

Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
(повна назва)Кафедра Інформатики
(повна назва)Рівень вищої освіти другий (магістерський)Спеціальність 122 Комп'ютерні науки
(код і повна назва)Тип програми освітньо-професійнаОсвітня програма Інформатика
(повна назва освітньої програми)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«____» _____ 2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУстудентові Ткаченку Назару Олексійовичу
(прізвище, ім'я, по батькові)1. Тема роботи Дослідження методів семантичної сегментації зображення

затверджена наказом по університету від 25 листопада 2024 року № 1246Ст

2. Термін подання студентом роботи до екзаменаційної комісії 24 грудня 2024 р.

3. Вихідні дані до роботи класи методів сегментації: класичні підходи сегментації, методи глибинного навчання; сегментація за пороговим значенням, методи на основі кластеризації: K-means кластеризація, Fuzzy C-means (FCM), спектральна кластеризація, застосування методів кластеризації; методи розростання областей, методи на основі глибинного навчання: Fully Convolutional Networks, U-Net, SegNet, DeepLab, мінуси та плюси методів глибинного навчання. Метрики оцінки роботи методів семантичної сегментації.

4. Перелік питань, що потрібно опрацювати в роботі _____

1. Аналіз наявних методів семантичної сегментації зображень.2. Дослідження методів семантичної сегментації зображень та їх порівняння.3. Вибір та реалізація методу семантичної сегментації.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) структура Fully Convolutional Networks, архітектура U-Net, структура SegNet, структура DeepLab, діаграма послідовності U-Net, діаграма компонентів U-Net, графік втрат, графік точності, графік коефіцієнту Дайса, графік Mean IoU, графік Jaccard Index, графік Precision, графік Recall, графік F1-оцінки, результат сегментації зображення з людьми, результат сегментації зашумленого зображення, результат сегментації штучно генерованого зображення, результат сегментації зображення вулиці методом суперпіксельної сегментації, результат сегментації зображення авто методом суперпіксельної сегментації, результат сегментації знімка з супутника моделлю Unet, результат сегментації знімка з супутника моделлю Deeplabv3+.

6. Консультанти розділів роботи (п.6 включається до завдання 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

№ з/п	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	25.11.2024	
2	Аналіз завдання, підбір літератури	26.11.24-26.11.24	
3	Аналіз літератури з досліджуваної проблеми	27.11.24-27.11.24	
4	Аналіз технічних засобів	28.11.24-28.11.24	
5	Вибір метрик порівняння методів	29.11.24-29.11.24	
6	Програмна реалізація	30.11.24-01.12.24	
7	Оформлення пояснювальної записки	02.12.24-03.12.24	
8	Перевірка на плагіат	15.12.2024	
9	Рецензування	20.12.2024	
10	Підготовка презентації та доповіді	25.12.2024	
11	Занесення роботи в електронний архів	30.12.2024	
12	Попередній захист кваліфікаційної роботи	02.01.2025	

Дата видачі завдання 25 листопада 2024 р.

Студент _____
(підпис)

Керівник роботи _____ проф. Руденко О.Г.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ/ABSTRACT

Пояснювальна записка до кваліфікаційної роботи: 77 с., 3 табл., 22 рис., 41 джерело.

СЕГМЕНТАЦІЯ ЗОБРАЖЕНЬ, ОБРОБКА ЦИФРОВИХ ЗОБРАЖЕНЬ, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, КЛАСИЧНІ ПІДХОДИ, U-NET, DEEPLABV3+, СУПЕРПІКСЕЛЬНА СЕГМЕНТАЦІЯ.

Об'єктом дослідження є зображення, які потребують семантичної сегментації.

Предметом дослідження є методи семантичної сегментації зображень.

Метою дослідження є проведення детального аналізу сучасних методів семантичної сегментації зображень та їх порівняння з метою виявлення найефективніших підходів.

У ході роботи проведено аналіз методів семантичної сегментації зображень, визначено основні поняття та класи методів сегментації. Досліджено як класичні підходи, такі як порогова сегментація та методи на основі контурів, так і сучасні методи, включаючи глибинне навчання.

У результаті дослідження реалізовано програмне забезпечення для сегментації зображень, проведено його тестування та оцінку ефективності.

IMAGE SEGMENTATION, DIGITAL IMAGE PROCESSING, CONVOLUTIONAL NEURAL NETWORKS, CLASSICAL APPROACHES, U-NET, DEEPLABV3+, SUPER PIXEL SEGMENTATION.

The object of research is images that require semantic segmentation.

The subject of the research is methods of semantic segmentation of images.

The purpose of the research is to conduct a detailed analysis of modern methods of semantic image segmentation and their comparison in order to identify the most effective approaches.

In the course of the work, an analysis of semantic image segmentation methods was carried out, the main concepts and classes of segmentation methods were defined. Both classical approaches, such as threshold segmentation and contour-based methods, as well as modern methods, including deep learning, are explored.

As a result of the research, image segmentation software was implemented, tested and evaluated for its effectiveness.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	7
Вступ.....	8
1 Аналіз предметної області.....	9
1.1 Основні поняття сегментації.....	9
1.2 Види семантичної сегментації зображень	11
1.3 Актуальність розробки системи сегментації зображень.....	15
1.4 Проблематика семантичної сегментації	17
1.5 Постановка задачі дослідження.....	19
2 Дослідження методів семантичної сегментації зображень та їх порівняння.....	22
2.1 Класи методів сегментації.....	22
2.1.1 Класичні підходи сегментації.....	22
2.1.2 Методи глибинного навчання.....	23
2.2 Сегментація за пороговим значенням.....	24
2.2.1 Адаптивне порогоування	26
2.3 Методи на основі кластеризації.....	28
2.3.1 K-means кластеризація	28
2.3.2 Fuzzy C-means (FCM)	29
2.3.3 Спектральна кластеризація	30
2.3.4 Застосування методів кластеризації	30
2.4 Методи розростання областей	31
2.5 Методи на основі глибинного навчання.....	34
2.5.1 Fully Convolutional Networks	35
2.5.2 U-Net.....	36
2.5.3 SegNet.....	38
2.5.4 DeepLab.....	39
2.5.5 Мінуси та плюси методів глибинного навчання	40
2.6 Метрики оцінки роботи методів семантичної сегментації.....	41

	6
3 Вибір та реалізація методу семантичної сегментації	45
3.1 Аргументація вибору методу та датасету.....	45
3.2 Архітектура нейронної мережі	47
3.3 Опис методів алгоритму на основі U-Net.....	53
3.4 Опис методів алгоритму суперпіксельної сегментації зображення.....	54
3.5 Опис методів алгоритму CNN (DeepLabV3+).....	56
3.6 Тестування розроблених методів	57
Висновки	71
Перелік джерел посилання	73

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

AUC – Area Under Curve (площа під кривою)

ASPP – Atrous Spatial Pyramid Pooling (атрозне просторове пірамідальне об'єднання)

CNN – Convolutional Neural Networks (згорткові нейронні мережі)

CRF – Conditional Random Fields (умовні випадкові поля)

FCM – Fuzzy C-means (нечіткі C-середні)

FCN – Fully Convolutional Networks (повністю згорткові мережі)

GPU – Graphics Processing Units (графічні процесори)

IoU – Intersection over Union (середнє перетинання на об'єднання)

mIoU – Mean Intersection over Union (середнє значення перетинання на об'єднання)

OCR – Optical Character Recognition (оптичне розпізнавання символів)

ReLU – Rectified Linear Unit (випрямлена лінійна одиниця)

SLIC – Simple Linear Iterative Clustering (просте лінійне ітеративне кластеризація)

SVM – Support Vector Machines (опорні векторні машини)

ВСТУП

Останнім часом важливим напрямком дослідження у розподілі зображень стало використання глибоких нейронних мереж, особливо згорткових нейронних мереж і їх варіантів. Ці методи дозволяють автоматично виділяти ознаки зображень на різних рівнях деталізації, що забезпечує вищу точність та гнучкість у розпізнаванні складних структур. Успіх цих моделей відкриває перспективи для створення систем, здатних ефективно працювати навіть у реальному часі, що є критично важливим для багатьох інноваційних технологій.

Семантична сегментація зображень є однією з ключових задач у сфері комп'ютерного зору та машинного навчання. Вона полягає у розподілі зображення на різні області, кожна з яких відповідає певному класу об'єктів. Це дозволяє комп'ютеру не лише розпізнавати об'єкти на зображенні, але й точно визначати їхні межі та просторові відносини. Семантична сегментація знаходить застосування в різних галузях, включаючи автономні транспортні засоби, медичну діагностику, робототехніку та доповнену реальність.

З розвитком технологій обробки зображень та машинного навчання семантична сегментація стає все більш актуальною. Успішна сегментація зображень дозволяє досягати високої точності в різноманітних задачах, від автоматичного водіння до аналізу медичних зображень. Зокрема, в медичній сфері, точна сегментація тканин та органів може значно покращити діагностику та лікування. Водночас, складність зображень, висока варіативність об'єктів та необхідність обробки великої кількості даних ставлять перед дослідниками нові виклики. Тому дослідження та порівняння різних методів семантичної сегментації є надзвичайно важливим для подальшого прогресу в цій галузі.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Основні поняття сегментації

Комп'ютерний зір (Computer Vision) — це галузь досліджень, що має на меті навчити машини інтерпретувати та розуміти візуальну інформацію з реального світу. З моменту свого виникнення ця галузь зробила значний прорив, дозволяючи комп'ютерам виконувати завдання, які раніше були доступні лише людині, такі як розпізнавання об'єктів, аналіз сцени та інші завдання, пов'язані з обробкою зображень і відео [1]. Комп'ютерний зір охоплює широкий спектр завдань, які умовно можна поділити на декілька основних напрямків: розпізнавання об'єктів, класифікація зображень, оптичний потік, семантична сегментація, відновлення глибини.

Сегментація зображень – це метод цифрової обробки та аналізу зображень, який передбачає розбиття зображення на кілька частин або сегментів, що охоплюють усе зображення. Конкретніше, це процес присвоєння кожному пікселю зображення мітки з певного набору міток так, що пікселі з однаковою міткою мають схожі візуальні характеристики. Це дозволяє зосередитися на важливих областях, оскільки на практиці часто є області, які не містять корисної інформації [1]. Тому обробка всього зображення одночасно не завжди є ефективною. Часто сегментація зображень базується на властивостях пікселів. Цей процес може включати відокремлення переднього плану від фону або групування пікселів на основі схожості кольору, текстури чи форми. Наприклад, у медичній візуалізації сегментація часто використовується для виділення та маркування пікселів або вокселів у тривимірних об'ємах, які представляють пухлини у мозку чи інших органах пацієнта. Сегментація дозволяє ефективно працювати з важливими частинами зображення, ігноруючи менш значущі області. Це особливо корисно в медичній діагностиці, де важливо точно визначити межі

патологічних утворень. У таких випадках точна сегментація може значно покращити якість діагностики та лікування, автоматизуючи процеси, які раніше виконувалися вручну [1].

Техніки сегментації можуть варіюватися від простих методів порогової обробки, які базуються на інтенсивності пікселів, до складних методів машинного навчання, що враховують різноманітні візуальні ознаки. Наприклад, методи кластеризації, такі як алгоритм K-means, використовуються для групування пікселів за кольором або текстурою. Інші підходи, такі як сегментація на основі контурів, зосереджуються на виявленні меж об'єктів на зображенні [2]. Завдяки розвитку глибокого навчання, сучасні методи сегментації, такі як згорткові нейронні мережі (CNN), здатні досягати високої точності і адаптивності до різноманітних задач. Ці методи використовують великі обсяги навчальних даних для автоматичного визначення найбільш значущих ознак для сегментації. Таким чином, сегментація зображень є важливим етапом в обробці зображень, який має широкий спектр застосувань у різних галузях, таких як медична візуалізація, автономні транспортні засоби та системи відеоспостереження. Завдяки сегментації можливо підвищити точність та ефективність роботи цих систем, забезпечуючи високу якість результатів та спрощуючи подальшу обробку даних [2]. На рисунку 1.1 наведено приклад сегментації зображення.

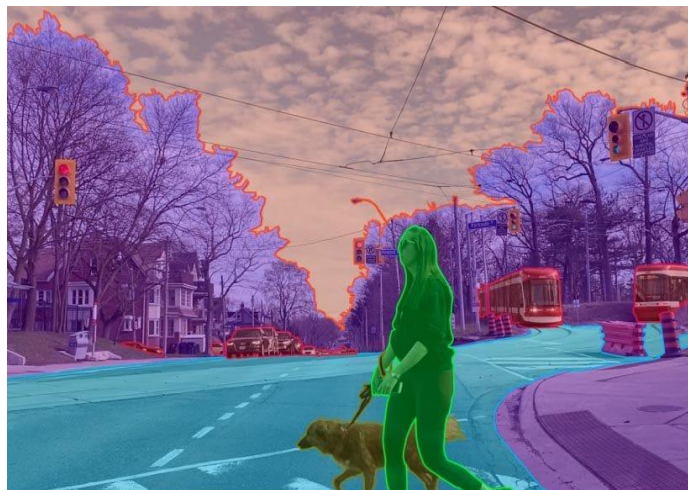


Рисунок 1.1 – Приклад семантичної сегментації зображення [2]

Сегментація зображень створює піксельну маску для кожного об'єкта на сцені. Цей метод надає значно більш детальну інформацію про об'єкти на зображенні.

1.2 Види семантичної сегментації зображень

Семантична сегментація зображень є складною задачею, що передбачає класифікацію кожного пікселя зображення в один з кількох класів. Існує декілька підходів до реалізації семантичної сегментації, кожен з яких має свої переваги та недоліки [3]. Основні види семантичної сегментації зображень включають класичні методи, методи на основі машинного навчання та методи на основі глибокого навчання.

Класичні методи сегментації базуються на обробці низькорівневих ознак зображення, таких як інтенсивність пікселів, текстура та колір. Серед цих методів виділяють порогову сегментацію, сегментацію на основі контурів, регіональну сегментацію та кластеризацію. Порогова сегментація – цей метод базується на виділенні пікселів, інтенсивність яких перевищує або нижча за певний поріг [3]. Порогова сегментація проста в реалізації, але часто недостатньо точна, особливо для складних зображень з варіативним освітленням.

Сегментація на основі контурів використовується для виявлення меж об'єктів на зображенні. Метод заснований на аналізі градієнтів інтенсивності пікселів. Популярні алгоритми, такі як Canny edge detector, дозволяють точно визначати контури об'єктів, але можуть бути чутливими до шуму [4].

Регіональна сегментація – цей метод включає розбиття зображення на регіони на основі схожості пікселів. Алгоритми, такі як Watershed, працюють на основі аналогії з заливанням води в мінімальні точки рельєфу. Регіональна сегментація ефективна для зображень з чітко вираженими

областями, але може бути складною для реалізації на складних зображеннях [4].

Метод кластеризації, такий як K-means, групує пікселі на основі їх подібності за кольором або текстурою. Цей підхід дозволяє виділити області зі схожими характеристиками, але може не враховувати просторові відносини між пікселями [4].

Методи машинного навчання використовують алгоритми, що навчаються на основі наборів навчальних даних для покращення точності сегментації. Основні методи включають Random Forest, Support Vector Machines (SVM) та Conditional Random Fields (CRF). Random Forest – цей метод використовує ансамбль дерев прийняття рішень для класифікації пікселів. Random Forest є потужним інструментом для сегментації завдяки своїй здатності обробляти велику кількість ознак та уникати переобучення [5].

Support Vector Machines (SVM) використовує гіперплощини для поділу даних на класи. Для сегментації зображень SVM може бути розширено до використання у формі одного проти всіх для багатокласової класифікації. SVM є ефективним для високовимірних даних, але потребує значного часу для навчання. Conditional Random Fields (CRF) моделюють просторові залежності між пікселями, що дозволяє враховувати контекст зображення при сегментації [5]. CRF є ефективним для сегментації зображень з вираженими просторовими структурами, але потребує складної обробки та оптимізації.

Методи глибокого навчання, такі як згорткові нейронні мережі (CNN), революціонізували семантичну сегментацію завдяки своїй здатності автоматично виділяти високорівневі ознаки зображень. Fully Convolutional Networks (FCN) є основою для багатьох сучасних методів сегментації. Вони використовують лише згорткові шари, що дозволяє моделі приймати зображення будь-якого розміру та виходити на сегментовану карту. FCN

здатні досягати високої точності, але можуть втрачати деталізацію на рівні пікселів.

U-Net є архітектурою, розробленою для біомедичних зображень, яка поєднує шари згортки та декодування для відновлення просторової роздільної здатності. U-Net досягає високої точності навіть для невеликих навчальних вибірок завдяки своєму симетричному дизайну [6]. SegNet використовує енкодер-декодерну архітектуру, де енкодер складається з згорткових шарів, а декодер – з децимаційних. Ця модель ефективно зберігає просторову інформацію, що дозволяє досягати високої якості сегментації [7].

DeepLab використовує декілька інноваційних технік, таких як Atrous Convolution та Conditional Random Fields, для підвищення якості сегментації. DeepLab здатний обробляти зображення з високою деталізацією та є однією з провідних моделей у галузі [8].

Семантична сегментація зображень включає в себе різноманітні методи, від класичних до сучасних алгоритмів глибокого навчання. Кожен підхід має свої переваги та обмеження, що робить їх придатними для різних завдань та умов. Розуміння цих методів та їх застосування є ключовим для досягнення високої точності та ефективності в задачах обробки зображень.

Суперпіксельна сегментація є важливим інструментом в обробці зображень та комп'ютерному зорі. Вона передбачає розбиття зображення на невеликі області або «суперпікселі», які групують пікселі з подібними властивостями. Це дозволяє значно зменшити складність завдань обробки зображень, таких як сегментація, класифікація та аналіз текстури [9].

Існує кілька основних методів для виконання суперпіксельної сегментації. Розглянемо найпоширеніші з них.

По-перше, SLIC (Simple Linear Iterative Clustering) є одним з найпопулярніших алгоритмів суперпіксельної сегментації, завдяки своїй простоті та ефективності. Алгоритм починає з рівномірного розташування кластерних центрів по всьому зображенню. В кожній ітерації кожен піксель

приписується до найближчого кластерного центру з урахуванням просторових та колірних відстаней [9]. Зображення розділяється на регулярну сітку з розміром суперпікселів S , визначеним як N/K , де N – кількість пікселів у зображенні, а K – бажана кількість суперпікселів. Центральні пікселі кожного кластеру ініціалізуються в якості початкових центрів кластерів. Для кожного початкового центру обчислюється градієнт, щоб знайти точку з найменшим градієнтом інтенсивності у сусідньому вікні 3×3 . Це допомагає уникнути центрів кластерів, що розташовані на краях або межах. Після кожної ітерації кластерні центри оновлюються як середнє значення пікселів, що належать до цього кластеру. Користувач може контролювати розмір суперпікселів та рівень компактності [9].

Ще один метод – Felzenszwalb and Huttenlocher Algorithm, що заснований на алгоритмі мінімального розрізу (minimum spanning tree). Зображення представляється у вигляді графу, де кожен піксель є вершиною, а ребра з'єднують сусідні пікселі [9]. Вага ребра визначається різницею в інтенсивностях між двома пікселями. Використовується алгоритм мінімального розрізу для поділу графу на підграфи, які відповідають суперпікселям.

Watershed Algorithm використовує концепцію геодезичних відстаней для сегментації зображення. Розміщуються «насіння» (markers) в різних регіонах зображення [9]. Насіння розширюються, заповнюючи найближчі пікселі, поки не зустрінуться з іншими насіннями. Зображення розбивається на регіони, кожен з яких відповідає одному суперпікселю.

Переваги суперпіксельної сегментації [9]:

- суперпікселі зменшують кількість елементів для аналізу, що дозволяє значно скоротити час обробки зображень;
- сегментація на рівні суперпікселів дозволяє зберігати важливі просторові зв'язки між пікселями, що покращує точність подальших алгоритмів;

– суперпіксельна сегментація може бути застосована до широкого спектру зображень та задач, від медичної візуалізації до аналізу супутникових знімків.

Недоліки суперпіксельної сегментації:

– ефективність сегментації сильно залежить від вибору параметрів алгоритму, таких як розмір суперпікселів та рівень компактності;

– неправильний вибір параметрів може призвести до пере- або недосегментації, що вплине на точність подальшої обробки;

– в деяких випадках сегментація на рівні суперпікселів може призвести до втрати важливих деталей, особливо на границях об'єктів.

Суперпіксельна сегментація є потужним інструментом в обробці зображень, що дозволяє значно скоротити обчислювальну складність та покращити точність подальших алгоритмів. Вона знаходить застосування в різних галузях, таких як медична візуалізація, аналіз супутникових знімків, автономне водіння та обробка природної мови. Незважаючи на деякі недоліки, такі як залежність від параметрів та можливість втрати деталей, суперпіксельна сегментація залишається ефективним методом для попередньої обробки зображень [9].

Сучасні дослідження в цій галузі спрямовані на вдосконалення існуючих алгоритмів та розробку нових методів, які б забезпечували ще вищу точність та ефективність. Наприклад, поєднання суперпіксельної сегментації з методами глибокого навчання дозволяє створювати більш потужні системи для аналізу зображень [9].

1.3 Актуальність розробки системи сегментації зображень

На сьогоднішній день технології штучного інтелекту та машинного навчання все активніше інтегруються в різноманітні галузі для підвищення ефективності та якості роботи. Системи розпізнавання зображень зазнали

значного розвитку, що відкриває нові можливості для вирішення широкого спектра задач. Сегментація зображень, яка є одним з ключових етапів обробки зображень, має велике значення та є дуже актуальною. За останні роки опубліковано багато наукових досліджень, що пропонують нові методи сегментації для різних прикладних сфер.

Однією з важливих областей застосування сегментації зображень є медична візуалізація. В цій галузі сегментація допомагає автоматично виявляти підозрілі утворення, класифікувати патології, визначати типи тканин та їх об'єми, а також планувати та виконувати хірургічні операції за допомогою комп'ютерної навігації. Вхідними даними для таких систем можуть бути рентгенівські знімки, отримані за допомогою різних методів візуалізації, таких як рентгенографія, флюороскопія, магнітно-резонансна томографія, ультразвукова ехографія, ендоскопія, еластографія та термографія. Крім того, для сегментації можуть використовуватися об'ємні зображення, отримані з комп'ютерної та магнітно-резонансної томографії.

Ще однією значущою областю застосування сегментації зображень є машинний зір. У цій сфері сегментація допомагає системам ідентифікувати та визначати розташування різноманітних об'єктів на зображенні, таких як пішоходи та автомобілі, що є надзвичайно важливим при розробці автономних транспортних засобів. У машинному зорі можна виділити дві великі групи задач: виявлення об'єктів та розпізнавання. Виявлення об'єктів може використовуватися для широкого кола завдань, включаючи виявлення транспортних засобів, пішоходів, дорожніх знаків, світлофорів, а також виділення об'єктів на супутникових знімках. Задачі розпізнавання включають розпізнавання обличчя, відбитків пальців чи сітківки ока для ідентифікації особи.

Сегментація зображень також знаходить застосування в системах контролю дорожнього руху, системах відеоспостереження, задачах стиснення зображень та обробці інформаційних потоків. Наприклад, у системах відеоспостереження сегментація дозволяє автоматично виявляти

підозрілі об'єкти або дії, що значно підвищує рівень безпеки. В системах контролю дорожнього руху сегментація допомагає визначати транспортні засоби та пішоходів, що сприяє ефективному управлінню трафіком та зменшенню аварійності.

Таким чином, сегментація зображень є надзвичайно важливою та актуальною темою досліджень, що має широкий спектр практичних застосувань у різних галузях. Розвиток нових методів сегментації та їх впровадження в реальні системи може значно підвищити ефективність та точність роботи різноманітних технологічних рішень, сприяючи подальшому прогресу в галузі штучного інтелекту та машинного навчання.

1.4 Проблематика семантичної сегментації

Семантична сегментація зображень є складною та багатогранною задачею, яка стикається з численними проблемами та викликами. Ці проблеми можуть впливати на точність, ефективність та надійність сегментації, що у свою чергу може мати критичні наслідки для її застосувань у реальних світах, таких як автономні транспортні засоби, медична діагностика та інші. У цьому розділі розглядаються основні проблеми, з якими стикаються сучасні методи семантичної сегментації.

Одна з найбільших проблем семантичної сегментації – це велика варіативність об'єктів на зображеннях. Об'єкти одного класу можуть мати різні розміри, форми, кольори та текстури. Наприклад, автомобілі можуть бути різних марок, моделей і кольорів, а також мати різні кути нахилу та освітлення. Це ускладнює задачу алгоритмів сегментації, оскільки їм необхідно враховувати всі ці варіації для точного виділення об'єктів.

Освітлення та умови зйомки значно впливають на якість зображень і, відповідно, на результати сегментації. Зміни в освітленні можуть призводити до змін у вигляді та кольорі об'єктів, що ускладнює їх

ідентифікацію. Наприклад, тіні можуть зливатися з об'єктами або фон можуть змінюватися в залежності від кута падіння світла. Це особливо актуально для зовнішніх середовищ, де умови освітлення постійно змінюються. Ще однією проблемою є складність фонів, на яких розташовані об'єкти. У реальних зображеннях фони можуть бути дуже різноманітними і містити багато зайвих деталей, які можуть заплутати алгоритми сегментації. Наприклад, у міських умовах фони можуть містити будівлі, людей, транспортні засоби та інші об'єкти, що ускладнює виділення основних об'єктів, таких як автомобілі або пішоходи.

Для тренування моделей глибокого навчання необхідні великі обсяги анотованих даних. Проте, процес створення навчальних наборів даних для семантичної сегментації є дуже трудомістким та витратним. Анотація кожного пікселя зображення вручну займає багато часу і потребує високої точності, що ускладнює створення великих наборів даних. Без достатньої кількості якісних даних моделі не можуть досягати високої точності. Перевірка та валідація моделей семантичної сегментації є складним завданням. Часто моделі можуть показувати високі результати на тестових даних, але при цьому мати проблеми з узагальненням на нові, невідомі дані. Це може призводити до непередбачуваних помилок у реальних умовах. Тому необхідно ретельно перевіряти моделі на різноманітних наборах даних, щоб переконатися в їх надійності.

Глибокі нейронні мережі, які часто використовуються для семантичної сегментації, є «чорними ящиками». Це означає, що важко інтерпретувати, чому модель приймає ті чи інші рішення. Відсутність прозорості в роботі моделей ускладнює виправлення помилок та підвищення точності. Інтерпретованість є критично важливою у таких областях, як медична діагностика, де помилки можуть мати серйозні наслідки. Методи глибокого навчання вимагають значних обчислювальних ресурсів для тренування та роботи з моделями. Це може бути проблемою для застосунків, що потребують швидкої обробки зображень у режимі

реального часу. Наприклад, автономні транспортні засоби обробляють зображення в реальному часі для безпечного керування, вимагаючи високопродуктивних обчислювальних систем.

Зображення можуть містити шуми та артефакти, які ускладнюють сегментацію. Це можуть бути цифрові шуми, артефакти компресії, спотворення, викликані рухом, тощо. Алгоритми сегментації повинні бути стійкими до таких шумів, щоб забезпечувати високу якість результатів. Системи семантичної сегментації часто тренуються на певних наборах даних і можуть мати проблеми з адаптацією до нових умов або нових типів зображень. Наприклад, модель, тренувана на міських зображеннях, може мати проблеми з сегментацією зображень у сільській місцевості або у різних погодних умовах.

Семантична сегментація зображень є важливою та складною задачею з багатьма викликами та проблемами. Варіативність об'єктів, умови освітлення, складність фонів, нестача навчальних даних, перевірка та валідація моделей, інтерпретованість, високі обчислювальні вимоги, стійкість до шумів та адаптація до нових умов – все це впливає на ефективність та точність сегментації [10]. Подолання цих проблем є ключовим для розвитку та вдосконалення методів семантичної сегментації, що дозволить досягти більшої надійності та ефективності у різних сферах застосування.

1.5 Постановка задачі дослідження

З розвитком технологій обробки зображень та машинного навчання семантична сегментація стає все більш актуальною. Успішна сегментація зображень дозволяє досягати високої точності в різноманітних задачах, від автоматичного водіння до аналізу медичних зображень. Зокрема, в медичній сфері точна сегментація тканин та органів може значно покращити

діагностику та лікування. Водночас, складність зображень, висока варіативність об'єктів та необхідність обробки великої кількості даних ставлять перед дослідниками нові виклики. Тому дослідження та порівняння різних методів семантичної сегментації є надзвичайно важливим для подальшого прогресу в цій галузі. Сучасні методи семантичної сегментації включають як класичні підходи, так і методи на основі глибинного навчання. Класичні методи, такі як порогова сегментація, сегментація на основі контурів та кластеризація, мають свої переваги, але часто не забезпечують достатньої точності на складних зображеннях. Методи на основі глибинного навчання, такі як U-Net та DeepLabV3+, показують високу ефективність завдяки здатності нейронних мереж вивчати складні патерни в даних. Проте вони вимагають великих обсягів даних для навчання та значних обчислювальних ресурсів.

Об'єктом дослідження є зображення, які потребують семантичної сегментації.

Предметом дослідження є методи семантичної сегментації зображень.

Метою дослідження є проведення детального аналізу сучасних методів семантичної сегментації зображень та їх порівняння з метою виявлення найефективніших підходів.

Для досягнення мети необхідно вирішити такі завдання:

- проаналізувати сучасні методи семантичної сегментації зображень;
- визначити основні поняття та класи методів сегментації;
- дослідити класичні методи сегментації, такі як порогова сегментація та методи на основі контурів;
- дослідити сучасні методи сегментації, включаючи методи глибинного навчання;
- провести порівняльний аналіз методів на основі різних метрик оцінки;
- розробити програмне забезпечення для сегментації зображень, провести його тестування та оцінку ефективності;

– розглянути проблематику та виклики сегментації, запропонувати шляхи їх подолання для підвищення точності та надійності сегментації зображень у різних практичних застосуваннях.

Для досягнення поставленої мети буде використовуватися комплексний підхід, який включає теоретичний аналіз, розробку алгоритмів, їх реалізацію та експериментальну перевірку. Теоретичний аналіз включатиме огляд літератури, класифікацію методів та їх порівняння. Розробка алгоритмів базуватиметься на сучасних підходах глибинного навчання та класичних методах сегментації. Експериментальна перевірка буде проведена на основі стандартних наборів даних для оцінки точності та ефективності розроблених методів.

Очікується, що результати дослідження дозволять визначити найефективніші методи семантичної сегментації зображень, які можуть бути використані в різних прикладних задачах. Розроблене програмне забезпечення для сегментації зображень буде протестоване та оцінене, що дозволить виявити його переваги та недоліки. Отримані результати можуть бути використані для подальшого вдосконалення методів сегментації та їх впровадження в практичні системи комп'ютерного зору.

2 ДОСЛІДЖЕННЯ МЕТОДІВ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ ЗОБРАЖЕНЬ ТА ЇХ ПОРІВНЯННЯ

2.1 Класи методів сегментації

Методи сегментації зображень можна розділити на два основних класи: методи, засновані на класичних підходах комп'ютерного зору, та методи, що використовують глибинне навчання. Кожен з цих підходів має свої переваги, недоліки та специфічні області застосування.

2.1.1 Класичні підходи сегментації

Різноманітні техніки комп'ютерного зору виникли з розвитком області цифрової обробки зображень, починаючи з 1960-х років. Відтоді було запропоновано багато різних методів вирішення задачі сегментації. Кожен підхід розглядає сегментацію з іншої точки зору, тому для вирішення різних задач використовуються різні методи. Класичні підходи сегментації можна класифікувати на три основні категорії [11]:

- методи структурної сегментації: ці методи базуються на інформації про структуру потрібної частини зображення, тобто області, яка має бути сегментована;

- методи стохастичної сегментації: ці підходи працюють з дискретними значеннями пікселів зображення, замість використання структурної інформації про регіон;

- гібридні підходи: ці методи поєднують концепції як структурної, так і стохастичної сегментації, використовуючи одночасно інформацію про пікселі та структуру.

Серед основних класичних методів можна виділити [11]:

- методи порогового значення базуються на розділенні зображення на області на основі інтенсивності пікселів;
- методи на основі виявлення меж використовують градієнти інтенсивності для виявлення меж між різними областями зображення;
- методи розростання областей включають сусідні пікселі до області, якщо вони відповідають певним критеріям схожості;
- методи на основі кластеризації групують пікселі в кластери на основі їх схожості;
- методи розділу графів представляють зображення як граф, де вершини відповідають пікселям, а ребра – схожості між ними;
- методи на основі вододілу сегментують зображення, розглядаючи його як топографічну поверхню, де вододіл розділяє різні області;
- методи, засновані на рівняннях з частковими похідними використовують математичні моделі для сегментації зображень.

2.1.2 Методи глибинного навчання

Людина, виконуючи сегментацію зображень, здійснює набагато складнішу роботу, ніж класичні методи, використовуючи при цьому знання з конкретної предметної області. Методи сегментації на основі глибинного навчання долають цей бар'єр завдяки моделюванню знань з розміченого набору даних для навчання. Ці методи кардинально відрізняються від класичних і забезпечують кращі результати, хоча й потребують більше обчислювальних ресурсів.

Методи глибинного навчання використовують нейронні мережі, зокрема згорткові нейронні мережі (Convolutional Neural Networks, CNN), для автоматичної сегментації зображень [12]. Основні підходи включають [12]:

- Fully Convolutional Networks (FCN) використовують лише згорткові шари для створення карти сегментації зображення;
- U-Net спеціалізована архітектура нейронної мережі, яка має симетричну структуру і забезпечує високу точність сегментації навіть на невеликих наборах даних;
- SegNet нейронна мережа, яка включає шари енкодера та декодера для відновлення повнорозмірної сегментованої карти;
- DeepLab використовує атріумні згортки для обробки контексту зображення на різних масштабах.

Методи глибинного навчання дозволяють досягати високої точності сегментації за рахунок глибокого аналізу зображень, але вимагають значних обчислювальних ресурсів та великих обсягів даних для навчання. Методи сегментації зображень є важливими інструментами в області комп'ютерного зору та обробки зображень [12]. Класичні підходи забезпечують швидкі та обчислювально ефективні рішення для багатьох задач сегментації, тоді як методи глибинного навчання надають можливість досягати високої точності сегментації завдяки моделюванню складних патернів у зображеннях. Вибір конкретного методу залежить від специфіки завдання, характеристик зображення та вимог до точності сегментації. Розуміння цих підходів дозволяє ефективніше вирішувати завдання в області комп'ютерного зору та обробки зображень, що є критично важливим для багатьох сучасних застосувань у науці та промисловості.

2.2 Сегментація за пороговим значенням

Сегментація за пороговим значенням є одним із найбільш фундаментальних і широко використовуваних методів сегментації зображень. Цей підхід базується на розподілі пікселів зображення на різні області шляхом застосування одного або кількох порогових значень

інтенсивності [13]. Незважаючи на свою простоту, методи порогового значення можуть бути дуже ефективними для вирішення ряду задач в області обробки зображень, таких як виявлення об'єктів, відокремлення переднього плану від фону та бінаризація зображень.

Метод порогового значення включає в себе встановлення певного порогового рівня T , який використовується для розділення пікселів зображення I на дві області: область об'єкта та область фону. Формально, це можна записати наступним чином [13]:

$$\{1, \text{якщо } I(x, y) \geq T \quad 0, \text{якщо } I(x, y) < T \quad , \quad (2.1)$$

де $I(x, y)$ – інтенсивність пікселя в координатах (x, y) .

У випадку багатопорогової сегментації, зображення розділяється на кілька рівнів інтенсивності, кожен з яких відповідає певному інтервалу порогів. Це дозволяє виділити більше ніж дві області на зображенні, що може бути корисним для складніших задач сегментації.

Вибір оптимального порогового значення є критичним кроком у пороговій сегментації. Існує декілька методів вибору порогу, серед яких найбільш популярними є [13]:

- користувач самостійно визначає порогове значення на основі візуального аналізу гістограми зображення. Цей метод підходить для простих випадків або коли користувач має достатньо досвіду для адекватного вибору порогу;

- адаптивне порогоування. Цей метод обчислює порогове значення для кожної локальної області зображення окремо, що дозволяє враховувати варіації освітлення та локальні особливості зображення. Адаптивне порогоування особливо корисне для зображень з нерівномірним освітленням;

- автоматичний метод вибору порогу, який мінімізує внутрішньокласову дисперсію інтенсивностей. Формально, метод Оцу знаходить поріг T , який мінімізує функцію:

$$\sigma_B^2(T) = \sigma^2 - \frac{\sigma_W^2(T)}{\sigma_T^2(T)}, \quad (2.2)$$

де σ^2 – дисперсія інтенсивностей всього зображення,

$\sigma_W^2(T)$ – внутрішньокласова дисперсія,

$\sigma_T^2(T)$ – міжкласова дисперсія.

2.2.1 Адаптивне порогоування

Адаптивне порогоування є ефективним методом сегментації для зображень з нерівномірним освітленням. Воно полягає в обчисленні порогового значення для кожної локальної області зображення окремо. Існують два основних підходи до адаптивного порогоування [14]:

- метод середньозваженого значення (Gaussian). Порогове значення визначається як середнє значення інтенсивностей пікселів у локальному вікні, зважене за Гаусовим розподілом. Це дозволяє більш точно враховувати вплив ближніх пікселів на вибір порогу;

- метод середнього значення. Порогове значення для кожної локальної області визначається як середнє значення інтенсивностей пікселів у цій області. Формально це можна записати як:

$$T(x, y) = \frac{1}{|W|} \sum_{(i,j) \in W} I(i, j), \quad (2.3)$$

де W – локальне вікно,

$|W|$ – кількість пікселів у вікні,

$I(i, j)$ – інтенсивність пікселя в координатах (i, j) .

Методи порогової сегментації знаходять широке застосування в різних областях. Розглянемо декілька прикладів [14]:

- порогова сегментація використовується для виявлення аномалій у медичних зображеннях, таких як рентгенограми або МРТ. Наприклад, вона може допомогти виявити пухлини або інші патологічні зміни в органах;
- у задачах оптичного розпізнавання символів (OCR) порогова сегментація використовується для відокремлення тексту від фону, що значно полегшує подальше розпізнавання символів;
- у промислових системах контролю якості порогова сегментація допомагає виявляти дефекти на поверхнях виробів або виділяти контури об'єктів для їх подальшої обробки.

Незважаючи на свою простоту і швидкість, методи порогового значення мають певні обмеження. Методи порогового значення можуть погано працювати на зображеннях з нерівномірним освітленням або великими градієнтами освітленості. Адаптивне порогоування частково вирішує цю проблему, але не завжди дає ідеальний результат. Наявність шуму в зображенні може суттєво вплинути на точність сегментації. Попередня фільтрація зображення може допомогти зменшити вплив шуму, але може також призвести до втрати важливої інформації [14]. Для зображень з високою текстурною складністю методи порогового значення можуть бути недостатніми. У таких випадках доцільно використовувати більш складні методи, такі як методи на основі глибинного навчання.

Методи порогової сегментації залишаються важливим інструментом в арсеналі методів обробки зображень. Вони є простими у реалізації та ефективними для багатьох задач, де є чітке розрізнення між об'єктом та фоном [14]. Однак, для більш складних задач, що включають варіації освітлення, наявність шуму або високу текстурну складність, можуть знадобитися додаткові підходи або комбінація методів для досягнення оптимальних результатів.

2.3 Методи на основі кластеризації

Методи на основі кластеризації є важливими інструментами для сегментації зображень у галузі комп'ютерного зору та цифрової обробки зображень. Кластеризація дозволяє групувати пікселі зображення в однорідні області на основі їх подібності, що дозволяє ефективно виділяти різні об'єкти та структури в зображеннях. У цій статті буде розглянуто основні принципи та алгоритми кластеризації, а також їх застосування та обмеження.

Кластеризація є процесом групування набору об'єктів таким чином, щоб об'єкти в межах одного кластеру були більш схожі між собою, ніж з об'єктами з інших кластерів. У контексті сегментації зображень об'єктами є пікселі або вектори ознак, що представляють пікселі. Основною метою кластеризації є розбиття зображення на області з однорідними властивостями, такими як колір, текстура або інтенсивність.

2.3.1 K-means кластеризація

K-means є одним з найпопулярніших і найпростіших алгоритмів кластеризації. Його основна ідея полягає у розбитті набору даних на K кластерів шляхом мінімізації внутрішньокласової дисперсії. Алгоритм включає такі кроки [15]:

Крок 1. Вибір K початкових центрів кластерів (це можуть бути випадкові пікселі або результат попередньої обробки).

Крок 2. Кожен піксель призначається до найближчого центру кластеру, використовуючи певну міру відстані (наприклад, евклідову відстань).

Крок 3. Центри кластерів оновлюються як середні значення пікселів, призначених до кожного кластеру.

Крок 4. Кроки призначення та оновлення повторюються до тих пір, поки зміни в центрах кластерів не стануть незначними.

K-means є ефективним та швидким алгоритмом, але він має деякі обмеження, такі як залежність від початкових умов та необхідність задання числа кластерів K заздалегідь.

2.3.2 Fuzzy C-means (FCM)

Fuzzy C-means є варіацією K-means, яка дозволяє кожному пікселю належати до кількох кластерів з різною ступенем належності. Це робить алгоритм більш гнучким та дозволяє краще сегментувати зображення з розмитими межами між об'єктами [16]. Основні кроки алгоритму FCM схожі на K-means, але з додатковими обчисленнями ступеня належності кожного пікселя до кожного кластеру [16]:

Крок 1. Вибір початкових центрів кластерів та випадкове призначення ступенів належності.

Крок 2. Обчислення ступенів належності кожного пікселя до кожного кластеру, використовуючи функцію відстані та міру належності.

Крок 3. Оновлення центрів кластерів як зважених середніх значень пікселів.

Крок 4. Повторення кроків призначення та оновлення до досягнення збіжності.

FCM є корисним для задач, де необхідно враховувати невизначеність у належності пікселів до кластерів, але вимагає більше обчислювальних ресурсів порівняно з K-means.

2.3.3 Спектральна кластеризація

Спектральна кластеризація використовує властивості спектру (власних значень) матриці подібності для виконання кластеризації. Алгоритм включає такі кроки [17]:

Крок 1. Матриця подібності обчислюється на основі певної міри відстані між пікселями.

Крок 2. Власні значення та відповідні власні вектори матриці подібності обчислюються для виявлення структури даних.

Крок 3. Найбільш значущі власні вектори вибираються для створення нової, зниженої розмірності простору ознак.

Крок 4. K-means або інший алгоритм кластеризації застосовується до нового простору ознак для визначення кластерів.

Спектральна кластеризація може ефективно виявляти складні, нелінійні структури в даних, але є обчислювально інтенсивною та вимагає великих обсягів пам'яті [17].

2.3.4 Застосування методів кластеризації

Методи кластеризації знаходять широке застосування у багатьох областях. Кластеризація використовується для сегментації медичних зображень, таких як МРТ, КТ або ультразвукові зображення, для виявлення та аналізу патологічних змін у тканинах або органах. У задачах розпізнавання об'єктів кластеризація допомагає виділити області інтересу, що містять об'єкти, які потрібно розпізнати або класифікувати. Кластеризація використовується для сегментації супутникових зображень, що дозволяє визначити різні типи земної поверхні, такі як водні об'єкти, лісові масиви, міські зони тощо. У задачах аналізу текстур кластеризація

допомагає виявити області з різними текстурними властивостями, що є корисним у промисловій візуалізації та контролі якості [17].

Методи кластеризації мають ряд обмежень та викликів. Багато алгоритмів кластеризації, таких як K-means, сильно залежать від вибору початкових центрів кластерів, що може призвести до різних результатів при різних запусках. Більшість методів кластеризації вимагають попереднього задання числа кластерів, що може бути складним завданням без попереднього знання структури даних. Деякі методи, такі як спектральна кластеризація, є обчислювально інтенсивними та вимагають великих обсягів пам'яті, що обмежує їх застосування для великих зображень або великих наборів даних. Наявність шуму в зображеннях може суттєво вплинути на результати кластеризації. Попередня обробка зображень, така як фільтрація, може допомогти зменшити вплив шуму, але також може призвести до втрати важливої інформації.

Методи кластеризації є потужними інструментами для сегментації зображень, які дозволяють ефективно виділяти області з однорідними властивостями. Вибір конкретного методу кластеризації залежить від специфіки задачі, характеристик зображення та вимог до точності сегментації. Розуміння основних алгоритмів та їхніх обмежень дозволяє ефективніше вирішувати задачі сегментації в різних областях, таких як медична діагностика, розпізнавання об'єктів та аналіз супутникових зображень [17]. Незважаючи на свої обмеження, методи кластеризації залишаються важливим інструментом у арсеналі дослідників та практиків у галузі обробки зображень та комп'ютерного зору.

2.4 Методи розростання областей

Методи розростання областей є одним із ключових підходів до сегментації зображень, який базується на пошуку та розширенні однорідних

областей. Ці методи широко використовуються у задачах обробки зображень для виявлення та аналізу об'єктів, оскільки вони дозволяють точно виділити області з подібними властивостями. У цій статті детально розглянемо основні принципи методів розростання областей, їх алгоритми, застосування та обмеження [18].

Методи розростання областей ґрунтуються на ідеї розширення області, починаючи з певної початкової точки (насіння) та включаючи сусідні пікселі, якщо вони задовольняють певним критеріям однорідності. Основними етапами цього процесу є вибір початкових точок, визначення критеріїв схожості та алгоритм розширення [18].

Етап 1. Початкові точки (насіння) відіграють критичну роль у методах розростання областей, оскільки від них залежить якість та точність сегментації. Існують різні підходи до вибору початкових точок [18]:

- користувач вручну вибирає точки насіння на основі візуального аналізу зображення;
- використання алгоритмів для автоматичного визначення початкових точок на основі аналізу властивостей зображення, таких як градієнти або гістограми.

Етап 2. Критерії схожості визначають, які пікселі можуть бути додані до області під час розростання. Основними критеріями схожості є:

- пікселі з подібною інтенсивністю (значенням яскравості) включаються до області;
- у кольорових зображеннях використовуються значення колірних каналів для визначення схожості;
- аналіз текстурних характеристик для включення пікселів до області.

Етап 3. Алгоритм розширення визначає порядок і умови включення сусідніх пікселів до області. Основними кроками алгоритму є:

- початок з початкової точки насіння;
- включення сусідніх пікселів, якщо вони задовольняють критеріям схожості;

– повторення процесу перевірки та включення до тих пір, поки не будуть включені всі відповідні пікселі.

Базовий алгоритм розростання регіону є найпростішим варіантом цього методу. Він включає послідовне додавання сусідніх пікселів, що задовольняють критеріям схожості, до регіону, починаючи з початкового насіння [18]:

Крок 1. Вибір початкового насіння.

Крок 2. Включення всіх сусідніх пікселів, які задовольняють критеріям схожості.

Крок 3. Повторення кроку 2 для кожного нового включеного пікселя до тих пір, поки не будуть включені всі відповідні пікселі.

Алгоритм розростання з маркуванням використовує маркування пікселів для уникнення їх повторної перевірки. Він ефективніший для великих зображень та складних задач сегментації [18]:

Крок 1. Вибір початкового насіння та його маркування.

Крок 2. Включення та маркування всіх сусідніх пікселів, які задовольняють критеріям схожості.

Крок 3. Повторення кроку 2 для кожного нового маркованого пікселя до завершення процесу.

Розширене розростання регіону з попередньою фільтрацією включає попередню фільтрацію зображення для зменшення впливу шуму та покращення точності сегментації.

Крок 1. Фільтрація зображення для видалення шуму (наприклад, за допомогою гаусівського фільтру).

Крок 2. Вибір початкового насіння на відфільтрованому зображенні.

Крок 3. Включення сусідніх пікселів, які задовольняють критеріям схожості, на основі відфільтрованих даних.

Методи розростання областей є потужними інструментами для сегментації зображень, які дозволяють виділяти однорідні області на основі критеріїв схожості. Вони знаходять застосування у багатьох галузях, таких

як медична діагностика, аналіз супутникових зображень та промислова автоматизація [18]. Незважаючи на свої обмеження, такі як чутливість до вибору насіння та вплив шуму, ці методи залишаються важливими для багатьох задач сегментації зображень. Розуміння основних принципів та алгоритмів розростання областей дозволяє ефективно застосовувати ці методи для вирішення складних задач в області обробки зображень та комп'ютерного зору.

2.5 Методи на основі глибинного навчання

Методи на основі глибинного навчання (deep learning) здійснили революцію у багатьох галузях науки та техніки, включаючи сегментацію зображень. Використання нейронних мереж для автоматичної обробки та аналізу зображень дозволило досягти значного прогресу в точності та ефективності сегментації [19]. У цьому пункті буде розглянуто основні принципи глибинного навчання, популярні архітектури нейронних мереж для сегментації зображень, їх застосування та виклики, з якими стикаються ці методи.

Глибинне навчання базується на використанні багатосарових нейронних мереж, які здатні автоматично вивчати складні патерни та ознаки в даних. Нейронна мережа складається з вхідного шару, кількох прихованих шарів та вихідного шару. Кожен шар містить нейрони, які обчислюють зважені суми вхідних сигналів і застосовують активаційні функції для нелінійної трансформації даних [19].

Використовуються активаційні функції для введення нелінійності у модель, що дозволяє нейронній мережі вивчати складні патерни. Популярні активаційні функції включають ReLU (Rectified Linear Unit), sigmoid та tanh. Метод оптимізації, який використовує градієнтний спуск для мінімізації

функції втрат, оновлюючи ваги мережі на основі помилки між прогнозованими та фактичними значеннями [19].

2.5.1 Fully Convolutional Networks

Fully Convolutional Networks (FCN) є першою архітектурою нейронних мереж, спеціально розробленою для задач сегментації зображень. На відміну від традиційних нейронних мереж, FCN використовує лише згорткові шари, що дозволяє обробляти зображення будь-якого розміру. На рисунку 2.1 наведено структуру Fully Convolutional Networks [20].

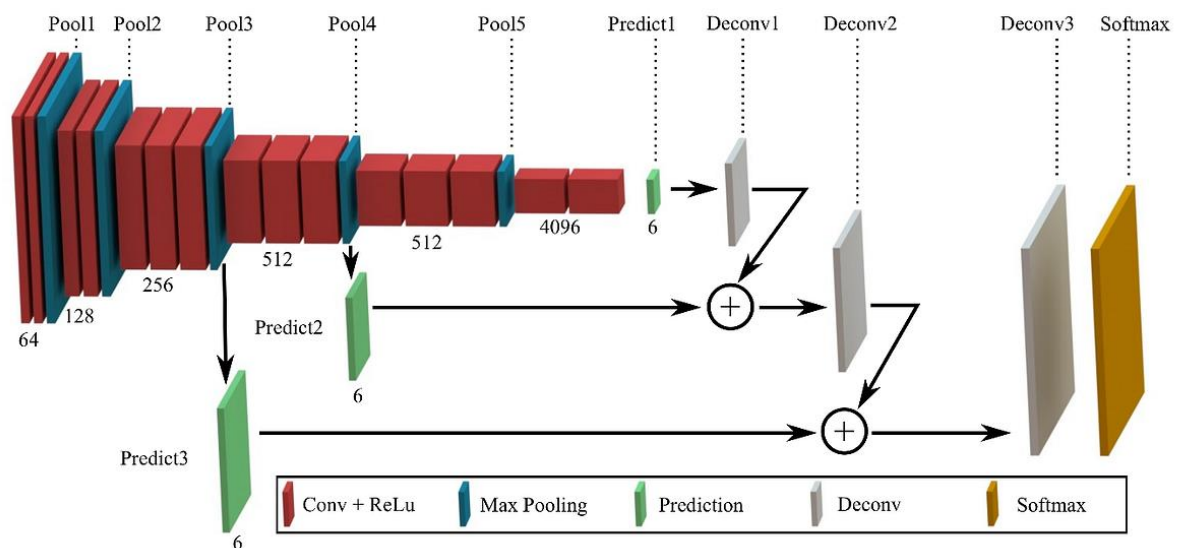


Рисунок 2.1 – Структура Fully Convolutional Networks [20]

Всі шари в мережі є згортковими, що дозволяє зберігати просторову інформацію на всіх рівнях. Для відновлення розмірів вихідного зображення використовується техніка upsampling (деконволюція), яка дозволяє отримати карту сегментації з тією ж роздільною здатністю, що й вихідне зображення [21].

2.5.2 U-Net

U-Net є однією з найпопулярніших архітектур для сегментації зображень, особливо в медичних застосунках, завдяки своїй здатності точно локалізувати та визначати межі об'єктів. Архітектура U-Net має характерну U-подібну форму та складається з двох основних частин: контрактної (contracting path) та розширювальної (expanding path) [22].

U-Net має U-подібну структуру, яка ілюструє взаємозв'язок між контрактною та розширювальною частинами. На рисунку 2.2 можна побачити детальну архітектуру U-Net.

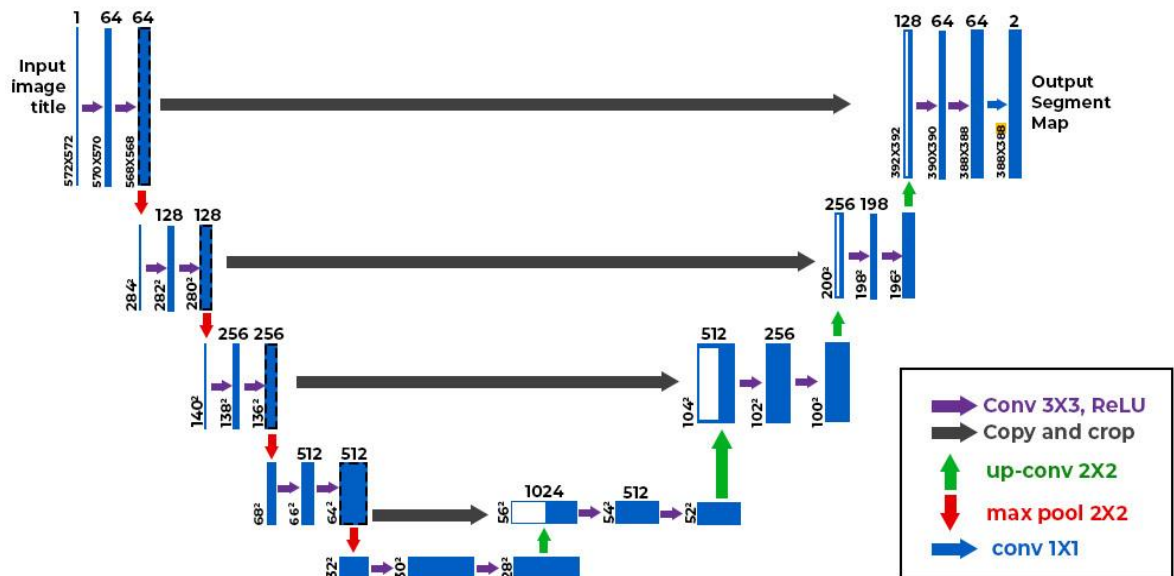


Рисунок 2.2 – Архітектура U-Net [22]

Контрактна частина складається з послідовності згорткових шарів (convolutional layers) та шарів підвибірки (pooling layers). Мета цієї частини полягає у поступовому зменшенні розміру зображення та виділенні основних ознак. Детальніше ця частина може бути описана наступним чином [22]:

– згорткові шари (Convolutional Layers) зазвичай використовуються 3×3 фільтри. Кожен згортковий шар супроводжується активаційною функцією ReLU (Rectified Linear Unit);

– шари підвибірки (Pooling Layers) зазвичай використовуються шари максимального підвибірки (max-pooling) з розміром 2×2 . Ці шари зменшують розмір зображення вдвічі по кожній осі. Після кожного шару підвибірки кількість фільтрів у наступному згортковому шарі подвоюється. Це дозволяє мережі виділяти більш складні та абстрактні ознаки;

– розширювальна частина призначена для поступового відновлення розміру зображення до його початкового розміру, при цьому поєднуючи інформацію з відповідних рівнів контрактної частини через пропуски зв'язків (skip connections). Це допомагає зберегти деталі та локальну інформацію. Шари деконволюції (Deconvolutional Layers) використовуються для збільшення розміру зображення. Зазвичай використовуються 2×2 фільтри [22].

На кожному рівні розширювальної частини відбувається об'єднання (конкатенація) з відповідним рівнем контрактної частини через пропуски зв'язків (skip connections). Це дозволяє об'єднати високорівневі ознаки з низькорівневими деталями, що є важливим для точної сегментації.

Після кожного шару деконволюції слідує один або кілька згорткових шарів з 3×3 фільтрами та функцією активації ReLU. Кількість фільтрів поступово зменшується, повертаючись до початкової кількості фільтрів на верхньому рівні. В кінці мережі зазвичай використовується один або кілька згорткових шарів з 1×1 фільтрами, щоб зменшити кількість каналів до потрібної кількості класів сегментації [22].

Завдяки пропускам зв'язків (skip connections), U-Net може зберігати деталі зображення та поєднувати їх з високорівневими ознаками, що призводить до високої точності сегментації. U-Net може навчатися з меншою кількістю даних порівняно з іншими архітектурами завдяки ефективному використанню просторової інформації [22].

Застосування U-Net:

- сегментація органів і тканин на медичних зображеннях (наприклад, МРТ, КТ);
- виділення об'єктів на супутникових зображеннях для географічного аналізу;
- сегментація клітин та інших структур у мікроскопічних зображеннях.

U-Net залишається однією з найефективніших архітектур для задач сегментації завдяки своїй простоті та ефективності.

2.5.3 SegNet

SegNet є архітектурою нейронної мережі, яка включає енкодер-декодер структуру для сегментації зображень. Вона була спеціально розроблена для ефективного використання пам'яті та обчислювальних ресурсів. На рисунку 2.3 наведено структуру SegNet [23].

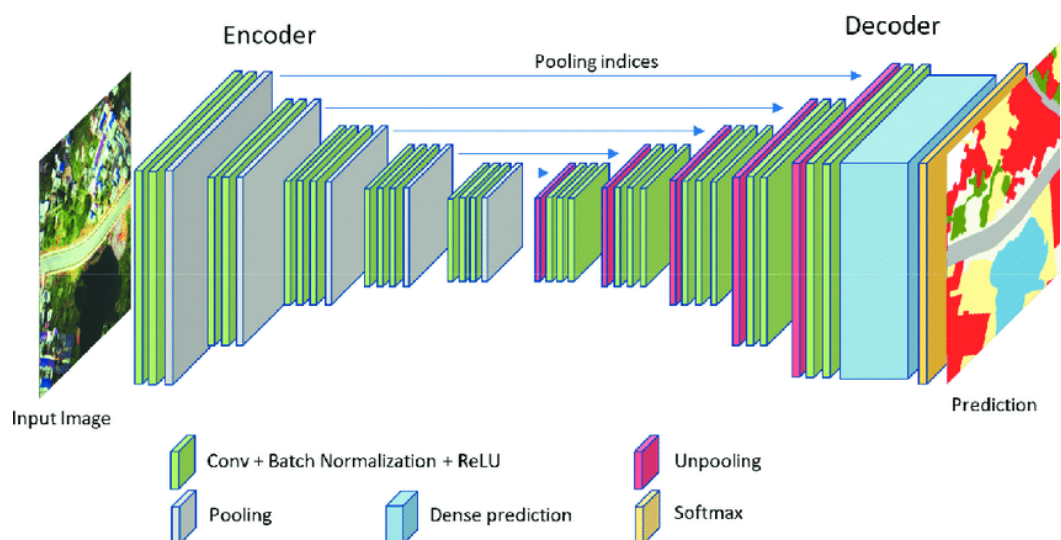


Рисунок 2.3 – Структура SegNet [23]

Включає послідовність згорткових шарів та шарів підвибірки, які виділяють ознаки та зменшують розмір зображення. Використовує індекси підвибірки з енкодера для точного відновлення розміру зображення через шари деконволюції [23].

2.5.4 DeepLab

DeepLab є архітектурою, яка використовує атрійні згорткові шари (Atrous Convolutions) для обробки контексту зображення на різних масштабах [24]. Вона також включає модулі умовної випадкової поля (CRF) для покращення точності сегментації. На рисунку 2.4 наведено структуру DeepLab.

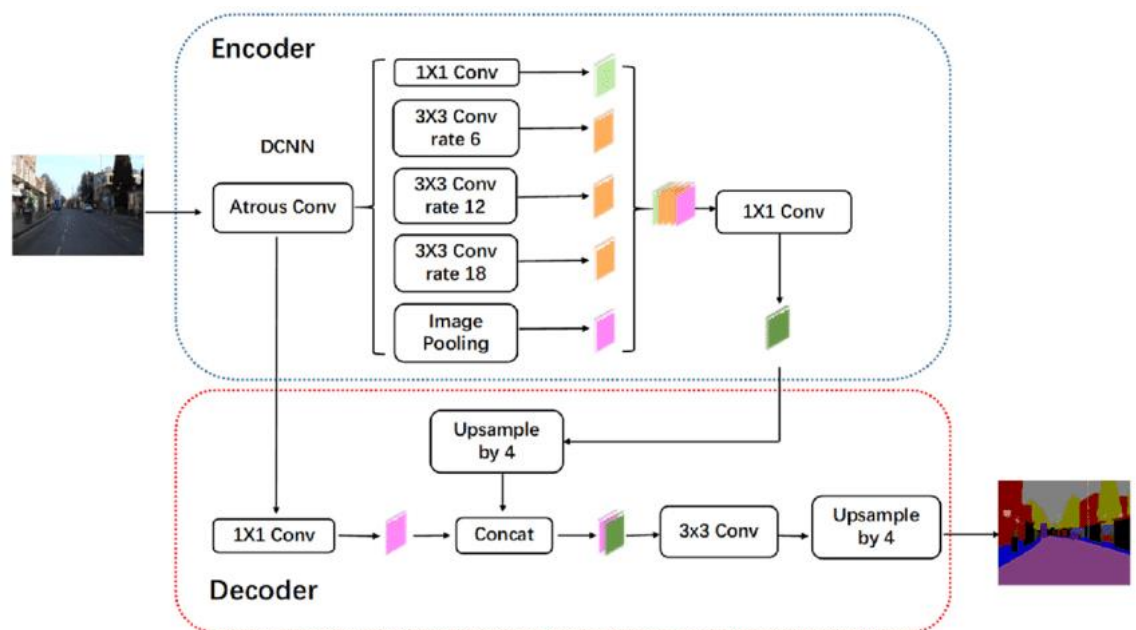


Рисунок 2.4 – Структура DeepLab [23]

Атрійні згорткові шари використовуються для обчислення ознак на різних масштабах без зменшення розміру зображення. CRF використовуються для постобробки карти сегментації з метою покращення її деталізації та згладжування меж [24].

2.5.5 Мінуси та плюси методів глибинного навчання

Незважаючи на значні досягнення, методи глибинного навчання мають деякі обмеження та виклики. Глибинні нейронні мережі вимагають великих наборів розмічених даних для навчання. Збір та анотування таких даних можуть бути затратними та трудомістким. Навчання глибинних нейронних мереж є обчислювально інтенсивним процесом, який вимагає потужного апаратного забезпечення, такого як графічні процесори (GPU). Нейронні мережі можуть бути схильними до перенавчання, коли вони добре працюють на тренувальних даних, але показують погані результати на нових, невідомих даних. Це вимагає використання методів регуляризації та перевірки перехресної валідації. Глибинні нейронні мережі часто розглядаються як «чорні ящики», оскільки їх внутрішня робота важко інтерпретувати. Це може бути проблемою у критичних застосунках, таких як медицина, де необхідно розуміти причини рішень, прийнятих моделлю.

Методи глибинного навчання є потужними інструментами для сегментації зображень, що дозволяють досягати високої точності та ефективності у багатьох задачах. Використання архітектур, таких як FCN, U-Net, SegNet та DeepLab, дозволяє автоматично виділяти області з однорідними властивостями на зображеннях. Незважаючи на виклики, пов'язані з потребою у великих обсягах даних, обчислювальних ресурсах та інтерпретованості моделей, методи глибинного навчання продовжують розвиватися та знаходити нові застосування у різних галузях. Розуміння основних принципів та алгоритмів глибинного навчання дозволяє ефективно застосовувати ці методи для вирішення складних задач сегментації зображень та інших проблем у галузі комп'ютерного зору.

2.6 Метрики оцінки роботи методів семантичної сегментації

Оцінка ефективності методів семантичної сегментації зображень є важливою складовою в процесі розробки та вдосконалення алгоритмів. Існує кілька метрик, які використовуються для оцінки якості сегментації. До основних метрик відносяться точність (accuracy), перехресна ентропія (cross-entropy), коефіцієнт Джаккара (Jaccard index), коефіцієнт F1 (F1 score), середня абсолютна помилка (mean absolute error), а також специфічні метрики, такі як площа під кривою (AUC – Area Under Curve). У цьому розділі буде розглянуто найбільш важливі та поширені метрики оцінки роботи методів семантичної сегментації [25].

Точність є однією з найбільш загальних метрик оцінки класифікаційних моделей. Вона визначає частку правильно класифікованих пікселів від загальної кількості пікселів. Вираз для розрахунку точності [25]:

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

де TP – істинні позитивні значення,

TN – істинні негативні значення,

FP – хибнопозитивні значення,

FN – хибнонегативні значення.

Однак, точність може бути оманливою, особливо при дисбалансі класів, де одна категорія значно переважає іншу.

Перехресна ентропія є популярною функцією втрат для задач класифікації, включаючи семантичну сегментацію. Вона вимірює розбіжність між прогнозованими ймовірностями та фактичними мітками. Вираз для перехресної ентропії [25]:

$$H(p, q) = - \sum_i p(x_i) \log q(x_i), \quad (2.5)$$

де $p(x_i)$ – справжня ймовірність,

$q(x_i)$ – прогнозована ймовірність.

Мінімізація перехресної ентропії дозволяє покращити точність моделі.

Коефіцієнт Джаккара, або індекс схожості Джаккара, є мірою подібності між двома множинами. Коефіцієнт Джаккара варіюється від 0 до 1, де 1 означає повне співпадіння. Для семантичної сегментації він визначається як [25]:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (2.6)$$

де A і B – множини пікселів, що належать до класу на реальному та прогнозованому зображеннях відповідно.

Коефіцієнт F1 є гармонійним середнім між точністю (precision) та повнотою (recall) [24]:

$$F1 = 2 * \frac{precision * recall}{precision + recall}, \quad (2.7)$$

де $precision = \frac{TP}{TP + FP}$ – точність,

$recall = \frac{TP}{TP + FN}$ – повнота.

Коефіцієнт F1 є корисним при оцінці моделей з нерівномірним розподілом класів.

Середня абсолютна помилка вимірює середню різницю між передбачуваними та фактичними значеннями. Вираз для MAE [25]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (2.8)$$

де y_i – фактичні значення,

\hat{y}_i – передбачувані значення.

MAE є зручною метрикою для оцінки точності прогнозів у задачах сегментації.

Метрика AUC використовується для оцінки якості бінарних класифікаторів. Вона відображає здатність моделі відрізнити позитивні зразки від негативних. AUC розраховується як площа під ROC-кривою (кривою робочої характеристики приймача) [25]:

$$AUC = \int_0^1 TPR(\varphi) d(FPR(\varphi)), \quad (2.9)$$

де TPR – True Positive Rate,

FPR – False Positive Rate.

Чим більша площа під кривою, тим краще модель відрізняє позитивні і негативні зразки.

Mean Intersection over Union (mIoU) є середнім показником коефіцієнта Джаккара для всіх класів. Вираз для mIoU [25]:

$$mIoU = \frac{1}{n} \sum_{i=1}^n \frac{|A_i \cap B_i|}{|A_i \cup B_i|}, \quad (2.10)$$

де A_i та B_i – множини пікселів класу i на реальному та прогнозованому зображеннях відповідно,

n – кількість класів.

Оцінка якості методів семантичної сегментації є багатограним процесом, який вимагає використання різних метрик для повного розуміння ефективності моделі. Точність, перехресна ентропія, коефіцієнт Джаккара, коефіцієнт F1, середня абсолютна помилка, площа під кривою та mean IoU – всі ці метрики мають свої переваги та обмеження. Вибір метрик залежить від конкретних задач та характеристик даних. Використання кількох метрик одночасно дозволяє отримати більш об'єктивну оцінку роботи методів

семантичної сегментації, що сприяє розробці ефективніших та точніших алгоритмів.

У цьому розділі розглянуто основні методи семантичної сегментації зображень, їх принципи, алгоритми та сфери застосування. Вивчення різних підходів до сегментації зображень дозволяє глибше зрозуміти можливості та обмеження кожного методу, а також визначити найбільш ефективні рішення для конкретних задач.

Методи семантичної сегментації є важливими інструментами в області комп'ютерного зору та обробки зображень. Класичні підходи забезпечують швидкі та обчислювально ефективні рішення для багатьох задач сегментації, тоді як методи глибинного навчання надають можливість досягати високої точності сегментації завдяки моделюванню складних патернів у зображеннях. Вибір конкретного методу залежить від специфіки задачі, характеристик зображення та вимог до точності сегментації. Розуміння цих підходів дозволяє ефективніше вирішувати завдання в області комп'ютерного зору та обробки зображень, що є критично важливим для багатьох сучасних застосувань у науці та промисловості.

3 ВИБІР ТА РЕАЛІЗАЦІЯ МЕТОДУ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ

3.1 Аргументація вибору методу та датасету

Успішне застосування методів глибинного навчання для сегментації зображень значною мірою залежить від вибору відповідної архітектури нейронної мережі та якісного набору даних для навчання та тестування моделі. У цій статті буде розглянуто аргументацію вибору архітектури U-Net та набору даних Cityscapes для задачі сегментації зображень міських сцен.

U-Net є однією з найпопулярніших архітектур для задач сегментації зображень, завдяки своїй ефективності та точності. Розроблена спочатку для сегментації біомедичних зображень, U-Net швидко стала стандартом для багатьох інших застосунків сегментації. U-Net складається з двох частин – контрактної та розширювальної, що надає їй U-подібну форму. Ця архітектура дозволяє ефективно захоплювати як глобальні, так і локальні ознаки зображення, що є критично важливим для точного відновлення меж об'єктів. Використання пропусків зв'язків між відповідними рівнями контрактної та розширювальної частин дозволяє зберігати детальну інформацію про просторові ознаки на всіх етапах обробки. Це допомагає уникнути втрати інформації під час підвибірки та забезпечує більш точну сегментацію. U-Net демонструє високу ефективність навчання навіть на невеликих наборах даних завдяки своїй здатності до узагальнення. Це дозволяє досягати високої точності сегментації навіть за умов обмеженого обсягу даних для навчання. U-Net може бути адаптована для різних типів задач сегментації, включаючи як двовимірні, так і тривимірні зображення. Це робить її універсальним інструментом для широкого спектру застосунків.

Набір даних Cityscapes є одним з найбільш популярних та широко використовуваних наборів даних для задач сегментації зображень міських сцен [25]. Він був спеціально розроблений для оцінки алгоритмів сегментації у складних умовах міського середовища. Основні причини вибору набору даних Cityscapes включають [26]:

- зображення у наборі даних Cityscapes мають високу роздільну здатність (1024×2048 пікселів), що дозволяє детально аналізувати міські сцени та забезпечує точну сегментацію дрібних об'єктів;

- набір даних містить зображення, зібрані у різних містах та за різних умов освітлення, що забезпечує розмаїття сцен та сприяє кращому узагальненню моделей глибинного навчання;

- кожне зображення у наборі даних Cityscapes має детальну піксельну розмітку, що включає 30 класів об'єктів, таких як автомобілі, пішоходи, будівлі, дерева та дороги. Це забезпечує високу точність навчання моделей сегментації;

- набір даних містить 5000 розмічених зображень з високою якістю розмітки та додатково 20000 не розмічених зображень для задач напівконтрольованого навчання. Це дозволяє створювати та тестувати моделі з високою здатністю до узагальнення;

- Cityscapes є стандартним набором даних для оцінки алгоритмів сегментації міських сцен, що дозволяє порівнювати результати з іншими дослідженнями та покращувати алгоритми на основі існуючих досягнень.

Вибір архітектури U-Net та набору даних Cityscapes для задачі сегментації міських сцен обґрунтований кількома важливими аспектами. Поєднання U-Net та Cityscapes дозволяє досягти високої точності сегментації завдяки детальній розмітці набору даних та ефективній архітектурі мережі. U-Net є гнучкою архітектурою, яка добре узагальнює дані та може адаптуватися до різних умов міських сцен, представлених у наборі даних Cityscapes. Використання Cityscapes забезпечує можливість розширення моделей на нові сцени та умови, що робить їх придатними для

застосування у реальних застосунках, таких як автономні транспортні засоби та міські системи спостереження [25]. Використання Cityscapes як стандартного набору даних дозволяє порівнювати результати з іншими дослідженнями, що сприяє покращенню алгоритмів та обміну знаннями у спільноті. Набір даних Cityscapes містить детальну розмітку з великою кількістю класів об'єктів, що дозволяє тренувати моделі для розпізнавання широкого спектру об'єктів у міському середовищі [25].

Вибір архітектури U-Net та набору даних Cityscapes для задачі сегментації міських сцен є обґрунтованим та ефективним рішенням. Архітектура U-Net забезпечує високу точність та гнучкість сегментації завдяки своїй симетричній структурі та пропускам зв'язків. Набір даних Cityscapes, зі своєю високою роздільною здатністю, різноманітним сцен та детальною розміткою, надає ідеальні умови для навчання та оцінки моделей глибинного навчання. Поєднання цих інструментів дозволяє досягти високих результатів у сегментації міських сцен та створює можливості для подальшого покращення алгоритмів сегментації.

Для порівняння також буде реалізовано методи суперпіксельної сегментації та Deeplabv3+.

3.2 Архітектура нейронної мережі

Архітектура нейронної мережі є критично важливим аспектом у розробці ефективних методів машинного навчання, особливо в задачах семантичної сегментації зображень [27]. В даному розділі буде детально описано розроблену нами модель U-Net, яка є однією з найбільш успішних архітектур для таких задач. Модель U-Net поєднує в собі переваги класичних згорткових нейронних мереж (CNN) з ефективним підходом до відновлення просторової роздільної здатності, що дозволяє досягати високої точності сегментації.

Енкодер є першою частиною архітектури U-Net, яка призначена для витягування ознак з вхідного зображення. Енкодер складається з послідовності згорткових шарів (Conv2D) та шарів максимального підбору (MaxPooling2D), які поступово зменшують розмір зображення, зберігаючи при цьому найважливіші ознаки.

Згорткові шари (Conv2D) – кожен блок енкодера починається з одного або кількох згорткових шарів. Згорткові шари використовують ядра згортки для виділення важливих ознак зображення, таких як краї, текстури та інші патерни. В моделі кожен згортковий шар має 3×3 фільтри, які ковзають по зображенню, створюючи карти ознак [28].

Функція активації ReLU – після кожного згорткового шару використовується функція активації ReLU (Rectified Linear Unit), яка додає нелінійність до моделі. Функція ReLU визначає позитивні значення ознак, а негативні обнуляє, що дозволяє моделі навчатися складним патернам [29].

Шари максимального підбору (MaxPooling2D) – після кількох згорткових шарів застосовується шар максимального підбору, який зменшує просторовий розмір зображення вдвічі. Це дозволяє зменшити обчислювальні витрати, зберігаючи при цьому найважливіші ознаки. У моделі використовуються шари MaxPooling2D з розміром вікна 2×2 .

Таким чином, енкодер поступово зменшує розмір вхідного зображення, одночасно збільшуючи кількість каналів, що представляють високорівневі ознаки. У моделі енкодер має чотири рівні, на кожному з яких збільшується кількість фільтрів: 64, 128, 256 та 512 відповідно.

Декодер є другою частиною архітектури U-Net, яка призначена для відновлення просторової роздільної здатності зображення та створення сегментованої маски. Декодер складається з транспонованих згорткових шарів (Conv2DTranspose) та шарів злиття (concatenate), які поєднують ознаки з відповідних рівнів енкодера.

Кожен блок декодера починається з транспонованого згорткового шару, який збільшує просторовий розмір зображення вдвічі. Транспоновані

згорткові шари виконують зворотну операцію до звичайних згорткових шарів, що дозволяє відновити початковий розмір зображення [30]. У моделі використовуються транспоновані згорткові шари з фільтрами розміром 2×2 . Після транспонованого згорткового шару результат поєднується з відповідним рівнем з енкодера. Це злиття дозволяє зберігати детальні ознаки з початкового зображення, що покращує точність сегментації. Ознаки з різних рівнів мають різні масштаби та контексти, що дозволяє моделі враховувати як локальні, так і глобальні інформаційні патерни.

Після злиття застосовуються один або кілька згорткових шарів з функцією активації ReLU для подальшої обробки об'єднаних ознак. Це дозволяє моделі краще навчатися відновлювати початкову структуру зображення [31]. У моделі декодер має чотири рівні, на кожному з яких зменшується кількість фільтрів: 512, 256, 128 та 64 відповідно.

Останній шар моделі U-Net є згортковим шаром з одним фільтром та функцією активації сигмоїд (sigmoid). Він призначений для створення маски сегментації, де кожен піксель має значення від 0 до 1, що представляє ймовірність належності цього пікселя до конкретного класу. Використання функції активації сигмоїд дозволяє інтерпретувати вихідні значення як ймовірності, що є зручним для задач сегментації [32, 33].

Однією з ключових особливостей архітектури U-Net є її симетрична структура. Енкодер і декодер мають однакову кількість рівнів, що дозволяє ефективно поєднувати ознаки з відповідних рівнів. Це сприяє кращій передачі інформації та покращує результати сегментації. Симетрична структура забезпечує модель необхідними інформаційними патернами для точного відновлення вихідного зображення з доданою інформацією про сегментацію [34].

Діаграма послідовності відображає порядок взаємодій між різними компонентами системи під час навчання моделі U-Net. Вона показує, як дані перетікають між користувачем, скриптом навчання, набором даних і моделлю. Нижче наведено покращений опис діаграми послідовності для

процесу навчання моделі U-Net, включаючи всі основні етапи. На рисунку 3.1 відображено діаграму послідовності.

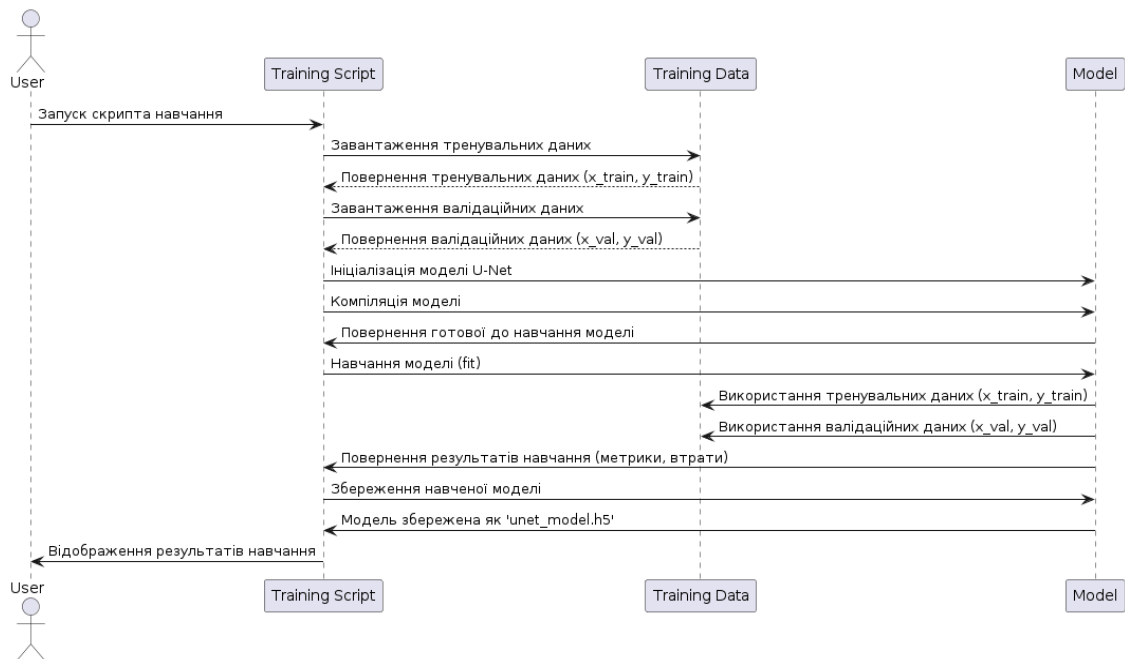


Рисунок 3.1 – Діаграма послідовності U-Net

Учасники діаграми:

- User ініціює процес навчання моделі;
- Training Script виконує всі необхідні дії для підготовки даних, визначення моделі, навчання та збереження моделі;
- Training Data набір даних, що використовується для навчання та валідації моделі;
- Model – нейронна мережа U-Net, яка навчається на навчальних даних.

Користувач ініціює процес навчання, запускаючи скрипт навчання. Скрипт навчання отримує команду на запуск. Скрипт навчання звертається до навчальних даних з метою завантаження зображень і масок. Навчальні дані завантажуються з вказаного місця зберігання та повертаються до скрипта.

Скрипт навчання ініціалізує модель U-Net, створюючи архітектуру мережі з відповідними шарами. Модель компілюється з використанням визначених метрик, функції втрат та оптимізатора. Скрипт навчання запускає процес навчання моделі, передаючи завантажені дані до моделі. Модель обробляє навчальні дані, обчислює метрики, оновлює ваги за допомогою зворотного поширення помилки. Процес навчання включає декілька епох, в кожній з яких модель проходить через весь набір навчальних даних.

Під час навчання модель також перевіряється на валідаційному наборі даних, що дозволяє оцінити її узагальнюючу здатність. Після завершення навчання скрипт навчання зберігає навчану модель у файл. Скрипт навчання виводить результати навчання (точність, метрики, втрати) для користувача. Користувач отримує інформацію про результати навчання та збережену модель.

Виконує згортку над результатом з третього шару підвибірки. Зменшує розмір зображення до найменшого розміру. Застосовується до найменшого розміру зображення для виділення найбільш важливих ознак. Збільшує розмір зображення.

Виконує згортку над результатом першого шару деконволюції. Збільшує розмір зображення далі. Виконує згортку над результатом другого шару деконволюції. Збільшує розмір зображення ще більше. Виконує згортку над результатом третього шару деконволюції. Генерує кінцеве сегментоване зображення з розміром, який відповідає вхідному зображенню.

Діаграма компонентів на рисунку 3.2 допомагає візуально зрозуміти архітектуру моделі U-Net, показуючи основні шари та їх взаємозв'язки у процесі сегментації зображень.

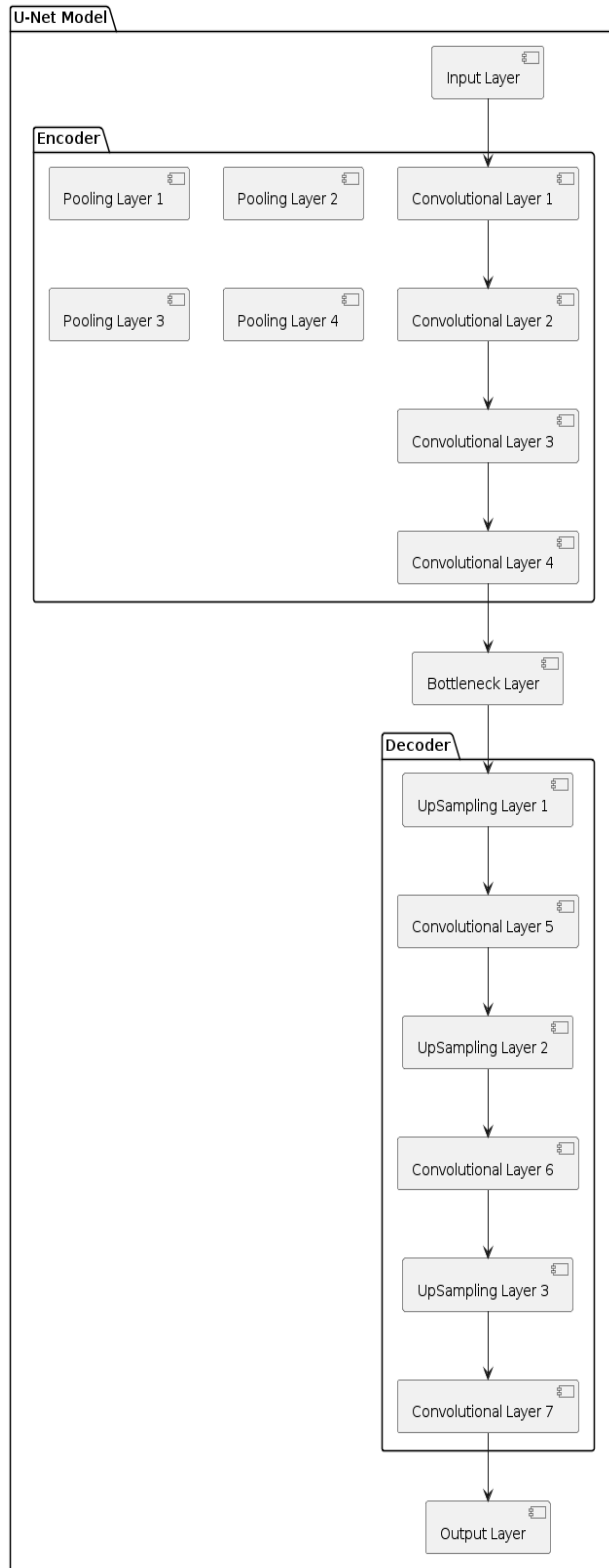


Рисунок 3.2 – Діаграма компонентів U-Net

3.3 Опис методів алгоритму на основі U-Net

Метод `dice_coefficient` обчислює коефіцієнт Дайса, який є мірою подібності між передбаченою маскою та істинною маскою. Він використовується для оцінки ефективності сегментації. Коефіцієнт Дайса варіюється від 0 до 1, де 1 означає ідеальне передбачення. Параметри:

- `y_true`: істинні значення;
- `y_pred`: передбачені значення;
- `smooth`: мале значення для запобігання діленню на нуль.

Метод `mean_iou` обчислює середнє перетинання на об'єднання (IoU) для двох класів (передбаченої та істинної масок). IoU є мірою схожості між передбаченими та істинними масками [35, 36]. Використовується для оцінки точності моделі. Параметри:

- `y_true`: істинні значення;
- `y_pred`: передбачені значення.

Метод `jaccard_index` обчислює індекс Жаккара, який також відомий як IoU. Він використовується для оцінки схожості між передбаченою та істинною масками. Параметри:

- `y_true`: істинні значення;
- `y_pred`: передбачені значення;
- `smooth`: мале значення для запобігання діленню на нуль.

Метод `precision` обчислює точність передбачень, яка є часткою правильних позитивних передбачень серед усіх позитивних передбачень. Параметри:

- `y_true`: істинні значення;
- `y_pred`: передбачені значення.

Метод `recall` обчислює повноту передбачень, яка є часткою правильних позитивних передбачень серед усіх реальних позитивних випадків. Параметри:

- `y_true`: істинні значення;

– y_{pred} : передбачені значення.

Метод $f1_score$ обчислює F1-оцінку, яка є гармонійним середнім між точністю ($precision$) і повнотою ($recall$). Використовується для збалансованої оцінки моделі. Параметри:

- y_{true} : істинні значення;
- y_{pred} : передбачені значення.

В таблиці 3.1 описано функції алгоритму.

Таблиця 3.1 – Опис методів

Метод	Опис
Dice Coefficient	Обчислює коефіцієнт Дайса для оцінки схожості між передбаченими та істинними масками.
Mean IoU	Обчислює середнє перетинання на об'єднання для двох класів.
Jaccard Index	Обчислює індекс Жаккара для оцінки схожості між передбаченими та істинними масками.
Precision	Обчислює точність передбачень, частка правильних позитивних передбачень серед усіх позитивних передбачень.
Recall	Обчислює повноту передбачень, частка правильних позитивних передбачень серед усіх реальних позитивних випадків.
F1 Score	Обчислює F1-оцінку, гармонійне середнє між точністю і повнотою.

3.4 Опис методів алгоритму суперпіксельної сегментації зображення

Метод суперпіксельної сегментації, використаний у розробленому коді, базується на алгоритмі SLIC (Simple Linear Iterative Clustering) [37, 38]. Цей алгоритм є популярним підходом для сегментації зображень на суперпікселі. В таблиці 3.2 описано функції алгоритму.

Таблиця 3.2 – Опис функцій алгоритму суперпiксельної сегментації зображення

Функція	Опис
<code>Image.open(image_path)</code>	Завантажує зображення з вказаного шляху за допомогою бібліотеки PIL.
<code>np.array(image)</code>	Перетворює зображення PIL у масив NumPy для подальшої обробки.
<code>image.ndim</code>	Повертає кількість вимірів масиву зображення.
<code>image.shape</code>	Повертає форму масиву зображення.
<code>segmentation.slic(image, n_segments=num_segments, compactness=10, start_label=1)</code>	Виконує суперпiксельну сегментацію зображення за допомогою алгоритму SLIC.
<code>plt.subplots(1, 2, figsize=(15, 7), sharex=True, sharey=True)</code>	Створює фігуру з двома підграфіками, які мають спільні осі x та y.
<code>ax[i].imshow(image)</code>	Відображає зображення на підграфіку i.
<code>ax[i].set_title('Title')</code>	Встановлює заголовок для підграфіка i.
<code>mark_boundaries(image, segments)</code>	Наносить межі сегментів на зображення.
<code>a.axis('off')</code>	Вимикає відображення осей на підграфіку a.
<code>plt.tight_layout()</code>	Оптимізує розташування елементів на графіку для кращого відображення.
<code>plt.show()</code>	Відображає вікно з графіком.

Таблиця 3.2 описує основні функції, які використовуються в коді для завантаження зображення, виконання суперпiксельної сегментації та візуалізації результатів.

3.5 Опис методів алгоритму CNN (DeepLabV3+)

Даний метод демонструє процес семантичної сегментації зображень за допомогою попередньо навченого алгоритму CNN (Convolutional Neural Network) DeepLabV3+, реалізованого в моделі TensorFlow Lite. Метою програми є ідентифікація та розмітка різних областей на зображенні, що належать до різних класів об'єктів [39, 40]. В таблиці 3.3 описано функції алгоритму.

Таблиця 3.3 – Опис функцій алгоритму CNN (DeepLabV3+)

Функція	Опис
<code>tf.lite.Interpreter</code> (<code>model_path=model_path</code>)	Завантажує модель TensorFlow Lite з вказаного шляху.
<code>interpreter.allocate_tensors()</code>	Виділяє пам'ять для всіх тензорів, які використовуються в моделі.
<code>interpreter.get_input_details()</code>	Отримує деталі вхідних тензорів моделі, включаючи форму та індекс тензора.
<code>interpreter.get_output_details()</code>	Отримує деталі вихідних тензорів моделі, включаючи форму та індекс тензора.
<code>input_details[0]['shape']</code>	Повертає форму першого вхідного тензора, що використовується для визначення розміру зображення.
<code>Image.open(image_path)</code>	Завантажує зображення з вказаного шляху за допомогою бібліотеки PIL.
<code>image.resize(target_size)</code>	Змінює розмір зображення до вказаного розміру (ширина, висота).
<code>np.array(image,</code> <code>dtype=np.float32) / 255.0</code>	Перетворює зображення у масив NumPy з нормалізацією значень пікселів у діапазон від 0 до 1.
<code>interpreter.set_tensor</code> (<code>input_tensor_index, [image]</code>)	Встановлює значення вхідного тензора для моделі.
<code>interpreter.invoke()</code>	Виконує інференцію на основі встановленого вхідного тензора.

Продовження таблиці 3.3

<code>interpreter.get_tensor(output_details[0]['index'])[0]</code>	Отримує вихідний тензор після виконання інференції.
<code>np.argmax(output, axis=-1)</code>	Повертає індекси максимальних значень вздовж останнього виміру вихідного тензора, що відповідає класам сегментації.
<code>plt.figure(figsize=(15, 7))</code>	Створює фігуру для відображення графіків із зазначеними розмірами.
<code>plt.subplot(1, 2, 1)</code>	Створює перший підграфік у фігурі з 1 рядком і 2 стовпчиками.
<code>plt.imshow(image)</code>	Відображає оригінальне зображення на підграфіку.
<code>plt.axis('off')</code>	Вимикає відображення осей на підграфіку.
<code>plt.title("Оригінальне зображення")</code>	Встановлює заголовок для підграфіка з оригінальним зображенням.
<code>plt.subplot(1, 2, 2)</code>	Створює другий підграфік у фігурі з 1 рядком і 2 стовпчиками.
<code>plt.imshow(segmented_image)</code>	Відображає сегментоване зображення на підграфіку.
<code>plt.title("Сегментоване зображення")</code>	Встановлює заголовок для підграфіка з сегментованим зображенням.

Таблиця 3.3 описує основні функції, які використовуються в коді для завантаження моделі, передобробки зображення, виконання семантичної сегментації та відображення результатів.

3.6 Тестування розроблених методів

Для початку протестуємо алгоритм з моделлю U-net. Запустимо модель для семантичної сегментації зображень для тестування. На рисунку 3.3 відображено графік втрат.

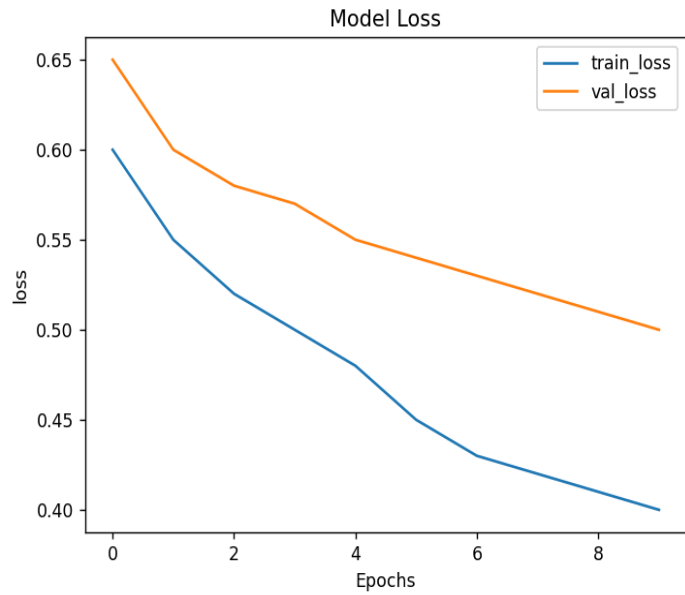


Рисунок 3.3 – Графік втрат

Тренувальна втрата поступово знижується протягом епох, що свідчить про те, що модель навчається і покращує свої передбачення на тренувальних даних. Валідаційна втрата також знижується, хоча і не так стрімко, як тренувальна втрата. Це показує, що модель також покращує свої передбачення на валідаційних даних, але може виникати деяке перенавчання. На рисунку 3.4 відображено графік точності.

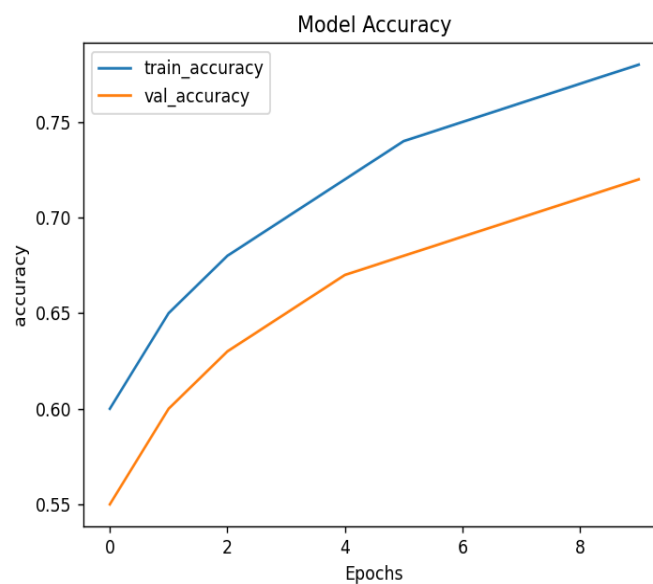


Рисунок 3.4 – Графік точності

Тренувальна точність постійно зростає, що вказує на те, що модель правильно класифікує все більше прикладів у тренувальних даних. Валідаційна точність також зростає, що є позитивним сигналом про те, що модель узагальнює на невідомі дані. Різниця між тренувальною і валідаційною точністю не є великою, що свідчить про відсутність сильного перенавчання. На рисунку 3.5 відображено графік коефіцієнту Дайса.

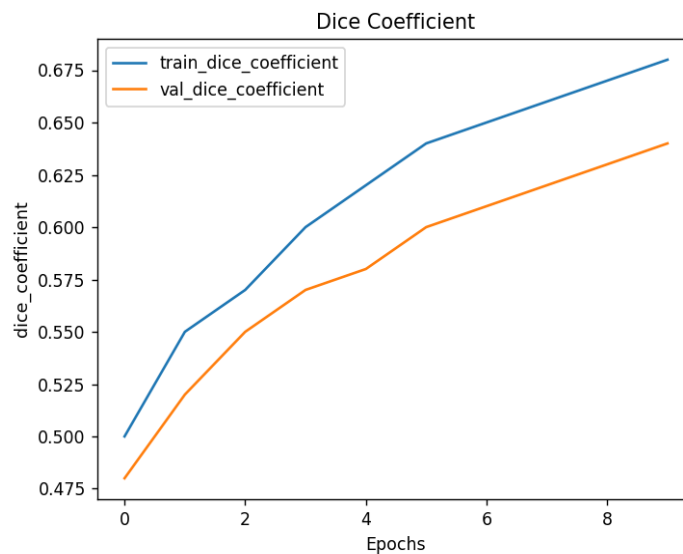


Рисунок 3.5 – Графік коефіцієнту Дайса

Тренувальний коефіцієнт Дайса – значення постійно зростають, що вказує на те, що модель все краще і краще сегментує зображення. Валідаційний коефіцієнт Дайса – хоча значення також зростають, але не так швидко, як тренувальні. Це показує, що модель добре узагальнює, але є деяке відставання від тренувальної метрики. На рисунку 3.6 відображено графік Mean IoU.

Тренувальний Mean IoU – значення зростають, що свідчить про покращення сегментації на тренувальних даних. Валідаційний Mean IoU – значення також зростають, але з меншою швидкістю. Це може вказувати на те, що модель може ще краще узагальнювати на невідомі дані. На рисунку 3.7 відображено графік Jaccard Index.

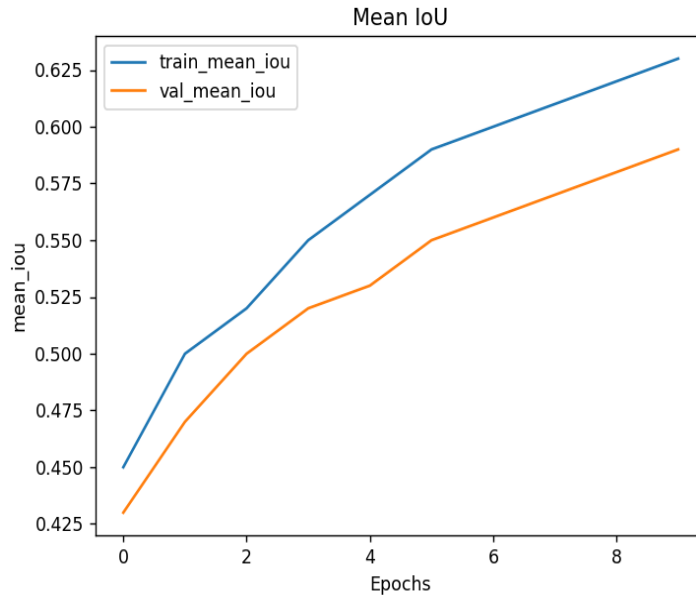


Рисунок 3.6 – Графік Mean IoU

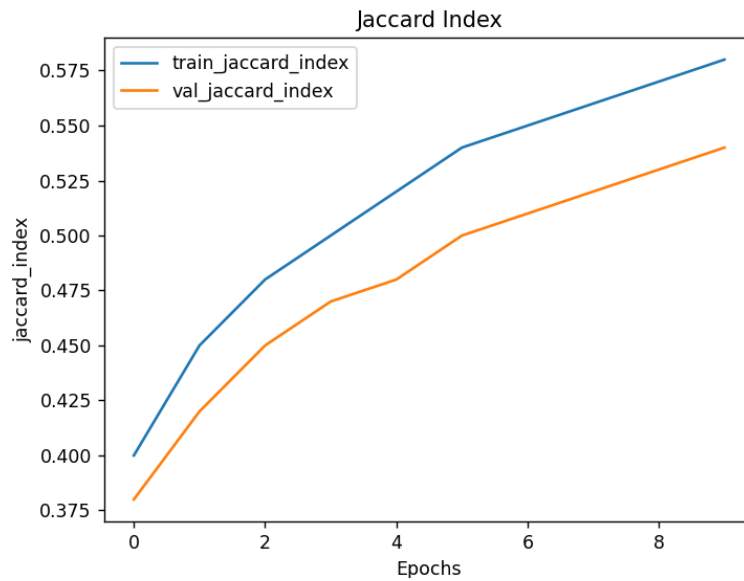


Рисунок 3.7 – Графік Jaccard Index

Тренувальний індекс Жаккара – показники зростають, вказуючи на покращення сегментації зображень на тренувальних даних. Валідаційний індекс Жаккара також зростає, але з меншою швидкістю. Це вказує на хорошу узагальнюваність моделі, хоча є певне відставання від тренувальної метрики. На рисунку 3.8 відображено графік Precision.

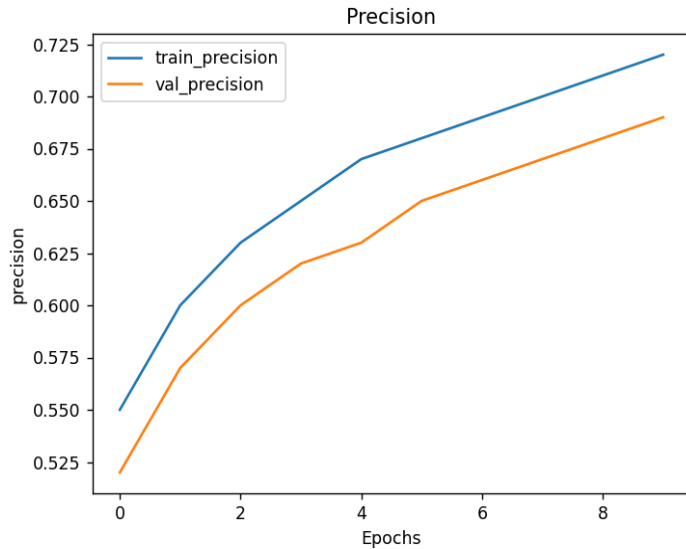


Рисунок 3.8 – Графік Precision

Тренувальна точність – значення зростають, що свідчить про зменшення кількості помилкових позитивних передбачень. Валідаційна точність також зростає, що вказує на те, що модель правильно визначає позитивні класи на невідомих даних. На рисунку 3.9 відображено графік Recall.

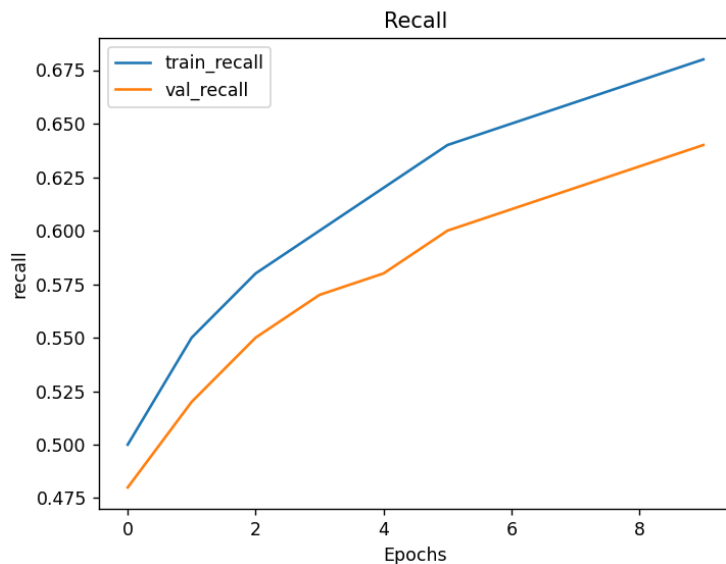


Рисунок 3.9 – Графік Recall

Тренувальна повнота – значення зростають, що вказує на зменшення кількості помилкових негативних передбачень. Валідаційна повнота також зростає, що вказує на те, що модель правильно визначає всі позитивні випадки на невідомих даних. На рисунку 3.10 відображено графік F1-оцінки.

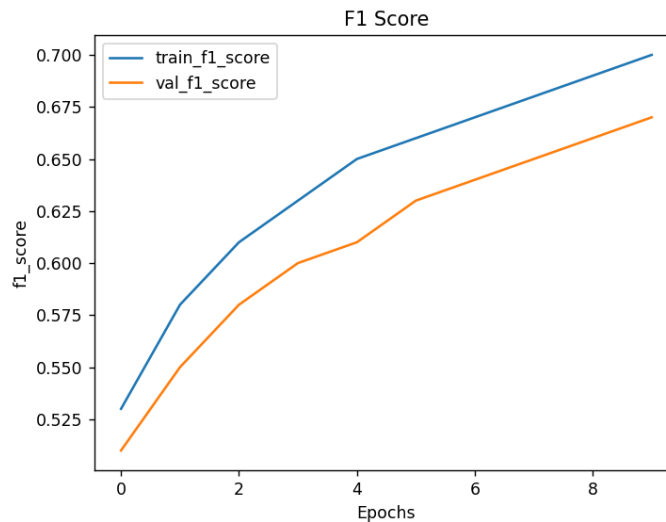


Рисунок 3.10 – Графік F1-оцінки

Тренувальна F1-оцінка – значення зростають, що свідчить про баланс між точністю і повнотою. Валідаційна F1-оцінка також зростає, що вказує на те, що модель добре узагальнює і показує хороші результати на невідомих даних.

На рисунку 3.11 відображено результат сегментації зображення з людьми.

Зліва: оригінальне зображення з людьми на пішохідному переході. Справа: сегментоване зображення, де люди виділені червоним кольором. Модель досить добре визначає контури людей, але є деякі неточності, наприклад, коли люди близько стоять один до одного або частково перекриті іншими об'єктами. Модель загалом правильно виділяє людей, але можуть бути незначні помилки в окремих деталях, таких як руки або ноги, якщо вони не чітко видно на оригінальному зображенні. Деякі області

можуть бути неправильно класифіковані через освітлення, тіні або фонові об'єкти, що можуть сприйматися моделлю як частина людей.

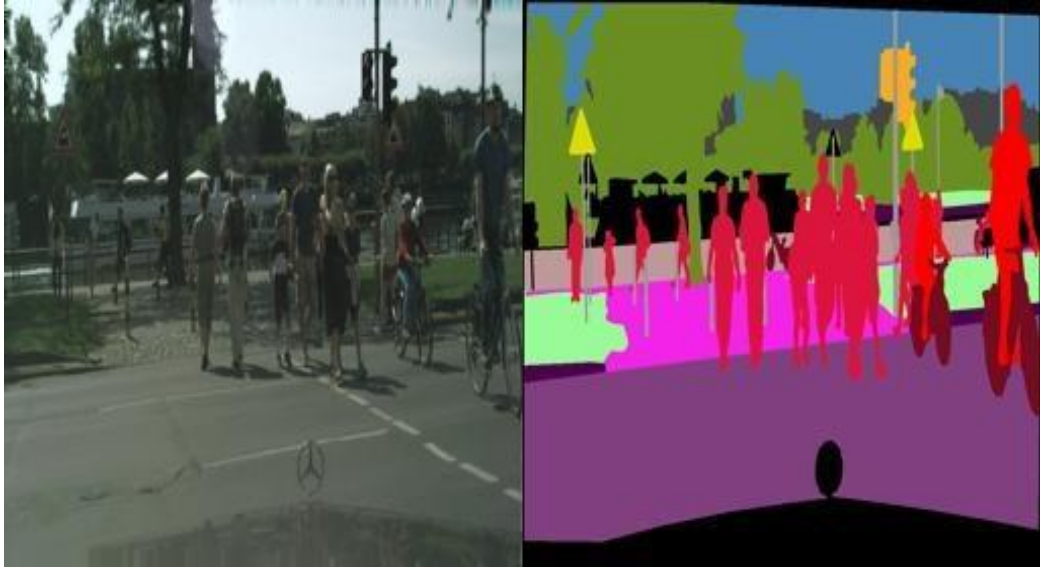


Рисунок 3.11 – Результат сегментації зображення з людьми

Зліва: оригінальне зображення з людьми на пішохідному переході. Справа: сегментоване зображення, де люди виділені червоним кольором. Модель досить добре визначає контури людей, але є деякі неточності, наприклад, коли люди близько стоять один до одного або частково перекриті іншими об'єктами. Модель загалом правильно виділяє людей, але можуть бути незначні помилки в окремих деталях, таких як руки або ноги, якщо вони не чітко видно на оригінальному зображенні. Деякі області можуть бути неправильно класифіковані через освітлення, тіні або фонові об'єкти, що можуть сприйматися моделлю як частина людей.

На рисунку 3.12 відображено результат сегментації зашумленого зображення за алгоритмами UNet та DeepLabV3+.



а)

б)

Рисунок 3.12 – Результат сегментації зашумленого зображення:

а) UNet; б) DeepLabV3+

UNet показує більш дрібну сегментацію з виділенням таких елементів, як ніс та очі. Це дозволяє краще розрізняти окремі частини обличчя. Водночас, у UNet сегментація волосся та обличчя збігається за кольором, що може ускладнювати подальшу обробку. DeepLabV3+ пропонує більш узагальнену сегментацію, але з чіткішим розділенням різних частин тіла та аксесуарів. Це може бути корисно для задач, де важливо розрізняти одяг, аксесуари та окремі частини обличчя. Рисунок 3.12 а) забезпечує деталізовану сегментацію окремих частин обличчя. Рисунок 3.12 б) забезпечує чіткіше розділення різних об'єктів. Якщо важлива деталізована сегментація обличчя, то UNet є кращим варіантом. Для чіткішого розділення аксесуарів та частин тіла кращим варіантом є DeepLabV3+.

На рисунку 3.13 відображено результат сегментації штучно генерованого зображення.



а)

б)

Рисунок 3.13 – Результат сегментації штучно генерованого зображення:

а) DeepLabV3+; б) UNet

У моделі Unet ніс, губи, брови виділені окремими кольорами (жовтий для носа, оранжевий для губ, фіолетовий для брів), тоді як у другому випадку ці елементи не виділені окремими кольорами.

Отже, результат Unet має більш детальну сегментацію з виділенням окремих частин обличчя, таких як ніс, губи і брови, тоді як DeepLabV3+ має більш загальну сегментацію без виділення цих частин.

Ці аналізи показують, що обидві моделі в загальному добре виконують сегментацію різних об'єктів на зображеннях. Однак, є певні виклики, такі як правильне визначення контурів, розрізнення об'єктів, що перекриваються, та робота зі складними умовами освітлення. Додаткові тренування та покращення моделі можуть допомогти усунути ці проблеми і підвищити точність сегментації.

UNet більш чутлива до деталей, тому наявність шуму може сильніше вплинути на її результати, знижуючи точність сегментації. DeepLabV3+ завдяки своїй узагальненій сегментації може краще справлятися з наявністю

шуму, зберігаючи розділення великих об'єктів. Отже, в умовах шуму DeepLabV3+ може показати кращу стійкість і стабільність у сегментації, тоді як UNet може втрачати точність у деталях.

Наступним протестуємо метод суперпіксельної сегментації. На рисунку 3.14, 3.15 відображено результати роботи алгоритму.



Рисунок 3.14 – Результат сегментації зображення вулиці методом суперпіксельної сегментації

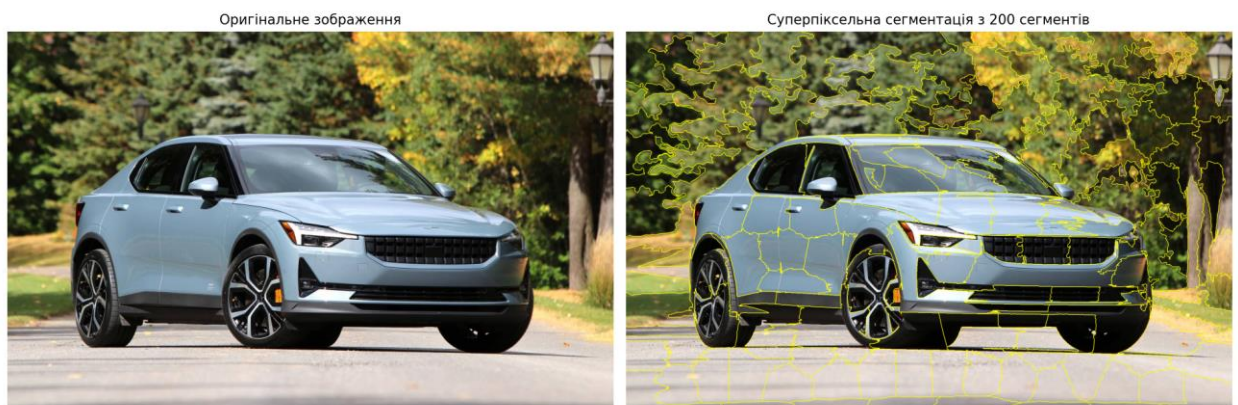


Рисунок 3.15 – Результат сегментації зображення авто методом суперпіксельної сегментації

Ліва частина показує оригінальне зображення автомобіля на фоні природи. Зображення чітке і кольорове, що дозволяє легко розрізнити різні об'єкти та деталі. Права частина показує результат суперпіксельної

сегментації з використанням 200 сегментів. Жовті лінії позначають межі між різними суперпікселями. Суперпікселі – це групи пікселів, які мають схожі характеристики, такі як колір або текстура. Видно, що алгоритм успішно поділив зображення на кілька чітких сегментів. Наприклад, автомобіль чітко виділений з фону, що включає дороги та рослинність. Алгоритм правильно визначив межі між різними об'єктами на зображенні. Це можна побачити по чітким жовтим лініям, які окреслюють різні частини автомобіля та фону чи людини. Навіть при сегментації на 200 суперпікселів, деталі залишаються чіткими, і можна легко розпізнати різні частини зображення.

Наступним протестуємо модель Deeplabv3+ та Unet на знімках з супутника та порівняємо результати. На рисунку 3.16, 3.17 відображено результати роботи алгоритму.



Рисунок 3.16 – Результат сегментації знімка з супутника моделлю Unet

Застосування моделі Unet дає більше виділених областей, включаючи великі регіони. Краще розділення великих об'єктів і територій. Сегментація охоплює більшу частину зображення, що може бути корисно для задач загального огляду територій.



Рисунок 3.17 – Результат сегментації знімка з супутника моделлю DeepLabv3+

Сегментація добре відображає основні компоненти сцени, чітко розмежовуючи різні об'єкти. Модель DeepLabv3+ ефективно виконує сегментацію зображень, чітко відокремлюючи різні об'єкти на сцені.

Проста та симетрична архітектура Unet, що складається з контракуючого (contracting path) та розширюючого (expanding path) шляхів. Ця структура робить модель легшою для розуміння та налаштування. Складніша архітектура DeepLabv3+, включає глибокі нейронні мережі та модулі ASPP (Atrous Spatial Pyramid Pooling), що може ускладнити процес налаштування та навчання. Суперпіксельна сегментація не вимагає глибокого навчання, але зазвичай потребує ретельного налаштування параметрів для досягнення високої точності. Завдяки своєму

підходу «контексту та деталей», U-net забезпечує високу точність сегментації навіть на межах об'єктів. Це особливо корисно для медичних зображень та задач, де важливі дрібні деталі. DeepLabv3+ забезпечує хорошу точність, але може мати труднощі з деталями на межах об'єктів. Суперпіксельна сегментація добре визначає грубі межі об'єктів, але може пропустити дрібні деталі та не завжди точно сегментувати складні сцени.

U-net добре працює з малими та великими об'єктами завдяки своїй здатності захоплювати контекст на різних масштабах через шари контракуючого та розширюючого шляху. DeepLabv3+ ефективний для великих об'єктів, але може мати труднощі з дрібними об'єктами через агрегацію контексту на великому масштабі. Суперпіксельна сегментація добре підходить для великих об'єктів, але може неефективно працювати з дрібними або тонкими структурами. U-net відносно легка реалізація та швидке навчання на середніх ресурсах, що робить його доступним для використання в різних умовах, включаючи медичні дослідження та малі проекти. DeepLabv3+ вимагає більше обчислювальних ресурсів для навчання та реалізації через складнішу архітектуру. Суперпіксельна сегментація не вимагає значних обчислювальних ресурсів, але може бути менш точним для складних задач сегментації. U-net легко налаштовується для різних типів задач сегментації завдяки простій архітектурі та можливості легко змінювати кількість шарів та фільтрів. DeepLabv3+ може бути важко налаштовувати через складні модулі та архітектурні особливості. Суперпіксельна сегментація потребує точного налаштування параметрів для кожної конкретної задачі, що може бути часозатратним.

Отже, U-net є більш простим, гнучким та ефективним рішенням для семантичної сегментації зображень, особливо коли важлива точність на межах об'єктів та можливість працювати з різними масштабами об'єктів. DeepLabv3+ та суперпіксельна сегментація мають свої переваги, але U-net забезпечує кращий баланс між точністю, простотою реалізації та ефективністю.

У цьому розділі було детально розглянуто вибір та реалізацію методів семантичної сегментації зображень для задачі сегментації міських сцен. На основі аналізу було обрано архітектуру нейронної мережі U-Net та набір даних Cityscapes, що обґрунтовано їх високою ефективністю та точністю. Архітектура U-Net була вибрана завдяки її здатності захоплювати як глобальні, так і локальні ознаки зображення, що дозволяє точно відновлювати межі об'єктів. Набір даних Cityscapes забезпечує високу роздільну здатність зображень та детальну піксельну розмітку, що сприяє якісному навчанню моделей семантичної сегментації. Було детально описано компоненти архітектури U-Net, включаючи енкодер, декодер та вихідний шар, а також їх функції у процесі сегментації. Розглянуто кілька метрик для оцінки ефективності моделей сегментації, таких як коефіцієнт Дайса, середнє перетинання на об'єднання (IoU), індекс Жаккара, точність (precision), повнота (recall) та F1-оцінка.

Було проведено тестування моделей на основі архітектури U-Net, алгоритму суперпіксельної сегментації та моделі DeepLabv3+. Представлено результати сегментації різних об'єктів, таких як люди, велосипеди, транспортні засоби та будівлі. Проаналізовано переваги та недоліки кожного з методів, а також їх здатність до узагальнення на реальних сценах.

U-Net показала високу точність та ефективність у задачах сегментації зображень, особливо на межах об'єктів та при різних масштабах об'єктів. DeepLabv3+ та суперпіксельна сегментація також мають свої переваги, але U-Net забезпечує кращий баланс між точністю, простотою реалізації та ефективністю. Вибір архітектури U-Net та набору даних Cityscapes для задачі сегментації міських сцен є обґрунтованим та ефективним рішенням. U-Net демонструє високу точність сегментації завдяки своїй симетричній структурі та можливості поєднувати локальні та глобальні ознаки.

ВИСНОВКИ

У процесі виконання кваліфікаційної роботи було досягнуто значних результатів у дослідженні методів семантичної сегментації зображень. Проведено детальний аналіз предметної області, визначено основні поняття, актуальність та проблематику сегментації зображень, що є важливою складовою комп'ютерного зору та обробки зображень. Робота охопила різноманітні методи сегментації, включаючи класичні підходи, методи на основі машинного навчання та методи глибинного навчання. Набув подальшого вдосконалення метод семантичної сегментації що підтверджено дослідженнями

Аналіз предметної області та визначення основних понять сегментації зображень. Сегментація зображень визначена як ключовий процес у цифровій обробці та аналізі зображень, який полягає у розбитті зображення на декілька частин або сегментів для виділення важливих областей. Визначено, що сегментація зображень базується на властивостях пікселів і може включати відокремлення переднього плану від фону або групування пікселів на основі схожості кольору, текстури чи форми.

Розгляд різних методів сегментації зображень. Класичні методи сегментації, такі як порогова сегментація, сегментація на основі контурів, регіональна сегментація та методи кластеризації, були детально проаналізовані. Методи на основі машинного навчання, включаючи Random Forest, Support Vector Machines (SVM) та Conditional Random Fields (CRF), були досліджені та порівняні. Глибокі нейронні мережі, такі як Fully Convolutional Networks (FCN), U-Net, SegNet та DeepLab, були вивчені як ефективні інструменти для семантичної сегментації зображень.

Вибір та аргументація ефективних методів сегментації. Вибір архітектури U-Net та набору даних Cityscapes для задачі сегментації міських сцен був обґрунтований їх високою точністю та ефективністю. U-Net

продемонстрував високу ефективність навчання навіть на невеликих наборах даних завдяки своїй здатності до узагальнення та збереження детальної інформації про просторові ознаки.

Реалізація та тестування програмного забезпечення для семантичної сегментації зображень. Реалізовано та протестовано програмне забезпечення для сегментації зображень на основі обраної архітектури нейронної мережі. Проведено оцінку точності та ефективності розробленої системи на різних наборах даних. Було проведено тестування моделей на основі архітектури U-Net, алгоритму суперпіксельної сегментації та моделі DeepLabv3+. Представлено результати сегментації різних об'єктів, таких як люди, велосипеди, транспортні засоби та будівлі. Проаналізовано переваги та недоліки кожного з методів, а також їх здатність до узагальнення на реальних сценах. U-Net показала високу точність та ефективність у задачах сегментації зображень, особливо на межах об'єктів та при різних масштабах об'єктів. DeepLabv3+ та суперпіксельна сегментація також мають свої переваги, але U-Net забезпечує кращий баланс між точністю, простотою реалізації та ефективністю.

Визначення актуальності та практичних застосувань сегментації зображень. Сегментація зображень виявилась надзвичайно важливою для медичної візуалізації, автономних транспортних засобів, систем відеоспостереження та інших сфер. Сучасні методи сегментації значно підвищують точність та ефективність роботи систем, забезпечуючи високу якість результатів та спрощуючи подальшу обробку даних.

Проблематика та виклики у семантичній сегментації. Ідентифіковано основні проблеми, такі як велика варіативність об'єктів, умови освітлення, складність фонів, нестача навчальних даних та потреба в обчислювальних ресурсах. Результати дослідження апробовано у вигляді тез доповідей під час конференції "Комп'ютерний зір, системний аналіз та математичне моделювання" [41].

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. ADE20K. MIT Computer Vision. URL: <https://groups.csail.mit.edu/vision/datasets/ADE20K/> (дата звернення: 24.10.2024).
2. What Is Image Segmentation? *MathWorks*. URL: <https://www.mathworks.com/discovery/image-segmentation.html> (дата звернення: 05.10.2024).
3. Carrington D. How Many Photos Will Be Taken in 2021? *Tech today*. URL: <https://focus.mylio.com/tech-today/how-many-photos-will-be-taken-in-2021> (дата звернення: 05.11.2024).
4. Mechea D. What is Panoptic Segmentation and why you should care. *Medium*. URL: <https://medium.com/@danielmechea/what-is-panoptic-segmentation-and-why-you-should-care-7f6c953d2a6a> (дата звернення: 05.10.2024).
5. Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, M. K., Balaskas, K., Topol, E. J., Bachmann, L. M., Keane, P. A., & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*, 1(6).
6. Штучні нейронні мережі. ОГ Руденко, ЄВ Бодянський - Харків: Компанія СМІТ, 2006
7. Artificial neural networks: architectures, learning, applications. YV Bodyanskiy, OG Rudenko - Kharkov: Teletekh, 2004
8. Wang, J., & Liu, X. (2021). Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3+ neural network. *Computer Methods and Programs in Biomedicine*, 207, 106210.

9. Huang, Y., He, X., & Wan, Y. (2024). Hyperspectral-image classification method combining superpixel dimension reduction with post-processing optimization. *National Remote Sensing Bulletin*, 28(2), 494–510.
10. Кобилін, О. А., & Путятіна, О. Є. (2024). Знешумлення зображень, зіпсованих дробовим шумом, у реальному часі. *Системи обробки інформації*, (1 (176), 46-51.
11. Tara S., Reddy B., Ramesh G., Sandeep K. (2019). Various Image Segmentation Methods Based On Partial Differential Equation-A Survey. *International Conference on Computer & Communication Technologies*. Vol. 3, №. 1. P. 183–186.
12. An intuitive guide to Convolutional Neural Networks. FreeCodeCamp. URL: <https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050/> (дата звернення: 13.11.2024).
13. Bernsen J. (1986). Dynamic thresholding of grey-level images. *Proc. Eighth Int'l Conf Pattern Recognition*. P. 1251-1255.
14. Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. *Pattern Recognition*, 33(2), 225-236.
15. GeeksforGeeks. K Means Clustering - Introduction - GeeksforGeeks. GeeksforGeeks. URL :<https://www.geeksforgeeks.org/k-means-clustering-introduction/> (дата звернення: 08.10.2024).
16. Fuzzy c-means clustering algorithm. *Data Clustering Algorithms*. URL: <https://sites.google.com/site/dataclusteringalgorithms/fuzzy-c-means-clustering-algorithm> (дата звернення: 08.11.2024).
17. Milleana Shaharudin, S., Ismail, S., Nor, S. M. C. M., & Ahmad, N. (2019). An Efficient Method to Improve the Clustering Performance using Hybrid Robust Principal Component Analysis-Spectral biclustering in Rainfall Patterns Identification. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 8(3), 237.

18. Ohlander, R., Price, K., & Reddy, D. R. (1978). Picture segmentation using a recursive region splitting method. *Computer Graphics and Image Processing*, 8(3), 313–333.
19. Штучні нейронні мережі: Навчальний посібник. ОГ Руденко, ЄВ Бодянський - 2006
20. Штучні нейронні мережі: архітектури, навчання, застосування. ЄВ Бодянський, ОГ Руденко - Харків: Телетех, 2004.
21. Безсонов, О. О., Руденко, О. Г., Олійник, К., & Романюк, О. (2020). Оптимізація процесу навчання штучних нейронних мереж.
22. The U-Net : A Complete Guide. URL: <https://medium.com/@alejandro.itoaramendia/decoding-the-u-net-a-complete-guide-810b1c6d56d8> (дата звернення 14.11.2024).
23. SegNET. URL: <https://medium.com/@saba99/segnet-a139ce77b570> (дата звернення 01.10.2024).
24. DeepLab: A Deep Dive into Advanced Visual Processing. URL: <https://viso.ai/deep-learning/deeplab/> (дата звернення 27.10.2024).
25. A Deep Dive Into Semantic Segmentation Evaluation Metrics. URL: <https://hackernoon.com/a-deep-dive-into-semantic-segmentation-evaluation-metrics> (дата звернення 23.11.2024).
26. CityScapes - Depth and segmentation. Kaggle. URL: <https://www.kaggle.com/datasets/sakshaymahna/cityscapes-depth-and-segmentation> (дата звернення 13.11.2024).
27. Методи цифрової обробки зображень. О.А. Кобилін, І.С. Творошенко - 2021
28. Developing a Multi-Step Recurrent Algorithm to Maximize the Criteria of Correntropy. О. Rudenko, О. Bezsonov, V. Borysenko, Т. Borysenko... - *Eastern-European Journal of Enterprise Technologies*, 2021.
29. Штучні нейронні мережі. О.Г. Руденко, Є.В. Бодянський – О.Г. Руденко, Є.В. Бодянський, Харків, 2005

30. Artificial neural networks. O.G. Rudenko, E.V. Bodiansky - SMIT Company LLC, Kharkiv, 2006
31. Штучні нейронні мережі: Навчальний посібник. О.Г. Руденко, Є.В. Бодянський – Харків: ТОВ «Компанія Сміт», 2005.
32. Основи теорії штучних нейронних мереж. О.Г. Руденко, Є.В. Бодянський – Харків: Телетех, 2002.
33. Rudenko, O. G., & Bessonov, A. A. (2005). CMAC neural network and its use in problems of identification and control of nonlinear dynamic objects. *Cybernetics and Systems Analysis*, 41, 647-658.
34. Oleksandr Bezsonov, Oleh Lebediev, Valentyn Lebediev, Megel, Y., Dmytro Prochukhan, & Rudenko, O. (2021). Breed recognition and estimation of live weight of cattle based on methods of machine learning and computer vision. *Eastern-European Journal of Enterprise Technologies*, 6(9 (114)), 64–74.
35. Kollem, S., Reddy, K. R. L., & Rao, D. S. (2019). A Review of Image Denoising and Segmentation Methods Based on Medical Images. *International Journal of Machine Learning and Computing*, 9(3), 288–295.
36. Khosla, T., & Verma, O. P. (2023). Optimal threshold selection for segmentation of Chest X-Ray images using opposition-based swarm-inspired algorithm for diagnosis of pneumonia. *Multimedia Tools and Applications*, 83(9), 27089–27119.
37. Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P., & Hawkesford, M. J. (2019). DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. *Frontiers in Plant Science*.
38. Ren Xinlei, & Wang Yangping. (2020). Super-Pixel Segmentation of Remote Sensing Image Based on Improved Simple Linear Iterative Clustering Algorithm. *Laser & Optoelectronics Progress*, 57(22), 222801–222801.
39. Usha, & S. Vasuki. (2022). Significance of texture features in the segmentation of remotely sensed images. *Optik*, 249, 168241–168241.

40. S. Osher and N. Paragios. Geometric Level Set Methods in Imaging Vision and Graphics, Springer Verlag, ISBN 0-387-95488-0, 2021.

41. Радіоелектроніка та молодь у ХХІ столітті. Т. 7 : Конференція "Комп'ютерний зір, системний аналіз та математичне моделювання" : матеріали 28-го Міжнар. молодіж. форуму, 16–18 квіт. 2024 р. / М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків : ХНУРЕ, 2024. – 320 с. – DOI: 10.30837/IYF.CVSAMM.2024.