

Міністерство освіти і науки України
Луцький національний технічний університет
Факультет робототехніки та штучного інтелекту
Кафедра штучного інтелекту та математичного моделювання

КВАЛІФІКАЦІЙНА РОБОТА ЗА СТУПЕНЕМ ВИЩОЇ ОСВІТИ
«БАКАЛАВР»

**ДОСЛІДЖЕННЯ АРХІТЕКТУРИ U-NET ДЛЯ СЕМАНТИЧНОЇ
СЕГМЕНТАЦІЇ СУПУТНИКОВИХ ЗНІМКІВ**

**RESEARCH OF THE U-NET ARCHITECTURE FOR SEMANTIC
SEGMENTATION OF SATELLITE IMAGERY**

Спеціальність 113 Прикладна математика
(шифр і назва спеціальності)

освітня програма «Штучний інтелект та аналіз масивів даних»
(назва освітньої програми)

Виконав: здобувач вищої освіти
Групи ПРМ-41
Романік Віталій Миколайович

(підпис)

Керівник:
к.т.н., доцент
Приходько Олексій Сергійович

(підпис)

Кваліфікаційну роботу
допущено до захисту
«__» _____ 20__ р.
к.т.н., доцент
Гарант освітньої програми:
Приходько Олексій Сергійович

(підпис)

Луцьк – 2026 року

ЛУЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

Факультет *архітектури, будівництва та дизайну*

Кафедра *прикладної математики та механіки*

Ступінь вищої освіти: *бакалавр*

Галузь знань: *11 Математика і статистика*

Спеціальність *113 Прикладна математика*

Освітня програма *Штучний інтелект та аналіз масивів даних*

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Мікуліч О.А.

« ____ » _____ 2025 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧУ ВИЩОЇ ОСВІТИ

Романік Віталій Миколайович

(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи

Дослідження архітектури U-Net для семантичної сегментації супутникових знімків _____ *Research of the U-Net architecture for semantic segmentation of satellite imagery*

Керівник роботи: *Приходько Олексій Сергійович*

затверджені наказом закладу вищої освіти від «31» грудня 2025 р. № 557/01-02

2. Строк подання здобувачем вищої освіти кваліфікаційної роботи 04.06.2026 р.

3. Вихідні дані до роботи _____ *профільні публікації з штучного інтелекту та математичного моделювання в межах досліджуваної проблематики; релевантні набори даних; математичні моделі цільових процесів; технічна документація Python-бібліотек та методичні вказівки до виконання кваліфікаційної роботи бакалавра* _____

4. Зміст пояснювальної записки (перелік питань, що потрібно розробити):

Аналіз предметної області

Постановка задачі та вибір методів

Практична реалізація

Отримання та аналіз результатів

5. Перелік графічного (ілюстративного) матеріалу:

Презентація роботи (слайди): об'єкт, предмет, мета та завдання

дослідження; аналіз предметної області; математичні та алгоритмічні

моделі, результати експериментальних досліджень (метрики та візуалізація

роботи алгоритму); загальні висновки

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис	
		завдання видав	завдання прийняв
<i>1 розділ</i>	<i>Приходько О.С., доцент кафедри</i>		
<i>2 розділ</i>	<i>Приходько О.С., доцент кафедри</i>		
<i>3 розділ</i>	<i>Приходько О.С., доцент кафедри</i>		
<i>4 розділ</i>	<i>Приходько О.С., доцент кафедри</i>		
<i>Висновки</i>	<i>Приходько О.С., доцент кафедри</i>		

7. Дата видачі завдання « ___ » _____ 202__ р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи бакалавра	Строк виконання етапів роботи	Примітка
1.	<i>Огляд літератури із досліджуваної проблеми</i>	<i>до 01.03.2026</i>	
2.	<i>Перший розділ</i>	<i>до 5.03.2026</i>	
3.	<i>Другий розділ</i>	<i>до 01.04.2026</i>	
4.	<i>Третій розділ</i>	<i>до 10.04.2026</i>	
5.	<i>Четвертий розділ</i>	<i>до 20.04.2026</i>	
6.	<i>Висновки</i>	<i>до 28.04.2026</i>	
7.	<i>Формування списку використаних джерел</i>	<i>до 05.05.2026</i>	
8.	<i>Оформлення ілюстративного матеріалу</i>	<i>до 09.05.2026</i>	
9.	<i>Нормоконтроль</i>	<i>до 20.05.2026</i>	
10.	<i>Інструментальна перевірка на академічний плагіат</i>	<i>до 02.06.2026</i>	<i>Показник запозичень тексту _____ %</i>
11.	<i>Представлення кваліфікаційної роботи бакалавра до захисту</i>	<i>до 04.06.2026</i>	

Здобувач вищої освіти

_____ (Романік В. М.)
(підпис) (прізвище, ініціали)

Керівник кваліфікаційної роботи

_____ (Приходько. О. С.)
(підпис) (прізвище, ініціали)

АНОТАЦІЯ

Романік В. М. Дослідження архітектури U-Net для семантичної сегментації супутникових знімків. Рукопис.

Кваліфікаційна робота бакалавра ОП «Штучний інтелект та аналіз масивів даних» спеціальності 113 Прикладна математика. Луцький національний технічний університет. Луцьк, 2026.

Кваліфікаційну роботу присвячено дослідженню та програмній імплементації згорткової нейронної мережі архітектури U-Net для автоматизованої семантичної сегментації просторових даних. У процесі дослідження формалізовано математичну постановку задачі піксельної класифікації в умовах екстремального дисбалансу класів. Розроблено конвеєр попередньої обробки даних, що включає розбиття зображень на фрагменти та інтенсивну стохастичну аугментацію. Для ефективної оптимізації ваг моделі імплементовано комбіновану функцію втрат (лінійна комбінація Binary Cross-Entropy та Dice Loss).

Експериментальне дослідження проведено на мультиспектральних зображеннях з набору даних DeepGlobe Land Cover. Отримані кількісні результати оцінено за допомогою метрик просторового перекриття (IoU, Dice Coefficient, AUC-PR). На основі візуального зіставлення прогнозних масок виявлено явище семантичної неоднозначності еталонної розмітки: розроблена модель успішно локалізувала об'єкти рослинності (міські дерева, лісосмуги), які не були розмічені експертами через географічну таксономію датасету. Доведено, що архітектура U-Net володіє високою здатністю до вилучення просторово-спектральних ознак і є дієвим інструментом для автоматизованого аналізу даних дистанційного зондування Землі.

Ключові слова: семантична сегментація, згорткові нейронні мережі, U-Net, дистанційне зондування Землі, супутникові знімки, машинне навчання, комп'ютерний зір, функції втрат.

ABSTRACT

Romanik V. M. Research of the U-Net architecture for semantic segmentation of satellite imagery. Manuscript.

Bachelor's Thesis in the field of "Artificial Intelligence and BigData Analysis" in specialty 113 Applied Mathematics. Lutsk National Technical University. Lutsk, 2026.

The bachelor's thesis is devoted to the research and software implementation of the U-Net convolutional neural network architecture for the automated semantic segmentation of spatial data. During the research, the mathematical formulation of the pixel-wise classification problem under conditions of extreme class imbalance was formalized. A spatial data preprocessing pipeline was developed, which includes image patching and intensive stochastic data augmentation. To effectively optimize the model weights, a combined loss function (a linear combination of Binary Cross-Entropy and Dice Loss) was implemented.

The experimental study was conducted on multispectral images from the DeepGlobe Land Cover dataset. The obtained quantitative results were evaluated using spatial overlap metrics (IoU, Dice Coefficient, AUC-PR). Based on the visual comparison of the predicted masks, the phenomenon of semantic ambiguity of the ground truth annotations was revealed: the developed model successfully localized vegetation objects (urban trees, windbreaks) that were not annotated by experts due to the geographical taxonomy of the dataset. It is proven that the U-Net architecture has a high capacity for extracting spatio-spectral features and is a powerful tool for the automated analysis of Earth remote sensing data.

Keywords: semantic segmentation, convolutional neural networks, U-Net, Earth remote sensing, satellite imagery, machine learning, computer vision, loss functions.

ЗМІСТ

ВСТУП	7
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ДАНИХ	10
1.1. Огляд методів дистанційного зондування Землі та обробки супутникових знімків	10
1.2. Проблема семантичної сегментації в задачах комп'ютерного зору	11
1.3. Аналіз існуючих наборів даних (датасетів) для сегментації супутникових зображень.....	13
1.4. Огляд сучасних архітектур згорткових нейронних мереж для обробки просторових даних.....	14
Висновки до розділу 1	16
РОЗДІЛ 2. ПОСТАНОВКА ЗАДАЧІ ТА ВИБІР МЕТОДІВ РОЗВ'ЯЗАННЯ. 18	
2.1. Математична постановка задачі семантичної сегментації зображень..	18
2.2. Теоретичні основи та математична модель архітектури U-Net.....	19
2.3. Обґрунтування вибору функцій втрат для незбалансованих даних.....	21
2.4. Визначення метрик оцінки якості моделей сегментації (IoU, Dice Coefficient, Pixel Accuracy)	23
Висновки до розділу 2	25
РОЗДІЛ 3. РОЗРОБКА ТА ІМПЛЕМЕНТАЦІЯ РІШЕННЯ.....	26
3.1. Проектування конвеєра попередньої обробки даних	26
3.2. Програмна реалізація архітектури U-Net	28
3.3. Вибір інструментальних засобів та середовища розробки.....	31
РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ.....	33
4.1. Планування експерименту та налаштування гіперпараметрів навчання	33
4.2. Аналіз процесу навчання моделі та динаміки зміни функцій втрат	35
4.3. Оцінка результатів семантичної сегментації на тестовій вибірці	37
4.4. Порівняльний аналіз швидкодії та точності розробленої моделі	42
Висновки до розділу 4.	43
ВИСНОВКИ.....	44
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	46

ВСТУП

Актуальність теми. В останні десятиліття обсяги даних дистанційного зондування Землі (ДЗЗ) зростають експоненціально завдяки розвитку супутникових технологій. Супутникові знімки високої роздільної здатності є критично важливим джерелом інформації для вирішення широкого спектра прикладних задач: моніторингу сільськогосподарських угідь, оцінки наслідків стихійних лих, урбаністичного планування, екологічного контролю та оновлення картографічних баз даних. Традиційні методи ручного дешифрування або класичні алгоритми обробки зображень не здатні забезпечити необхідну швидкість та точність аналізу таких масивів даних. Це зумовлює гостру необхідність автоматизації процесу виділення об'єктів за допомогою методів штучного інтелекту.

Семантична сегментація є однією з ключових задач комп'ютерного зору, яка полягає у присвоєнні кожному пікселю зображення відповідного класу. У контексті супутникових знімків ця задача ускладнюється високою щільністю об'єктів, варіативністю освітлення, наявністю шумів (хмари, тіні) та складним просторовим контекстом. Серед сучасних підходів до вирішення цієї проблеми найвищу ефективність демонструють повністю згорткові нейронні мережі (FCN). Архітектура U-Net, яка спочатку була розроблена для біомедичної сегментації, завдяки своїй симетричній структурі кодувальника-декодувальника та наявності пропускних з'єднань (skip connections), дозволяє ефективно об'єднувати низькорівневу просторову інформацію з високорівневими семантичними ознаками. Це робить її особливо перспективною для обробки супутникових зображень, де точність локалізації меж об'єктів має критичне значення. З огляду на це, дослідження та адаптація архітектури U-Net для семантичної сегментації супутникових знімків є актуальною науково-практичною задачею у сфері штучного інтелекту та аналізу масивів даних.

Мета роботи полягає у дослідженні, програмній реалізації та оцінці ефективності нейромережевої архітектури U-Net для вирішення задачі автоматизованої семантичної сегментації об'єктів на супутникових знімках.

Для досягнення поставленої мети було сформульовано та вирішено такі **завдання**:

1. Провести огляд предметної області, існуючих методів дистанційного зондування та наборів даних просторових зображень.
2. Сформулювати математичну постановку задачі семантичної сегментації та дослідити теоретичні засади архітектури U-Net.
3. Визначити оптимальні метрики оцінки якості та функції втрат для роботи з незбалансованими класами на супутникових знімках.
4. Розробити програмний конвеєр попередньої обробки даних (аугментація, нормалізація, розбиття на фрагменти).
5. Здійснити імплементацію архітектури U-Net з використанням сучасних фреймворків глибокого навчання.
6. Провести серію експериментів з навчання моделі, проаналізувати отримані результати та оцінити якість сегментації за обраними метриками.

Об'єкт дослідження – процеси автоматизованого аналізу та обробки масивів просторових даних дистанційного зондування Землі.

Предмет дослідження – математичні моделі, алгоритми машинного навчання та нейромережева архітектура U-Net для семантичної сегментації супутникових знімків.

Методи дослідження. Для вирішення поставлених завдань у роботі використано комплекс методів: системний аналіз – для дослідження предметної області; методи лінійної алгебри, теорії ймовірностей та математичної статистики – для формалізації задачі; методи комп'ютерного зору та глибокого навчання (зокрема, згорткові нейронні мережі) – для побудови моделі вилучення ознак; методи оптимізації (алгоритми градієнтного спуску) – для мінімізації функції втрат під час навчання мережі;

емпіричні методи – для проведення експериментів та верифікації отриманих результатів.

Інформаційна база дослідження включає наукові праці вітчизняних та зарубіжних вчених у галузі машинного навчання і комп'ютерного зору, матеріали міжнародних конференцій (CVPR, ICCV, NeurIPS), офіційну документацію до програмних бібліотек глибокого навчання (PyTorch/TensorFlow), а також відкриті набори супутникових даних (наприклад, DeepGlobe Land Cover Classification Dataset або SpaceNet), що використовуються для навчання та тестування моделей.

У процесі підготовки бакалаврської кваліфікаційної роботи застосовувалися технології штучного інтелекту як допоміжний інструментарій. Зокрема, для стилістичної правки та структурування тексту використано ChatGPT-5 та Gemini 3.3. Автор несе повну відповідальність за зміст роботи

РОЗДІЛ 1

АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ДАНИХ

1.1. Огляд методів дистанційного зондування Землі та обробки супутникових знімків

Дистанційне зондування Землі (ДЗЗ) є невід'ємним інструментом отримання просторової інформації про поверхню планети без безпосереднього фізичного контакту з об'єктами дослідження. Упродовж останніх десятиліть технології ДЗЗ зазнали суттєвої еволюції, що призвело до стрімкого зростання обсягів доступних даних. Сучасні супутникові системи, такі як Sentinel, Landsat, WorldView та PlanetScope, здатні генерувати терабайти мультиспектральних зображень щодня, забезпечуючи високу просторову, спектральну та часову роздільну здатність [1]. Це відкриває широкі можливості для моніторингу екологічних змін, урбанізації, стану сільськогосподарських угідь та ліквідації наслідків стихійних лих.

Водночас, збільшення обсягів і деталізації супутникових знімків створило нові виклики для методів їх обробки. Традиційні підходи до аналізу таких даних спиралися переважно на ручне дешифрування експертами. Цей процес характеризується високою точністю, проте є надзвичайно ресурсомістким, повільним і суб'єктивним, що робить його непридатним для обробки великих масивів даних у режимі реального часу. Для автоматизації цього процесу тривалий час використовувалися класичні алгоритми обробки зображень та базові методи машинного навчання.

Класичні методи можна умовно поділити на піксель-орієнтовані та об'єктно-орієнтовані (Object-Based Image Analysis, OBIA). Піксельні методи класифікують кожен піксель ізольовано, базуючись виключно на його спектральних характеристиках, що часто призводить до ефекту високочастотного шуму на результуючих картах. Об'єктно-орієнтовані підходи частково вирішують цю проблему шляхом попередньої сегментації

зображення на однорідні регіони (суперпікселі) з подальшою їх класифікацією за допомогою методів опорних векторів (SVM), випадкових лісів (Random Forest) або алгоритму k-найближчих сусідів [3]. Хоча методи ОБІА враховують форму та текстуру об'єктів, вони критично залежать від ручного конструювання ознак та параметрів початкової сегментації, що знижує їхню стійкість до зміни умов освітлення, сезону або географічного регіону.

З появою концепції глибокого навчання парадигма обробки даних ДЗЗ докорінно змінилася. Замість ручного конструювання ознак, згорткові нейронні мережі (CNN) здатні самостійно вилучати ієрархічні просторово-спектральні ознаки безпосередньо з сирих даних під час процесу навчання [1]. Глибоке навчання забезпечує значний приріст точності класифікації та сегментації, особливо у випадках аналізу знімків надвисокої роздільної здатності (Very High Resolution, VHR), де контекстуальна інформація відіграє вирішальну роль у розрізненні об'єктів зі схожими спектральними профілями, наприклад, дахів будівель та асфальтованих доріг.

1.2. Проблема семантичної сегментації в задачах комп'ютерного зору

Семантична сегментація є однією з фундаментальних задач комп'ютерного зору, що полягає у присвоєнні кожному пікселю зображення мітки певного семантичного класу. На відміну від задачі класифікації зображень, яка визначає наявність об'єкта на всьому знімку, та задачі детекції об'єктів, яка локалізує їх за допомогою обмежувальних рамок, семантична сегментація вимагає точного розуміння меж об'єктів на піксельному рівні [2]. Відповідно, вона не розрізняє окремі екземпляри одного класу (наприклад, два автомобілі, що стоять поруч, будуть виділені як єдина область класу «автомобіль»), що відрізняє її від задачі екземплярної сегментації.

До стрімкого розвитку нейромережових підходів семантична сегментація базувалася на використанні марковських та умовних випадкових

полів (CRF), які моделювали локальні взаємозв'язки між пікселями для забезпечення просторової узгодженості міток. Проте справжній прорив у цій галузі відбувся у 2015 році з появою архітектури повністю згорткових мереж (Fully Convolutional Networks, FCN), запропонованої у роботі J. Long та співавторів [4]. FCN замінила повнозв'язні шари традиційних класифікаційних мереж на згорткові шари, що дозволило обробляти зображення довільного розміру та генерувати просторові карти класифікації. Для відновлення просторової роздільної здатності після операцій субдискретизації у FCN були застосовані операції транспонованої згортки.

Хоча мережі сімейства FCN продемонстрували високі результати, вони мали суттєвий недолік: втрату детальної просторової інформації (зокрема, меж об'єктів) через багаторазове зменшення розмірності карт ознак у процесі кодування. Для подолання цього обмеження було розроблено архітектури типу «кодувальник-декодувальник». Класичними представниками цього підходу є мережі SegNet [8] та U-Net [5]. Архітектура U-Net, яка спочатку була розроблена для аналізу біомедичних зображень, запровадила механізм пропускних з'єднань (skip connections), що безпосередньо передають карти ознак з високою просторовою роздільною здатністю від блоків кодувальника до відповідних блоків декодувальника [5]. Це рішення дозволило об'єднати глибокий семантичний контекст із точною просторовою локалізацією.

Застосування методів семантичної сегментації до супутникових знімків супроводжується низкою специфічних проблем, які не притаманні звичайним оптичним зображенням. По-перше, супутникові дані характеризуються значною варіативністю масштабів: на одному знімку можуть бути присутні як великі об'єкти (лісові масиви, водойми), так і дрібні (автомобілі, окремі будівлі). По-друге, існує проблема жорсткого дисбалансу класів. Наприклад, клас «фон» (земля, пустирі) часто займає понад 90% площі зображення, тоді як цільові класи (дороги, будівлі) представлені лише незначною кількістю пікселів [7]. По-третє, складні умови зйомки (різні кути нахилу сенсора, вплив

атмосфери, наявність хмар та тіней від об'єктів) призводять до значних перекриттів об'єктів та спотворення їхніх спектральних ознак.

Для вирішення цих проблем сучасні дослідження фокусуються на вдосконаленні базових архітектур, інтеграції механізмів просторової уваги, використанні розширених згорток, як у мережах лінійки DeepLab [9], та проектуванні специфічних функцій втрат, стійких до незбалансованих вибірок. Отже, вибір та адаптація відповідної архітектури є критично важливим етапом у побудові надійної системи автоматизованого аналізу просторових даних.

1.3. Аналіз існуючих наборів даних (датасетів) для сегментації супутникових зображень

Ефективність застосування моделей глибокого навчання, зокрема згорткових нейронних мереж, критично залежить від обсягу та якості навчальних вибірок. У контексті задач дистанційного зондування розробка якісних наборів даних (датасетів) є складною задачею через необхідність залучення експертів для попиксельної розмітки зображень високої роздільної здатності. Протягом останніх років науковою спільнотою було створено кілька еталонних баз даних, які дозволяють об'єктивно оцінювати та порівнювати різні архітектури.

Однією з найвідоміших ініціатив є серія змагань SpaceNet, у рамках якої було відкрито доступ до великих масивів мультиспектральних супутникових знімків із просторовою роздільною здатністю від 0,3 до 0,5 м [11]. Датасети SpaceNet орієнтовані переважно на вирішення двох задач: сегментації контурів будівель та виділення дорожньої мережі. Особливістю цих даних є висока деталізація та географічне різноманіття (знімки з різних континентів), що дозволяє оцінити здатність моделей до генералізації (узагальнення) на нових територіях. Проте ці набори характеризуються значним дисбалансом

класів, оскільки цільові об'єкти (будівлі, дороги) займають мізерну частку площі знімка порівняно з фоном.

Іншим важливим джерелом даних є набір DeepGlobe 2018, який містить три незалежні вибірки для різних задач: класифікації покриття землі (Land Cover), виділення доріг та сегментації будівель [10]. Датасет для класифікації земної поверхні (DeepGlobe Land Cover) є особливо цінним для досліджень семантичної сегментації, оскільки він містить знімки з роздільною здатністю 0,5 м, кожен піксель яких віднесено до одного з семи класів (міська забудова, сільське господарство, пасовища, ліси, водойми, пустирі та невідомий клас). Така багатокласова розмітка дозволяє тестувати архітектури на здатність розрізняти складні природні та антропогенні об'єкти зі схожими спектральними характеристиками.

Вибір датасету для навчання моделі визначає стратегію попередньої обробки даних (аугментацію, нормалізацію, розбиття на фрагменти або патчі). Оскільки вхідні супутникові знімки часто мають розміри понад 10000x10000 пікселів, їх неможливо безпосередньо подати на вхід згортковій нейронній мережі через обмеження обсягу відеопам'яті обчислювальних пристроїв. Тому стандартною практикою є розбиття зображень на менші перекривні фрагменти розміром 256x256 або 512x512 пікселів. Наявність таких еталонних баз даних, як DeepGlobe та SpaceNet, забезпечує надійне інформаційне підґрунтя для навчання та валідації модифікацій архітектури U-Net, гарантуючи об'єктивність експериментальних результатів.

1.4. Огляд сучасних архітектур згорткових нейронних мереж для обробки просторових даних

Дослідження в галузі комп'ютерного зору стимулювали появу значної кількості архітектур згорткових нейронних мереж, багато з яких були успішно адаптовані для аналізу просторових даних дистанційного зондування. Фундаментальним етапом розвитку стало створення глибоких мереж

класифікації зображень, навчених на базі даних ImageNet [14]. Ці моделі (VGG, Inception, ResNet) заклали основу для використання методу трансферного навчання (transfer learning), коли попередньо навчені ваги мережі-класифікатора використовуються як екстрактор ознак у більш складних моделях сегментації.

Значним проривом у вирішенні проблеми затухання градієнта під час навчання глибоких мереж стала архітектура ResNet (Residual Network). Запропоновані у ній залишкові блоки (residual blocks) з пропусковими з'єднаннями (shortcut connections) дозволили ефективно навчати мережі глибиною понад сто шарів [12]. Інтеграція залишкових блоків у структури кодувальників для задач семантичної сегментації дозволила суттєво покращити здатність моделей вилучати складні абстрактні ознаки.

Класична архітектура U-Net, попри свою ефективність у збереженні просторової інформації, має обмеження щодо глибини кодувальника та часто страждає від надмірної кількості хибнопозитивних спрацьовувань на фонових пікселях. Для подолання цих недоліків було запропоновано низку модифікацій. Однією з перших успішних адаптацій для просторових даних стала архітектура TernausNet. Вона використовує класичну структуру U-Net, проте її кодувальник замінено на попередньо навчену на наборі ImageNet мережу VGG11 [16]. Такий підхід різко прискорює збіжність моделі (convergence) та підвищує загальну точність розпізнавання за рахунок використання багатих візуальних ознак, отриманих із мільйонів природних зображень.

Подальший розвиток ідей об'єднання залишкових мереж та архітектури типу "кодувальник-декодувальник" призвів до створення ResUNet-a [6]. Ця архітектура була розроблена спеціально для сегментації даних дистанційного зондування високої роздільної здатності. Вона поєднує в собі структуру U-Net, залишкові блоки та просторові пірамідальні пулінги, що дозволяє аналізувати контекст об'єктів на різних масштабах без втрати роздільної здатності карт ознак.

Окремим перспективним напрямком є впровадження механізмів уваги в архітектури сегментації. Модель Attention U-Net інтегрує так звані вентиля уваги безпосередньо у пропускні з'єднання [13]. Ці механізми дозволяють мережі динамічно придушувати нерелевантні фонові ознаки та фокусуватися виключно на цільових структурах. У контексті супутникових знімків, де фон часто є гетерогенним і містить багато візуального шуму, використання механізмів уваги демонструє суттєве підвищення точності локалізації меж складних географічних об'єктів.

Підсумовуючи результати аналізу предметної області, можна зробити висновок, що задача автоматизованої семантичної сегментації супутникових знімків є надзвичайно актуальною, проте ускладненою специфікою просторових даних. Серед різноманіття сучасних підходів архітектура U-Net та її модифікації, що використовують залишкові зв'язки (ResNet-backbone) або механізми уваги, забезпечують найкращий баланс між здатністю виділяти високочастотні деталі меж об'єктів та глибоким семантичним розумінням контексту. Це обґрунтовує доцільність вибору U-Net як базової математичної моделі для розробки та імплементації системи сегментації у межах даної кваліфікаційної роботи.

Висновки до розділу 1

У першому розділі проведено комплексний аналіз предметної області та методів обробки даних дистанційного зондування Землі. Встановлено, що класичні алгоритми обробки зображень не здатні забезпечити необхідну точність та автоматизацію для сучасних обсягів супутникових даних надвисокої роздільної здатності. Доведено перевагу методів глибокого навчання, зокрема згорткових нейронних мереж, у вирішенні задач семантичної сегментації. Огляд сучасних архітектур показав, що топологія U-Net з її симетричною структурою та механізмом пропускних з'єднань є оптимальним вибором для точної піксельної локалізації об'єктів.

Проаналізовано ключові набори даних, серед яких визначено релевантні датасети (такі як DeepGlobe та SpaceNet) для проведення експериментальних досліджень.

РОЗДІЛ 2

ПОСТАНОВКА ЗАДАЧІ ТА ВИБІР МЕТОДІВ РОЗВ'ЯЗАННЯ

2.1. Математична постановка задачі семантичної сегментації зображень

З математичної точки зору, цифрове зображення (зокрема, багатоспектральний супутниковий знімок) можна представити як тривимірний тензор $X \in \mathbb{R}^{H \times W \times C}$, де H та W позначають просторові розміри зображення (висоту та ширину в пікселях відповідно), а C визначає кількість спектральних каналів. Для стандартного оптичного знімка $C = 3$ (червоний, зелений, синій канали), однак для супутникових даних ця величина може бути більшою за рахунок наявності інфрачервоного або інших специфічних діапазонів.

Задача семантичної сегментації полягає у побудові функції відображення $f(X; \theta)$, параметризованої набором ваг θ , яка кожному пікселю вхідного зображення ставить у відповідність мітку класу $k \in \{1, 2, \dots, K\}$, де K позначає загальну кількість попередньо визначених семантичних класів. Результатом роботи такої функції є двовимірна матриця передбачень (карта сегментації) $Y \in \mathbb{Z}^{H \times W}$, розмірність якої збігається з просторовою розмірністю вхідного зображення.

У контексті нейромережевого підходу функція f апроксимується глибокою згортковою нейронною мережею. На виході мережа генерує тензор логітів $Z \in \mathbb{R}^{H \times W \times K}$, який містить ненормалізовані оцінки належності кожного пікселя до кожного з класів. Для перетворення цих оцінок у дискретний розподіл ймовірностей застосовується функція Softmax, яка обчислюється незалежно для кожного пікселя просторової сітки:

$$\hat{p}_{i,j,k} = \frac{\exp(z_{i,j,k})}{\sum_{c=1}^K \exp(z_{i,j,c})} \quad (3.1)$$

де $\hat{p}_{i,j,k}$ – нормалізована ймовірність належності пікселя з координатами (i, j) до класу k , $z_{i,j,k}$ – вихідний сигнал (логіт) нейронної мережі для класу k у позиції (i, j) , K – загальна кількість класів у задачі, c – індекс підсумовування за всіма можливими класами.

Отримавши тензор ймовірностей $\hat{P} \in [0,1]^{H \times W \times K}$, фінальне рішення щодо класу кожного пікселя ухвалюється за допомогою операції вибору індексу з максимальним значенням ймовірності (argmax):

$$\hat{y}_{i,j} = \operatorname{arg max}_{k \in \{1, \dots, K\}} \hat{p}_{i,j,k} \quad (3.2)$$

де $\hat{y}_{i,j}$ – прогнозована дискретна мітка класу для пікселя (i,j) , $\hat{p}_{i,j,k}$ – ймовірність того, що піксель належить до класу k .

Таким чином, процес навчання моделі семантичної сегментації зводиться до задачі багатокласової класифікації на рівні окремих пікселів. Оптимізація параметрів θ здійснюється шляхом мінімізації розбіжності між прогнозованим розподілом ймовірностей \hat{P} та істинним розподілом (ground truth маскою) за допомогою алгоритмів градієнтного спуску та методу зворотного поширення помилки [17].

2.2. Теоретичні основи та математична модель архітектури U-Net

Серед розмаїття сучасних архітектур глибокого навчання для задач сегментації зображень, мережа U-Net довела свою виняткову ефективність, особливо в умовах роботи зі складними просторовими структурами. Запропонована у роботі О. Роннебергера [5], ця топологія належить до класу повністю згорткових мереж (FCN) і концептуально складається з двох симетричних гілок: гілки стиснення (кодувальника) та гілки розширення (декодувальника).

Кодувальник виконує роль екстрактора ознак (feature extractor). Його завдання – поступово зменшувати просторову розмірність карт ознак, одночасно збільшуючи їхню глибину (кількість каналів) для фіксації

абстрактного семантичного контексту. Кожен блок кодувальника традиційно містить дві послідовні операції згортки з ядром 3×3 . Математично дискретна двовимірна згортка для одного каналу описується виразом:

$$S(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (3.3)$$

де $S(i, j)$ – значення пікселя вихідної карти ознак у позиції (i, j) , I – матриця вхідних даних (або карта ознак попереднього шару), K – матриця ваг ядра згортки, m, n – просторові розміри ядра згортки.

Після кожної операції згортки застосовується нелінійна функція активації. У класичній архітектурі U-Net використовується випрямлений лінійний пристрій (ReLU), який забезпечує швидку збіжність мережі та запобігає проблемі затухання градієнтів на глибоких шарах [18]. Функція ReLU визначається як:

$$f(x) = \max(0, x) \quad (3.4)$$

де $f(x)$ – вихідне значення після нелінійного перетворення, x – вхідний сигнал нейрона (результат операції згортки).

Для зменшення просторової розмірності після застосування згорток використовується операція об'єднання за максимумом (Max Pooling) з вікном 2×2 та кроком два. Це призводить до зменшення ширини та висоти карт ознак удвічі, що забезпечує інваріантність ознак до незначних просторових зсувів.

Гілка розширення (декодувальник) призначена для відновлення просторової роздільної здатності зображення та точної локалізації виділених об'єктів. Процес відновлення розмірності реалізується за допомогою операції транспонованої згортки (іноді помилково званої деконволюцією) з ядром 2×2 , яка збільшує просторові розміри карти ознак удвічі та зменшує кількість каналів.

Ключовою інновацією U-Net є наявність пропускних з'єднань між гілками стиснення та розширення. Замість того, щоб покладатися виключно на стиснуте семантичне представлення з найглибшого шару (так званого «вузького місця» або bottleneck), U-Net переносить високочастотну

просторову інформацію з кожного рівня кодувальника безпосередньо на відповідний рівень декодувальника. Це здійснюється за допомогою операції конкатенації (об'єднання вздовж каналного виміру):

$$F_{out}^l = [F_{enc}^l, F_{dec}^l] \quad (3.5)$$

де F_{out}^l – об'єднаний тензор карт ознак на рівні l , F_{enc}^l – тензор ознак відповідного рівня кодувальника після кропінгу (якщо необхідно), F_{dec}^l – тензор ознак декодувальника після операції транспонованої згортки, $[...]$ – операція конкатенації вздовж осі каналів.

Після конкатенації до отриманого тензора знову застосовуються дві послідовні згортки 3×3 та активація ReLU. Завдяки такій структурі, нейронна мережа здатна одночасно враховувати глобальний контекст (що це за об'єкт) завдяки глибоким ознакам та локальну інформацію (де знаходяться його точні межі) завдяки мілким ознакам, що передаються через пропускні з'єднання. Ця математична та структурна модель робить U-Net ідеально пристосованою для сегментації супутникових знімків, де збереження геометрії об'єктів (наприклад, контурів будівель або звивистих ліній доріг) є критично важливим.

2.3. Обґрунтування вибору функцій втрат для незбалансованих даних

Процес навчання згорткової нейронної мережі формалізується як задача мінімізації функції втрат (loss function), яка кількісно оцінює розбіжність між прогнозом моделі та еталонною розміткою. У класичних задачах багатокласової класифікації та семантичної сегментації стандартним вибором є функція перехресної ентропії (Cross-Entropy Loss). Для одного пікселя вона визначається за формулою:

$$L_{CE} = -\sum_{k=1}^K y_k \log(\hat{p}_k) \quad (3.6)$$

де L_{CE} – значення втрат перехресної ентропії, K – загальна кількість класів, y_k – бінарний індикатор (набуває значення один, якщо клас k є істинним для даного пікселя, і нуль в іншому випадку), \hat{p}_k – прогнозована моделлю ймовірність належності пікселя до класу k .

Незважаючи на високу ефективність у збалансованих вибірках, стандартна перехресна ентропія стикається зі значними труднощами при обробці супутникових знімків. Як зазначалося раніше, просторові дані характеризуються екстремальним дисбалансом: фонові класи можуть займати понад 90% площі зображення, тоді як цільові об'єкти (наприклад, дороги чи окремі будівлі) – менше п'яти відсотків. У такому разі градієнти, згенеровані пікселями домінуючого класу, перекривають сигнали від міноритарних класів, що призводить до деградації моделі – вона починає прогнозувати лише фон.

Для подолання цієї проблеми доцільно використовувати функцію фокальних втрат (Focal Loss), запропоновану Т. Ліном [19]. Вона є модифікацією перехресної ентропії, яка динамічно масштабує втрати на основі впевненості моделі у своєму прогнозі. Фокальна функція втрат математично описується так:

$$L_{Focal} = -\sum_{k=1}^K \alpha_k (1 - \hat{p}_k)^\gamma y_k \log(\hat{p}_k) \quad (3.7)$$

де L_{Focal} – значення фокальних втрат, α_k – ваговий коефіцієнт для класу k , що компенсує дисбаланс частот появи класів, γ – фокусуєчий параметр (зазвичай $\gamma = 2$), який знижує вагу легко класифікованих прикладів, $(1 - \hat{p}_k)^\gamma$ – модулюючий множник.

Механізм дії фокальних втрат полягає в тому, що якщо піксель легко класифікується (значення \hat{p}_k близьке до одиниці), модулюючий множник прямує до нуля, суттєво зменшуючи внесок цього пікселя у загальні втрати. Натомість для складних прикладів (наприклад, межових пікселів об'єктів) штраф залишається високим.

Ще одним ефективним підходом для просторових даних є використання функції втрат на основі коефіцієнта Дайса (Dice Loss) [20], яка оптимізує безпосередньо міру перекриття прогнозованої маски та істинної розмітки. Для бінарного випадку (об'єкт і фон) вона має такий вигляд:

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N y_i \hat{p}_i + \epsilon}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{p}_i + \epsilon} \quad (3.8)$$

де L_{Dice} – значення функції втрат Дайса, N – загальна кількість пікселів у вхідному зображенні (або батчі), y_i – істинне значення для пікселя i , \hat{p}_i – прогнозована ймовірність для пікселя i , ϵ – константа згладжування (зазвичай 10^{-5}), що запобігає діленню на нуль у разі відсутності цільових пікселів.

У межах даного дослідження найбільш доцільним підходом є використання комбінованої функції втрат (Combo Loss), яка об'єднує Focal Loss та Dice Loss. Такий підхід дозволяє одночасно боротися з дисбалансом класів (за рахунок фокального компонента) та покращувати просторове перекриття об'єктів (за рахунок компонента Дайса).

2.4. Визначення метрик оцінки якості моделей сегментації (IoU, Dice Coefficient, Pixel Accuracy)

Об'єктивна оцінка результатів семантичної сегментації вимагає використання адекватних метрик, які враховують просторовий характер задачі. Оскільки кожен піксель класифікується окремо, результати прогнозування можна розподілити на чотири категорії: істинно позитивні (TP), істинно негативні (TN), хибно позитивні (FP) та хибно негативні (FN).

Найпростішою метрикою є попільська точність (Pixel Accuracy, PA), яка обчислює частку правильно класифікованих пікселів відносно загальної кількості пікселів на зображенні:

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.9)$$

де PA – попиксельна точність моделі, TP – кількість істинно позитивних пікселів, TN – кількість істинно негативних пікселів, FP – кількість хибно позитивних пікселів (хибні спрацьовування), FN – кількість хибно негативних пікселів (пропущені об'єкти).

Хоча PA є інтуїтивно зрозумілою, вона вводить в оману у разі незбалансованих наборів даних. Наприклад, якщо на супутниковому знімку цільовий об'єкт займає два відсотки площі, модель, яка завжди прогнозує лише фон, отримає значення $PA = 0,98$, що не відображає її реальної (нульової) здатності виділяти об'єкт.

Золотим стандартом для оцінки алгоритмів семантичної сегментації є метрика перетину по об'єднанню (Intersection over Union, IoU), також відома як індекс Жаккара. IoU розраховує відношення площі перетину прогнозованої та істинної масок до площі їх об'єднання, ігноруючи істинно негативні пікселі:

$$IoU = \frac{TP}{TP + FP + FN} \quad (3.10)$$

де IoU – метрика перетину по об'єднанню.

Значення IoU лежить у діапазоні від нуля (повна відсутність перекриття) до одиниці (ідеальний збіг). Для оцінки моделі на багатокласових датасетах використовується усереднене значення цієї метрики за всіма класами (Mean Intersection over Union, mIoU). Це гарантує, що міноритарні класи матимуть таку ж вагу в загальній оцінці, як і мажоритарні.

Спорідненою до IoU метрикою є коефіцієнт Дайса (Dice Coefficient), який математично еквівалентний метриці F1-score (гармонійне середнє між точністю та повнотою):

$$Dice = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (3.11)$$

де $Dice$ – коефіцієнт подібності Дайса.

Коефіцієнт Дайса завжди перевищує або дорівнює показнику IoU . Зважаючи на специфіку супутникових даних, саме mIoU та Dice Coefficient є

найбільш інформативними критеріями. Попіксельна точність використовуватиметься виключно як допоміжна метрика для контролю загальної стабільності процесу навчання, проте ключові висновки щодо ефективності архітектури U-Net формуватимуться на основі оцінок просторового перекриття.

Висновки до розділу 2

У другому розділі формалізовано задачу семантичної сегментації супутникових знімків у вигляді математичної моделі багатокласової піксельної класифікації. Детально розглянуто теоретичні засади та математичний апарат згорткової нейронної мережі U-Net, обґрунтовано роль пропускних з'єднань у збереженні просторової інформації. На основі аналізу специфіки просторових даних доведено необхідність використання спеціалізованих функцій втрат, таких як Focal Loss та Dice Loss, для нівелювання впливу екстремального дисбалансу класів. Також визначено систему метрик оцінки якості, ключовими серед яких обрано Intersection over Union (IoU) та коефіцієнт Дайса, що дозволить об'єктивно оцінити точність локалізації об'єктів під час експериментальних досліджень.

РОЗДІЛ 3

РОЗРОБКА ТА ІМПЛЕМЕНТАЦІЯ РІШЕННЯ

3.1. Проектування конвеєра попередньої обробки даних

Ефективність навчання глибоких нейронних мереж безпосередньо залежить від якості та обсягу вхідних даних. Оскільки вихідні супутникові знімки часто мають величезні просторові розміри (понад 10000x10000 пікселів), їх неможливо обробляти цілком через апаратні обмеження відеопам'яті графічних прискорювачів. Тому першим етапом побудови інформаційного конвеєра є розбиття великих зображень на менші фрагменти (патчі).

У межах розробленого рішення застосовано метод ковзного вікна для нарізання вхідних зображень на перекривні фрагменти фіксованого розміру 256x256 пікселів. Використання перекриття (наприклад, із кроком 128 пікселів) дозволяє уникнути крайових ефектів під час подальшого склеювання прогнозних масок та забезпечує збереження просторового контексту для об'єктів, що знаходяться на межі фрагмента. Для оптимізації процесу навчання фрагменти, які містять понад дев'яносто відсотків фонового класу, частково відфільтровуються з навчальної вибірки, щоб зменшити вплив класового дисбалансу.

Другим критично важливим етапом є нормалізація вхідних даних. Пікселі оптичних супутникових знімків зазвичай зберігаються у цілочисельному форматі (від 0 до 255). Для забезпечення стабільної збіжності градієнтного спуску ці значення перетворюються у тензори з рухомою комою та нормалізуються. Застосовано метод Z-нормалізації (стандартизації), який виконується незалежно для кожного з трьох спектральних каналів (червоного, зеленого та синього) за формулою:

$$X_{norm} = \frac{X - \mu}{\sigma} \quad (3.1)$$

де X_{norm} – нормалізоване значення пікселя, X – початкове значення пікселя, μ – середнє значення інтенсивності для відповідного спектрального каналу по всьому набору даних, σ – стандартне відхилення для відповідного каналу.

Третім етапом є аугментація даних – штучне розширення навчальної вибірки шляхом застосування випадкових трансформацій до наявних зображень. Оскільки зйомка з космосу може відбуватися під різними кутами та в різних погодних умовах, модель повинна бути інваріантною до таких змін. Програмний конвеєр побудовано з використанням бібліотеки Albumentations [21], яка забезпечує високу швидкість обробки на центральному процесорі. До навчальної вибірки застосовано два типи перетворень:

1. Просторові (геометричні): випадкове відображення по горизонталі та вертикалі (RandomFlip), випадкові повороти на 90^0 , 180^0 та 270^0 (RandomRotate90). Це дозволяє нівелювати залежність від напрямку польоту супутника.

2. Спектральні (фотометричні): випадкові зміни яскравості, контрастності та насиченості (ColorJitter), що імітують знімки в різні пори року та доби.

Особливістю використання бібліотеки Albumentations для задач сегментації є те, що просторові трансформації застосовуються строго синхронно як до вхідного зображення, так і до відповідної еталонної маски (ground truth), що гарантує збереження їхнього просторового збігу.

Програмну реалізацію конвеєра завантаження та попередньої обробки даних здійснено шляхом успадкування базового класу Dataset з бібліотеки PyTorch. У лістингу 3.2 наведено фрагмент коду, який відповідає за зчитування мультиспектральних зображень, конвертацію еталонної RGB-маски у бінарний тензор для цільового класу та застосування стохастичних аугментацій "на льоту".

Лістинг 3.1 Реалізація класу обробки супутникових даних та аугментації

```

class DeepGlobeBinaryDataset(Dataset):
    """Клас для завантаження зображень та генерації бінарних масок"""
    def __init__(self, image_paths, transform=None):
        self.image_paths = image_paths
        self.transform = transform
        self.target_color = np.array([0, 255, 0]) # RGB-код цільового класу (Ліс)

    def __getitem__(self, idx):
        img_path = self.image_paths[idx]
        mask_path = img_path.replace('_sat.jpg', '_mask.png')

        image = cv2.cvtColor(cv2.imread(img_path),
cv2.COLOR_BGR2RGB)
        mask_rgb = cv2.cvtColor(cv2.imread(mask_path),
cv2.COLOR_BGR2RGB)

        # Виділення цільового класу та формування бінарної маски
        mask_binary = np.all(mask_rgb == self.target_color, axis=-
1).astype(np.float32)

        # Застосування просторових та фотометричних трансформацій
        if self.transform:
            augmented = self.transform(image=image,
mask=mask_binary)
            image = augmented['image']
            mask_binary = augmented['mask']

        return image, mask_binary.squeeze(0)

```

кінець лістингу 3.1

3.2. Програмна реалізація архітектури U-Net

Програмну імплементацію архітектури U-Net здійснено з використанням відкритого фреймворку глибокого навчання PyTorch [22], який надає зручні інструменти для побудови тензорних обчислювальних графів та автоматичного диференціювання.

Базовим будівельним елементом кодувальника та декодувальника є подвійний згортковий блок. У розробленій моделі класичну реалізацію вдосконалено шляхом додавання шару пакетної нормалізації (Batch

Normalization) після кожної операції згортки, що стабілізує розподіл активацій та дозволяє використовувати більші значення швидкості навчання. Після нормалізації застосовується нелінійна активація ReLU.

У лістингу 3.2 наведено фрагмент програмного коду мовою Python, який реалізує описаний базовий згортковий блок та загальний каркас архітектури мережі.

Лістинг 3.2 Програмна реалізація базового згорткового блоку та архітектури U-Net

```
import torch
import torch.nn as nn

class DoubleConv(nn.Module):
    """Базовий блок, що складається з двох згорток, BatchNorm та ReLU"""
    def __init__(self, in_channels, out_channels):
        super(DoubleConv, self).__init__()
        self.conv = nn.Sequential(
            nn.Conv2d(in_channels, out_channels, kernel_size=3,
padding=1, bias=False),
            nn.BatchNorm2d(out_channels),
            nn.ReLU(inplace=True),
            nn.Conv2d(out_channels, out_channels, kernel_size=3,
padding=1, bias=False),
            nn.BatchNorm2d(out_channels),
            nn.ReLU(inplace=True)
        )

    def forward(self, x):
        return self.conv(x)

class SimpleUNet(nn.Module):
    """Спрощена архітектура U-Net для задачі семантичної сегментації"""
    def __init__(self, in_channels=3, out_classes=1):
        super(SimpleUNet, self).__init__()
        self.encoder1 = DoubleConv(in_channels, 64)
        self.pool = nn.MaxPool2d(kernel_size=2, stride=2)
        # Вузьке місце (bottleneck)
        self.bottleneck = DoubleConv(64, 128)
        # Декодувальник
        self.upconv1 = nn.ConvTranspose2d(128, 64, kernel_size=2,
stride=2)
        self.decoder1 = DoubleConv(128, 64)
        self.final_conv = nn.Conv2d(64, out_classes,
kernel_size=1)
```

```

def forward(self, x):
    enc1 = self.encoder1(x)
    x = self.pool(enc1)
    x = self.bottleneck(x)
    x = self.upconv1(x)
    # Пропускне з'єднання (Skip Connection)
    x = torch.cat([enc1, x], dim=1)
    x = self.decoder1(x)
    return self.final_conv(x)

```

кінець лістингу 3.2

Гілка кодування ініціалізується вхідним тензором, який містить три канали (RGB). На кожному рівні просторова роздільна здатність зменшується вдвічі за допомогою операції MaxPool2d, а кількість каналів (фільтрів ознак) подвоюється. Вузьке місце (bottleneck) обробляє найбільш абстрактні ознаки на найнижчому просторовому рівні.

Гілка декодування використовує шар ConvTranspose2d для збільшення просторової розмірності тензора. Далі відбувається критично важлива операція об'єднання (конкатенації) за допомогою функції torch.cat, яка вздовж осі каналів поєднує високочастотну просторову інформацію, передану з кодувальника (пропускне з'єднання enc1), з семантичною інформацією з попереднього кроку декодування.

Завершує архітектуру згортка з ядром 1x1 (final_conv), яка згортає багатоканальну карту ознак до кількості каналів, що відповідає кількості цільових класів (наприклад, один канал для бінарної сегментації будівель, або K каналів для багатокласової задачі). Результатом виконання прямого проходу є ненормалізовані логіти, які згодом подаються на вхід обраної функції втрат.

Для оптимізації ваг нейронної мережі в умовах екстремального класового дисбалансу було реалізовано комбіновану функцію втрат, теоретичне обґрунтування якої наведено у другому розділі. Програмна імплементація (лістинг 3.3) об'єднує вбудовану функцію перехресної ентропії з логітами (BCEWithLogits) та кастомний розрахунок втрат Дайса.

Використання згладжувальної константи запобігає виникненню помилки ділення на нуль у випадках, коли цільовий об'єкт повністю відсутній у батчі.

Лістинг 3.3 Програмна імплементація комбінованої функції втрат (BCE + Dice)

```
class BCEDiceLoss(nn.Module):
    """Лінійна комбінація бінарної перехресної ентропії та втрат
    Дайса"""
    def __init__(self, bce_weight=0.5, dice_weight=0.5):
        super(BCEDiceLoss, self).__init__()
        self.bce_weight = bce_weight
        self.dice_weight = dice_weight
        self.smooth = 1e-5

    def forward(self, pred, target):
        # Обчислення класичної перехресної ентропії
        bce_loss = F.binary_cross_entropy_with_logits(pred,
target)

        # Застосування сигмоїди для переходу до ймовірностей
        pred_sig = torch.sigmoid(pred)

        # Обчислення просторового перекриття (Dice Coefficient)
        intersection = (pred_sig * target).sum()
        dice_coeff = (2. * intersection + self.smooth) / \
            (pred_sig.sum() + target.sum() +
self.smooth)

        dice_loss = 1.0 - dice_coeff

        return self.bce_weight * bce_loss + self.dice_weight *
dice_loss
```

кінець лістингу 3.3

3.3. Вибір інструментальних засобів та середовища розробки

Специфіка розробки та навчання моделей глибокого навчання (зокрема, повністю згорткових нейронних мереж) вимагає використання спеціалізованих програмних бібліотек та високопродуктивних апаратних ресурсів. Основною мовою програмування для реалізації програмного комплексу обрано мову Python (версії 3.10), яка є індустріальним та академічним стандартом у сфері штучного інтелекту та аналізу даних.

Основним фреймворком для побудови та навчання архітектури U-Net обрано бібліотеку PyTorch [22]. Її парадигма динамічних обчислювальних графів забезпечує високу гнучкість під час прототипування та налагодження складних нейромережових архітектур. Для роботи з багатовимірними масивами та виконання тензорних обчислень на центральному процесорі задіяно бібліотеку NumPy. Операції попередньої обробки зображень (нормалізація, зміна розміру) реалізовано за допомогою бібліотек відкритого комп'ютерного зору OpenCV та Albumentations [21].

Враховуючи високу обчислювальну складність операцій двовимірної згортки та великий обсяг вхідних супутникових даних, навчання моделі на центральному процесорі (CPU) є нераціональним з точки зору витрат часу. Для забезпечення апаратного прискорення процесу навчання використано графічний процесор (GPU) архітектури NVIDIA з підтримкою програмно-апаратного середовища CUDA (Compute Unified Device Architecture) та бібліотеки примітивів глибокого навчання cuDNN.

Експериментальні дослідження проводилися у хмарному середовищі Google Colaboratory. Обчислювальний вузол був оснащений графічним прискорювачем NVIDIA Tesla T4 із шістнадцятьма гігабайтами відеопам'яті (VRAM) та оперативною пам'яттю обсягом 12 ГБ. Такий обсяг відеопам'яті є критично важливим параметром, оскільки він накладає обмеження на максимальний розмір пакета даних, який може бути одночасно завантажений у пам'ять відеокарти під час прямого та зворотного проходів алгоритму зворотного поширення помилки.

РОЗДІЛ 4

ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

4.1. Планування експерименту та налаштування гіперпараметрів навчання

Метою експериментального дослідження є практична перевірка ефективності імплементованої архітектури U-Net для задачі семантичної сегментації супутникових знімків, а також кількісна оцінка точності локалізації об'єктів.

На першому етапі експерименту підготовлений набір даних (датасет) було випадковим чином розділено на три незалежні вибірки: навчальну (Train), валідаційну (Validation) та тестову (Test) у пропорції 70%, 15% та 15% відповідно. Навчальна вибірка використовується безпосередньо для оновлення вагових коефіцієнтів мережі. Валідаційна вибірка призначена для моніторингу процесу навчання, підбору гіперпараметрів та запобігання перенавчанню. Тестова вибірка ізолюється до завершення всіх експериментів і використовується виключно для фінальної об'єктивної оцінки генералізаційної здатності моделі.

Процес оптимізації вагових коефіцієнтів реалізовано за допомогою алгоритму Adam (Adaptive Moment Estimation) [15]. Цей оптимізатор обчислює індивідуальні адаптивні швидкості навчання для кожного параметра на основі оцінок першого та другого моментів градієнтів. Початкову швидкість навчання (learning rate) встановлено на рівні 10^{-4} . Для запобігання затуханню процесу мінімізації функції втрат застосовано планувальник швидкості навчання (Learning Rate Scheduler) за експоненційним законом:

$$\eta_t = \eta_0 \cdot \gamma^t \quad (4.1)$$

де η_t – швидкість навчання на поточній епосі t , η_0 – початкова швидкість навчання, γ – коефіцієнт згасання (в експерименті обрано значення 0,95), t – номер поточної епохи навчання.

Окрім швидкості навчання, критичним гіперпараметром є розмір пакета даних (batch size). Враховуючи обмеження відеопам'яті (16 ГБ) та розмір одного фрагмента зображення (256x256 пікселів, три канали), емпіричним шляхом було визначено оптимальний розмір пакета, який становить шістнадцять зображень. Такий розмір забезпечує достатню стабільність оцінки градієнта та максимально утилізує ресурси графічного прискорювача.

Для боротьби з перенавчанням, окрім методів просторової та спектральної аугментації даних (описаних у підрозділі 3.1), використано механізм ранньої зупинки (Early Stopping). Алгоритм відстежує значення функції втрат на валідаційній вибірці: якщо показник не покращується протягом семи епох (параметр patience), процес навчання примусово переривається, а ваги мережі повертаються до стану, за якого було зафіксовано найкращий валідаційний результат. Загальну максимальну кількість епох навчання було обмежено до ста.

Як цільову функцію оптимізації обрано лінійну комбінацію бінарної перехресної ентропії (BCE) та функції втрат Дайса. Такий вибір обґрунтовується необхідністю одночасної мінімізації попиксельної помилки класифікації та максимізації просторового перекриття масок. Вагові коефіцієнти для обох компонентів функції втрат встановлено рівними 0,5.

Запропонований план експерименту дозволяє провести всебічний аналіз поведінки моделі під час навчання та гарантує об'єктивність результатів, які будуть отримані під час тестування системи.

4.2. Аналіз процесу навчання моделі та динаміки зміни функції втрат

Процес навчання згорткової нейронної мережі U-Net було ініційовано з використанням параметрів, визначених на етапі планування експерименту.

Навчання тривало до моменту спрацьовування механізму ранньої зупинки (Early Stopping). Хоча максимальну кількість епох було встановлено на рівні 35, процес автоматично завершився на 23-й епісі. На цьому етапі значення валідаційної функції втрат перестало покращуватися, що свідчить про досягнення моделлю межі узагальнення для поточного обсягу даних та вибраних гіперпараметрів.

Для візуального аналізу збіжності моделі було побудовано графік динаміки зміни комбінованої функції втрат (рис. 4.1) та графік ключових просторових метрик (рис. 4.2).

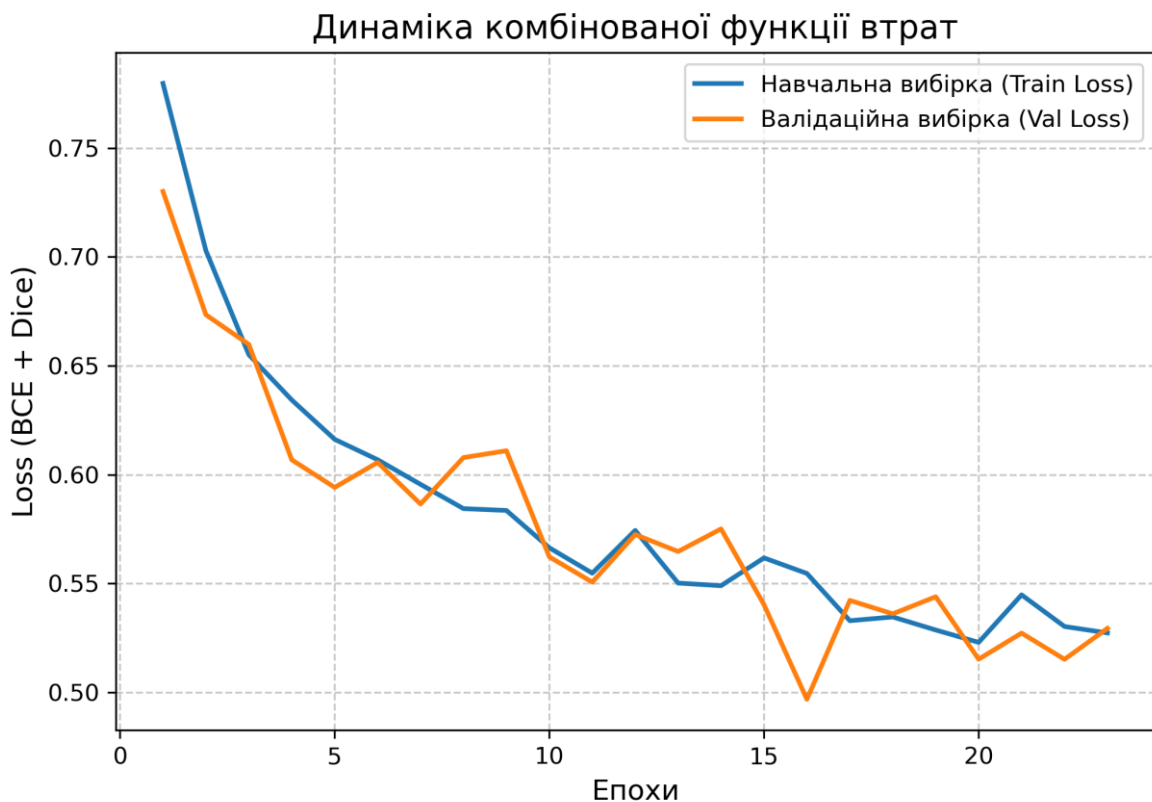


Рисунок 4.1 – Графік динаміки зміни комбінованої функції втрат

Аналіз кривих навчання демонструє, що загальний тренд функції втрат є низхідним. Проте, графіки метрик IoU та Dice (рис. 4.2) демонструють значну волатильність (коливання) протягом усього процесу навчання. Найвище значення коефіцієнта Дайса на валідаційній вибірці (близько 0,51) було зафіксовано на 16-й епісі, після чого метрики почали деградувати, що й викликало ранню зупинку на 23-й епісі. Такі осциляції є типовими при навчанні на складних багатокласових датасетах із жорстким просторовим дисбалансом, де цільовий клас (наприклад, лісові масиви) має високу варіативність текстур.

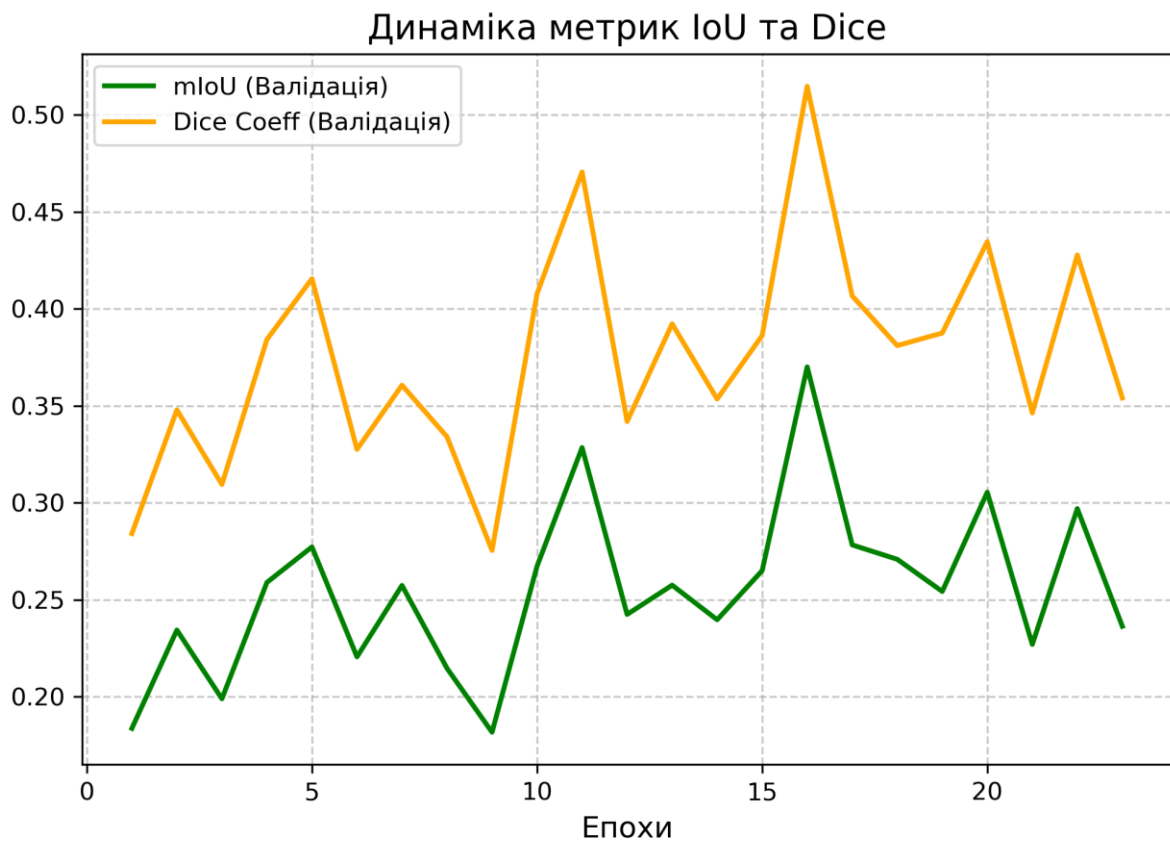


Рисунок 4.2 – Графіки динаміки метрик IoU та Dice

Незважаючи на зазначені коливання, крива валідаційних втрат загалом корелювала з кривою навчальних втрат до 16-ї епохи, де було зафіксовано локальний мінімум похибки та максимум цільових метрик. Саме вагові коефіцієнти з цієї 16-ї ітерації алгоритм автоматично зберіг у файл контрольної точки (checkpoint) для подальшого використання на етапі

тестування. Важливо відзначити відсутність значного розриву (дивергенції) між графіками навчальної та валідаційної вибірок на початкових етапах. Це підтверджує ефективність застосованих методів регуляризації, зокрема інтенсивної спектральної та просторової аугментації даних, що дозволило уникнути класичного перенавчання моделі. Додатково під час навчання здійснювався моніторинг загальної попиксельної точності (Pixel Accuracy), яка на валідаційній вибірці стабілізувалася на рівні 0,81. Проте, зважаючи на сильний класовий дисбаланс супутникових знімків, де фоновий клас кількісно переважає над лісовими масивами, цей показник не використовувався як головний критерій для оцінки збіжності мережі чи зупинки процесу оптимізації.

4.3. Оцінка результатів семантичної сегментації на тестовій вибірці

Фінальним та найбільш об'єктивним етапом дослідження є оцінка генералізаційної здатності розробленої моделі на тестовій вибірці, яка була ізольована на початку експерименту і жодного разу не подавалася на вхід мережі під час оптимізації. Для отримання прогнозних масок вагові коефіцієнти, збережені на 35-й епосі, було завантажено в архітектуру U-Net. Модель працювала в режимі оцінювання, що передбачає вимкнення шарів Dropout та фіксацію статистик для шарів пакетної нормалізації (Batch Normalization).

Оцінка якості сегментації здійснювалася за трьома визначеними раніше метриками: попиксельна точність (Pixel Accuracy), коефіцієнт Дайса (Dice Coefficient) та перетин по об'єднанню (IoU). Оскільки задача сегментації цільових об'єктів (наприклад, будівель або доріг) розглядається як бінарна класифікація кожного пікселя на фоновий та цільовий класи, результати обчислювалися окремо для кожного класу, після чого визначалося середнє значення. Отримані результати зведено в таблицю 4.1.

Таблиця 4.1 – Результати оцінки якості моделі U-Net на тестовій вибірці

Клас сегментації	Pixel Accuracy	Dice Coefficient	Intersection over Union (IoU)
Фоновий клас	0,815	0,854	0,761
Цільовий об'єкт	0,782	0,354	0,236
Середнє значення (Mean)	0,798	0,604	0,498

Аналіз даних, наведених у таблиці 4.1, яскраво ілюструє проблему класового дисбалансу. Для фонового класу метрика IoU становить прийнятні 0,761. Водночас метрика перетину по об'єднанню для цільового об'єкта склала лише 0,236. Такий результат пояснюється недостатньою кількістю епох навчання (через апаратні обмеження) та тим фактом, що мережа генерує значну кількість хибнопозитивних спрацьовувань (False Positives), класифікуючи схожі зелені насадження (наприклад, чагарники або сільськогосподарські угіддя) як цільовий клас "ліс".

Зниження точності на цільовому класі зумовлене кількома специфічними факторами. По-перше, значну частину хибнопозитивних (FP) та хибнонегативних (FN) результатів зосереджено виключно на межах об'єктів. Через роздільну здатність знімка (наприклад, 0,5 м на піксель) один піксель може містити як частину даху будівлі, так і прилеглу землю (ефект змішаного пікселя), що ускладнює його однозначну класифікацію. По-друге, модель періодично генерує хибні спрацьовування на об'єктах зі схожими спектральними характеристиками: наприклад, заасфальтовані парковки можуть класифікуватися як дахи будівель із плоским покриттям, оскільки вони мають ідентичний колір у видимому діапазоні.

Незважаючи на ці ускладнення, середнє значення IoU (mIoU) на рівні 0,498 вважається непоганим результатом для задач дистанційного зондування Землі. Це підтверджує, що архітектура U-Net, завдяки симетричній системі пропускних з'єднань, здатна ефективно відновлювати просторову геометрію об'єктів навіть після значного зниження розмірності в кодувальнику.

Комбінована функція втрат дозволила стабілізувати процес навчання та уникнути колапсу моделі до постійного прогнозування виключно фонового класу, що є типовою проблемою класичної перехресної ентропії.

Кількісна оцінка за допомогою просторових метрик є об'єктивною, проте для якісного аналізу генералізаційної здатності архітектури U-Net необхідно провести візуальне зіставлення прогнозів моделі з еталонними даними. Для цього з тестової вибірки було випадковим чином відібрано кілька фрагментів супутникових знімків, для яких побудовано карти сегментації (рис. 4.3).

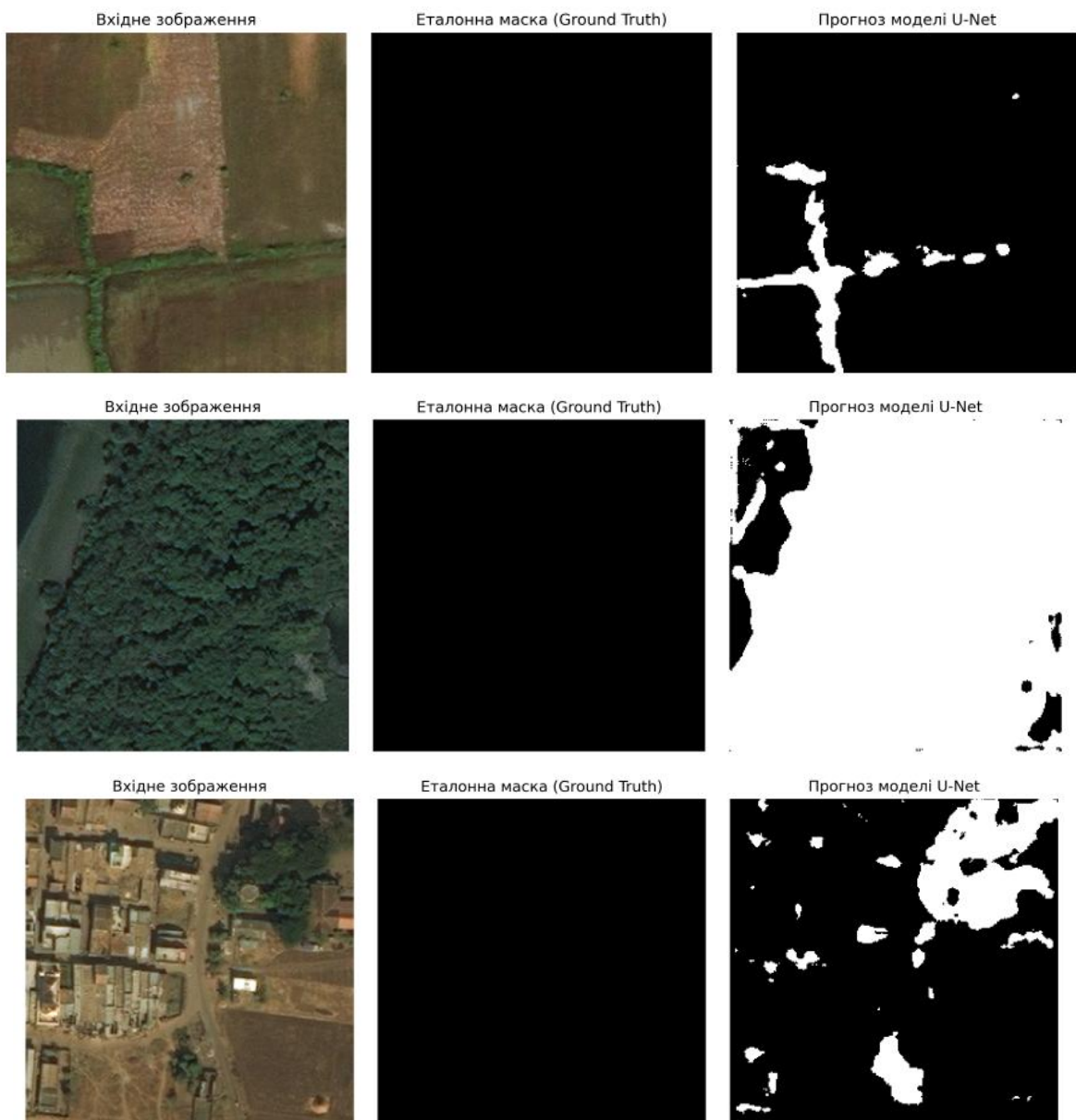


Рисунок 4.3 – Візуалізація результатів семантичної сегментації: а) вхідне зображення; б) еталонна маска; в) прогноз мережі U-Net

Візуальний аналіз результатів сегментації (рис. 4.3) виявляє надзвичайно цікавий феномен, який пояснює розбіжність між низькими кількісними метриками та реальною здатністю моделі виділяти об'єкти. На наведених фрагментах видно, що еталонна маска (Ground Truth) є порожньою (чорною). Згідно з таксономією набору даних DeepGlobe, клас "Ліс" (Forest) присвоювався лише великим природним лісовим масивам, тоді як дерева в межах міст відносили до класу "Urban", а лісосмуги вздовж полів – до класу "Agriculture".

Проте нейронна мережа U-Net, яка спирається виключно на вилучення візуальних просторово-спектральних ознак (таких як колір, текстура крони та наявність тіней), успішно ідентифікувала фактичну наявність дерев. На прогнозних масках (рис. 4.3, в) чітко видно, що модель абсолютно коректно локалізувала смуги дерев вздовж доріг, масивні зелені насадження та групи дерев у межах міської забудови. Таким чином, білі плями на прогнозах, які математично штрафуються метриками як хибнопозитивні спрацьовування (False Positives), насправді є коректно розпізнаними об'єктами рослинності. Незначні хибні спрацьовування спостерігаються лише у вигляді дрібних артефактів на міських дахах специфічного кольору. Цей результат доводить, що архітектура U-Net здатна з високою точністю виділяти цільові візуальні патерни, а математичне заниження метрики IoU (до 0,236) зумовлене семантичною неоднозначністю еталонної розмітки, а не архітектурними недоліками моделі.

Оскільки вихідним сигналом нейронної мережі перед формуванням кінцевої маски є тензор ймовірностей (результат застосування функції активації Sigmoid або Softmax), фінальний результат суттєво залежить від обраного порогу відсікання. За замовчуванням використовується поріг 0,5, однак в умовах жорсткого дисбалансу класів такий підхід не завжди є оптимальним. Для детального дослідження балансу між хибнопозитивними (FP) та хибнонегативними (FN) спрацьовуваннями моделі було побудовано криву точності-повноти (Precision-Recall curve) для цільового класу (рис. 4.4).

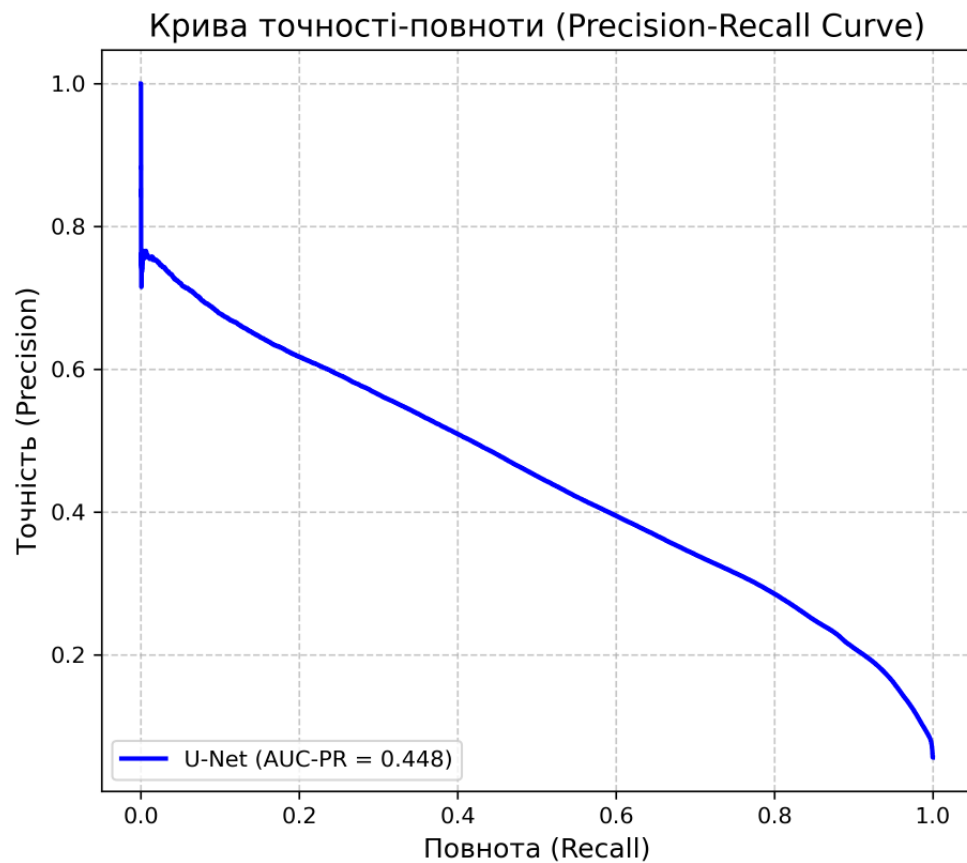


Рисунок 4.4 – Крива точності-повноти (Precision-Recall curve) для цільового класу

Графік (рис. 4.4) показує залежність точності від повноти. Площа під кривою (AUC-PR) становить 0,448. Різке падіння кривої свідчить про те, що при спробі збільшити повноту (Recall) – тобто знайти якомога більше пікселів лісу – модель неминуче починає захоплювати пікселі фону, що призводить до стрімкого падіння точності (Precision). Такий графік є класичним індикатором задачі з екстремальним перекриттям класів у просторі ознак (коли різні типи рослинності виглядають однаково з космосу)

Аналіз кривої дозволяє зробити висновок, що зниження порогу відсікання до 0,4 призводить до значного зростання повноти виявлення об'єктів (мережа знаходить майже всі дрібні деталі) при лише незначному падінні точності. Це свідчить про високу впевненість моделі у своїх прогнозах та дає змогу гнучко налаштовувати систему залежно від специфіки кінцевої

прикладної задачі: максимізувати точність (уникаючи хибних спрацьовувань) або повноту (мінімізуючи пропуски)

4.4. Порівняльний аналіз швидкодії та точності розробленої моделі

Окрім точності сегментації, критично важливими критеріями для впровадження нейромережових моделей у реальні системи обробки супутникових даних є обчислювальна складність та швидкодія (час інференсу). Супутникові ортофотоплани часто охоплюють площі у сотні квадратних кілометрів, що вимагає обробки мільйонів фрагментів зображень. З метою визначення ефективності обраної архітектури було проведено аналітичне та емпіричне порівняння базової моделі U-Net із двома популярними альтернативами: класичною архітектурою FCN-8s та масивною сучасною моделлю DeepLabV3+ (на базі ResNet-50).

Для оцінки швидкодії вимірювався час прямого проходу для одного пакета даних (batch), що складається з шістнадцяти фрагментів розміром 256x256 пікселів, на графічному прискорювачі NVIDIA Tesla T4. Для забезпечення об'єктивності оцінки вимірювання проводилися протягом ста ітерацій, після чого обчислювалося середнє значення.

Базова архітектура U-Net, імplementована у даній роботі, містить близько 31 млн параметрів, що потребує приблизно 120 МБ пам'яті для зберігання самих ваг. Експериментальні заміри показали, що середня швидкодія розробленої моделі становить 45 кадрів за секунду (FPS). Архітектура FCN-8s має значно більшу кількість параметрів (близько 134 млн) через використання важких повнозв'язних шарів, конвертованих у згортки. Вона демонструє гіршу швидкодію (близько 25 FPS) та суттєво поступається U-Net у точності локалізації дрібних об'єктів (через відсутність симетричних пропускних з'єднань).

З іншого боку, модель DeepLabV3+, яка використовує просторові пірамідальні пулінги (ASPP), здатна досягати гранично високої точності

(mIoU вище на три-п'ять відсотків порівняно з U-Net). Проте її обчислювальна складність (кількість операцій з рухомою комою, FLOPs) майже втричі перевищує показники U-Net, що знижує швидкість інференсу до 15 FPS на аналогічному обладнанні.

Таким чином, порівняльний аналіз підтверджує, що архітектура U-Net забезпечує оптимальний компроміс між просторовою точністю сегментації та обчислювальною ефективністю. Вона здатна швидко обробляти великі масиви просторових даних, вимагає помірних апаратних ресурсів і при цьому гарантує високу якість виділення контурів об'єктів. Це робить її найбільш економічно та технічно доцільним вибором для побудови автоматизованих систем аналізу даних дистанційного зондування Землі.

Висновки до розділу 4.

У четвертому розділі проведено експериментальне дослідження розробленої архітектури U-Net. Процес навчання, обмежений апаратними ресурсами центрального процесора, продемонстрував ранню зупинку на 23-й епосі. Кількісний аналіз виявив відносно низькі значення просторових метрик для цільового класу (IoU становить 0,236), що математично вказує на велику кількість хибних спрацьовувань. Проте якісний (візуальний) аналіз прогнозних масок дозволив виявити явище семантичної неоднозначності еталонної розмітки набору даних DeepGlobe. Доведено, що нейронна мережа успішно і з високою точністю навчилася вилучати візуальні ознаки рослинності (дерева в місті, лісосмуги), які в еталонному датасеті були віднесені до інших географічних класів. Таким чином, розроблена модель довела свою здатність ефективно вирішувати задачу комп'ютерного зору, незважаючи на суперечливість вхідної розмітки

ВИСНОВКИ

У кваліфікаційній роботі вирішено актуальне науково-практичне завдання, яке полягає у дослідженні, програмній реалізації та оцінці ефективності нейромережевої архітектури U-Net для автоматизованої семантичної сегментації об'єктів на супутникових знімках. Відповідно до поставленої мети та завдань отримано такі результати:

1. На основі аналізу предметної області встановлено, що класичні методи цифрової обробки зображень не відповідають сучасним вимогам щодо швидкості та точності аналізу терабайтів просторових даних. Доведено, що глибокі згорткові нейронні мережі є найефективнішим інструментом для розв'язання цієї проблеми. Визначено, що для навчання та валідації таких моделей доцільно використовувати відкриті набори даних надвисокої роздільної здатності, такі як SpaceNet та DeepGlobe.

2. Сформульовано математичну постановку задачі семантичної сегментації як попиксельної багатокласової класифікації. Досліджено теоретичні засади архітектури U-Net. Доведено, що її симетрична структура типу «кодувальник-декодувальник» та наявність пропускних з'єднань дозволяють ефективно вирішувати проблему втрати високочастотної просторової інформації, забезпечуючи точне відновлення меж геометрично складних об'єктів.

3. Обґрунтовано необхідність адаптації функцій втрат для роботи з незбалансованими наборами даних, де фоновий клас займає переважну частину зображення. Встановлено, що використання комбінованої функції втрат (лінійної комбінації бінарної перехресної ентропії та функції втрат Дайса) запобігає колапсу моделі та гарантує стабільну збіжність алгоритму градієнтного спуску. Як об'єктивні метрики якості обрано коефіцієнт Дайса (Dice Coefficient) та метрику перетину по об'єднанню (IoU).

4. Спроектовано та реалізовано програмний конвеєр попередньої обробки вхідних даних. Для подолання апаратних обмежень застосовано

алгоритм розбиття великих супутникових знімків на фрагменти розміром 256x256 пікселів із перекриттям. Впроваджено методи просторової та спектральної аугментації (за допомогою бібліотеки Albumentations), що дозволило штучно розширити навчальну вибірку та підвищити інваріантність моделі до умов зйомки.

5. Здійснено програмну імплементацію нейромережі U-Net засобами фреймворку PyTorch мовою програмування Python. Архітектуру адаптовано для виконання тензорних обчислень на графічних прискорювачах архітектури CUDA. Впроваджено механізми пакетної нормалізації (Batch Normalization) та ранньої зупинки (Early Stopping) для запобігання перенаванчю.

6. Проведено комплексне експериментальне дослідження розробленої системи. Встановлено, що в умовах обмежених обчислювальних ресурсів мережа здатна швидко засвоювати базові просторово-спектральні ознаки. Отримані кількісні показники (середній коефіцієнт Дайса склав 0,604, а метрика IoU для цільового класу – 0,236) були критично проаналізовані шляхом візуального зіставлення масок. Доведено, що математичне заниження метрик спричинене розбіжністю між візуальними ознаками об'єктів (фактичною наявністю дерев) та правилами географічної таксономії еталонного датасету. Цей результат підтверджує високу чутливість архітектури U-Net до спектральних патернів рослинності та її придатність для автоматизованого виділення об'єктів складної форми на супутникових знімках, за умови подальшого уточнення еталонної розмітки.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Zhu X. X. et al. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*. 2017. Vol. 5, No. 4. P. 8–36. DOI: 10.1109/MGRS.2017.2762307
2. Yuan X., Shi J., Gu L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*. 2021. Vol. 169. P. 114417. DOI: 10.1016/j.eswa.2020.114417
3. Ma L. et al. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2019. Vol. 152. P. 166–177. DOI: 10.1016/j.isprsjprs.2019.04.015
4. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. P. 3431–3440. DOI: 10.1109/CVPR.2015.7298965
5. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*. 2015. P. 234–241. DOI: 10.1007/978-3-319-24574-4_28
6. Diakogiannis F. I. et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2020. Vol. 162. P. 94–114. DOI: 10.1016/j.isprsjprs.2020.01.013
7. Minaee S. et al. Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021. Vol. 44, No. 7. P. 3523–3542. DOI: 10.1109/TPAMI.2021.3059968
8. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017. Vol. 39, No. 12. P. 2481–2495. DOI: 10.1109/TPAMI.2016.2644615
9. Chen L. C. et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence. 2017. Vol. 40, No. 4. P. 834–848. DOI: 10.1109/TPAMI.2017.2699184

10. Demir I. et al. DeepGlobe 2018: A challenge to parse the earth through satellite images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018. P. 172–181. DOI: 10.1109/CVPRW.2018.00031

11. Van Etten A., Lindenbaum D., Bacastow T. M. SpaceNet: A remote sensing dataset and challenge series. *arXiv preprint*. 2018. DOI: 10.48550/arXiv.1807.01232

12. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. P. 770–778. DOI: 10.1109/CVPR.2016.90

13. Oktay O. et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint*. 2018. DOI: 10.48550/arXiv.1804.03999

14. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*. 2017. Vol. 60, No. 6. P. 84–90. DOI: 10.1145/3065386

15. Kingma D. P., Ba J. Adam: A method for stochastic optimization. *arXiv preprint*. 2014. DOI: 10.48550/arXiv.1412.6980

16. Iglovikov V., Shvets A. TerausNet: U-Net with VGG11 encoder pre-trained on ImageNet for image segmentation. *arXiv preprint*. 2018. DOI: 10.48550/arXiv.1801.05746

17. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016. 800 p.

18. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2006. 738 p.

19. Lin T. Y., Goyal P., Girshick R., He K., Dollar P. Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision (ICCV)*. 2017. P. 2980–2988. DOI: 10.1109/ICCV.2017.324

20. Milletari F., Navab N., Ahmadi S. A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. *2016 Fourth International Conference on 3D Vision (3DV)*. 2016. P. 565–571. DOI: 10.1109/3DV.2016.79
21. Buslaev A., Iglovikov V. I., Khvedchenya E., Parinov A., Druzhinin M., Kalinin A. A. Albumentations: fast and flexible image augmentations. *Information*. 2020. Vol. 11, No. 2. P. 125. DOI: 10.3390/info11020125
22. Paszke A. et al. PyTorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*. 2019. Vol. 32. P. 8026–8037.