

Харківський національний університет радіоелектроніки

Факультет _____ комп'ютерної інженерії та управління _____

Кафедра _____ електронних обчислювальних машин _____

Рівень вищої освіти _____ перший (бакалаврський) _____

Спеціальність _____ 123 «Комп'ютерна інженерія» _____
(код і повна назва)

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

Освітня програма _____ Комп'ютерна інженерія _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

“ _____ ” _____ 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _____ Абрагамовичу Вадиму Олександровичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Комп'ютерна система класифікації та сегментації зображень з
використанням машинного навчання _____

затверджена наказом по університету від “ 26 ” травня 2025 р. № 424 Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії _____ 17 червня 2025 р.

3. Вхідні дані до роботи _____

набори зображень _____

Google Colab _____

розпізнавання зображень _____

класифікація _____

сегментація _____

4. Перелік питань, що потрібно опрацювати у роботі _____

Аналіз предметної області _____

Штучні нейронні мережі в задачах обробки зображень _____

Реалізація програмних засобів _____

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій 15 слайдів

6. Консультанти розділів роботи (заповнюється за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / терміни виконання етапів роботи	Примітка
1	Отримання завдання та аналіз літератури	26.05.2025–30.05.2025	
2	Огляд існуючих рішень та алгоритмів	31.05.2025–03.06.2025	
3	Вибір датасетів та архітектури системи	04.06.2025–06.06.2025	
4	Вибір програмних засобів	07.06.2025–08.06.2025	
5	Програмна реалізація	09.06.2025–11.06.2025	
6	Аналіз отриманих результатів	12.06.2025–13.06.2025	
7	Оформлення записки	14.06.2025–16.06.2025	

Дата видачі завдання “ 26 ” травня 2025 р.

Здобувач _____
(підпис)

Керівник роботи _____ ст. викл. В'ячеслав РАДЧЕНКО
(підпис) (посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 60 с., 11 рис., 2 дод., 7 джерел.

ОБРОБКА ЗОБРАЖЕНЬ, КЛАСИФІКАЦІЯ, СЕМАНТИЧНА СЕГМЕНТАЦІЯ, МАШИННЕ НАВЧАННЯ, ГЛИБОКЕ НАВЧАННЯ, ЗГОРТКОВА НЕЙРОННА МЕРЕЖА, CNN, U-NET, TENSORFLOW, GOOGLE COLAB, PYTHON, КОМП'ЮТЕРНИЙ ЗІР, ІНТЕЛЕКТУАЛЬНА СИСТЕМА, ВІЗУАЛІЗАЦІЯ РЕЗУЛЬТАТІВ.

Метою кваліфікаційної роботи є розробка та реалізація комп'ютерної системи, здатної виконувати класифікацію та сегментацію зображень на основі методів машинного навчання, зокрема глибоких згорткових нейронних мереж, з використанням відкритих платформ, таких як Google Colab.

У ході виконання кваліфікаційної роботи проведено огляд сучасних наукових підходів до обробки візуальної інформації, здійснено аналіз особливостей традиційних та глибоких методів машинного навчання, а також розглянуто переваги гібридних архітектур. Для реалізації системи було використано середовище Google Colab, що забезпечило доступ до ресурсів GPU та сучасних бібліотек TensorFlow і Keras.

Практична частина роботи включає підготовку двох датасетів: Fashion-MNIST для класифікації та Oxford-IIIT Pet для сегментації; проектування та навчання відповідних моделей, оцінку якості результатів та візуалізацію роботи системи. Особливу увагу приділено інтеграції класифікаційної та сегментаційної моделей у єдиний конвеєр, що забезпечує послідовну обробку зображень у двох режимах.

ABSTRACT

Bachelor's thesis: 60 pages, 11 figures, 2 appendices, 7 sources.

IMAGE PROCESSING, CLASSIFICATION, SEMANTIC SEGMENTATION, MACHINE LEARNING, DEEP LEARNING, CONVOLUTIONAL NEURAL NETWORK, CNN, U-NET, TENSORFLOW, GOOGLE COLAB, PYTHON, COMPUTER VISION, INTELLIGENT SYSTEM, RESULT VISUALIZATION.

The major goal of this thesis is the development and implementation of a computer system capable of performing image classification and segmentation based on machine learning methods, particularly deep convolutional neural networks, using open platforms such as Google Colab.

In order to a comprehensive review of current scientific approaches to visual data processing was conducted. The analysis covered both traditional and deep learning methods, with special emphasis on the advantages of hybrid architecture. The system was implemented using the Google Colab environment, which provides access to GPU resources and modern libraries such as TensorFlow and Keras.

The practical part of the project includes the preparation of two datasets: Fashion-MNIST for classification tasks and Oxford-IIIT Pet for segmentation purposes; the design and training of appropriate models; the evaluation of performance metrics; and visualization of the system's outputs. Particular attention was paid to the integration of classification and segmentation models into a unified pipeline, ensuring a consistent two-stage image processing workflow.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ	8
ВСТУП	9
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	11
1.1 Основи обробки зображень	11
1.2 Методи машинного навчання для класифікації зображень	13
1.3 Методи глибокого навчання: CNN, U-Net, SegNet	16
1.4 Порівняння традиційних та глибоких методів сегментації	18
1.5 Огляд публікацій і сучасних підходів до задач класифікації та сегментації	22
2 ВИБІР ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ТА СЕРЕДОВИЩА РОЗРОБКИ	25
2.1 Обґрунтування використання Google Colab	25
2.2 Бібліотеки TensorFlow, Keras, OpenCV, scikit-image	27
2.3 Інструменти візуалізації та обробки результатів	29
3 РЕАЛІЗАЦІЯ ПРОГРАМНИХ ЗАСОБІВ	32
3.1 Підготовка датасетів	32
3.2 Побудова моделі для класифікації зображень	32
3.3 Реалізація моделі сегментації зображень	34
3.4 Навчання моделей і обробка результатів	36
3.5 Візуалізація результатів класифікації і сегментації	39
3.6 Інтеграція класифікації та сегментації в єдину систему або до висновків по розділу	41
ВИСНОВКИ	43
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	45
ДОДАТОК А Графічний матеріал кваліфікаційної роботи	46
ДОДАТОК Б Програмний код	55
Б.1 Лістинг коду	55

Б.1.1 Імпорт бібліотек	55
Б.1.2 Побудова моделі для класифікації зображень	56
Б.1.3 Реалізація моделі сегментації зображень	56
Б.1.4 Навчання моделей і обробка результатів	58
Б.1.5 Візуалізація результатів класифікації і сегментації	58
Б.1.6 Інтеграція класифікації та сегментації в єдину систему	59

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

CNN – згорткова нейронна мережа

Colab – Google Colaboratory

CPU – центральний процесор

CIFAR-10 – набір зображень із 10 класами

DL – глибоке навчання

GPU – графічний процесор

MNIST – база рукописних цифр

OpenCV – бібліотека для обробки зображень і відео

ReLU – функція активації

SGD – стохастичний градієнтний спуск

TFDS –модуль готових датасетів

U-Net – архітектура нейронної мережі для семантичної сегментації

VAL – валідаційна вибірка

YOLO – архітектура для об'єктного детектування

ВСТУП

У сучасному світі, що характеризується стрімким зростанням обсягів цифрових даних та широким розповсюдженням візуальної інформації, обробка зображень стала однією з ключових задач у сфері інформаційних технологій. Особливого значення набуває застосування методів машинного навчання, зокрема глибоких нейронних мереж, для автоматизації процесів аналізу зображень, що дозволяє значно підвищити ефективність обробки, точність класифікації об'єктів та сегментації областей інтересу.

Серед прикладних напрямів використання технологій комп'ютерного зору варто виокремити медичну діагностику, автономні транспортні системи, системи відеоспостереження, аграрні технології, а також промислову автоматизацію. В усіх цих сферах ключовою задачею є здатність комп'ютерної системи не лише розпізнати об'єкти на зображенні, а й визначити їхні просторові межі, тобто здійснити сегментацію. Таким чином, поєднання задач класифікації та сегментації формує базис для побудови повноцінних систем штучного інтелекту, здатних до високоточної інтерпретації візуальних даних.

Сучасні досягнення в галузі глибокого навчання, зокрема розробка згорткових нейронних мереж (CNN) для класифікації та архітектур U-Net, SegNet для сегментації, забезпечують принципово новий рівень точності обробки зображень. Такі моделі продемонстрували високу ефективність у багатьох конкурсах і практичних застосуваннях, що зумовлює їх доцільність для впровадження у різноманітних галузях.

Розробка та реалізація подібних систем потребує ефективного програмного середовища, здатного забезпечити інтеграцію компонентів машинного навчання, обробки зображень та візуалізації результатів. У цьому контексті хмарна платформа Google Colab є зручним та потужним інструментом, що дозволяє проводити експерименти з використанням GPU-

акселерації, бібліотек TensorFlow та Keras, а також інтегрувати аналітичні дослідження з відтворюваним програмним кодом.

Обрана тема є актуальною не лише з наукової точки зору, а й з погляду практичного застосування в умовах цифрової трансформації, де якісна обробка візуальної інформації є основою для автоматизованого прийняття рішень. Тому створення комп'ютерної системи, що об'єднує методи класифікації та сегментації зображень із використанням машинного навчання, становить значний інтерес як для науковців, так і для практиків.

Метою кваліфікаційної роботи є розробка та реалізація комп'ютерної системи, здатної виконувати класифікацію та сегментацію зображень на основі методів машинного навчання, зокрема глибоких згорткових нейронних мереж, з використанням відкритих платформ, таких як Google Colab.

Завдання:

- провести огляд сучасних підходів до класифікації та сегментації зображень із використанням машинного навчання;
- обґрунтувати вибір середовища розробки та інструментів, зокрема Google Colab, TensorFlow та Keras;
- реалізувати модель класифікації зображень на базі згорткової нейронної мережі;
- реалізувати модель сегментації зображень із використанням архітектури U-Net;
- провести навчання моделей на відкритих датасетах та проаналізувати отримані результати;
- оцінити точність роботи системи, здійснити її візуальне тестування та вивести ефективні метрики.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Основи обробки зображень

Обробка зображень як окрема науково-технічна дисципліна є складовою комп'ютерного зору і охоплює сукупність методів та алгоритмів, спрямованих на покращення візуального представлення даних, вилучення інформації з зображень та підготовку їх до подальшого аналізу. Зображення у цифровій формі розглядається як двовимірна дискретна функція, де кожному пікселю відповідає певне числове значення, що описує інтенсивність яскравості або колір. Виходячи з цього, цифрове зображення може бути представлено у вигляді матриці піксельних значень, над якою здійснюються обчислювальні операції.

На початковому етапі цифрової обробки часто виконується передобробка зображень, мета якої полягає у зменшенні шумів, нормалізації освітлення, підвищенні контрасту та фільтрації небажаних артефактів. Цей етап є критично важливим, оскільки якість вхідних даних істотно впливає на ефективність подальших операцій розпізнавання або класифікації. У процесі передобробки застосовуються як прості операції зображення, такі як лінійне масштабування яскравості чи гістограмне вирівнювання, так і складні алгоритми згортки з використанням різноманітних фільтрів.

Наступним важливим аспектом є виявлення ключових ознак або характеристик зображення, що дозволяють виокремити значущі елементи для подальшого аналізу. До таких характеристик належать геометричні форми, контури, текстури, кольорові компоненти, топологічні властивості, а також статистичні розподіли піксельних значень. Методи виявлення ознак можуть бути як локальними, що фокусуються на окремих фрагментах зображення, так і глобальними, що відображають загальні властивості всього зображення. Ці ознаки формують вектор ознак – математичне представлення візуальної

інформації у вигляді числового опису, придатного для введення в алгоритми машинного навчання.

У сучасних системах все частіше застосовуються автоматизовані підходи до вилучення ознак, які не потребують ручної інженерії ознак. Цей підхід є характерним для методів глибокого навчання, зокрема згорткових нейронних мереж, які здатні самостійно навчатися відповідним ознаковим представленням шляхом обробки великої кількості зображень. Проте навіть при використанні таких моделей розуміння класичних принципів обробки зображень залишається необхідним для коректного проектування архітектури мережі, підготовки вхідних даних, інтерпретації результатів і оптимізації моделі.

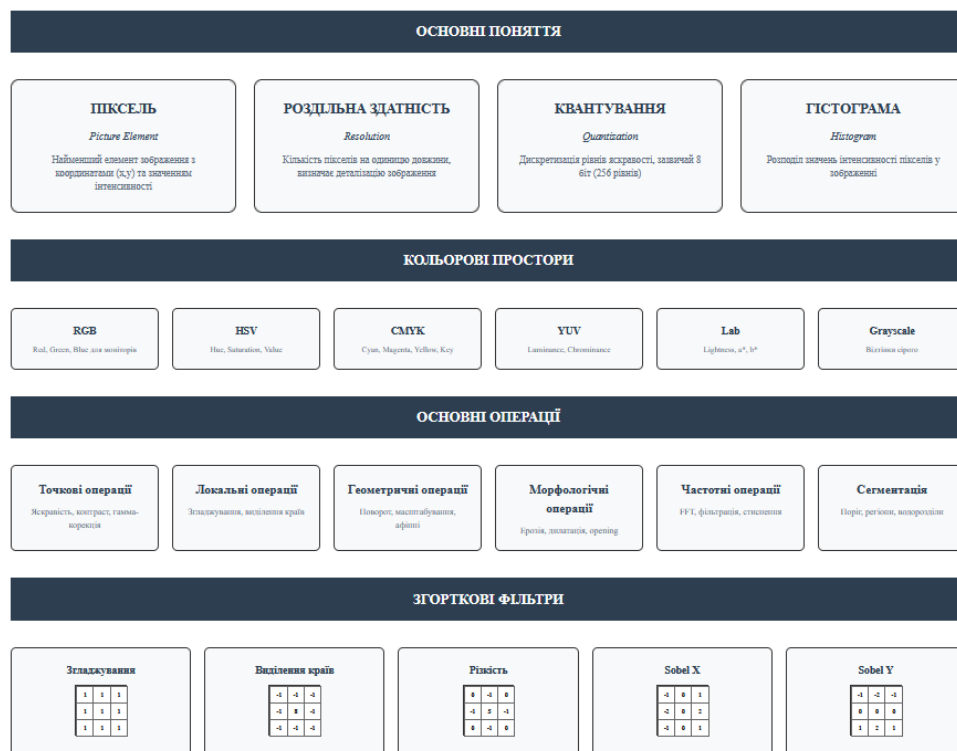


Рисунок 1.1 – Основи обробки зображень

Ще одним концептуальним напрямом є сегментація зображень, що передбачає поділ зображення на однорідні області, кожна з яких відповідає певному об'єкту, фону або функціонально значущій частині сцени. На відміну від класифікації, яка надає одне узагальнене рішення для всього

зображення, сегментація дозволяє аналізувати його на піксельному або об'єктному рівні, що забезпечує глибше розуміння структури сцени та підвищує точність і релевантність систем розпізнавання.

Важливе місце в обробці зображень займає також трансформація простору зображення. Часто для покращення розпізнавання або спрощення обчислень зображення перетворюються до іншого представлення – як, наприклад, перехід до колірному простору HSV, виділення країв за допомогою оператора Собеля або застосування перетворення Фур'є. Такі операції дозволяють виділити ті характеристики, які найменш чутливі до змін освітлення, масштабування або позиції об'єкта в кадрі.

Загалом, обробка зображень виступає як основоположна технологія, яка забезпечує підготовку та уніфікацію вхідних даних для задач комп'ютерного зору, створення векторного простору ознак, формування високорівневих представлень візуальної інформації та оптимізацію процесів навчання моделей. Саме тому розуміння її принципів є необхідним етапом для побудови ефективних систем класифікації та сегментації на основі машинного навчання.

1.2 Методи машинного навчання для класифікації зображень

Класифікація зображень є однією з найважливіших задач комп'ютерного зору, суть якої полягає у визначенні класу або категорії, до якого належить зображення на основі його візуального змісту. В основі цієї задачі лежить процес виявлення закономірностей між вхідними даними у вигляді зображень та відповідними мітками класів. Методи машинного навчання забезпечують побудову моделей, які здатні автоматично навчатися таким залежностям, використовуючи наявний набір розмічених прикладів.

На початкових етапах розвитку машинного навчання для класифікації зображень застосовувались здебільшого алгоритми традиційного (немережевого) машинного навчання, для яких ключову роль відіграла

інженерія ознак. Зображення перетворювалися на вектори ознак за допомогою детекторів контурів, гістограм орієнтованих градієнтів, методів вилучення текстур або кольорових характеристик. Після побудови векторного представлення застосовувались класифікаційні моделі, такі як метод опорних векторів, дерева рішень, наївні байєсівські класифікатори або методи k-ближчих сусідів. Незважаючи на свою обчислювальну ефективність і відносну простоту реалізації, ці підходи мали суттєве обмеження в узагальненні на складні та високорозмірні візуальні дані.

Суттєвий прорив у класифікації зображень став можливим із впровадженням глибоких нейронних мереж, зокрема згорткових нейронних мереж (CNN). Вони забезпечили автоматичне вилучення ієрархічно структурованих ознак без необхідності попереднього проектування дескрипторів. Основною перевагою CNN є їхня здатність формувати багаторівневі представлення зображень, починаючи від простих фільтрів країв і завершуючи високорівневими абстракціями, релевантними до цільового класу. Це дозволяє мережі навчатись на складних вибірках, зберігаючи чутливість до важливих візуальних патернів та ігноруючи другорядні деталі.

Процес навчання таких моделей здійснюється на основі великих датасетів, де кожне зображення має відповідну мітку класу. У процесі оптимізації вагових коефіцієнтів мережа мінімізує функцію втрат, яка відображає розбіжність між передбаченими результатами та справжніми мітками. Найчастіше для задач класифікації застосовується функція крос-ентропії, яка забезпечує ефективне навчання в багатокласових умовах. Навчання реалізується шляхом зворотного поширення помилки з використанням методів стохастичного градієнтного спуску та його вдосконалень, таких як Adam або RMSprop.

Архітектурні рішення в класифікаційних моделях постійно вдосконалюються. Історично значущими стали такі архітектури, як LeNet, AlexNet, VGG, ResNet, які заклали основи для подальшого розвитку глибоких

моделей. Зокрема, ResNet вирішила проблему зникнення градієнта в дуже глибоких мережах, що дало змогу будувати надзвичайно глибокі моделі без втрати якості навчання. Також відзначається поступове зростання ролі трансформерів у задачах візуальної класифікації, що демонструють конкурентні результати порівняно з класичними CNN.

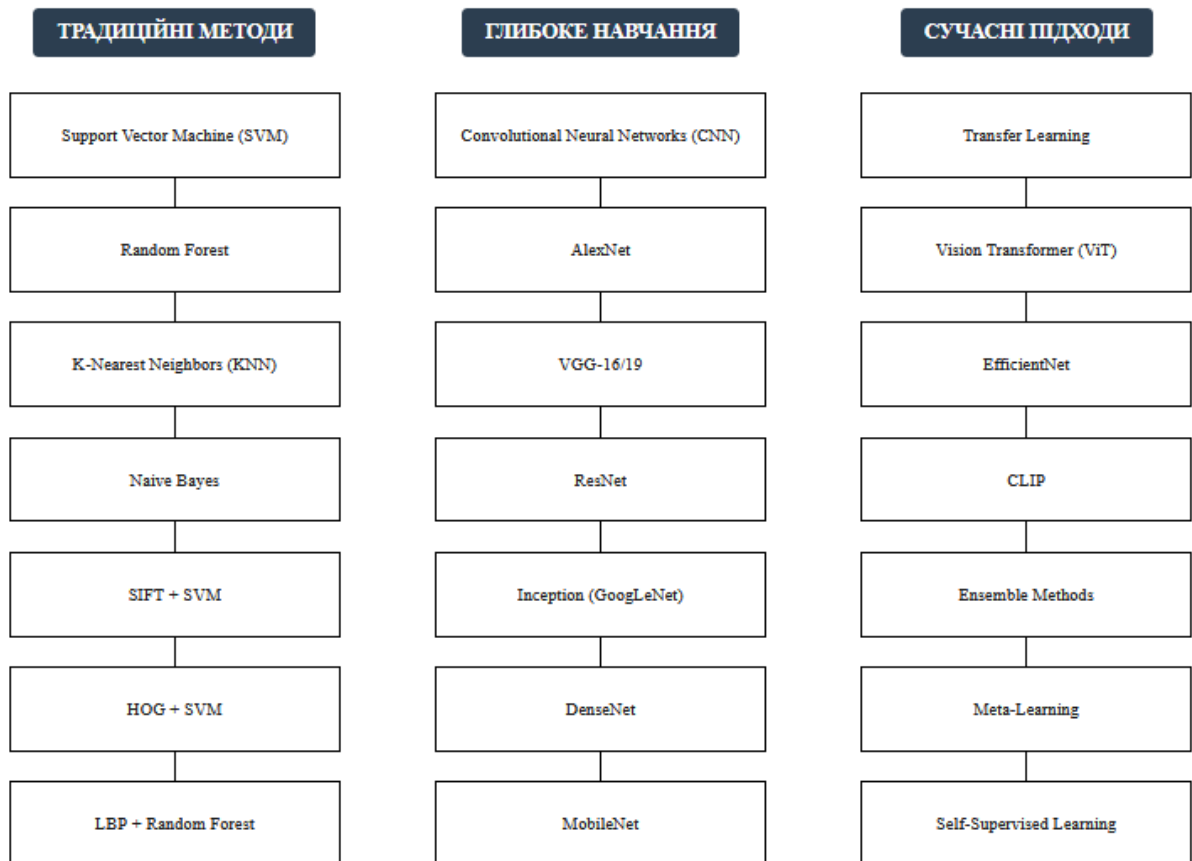


Рисунок 1.2 – Методи машинного навчання для класифікації зображень

Окремої уваги заслуговує процес оцінювання якості класифікаційних моделей. Основними метриками є точність (accuracy), повнота (recall), точність позитивного прогнозу (precision), F1-міра, а також аналіз матриці неточностей, яка дозволяє виявити специфіку помилок моделі для кожного класу. У випадках, коли існує суттєвий класовий дисбаланс, важливо використовувати стратифіковані вибірки, вагове коригування функції втрат або методи аугментації даних. Для ресурсозалежних застосунків, орієнтованих на швидку обробку даних.

1.3 Методи глибокого навчання: CNN, U-Net, SegNet

Глибоке навчання посідає ключове місце в сучасних системах обробки зображень, забезпечуючи принципово новий рівень автоматизації, узагальнення та точності в задачах класифікації, сегментації, виявлення об'єктів та генерації візуального контенту. Його ефективність зумовлена здатністю будувати багаторівневі ієрархії ознак без потреби в ручному конструюванні дескрипторів. Серед широкого спектра архітектур глибокого навчання особливу роль відіграють згорткові нейронні мережі (CNN), а також їхні модифікації, орієнтовані на сегментацію, зокрема U-Net і SegNet.

Згорткові нейронні мережі (CNN) стали основою сучасного глибокого аналізу зображень. Їхня структура побудована за принципом локальної рецептивності, згорткових фільтрів, підвбірок (пулінгу) та повнозв'язних шарів, що дозволяє моделі виявляти шаблони, інваріантні до зсувів, масштабів і деформацій. Згорткові шари автоматично вивчають фільтри, які реагують на елементарні візуальні патерни, такі як краї, текстури або геометричні структури, тоді як глибші шари формують все більш абстрактні представлення об'єктів. Основною перевагою CNN є їхня здатність узагальнювати закономірності на великих вибірках та зберігати просторову структуру вхідного сигналу.

Попри свою ефективність у задачах класифікації, класичні CNN не є придатними для повноцінної сегментації, оскільки повнозв'язні шари призводять до втрати просторової роздільності. Для розв'язання цього питання було запропоновано архітектури, орієнтовані на піксельну або семантичну сегментацію, які використовують симетричні структури з енкодером і декодером. Серед них особливою популярністю користується модель U-Net, яка була спочатку розроблена для біомедичної сегментації, але згодом набула широкого застосування в різних галузях.

U-Net поєднує ідеї згорткової архітектури з механізмами апсемплінгу, формуючи структуру у вигляді симетричного «U»-подібного графа. Енкодер

здійснює поступове зменшення роздільної здатності зображення з одночасним нарощуванням кількості фільтрів, у той час як декодер відновлює просторову структуру за допомогою транспонованих згорток. Унікальною особливістю архітектури є наявність пропускових зв'язків (skip connections), які з'єднують відповідні рівні екодера і декодера. Це дозволяє уникнути втрати детальної інформації та суттєво покращує якість сегментації, зокрема на границях об'єктів.

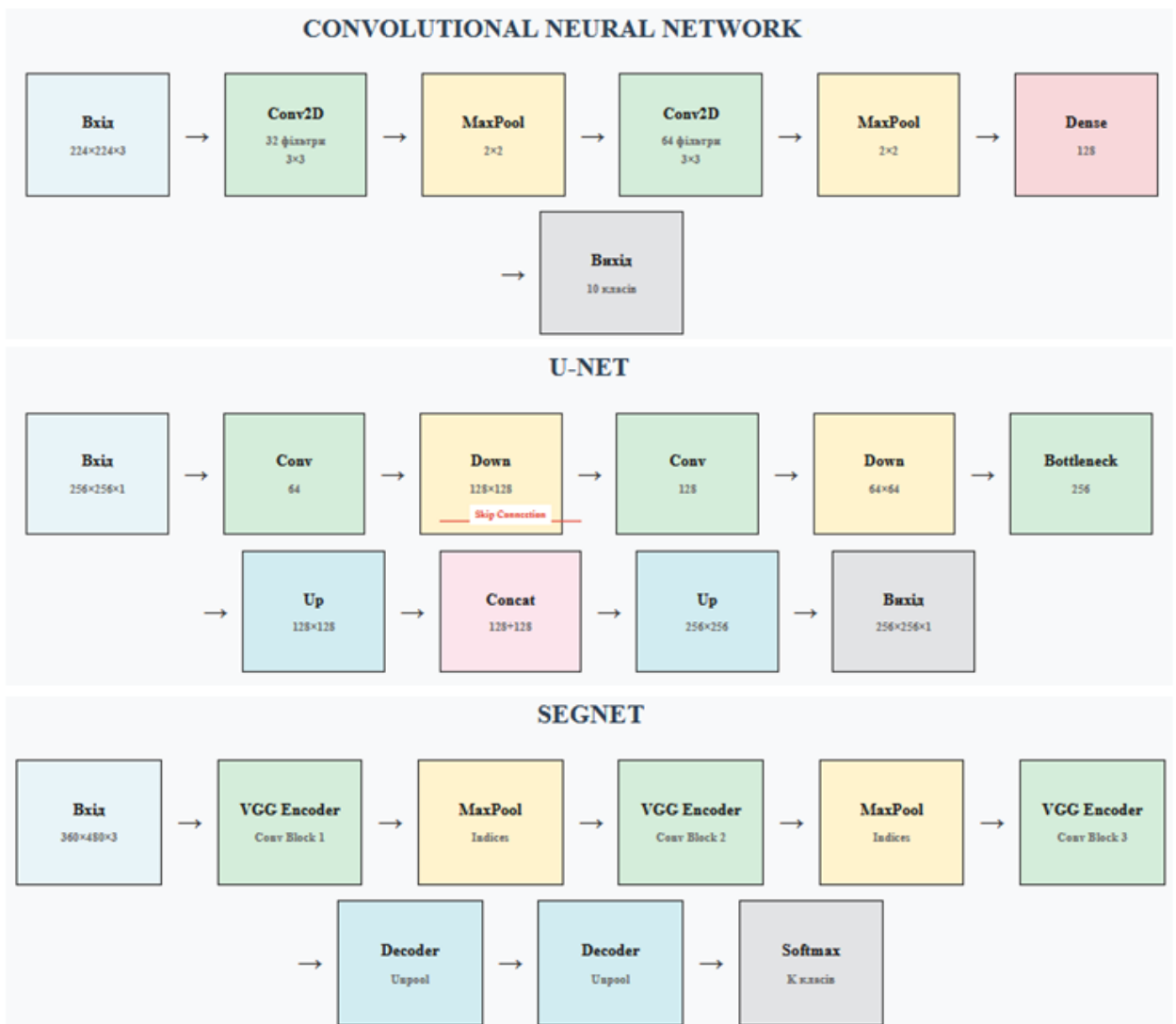


Рисунок 1.3 – CNN, U-Net та SegNet

Іншою популярною архітектурою для сегментації є SegNet, яка також базується на концепції симетричної структури, але реалізує деякі інші принципи відновлення зображення. Замість повного відновлення з

використанням транспонованих згорток, SegNet зберігає індекси максимумів під час пулінгу в енкодері і використовує їх у відповідному декодері для точної реконструкції просторового розташування об'єктів. Такий підхід забезпечує більшу ефективність у сенсі обчислювальної складності та дозволяє уникати розмиття, властивого деяким методам відновлення зображення.

Обидві архітектури – U-Net і SegNet – добре масштабуються до великих зображень і мають високу адаптивність до різних типів сегментаційних задач, включаючи бінарну, багатокласову та інстанс-сегментацію. Вони широко застосовуються не лише у медичних дослідженнях, а й в агрономії, аерофотозйомці, автономному водінні, цифровому картографуванні та промисловій візуальній діагностиці. За своєю суттю ці моделі ілюструють поступовий перехід від глобальної обробки зображення до локальної високоточної інтерпретації на рівні кожного пікселя.

Методи глибокого навчання для обробки зображень, зокрема CNN, U-Net та SegNet, демонструють значну точність, універсальність та здатність до перенавчання на нові домени. Вони суттєво підвищують рівень автоматизації аналізу зображень, а також дозволяють реалізовувати моделі з дедалі більшим рівнем складності, які здатні враховувати як просторову, так і семантичну інформацію. Їхня ефективність залежить від наявності якісно розмічених вибірок, правильного вибору архітектури, а також тонкого налаштування параметрів навчання та регуляризації.

1.4 Порівняння традиційних та глибоких методів сегментації

Сегментація зображень як фундаментальна задача комп'ютерного зору передбачає поділ вхідного зображення на структуровані області, що відповідають об'єктам або їхнім частинам. Упродовж десятиліть у цій сфері домінували класичні (традиційні) методи, які базувалися на евристичних

алгоритмах, геометричних ознаках, градієнтних характеристиках і просторово-статистичних моделях. Згодом, із розвитком методів глибокого навчання, традиційні підходи почали поступатися місцем автоматизованим моделям, здатним до складного навчання ієрархічних представлень даних. Порівняння цих двох парадигм виявляє не лише відмінності в алгоритмічному рівні, а й у концептуальному підході до вирішення задачі.

Традиційні методи сегментації здебільшого ґрунтуються на локальних або глобальних евристичних методах. Серед них ключову роль відіграють методи порогової обробки, сегментація на основі градієнта (методи виділення контурів), сегментація за кольором або текстурою, а також алгоритми кластеризації, такі як k-середніх або алгоритм Ватершада. В основі цих методів лежить припущення, що сусідні пікселі, які мають схожі характеристики, повинні належати до однієї області. Зазначені підходи є обчислювально простими, легко інтерпретуються, не потребують навчання та застосовуються у випадках з низькою складністю сцени або на етапах попередньої обробки.

Однак основною слабкістю традиційних методів є їхня обмежена здатність до узагальнення та низька стійкість до варіативності зображень. Вони чутливі до змін освітлення, шумів, варіацій форми об'єктів та складної текстури фону. Крім того, такі методи потребують ручного налаштування параметрів або введення апріорної інформації, що обмежує їхню придатність для широкомасштабних автоматизованих систем.

Натомість методи глибокого навчання, які набули значного поширення в останнє десятиліття, забезпечують принципово новий рівень точності та адаптивності. Глибокі нейронні мережі, зокрема архітектури U-Net, SegNet, DeepLab та інші, дозволяють моделі самостійно навчатися релевантним ознакам без потреби в ручному втручанні. Їхня багаторівнева структура забезпечує виявлення як низькорівневих характеристик (наприклад, краї), так і високорівневих семантичних ознак (наприклад, клас об'єкта). Завдяки цьому такі моделі можуть точно сегментувати об'єкти складної форми,

відокремлювати об'єкти на складному фоні та забезпечувати узагальнення на нових даних.



Рисунок 1.4 – Традиційні методи сегментації

На відміну від традиційних методів, глибокі підходи є обчислювально інтенсивними, потребують значних обсягів розмічених даних для навчання, GPU-прискорення та тонкого налаштування гіперпараметрів. Незважаючи на це, результати, яких вони дозволяють досягти, значно перевершують традиційні підходи за всіма основними метриками якості, включаючи точність сегментації, повноту, F1-міру та індекс перекриття (IoU). Крім того,

глибокі моделі здатні до трансферного навчання, що дозволяє адаптувати їх до нових предметних областей без потреби у повному перенавчанні з нуля.



Рисунок 1.5 – Глибокі методи сегментації

У прикладних задачах глибокі методи довели свою ефективність у таких складних сферах, як медична візуалізація, супутникова аналітика, розпізнавання дорожніх сцен і промисловий контроль якості. При цьому традиційні методи залишаються актуальними у випадках, коли системі потрібна висока швидкість обробки зображень у режимі реального часу, мінімальні обчислювальні ресурси або пояснюваність на рівні простих алгоритмів.

Загалом порівняння традиційних і глибоких методів сегментації свідчить про глибокий перехід у парадигмі обробки зображень – від ручного

налаштування алгоритмів до навчання моделей, які здатні самостійно виділяти закономірності у складному візуальному середовищі. Вибір підходу в кожному конкретному випадку має ґрунтуватися на аналізі вимог до точності, швидкості, обчислювальних ресурсів і наявності навчальних даних.

Порівняльний аналіз традиційних та глибоких методів сегментації зображень засвідчив принципову різницю як у концептуальних підходах, так і в прикладних характеристиках. Традиційні методи залишаються релевантними в контексті задач із низькою складністю, обмеженими ресурсами або суворими вимогами до інтерпретованості алгоритмів. Вони добре підходять для швидкої обробки в режимі реального часу та вбудованих систем із низькою обчислювальною потужністю.

Натомість глибокі методи, зокрема ті, що базуються на архітектурах CNN, U-Net і SegNet, демонструють значно вищу точність, адаптивність до складних сценаріїв і здатність до узагальнення на нові вибірки. Їх використання є доцільним у випадках, коли пріоритетами є якість сегментації, стійкість до візуальних артефактів та варіативності сцени, а також можливість подальшого масштабування.

У сучасних практиках спостерігається тенденція до комбінування сильних сторін обох підходів – застосування класичних алгоритмів для попередньої обробки або аугментації даних, а глибоких моделей – для фінального семантичного аналізу. Таким чином, вибір оптимального підходу має ґрунтуватися на балансі між вимогами до точності, пояснюваністю, ресурсами та специфікою предметної області.

1.5 Огляд публікацій і сучасних підходів до задач класифікації та сегментації

Упродовж останнього десятиліття дослідження в галузі комп'ютерного зору, зокрема задач класифікації та сегментації зображень, стали однією з найбільш динамічно зростаючих сфер штучного інтелекту. Це зумовлено

стрімким розвитком глибоких нейронних мереж, збільшенням обсягів доступних даних і зростанням обчислювальних потужностей. Значна частина наукових публікацій демонструє тенденцію до інтеграції глибокого навчання в усі етапи обробки зображень, що суттєво підвищує точність, масштабованість та адаптивність відповідних систем.

Першим значним проривом у класифікації зображень стала публікація [1] про модель AlexNet, яка здобула перемогу в конкурсі ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Вона започаткувала нову епоху використання згорткових нейронних мереж у задачах комп'ютерного зору, продемонструвавши перевагу автоматизованого вилучення ознак над традиційними дескрипторами. Подальші роботи, [2, 3], розширили глибину архітектур і забезпечили ще вищу точність класифікації, зробивши їх стандартом для багатьох прикладних задач.

Сегментація зображень, як більш складна задача, що вимагає піксельного або об'єктного розділення сцени, отримала новий імпульс завдяки архітектурам U-Net [4] та [5]. U-Net було вперше запропоновано для медичної візуалізації, але вона швидко поширилася у сфері біоінформатики, супутникового моніторингу, агротехнологій і розпізнавання дорожніх сцен. Завдяки симетричній структурі енкодер-декодер і використанню пропускних зв'язків, модель досягла високої точності навіть при обмеженій кількості навчальних зразків. SegNet, зі свого боку, зробила акцент на економії пам'яті та ефективному відновленні просторової структури зображення, що зробило її придатною для вбудованих систем і мобільного аналізу.

Окремий напрям сучасних досліджень пов'язаний із застосуванням архітектур трансформерів у задачах обробки зображень. Робота [6] запропонувала Vision Transformer (ViT) – модель, яка замінює згортки на механізм самоуваги, що успішно демонструє конкурентну точність у задачах класифікації. Подальші розробки, зокрема Segmenter, SETR та Swin Transformer, адаптували цю парадигму для сегментаційних завдань, що стало потужною альтернативою класичним CNN-архітектурам.

Серед практичних платформ значну роль відіграють відкриті датасети, які використовуються як еталон для порівняння підходів. Наприклад, ImageNet застосовується для класифікації, Cityscapes, COCO, PASCAL VOC – для семантичної сегментації, а LUNA16 або BraTS – для медичних зображень. Дослідження, що базуються на цих наборах, дозволяють формувати стандартизовані метрики, такі як точність (Accuracy), індекс Джаккара (IoU), F1-міра або середньозважена пер-піксельна похибка, що сприяє об'єктивному аналізу моделей.

Водночас, у сучасних публікаціях посилено розглядаються проблеми, пов'язані з недостатністю даних, класовим дисбалансом, нечіткими межами об'єктів та необхідністю пояснюваності моделей. Пропонуються гібридні підходи, зокрема використання генеративних моделей (GAN) для синтезу додаткових даних, weakly-supervised learning – для зменшення потреби в розмітці, а також метамоделі для інтеграції доменних знань у процес прийняття рішень.

2 ВИБІР ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ТА СЕРЕДОВИЩА РОЗРОБКИ

2.1 Обґрунтування використання Google Colab

У процесі реалізації систем комп'ютерного зору, особливо таких, що базуються на глибокому навчанні, ключову роль відіграє вибір середовища розробки, яке має забезпечити як функціональну гнучкість, так і доступ до ресурсів високої продуктивності. Серед різноманітних інструментів, що застосовуються для досліджень і практичної реалізації моделей машинного навчання, особливої популярності набуло середовище Google Colaboratory (або Google Colab), яке поєднує інтерфейс Jupyter-ноутбуків із хмарною обчислювальною інфраструктурою.

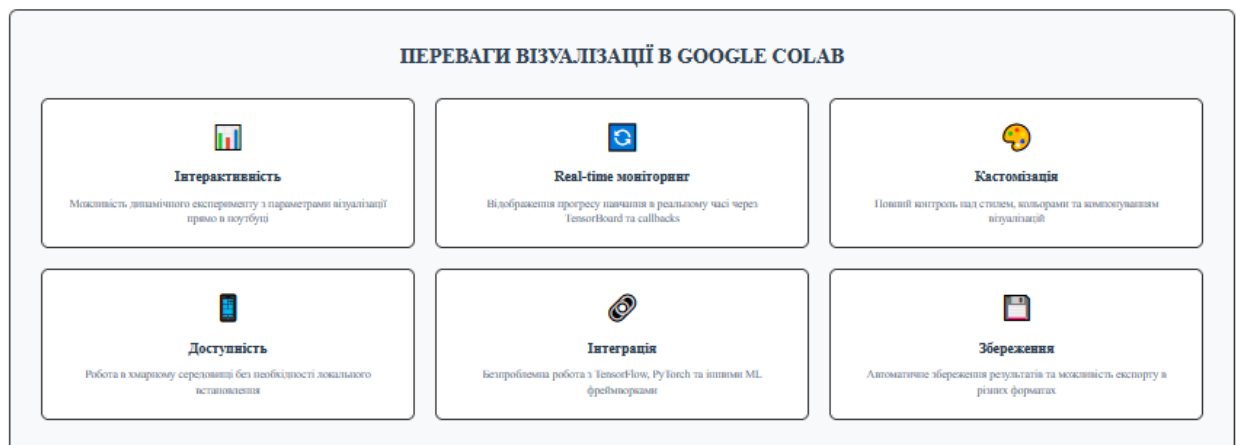


Рисунок 2.1 – Переваги Google Colab

Однією з ключових переваг Google Colab є можливість безкоштовного використання графічних процесорів (GPU) та тензорних процесорів (TPU), що є критично важливим для прискорення навчання глибоких нейронних мереж, зокрема згорткових моделей. У задачах, де обробка зображень вимагає багатьох епох навчання, великих обсягів даних та складної архітектури моделей, наявність апаратного прискорення значно скорочує час виконання та розширює можливості експериментування.

Google Colab надає повну підтримку мовою програмування Python, яка є домінуючою у сфері машинного навчання завдяки своїй простоті, великій кількості бібліотек і активній спільноті розробників. Інтеграція з такими популярними бібліотеками, як TensorFlow, Keras, PyTorch, OpenCV, NumPy та Matplotlib, дозволяє здійснювати повний цикл побудови моделі: від імпорту та передобробки зображень до навчання, оцінювання та візуалізації результатів. Крім того, доступ до зовнішніх ресурсів, таких як Google Drive, забезпечує зручне збереження і повторне використання даних, що актуально при роботі з великими датасетами або довготривалими проєктами.

Варто підкреслити, що Google Colab підтримує інтерактивне середовище з можливістю коментування, поєднання текстових пояснень і програмного коду, що робить його надзвичайно зручним як для самостійного дослідження, так і для командної співпраці. Зокрема, у рамках освітніх програм або наукових досліджень, такий формат дозволяє поєднати опис теоретичних засад із практичною реалізацією методів, що значно підвищує якість навчального процесу та прозорість дослідницької роботи.

Іншою важливою перевагою є відсутність потреби в локальній інсталяції складного програмного забезпечення, драйверів чи бібліотек, що зменшує бар'єр входу для дослідників, які не мають доступу до потужної обчислювальної техніки. Хмарна природа Google Colab також забезпечує доступ до середовища з будь-якої точки світу, з будь-якого пристрою, що особливо актуально в умовах дистанційного навчання, хакатонів або колективних досліджень.

Таким чином, вибір Google Colab для реалізації комп'ютерної системи класифікації та сегментації зображень є повністю обґрунтованим з огляду на його безкоштовність, доступ до апаратного прискорення, інтеграцію з сучасними бібліотеками машинного навчання, простоту у використанні та зручність документування дослідницького процесу. Це середовище надає усі необхідні інструменти для реалізації проєкту у межах кваліфікаційної роботи, одночасно забезпечуючи відповідність сучасним стандартам

програмної інженерії та штучного інтелекту.

2.2 Бібліотеки TensorFlow, Keras, OpenCV, scikit-image

Для реалізації повноцінної системи класифікації та сегментації зображень у середовищі Google Colab необхідно застосовувати сучасні програмні бібліотеки, які забезпечують широкий набір функцій для роботи з нейронними мережами, обробкою зображень, візуалізацією та математичним аналізом даних. Серед них ключову роль відіграють TensorFlow, Keras, OpenCV та scikit-image, які становлять фундаментальний інструментарій для розв'язання задач у сфері комп'ютерного зору на основі глибокого навчання.



Рисунко 2.2 – Вибір програмних засобів

TensorFlow є фреймворком з відкритим кодом, розробленим компанією Google, що дозволяє створювати і навчати глибокі нейронні мережі різної складності. Ця бібліотека підтримує як низькорівневу побудову обчислювальних графів, так і високорівневі API, зокрема Keras, що значно спрощує процес створення моделей. TensorFlow забезпечує апаратну прискорюваність за допомогою GPU та TPU, що особливо важливо для прискорення обчислень під час навчання великих моделей сегментації, таких як U-Net чи SegNet. Крім того, TensorFlow має зручні засоби для збереження, експорту та розгортання моделей у вигляді сервісів, що дозволяє

реалізовувати концепції повного життєвого циклу моделей машинного навчання.

Keras є високорівневою обгорткою над TensorFlow, яка надає інтуїтивно зрозумілий інтерфейс для побудови моделей. Завдяки своїй модульній архітектурі, Keras дозволяє швидко створювати згорткові нейронні мережі, задавати функції активації, метрики, оптимізатори та шари без необхідності занурюватись у низькорівневу реалізацію. У контексті кваліфікаційної роботи використання Keras дозволяє створити гнучку, зрозумілу та ефективну модель класифікації або сегментації з мінімальними зусиллями. Вона також забезпечує сумісність із багатьма функціональними компонентами, такими як колбеки для моніторингу навчання, засоби візуалізації, аугментація даних тощо.

OpenCV (Open Source Computer Vision Library) є однією з найпоширеніших бібліотек для комп'ютерного зору та цифрової обробки зображень. Вона надає широкі можливості для базової обробки зображень: зчитування, перетворення кольорових просторів, фільтрація, зміна розмірів, геометричні трансформації, виявлення контурів та гістограм. У контексті даної роботи OpenCV використовується на етапі передобробки зображень, підготовки даних для навчання нейронних мереж, а також візуалізації результатів класифікації та сегментації. Її ефективність, швидкість роботи та багатоплатформеність роблять її невіддільною частиною програмного стеку у сфері комп'ютерного зору.

scikit-image є спеціалізованим модулем Python для обробки зображень, який базується на бібліотеках NumPy та SciPy. Вона містить набір функцій для морфологічної обробки, сегментації, виявлення країв, аналізу регіонів, а також обчислення статистичних характеристик зображень. У порівнянні з OpenCV, scikit-image орієнтована більше на наукові та дослідницькі задачі, має чітку структуру та забезпечує точність реалізації відповідно до сучасних алгоритмів обробки. У межах даної кваліфікаційної роботи вона може бути застосована для експериментального порівняння різних способів сегментації

або для побудови карт схожості між зображеннями.

Інтеграція цих бібліотек у межах одного середовища – Google Colab – забезпечує ефективну, масштабовану та відкриту платформу для створення, навчання, тестування і візуалізації результатів глибоких нейронних мереж. Вони дозволяють реалізовувати повний цикл машинного навчання, що включає зчитування зображень, їх передобробку, побудову архітектури моделей, тренування на датасетах, оцінювання метрик якості та формування висновків щодо ефективності алгоритмів.

2.3 Інструменти візуалізації та обробки результатів

Візуалізація результатів є ключовим компонентом процесу розробки й аналізу систем машинного навчання, особливо у задачах класифікації та сегментації зображень, де якість моделі повинна оцінюватися не лише за числовими метриками, але й за наочністю отриманих результатів. Глибокі нейронні мережі часто оперують великою кількістю параметрів і складною внутрішньою структурою, тому візуальне представлення вхідних і вихідних даних, а також проміжних результатів навчання, дозволяє краще зрозуміти функціонування моделі, виявити потенційні помилки та оптимізувати параметри.

У середовищі Google Colab реалізація візуалізації здійснюється за допомогою потужних Python-бібліотек, серед яких найбільш поширеними є Matplotlib, Seaborn, Plotly, а також функціональні компоненти самих бібліотек TensorFlow/Keras. Ці інструменти забезпечують побудову як стандартних графіків навчання (залежність втрат і точності від кількості епох), так і візуалізацію зображень до та після обробки, зокрема накладення сегментованих масок на оригінальні зображення, відображення меж об'єктів або карт активацій.

Matplotlib є базовою бібліотекою для побудови статичних графіків і зображень, яка широко використовується для виводу вхідних зображень,

результатів класифікації та порівняння етапів обробки. Вона дозволяє відображати зображення в сірій шкалі, кольоровому просторі RGB, а також здійснювати накладання різних масок або контурів, що є надзвичайно важливим у задачах сегментації.

Seaborn, як розширення Matplotlib, орієнтований на побудову інформативних статистичних візуалізацій, таких як теплові карти (heatmaps), гістограми, розподіли ознак і матриці плутанини. Його застосування дозволяє краще зрозуміти розподіл даних, зокрема вхідних класів, ступінь дисбалансу, варіації інтенсивності або колірних характеристик, що впливають на якість класифікації.

Plotly забезпечує інтерактивну візуалізацію даних у форматі HTML-графіків. Хоча вона менш поширена в базових проєктах, її можливості є цінними для демонстрацій і презентацій – зокрема при побудові 3D-графіків або інтерактивних сегментаційних результатів із повзунками для перегляду шарів зображення.

У контексті глибокого навчання важливу роль відіграє також візуалізація процесу навчання нейронної мережі. Завдяки вбудованим функціям Keras, такими як `model.fit()` з параметрами `history`, можливо виводити графіки зміни функції втрат і точності протягом епох, що дозволяє оцінити ефективність навчання, виявити проблему переобучення або недонавчання моделі. Також у TensorBoard – спеціальному модулі для TensorFlow – реалізовано розширену візуалізацію нейронних мереж, у тому числі граф обчислень, гістограми ваг і карт активацій, що дозволяє аналізувати внутрішні процеси в моделі.

У задачах сегментації особливої уваги потребує візуалізація отриманих масок: правильність розмітки, точність меж об'єктів, ступінь накладання на реальні зони інтересу. Для цього застосовується комбінація `imshow()` з можливістю накладання прозорої маски (`alpha blending`) або контуровання (`contour`). Це дозволяє візуально оцінити якість сегментації без необхідності занурення у числові метрики.

З метою порівняння моделей використовуються матриці неточностей (confusion matrices), що дозволяють оцінити частоту правильних і хибних класифікацій по класах. Вони особливо корисні у випадках багатокласової класифікації з нерівномірним розподілом об'єктів.

Загалом, візуалізація результатів є не лише інструментом для представлення даних, але й аналітичним засобом, що сприяє глибшому розумінню поведінки моделі, оптимізації параметрів і підвищенню інтерпретованості результатів. Її поєднання з числовими метриками формує комплексну основу для оцінювання ефективності систем глибокого навчання в задачах класифікації та сегментації зображень.

3 РЕАЛІЗАЦІЯ ПРОГРАМНИХ ЗАСОБІВ

3.1 Підготовка датасетів

У рамках реалізації комп'ютерної системи класифікації та сегментації зображень було обрано два різні датасети, кожен з яких відповідає специфічній задачі. Для класифікації зображень було застосовано набір даних Fashion-MNIST, що містить зображення предметів одягу у відтінках сірого. Він характеризується збалансованістю класів та доступністю для завантаження безпосередньо з бібліотеки TensorFlow. У свою чергу, для задачі семантичної сегментації використано датасет Oxford-III Pet, який містить кольорові зображення тварин із відповідними піксельними масками.

Підготовка датасету Fashion-MNIST передбачає нормалізацію значень пікселів для прискорення навчання та приведення до одноканального формату, сумісного зі згортковими мережами. Для набору Oxford Pets проводиться масштабування зображень до фіксованого розміру, нормалізація інтенсивностей і перетворення багатокласових масок у бінарні, що відповідають наявності чи відсутності об'єкта на зображенні. Такий підхід дозволяє забезпечити уніфікований вхідний формат для подальшого навчання моделей глибокого навчання.

3.2 Побудова моделі для класифікації зображень

У межах реалізації комп'ютерної системи для класифікації візуальних об'єктів було спроектовано згорткову нейронну мережу (CNN), що дозволяє ефективно розв'язувати задачу автоматичного розпізнавання класів зображень на основі їхніх просторових ознак. Архітектура такої мережі базується на фундаментальних принципах локального вилучення ознак, інваріантності до трансформацій та ієрархічної побудови ознакових

просторових уявлень.

Побудована модель має послідовну структуру, яка складається з двох згорткових блоків, кожен з яких включає згортковий шар із фільтрами малого розміру та шар підвибірки (пулінгу), що виконує просторову агрегацію інформації. У першому згортковому шарі використовуються 32 фільтри розміром 3×3 , які активуються функцією ReLU і реагують на локальні контурні та текстурні елементи. Наступний шар MaxPooling із розміром вікна 2×2 дозволяє зменшити просторову роздільну здатність, що забезпечує інваріантність до незначних зсувів та масштабів. Аналогічний принцип реалізовано в другому згортковому блоці, де кількість фільтрів збільшується до 64, що дає змогу виявляти складніші патерни на основі попередньо вивчених ознак.

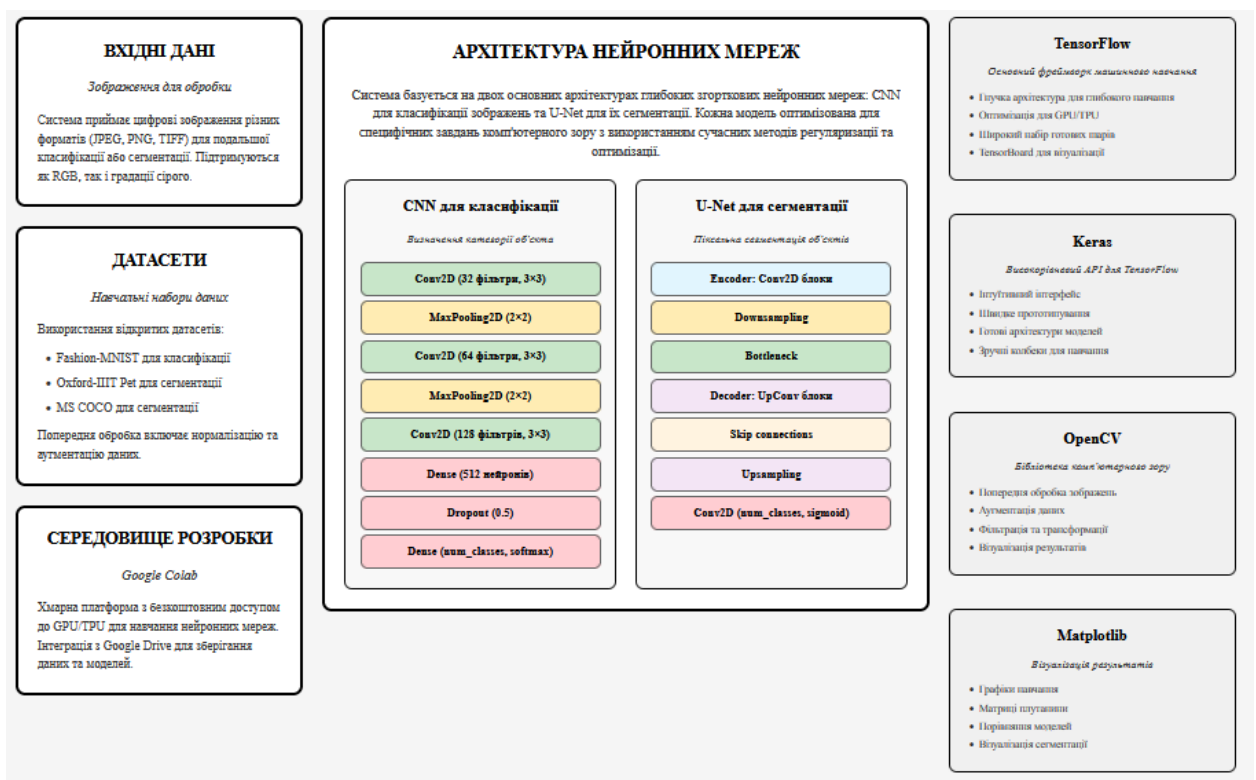


Рисунок 3.1 – Реалізація комп'ютерної системи

Після вилучення ознак застосовується шар Flatten, який трансформує тривимірний тензор ознак у вектор для подальшої обробки в повнозв'язних шарах. Завершальний блок моделі складається з одного щільного шару з 64

нейронами та функцією активації ReLU, що виконує нелінійне відображення ознакового простору, а також фінального вихідного шару з 10 нейронами та активацією softmax, яка формує ймовірнісний розподіл по класах.

Для навчання мережі було застосовано оптимізатор Adam, який поєднує переваги методів адаптивного градієнта та моменту, забезпечуючи стабільну та швидку збіжність. Як функцію втрат було обрано категоріальну крос-ентропію, яка є найбільш поширеною для багатокласових задач класифікації. Навчання проводилось упродовж 10 епох із валідацією на тестовій вибірці, що дозволяє відслідковувати здатність моделі до узагальнення та своєчасно виявляти симптоми перенавчання.

Результати навчання відображаються у вигляді графіка зміни точності на тренувальних і тестових даних, що надає наочну інформацію про динаміку процесу оптимізації. Показано, що модель досягає високих показників точності вже після кількох епох, що свідчить про її ефективність у розв'язанні задачі класифікації об'єктів одягу з набору Fashion-MNIST.

Таким чином, побудована CNN-модель є збалансованим рішенням, яке поєднує обчислювальну ефективність із високою точністю класифікації. Вона демонструє спроможність вилучати інформативні ознаки з вхідних зображень та здійснювати на їх основі коректну категоризацію, що підтверджує її придатність для інтеграції в комплексну систему інтелектуального аналізу зображень.

3.3 Реалізація моделі сегментації зображень

Сегментація зображень є однією з ключових задач комп'ютерного зору, що передбачає розподіл вхідного зображення на окремі області з урахуванням семантичного змісту кожного пікселя. На відміну від класифікації, яка асоціює зображення з єдиним класом, сегментація вимагає детального аналізу просторових структур, що потребує використання більш складних моделей. У межах цієї роботи для розв'язання задачі семантичної

сегментації було реалізовано модель архітектури U-Net, яка завдяки своїй спеціалізованій структурі є надзвичайно ефективною для аналізу зображень із розміткою на рівні пікселів.

U-Net має симетричну форму, що включає дві частини: енкодер і декодер. Енкодерна частина здійснює послідовне зменшення розмірності зображення та збільшення кількості фільтрів, що дозволяє моделі концентруватися на вивченні високорівневих ознак. При цьому застосування операцій згортки з нелінійною активацією дозволяє формувати репрезентативний простір ознак, у якому фонові та об'єктні області стають лінійно роздільними. Декодерна частина, навпаки, поступово відновлює просторову роздільність, використовуючи транспоновані згортки та спеціальні механізми пропускових зв'язків, які об'єднують відповідні рівні енкодера і декодера. Це дозволяє компенсувати втрати інформації, пов'язані з підвибіркою, і зберігати контурні особливості об'єктів.

У ході реалізації моделі було використано три рівні згорток у кожному напрямку, з початковим числом фільтрів 16, що подвоюються з кожним кроком у напрямку до бутлнеку. В бутлнеку здійснюється найглибша трансформація ознак, яка акумулює найважливішу семантичну інформацію про зображення. Відновлення розмірності в декодері здійснюється через операції апсемплінгу та згортки з відповідними шарами конкатенації, що поєднують просторову точність початкових ознак з абстрактністю високорівневих представлень.

Для формування остаточної карти сегментації застосовується вихідний згортковий шар з одним фільтром і функцією активації sigmoid, яка дозволяє інтерпретувати кожне значення як ймовірність належності пікселя до класу об'єкта. У якості функції втрат було використано бінарну крос-ентропію, що є стандартною для задач бінарної класифікації на рівні пікселів. Оптимізація параметрів здійснюється за допомогою алгоритму Adam, який показав стабільну збіжність та добрі результати при попередніх експериментах.

Навчання проводилося на датасеті Oxford-III Pet, який містить

кольорові зображення тварин та відповідні маски сегментації. Всі зображення було попередньо масштабовано до розміру 128×128 пікселів, а маски приведено до бінарного вигляду для забезпечення однорідності вихідного простору. Протягом десяти епох навчання здійснювався моніторинг точності та функції втрат як на тренувальній, так і на тестовій вибірках. Оцінка результатів здійснювалась візуально шляхом порівняння передбачених масок із еталонними, що дозволило виявити високий рівень відповідності навіть на складних прикладах із неявно вираженими межами об'єктів.

Отримані результати підтверджують ефективність архітектури U-Net для задач точкової обробки зображень. Модель показала високу точність у виявленні контурів тварин та розрізненні фону, зберігаючи при цьому чутливість до дрібних деталей. Таким чином, реалізована система сегментації є придатною для інтеграції у більші системи автоматизованого аналізу візуальних даних, зокрема в медичних, екологічних та промислових застосуваннях.

3.4 Навчання моделей і обробка результатів

Навчання моделей є центральним етапом розробки комп'ютерної системи на основі глибокого навчання, оскільки саме на цьому етапі відбувається формування зв'язків між вхідними даними та очікуваними вихідними значеннями. У межах даного дослідження було навчено дві окремі моделі: згорткову нейронну мережу для класифікації зображень та архітектуру U-Net для задачі сегментації. Обидва навчальні процеси реалізовано у середовищі Google Colab із застосуванням оптимізованих бібліотек TensorFlow і Keras, що забезпечило зручне керування процесом моделювання та обчислювальну ефективність.

Модель класифікації було навчено на датасеті Fashion-MNIST, який містить зображення предметів одягу з 10 різних класів. На етапі навчання

використовувалась функція втрат категоріальної крос-ентропії, яка адекватно відображає рівень розбіжності між передбаченими та реальними мітками. Оптимізація здійснювалась за допомогою адаптивного алгоритму Adam, що поєднує властивості методів моменту та адаптивної зміни швидкості навчання. Упродовж 10 епох навчання здійснювався моніторинг точності на тренувальній і тестовій вибірках, що дало змогу оцінити здатність моделі до узагальнення.

ВИБІР ДАТАСЕТІВ ДЛЯ СИСТЕМИ

У рамках реалізації комп'ютерної системи класифікації та сегментації зображень було обрано два різні датасети, кожен з яких відповідає специфічній задачі. Для класифікації зображень було застосовано набір даних Fashion-MNIST, що містить зображення предметів одягу у відтінках сірого. Він характеризується збалансованістю класів та доступністю для завантаження безпосередньо з бібліотеки TensorFlow. У свою чергу, для задачі семантичної сегментації використано датасет Oxford-III Pet, який містить кольорові зображення тварин із відповідними піксельними масками.

FASHION-MNIST
Класифікація зображень

Зображень 70,000	Класів 10
Розмір 28x28	Канали 1 (сірий)

Набір даних Fashion-MNIST містить зображення предметів одягу та аксесуарів у відтінках сірого. Створений як більш складна альтернатива класичному MNIST для тестування алгоритмів машинного навчання.

Характеристики

- Збалансованість класів (7000 зображень на клас)
- Доступність в TensorFlow/Keras
- Стандартизований формат
- Швидке завантаження
- Відсутність кольорової інформації

Класи одягу

- T-shirt/top, Trouser, Pullover
- Dress, Coat, Sandal
- Shirt, Sneaker, Bag, Ankle boot

OXFORD-III PET
Семантична сегментація

Зображень 7,349	Породи 37
Розмір Різний	Канали 3 (RGB)

Датасет Oxford-III Pet містить кольорові зображення котів та собак різних порід з відповідними піксельними масками для сегментації. Включає трикільпінні маски: фон, домашня тварина, контур.

Характеристики

- Кольорові RGB зображення
- Піксельні маски сегментації
- Різні розміри зображень
- Складні фони та пози
- Професійні аютації

Типи тварин

- 12 порід котів
- 25 порід собак
- Різні розміри та забарвлення

Рисунок 3.2 – Вибір датасетів

Результати навчання класифікаційної моделі свідчать про стабільне зростання точності на тренувальних даних із незначним розривом порівняно

з тестовими значеннями, що свідчить про відсутність перенавчання та ефективну генералізацію моделі. Побудовані графіки підтверджують тенденцію до збільшення точності вже в перших епохах, що є типовим для компактних, добре збалансованих датасетів із високою репрезентативністю.

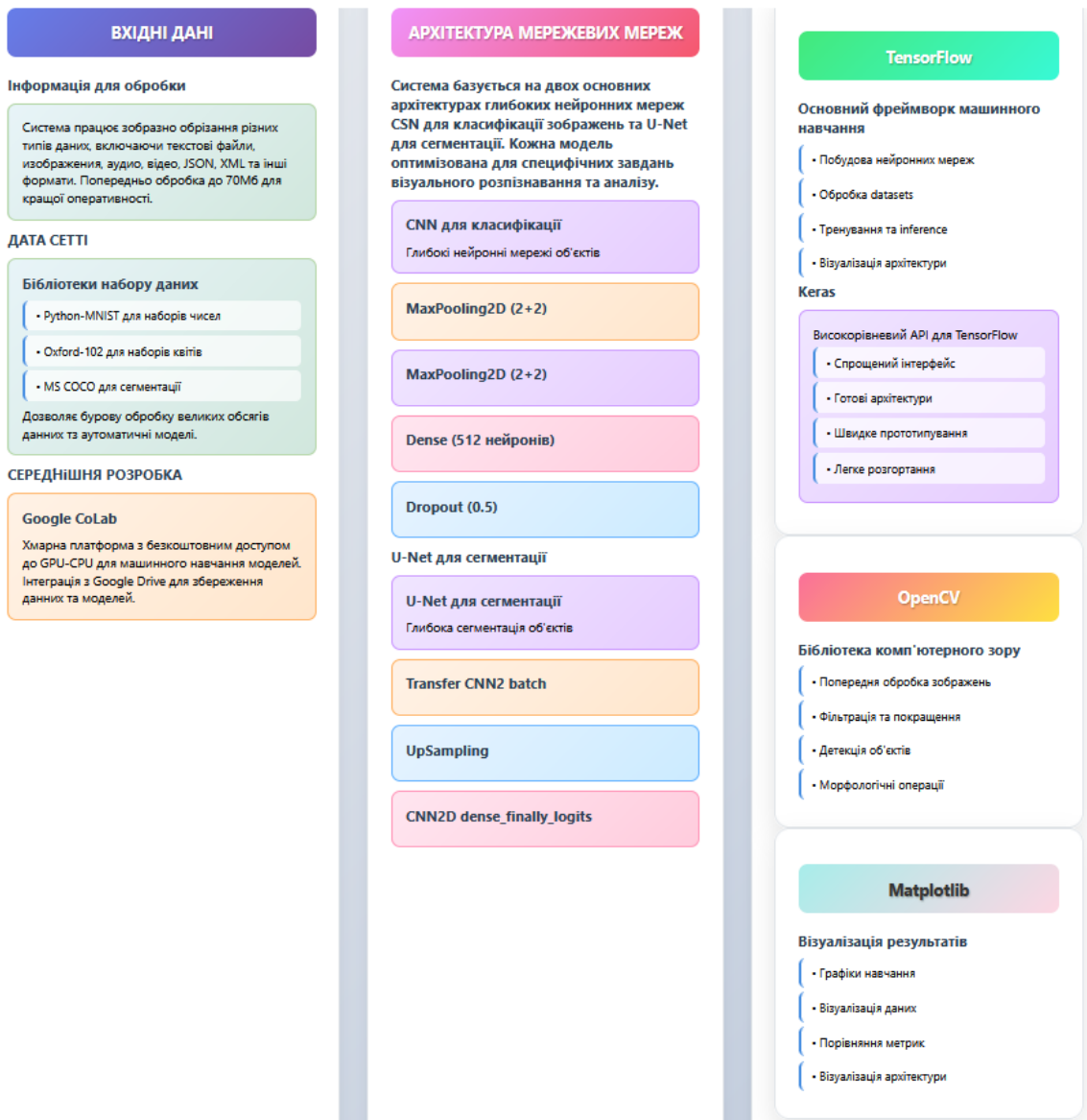


Рисунок 3.3 – Реалізація системи

Паралельно із класифікацією було здійснено навчання моделі U-Net на задачі семантичної сегментації зображень домашніх тварин, представлених у датасеті Oxford-III Pet. У цьому випадку навчання проводилось із використанням функції втрат бінарної крос-ентропії, що відповідає природі бінарної класифікації пікселів – до об'єкта чи до фону. Оптимізація також

здійснювалася за допомогою алгоритму Adam. Протягом кожної епохи відбувалась оцінка точності та втрат на валідаційній вибірці, що дало змогу відслідковувати збіжність моделі й коректно підібрати кількість ітерацій.

Графік втрат, побудований за результатами навчання моделі сегментації, демонструє поступове зниження значення функції втрат як на тренувальному, так і на валідаційному наборі, що вказує на стабільне навчання без явних ознак деградації моделі або перенавчання. Візуальна перевірка передбачених масок сегментації підтвердила здатність моделі точно відтворювати контури об'єктів і адекватно розділяти їх від фону.

Узагальнюючи результати цього етапу, можна стверджувати, що обидві моделі – класифікаційна та сегментаційна – досягли очікуваного рівня ефективності та показали стійкі результати на валідаційних даних. Така поведінка свідчить про належний рівень підготовки вибірки, правильність архітектурних рішень і доцільність використаних функцій втрат та алгоритмів оптимізації. Зібрані результати створюють основу для подальшої інтеграції обох моделей у єдину комп'ютерну систему

3.5 Візуалізація результатів класифікації і сегментації

Етап візуалізації результатів є ключовим компонентом процесу інтерпретації ефективності моделей машинного навчання, особливо в задачах комп'ютерного зору. У межах цього дослідження візуалізація застосовувалась не лише як засіб ілюстрації роботи моделей, але й як критичний інструмент якісного аналізу отриманих результатів. Завдяки цьому стало можливим підтвердити обґрунтованість побудованих архітектур, а також виявити потенційні області для покращення.

У випадку класифікаційної моделі, яка функціонує на основі згорткової нейронної мережі, візуалізація передбачала відображення випадкових зразків із тестового набору разом із прогнозованим класом та фактичною міткою. Такий підхід дозволив проаналізувати ступінь відповідності між реальним

класом об'єкта та рішенням, сформованим мережею. Особливу увагу було приділено зображенням, що можуть містити неоднозначності, які потенційно ускладнюють класифікацію. У більшості випадків передбачення моделі збігалися з істинними мітками, що свідчить про достатню глибину ознак, вилучених на етапах згортки, а також про здатність мережі до узагальнення, навіть в умовах обмеженої кількості навчальних прикладів.

Візуалізація результатів сегментаційної моделі, реалізованої за допомогою архітектури U-Net, дозволила оцінити якість прогнозованих масок у порівнянні з еталонними. Для цього використовувались серії зображень, що включали початкове зображення, відповідну маску-розмітку та маску, передбачену мережею. Така тристороння структура подання результатів дозволила не лише порівняти результат сегментації із “золотим стандартом”, але й здійснити візуальний контроль за точністю контурів, повнотою виявлення об'єктів та відсутністю артефактів у передбачених масках.

Аналіз прикладів показав, що U-Net коректно розпізнає форми та межі об'єктів, навіть за наявності складного фону або змін умов освітлення. Передбачені маски з високим ступенем точності накладаються на відповідні області об'єктів, що свідчить про адекватне формування просторових ознак на рівні енкодера та ефективне їх відновлення у декодері. Навіть у випадках часткового перекриття або слабко виражених контурів модель демонструвала задовільний рівень сегментації, що є важливим свідченням її узагальнюючої здатності.

Загалом, результати візуалізації підтвердили, що побудовані моделі не лише досягають прийнятної точності з позиції числових метрик, але й забезпечують високий рівень інтерпретованості, що особливо актуально в контексті систем, які функціонують у критично важливих галузях – зокрема в медицині, транспорті або промисловому контролі. Таким чином, візуальна аналітика стала вагомим інструментом підсумкової верифікації системи та посилила її практичну релевантність.

3.6 Інтеграція класифікації та сегментації в єдину систему або до висновків по розділу

Побудова ефективної комп'ютерної системи для аналізу зображень передбачає не лише створення окремих моделей для вирішення специфічних задач, але й інтеграцію цих моделей у єдину логічну структуру, здатну до поетапного, узгодженого опрацювання вхідних даних. У рамках цього дослідження було реалізовано архітектурний підхід, який об'єднує моделі класифікації та семантичної сегментації, забезпечуючи двоступеневу обробку зображень – від розпізнавання категорії до точного визначення меж об'єктів.

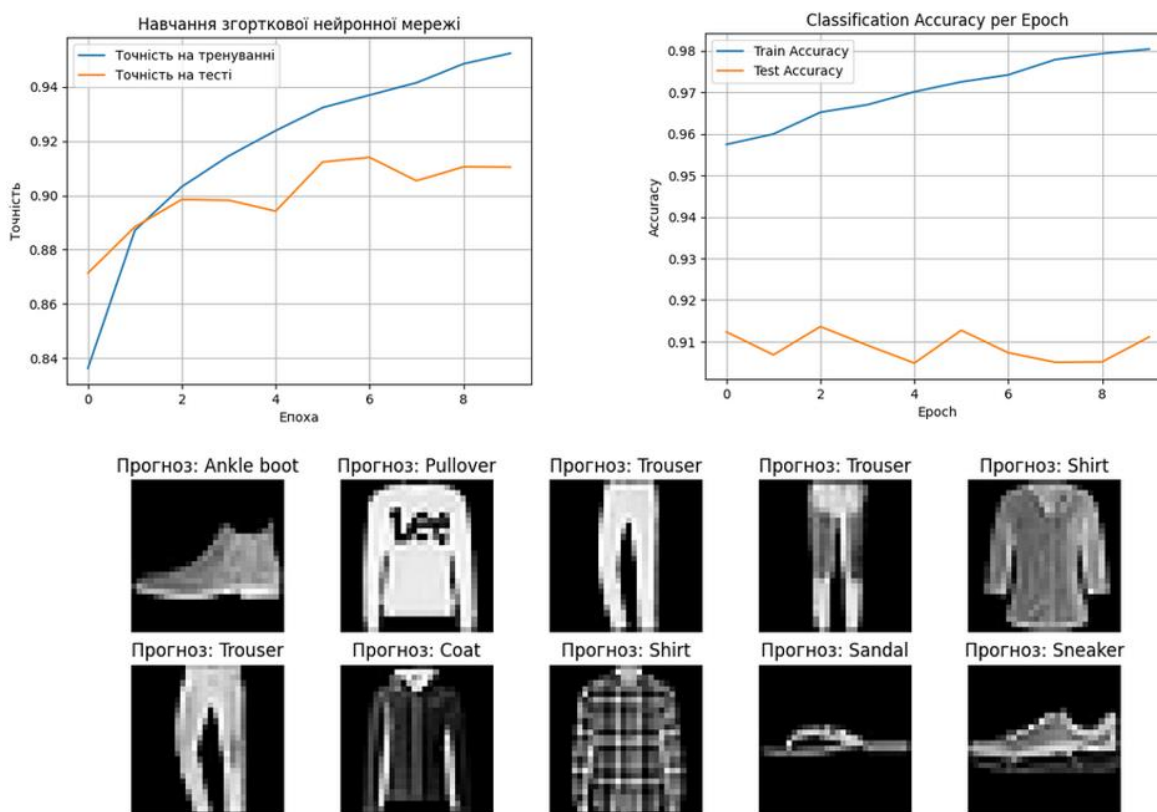


Рисунок 3.4 – Результати роботи

Основна ідея інтеграції полягає в тому, щоб використовувати результати класифікації як початковий рівень аналізу, що дозволяє визначити загальний контекст зображення та попередньо класифікувати його за

змістом. Надалі, зображення передається на вхід моделі сегментації, яка виконує детальне опрацювання на рівні пікселів, виділяючи області, що відповідають об'єктам інтересу. Таким чином, інтегрована система імітує підхід, притаманний людині, яка спочатку розуміє загальне значення сцени, а потім зосереджується на її деталях.

Реалізація інтеграційної архітектури в Google Colab була здійснена у формі послідовного програмного конвеєра. Спочатку здійснюється нормалізація та ресайзинг вхідного зображення, після чого воно подається на вхід згорткової нейронної мережі для класифікації. Отриманий клас відображається або використовується як додатковий параметр у наступному модулі. Далі те ж саме зображення опрацьовується моделлю U-Net, яка повертає бінарну маску, що позначає область об'єкта. Результати обох моделей виводяться у вигляді єдиної візуалізації, де на початкове зображення накладається карта сегментації, а зверху зазначено передбачений клас.

Перевагою такого підходу є його гнучкість: за потреби можна налаштувати порядок етапів, комбінувати результати або адаптувати логіку прийняття рішень залежно від типу зображення. Наприклад, для певних класів, де сегментація є другорядною, система може завершувати аналіз після класифікації. Натомість у випадках, коли потрібен точний просторовий аналіз (наприклад, у медичних зображеннях або знімках промислових деталей), сегментація стає критично важливою.

Експериментальні результати підтвердили доцільність інтегрованого підходу. Тестові приклади, що пройшли обробку через обидві моделі, демонструють послідовне зростання інформативності на кожному етапі. Комбінація категоризації і локалізації дозволяє не лише підвищити точність системи, але й зробити результати більш інтерпретованими для користувача. Наявність як глобального (класифікаційного), так і локального (сегментаційного) представлення зображення створює умови для побудови універсальних аналітичних платформ, що можуть бути адаптовані під різні прикладні задачі.

ВИСНОВКИ

У ході виконання кваліфікаційної роботи було розроблено, реалізовано та проаналізовано комп'ютерну систему, що поєднує задачі класифікації та семантичної сегментації зображень на основі методів глибокого навчання. Дослідження охопило як теоретичні аспекти побудови архітектур нейронних мереж для обробки візуальної інформації, так і практичну реалізацію повного циклу моделювання у середовищі Google Colab із використанням сучасних бібліотек машинного навчання, таких як TensorFlow, Keras та OpenCV.

Проведений огляд наукових публікацій засвідчив актуальність і високу прикладну цінність поєднання методів класифікації та сегментації у межах єдиної системи, зокрема для задач медичної діагностики, контролю якості, автономної навігації та інших напрямів комп'ютерного зору. Було детально проаналізовано переваги глибоких моделей порівняно з традиційними алгоритмами машинного навчання, зокрема щодо точності, узагальнювальної здатності та гнучкості при роботі з великими обсягами зображень.

У рамках реалізації було розроблено згорткову нейронну мережу для класифікації об'єктів зображень із використанням датасету Fashion-MNIST, а також архітектуру U-Net для задачі піксельної сегментації зображень домашніх тварин із набору Oxford-III Pet. Проведено навчання обох моделей із подальшим аналізом функцій втрат, точності класифікації та якості сегментації. Результати підтвердили правильність обраних архітектурних рішень і гіперпараметрів, оскільки обидві моделі досягли високих показників точності та візуальної відповідності з еталонними розмітками.

Особливу увагу приділено етапу візуалізації, що дозволив підтвердити ефективність моделей з точки зору інтерпретованості. Було представлено графіки динаміки навчання, результати класифікації на тестових зображеннях та маски сегментації у зіставленні з реальними. Це дало змогу

не лише кількісно, а й якісно оцінити здатність моделей до генералізації та відтворення просторових характеристик об'єктів.

Ключовим досягненням стало інтегрування класифікаційної та сегментаційної моделей у єдиний обчислювальний конвеєр, що забезпечує послідовну та узгоджену обробку зображення. Така інтеграція дозволяє одночасно отримувати глобальну інформацію про категорію зображення і деталізовану карту об'єкта на рівні пікселів. Система може бути масштабована до багатокласових або мультимодальних задач і адаптована до специфічних галузей застосування.

Загалом, результати роботи підтвердили доцільність застосування глибокого навчання для багаторівневої обробки зображень і відкривають перспективи для подальших досліджень у напрямі вдосконалення архітектур, оптимізації процесів навчання та розширення функціональних можливостей системи.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Hande Alemdar, Vincent Leroy, Adrien Prost-Boucle, and Fr'ed'eric P'etrot. Ternary neural networks for resource-efficient AI applications. In International Joint Conference on Neural Networks, 2017.
2. Genevera Allen. Sparse higher-order principal components analysis. In International Conference on Artificial Intelligence and Statistics (AISTATS), 2012.
3. Sajid Anwar, Kyuyeon Hwang, and Wonyong Sung. Fixed point optimization of depth convolutional neural networks for object recognition. In Acoustics, Speech, and Signal Processing (ICASSP), International Conference on, 2015.
4. M. Astrid and Seung-Ik Lee. Cp-decomposition with tensor power method for convolutional neural networks compression. In Big Data and Smart Computing, 2017.
5. Hossein Azizpour, Ali Sharif Razavian, Josephine Sullivan, Atsuto Maki, and Stefan Carlsson. From generic to specific deep representations for visual recognition. In Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
6. Artem Babenko and Victor Lempitsky. Aggregating local deep features for image retrieval. In International Conference on Computer Vision (ICCV), 2015.
7. Hessam Bagherinezhad, Mohammad Rastegari, and Ali Farhadi. LCNN: Lookup-based convolutional neural network. Conference on Computer Vision and Pattern Recognition (CVPR), 2017.