти з кількістю вихідних даних для генерації нових зображень: в експериментах ми взяли 100 зображень і на їх основі генерували нові дані, але можна провести тести, взявши 500 або 1000 вихідних зображень;

- продовжити експерименти з кількістю даних, що згенерували, на одне вихідне зображення;

- навчання інших архітектур нейронних мереж для вирішення задачі семантичної сегментації (наприклад, SegNet або DeepLab);

- розширити набір даних за рахунок генерації повністю синтетичних даних:
  - через підхід «txt2img» від Stable Diffusion згенерувати абсолютно нове зображення, що відповідає змісту нашого набору даних:
  - створити маску згенерованого зображення - наприклад, за допомогою розширення «depthmap2mask» від Stable Diffusion створити маску зображення (за рахунок спеціального алгоритму це розширення визначає глибину зображення і створює маску з різною прозорістю - як правило об'єкт у фокусі зафарбовується інтенсивно, а фон не зафарбовується) або іншим відповідним способом;
  - проте запропонований метод необхідно тестувати і підбирати відповідні параметри.

## **РОЗДІЛ 4. БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ОХОРОНИ ПРАЦІ**

### **4.1 Класифікація шкідливих та небезпечних виробничих факторів**

Шкідливий виробничий фактор – небажане явище, яке супроводжує виробничий процес і вплив якого на працюючого може призвести до погіршення самопочуття, зниження працездатності, захворювання, виробничо зумовленого чи професійного, і навіть смерті, як результату захворювання. Небезпечний виробничий фактор – небажане явище, яке супроводжує виробничий процес і дія якого за певних умов може призвести до травми або іншого раптового погіршення здоров'я працівника (гострого отруєння, гострого захворювання) і навіть до раптової смерті [41].

Поділ несприятливих чинників виробничого середовища на шкідливі та небезпечні зумовлене різним характером їх дії на людський організм, тим, що вони потребують різних заходів та засобів для боротьби з ними та профілактики викликаних ними ушкоджень, а також рядом причин організаційного характеру. В той же час між шкідливими та небезпечними виробничими факторами інколи важко провести чітку межу. Один і той же чинник може викликати травму і профзахворювання (наприклад, високий рівень іонізуючого або теплового випромінювання може викликати опік або навіть призвести до миттєвої смерті, а довготривала дія порівняно невисокого рівня цих же факторів – до хвороби; пилинки, що потрапили в око, спричиняє травму, а пил, що осідає в легенях, – захворювання, що зветься пневмоконіоз). Через це всі несприятливі виробничі чинники часто розглядаються як єдине поняття – небезпечний та шкідливий виробничий фактор (НШВФ) [41]. За своїм походженням та природою дії всі НШВФ можна поділити на 5 груп: фізичні, хімічні, біологічні, психофізіологічні та соціальні. До фізичних НШВФ відносяться машини та механізми або їх елементи, а також вироби, матеріали, заготовки тощо, які рухаються або обертаються; конструкції, які руйнуються; системи, устаткування або елементи обладнання, які знаходяться під підвищеним тиском; підвищена запиленість та

загазованість повітря; підвищена або понижена температура повітря, поверхонь приміщення, обладнання, матеріалів; підвищені рівні шуму, вібрації, ультразвуку, інфразвуку; підвищений або понижений барометричний тиск та його різкі коливання; підвищена та понижена вологість; підвищена швидкість руху та підвищена іонізація повітря; підвищений рівень іонізуючих випромінювань; підвищене значення напруги в електричній мережі; підвищені рівні статичної електрики, електромагнітних випромінювань; підвищена напруженість електричного, магнітного полів; відсутність або нестача світла; недостатня освітленість робочої зони; підвищена яскравість світла; понижена контрастність; прямий та віддзеркалений блиск; підвищена пульсація світлового потоку; підвищені рівні ультрафіолетової та інфрачервоної радіації; гострі крайки, зачипки, шершавість на поверхні заготовок, інструментів та обладнання; розташування робочого місця на значній висоті відносно землі (підлоги); слизька підлога; невагомість.

#### Хімічні НШВФ:

- за характером дії на організм людини поділяються на токсичні, задушливі, наркотичні, подразнюючі, сенсibiliзуючі, канцерогенні, мутагенні та такі, що впливають на репродуктивну функцію;

- за шляхами проникнення в організм людини поділяються на такі, що потрапляють через: 1) органи дихання; 2) шлунково-кишковий тракт; 3) шкіряні покриви та слизова оболонка;

- які перебувають у різному агрегатному стані: 1) твердому 2) газоподібному 3) рідкому.

Біологічні НШВФ – це: - патогенні мікроорганізми (бактерії, віруси, рикетсії, спірохети, грибки, найпростіші) та продукти їхньої життєдіяльності; - макроорганізми (тварини та рослини) та продукти їхньої життєдіяльності. До психофізіологічних НШВФ відносяться фізичні (статичні та динамічні) перевантаження і нервово-психічні перевантаження (розумове перенапруження, перенапруження аналізаторів, монотонність праці, емоційні перевантаження).

6 Соціальні НШВФ – це неякісна організація роботи, понаднормова робота, змушеність праці в колективі з поганими відносинами між його членами,

соціальна ізоляція з відривом від сім'ї, зміна біоритмів, незадоволеність роботою, фізична та/або словесна образа та її ризик, насильство та його ризик. Один і той же НШВФ за природою своєї дії може належати водночас до різних груп.

## **4.2 Вплив вібрації на людину**

Вібрація - це механічні коливання пружних тіл або коливальні рухи механічних систем. Для людини вібрація є видом механічного впливу, який має негативні наслідки для організму [42].

Причиною появи вібрації є невідновлені сили та ударні процеси в діючих механізмах. Створення високопродуктивних потужних машин і швидкісних транспортних засобів при одночасному зниженні їх матеріалоемності неминуче призводить до збільшення інтенсивності і розширення спектру вібраційних та віброакустичних полів. Цьому сприяє також широке використання в промисловості і будівництві високоефективних механізмів вібраційної та віброударної дії.

Дія вібрації може приводити до трансформування внутрішньої структури і поверхневих шарів матеріалів, зміни умов тертя і зносу на контактних поверхнях деталей машин, нагрівання конструкцій. Через вібрацію збільшуються динамічні навантаження в елементах конструкцій, стиках і сполученнях, знижується несуча здатність деталей, ініціюються тріщини, виникає руйнування обладнання. Усе це приводить до зниження строку служби устаткування, зростання імовірності аварійних ситуацій і зростання економічних витрат. Вважають, що 80% аварій в машинах і механізмах здійснюється внаслідок вібрації. Крім того, коливання конструкцій часто є джерелом небажаного шуму. Захист від вібрації є складною і багатоплановою в науково-технічному та важливою у соціально-економічному відношеннях проблемою нашого суспільства [42].

Вплив вібрації на людину залежить від її спектрального складу, напрямку дії, прикладення, тривалості впливу, а також від індивідуальних особливостей людини. При оцінці вібраційного впливу потрібно враховувати, що коливальні

процеси притаманні живому організму. В основі серцевої діяльності і кровообігу та біострумів мозку лежать ритмічні коливання. Внутрішні органи людини можна розглядати як коливальні системи з пружними зв'язками. Частоти їх власних коливань лежать у діапазоні 3..6 Гц. Частоти власних коливань плечового пояса, стегон і голови щодо опорної поверхні (положення стоячи) складають 4...6 Гц, голови щодо плечей (положення сидячи) 25...30 Гц.

При впливі на людину зовнішніх коливань (хитавиці, струсів, вібрації) відбувається їхня взаємодія з внутрішніми хвильовими процесами, виникнення резонансних явищ. Так, зовнішні коливання частотою менш 0,7 Гц утворюють хитавицю і порушують у людини нормальну діяльність вестибулярного апарата. Інфразвукові коливання (менш 16 Гц), впливаючи на людину, пригнічують центральну нервову систему, викликаючи почуття тривоги, страху. При певній інтенсивності на частоті 6..7 Гц інфразвукові коливання, втягуючи у резонанс внутрішні органи і систему кровообігу, здатні викликати травми, розриви артерій, тощо [42].

Вібрація, що діє на людину, має широкий діапазон – від десятих часток одного до декількох тисяч Гц. Характерними ознаками шкідливого впливу вібрації на людину є можливі зміни у функціональному стані: підвищена втома, збільшення часу моторної реакції, порушення вестибулярної реакції. Медичними дослідженнями встановлено, що вібрація є подразником периферичних нервових закінчень, розташованих на ділянках тіла людини, що сприймають зовнішні коливання. Адекватним фізичним критерієм оцінки її впливу на організм людини є коливальна енергія, що виникає на поверхні контакту, а також енергія, поглинена тканинами і передана опорно-руховому апарату та іншим органам. У результаті впливу вібрації виникають нервовосудинні розлади, ураження кістково-суглобної та інших систем організму. Відзначаються, наприклад, зміни функції щитовидної залози, сечостатевої системи, шлунково-кишкового тракту. Так, медичні дослідження показали, що у працюючих в умовах вібрації відбуваються значні зміни кістковосуглобної системи, які виражаються у функціональній перебудові кісткової тканини, регіональному остеопорозі, кістковидних утвореннях у кістках, асептичному некрозі кісток, хронічних

переломах. Відзначається, що терміни виникнення змін у кістках у працівників вібраційних професій коливається в межах від 6-8 місяців до 2-5 років.

Шкідливість вібрації збільшується при одночасному впливі на людину таких факторів, як знижена температура, підвищений шум, запиленість повітря, тривала статична напруга тощо. Сучасна медицина розглядає виробничу вібрацію як могутній стрес-фактор, що має негативний вплив на психомоторну працездатність, емоційну сферу і розумову діяльність, підвищує ймовірність виникнення різних захворювань і нещасних випадків. Особливо небезпечний тривалий вплив вібрації для жіночого організму. Широкий комплекс патологічних відхилень, викликаний впливом вібрації на організм людини, кваліфікується як віброзахворювання [42].

Вібрація як фізичний чинник виробничого середовища спостерігається в металообробній, гірничодобувній, металобудівній, машинобудівній, авіаційній та інших галузях народного господарства. Джерелом вібрації можуть бути різні механізми, вібраційне устаткування, віброінструменти, акустичні системи, транспортні та сільськогосподарські машини.

Загальна вібрація поділяється на транспортну вібрацію, яка діє на людину на робочих місцях в транспортних засобах (трактори сільськогосподарські та промислові, самохідні сільськогосподарські машини (комбайни), тягачі, грейдери ті інші); транспортно-технологічну вібрацію, яка діє на людину на робочих місцях машин з обмеженою рухливістю (екскаватори, крани промислові та будівельні, гірничі комбайни, транспорт виробничих приміщень та інші) та технологічну вібрацію, яка діє на людину на робочих місцях стаціонарних машин чи передається на робочі, де немає джерел вібрації (верстати та метало-деревообробне, пресувально-ковальське обладнання, ливарні машини, електричні машини, насосні агрегати та вентилятори, обладнання для буріння свердловин, бурові верстати, машини для тваринництва, очищення та сортування зерна (у тому числі сушарні), обладнання промисловості будматеріалів (крім бетоноукладачів), установки хімічної та нафтохімічної промисловості та інші.

Оператори машин, які зазнають у процесі трудової діяльності впливу вібрації, підлягають попереднім та періодичним медичним оглядам відповідно

до Порядку проведення медичних оглядів працівників певних категорій, затвердженого Наказом МОЗ України від 21.05.2007 р. №246. Обов'язкові попередні (під час прийняття на роботу) та періодичні (протягом трудової діяльності) медичні огляди дозволять визначити стан здоров'я працівника та можливість виконання без погіршення стану здоров'я професійних обов'язків, своєчасно виявити ранні ознаки хронічного професійного захворювання, забезпечує динамічне спостереження за станом здоров'я в умовах дії шкідливих та небезпечних факторів і трудового процесу, вирішує питання щодо можливості продовжувати роботу в умовах дії шкідливих та небезпечних факторів і трудового процесу [42].

За результатами періодичних медичних оглядів роботодавець забезпечує проведення відповідних оздоровчих заходів Заключного акта у повному обсязі та усуває причини, що призводять до професійних захворювань. Організовує проведення лабораторних досліджень умов праці на робочих місцях та вживає заходів до усунення небезпечних і шкідливих для здоров'я виробничих факторів.

До роботи операторами машин допускаються особи не молодші 18 років, які пройшли попередній медичний огляд, мають відповідну кваліфікацію та ознайомлені з характером впливу вібрації на організм.

## ВИСНОВКИ

У цій роботі було проведено дослідження ефективності генеративної аугментації зображень, виконаної за допомогою дифузійної моделі Stable Diffusion, для вирішення задачі семантичної сегментації на прикладі датасету Caltech-UCSD Birds-200-2011 .

Для цього було розглянуто основні підходи до вирішення задач сегментації зображень та основні групи методів аугментації. Спираючись на вихідну ідею дослідження про те, що за допомогою генеративної аугментації можна синтезувати нові зображення і тим самим збільшити різноманітність зображень у тренувальній вибірці та підвищити узагальнюючу здатність моделі, було розроблено структуру проведення експериментів.

Для вирішення задачі семантичної сегментації було обрано повнозгорткову нейронну мережу U-Net, з моделлю класифікації Resnet34 як кодувальник. Як функція втрат була використана функція втрат Тверського (Tversky loss). Оцінка ефективності моделей здійснювалася за допомогою метрик IoU та Accuracy.

Експерименти були проведені у три етапи. На першому етапі була навчена базова модель сегментації на повній навчальній вибірці з набору даних (8990 зображень). На другому етапі була навчена модель з тими ж параметрами, що і базова модель, але навчання проводилося на обмеженій вибірці (100 зображень, вибраних випадковим чином з 8900 зображень повної навчальної вибірки). Для реалізації третього етапу експерименту були попередньо за допомогою функції Inpaint від Stable Diffusion згенеровані кілька нових наборів даних на основі зображень з обмеженої вибірки і об'єднані з нею. Далі на отриманих даних були навчені моделі з тими самими параметрами, що й базова модель.

Тестування моделей з різним обсягом аугментованих даних стосовно вихідного набору показало, що збільшення обсягу синтезованих даних може бути корисним лише до певної межі. У нашому випадку такою межею став поріг 1000%. Після досягнення зазначеного обсягу даних спостерігалася насичення, і додаткове синтезування даних не призводило до значного зростання значення

цільової метрики. Таким чином, аугментація, що породжує, здатна урізноманітнити дані до певної межі, але повністю замінити або заповнити оригінальні дані не може.

Порівняння результатів роботи моделей на трьох наборах даних (повний, обмежений та аугментований) підтвердили ефективність використання запропонованого методу генеративної аугментації даних. Доповнення обмеженого датасета синтетичними даними дозволило покращити значення метрики IoU з 0,508 на обмеженій вибірці до 0,6565 на розширеній вибірці (тобто на 14,85%) та наблизитися до значення метрики IoU 0,8847 у базовому варіанті моделі.

В роботі був продемонстрований підхід до аугментації зображень, реалізований за рахунок розширення частини вихідного датасета шляхом генерування нових даних за допомогою дифузійної моделі і при цьому не вимагає збирання додаткових даних та трудомісткості розмітки зображень.

Отож застосування сучасних GAN для аугментації невеликих наборів даних чи наборів даних із дисбалансом класів дає змогу набагато поліпшити якість функціонування моделей, обійшовши при цьому можливі проблеми, пов'язані з використанням простих методів аугментації.

## ПЕРЕЛІК ДЖЕРЕЛ

1. Yeromina, N. & Koltun, Yu & Bessalyi, A. & Shmatko, Yu. (2024). Аналіз сучасних методів сегментації зображень в інтересах навігації мобільних роботів. Системи управління, навігації та зв'язку. Збірник наукових праць. 2. С. 82-86.
2. Goodfellow, I., Bengio, Y., & Courville, A. "Deep Learning." MIT Press, 2016. 788 p.
3. TENSORFLOW І МАШИННЕ НАВЧАННЯ. [Електронний ресурс] – Режим доступу: <https://foxminded.ua/tensorflow-shcho-tse/> (дата звертання: 03.05.2026).
4. Long J. Fully Convolutional Networks for Semantic Segmentation / J. Long, E. Shelhamer, T. Darrell // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. № 4. Vol. 39. P. 640–651.
5. Ronneberger O. U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, Ph. Fischer, T. Brox // International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI-2015). – 2015. Vol. 39. P. 234–241.
6. Badrinarayanan V. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation / V. Badrinarayanan, A. Kendall, R. Cipolla // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. № 13. 57 Vol. 39. P. 1–14.
7. Khalifa N. A comprehensive survey of recent trends in deep learning for digital images augmentation / N. Khalifa, M. Loey, S. Mirjalili // Artificial Intelligence Review. An International Science and Engineering Journal. 2022. № 55. P. 2351–2377.
8. Методичні вказівки до виконання кваліфікаційної роботи оп Бакалавр для студентів спеціальності 122 – Комп'ютерні науки, всіх форм навчання / укладачі: Готович В.А., Дуда О.М. Никитюк В.В. – Тернопіль: Тернопільський національний технічний університет імені Івана Пулюя, 2024. 43 с.

9. Alomar K. Data Augmentation in Classification and Segmentation: A Survey and New Strategies / K. Alomar, H.I. Aysel, X. Cai // Image Segmentation Techniques: Current Status and Future Directions. 2023. № 9 (2).

10. Lytvynenko, I., Lupenko, S., Nazarevych, O., Shymchuk, G., & Hotovych, V. (2021). Mathematical model of gas consumption process in the form of cyclic random process. 2021 IEEE 16th International Conference on Computer Sciences and Information

11. Як працюють моделі для генерації зображень. [Електронний ресурс] – Режим доступу: <https://dou.ua/forums/topic/53934/> (дата звертання: 03.05.2026).

12. Tanaka F. Data augmentation using gans / F. Tanaka, C. Aranha [Електронний ресурс] – Режим доступу: <https://arxiv.org/abs/1904.09135> (дата звертання: 03.05.2026).

13. Yamaguchi S. Effective data augmentation with multi-domain learning gans / S. Yamaguchi, S. Kanai, T. Eda // Proceedings of the AAAI Conference on Artificial Intelligence. 2020. № 34 (04). P. 6566–6574.

14. Lytvynenko I. V. Method of segmentation of determined cyclic signals for the problems related to their processing and modeling. Scientific journal of the Ternopil National Technical University. No. 4 (88). 2017. ISSN: 2522-4433. P. 153–169. [https://doi.org/10.33108/visnyk\\_tntu2017.04.153](https://doi.org/10.33108/visnyk_tntu2017.04.153)

15. Sohl-Dickstein J. Deep unsupervised learning using nonequilibrium thermodynamics / J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan // Proceedings of the 32nd International Conference on Machine Learning. 2015. Vol. 37. P. 2256–2265.

16. Ho J. Denoising Diffusion Probabilistic Models / J. Ho, A. Jain, P. Abbeel // Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 1–25.

17. He R. Is synthetic data from generative models ready for image recognition? / R. He, S. Sun, X. Yu, C. Xue, W. Zhang, P. Torr, S. Bai, X. Qi // Conference paper at 11th International Conference on Learning Representations. 2023. P. 1–24.

18. Trabucco B. Effective data augmentation with diffusion models / B. Trabucco, K. Doherty, M. Gurinas, R. Salakhutdinov [Електронний ресурс] – Режим доступу: <https://arxiv.org/pdf/2302.07944.pdf> (дата звертання: 04.05.2026).

19. Магія за пікселями: Пояснення дифузійних моделей для генерації AI-арту. [Електронний ресурс] – Режим доступу: <https://sider.ai/uk/blog/ai-tools/the-magic-behind-the-pixels-diffusion-models-explained-for-ai-art-generation> (дата звертання: 05.05.2026).

20. Leung K. Top Python libraries for Image Augmentation in Computer Vision [Електронний ресурс] – Режим доступу: <https://towardsdatascience.com/top-python-libraries-for-image-augmentation-in-computer-vision-2566bed0533e> (дата звертання: 05.05.2026).

21. Alumentations: A Python library for advanced Image Augmentation strategies [Електронний ресурс] – Режим доступу: <https://medium.com/ml-learning-ai/alumentations-a-python-library-for-advanced-image-augmentation-strategies-752bff3a3da0> (дата звертання: 05.05.2026).

22. Stable Diffusion Public Release [[Електронний ресурс] – Режим доступу: <https://stability.ai/blog/stablediffusion-public-release> (дата звертання: 07.05.2026).

23. Як використовувати стабільну дифузійну нейронну мережу [Електронний ресурс] – Режим доступу: <https://alexhost.com/uk/faq/how-to-use-the-stable-diffusion-neural-network/> (дата звертання: 07.05.2026).

24. He K. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification / K. He, X. Zhang, S. Ren, J. Sun // Conference paper at IEEE International Conference on Computer Vision (ICCV). 2015. P. 1–11.

25. Salehi S.S.M. Tversky loss function for image segmentation using 3D fully convolutional deep networks / S.S.M. Salehi, D. Erdogmus, A. Gholipour // Machine Learning in Medical Imaging. Proceedings of 8th International Workshop MLMI 2017. 2017. P. 379–387.

26. Loshchilov I. Decoupled Weight Decay Regularization / I. Loshchilov, F. Hutter // Conference paper at 7th International Conference on Learning Representations. 2019. P.1–8.

27. Stable-diffusion-webui [Електронний ресурс] – Режим доступу: <https://github.com/AUTOMATIC1111/stable-diffusion-webui> (дата звертання: 08.05.2026).

28. API guide [Електронний ресурс] – Режим доступу: <https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/API> (дата звертання: 08.05.2026).
29. Stable Diffusion with Diffusers [Електронний ресурс] – Режим доступу: [https://huggingface.co/blog/stable\\_diffusion](https://huggingface.co/blog/stable_diffusion) (дата звертання: 09.05.2026).
30. Stable Diffusion v1-5 Model Card [Електронний ресурс] – Режим доступу: <https://huggingface.co/runwayml/stable-diffusion-v1-5> (дата звертання: 09.05.2026).
31. Stable-Diffusion-Inpainting [Електронний ресурс] – Режим доступу: <https://huggingface.co/runwayml/stable-diffusion-inpainting> (дата звертання: 09.05.2026).
32. Wah C. The Caltech-UCSD Birds-200-2011 Dataset / C. Wah, S. Branson, P. Welinder, P. Perona, S. Belongie // Caltech Vision Lab. 2011. P. 1–8.
33. Caltech-UCSD Birds-200-2011 (CUB-200-2011) [Електронний ресурс] – Режим доступу: [https://www.vision.caltech.edu/datasets/cub\\_200\\_2011/](https://www.vision.caltech.edu/datasets/cub_200_2011/) (дата звертання: 10.05.2026).
34. CUB-200-2011 Segmentations [Електронний ресурс] – Режим доступу: <https://data.caltech.edu/records/w9d68-gec53> (дата звертання: 10.05.2026).
35. Pytorch [Електронний ресурс] – Режим доступу: <https://github.com/pytorch/vision> (дата звертання: 10.05.2026).
36. Segmentation\_models.pytorch [Електронний ресурс] – Режим доступу: [https://github.com/qubvel/segmentation\\_models.pytorch/tree/master](https://github.com/qubvel/segmentation_models.pytorch/tree/master) (дата звертання: 10.05.2026).
37. TensorBoard: набір інструментів для візуалізації TensorFlow [Електронний ресурс] – Режим доступу: <https://www.tensorflow.org/> (дата звертання: 11.05.2026).
38. Improving Deep Learning Models with Data Augmentation [Електронний ресурс] – Режим доступу: [https://medium.com/@ShortHills\\_Tech/improving-deep-learning-models-withdata-augmentation-d4e3d0a9301b](https://medium.com/@ShortHills_Tech/improving-deep-learning-models-withdata-augmentation-d4e3d0a9301b) (дата звертання: 11.05.2026).
39. Lupenko, S. A., Lytvynenko, I. V., Sverstiuk, A., Shelestovskyi, B., & Horkunenko, A. (2021). Software for Statistical Processing and Modeling of a Set of

Synchronously Registered Cardio Signals of Different Physical Nature. CMIS, 194-205.

40. Bodnarchuk, I., Skorenkyu, Y., Kramar, T., Duda, O., & Nykytyuk, V. (2022). Use of Analytical Hierarchy Process in Scenarios Design for a Digital Museum with XR components. ITTAP, 414–425

41. Заїкіна Д., Глива В. Основи охорони праці та безпека життєдіяльності. 2019. [Електронний ресурс]. – Режим доступу: <https://doi.org/10.31435/rsglobal/001> (дата звернення: 27.05.2026).

42. Безпека в надзвичайних ситуаціях. Методичний посібник для здобувачів освітнього ступеня «магістр» всіх спеціальностей денної та заочної (дистанційної) форм навчання / укл.: Стручок В. С. Тернопіль: ФОП Паляниця В. А., 2022. 156 с.