Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Розробка масштабованої глибокої моделі для багатокласової сегментації супутникових зображень в умовах обмеженої вибірки даних
(тема)

Виконав:
здобувач другого року навчання,
групи СШІм-23-1
Богдан Ткаченко
(власне ім'я, прізвище)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)
Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Системи штучного інтелекту
(повна назва освітньої програми)
Керівник проф. Євгеній Бодянський
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ШІ _____ Олег ЗОЛОТУХІН
(підпис) (власне ім'я, прізвище)

2025 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

Кафедра Штучного інтелекту

Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

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

Освітня програма Системи штучного інтелекту
(повна назва)

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

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Ткаченку Богдану Сергійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Розробка масштабованої глибокої моделі для багатокласової сегментації супутникових зображень в умовах обмеженої вибірки даних

затверджена наказом університету від 21 квітня 20 25 р. № 295Ст

2. Термін подання студентом роботи до екзаменаційної комісії 6 червня 20 25 р.

3. Вихідні дані до роботи Науково-технічні публікації, що стосуються глибокого навчання та методів сегментації супутникових зображень, дані Інтернет-джерел та актуальних наукових проєктів у сфері штучного інтелекту, документація Python та бібліотек TensorFlow/Keras, набір зображень для тренування, валідації та тестування моделі сегментації, моделі попереднього навчання (InceptionResNetV2) з відкритим доступом

4. Перелік питань, що потрібно опрацювати в роботі
1) Аналіз предметної галузі та постановка задачі дослідження
2) Методологія сегментації супутникових зображень
3) Експериментальні результати та оцінка системи

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	21.04.2025	виконано
2	Аналіз предметної галузі	27.04.2025	виконано
3	Формулювання проблеми	29.04.2025	виконано
4	Збір та попередня обробка даних	12.05.2025	виконано
5	Програмна реалізація глибокої моделі	15.05.2025	виконано
6	Тестування розробленої моделі	18.05.2025	виконано
7	Написання пояснювальної записки	21.05.2025	виконано
8	Захист перед ЕК	06.06.2025	

Дата видачі завдання 21 квітня 20 25 р.

Здобувач _____
(підпис)

Керівник роботи _____ проф. Євгеній Бодянський
(підпис) (посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка: 83 с., 16 рис., 1 табл., 1 дод., 30 джерел.

ГЛИБОКЕ НАВЧАННЯ, ЗГОРТКОВА НЕЙРОННА МЕРЕЖА, КОМП'ЮТЕРНИЙ ЗІР, СЕМАНТИЧНА СЕГМЕНТАЦІЯ СУПУТНИКОВИХ ЗОБРАЖЕНЬ.

Об'єкт дослідження – системи автоматичної сегментації та класифікації об'єктів на супутникових зображеннях.

Предмет дослідження – методи та алгоритми багатокласової сегментації супутникових зображень за допомогою глибоких нейронних мереж в умовах обмеженої вибірки даних.

Мета роботи – розробка та дослідження масштабованої глибокої моделі для багатокласової сегментації супутникових зображень, що забезпечує високу точність сегментації при обмеженій кількості навчальних даних.

Методи дослідження – аналіз існуючих рішень, математичне моделювання, експериментальне дослідження, статистичний аналіз, методи аугментації даних, трансферне навчання, комп'ютерне моделювання та програмна реалізація.

Проведено комплексний аналіз сучасних методів сегментації супутникових зображень, досліджено архітектури згорткових нейронних мереж та методи їх оптимізації. Використано спеціалізований набір даних для навчання мережі, що містить всього 8 супутникових зображень високої роздільної здатності.

На основі проведених досліджень розроблено оригінальну архітектуру CNN, що досягає точності в 92% для використаних класів сегментації.

ABSTRACT

Master’s thesis contains: 83 pp., 16 fig., 1 tabl., 1 ann., 30 references.

COMPUTER VISION, CONVOLUTIONAL NEURAL NETWORK, DEEP LEARNING, SEMANTIC SEGMENTATION OF SATELLITE IMAGES.

Object of the study – systems for automatic segmentation and classification of objects in satellite images.

Subject of the study – methods and algorithms for multiclass segmentation of satellite images using deep neural networks under limited data conditions.

Purpose of the work – to develop and investigate a scalable deep model for multiclass satellite image segmentation that ensures high segmentation accuracy with a limited amount of training data.

Research methods – analysis of existing solutions, mathematical modeling, experimental research, statistical analysis, data augmentation methods, transfer learning, computer modeling, and software implementation.

A comprehensive analysis of modern methods for satellite image segmentation was conducted. Architectures of convolutional neural networks and their optimization methods were studied. A specialized dataset was used to train the network, containing only 8 high-resolution satellite images.

Based on the conducted research, an original CNN architecture was developed, achieving 92% accuracy for the segmentation classes used.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів 8
Вступ 10
1 Аналіз предметної галузі та постановка задачі дослідження 12
1.1 Завдання сегментації супутникових зображень 12
1.1.1 Роль семантичної сегментації для дистанційного зондування . 12
1.1.2 Виклики пов‘язані з сегментації супутникових зображень 16
1.2 Огляд сучасних методів багатокласової сегментації 24
1.2.1 Традиційні методи багатокласової сегментації 24
1.2.2 Глибокі нейронні мережі в багатокласовій сегментації 28
1.3 Критерії оцінки методів багатокласової сегментації 35
1.4 Постановка задачі дослідження 38
2 Методологія сегментації супутникових зображень 41
2.1 Архітектура моделі: U-Net з InceptionResNetV2 41
2.1.1 Структура енкодера та декодера 41
2.1.2 Виділення ознак на різних масштабах 45
2.2 Передобробка та аугментація даних 49
2.3 Стратегія навчання 55
3 Експериментальні результати та оцінка системи 60
3.1 Огляд використаного набору даних 60
3.1.1 Аналіз набору даних та розбивка на підвиборки 60
3.1.2 Методика аугментації набору даних 62
3.2 Налаштування середовища експерименту 64
3.3 Навчання моделі та метрики оцінки 66
3.3.1 Опис процесу навчання моделі 66
3.3.2 Аналіз результатів навчання 69
3.4 Оцінка роботи моделі на знімках різного масштабу 72
3.5 Напрями подальших досліджень 74
Висновки 76

Перелік джерел посилання 78
Додаток А Відомість кваліфікаційної роботи..... 83

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ГІС – геоінформаційна система;

ОАЕ – Об‘єднані Арабські Емірати;

API – Application Programming Interface – інтерфейс прикладного програмування;

ASPP – Atrous Spatial Pyramid Pooling – проріджене просторове пірамідальне згорткове об'єднання;

CLAHE – Contrast Limited Adaptive Histogram Equalization – адаптивне вирівнювання гістограми з обмеженням контрасту;

CNN – Convolutional Neural Network – згорткова нейронна мережа;

CPU – Central Processing Unit – центральний процесор;

FCN – Fully Convolutional Network – повністю згорткова нейронна мережа;

FLOPS – Floating Point Operations Per Second – операції з рухомою комою за секунду;

FPN – Feature Pyramid Network – мережа піраміди ознак;

GAN – Generative Adversarial Network – генеративна змагальна мережа;

GPU – Graphics Processing Unit – графічний процесор;

IoU – Intersection over Union – перетин над об‘єднанням;

KNN – K Nearest Neighbours – K найближчих сусідів;

NIR – Near-Infrared Spectroscopy – ближня інфрачервона спектроскопія;

PCA – Prinipal Component Analysis – аналіз головних компонент;

PIL – Python Imaging Library – бібліотека обробки зображень для Python;

PPM – Pyramid Pooling Module – пірамідальний модуль об‘єднання;

ReLU – Rectified Linear Unit – випрямлена лінійна функція активація;

RGB – Red Green Blue – модель кольору, що використовує червоний, зелений і синій кольори для створення зображень;

SGD – Stohastic Gradient Descent – стохастичний градієнтний спуск;

SVM – Support Vector Machine – метод опорних векторів;

SWIR – Short-Wavelength Infrared – короткохвильове інфрачервоне світло.

ВСТУП

На сьогоднішній день, супутникові зображення являють собою важливий інструмент для моніторингу та оцінки великих територій у різних сферах діяльності. Супутникові зображення активно використовуються для аналізу стану навколишнього середовища, зокрема для оцінки змін у землекористуванні, моніторингу сільськогосподарських угідь, виявлення наслідків природних катастроф та антропогенного впливу. Однак ефективне застосування глибокого навчання для автоматизації цих процесів ускладнюється обмеженістю доступних розмічених даних, особливо в умовах багатокласової сегментації, коли необхідно класифікувати об’єкти на зображеннях на багато категорій.

Розробка масштабованої глибокої моделі, здатної ефективно працювати в умовах обмеженої вибірки даних, має важливе значення для подолання цих обмежень. Така модель дозволить значно покращити точність та ефективність аналізу супутникових зображень, що є особливо актуальним для регіонів із обмеженими ресурсами або специфічними умовами. Наприклад, у регіонах, що зазнають значних природних змін або урбанізаційного впливу, супутникові зображення використовуються для моніторингу стану лісових масивів, водних ресурсів та міських територій. Вони допомагають оцінювати масштаби змін, виявляти потенційні загрози довкіллю та приймати обґрунтовані рішення щодо екологічного планування та управління природними ресурсами.

Такі системи також можуть стати інструментом для планування відновлення територій, моніторингу екологічних наслідків та оптимізації використання земельних ресурсів.

Таким чином, дослідження проблеми аналізу супутникових зображень та розробка ефективних методів глибокого навчання для їхньої обробки є актуальним і практично значущим завданням. Створення масштабованих моделей, здатних працювати в умовах обмежених даних,

сприятиме покращенню точності моніторингу природних та антропогенних змін, що має критичне значення як для окремих регіонів, так і для глобальних екологічних та соціально-економічних процесів. Впровадження таких рішень дозволить не лише оптимізувати використання супутникових даних, а й забезпечити ефективне прийняття рішень у сферах екологічного управління, міського планування та відновлення територій після кризових ситуацій.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

1.1 Завдання сегментації супутникових зображень

1.1.1 Роль семантичної сегментації для дистанційного зондування

У галузі дистанційного зондування Землі семантична сегментація зображень відіграє визначальну роль, виступаючи фундаментальним інструментом для автоматизованого аналізу значних обсягів даних, отриманих з супутникових систем. На відміну від інших методів комп'ютерного зору, таких як класифікація зображень чи виявлення об'єктів, семантична сегментація забезпечує найвищий рівень деталізації та інформативності аналізу, що є критично важливим для широкого спектру практичних застосувань. Детальний аналіз значення семантичної сегментації для дистанційного зондування вимагає розгляду багатьох аспектів, пов'язаних з обробкою, аналізом та інтерпретацією супутникових знімків.

Насамперед, семантична сегментація дозволяє здійснювати повну пікселізовану ідентифікацію різних типів земного покриву та землекористування. На відміну від класифікації, де кожному зображенню або його фрагменту призначається єдина мітка, семантична сегментація передбачає присвоєння класу кожному пікселю зображення. Таким чином, результатом обробки є детальна карта, де кожен піксель має чітко визначену семантичну категорію. Це дозволяє отримувати високоточні карти земного покриву з чіткими межами різних класів об'єктів, що є надзвичайно важливим для задач моніторингу природних ресурсів, екологічного контролю та управління територіями.

Точність визначення меж різних об'єктів на земній поверхні є одним з ключових переваг семантичної сегментації. Алгоритми, засновані на

глибоких нейронних мережах, здатні виявляти складні переходи між різними типами покриву, навіть якщо ці переходи нечіткі або розмиті. Це особливо важливо при моніторингу урбанізованих територій, де щільна забудова та різноманітність об'єктів створюють складний ландшафт для аналізу. Точне визначення меж будівель, доріг, зелених насаджень та інших елементів міської інфраструктури дозволяє проводити детальний аналіз міської морфології та її динаміки [1] (рисунок 1.1).

Рисунок 1.1 – Приклад семантичної сегментації супутникового зображення

Масштабованість методів семантичної сегментації дозволяє застосовувати їх до зображень різної роздільної здатності та географічного охоплення. Це надає можливість аналізувати як локальні території, наприклад, окремі населені пункти, так і великі регіони, включаючи цілі країни або континенти. При цьому зберігається високий рівень деталізації та точності, що є ключовим фактором для багатьох прикладних задач. Особливо важливим це є для глобальних проектів моніторингу, які вимагають обробки та аналізу величезних масивів даних.

Мультитемпоральний аналіз, тобто вивчення змін на земній поверхні з плином часу, є ще одним важливим застосуванням семантичної сегментації в дистанційному зондуванні. Порівнюючи результати

сегментації знімків, отриманих у різні періоди, можна виявляти та кількісно оцінювати зміни в землекористуванні, відстежувати процеси урбанізації, дефорестації, опустелювання та інші динамічні явища. Це дозволяє не лише констатувати факт змін, але й прогнозувати їх майбутній розвиток, що є надзвичайно важливим для прийняття обґрунтованих рішень у сфері управління територіями та природними ресурсами.

У контексті моніторингу сільськогосподарських угідь семантична сегментація дозволяє точно визначати межі полів, ідентифікувати типи сільськогосподарських культур та оцінювати їх стан. Це забезпечує можливість прогнозування врожайності, виявлення зон стресу рослин, оптимізації використання добрив та пестицидів, а також планування зрошення. Такий підхід лежить в основі точного землеробства, яке спрямоване на максимізацію ефективності використання сільськогосподарських земель при мінімізації негативного впливу на навколишнє середовище.

Для екологічного моніторингу та охорони природи семантична сегментація надає інструменти для відстеження стану та динаміки природних екосистем, таких як ліси, водно-болотні угіддя, гірські системи та прибережні зони. Точне картування різних типів рослинності, водних об'єктів та геологічних формацій дозволяє виявляти деградацію природних екосистем, оцінювати вплив господарської діяльності та природних катаклізмів, планувати заходи з відновлення та охорони цінних природних територій. Особливо важливим є моніторинг лісових масивів для виявлення незаконних вирубок, пожеж та інших факторів деградації лісів.

Управління водними ресурсами також значно вдосконалюється завдяки застосуванню методів семантичної сегментації. Точне картування водних об'єктів, включаючи річки, озера, водосховища та затоплені території, дозволяє оцінювати запаси прісної води, планувати заходи з управління водними ресурсами, прогнозувати та мінімізувати наслідки паводків та засух. Крім того, аналіз супутникових знімків дозволяє виявляти

забруднення водних об'єктів, оцінювати їх екологічний стан та планувати заходи з очищення та відновлення.

У сфері управління ризиками та реагування на надзвичайні ситуації семантична сегментація супутникових знімків надає можливість швидко оцінювати масштаби та наслідки природних та техногенних катастроф, таких як повені, землетруси, лісові пожежі, розливи нафти та інші. Це дозволяє оперативно приймати рішення щодо евакуації населення, направлення рятувальних служб та організації гуманітарної допомоги. Крім того, аналіз знімків, отриманих до і після катастрофи, дозволяє оцінювати завдані збитки та планувати відновлювальні роботи.

Інтеграція результатів семантичної сегментації з ГІС дозволяє створювати комплексні просторові бази даних, які поєднують результати дистанційного зондування з іншими джерелами просторової інформації. Це забезпечує можливість проведення складних аналітичних досліджень, які враховують різноманітні просторові фактори та їх взаємозв'язки. Такий підхід є основою для створення систем підтримки прийняття рішень у сфері територіального планування, управління природними ресурсами та оцінки екологічних ризиків.

Розвиток методів глибокого навчання значно підвищив точність та ефективність семантичної сегментації супутникових знімків. Сучасні архітектури нейронних мереж, такі як U-Net, DeepLab, PSPNet та інші, демонструють вражаючі результати в задачах сегментації різних типів об'єктів на земній поверхні. Це відкриває нові можливості для автоматизації процесів аналізу та інтерпретації даних дистанційного зондування, що особливо важливо в умовах постійного зростання обсягів доступної супутникової інформації.

Однак, незважаючи на значний прогрес, залишаються серйозні виклики, пов'язані з обмеженістю навчальних даних, варіативністю умов зйомки та складністю деяких типів ландшафту. Ці проблеми особливо гостро проявляються при сегментації супутникових знімків високої

роздільної здатності, де деталізація зображень дозволяє виявляти дрібні об'єкти, але вимагає більш складних моделей та більшої кількості розмічених даних для навчання. Розробка методів, які ефективно працюють в умовах обмеженої вибірки даних, є одним з ключових напрямків досліджень у цій галузі [2].

Таким чином, семантична сегментація відіграє критично важливу роль у сучасному дистанційному зондуванні, забезпечуючи високоточний автоматизований аналіз супутникових знімків для широкого спектру прикладних задач. Постійне вдосконалення методів сегментації, розвиток нових архітектур нейронних мереж та збільшення обсягів доступних навчальних даних відкривають нові можливості для більш ефективного використання даних дистанційного зондування в науці, економіці та управлінні.

1.1.2 Виклики пов‘язані з сегментації супутникових зображень

Незважаючи на значущість сегментації супутникових зображень, ця сфера стикається з численними викликами, які суттєво ускладнюють ефективну реалізацію методів сегментації. Розглянемо детально основні проблеми, з якими стикаються дослідники та практики у цій галузі.

Проблема великих зображень є однією з найбільш значущих перешкод у сегментації супутникових даних. Супутникові знімки зазвичай мають надзвичайно високу роздільну здатність, що призводить до формування файлів величезного розміру, які можуть досягати десятків гігабайтів. Наприклад, зображення із супутників серії Sentinel-2 можуть мати розмір до 10980x10980 пікселів з просторовою роздільною здатністю до 10 метрів на піксель, що спричиняє значні обчислювальні навантаження [3]. Обробка таких масивних даних вимагає потужних обчислювальних ресурсів, які не завжди доступні для дослідницьких груп або комерційних компаній. Традиційні алгоритми машинного навчання та

глибокі нейронні мережі розроблені для роботи із зображеннями значно меншого розміру, тому їх безпосереднє застосування до супутникових знімків часто неможливе через обмеження пам'яті GPU або CPU.

Вирішення проблеми великих зображень вимагає складних підходів, таких як розбиття зображень на менші фрагменти (патчі) з подальшим об'єднанням результатів, що створює додаткові складнощі на етапі постобробки. При такому підході виникають артефакти на межах патчів, втрачається контекстна інформація між фрагментами, а також зростає складність алгоритмів обробки. Крім того, час, необхідний для обробки повного супутникового зображення, значно збільшується, що обмежує можливості аналізу даних у режимі реального часу або близького до нього [4].

Різнорідність даних представляє інший фундаментальний виклик у сегментації супутникових зображень. Супутникові дані надходять із різних джерел, отриманих за допомогою різних типів сенсорів і платформ, що працюють у різних діапазонах електромагнітного спектру. Наприклад, оптичні супутники (як Landsat чи Sentinel-2) фіксують видиме та інфрачервоне випромінювання, тоді як радарні системи (як Sentinel-1) використовують радіохвилі для створення зображень незалежно від хмарності та часу доби [5]. Ці різноманітні джерела даних мають відмінні характеристики шуму, просторову та спектральну роздільну здатність, що ускладнює їх інтеграцію та спільне використання.

Окрім відмінностей між типами сенсорів, існує також проблема різночасових даних. Супутникові знімки однієї й тієї ж території, отримані в різний час, можуть суттєво відрізнятися через сезонні зміни рослинності, погодні умови, зміни у землекористуванні або різні кути спостереження супутників. Ця темпоральна мінливість створює додаткові труднощі для алгоритмів сегментації, які мають бути достатньо надійними, щоб правильно класифікувати об'єкти, незважаючи на ці зміни.

Інтеграція різнорідних даних вимагає складних методів злиття на різних рівнях обробки – від попередньої обробки до моделювання та постобробки. Розроблення ефективних методів злиття мультимодальних даних залишається активною сферою досліджень, що вимагає глибокого розуміння як властивостей даних, так і специфіки задач сегментації.

Проблема обмеженої анотації є особливо гострою для супутникових зображень. Створення якісних навчальних наборів даних вимагає експертної розмітки зображень, що є надзвичайно трудомістким і дорогим процесом. Наприклад, для точної сегментації міських територій експерт має вручну розмітити будівлі, дороги, зелені зони та інші об'єкти на зображеннях. Для супутникових знімків високої роздільної здатності цей процес може займати багато годин для одного зображення. Більше того, для деяких специфічних задач, таких як моніторинг рідкісних стихійних лих (повеней, пожеж, зсувів) або специфічних типів землекористування, кількість доступних анотованих прикладів може бути надзвичайно малою.

Недостатня кількість анотованих даних призводить до обмежених можливостей для навчання глибоких нейронних мереж, які потребують великих обсягів даних для ефективної генералізації [6]. У контексті супутникових зображень ця проблема посилюється через велику варіативність об'єктів на знімках залежно від географічного регіону, сезону та інших факторів. Наприклад, модель, навчена на розпізнавання міських структур у європейських містах, може показувати незадовільні результати при застосуванні до міст Африки або Азії через відмінності в архітектурних стилях, плануванні міст та щільності забудови.

Для подолання обмежень анотації розробляються методи навчання з обмеженим наглядом, напівавтоматичного анотування, активного навчання та слабо-супервізованого навчання. Ці підходи спрямовані на максимальне використання доступних анотованих даних та залучення неанотованих даних для покращення моделей. Однак, незважаючи на прогрес у цих напрямках, питання ефективного навчання з обмеженими анотаціями

залишається однією з найскладніших проблем у сегментації супутникових зображень.

Радіометрична варіативність також представляє значний виклик для сегментації супутникових зображень. Супутникові знімки однієї території, отримані за різних умов освітлення, атмосферних умов або з різних сенсорів, можуть мати суттєво відмінні радіометричні характеристики. Ці відмінності виникають через різні кути сонця, сезонні зміни, атмосферне розсіювання, поглинання та відбиття, що впливає на інтенсивність та спектральні характеристики пікселів. Наприклад, наявність хмар, димки або туману може значно змінити яскравість та контраст зображення, ускладнюючи ідентифікацію та сегментацію об'єктів.

Ефекти атмосферної дисперсії та розсіювання також призводять до зниження контрасту та чіткості зображень, особливо для віддалених від надиру частин знімка. Ці ефекти можуть суттєво погіршити якість сегментації, особливо для об'єктів з подібними спектральними характеристиками, як, наприклад, різні типи сільськогосподарських культур або різні типи міської забудови. Для подолання радіометричної варіативності застосовуються методи атмосферної корекції та радіометричного калібрування, але ці процеси самі по собі складні та можуть вносити додаткові артефакти або шуми. Більш того, для багатьох архівних зображень або зображень з обмеженими метаданими, точна атмосферна корекція може бути неможливою.

Просторова гетерогенність об'єктів на супутникових зображеннях є ще одним важливим викликом. Об'єкти одного класу можуть мати різні розміри, форми та орієнтації залежно від географічного регіону та контексту. Наприклад, будівлі в історичних центрах міст мають інші характеристики, ніж у промислових зонах або сучасних житлових районах. Ця гетерогенність ускладнює створення універсальних моделей сегментації, які добре працюють у різних контекстах [7].

Більше того, деякі об'єкти, особливо лінійні структури, такі як дороги, річки або лінії електропередач, можуть бути дуже тонкими відносно загальної роздільної здатності зображення і тому складними для точної сегментації. Ці об'єкти часто фрагментуються у процесі сегментації, що вимагає складних методів постобробки для відновлення їхньої цілісності.

Проблема просторових відносин між об'єктами також суттєво впливає на якість сегментації. У багатьох випадках правильна класифікація об'єкта залежить не лише від його власних характеристик, але й від його оточення та контексту. Наприклад, невеликі водні об'єкти в міських парках можуть бути візуально подібними до тіней від високих будівель, і їх правильна класифікація вимагає аналізу навколишнього контексту.

Геометрична складність і зміни масштабу також представляють значний виклик. Супутникові зображення часто охоплюють великі території з різноманітними об'єктами різних масштабів. Наприклад, на одному зображенні можуть бути присутні як великі сільськогосподарські поля, так і дрібні деталі міської забудови. Стандартні алгоритми сегментації часто оптимізовані для роботи з об'єктами певного масштабу, що ускладнює одночасну ефективну сегментацію малих і великих об'єктів.

Цю проблему посилює різна просторова роздільна здатність різних спектральних каналів у багатьох супутникових системах. Наприклад, у системі Sentinel-2 видимі та ближні інфрачервоні канали мають роздільну здатність 10 метрів на піксель, тоді як деякі інфрачервоні канали – лише 20 або 60 метрів [3]. Ця неоднорідність роздільної здатності вимагає додаткової обробки для узгодження даних різних каналів.

Дисбаланс класів є ще одним фундаментальним викликом для систем сегментації супутникових зображень. У багатьох практичних застосуваннях існує значна нерівномірність у представленості різних класів об'єктів. Наприклад, при моніторингу міських територій площа доріг та будівель зазвичай значно перевищує площу водних об'єктів або специфічних типів інфраструктури. Ще більш екстремальні випадки дисбалансу

спостерігаються при завданнях виявлення рідкісних об'єктів або подій, таких як вирубки лісів, нелегальні звалища або пошкодження інфраструктури після стихійних лих.

Такий дисбаланс класів призводить до зміщення моделей у бік домінуючих класів, що значно знижує точність розпізнавання рідкісних, але часто критично важливих об'єктів. Традиційні метрики оцінки, такі як загальна точність, можуть приховувати погану продуктивність для рідкісних класів, створюючи ілюзію хорошої роботи моделі, яка насправді може повністю ігнорувати деякі важливі класи.

Для подолання проблеми дисбалансу класів застосовуються різноманітні підходи, включаючи зважування класів під час навчання, специфічні функції втрат (наприклад, фокальна втрата або зважена перехресна ентропія), техніки відбору зразків і аугментації даних. Однак ефективне вирішення цієї проблеми залишається складним, особливо в контексті взаємодії з іншими викликами, такими як обмежена анотація та висока варіативність об'єктів.

Часові зміни представляють один з найскладніших аспектів сегментації супутникових зображень, особливо в контексті моніторингу довгострокових процесів або виявлення змін. Земна поверхня постійно змінюється як через природні процеси (зміна сезонів, ріст рослинності, ерозія), так і через антропогенну діяльність (будівництво, сільське господарство, видобуток ресурсів). Ці зміни відбуваються з різною швидкістю та в різних масштабах, що створює складний динамічний контекст для систем сегментації.

Моделі, навчені на даних одного періоду, часто демонструють знижену ефективність при застосуванні до даних інших періодів через зміни в характеристиках об'єктів. Наприклад, листяні ліси виглядають абсолютно по-різному влітку та взимку, сільськогосподарські поля змінюють свої спектральні характеристики протягом вегетаційного періоду, а міські території еволюціонують з розвитком інфраструктури.

Часова узгодженість є особливо важливою для завдань моніторингу та виявлення змін, де потрібно відрізнити реальні зміни від артефактів, спричинених різними умовами зйомки або сезонною варіативністю. Розроблення моделей, стійких до часових змін або спеціально розроблених для виявлення та характеризації таких змін, залишається активною сферою досліджень із багатьма відкритими питаннями.

Обчислювальна складність є практичним викликом, який часто недооцінюється в академічних дослідженнях, але має критичне значення для реальних застосувань. Сучасні методи глибокого навчання для сегментації зображень, такі FCN, U-Net, DeepLab та їхні варіанти, вимагають значних обчислювальних ресурсів як для навчання, так і для виведення. У контексті надвеликих супутникових зображень ці вимоги стають ще більш суттєвими [8].

Навчання складних моделей сегментації на великих наборах супутникових даних може вимагати тижнів обчислень на потужних графічних процесорах або спеціалізованих кластерах. Навіть після навчання, виведення результатів для нових зображень може бути обчислювально складним процесом, особливо якщо потрібна обробка в реальному часі або близькому до нього режимі.

Додаткові обчислювальні витрати виникають через необхідність попередньої обробки (корекція атмосферних спотворень, нормалізація, узгодження даних з різних джерел) та постобробки (фільтрація шуму, згладжування меж, векторизація результатів). У контексті аналізу часових серій або виявлення змін ці обчислювальні витрати множаться на кількість часових точок, що робить завдання ще складнішим.

Розроблення обчислювально ефективних алгоритмів, оптимізація архітектур моделей та пошук балансу між точністю та швидкістю є важливими напрямками досліджень. Крім того, ефективне використання розподілених обчислень, хмарних сервісів та спеціалізованого апаратного

забезпечення стає критичним для практичного застосування методів сегментації супутникових зображень.

Узагальнення та перенесення моделей між різними географічними регіонами та джерелами даних є фундаментальним викликом для систем сегментації. Моделі, навчені на даних одного регіону, часто демонструють значне зниження продуктивності при застосуванні до інших регіонів через відмінності в ландшафтах, архітектурних стилях, типах рослинності та інших географічно специфічних особливостях. Наприклад, модель, оптимізована для розпізнавання міських структур в європейських містах, може давати незадовільні результати при застосуванні до міст Азії через різні патерни забудови та використання простору.

Аналогічно, моделі, навчені на даних одного супутникового сенсора, часто погано узагальнюються на дані з інших сенсорів через відмінності в спектральних характеристиках, роздільній здатності та геометричних особливостях. Ця проблема стає ще більш складною при спробах об'єднати дані різних типів (оптичні, радарні, лідарні) або різних часових періодів.

Складність об'єднання різних аспектів якості в єдину метрику та необхідність врахування специфічних вимог конкретних застосувань робить оцінку та порівняння різних методів сегментації супутникових зображень непростим завданням. Розроблення більш адекватних та інформативних метрик оцінки, які враховують специфіку супутникових даних та практичних застосувань, залишається важливим напрямком досліджень.

Інтерпретованість результатів сегментації є критично важливою для багатьох практичних застосувань, особливо тих, що пов'язані з прийняттям рішень у таких сферах, як управління надзвичайними ситуаціями, екологічний моніторинг або містобудування. Кінцеві користувачі систем сегментації супутникових зображень часто потребують не лише самих результатів, але й розуміння їхньої достовірності, обмежень та факторів, що впливають на якість сегментації.

Незважаючи на значний прогрес у останні роки, зокрема з розвитком глибокого навчання та обчислювальних технологій, багато з цих викликів залишаються актуальними та потребують подальших досліджень. Розроблення комплексних підходів, які враховують специфіку супутникових даних та практичних застосувань, є критично важливим для розширення можливостей та впливу технологій сегментації супутникових зображень на різноманітні сфери людської діяльності.

1.2 Огляд сучасних методів багатокласової сегментації

1.2.1 Традиційні методи багатокласової сегментації

Традиційні методи сегментації супутникових зображень тривалий час були основою аналізу даних дистанційного зондування Землі. Ці підходи, що сформувалися до ери глибокого навчання, продемонстрували свою ефективність і продовжують застосовуватися в багатьох практичних задачах. Дані методи характеризуються відносною простотою імплементації, інтерпретованістю результатів та меншими вимогами до обчислювальних ресурсів у порівнянні з сучасними нейромережевими підходами. Серед ключових традиційних методів можна навести ті, що засновані на класифікації пікселів та кластеризації, наприклад метод максимальної правдоподібності, метод мінімальної відстані, метод паралелепіпедів, k-найближчих сусідів та метод опорних векторів.

Піксельні методи класифікації розглядають кожен піксель зображення як незалежну одиницю аналізу. Такий підхід став одним із перших методів, використаних для сегментації супутникових зображень, та заснований на припущенні, що спектральні характеристики пікселя несуть достатню інформацію для його категоризації.

Метод максимальної правдоподібності є одним із найбільш поширених параметричних методів класифікації у дистанційному

зондуванні. Даний метод базується на припущенні, що спектральні значення для кожного класу розподілені за нормальним законом. Використовуючи навчальну вибірку, алгоритм оцінює параметри розподілу (середнє значення та коваріаційну матрицю) для кожного класу та обчислює ймовірність належності невідомого пікселя до кожного з класів. Піксель призначається до класу з найвищою ймовірністю. Перевагами методу є його статистична обґрунтованість та можливість отримання ймовірнісних оцінок належності пікселя до певного класу. Однак, метод чутливий до якості та репрезентативності навчальної вибірки, вимагає достатньої кількості зразків для кожного класу і може стикатися з проблемами при роботі з даними, що не відповідають нормальному розподілу.

Метод мінімальної відстані – простий та обчислювально ефективний підхід, який базується на вимірюванні евклідової відстані між спектральним вектором пікселя та центроїдами класів, визначеними на основі навчальної вибірки. Піксель відноситься до класу з найменшою відстанню. Даний метод є менш чутливим до розміру навчальної вибірки порівняно з методом максимальної правдоподібності, але не враховує дисперсію та коваріацію між спектральними каналами, що може призводити до зниження точності класифікації в умовах спектральної схожості різних класів.

Метод паралелепіпедів встановлює для кожного класу межі прийняття рішень у вигляді паралелепіпеда в n-вимірному просторі ознак (де n представляє кількість спектральних каналів). Межі визначаються мінімальними та максимальними значеннями для кожного спектрального каналу в навчальній вибірці, часто з доданим пороговим значенням або стандартним відхиленням. Піксель класифікується до певного класу, якщо його спектральний вектор потрапляє в межі відповідного паралелепіпеда. Перевагою методу є його простота та обчислювальна ефективність, проте він страждає від потенційного перекриття класів та нечутливості до структури коваріації даних. Часто цей

метод використовується як попередній етап класифікації з подальшим уточненням результатів іншими методами.

K-найближчих сусідів – непараметричний метод, який класифікує піксель на основі більшості класів серед K найближчих навчальних зразків у просторі ознак. Цей метод не потребує попередніх припущень щодо розподілу даних і легко адаптується до складних меж між класами. Для реалізації KNN у контексті сегментації супутникових зображень необхідно визначити відповідну метрику відстані (часто використовується евклідова відстань) та оптимальне значення K, яке може суттєво впливати на результати класифікації [9]. Більші значення K зменшують вплив шуму, але можуть згладжувати межі між класами. Перевагами KNN є його простота, відсутність фази навчання та здатність адаптуватися до локальних особливостей даних. Недоліками є обчислювальна складність при великих навчальних наборах та чутливість до проблеми прокляття розмірності при аналізі гіперспектральних зображень з великою кількістю каналів.

Метод опорних векторів (Support Vector Machine) виявився ефективним підходом для класифікації супутникових зображень, особливо в умовах обмеженої кількості навчальних даних. SVM знаходить оптимальну гіперплощину, яка максимізує відстань (margin) між класами у просторі ознак. Для роботи з нелінійно роздільними даними SVM використовує ядерні функції (kernel functions), які проєктують вхідні дані у простір вищої розмірності, де класи стають лінійно роздільними. У контексті сегментації супутникових зображень SVM демонструє високу точність, особливо при роботі з мультиспектральними та гіперспектральними даними. Найбільш поширеними ядерними функціями є радіальна базисна функція, поліноміальне ядро та сигмоїдне ядро [10]. Переваги SVM включають високу точність класифікації, стійкість до перенавчання та ефективність при роботі з обмеженою навчальною вибіркою. Недоліками є необхідність ретельного підбору

параметрів (параметри ядра, параметр регуляризації) та обчислювальна складність при великих наборах даних.

На відміну від методів класифікації, що вимагають наявності навчальної вибірки, методи кластеризації здійснюють сегментацію зображень у режимі навчання без учителя, базуючись лише на властивостях самих даних. Ці методи автоматично групують пікселі зображення у кластери на основі їх спектральної схожості. Кластеризація особливо цінна при відсутності або обмеженості апріорної інформації про досліджувану територію.

K-середніх – один із найпростіших та найпоширеніших алгоритмів кластеризації, який розділяє пікселі зображення на K кластерів. Алгоритм ітеративно оптимізує розташування центроїдів кластерів та призначення пікселів до кластерів, мінімізуючи сумарну внутрішньокластерну дисперсію [11]. Перевагами цього методу є простота, обчислювальна ефективність та інтуїтивна інтерпретація результатів. Недоліками є необхідність попереднього визначення кількості кластерів, чутливість до початкової ініціалізації центроїдів, тенденція до створення кластерів приблизно однакового розміру та обмеженість у роботі з кластерами складної форми. У контексті сегментації супутникових зображень K-середніх ефективно використовується для попередньої сегментації та створення первинної карти земного покриву, яка може бути в подальшому уточнена іншими методами.

Ієрархічна кластеризація створює дендрограму (деревоподібну структуру), яка представляє вкладеність кластерів на різних рівнях. Існують два основні підходи: агломеративний (знизу-вгору), який починає з розгляду кожного пікселя як окремого кластера і послідовно об'єднує найближчі кластери, та дивізимний (згори-вниз), який починає з одного кластера, що містить усі пікселі, і рекурсивно розділяє його. У контексті сегментації супутникових зображень ієрархічна кластеризація дозволяє аналізувати структуру даних на різних рівнях деталізації та виявляти

підкласи в межах більших кластерів [12]. Проте через високу обчислювальну складність цей метод рідко застосовується до повних супутникових сцен, а частіше використовується для аналізу попередньо відібраних зразків або регіонів інтересу.

Mean Shift – непараметричний алгоритм кластеризації, що визначає кластери шляхом ідентифікації локальних максимумів щільності розподілу даних. Процедура Mean Shift ітеративно зміщує кожну точку даних до області з найвищою щільністю точок, використовуючи ядерну функцію для оцінки щільності. В контексті сегментації супутникових зображень Mean Shift ефективно виявляє кластери різних розмірів та форм без необхідності попереднього визначення їх кількості. Алгоритм особливо корисний для виявлення об'єктів неправильної форми та природних границь на зображеннях [13]. Основним параметром алгоритму є ширина вікна пошуку (bandwidth), що визначає радіус, в якому враховуються сусідні точки при оцінці щільності. Великі значення bandwidth призводять до меншої кількості більших кластерів, тоді як малі значення створюють більшу кількість детальніших кластерів. Перевагами Mean Shift є автоматичне визначення кількості кластерів, здатність виявляти кластери довільної форми та стійкість до шумів. Недоліками є висока обчислювальна складність та чутливість до вибору параметра bandwidth.

1.2.2 Глибокі нейронні мережі в багатокласовій сегментації

За останнє десятиліття розвиток глибокого навчання призвів до значного прогресу в галузі семантичної сегментації, пропонуючи архітектури, які демонструють вражаючу точність та ефективність. У цьому підрозділі будуть детально розглянуті чотири провідні архітектури глибоких нейронних мереж для семантичної сегментації: FCN, U-Net, DeepLabv3+ та PSPNet. Ці моделі становлять основу сучасних підходів до

сегментації супутникових зображень і демонструють еволюцію методів глибокого навчання в цій сфері.

FCN стали революційним проривом у семантичній сегментації, коли були представлені у 2015 році Long et al. Головна ідея FCN полягає у трансформації класичних згорткових нейронних мереж у повністю згорткову архітектуру, здатну генерувати сегментаційні карти на рівні пікселів. Ключова інновація FCN – це заміна повнозв'язних шарів на згорткові шари з ядром розміром 1×1. Такий підхід дозволяє зберігати просторову інформацію протягом усієї мережі і працювати із зображеннями будь-якого розміру.

Типова архітектура FCN (рисунок 1.2) складається з двох основних компонентів. Енкодер – це послідовність згорткових та об'єднуючих шарів, які поступово зменшують просторову роздільну здатність і збільшують глибину ознак. Цей компонент зазвичай базується на архітектурах класифікаційних мереж, таких як VGG, ResNet або Inception, з видаленими повнозв'язними шарами. Декодер (висхідний шлях) включає транспоновані згортки (deconvolution або upsampling), які відновлюють просторову роздільну здатність для отримання сегментаційної карти того ж розміру, що й вхідне зображення.

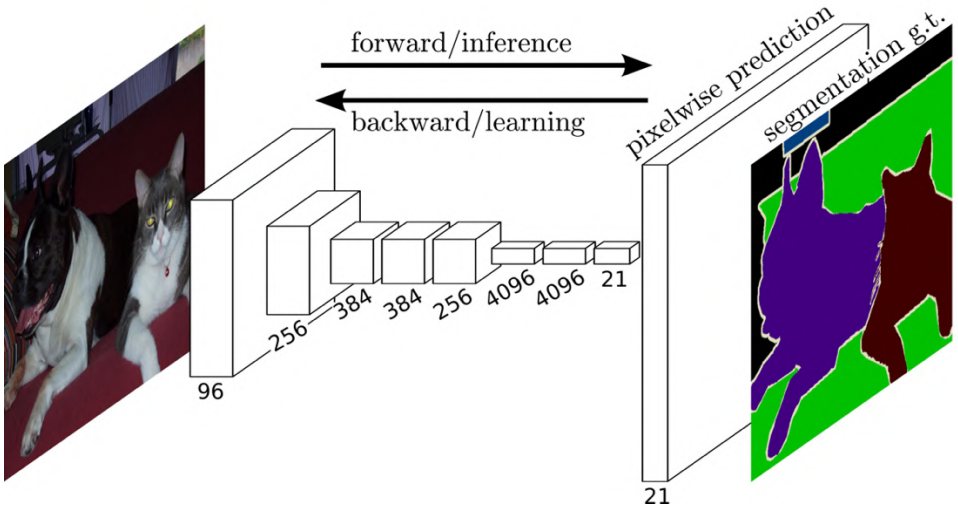

Рисунок 1.2 – Типова структура FCN

Одна з ключових особливостей FCN – використання skip connections (обхідних з'єднань), які об'єднують низькорівневі ознаки з високорівневими, що дозволяє зберегти детальну просторову інформацію. Автори оригінальної роботи запропонували три варіанти архітектури: FCN 32s, FCN-16s та FCN-8s, які відрізняються способом об'єднання обхідних з'єднань та відповідно точністю сегментації.

FCN продемонстрували високу ефективність для сегментації супутникових зображень з кількох причин. Здатність зберігати просторовий контекст надзвичайно важлива для аналізу супутникових зображень, де контекст часто визначає клас об'єкта. Масштабованість дозволяє FCN працювати із зображеннями високої роздільної здатності, що критично для супутникових знімків. Адаптивність архітектури дозволяє легко інтегрувати різні базові моделі для вилучення ознак [14]. Однак, FCN мають певні обмеження при роботі з супутниковими зображеннями. Послідовні згортки та операції pooling можуть призвести до втрати дрібних деталей, які часто критичні у супутникових знімках. Також FCN обмежені у здатності інтегрувати інформацію з різних масштабів, що важливо для розрізнення об'єктів різного розміру.

U-Net була розроблена Ronneberger et al. у 2015 році спеціально для сегментації біомедичних зображень, але швидко знайшла застосування в різних областях, включаючи сегментацію супутникових знімків. Архітектура отримала свою назву через U-подібну форму, яка візуально представляє її структуру. Головні концептуальні особливості U-Net включають симетричну структуру, де мережа складається з симетричних енкодера та декодера з однаковою кількістю рівнів. Розгалужені обхідні з'єднання, на відміну від FCN, U-Net використовує прямі конкатенації карт ознак з енкодера до відповідних рівнів декодера. Архітектура спроектована для одночасного збереження як локальних деталей, так і глобального контексту.

Архітектура U-Net складається з двох симетричних шляхів. Енкодер (шлях стиснення) слідує типовій архітектурі CNN для класифікації, складаючись з повторюваних блоків, кожен з яких містить дві послідовні згортки 3×3 з активацією ReLU та операцію max-pooling 2×2 з кроком 2 для зменшення розміру. Кожен крок об'єднання зменшує просторові розміри удвічі і подвоює кількість каналів ознак. Декодер (шлях розширення) складається з повторюваних блоків, кожен з яких містить транспоновану згортку 2×2, яка збільшує просторові розміри вдвічі, конкатенацію з відповідною картою ознак з енкодера, та дві послідовні згортки 3×3 з активацією ReLU. Фінальний шар – згортка 1×1, яка відображає вектор ознак кожного пікселя у вихідні класи (рисунок 1.3).

Ключовим елементом архітектури U-Net є обхідні з'єднання (skip connections), які з'єднують відповідні рівні енкодера та декодера. Ці з'єднання дозволяють декодеру використовувати просторову інформацію, збережену на ранніх етапах енкодера, що вирішує проблему втрати просторових деталей під час стиснення. Математично операція конкатенації можна представити як об'єднання тензорів ознак, отриманих на різних рівнях мережі [15].

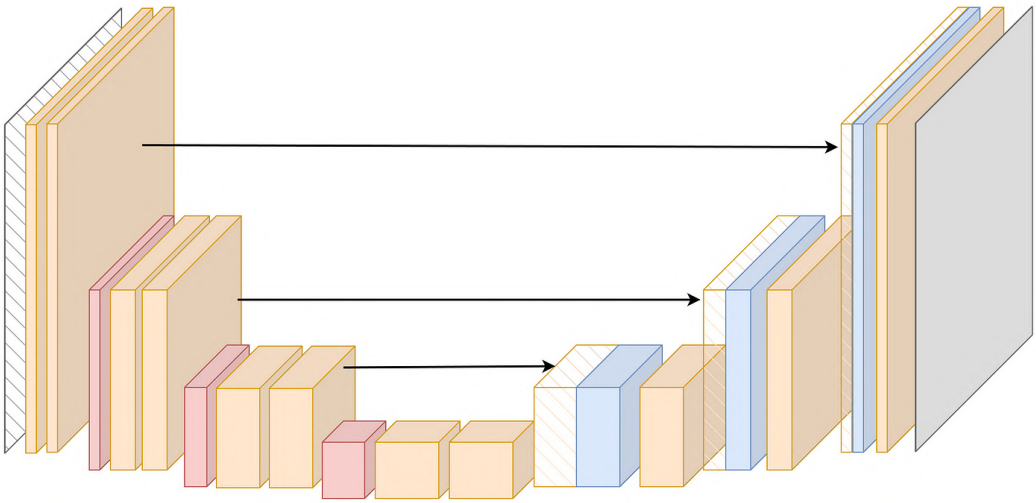

Рисунок 1.3 – Типова структура U-Net

U-Net особливо ефективна для сегментації супутникових зображень завдяки кільком перевагам. Збереження дрібних деталей через обхідні з'єднання дозволяє точно сегментувати малі об'єкти, такі як будівлі, дороги та окремі дерева. Ефективність при обмежених даних робить U-Net ідеальною для задач дистанційного зондування, де часто доступні лише обмежені набори розмічених даних. Симетрична структура забезпечує оптимальний баланс між точністю та обчислювальною ефективністю. Проте, U-Net має певні обмеження при роботі з супутниковими зображеннями. Фіксоване рецептивне поле обмежує здатність мережі враховувати широкий контекст, що може бути проблемою для великомасштабних об'єктів. Чутливість до шуму та артефактів може впливати на якість сегментації, особливо при роботі з радарними або низькоякісними оптичними знімками.

DeepLabv3+ є передовою архітектурою для семантичної сегментації, розробленою Chen et al. у 2018 році, яка поєднує переваги попередніх версій DeepLab з додатковими інноваціями. DeepLabv3+ представляє собою енкодер-декодерну структуру, подібну до U-Net, але з важливими вдосконаленнями. Архітектура DeepLabv3+ базується на двох ключових інноваціях: використанні згорток з отворами (atrous/dilated convolutions) та просторовій пірамідальній об'єднуючій архітектурі ASPP.

Згортки з отворами (dilated convolutions) дозволяють розширити рецептивне поле без збільшення кількості параметрів або зменшення просторової роздільної здатності. Математично dilated convolution з коефіцієнтом розширення r можна описати наступним чином. У стандартній 2D згортці ядро застосовується до суміжних пікселів вхідної карти ознак. У згортці з отворами вводиться проміжок r-1 між послідовними точками вибірки в ядрі. Це дозволяє захоплювати більш широкий контекст без збільшення кількості параметрів [16].

Модуль ASPP застосовує паралельні згортки з отворами з різними коефіцієнтами розширення, що дозволяє ефективно агрегувати

багатомасштабну контекстну інформацію. Модуль ASPP складається з однієї згортки 1×1 та кількох згорток 3×3 з різними коефіцієнтами розширення, наприклад, r = {6, 12, 18}. Всі паралельні гілки обробляють одну й ту ж вхідну карту ознак, і їхні виходи потім конкатенуються. Додатково, ASPP включає глобальний контекстний модуль, який виконує глобальне усереднене об'єднання (global average pooling), згортку 1×1, а потім двовимірне білінійне інтерполювання для відновлення просторових розмірів.

Архітектура DeepLabv3+ складається з енкодера, розширеного модуля ASPP та декодера. Енкодер може використовувати різні базові мережі, такі як ResNet, Xception або MobileNet, модифіковані для включення згорток з отворами на пізніх стадіях. Це дозволяє отримувати карти ознак з високим рецептивним полем без надмірного зменшення просторової роздільної здатності. Декодер поступово відновлює деталі об'єктів. Спочатку виконується білінійне інтерполювання виходу енкодера/ASPP до 1/4 розміру вхідного зображення. Потім ці ознаки конкатенуються з відповідними низькорівневими ознаками з ранніх шарів енкодера (після 1×1 згортки для зменшення кількості каналів). Після цього застосовуються кілька згорток 3×3, за якими слідує ще одне білінійне інтерполювання для отримання сегментаційної карти в оригінальному розмірі зображення.

DeepLabv3+ показує виняткову ефективність у сегментації супутникових зображень з кількох причин. Покращене багатомасштабне моделювання дозволяє ефективно сегментувати об'єкти різних розмірів, від великих (поля, ліси) до дрібних (будівлі, транспортні засоби). Збереження деталей завдяки комбінації згорток з отворами та енкодер-декодерної архітектури дозволяє точно визначати межі об'єктів. Модульність і гнучкість архітектури дає можливість використовувати різні базові мережі залежно від вимог до швидкості та точності. Проте, DeepLabv3+ має і певні обмеження для супутникових застосувань. Обчислювальна складність може бути проблемою для обробки великих супутникових знімків у режимі

реального часу. Чутливість до якості вхідних даних вимагає ретельної попередньої обробки, особливо для супутникових знімків з різними атмосферними умовами або кутами зйомки.

PSPNet розроблена Zhao et al. у 2017 році, є ще однією впливовою архітектурою для семантичної сегментації, яка фокусується на інтеграції глобального контексту для покращення точності. PSPNet вирішує проблему обмеженого рецептивного поля та недостатнього врахування глобального контексту, які властиві багатьом архітектурам для сегментації. Концептуальна основа PSPNet полягає в ієрархічному аналізі глобального контексту через PPM, який агрегує інформацію з різних регіонів і масштабів.

Архітектура PSPNet складається з трьох основних компонентів. Базова мережа для вилучення ознак – як правило, ResNet або подібні глибокі CNN з модифікованим вихідним кроком (output stride), щоб забезпечити більшу просторову роздільну здатність карт ознак. PPM використовує різні масштаби усередненого об'єднання, щоб отримати ознаки з різних субрегіонів зображення. Типово PPM виконує операції об'єднання в чотирьох різних масштабах: 1×1 (глобальне об'єднання), 2×2, 3×3 та 6×6. Виходи кожного рівня проходять через згортку 1×1 для зменшення розмірності, потім інтерполюються до оригінального розміру карти ознак і конкатенуються. Фінальний сегментаційний модуль об'єднує оригінальні ознаки з виходом PPM, обробляє їх за допомогою згортки та формує вихідну сегментаційну карту [17].

Ключовий внесок PSPNet – PPM, який забезпечує збір контекстної інформації на різних масштабах. PSPNet пропонує кілька переваг для сегментації супутникових зображень. Покращене розуміння глобального контексту дозволяє правильно класифікувати ділянки, які можуть бути неоднозначними при розгляді лише локальної інформації. Збереження просторових деталей шляхом використання модифікованих базових мереж з меншим вихідним кроком забезпечує високу роздільну здатність

сегментаційних карт. Стійкість до варіацій масштабу допомагає точно сегментувати об'єкти різних розмірів, що особливо важливо для супутникових знімків, де об'єкти інтересу можуть сильно відрізнятися за розміром. Однак, PSPNet має і деякі обмеження для супутникових застосувань. Висока обчислювальна складність потребує потужного апаратного забезпечення, особливо для обробки супутникових знімків з високою просторовою роздільною здатністю. Необхідність ретельного налаштування PPM для конкретних типів супутникових даних може ускладнювати процес розробки ефективних моделей.

1.3 Критерії оцінки методів багатокласової сегментації

При створенні моделей семантичної сегментації також необхідно детально розглянути ключові аспекти, що дозволяють об'єктивно порівнювати та оцінювати ефективність різних алгоритмів. Основними критеріями оцінки є точність, ефективність при малих вибірках та обчислювальна складність.

Точність є фундаментальним критерієм оцінки якості сегментації зображень. У контексті сегментації супутникових знімків вона відображає ступінь відповідності між отриманими результатами та еталонними даними. Оцінка точності може здійснюватися за допомогою різних метрик, кожна з яких висвітлює певні аспекти продуктивності моделі. Однією з найпоширеніших метрик є IoU, також відома як коефіцієнт Жаккара. Ця метрика обчислюється як відношення площі перетину передбаченої та еталонної сегментації до площі їх об'єднання. IoU надає інтуїтивно зрозумілу оцінку якості сегментації, оскільки її значення знаходиться в діапазоні від 0 до 1, де 1 означає ідеальний збіг передбаченої та еталонної масок. У контексті супутникових зображень IoU є особливо важливою метрикою, оскільки вона дозволяє об'єктивно оцінити ступінь накладання

сегментованих об'єктів, таких як будівлі, дороги чи водойми, з їх справжніми межами.

Dice coefficient, або F1-score, є ще однією важливою метрикою для оцінки точності сегментації. Він обчислюється як подвоєне відношення площі перетину до суми площ передбаченої та еталонної сегментації. Dice coefficient також варіюється від 0 до 1, де вищі значення вказують на кращу якість сегментації. Порівняно з IoU, Dice coefficient зазвичай надає дещо вищі значення, особливо для дрібних об'єктів, що робить його корисним при роботі з супутниковими зображеннями, де часто потрібно виділити невеликі об'єкти на фоні великих територій. Ця метрика є особливо чутливою до правильного виявлення позитивних випадків, що робить її цінною для задач, де важливо мінімізувати пропуск об'єктів інтересу.

Accuracy (точність класифікації) є більш загальною метрикою, яка обчислюється як відношення кількості правильно класифікованих пікселів до загальної кількості пікселів у зображенні. Хоча ця метрика є інтуїтивно зрозумілою, вона може бути менш інформативною для задач сегментації супутникових зображень через незбалансованість класів. Наприклад, при сегментації рідкісних об'єктів на великій території модель може досягти високої accuracy, просто класифікуючи всі пікселі як фон. Тому для більш повної оцінки часто використовують додаткові метрики, такі як precision (точність), recall (повнота) та F1-score.

Ефективність при малих вибірках є особливо важливим критерієм для задач сегментації супутникових зображень, оскільки створення анотованих наборів даних для таких зображень часто вимагає значних ресурсів і експертних знань. Моделі, які здатні ефективно навчатися на обмеженій кількості прикладів, мають значну практичну перевагу. Для оцінки ефективності при малих вибірках можна використовувати криві навчання, які відображають залежність між розміром навчальної вибірки та продуктивністю моделі. Крім того, важливо враховувати здатність моделі до генералізації, тобто перенесення знань, отриманих на одному наборі

даних, на інші, подібні набори. У контексті супутникових зображень це може означати здатність моделі, навченої на знімках одного регіону, коректно сегментувати об'єкти на знімках інших регіонів з різними географічними та кліматичними умовами.

Для підвищення ефективності при малих вибірках часто застосовують методи передачі навчання (transfer learning) та аугментації даних. Передача навчання дозволяє використовувати знання, отримані моделлю при навчанні на великих наборах даних, для вирішення нових задач з обмеженою кількістю прикладів. Аугментація даних, у свою чергу, дозволяє штучно розширити навчальну вибірку за рахунок застосування різних трансформацій до наявних зображень, таких як поворот, масштабування, зміна контрасту тощо.

Обчислювальна складність є критичним фактором при виборі методів сегментації супутникових зображень, особливо для практичних застосувань. Вона включає часову складність (час, необхідний для навчання моделі та отримання результатів сегментації) та просторову складність (обсяг пам'яті, необхідний для зберігання моделі та проміжних результатів). Супутникові зображення часто мають високу роздільну здатність та охоплюють великі території, що може призводити до значних обчислювальних витрат при їх обробці. Тому важливо оцінювати ефективність алгоритмів не лише з точки зору точності, але й з урахуванням їх обчислювальних вимог.

Для оцінки обчислювальної складності можна використовувати такі метрики, як FLOPS, час виконання на стандартному апаратному забезпеченні, обсяг використовуваної пам'яті та енергоспоживання. Ці метрики дозволяють об'єктивно порівнювати різні алгоритми та обирати оптимальні рішення для конкретних задач та апаратних обмежень.

У практичних застосуваннях часто доводиться шукати компроміс між точністю сегментації та обчислювальною ефективністю. Наприклад, складні глибокі нейронні мережі, такі як U-Net або DeepLab, можуть

забезпечувати високу точність сегментації, але вимагають значних обчислювальних ресурсів. З іншого боку, більш легкі моделі, такі як EfficientNet або MobileNetV3, можуть бути більш придатними для застосувань з обмеженими ресурсами, хоча й можуть поступатися у точності.

Комплексна оцінка методів сегментації супутникових зображень повинна враховувати всі три розглянуті критерії: точність, ефективність при малих вибірках та обчислювальну складність. Залежно від конкретного завдання та доступних ресурсів, різним критеріям може надаватися різна вага. Наприклад, для критичних застосувань, таких як моніторинг стихійних лих, пріоритетом може бути точність, тоді як для систем реального часу з обмеженими ресурсами важливішою може бути обчислювальна ефективність.

1.4 Постановка задачі дослідження

Для ефективного вирішення цієї задачі необхідно обрати підхід, який враховує специфіку супутникових даних та забезпечує високу точність сегментації при обмежених наборах навчальних даних.

З огляду на характер задачі сегментації супутникових зображень та обмеженість доступних розмічених даних, постає необхідність використання механізму трансферного навчання. Трансферне навчання є критично важливим підходом при роботі з обмеженими наборами даних у сфері глибинного навчання, особливо для задач сегментації супутникових зображень, де збір та розмітка даних є ресурсовитратним процесом. Оптимальне рішення повинно включати можливість використання попередньо навчених моделей, які були треновані на великих наборах даних, таких як ImageNet. Такий підхід дозволяє використовувати вже сформовані фільтри та ваги для розпізнавання базових елементів зображень, які є універсальними для багатьох типів візуальних даних, включаючи

супутникові знімки. При цьому модель отримує можливість розпізнавати низькорівневі та середньорівневі ознаки, такі як краї, текстури та прості геометричні форми без потреби навчатися «з нуля», що значно прискорює процес збіжності оптимізаційного алгоритму та знижує ризик перенавчання на малих наборах даних.

Важливою вимогою до архітектурного рішення є також наявність механізмів, які дозволяють ефективно зберігати та використовувати просторову інформацію. При обробці супутникових зображень критично важливо не втрачати деталі та контекст під час згортання та розгортання даних у мережі. Для цього необхідно забезпечити наявність обхідних з'єднань (skip-зв'язків) між відповідними шарами згортки та розгортки. Такі зв'язки дозволяють передавати детальну просторову інформацію безпосередньо з ранніх шарів мережі до пізніх шарів, минаючи середні шари, що особливо важливо для збереження чітких меж об'єктів, дрібних елементів ландшафту та текстурних особливостей поверхні на супутникових знімках.

Використання skip-зв'язків також суттєво покращує градієнтний потік у глибоких нейронних мережах, які часто стикаються з проблемою затухання градієнтів під час зворотного поширення помилки. Створюючи додаткові шляхи для потоку градієнтів, такі зв'язки сприяють більш ефективному навчанню мережі та дозволяють будувати глибші архітектури без втрати якості. Поєднання семантичної інформації з глибоких шарів та просторової інформації з ранніх шарів завдяки skip-зв'язкам підвищує точність локалізації об'єктів, що є одним із ключових критеріїв якості роботи алгоритмів сегментації.

Ще одним необхідним компонентом ефективного рішення для сегментації супутникових зображень є механізм багатомасштабного виділення ознак. Супутникові знімки характеризуються значною варіативністю об'єктів за розміром – від великих географічних об'єктів, таких як водойми та лісові масиви, до дрібних антропогенних структур,

наприклад, окремих будівель чи доріг. Тому архітектурне рішення має забезпечувати одночасну обробку вхідних даних за допомогою згорткових фільтрів різних розмірів, що дозволяє ефективно виділяти ознаки на різних рівнях деталізації.

Для підвищення стійкості моделі до проблеми затухання градієнтів доцільно використовувати залишкові з'єднання (residual connections), які дозволяють будувати більш глибокі мережі без втрати якості навчання. Такі з'єднання, в поєднанні з механізмами багатомасштабного виділення ознак, забезпечують формування ієрархічного представлення ознак, де кожен рівень мережі відповідає за виділення ознак різного рівня абстракції – від низькорівневих ознак до високорівневих семантичних концепцій.

Особливу увагу слід приділити можливості часткового заморожування шарів попередньо навченої моделі, що дозволить зберегти низькорівневі ознаки, важливі для розпізнавання базових елементів зображень, і водночас адаптувати модель до специфіки супутникових даних. Такий підхід дозволяє знайти оптимальний баланс між використанням попередньо вивчених загальних ознак та навчанням специфічних для конкретної задачі характеристик.

З огляду на вищезазначені вимоги, оптимальне архітектурне рішення для сегментації супутникових зображень має поєднувати енкодерно-декодерну структуру з попередньо навченим енкодером, який забезпечує ефективне виділення ознак на різних масштабах, та skip-зв'язки між енкодером і декодером для збереження просторової інформації. Така архітектура дозволить досягти високої точності сегментації навіть на складних наборах даних з обмеженою кількістю розмічених зображень, що є важливим фактором для практичного застосування в галузі дистанційного зондування Землі. Впровадження такого рішення створить потужний інструмент для автоматизованого аналізу супутникових даних та вирішення широкого спектру прикладних задач, від моніторингу стану навколишнього середовища до міського планування та картографування.

2 МЕТОДОЛОГІЯ СЕГМЕНТАЦІЇ СУПУТНИКОВИХ ЗОБРАЖЕНЬ

2.1 Архітектура моделі: U-Net з InceptionResNetV2

Архітектура моделі U-Net з InceptionResNetV2 для задачі семантичної сегментації супутникових зображень представляє собою потужну конструкцію, що складається з двох основних компонентів: енкодера та декодера, з'єднаних між собою пропускними зв'язками. Ця архітектура була обрана для дослідження завдяки її доведеній ефективності у вирішенні задач сегментації зображень у різних доменах, зокрема у галузі дистанційного зондування Землі.

2.1.1 Структура енкодера та декодера

Енкодер моделі базується на архітектурі InceptionResNetV2, яка є гібридом двох популярних підходів до побудови згорткових нейронних мереж: концепції Inception-модулів та залишкових з'єднань. Така комбінація дозволяє ефективно екстрагувати ієрархічні ознаки з вхідних супутникових зображень, зберігаючи при цьому компактність моделі та знижуючи ризик деградації під час навчання глибоких шарів мережі. Архітектура моделі наведена на рисунку 2.1.

InceptionResNetV2 як основа енкодера починається зі вхідного блоку, який приймає супутникове зображення та пропускає його через набір згорткових шарів. Вхідний блок складається з початкового згорткового шару з фільтром розміром 3×3 та кроком 2, за яким слідує шар нормалізації даних (batch normalization) та функція активації ReLU. Далі слідує ще один згортковий шар з аналогічними параметрами та ще один шар нормалізації з активацією. Завершується вхідний блок шаром максимального об'єднання (max pooling) з розміром вікна 3×3 та кроком 2, який зменшує просторову розмірність карт ознак.

Після вхідного блоку енкодер InceptionResNetV2 складається з п'яти послідовних секцій, кожна з яких містить набір Inception-ResNet блоків. Перша секція починається з шару згортки 3×3 з кроком 2, за яким слідує блок нормалізації та активації. Далі йдуть декілька Inception-ResNet блоків типу A, кожен з яких має структуру з паралельних гілок згорток різних розмірів (1×1, 3×3) та залишкових з'єднань, що додають вхід блоку до його виходу після відповідних перетворень.

Друга секція енкодера починається з редукційного блоку, який зменшує просторову розмірність карт ознак, але збільшує їх кількість. За ним слідує серія Inception-ResNet блоків типу B, які мають більш складну структуру з додатковими гілками згорток та більшою кількістю фільтрів. Ці блоки обробляють дані на середньому рівні абстракції, виявляючи більш складні патерни на зображенні. Третя, четверта та п'ята секції енкодера продовжують обробку даних, поступово збільшуючи рівень абстракції та зменшуючи просторову розмірність. У фінальній частині енкодера дані проходять через глобальний шар усереднення (global average pooling), який зводить просторові виміри до єдиного вектора ознак високого рівня [18].

На виході енкодера формується набір карт ознак різного рівня, які потім передаються до декодера через прямі та пропускні з'єднання. Важливо відзначити, що на кожному етапі стиснення в енкодері зберігаються проміжні карти ознак для подальшої передачі у відповідні шари декодера.

Завдяки такій архітектурі енкодер InceptionResNetV2 забезпечує ефективне вилучення як низькорівневих, так і високорівневих ознак, що є критично важливим для завдань семантичної сегментації та розпізнавання складних об’єктів. Поєднання залишкових з’єднань із багатогілковими структурами згорток дозволяє мережі глибше аналізувати вхідне зображення без суттєвих втрат градієнта під час навчання, покращуючи як швидкість збіжності, так і точність моделі.

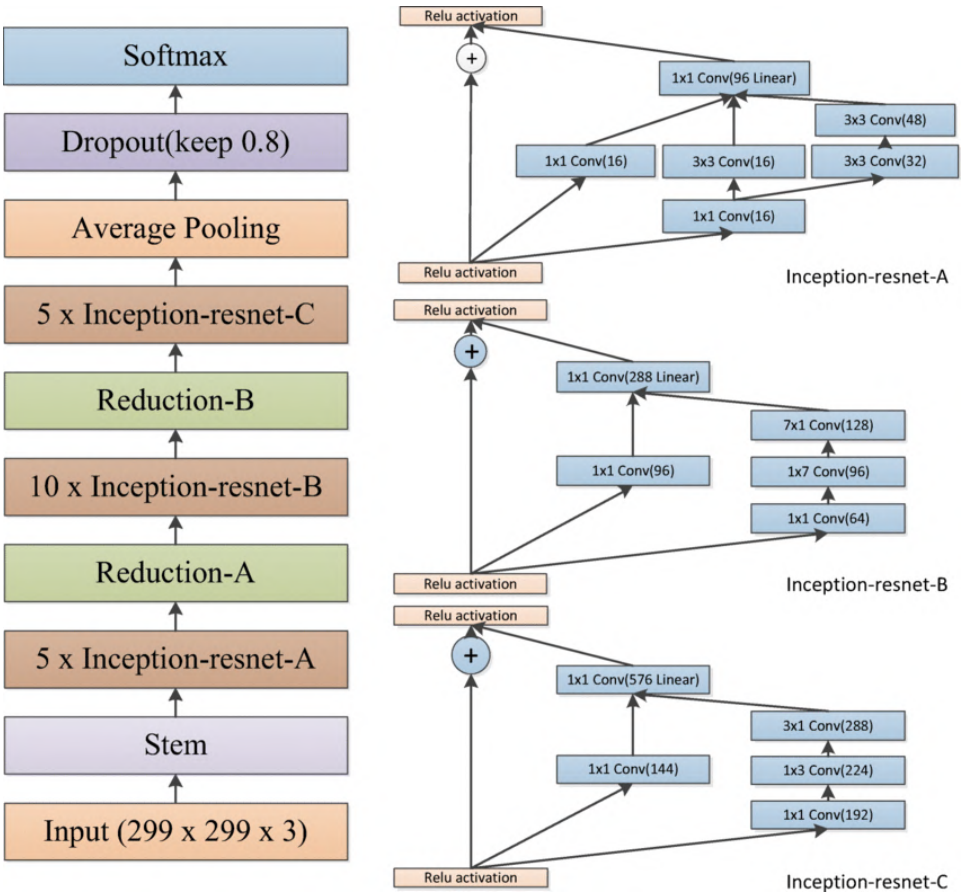

Рисунок 2.1 – Архітектура InceptionResNetV2

Декодер U-Net архітектури розроблений для поступового відновлення просторового розміру сегментаційної маски з набору стиснутих карт ознак, отриманих від енкодера. Декодер складається з послідовності блоків розширення, кожен з яких виконує збільшення просторового розміру та зменшення кількості каналів ознак.

Кожен блок декодера починається з операції підвищення дискретизації (upsampling), яка подвоює просторовий розмір карт ознак. Це може бути реалізовано за допомогою транспонованої згортки (transposed convolution) або методу інтерполяції (наприклад, білінійної) з подальшою згорткою 2×2. Після операції підвищення дискретизації відбувається конкатенація отриманих карт ознак з відповідними картами з енкодера, що передаються через пропускні з'єднання. Цей механізм є критично важливим

для архітектури U-Net, оскільки він дозволяє зберегти детальну просторову інформацію, яка могла бути втрачена під час процесу стиснення в енкодері.

Після конкатенації карти ознак обробляються двома послідовними згортковими шарами 3×3 з відповідними шарами нормалізації та активації ReLU. Ці згортки відповідають за інтеграцію локальної та глобальної інформації для формування точної сегментаційної маски. Кількість фільтрів у кожному наступному блоці декодера зменшується вдвічі, що відповідає симетричній структурі U-Net архітектури.

У типовій конфігурації U-Net з InceptionResNetV2 декодер має п'ять блоків розширення, що відповідають п'яти секціям енкодера. На кожному етапі декодера просторова розмірність карт ознак збільшується вдвічі, а кількість каналів зменшується. Фінальний блок декодера містить згортку 1×1, яка перетворює карти ознак у сегментаційну маску з кількістю каналів, що дорівнює кількості класів у задачі семантичної сегментації супутникових зображень [19]. До цього шару застосовується функція активації softmax, яка нормалізує вихідні значення таким чином, щоб вони представляли ймовірності належності кожного пікселя до відповідного класу.

Архітектурні особливості U-Net з InceptionResNetV2 роблять цю модель особливо придатною для семантичної сегментації супутникових зображень, де необхідно зберігати балавс між глобальним контекстом та локальними деталями. Енкодер InceptionResNetV2 ефективно знижує розмірність вхідних даних та виявляє важливі високорівневі ознаки, а декодер, посилений пропускними з'єднаннями, забезпечує точну локалізацію об'єктів на зображенні.

Слід також зазначити особливості реалізації пропускних з'єднань між енкодером та декодером. У класичній архітектурі U-Net пропускні з'єднання просто передають карти ознак з енкодера до відповідного рівня декодера без додаткової обробки. Однак, у модифікованих версіях, включаючи U-Net з InceptionResNetV2, ці з'єднання можуть включати додаткові згорткові

шари 1×1 для узгодження розмірності каналів або шари просторової уваги (spatial attention), які дозволяють моделі фокусуватися на найбільш інформативних регіонах карт ознак.

Загалом, архітектура U-Net з InceptionResNetV2 представляє собою потужний інструмент для точної та ефективної семантичної сегментації супутникових зображень. Глибинна структура енкодера дозволяє виявляти складні абстрактні патерни на зображеннях, а декодер з пропускними з'єднаннями забезпечує точну локалізацію та сегментацію різних класів об'єктів на супутникових знімках.

2.1.2 Виділення ознак на різних масштабах

Виділення ознак на різних масштабах є критично важливим аспектом архітектури U-Net з InceptionResNetV2 для задачі семантичної сегментації супутникових зображень. Саме здатність ефективно виявляти та інтегрувати ознаки різного рівня абстракції та просторового розміру дозволяє досягти високої точності сегментації навіть для складних сцен на супутникових знімках з неоднорідними об'єктами.

Супутникові зображення характеризуються значною варіативністю масштабів об'єктів інтересу. Наприклад, на одному знімку можуть бути присутні як великі природні об'єкти (водойми, ліси, поля), так і дрібні антропогенні структури (окремі будівлі, дороги, транспортні засоби). Крім того, навіть об'єкти одного класу можуть значно відрізнятися за розміром – міські квартали різної щільності забудови, водойми різного розміру, сільськогосподарські угіддя різної площі. Це створює унікальні виклики для алгоритмів семантичної сегментації, які повинні однаково ефективно працювати з об'єктами різного масштабу.

У контексті архітектури U-Net з InceptionResNetV2 механізм виділення ознак на різних масштабах реалізується через ряд взаємопов'язаних архітектурних рішень. Першим ключовим компонентом є

сама структура Inception-модулів, що лежить в основі енкодера InceptionResNetV2. Inception-модулі за своєю природою розроблені для одночасного аналізу вхідних даних на різних масштабах завдяки використанню паралельних згорткових шарів з різними розмірами ядер. Типовий Inception-модуль включає кілька паралельних гілок: згортки 1×1 для точкового аналізу, згортки 3×3 для аналізу локального контексту та згортки 5×5 (або дві послідовні 3×3) для охоплення більшого рецептивного поля [20].

У випадку InceptionResNetV2, що використовується як енкодер у нашій моделі, ця концепція розвивається ще далі. Тут застосовуються більш складні Inception-ResNet блоки, які поєднують паралельні згортки різних розмірів з залишковими з'єднаннями. Залишкові з'єднання додають вхідні дані блоку до його виходу після перетворення, що сприяє більш ефективному потоку градієнтів під час навчання та дозволяє будувати надглибокі архітектури без деградації точності. Такі блоки організовані в кілька послідовних секцій (зазвичай три: Inception-ResNet-A, Inception-ResNet-B, Inception-ResNet-C), кожна з яких має свою специфічну архітектуру з різними комбінаціями згорток та різною кількістю фільтрів.

Проходження супутникового зображення через енкодер InceptionResNetV2 забезпечує поступове зменшення просторової розмірності карт ознак та збільшення їх семантичної глибини. На початкових шарах мережа виявляє прості низькорівневі ознаки, такі як краї, контури та текстури. Ці ознаки мають високу просторову роздільність і відображають детальну локальну структуру зображення. З просуванням вглиб мережі, через послідовні Inception-ResNet блоки та редукційні шари, карти ознак стають все більш абстрактними та семантично насиченими, але втрачають просторову деталізацію.

Такий ієрархічний підхід до вилучення ознак дозволяє моделі формувати багаторівневе розуміння сцени. Наприклад, на нижчих рівнях мережа може виявляти базові візуальні елементи, такі як контури та

текстури різних ландшафтів, на середніх рівнях – більш складні структури, такі як частини будівель або межі полів, а на вищих рівнях – цілісне представлення великих об'єктів, таких як міські квартали або водні системи.

Однак сама по собі ієрархічна структура енкодера не вирішила б проблему точної локалізації об'єктів, оскільки карти ознак на глибоких рівнях мережі, хоч і семантично багаті, але мають низьку просторову роздільність. Саме тут на допомогу приходить архітектура U-Net з її унікальною структурою декодера та пропускними з'єднаннями.

Декодер U-Net, як було згадано раніше, поступово відновлює просторову розмірність карт ознак, використовуючи операції підвищення дискретизації (upsampling). На кожному рівні декодера відбувається конкатенація (об'єднання) поточних карт ознак із відповідними картами з енкодера, які передаються через пропускні з'єднання. Це дозволяє комбінувати високорівневу семантичну інформацію з глибоких шарів мережі з детальною просторовою інформацією з ранніх шарів.

Така інтеграція ознак різних масштабів у декодері є другим ключовим компонентом механізму мультимасштабного аналізу в моделі U-Net з InceptionResNetV2. Пропускні з'єднання не просто допомагають розповсюдженню градієнтів під час навчання, а й забезпечують доступ декодера до повного спектру ознак різного рівня абстракції – від низькорівневих деталей до високорівневих семантичних концепцій.

Важливим аспектом цих пропускних з'єднань є те, що вони можуть бути реалізовані з додатковою обробкою для більш ефективної інтеграції ознак різних масштабів. Наприклад, можуть застосовуватися проміжні згорткові шари 1×1 для узгодження розмірності каналів або механізми уваги для виділення найбільш інформативних регіонів карт ознак.

Для подальшого вдосконалення виділення ознак на різних масштабах у сучасних реалізаціях U-Net з InceptionResNetV2 часто застосовуються додаткові техніки. Однією з таких технік є використання атрузівних (розширених) згорток (atrous or dilated convolutions). Атрузівні

згортки дозволяють розширити рецептивне поле нейронів без збільшення кількості параметрів або зменшення просторової розмірності. Це досягається шляхом введення «дірок» або проміжків між елементами фільтра згортки. Коефіцієнт розширення визначає відстань між сусідніми елементами фільтра. Наприклад, атрузівна згортка 3×3 з коефіцієнтом розширення 2 матиме ефективне рецептивне поле 5×5, але використовуватиме лише 9 параметрів замість 25.

Важливим розширенням концепції атрузівних згорток є ASPP. Ця техніка застосовує кілька паралельних атрузівних згорток з різними коефіцієнтами розширення до одних і тих самих вхідних даних, а потім об'єднує отримані карти ознак. Це дозволяє ефективно аналізувати контекст на різних масштабах без втрати просторової роздільності. У контексті U-Net з InceptionResNetV2 модуль ASPP може бути інтегрований у найглибшу частину мережі між енкодером та декодером для покращення виділення ознак на різних масштабах [16].

Ще одним ефективним підходом до мультимасштабного аналізу є використання пірамідальних структур, таких як FPN. В таких структурах карти ознак різних рівнів енкодера не просто передаються в декодер, а спочатку збагачуються інформацією з більш глибоких рівнів шляхом операцій підвищення дискретизації та поелементного додавання. Це створює ієрархію багатомасштабних карт ознак, кожна з яких містить як детальну просторову інформацію, так і високорівневий семантичний контекст.

Механізми уваги також відіграють важливу роль у вдосконаленні виділення ознак на різних масштабах. Модулі самоуваги (self-attention) або не-локальні блоки (non-local blocks) дозволяють моделі зосередитися на найбільш інформативних регіонах карт ознак та встановити довгі залежності між просторово віддаленими елементами зображення. Це особливо корисно для аналізу супутникових знімків, де об'єкти одного класу можуть бути розподілені по всьому зображенню.

На практиці комбінація всіх цих технік дозволяє моделі U-Net з InceptionResNetV2 ефективно виділяти та інтегрувати ознаки на різних масштабах, що критично важливо для точної семантичної сегментації супутникових зображень. Така здатність робить модель стійкою до варіацій розміру, орієнтації та контексту об'єктів, що особливо важливо для аналізу супутникових даних, де умови зйомки, масштаб та характеристики об'єктів можуть суттєво відрізнятися від знімка до знімка.

Крім того, для покращення якості виділення ознак на різних масштабах можуть застосовуватися специфічні стратегії навчання, такі як мультимасштабне навчання (multi-scale training), коли на різних ітераціях навчання модель отримує зображення різного розміру, або каскадне навчання (curriculum learning), коли модель спочатку навчається на простіших прикладах з поступовим переходом до більш складних випадків.

У контексті семантичної сегментації супутникових зображень ефективне виділення ознак на різних масштабах дозволяє моделі U-Net з InceptionResNetV2 точно сегментувати як великі природні об'єкти (ліси, водойми, гірські масиви), так і дрібні антропогенні структури (будівлі, дороги, автомобілі). Це робить таку архітектуру особливо придатною для аналізу різноманітних сцен у задачах дистанційного зондування Землі.

2.2 Передобробка та аугментація даних

Процес передобробки та аугментації даних при семантичній сегментації супутникових зображень є фундаментальним етапом, що визначає успішність подальшого навчання нейронної мережі.

Типовий набір даних для семантичної сегментації супутникових зображень складається з двох основних компонентів: вихідних супутникових знімків та відповідних масок сегментації. Супутникові знімки характеризуються специфічними особливостями, що вирізняють їх серед інших типів зображень. По-перше, вони зазвичай мають мультиспектральну

природу, що означає наявність інформації в декількох діапазонах електромагнітного спектру. Типовий супутниковий знімок може містити від 3–4 до десятків спектральних каналів: RGB у видимому спектрі, NIR, SWIR, термальні інфрачервоні канали та інші. Кожен із цих каналів несе унікальну інформацію про об'єкти на земній поверхні – наприклад, NIR особливо чутливий до рослинності, що дозволяє ефективно визначати її стан та тип [21].

Просторова роздільна здатність супутникових знімків може значно варіюватися залежно від сенсора: від десятків сантиметрів на піксель (для комерційних супутників високої роздільної здатності, таких як WorldView або Pleiades) до десятків метрів (для вільно доступних даних, таких як Landsat або Sentinel). При роботі з даними різної роздільної здатності необхідно враховувати, що деталізація об'єктів буде суттєво відрізнятися, і це потрібно брати до уваги при проектуванні архітектури нейронної мережі.

Важливою характеристикою супутникових даних є їхня часова складова – знімки однієї території можуть бути зроблені в різні сезони, часи доби або за різних погодних умов, що призводить до значних варіацій у зовнішньому вигляді одних і тих самих об'єктів. Це створює додаткові складнощі для моделі, яка повинна навчитися розпізнавати об'єкти незалежно від цих варіацій.

Нормалізація супутникових зображень є необхідною для забезпечення стабільності та ефективності навчання нейронних мереж. Існує кілька підходів до нормалізації, і вибір конкретного методу залежить від характеристик даних та архітектури мережі. Найпростіший метод – це лінійна нормалізація до діапазону [0, 1], що досягається діленням значень пікселів на максимально можливе значення для даного формату (наприклад, 255 для 8-бітних зображень або 65535 для 16-бітних). Однак цей підхід не враховує реальний розподіл значень, який може бути зміщеним або мати викиди.

Більш ефективним є підхід стандартизації, при якому від кожного значення віднімається середнє і результат ділиться на стандартне відхилення. Це приводить розподіл даних до стандартного нормального розподілу з нульовим середнім та одиничною дисперсією. Причому для супутникових зображень критично важливо обчислювати ці параметри окремо для кожного спектрального каналу, оскільки вони можуть мати суттєво різні діапазони та розподіли значень.

Іншою важливою частиною передобробки є сегментація зображень на патчі (фрагменти) оптимального розміру для подачі в нейронну мережу. Супутникові знімки часто мають дуже великі розміри (тисячі пікселів у ширину та висоту), що робить неможливим їх безпосереднє використання через обмеження пам'яті GPU. Тому зображення розбивають на менші патчі, зазвичай розміром від 256×256 до 1024×1024 пікселів, залежно від доступних обчислювальних ресурсів. При цьому важливо враховувати граничні ефекти – об'єкти на межі патчів можуть бути розділені, що ускладнює їх коректну сегментацію. Для вирішення цієї проблеми часто використовують підходи з перекриттям патчів (overlapping patches), коли сусідні патчі частково перекриваються, а потім результати сегментації зшиваються з урахуванням цього перекриття.

Аугментація даних є критично важливим етапом для підвищення розмірності та різноманітності навчального набору даних. Для супутникових зображень існують специфічні техніки аугментації, що враховують їхню природу та особливості задач сегментації.

Геометричні перетворення (наведені на рисунку 2.2) є найбільш поширеними методами аугментації для супутникових зображень. Вони включають обертання зображення на довільний кут або на кути, кратні 90 градусам (що відповідає різним орієнтаціям супутника при зйомці), горизонтальне та вертикальне відображення, масштабування (zoom-in та zoom-out), випадкові обрізки, зсуви та деформації [22]. Ці перетворення повинні застосовуватися ідентично як до вхідного зображення, так і до

маски сегментації, щоб зберегти точну відповідність між ними. Обертання є особливо ефективним для супутникових зображень, оскільки орієнтація об'єктів на земній поверхні може бути будь-якою, і модель має бути інваріантною до цього.

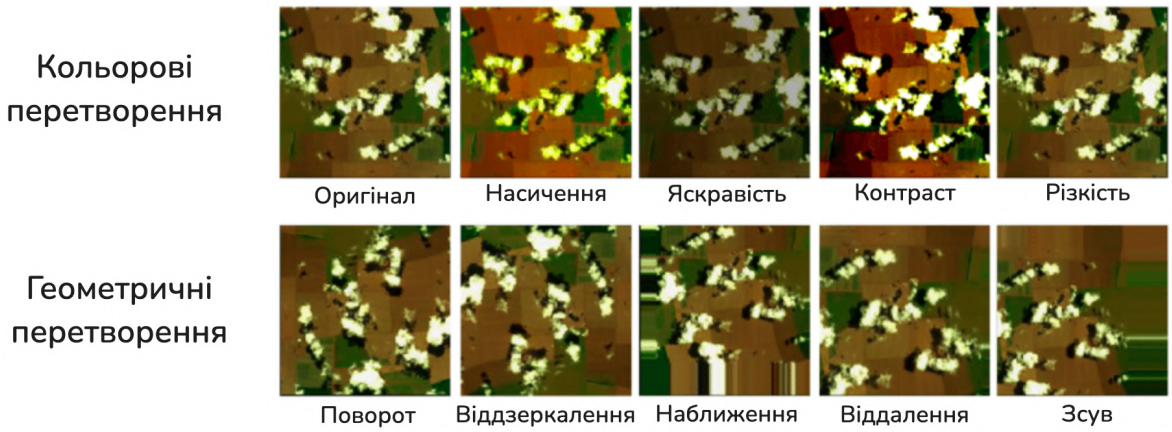

Рисунок 2.2 – Візуалізація геометричних перетворень зображень

Кольорові та спектральні трансформації включають зміну яскравості, контрасту, насиченості, випадкові зсуви кольорового балансу. Однак для мультиспектральних зображень ці перетворення повинні застосовуватися з обережністю, щоб не порушити фізичний зміст даних у спеціалізованих каналах. Наприклад, перетворення, що змінюють співвідношення між NIR та червоним каналами, можуть негативно вплинути на здатність моделі розпізнавати типи рослинності. Тому часто застосовуються більш спеціалізовані спектральні трансформації, такі як PCA з подальшим додаванням шуму до головних компонент, або незалежне масштабування значень у різних спектральних каналах.

Технікою аугментації, специфічною для задач сегментації, є міксап (mixup) та розрізання (cutmix), коли два зображення з відповідними масками сегментації комбінуються за певними правилами [23]. Візуалізація двох вищезгаданих методів аугментації наведена на рисунку 2.3. Ці техніки

дозволяють моделі краще навчитися розпізнавати об'єкти в різних контекстах та підвищує її узагальнюючу здатність.

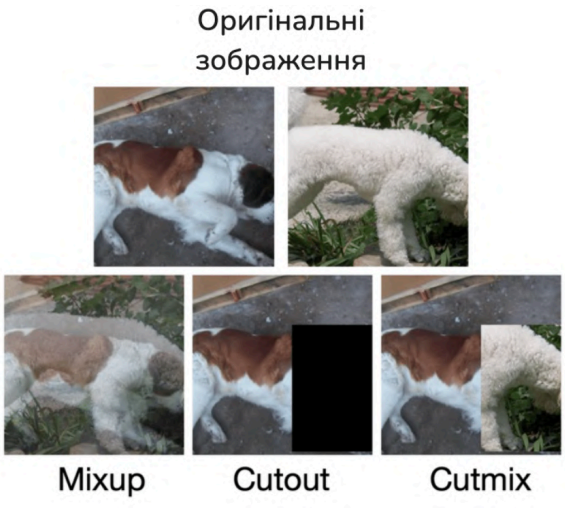

Рисунок 2.3 – Візуалізація технік аугментації mixup та cutmix

Ще одним важливим аспектом аугментації є симуляція різної просторової роздільної здатності. Шляхом зменшення роздільної здатності вхідного зображення з подальшим відновленням його розміру можна імітувати дані від різних сенсорів, що підвищує рівень узагальнення моделі при роботі з даними різної якості.

Додавання шуму є ефективним методом аугментації, особливо для супутникових зображень, де сигнал може бути спотворений різними факторами. Гаусівський шум, сіль-і-перець шум, шум Пуассона (характерний для сенсорів) додаються до зображень для імітації різних типів сенсорних артефактів. Рівень шуму може варіюватися для різних спектральних каналів, відображаючи різну чутливість сенсорів у різних діапазонах.

Генерація синтетичних даних за допомогою генеративних моделей, таких як GAN або дифузійні моделі, стає все більш популярним методом аугментації. Ці моделі можуть генерувати реалістичні супутникові зображення разом із відповідними масками сегментації, збагачуючи

навчальний набір даних. Особливо цінними є умовні генеративні моделі, які можуть створювати зображення з певними заданими характеристиками, наприклад, з певним типом ландшафту або міської забудови, що дозволяє цілеспрямовано розширювати набір даних у напрямках, де він має недостатню репрезентативність [24].

При роботі з супутниковими зображеннями важливо також враховувати проблему незбалансованості класів. У типовому наборі даних для семантичної сегментації супутникових зображень розподіл класів часто буває вкрай нерівномірним. Деякі класи, такі як «рослинність» або «голий ґрунт», можуть займати більшу частину площі зображення, тоді як інші класи, такі як «автомобілі» або «будівлі певного типу», можуть бути представлені лише на невеликій частині пікселів. Це призводить до того, що модель може ігнорувати рідкісні класи, оскільки їхній внесок у загальну функцію втрат є незначним.

Для боротьби з цією проблемою під час передобробки даних можна застосовувати різні стратегії: збалансування за допомогою випадкової вибірки (oversampling рідкісних класів та undersampling поширених); стратифікована вибірка патчів, щоб забезпечити більш рівномірне представлення різних класів; застосування зважених функцій втрат, де вага кожного класу обернено пропорційна його частоті в наборі даних; фокусування на граничних областях між різними класами, які часто містять більше інформації для навчання моделі.

В цілому, процес передобробки та аугментації даних для семантичної сегментації супутникових зображень є комплексним і багатогранним. Він вимагає глибокого розуміння природи супутникових даних, специфіки задачі сегментації та особливостей об'єктів, що підлягають розпізнаванню. Правильно організований процес передобробки та аугментації даних є запорукою успішного навчання нейронної мережі та отримання високоякісних результатів сегментації.

2.3 Стратегія навчання

При роботі над завданнями семантичної сегментації супутникових знімків важливо розробити комплексну стратегію навчання глибинних нейронних мереж, що забезпечить високу точність результатів. Трансферне навчання становить собою фундаментальний підхід до вирішення таких завдань, особливо коли кількість розмічених даних обмежена, що часто трапляється у задачах дистанційного зондування Землі. Сутність трансферного навчання полягає у використанні знань, отриманих моделлю при навчанні на одному наборі даних, для прискорення та покращення процесу навчання на іншому, цільовому наборі [25].

Реалізація трансферного навчання в контексті семантичної сегментації супутникових зображень може бути здійснена кількома способами. Найпоширенішим є використання попередньо навчених на великих набрах даних енкодерів, які вже мають сформовані ефективні фільтри для виділення низькорівневих і високорівневих ознак на зображеннях. При цьому важливо враховувати специфіку доменів: хоча фільтри, навчені на природних зображеннях, можуть ефективно виділяти базові елементи, такі як краї та текстури, супутникові зображення мають свої особливості – специфічний масштаб об'єктів, відмінну текстуру поверхонь, особливі спектральні характеристики.

За умови застосування трансферного навчання доцільно спочатку зафіксувати ваги попередньо навченого енкодера, зосередивши навчання на декодері, який відповідає за відновлення просторової інформації та формування сегментаційної маски. Після досягнення певного рівня збіжності можна перейти до етапу тонкого налаштування, коли розморожуються верхні шари енкодера. Поступове розморожування дозволяє моделі адаптувати високорівневі абстракції до специфіки супутникових даних без втрати цінної інформації, закодованої в нижніх шарах мережі.

Для поліпшення цього процесу рекомендується застосовувати диференційовані швидкості навчання для різних частин мережі. Нижні шари, що містять більш універсальні фільтри, навчаються з меншою швидкістю, тоді як верхні шари та декодер, які потребують більшої адаптації, навчаються з вищою швидкістю. Такий підхід забезпечує збереження корисних низькорівневих ознак при одночасній адаптації високорівневих абстракцій до нового домену [26].

Налаштування гіперпараметрів становить критичний етап у процесі навчання глибинних моделей для семантичної сегментації. Коректний вибір гіперпараметрів може значно підвищити точність моделі та скоротити час навчання. При роботі з супутниковими зображеннями особливу увагу слід приділяти таким параметрам як швидкість навчання, розмір міні-батчу, параметри регуляризації та планування швидкості навчання.

Швидкість навчання є одним з найважливіших гіперпараметрів. Високі значення можуть призвести до нестабільності процесу навчання, тоді як надто низькі значення зумовлюють повільну збіжність. Для задач семантичної сегментації супутникових зображень оптимальним зазвичай є початкове значення в діапазоні від одної десятитисячної до однієї тисячної, з подальшим поступовим зменшенням під час навчання.

Розмір міні-батчу також суттєво впливає на процес навчання. При роботі з супутниковими зображеннями високої роздільної здатності вибір розміру міні-батчу обмежується доступними обчислювальними ресурсами. Зазвичай використовуються значення від восьми до шістнадцяти зображень у батчі. При цьому варто враховувати, що менші розміри батчу призводять до більш шумних градієнтів, але можуть сприяти кращій генералізації моделі.

Значна увага має приділятися планувальникам швидкості навчання. Найпростішим є покрокове зменшення швидкості навчання, коли значення множиться на коефіцієнт (зазвичай 0,1) після певної кількості епох або при досягненні плато у значенні функції втрат. Більш складні стратегії

включають косинусне згасання швидкості навчання або циклічну зміну швидкості. Останній підхід передбачає періодичну зміну швидкості навчання між визначеними мінімальним і максимальним значеннями, що дозволяє моделі долати локальні мінімуми функції втрат [27].

Регуляризація є важливим інструментом запобігання перенавчанню моделі. Для задач семантичної сегментації супутникових зображень ефективною є комбінація кількох методів регуляризації. L2-регуляризація, що додає до функції втрат штраф за великі значення вагів, використовується з коефіцієнтом від однієї стотисячної до однієї десятитисячної. Техніка відсіювання (dropout) застосовується переважно в декодері мережі з імовірністю від 0,3 до 0,5.

Концепція ранньої зупинки є ефективним механізмом запобігання перенавчанню моделі при роботі з обмеженими наборами даних, що часто зустрічається при аналізі супутникових знімків. Сутність цього підходу полягає у моніторингу продуктивності моделі на валідаційному наборі та припиненні навчання, коли не спостерігається поліпшення протягом визначеної кількості епох.

У контексті семантичної сегментації супутникових зображень доцільно використовувати як метрику для ранньої зупинки не лише функцію втрат, але й більш інтерпретовані показники якості сегментації, такі як коефіцієнт Дайса або метрика перетину над об'єднанням. Це зумовлено тим, що мінімізація функції втрат не завжди безпосередньо корелює з покращенням якості сегментації, особливо за наявності незбалансованих класів.

Окрім збереження моделі з найкращим значенням метрики, корисною практикою є також зберігання контрольних точок моделі через регулярні інтервали епох. Це забезпечує можливість повернення до попередніх станів моделі та аналізу динаміки навчання. Для обмеження обсягу використовуваного дискового простору зазвичай зберігається обмежена кількість останніх контрольних точок.

Вибір відповідної функції втрат є критичним аспектом у розробці ефективних моделей семантичної сегментації супутникових зображень. Враховуючи особливості таких зображень, зокрема значний дисбаланс класів та варіативність об'єктів інтересу, використання класичних функцій втрат може бути недостатньо ефективним.

Категоріальна крос-ентропія є базовою функцією втрат для задач мультикласової сегментації. Вона вимірює розбіжність між прогнозованим розподілом ймовірностей класів та фактичним розподілом для кожного пікселя зображення. Однак у випадку супутникових зображень, де певні класи (наприклад, будівлі або дороги) можуть займати значно меншу площу порівняно з фоновими класами (такими як рослинність або вода), використання немодифікованої крос-ентропії призводить до зміщення моделі в бік домінуючих класів.

Для подолання цієї проблеми застосовується зважена крос-ентропія, де кожному класу призначається вага, обернено пропорційна його представленості у навчальному наборі. Такий підхід забезпечує збалансований внесок різних класів у функцію втрат. Ваги можуть бути статичними, визначеними на основі апріорного знання про розподіл класів, або динамічними, що коригуються в процесі навчання.

Іншим підходом до вирішення проблеми дисбалансу класів є використання функції втрат на основі коефіцієнта Дайса, яка безпосередньо оптимізує метрику перетину над об'єднанням. Функція втрат Дайса менш чутлива до дисбалансу класів, оскільки вона зосереджена на площі перетину прогнозованої та фактичної масок, нормалізованій їх сумою. Однак ця функція може бути нестабільною на початкових етапах навчання, коли прогнози моделі далекі від істинних значень.

Вибір оптимізатора має суттєвий вплив на швидкість та якість навчання моделі семантичної сегментації. Традиційний стохастичний градієнтний спуск із моментом залишається конкурентоспроможним вибором для багатьох задач комп'ютерного зору. При використанні SGD

рекомендується встановлювати початкову швидкість навчання в діапазоні від однієї сотої до однієї десятої, значення моменту 0,9 та застосовувати алгоритм Нестерова для прискорення збіжності.

Проте для складних архітектур, що застосовуються у семантичній сегментації супутникових зображень, адаптивні методи оптимізації часто демонструють кращі результати. Оптимізатор Adam, що поєднує ідеї моменту та адаптивних швидкостей навчання для різних параметрів, є найпопулярнішим вибором. Він забезпечує швидку збіжність навіть у складних ландшафтах функції втрат. Рекомендовані параметри включають швидкість навчання від однієї десятитисячної до однієї тисячної та стандартні значення експоненціальних швидкостей затухання для оцінок першого та другого моментів.

Модифікація Adam, відома як AdamW, забезпечує більш ефективну реалізацію L2-регуляризації, що робить її привабливим вибором для моделей, схильних до перенавчання. Інші варіації Adam, такі як Radam та Lookahead, також можуть покращити процес навчання за певних умов.

Важливим аспектом вибору оптимізатора є його взаємодія з планувальником швидкості навчання. Для оптимізатора Adam та його модифікацій ефективним є покрокове зменшення швидкості навчання або косинусне згасання. Для SGD хороші результати показує комбінація з циклічною швидкістю навчання, що дозволяє подолати локальні мінімуми функції втрат [28].

3 ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ ТА ОЦІНКА СИСТЕМИ

3.1 Огляд використаного набору даних

3.1.1 Аналіз набору даних та розбивка на підвиборки

Набір даних становить собою колекцію супутникових знімків міста Дубай, створену та анотовану в рамках спільного проекту організації Humans in the Loop та космічного центру імені Мохаммеда Рашида в Дубаї, ОАЕ [29]. Цей набір даних є відкритим ресурсом для дослідників та спеціалістів у галузі комп'ютерного зору та дистанційного зондування Землі.

Основний вміст набору даних складається з високоякісних аерокосмічних знімків Дубая, отриманих супутниками космічного центру імені Мохаммеда Рашида, що супроводжуються детальною семантичною сегментацією на рівні пікселів. Знімки були зроблені при різних погодних умовах (сніг, дощ, відсутність опадів), що додає варіативності набору даних. Анотації включають шість семантичних класів, що охоплюють основні типи міського та природного ландшафту. Семантичні класи представлені наступними категоріями:

- будівля;
- земля;
- дорога;
- рослинність;
- водний об'єкт;
- некласифікований об‘єкт.

Для створення набору даних використовувався багатоетапний процес обробки та сегментації. Початкові супутникові знімки були представлені у вигляді 8 великих тайлів роздільною здатністю 3264x3264 пікселів, де один піксель покриває 0.75 метри. Тайли були розділені на менші фрагменти

розміром 256×256 пікселів, із застосуванням методики перекриття на краях зображень для забезпечення цілісності інформації на межах. Така техніка дозволяє зменшити крайові ефекти при сегментації та підвищити точність розпізнавання об'єктів, що перетинають межі патчів. У результаті процесу розбиття було отримано загалом 1305 патчів розміром 256×256 пікселів. Приклад отриманих патчів наведений на рисунку 3.1.

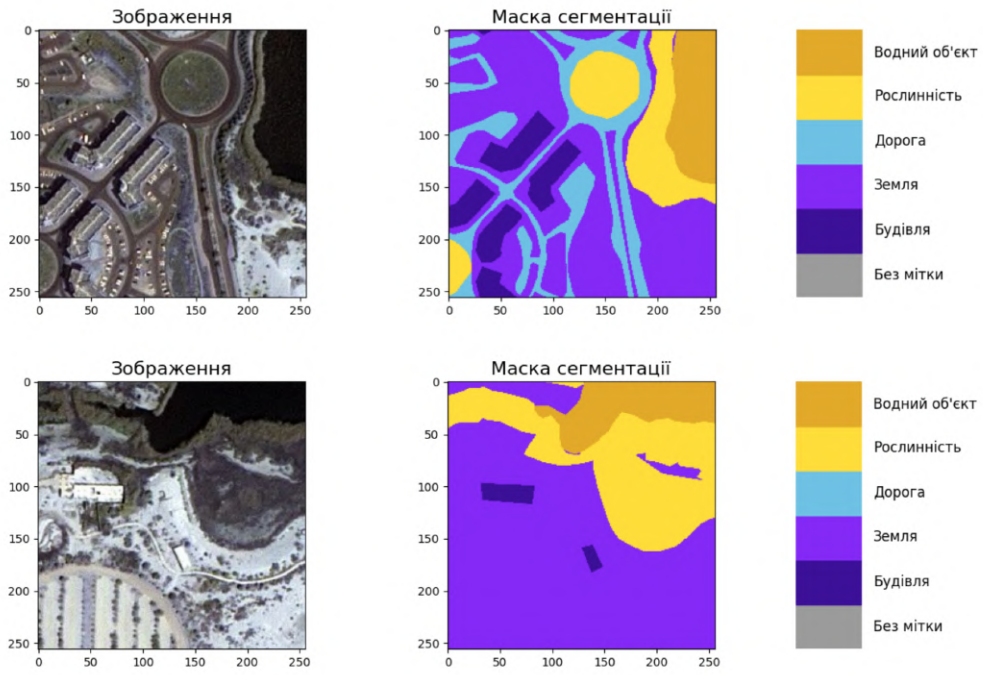

Рисунок 3.1 – Приклад отриманих патчів супутникових зображень

Для забезпечення можливості навчання та об'єктивної оцінки ефективності алгоритмів семантичної сегментації, набір даних був розділений на тренувальну та тестувальну вибірки у співвідношенні 80% до 20% відповідно. Така пропорція є стандартною практикою у галузі машинного навчання, що забезпечує достатній обсяг даних для навчання моделі з одночасним збереженням репрезентативної вибірки для тестування. Як видно на рисунку 3.2, тренувальна та тестувальні підвиборки є незбалансованими за розподілом пікселів по класам, що має бути

врахованим під час навчання моделі, проте цей розподіл є майже ідентичним між підвиборками.

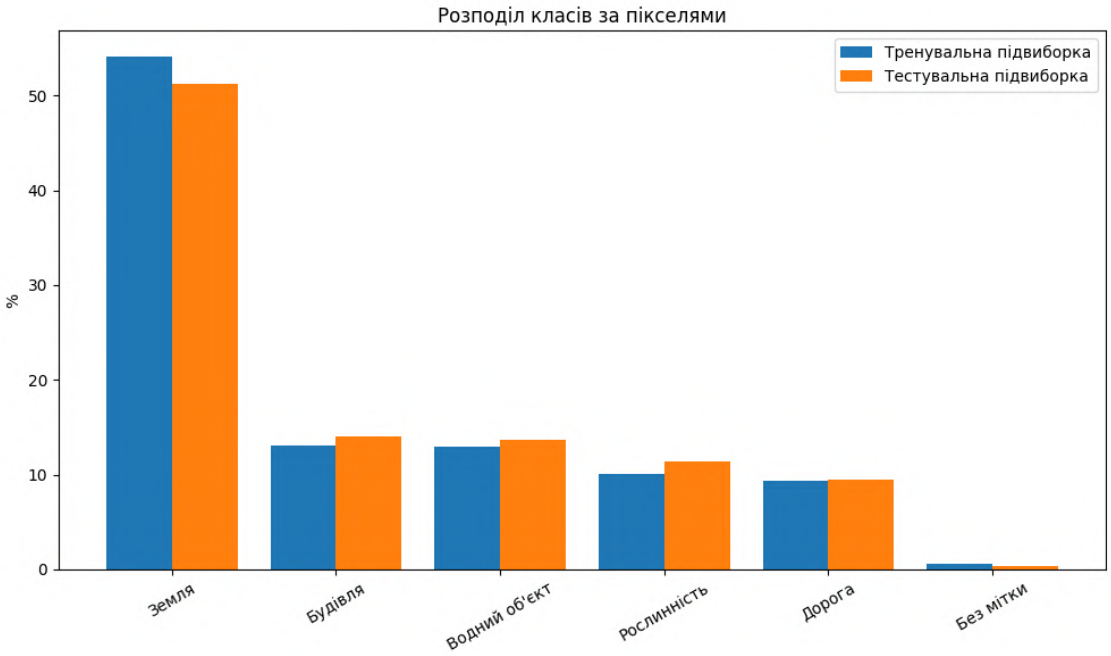

Рисунок 3.2 – Розподіл пікселів зображень набору даних за класами

3.1.2 Методика аугментації набору даних

Для підвищення ефективності навчання моделі та покращення її здатності до узагальнення було застосовано комплексну стратегію аугментації даних. Цей підхід дозволяє створювати велику кількість варіацій вихідних даних, що збагачує модель додатковою інформацією та покращує її стійкість до змін у вхідних даних. Реалізація аугментації була здійснена за допомогою бібліотеки Albumentations, яка є спеціалізованим інструментом для трансформації зображень у задачах комп'ютерного зору та має широкий набір ефективних операцій для обробки зображень, таких як обертання, відображення, зміну розміру та інші. Вона оптимізована для швидкого виконання та дозволяє застосовувати складні трансформації з високою швидкістю. Особливістю застосованого підходу є синхронна

трансформація як вхідних зображень, так і відповідних масок сегментації, що забезпечує збереження семантичної цілісності даних. Це особливо важливо у задачах сегментації зображень, де кожен піксель зображення пов'язаний з певним класом, і будь-які зміни, які відбуваються на зображенні, повинні бути відповідним чином відображені в масці.

Алгоритм аугментації включав кілька ключових трансформацій: горизонтальне та вертикальне відображення застосовувалися з імовірністю 50% для кожного зображення, що дозволило подвоїти різноманітність просторової орієнтації об'єктів. Такі геометричні трансформації є особливо ефективними для супутникових знімків, оскільки семантична інтерпретація більшості об'єктів не залежить від їх орієнтації у просторі.

Додатково застосовувалося обертання зображень в діапазоні від 0 до 180 градусів з високою імовірністю (80%). При цьому використовувався метод інтерполяції «найближчий сусід», що є оптимальним для масок сегментації, оскільки зберігає чіткі межі між класами та запобігає появі артефактів на кордонах сегментів.

Для імітації варіативності умов освітлення, що характерна для супутникових знімків, отриманих у різний час доби та за різних погодних умов, застосовувалася трансформація яскравості та контрасту. Яскравість змінювалася в діапазоні від -20% до +30%, а контраст варіювався в межах ±20%. Ця трансформація застосовувалася до кожного зображення, забезпечуючи стійкість моделі до варіацій освітлення.

Алгоритм також включав альтернативне застосування одного з двох більш складних перетворень. Перше – метод CLAHE з обмеженням контрасту 1.5 та розміром сітки 8×8 пікселів, що покращує локальний контраст зображення та виділення деталей у затінених або пересвітлених областях. Друге – сіткова дисторсія (GridDistortion), що імітує геометричні спотворення, які можуть виникати внаслідок кривизни земної поверхні або

недосконалостей оптичних систем супутників. Одне із цих перетворень застосовувалося з імовірністю 50%.

На рисунку 3.3 наведені приклади аугментації одного й того самого зображення за допомогою створеного алгоритму аугментації.

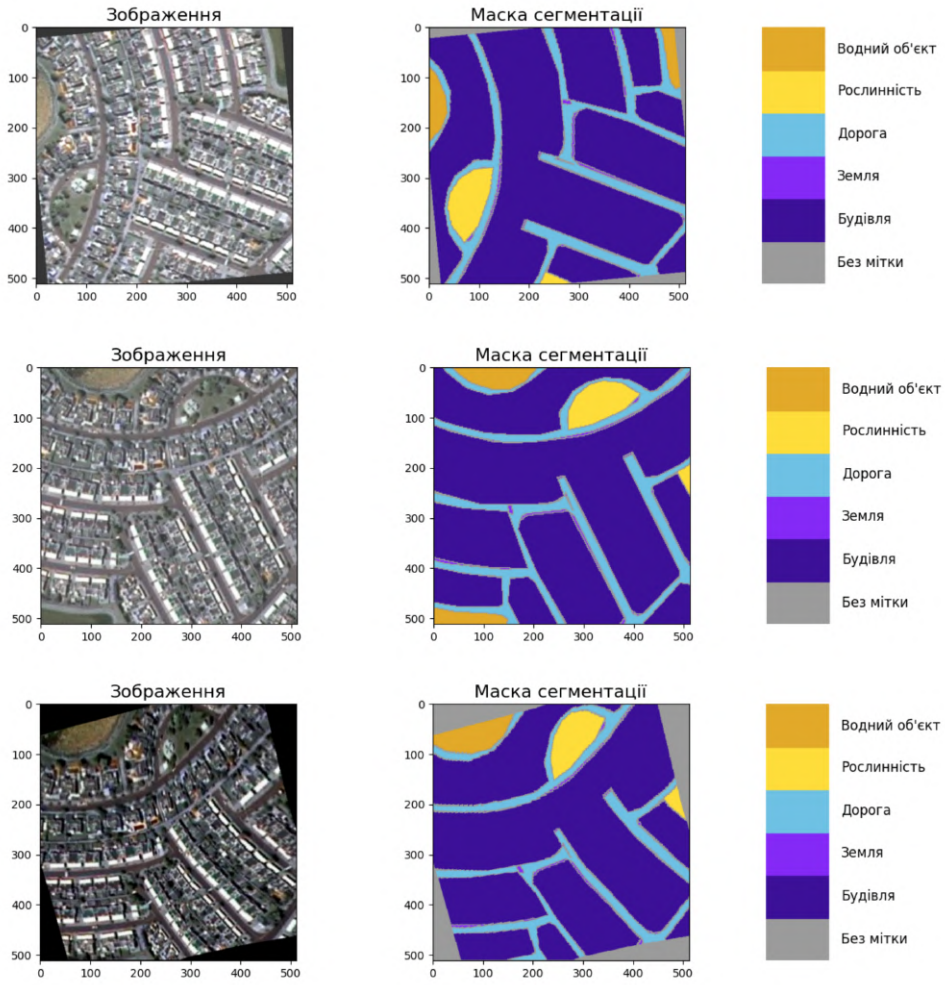

Рисунок 3.3 – Приклад аугментації тренувального зображення

3.2 Налаштування середовища експерименту

Для проведення досліджень було сформовано експериментальне середовище, що складалося з апаратного та програмного забезпечення, необхідного для ефективної розробки, навчання та тестування моделі.

Основним середовищем розробки було обрано Google Colaboratory (Google Colab) – хмарну платформу для досліджень у галузі машинного навчання та штучного інтелекту, що дозволяє виконувати обчислення на серверах Google з використанням GPU та TPU. Ця платформа надала значні переваги для реалізації проекту, зокрема безкоштовний доступ до потужних обчислювальних ресурсів, що критично важливо для навчання глибоких нейронних мереж [30]. Google Colab забезпечив доступ до GPU NVIDIA T4, що дозволило значно прискорити процес навчання моделі порівняно з навчанням на центральному процесорі.

Для розробки та навчання моделі використовувався мовний інтерпретатор Python версії 3.8, що є стандартним для середовища Google Colab. Python був обраний завдяки широкій екосистемі бібліотек для машинного навчання та обробки зображень, а також через простоту синтаксису та гнучкість у розробці. Важливою складовою програмного забезпечення стали фреймворки та бібліотеки для глибокого навчання. Зокрема, TensorFlow було використано як основну платформу для створення, навчання та оцінки моделі. Ці інструменти надали широкий спектр готових компонентів для побудови нейронних мереж, а також зручні API для їх конфігурації та оптимізації.

Для обробки та підготовки даних були застосовані бібліотеки NumPy та Pandas, що забезпечили ефективні структури даних та операції для маніпуляції з великими масивами інформації. Бібліотека OpenCV та PIL використовувалися для роботи з зображеннями – зчитування, попередньої обробки, трансформації та аугментації даних. Саме аугментація даних стала важливим компонентом дослідження, оскільки дозволила штучно розширити навчальний набір та покращити узагальнюючу здатність моделі. Для цього застосовувалися різноманітні трансформації: поворот, віддзеркалення, зміна яскравості, контрасту та інші методи, реалізовані за допомогою бібліотеки albumentations, що спеціалізується на аугментації зображень для задач комп'ютерного зору.

Для візуалізації результатів та аналізу прогресу навчання використовувалися бібліотеки Matplotlib та Seaborn, що дозволили створювати інформативні графіки та візуальні представлення даних. Scikit-learn надав інструменти для оцінки якості моделі та розрахунку метрик, таких як точність, повнота, F1-міра та IoU, що є особливо важливими для задач семантичної сегментації.

Для зберігання та організації даних було використано Google Drive, інтегрований з Google Colab, що забезпечило зручний доступ до великих наборів супутникових зображень, масок сегментації та проміжних результатів навчання. Використання хмарного сховища також дозволило зберігати чекпоїнти моделі під час довготривалого навчання, що є важливим для запобігання втраті прогресу у випадку збоїв системи.

3.3 Навчання моделі та метрики оцінки

3.3.1 Опис процесу навчання моделі

Навчання моделі проводилось з використанням генераторів даних, які дають змогу ефективно працювати з великими наборами зображень, не завантажуючи всю вибірку в оперативну пам’ять. Це особливо важливо в умовах обмежених апаратних ресурсів, коли розмір датасету не дозволяє здійснювати повну попередню обробку.

Було встановлено розмір вхідних зображень 512 x 512 пікселів з 3 кольоровими каналами (RGB), що відповідає типовому формату супутникових знімків високої роздільності. Такий розмір дозволяє зберігати достатній рівень просторової деталізації, водночас не перевантажуючи графічний процесор.

Процес навчання тривав протягом 30 епох з початковою швидкістю навчання 0.0001. Для покращення збіжності мережі та запобігання переобученню було реалізовано механізм експоненційного зменшення

швидкості навчання, який дозволяє поступово зменшувати крок оновлення вагів, наближаючись до локального оптимуму з високою точністю (рисунок 3.4).

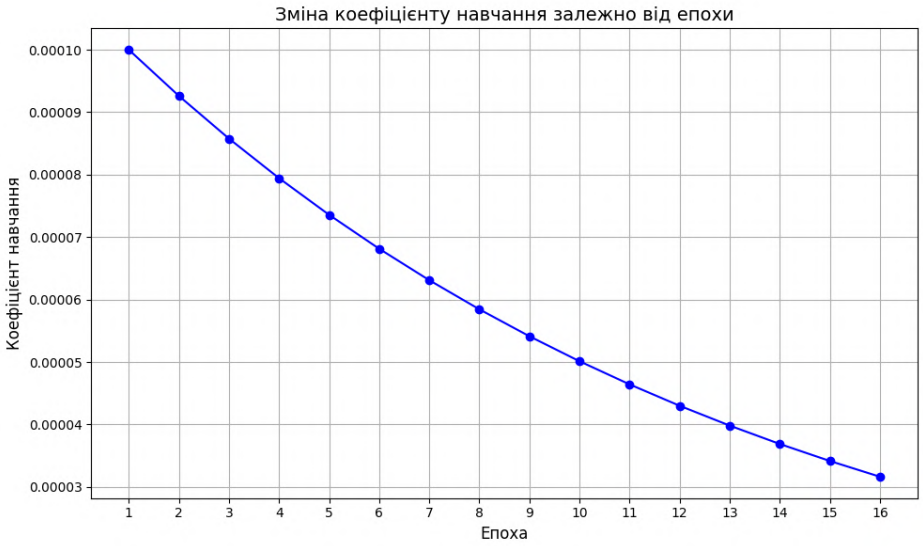

Рисунок 3.4 – Графік зміни коефіцієнту навчання в залежності від епохи

Для контролю процесу навчання та запобігання перенавчанню було використано набір callback-функцій, включаючи ModelCheckpoint для збереження найкращої моделі на основі показників на валідаційному наборі, EarlyStopping для зупинки навчання у випадку відсутності покращення протягом 5 епох, а також CSVLogger для документування процесу навчання.

Особливістю процесу навчання було використання попередньо навченої моделі InceptionResNetV2 як енкодера, що дозволило скористатись перевагами трансферного навчання. Спочатку базова модель була заморожена, що дозволило зберегти вагові коефіцієнти, отримані під час попереднього навчання на наборі даних ImageNet. Після 5 епохи модель енкодера була розморожена для більш точного налаштування її ваг.

Для запобігання перенавчанню в архітектурі було використано шар Dropout з коефіцієнтом відсіву 0.3 безпосередньо перед вихідним шаром.

Вихідний шар реалізований як згортковий шар з 6 фільтрами та функцією активації softmax, що відповідає кількості класів для сегментації.

На рисунку 3.5 представлено динаміку зміни значень функції втрат під час навчання моделі. Навчання моделі було завершено після 16 епохи через замалу зміну значення функції втрат тестувальної підвиборки між епохами. Як можна бачити з графіка, значення функції втрат стабільно зменшувалось протягом процесу навчання для тренувальної та тестувальної підвиборок. Цей графік демонструє, що модель успішно навчається та покращує результати з кожною епохою до певного моменту, після якого покращення стає незначним. Також необхідно зазначити, що модель не зазнала перенавчання, адже значення функції втрат для обох підвиборок є доволі близькими навіть під час останніх епох.

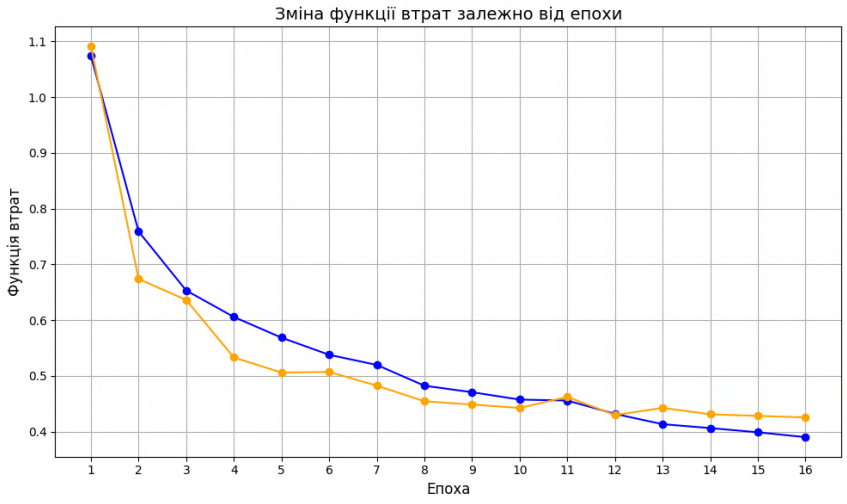

Рисунок 3.5 – Графік зміни значення функції втрат в залежності від епохи

На рисунку 3.6 представлена візуалізація зміни результатів роботи моделі на одному з зображень тестувальної підвиборки протягом кількох епох навчання. В результаті першої епохи моделі має дуже низьку точність – більшість пікселів не класифіковані через низький рівень впевненості моделі у результатах сегментації. Після восьмої епохи модель вже здатна відділяти об‘єкти один від одного, але їй бракує здатності чітко відділяти їх

межі один від одного. В результаті останньої епохи модель чітко відділяє об‘єкти один від одного з мінімальною кількістю помилок.

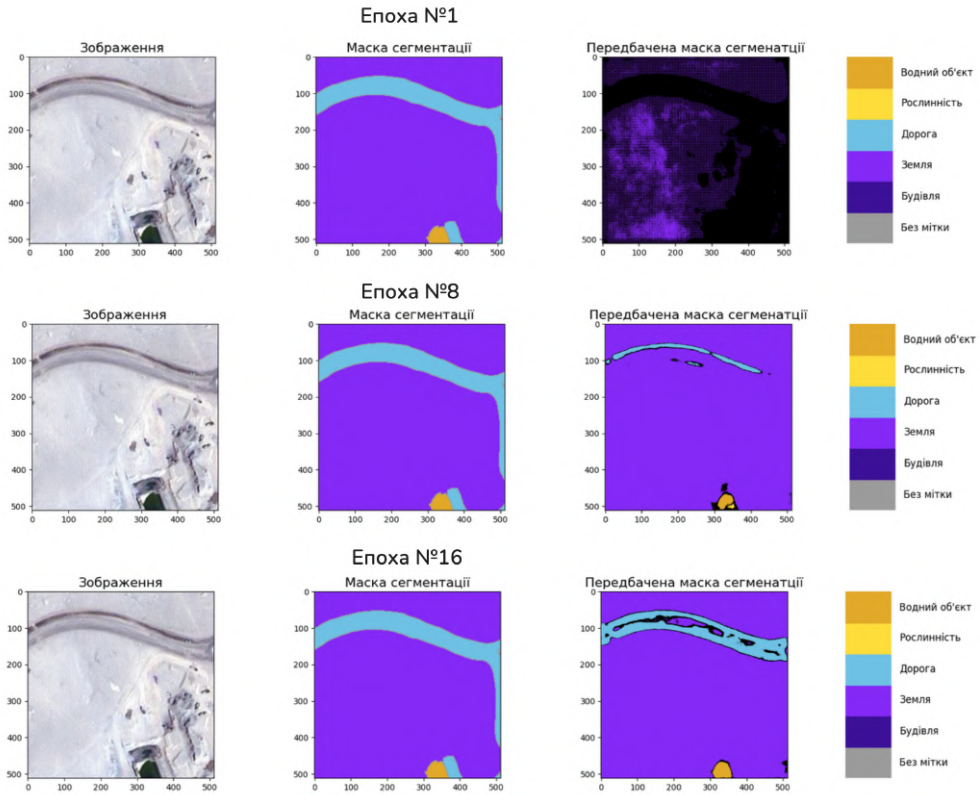

Рисунок 3.6 – Зміна результатів сегментації протягом навчання

3.3.2 Аналіз результатів навчання

Протягом навчання моделі з кожною епохою було виміряне не лише значення функції втрат моделі на тренувальній та тестувальній підвиборках, а і значення точності та коефіцієнту Дайса відповідно. Як можна побачити на рисунку 3.7, значення вищезгаданих метрик поступово збільшуються з кількістю пройдених епох, а навчання закінчується тоді, коли значення тестувальної підвиборки виходять на плато. Близкість метрик між підвиборками в результаті навчання моделі говорить про відсутність ефекту перенавчання – моделі вдалося зберегти здатність до узагальнення.

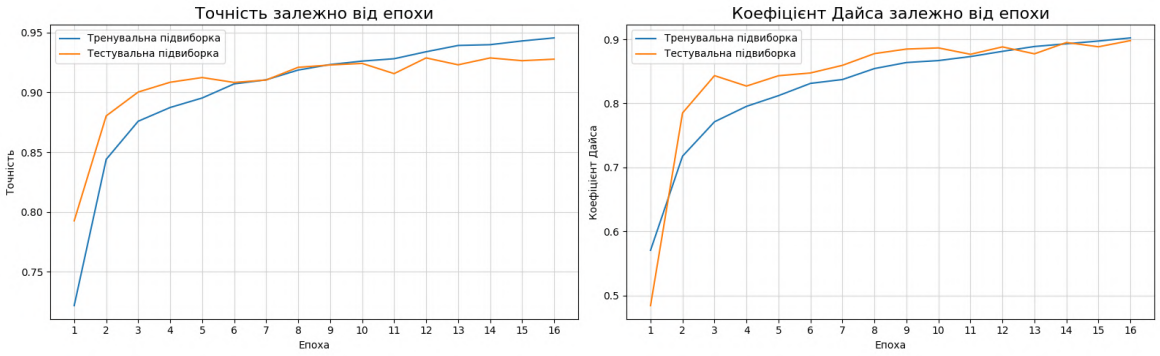

Рисунок 3.7 – Зміна метрик сегментації протягом навчання

На рисунку 3.8 наведено матрицю помилок моделі на зображеннях тестувальної підвиборки. Точність сегментації кожного з основних класів є високою (більше 90%). Єдиний клас, що має низьку точність є допоміжний клас, відповідаючий пікселям без мітки – це очікувано, адже цей клас передбачається моделлю у випадку, якщо їй бракує впевненості у тому, що окремий піксель належить до одного з основних класів. Найвищу точність сегментації мають класи землі та водного об‘єкту відповідно.

Рисунок 3.8 – Матриця помилок для тестувальних зображень

Враховуючи дисбаланс класів, на основі представленої матриці помилок було вираховано значення точності сегментації по всім класам задачі, воно склало 92%.

На рисунку 3.9 представлені значення коефіцієнту Дайса по кожному класу використаного набору даних тестувальної підвиборки. Високі значення метрики говорять про здатість моделі ефективно сегментувати супутникові зображення, припускаючись вкрай невеликої кількості помилок.

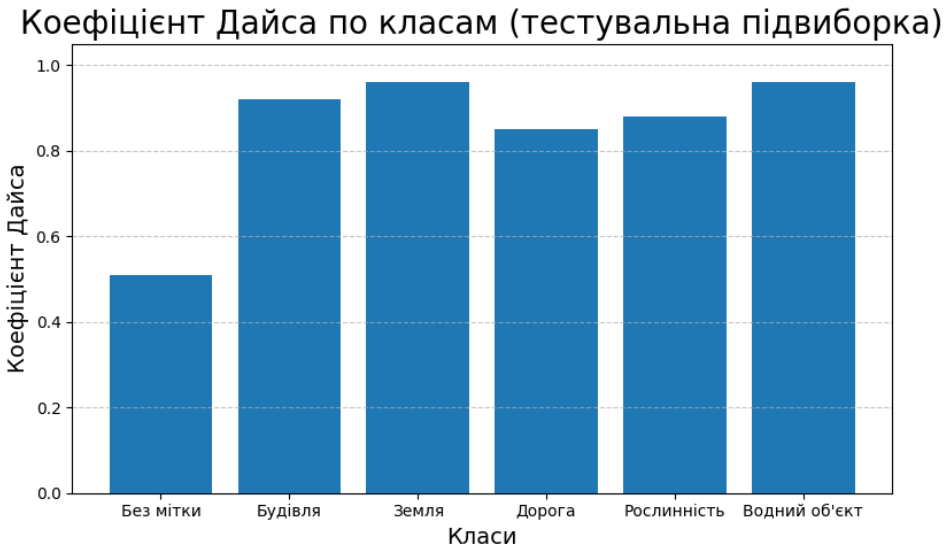

Рисунок 3.9 – Матриця помилок для тестувальних зображень

На рисунку 3.10 представлені приклади сегментації зображень тестувальної підвиборки за допомогою навченої моделі. Візуальний аналіз результатів підтверджує виміряний рівень точності – переважна більшість пікселів має правильне передбачення класу. Варто відмітити, що на деяких передбачених масках сегментації присутні певні артефакти, здебільшого на краях відповідних об‘єктів.

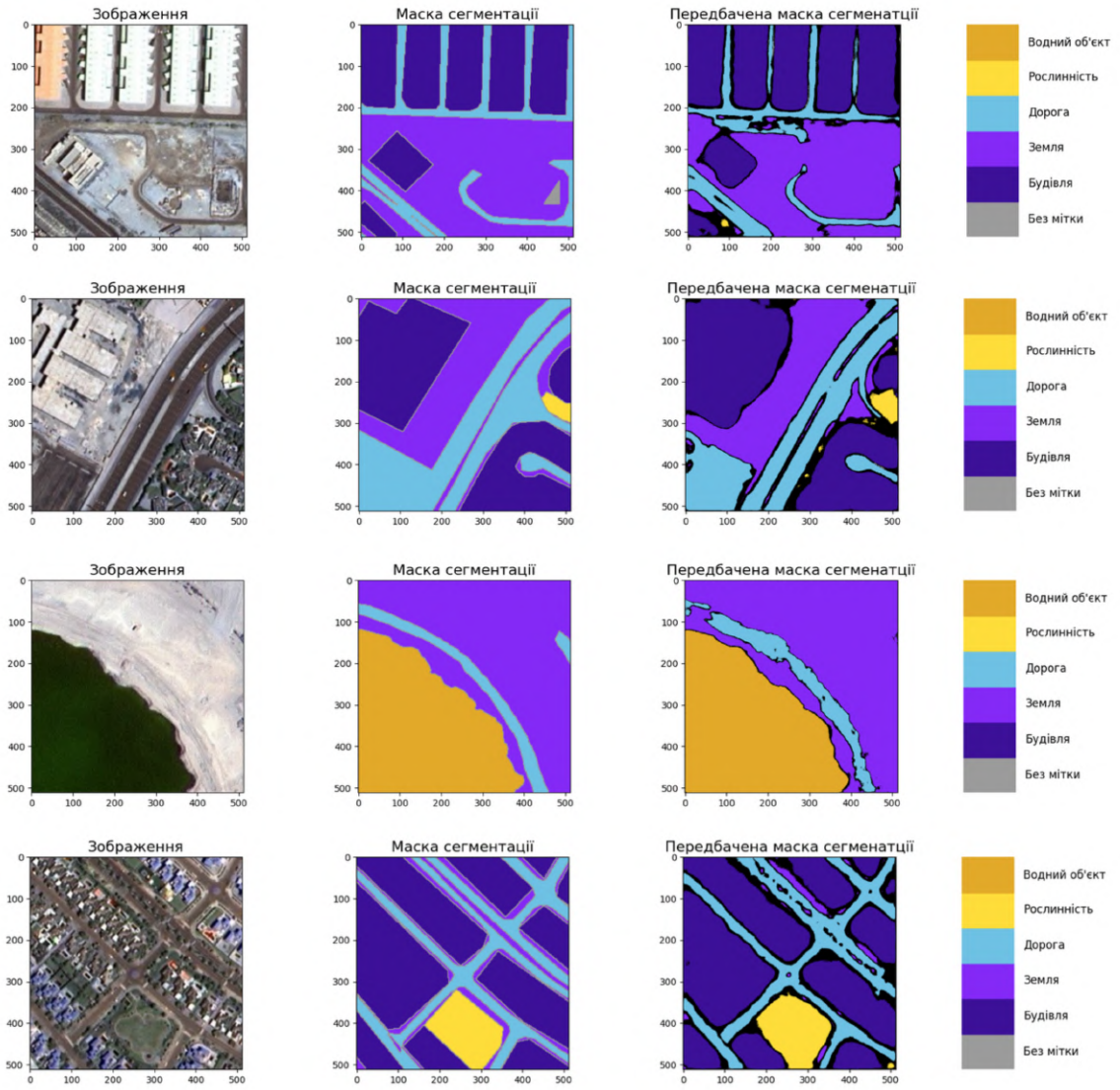

Рисунок 3.10 – Приклад сегментації тестувальних зображень

3.4 Оцінка роботи моделі на знімках різного масштабу

Дослідження масштабованості та узагальнення є важливим етапом оцінки розробленої моделі сегментації супутникових зображень. Ці характеристики визначають практичну цінність моделі для реальних застосувань.

Масштабованість моделі розглядається як її здатність ефективно працювати з супутниковими зображеннями різної просторової роздільної здатності. Під час тренування модель навчалася на даних з роздільною

здатністю 0.75 метри на піксель. Для оцінки масштабованості був проведено серію експериментів з використанням тестових зображень з іншими масштабами: 0.375 метри на піксель та 1.5 метри на піксель. Значення точності класифікації по кожному з класів в залежності від масштабу супутникового зображення наведені в таблиці 3.1.

Таблиця 3.1 – Точність сегментації в залежності від масштабу супутникових знімків

Масштаб (м/п)	Без мітки	Будівля	Земля	Дорога	Рослинність	Водний об‘єкт
0.375	40.1%	91.9%	97.2%	90.8%	92.3%	98%
0.75	**36.2%**	**90.5%**	**95%**	**90.2%**	**90%**	**97.2%**
1.5	26%	80.1%	82.3%	83.8%	79.9%	94%

Отримані результати демонструють, що модель сегментації має певну здатність до масштабування, тобто вона може працювати з супутниковими зображеннями різної роздільної здатності, навіть якщо під час навчання використовувались дані лише з одного масштабу. Найвищі показники точності були зафіксовані на зображеннях із вищою деталізацією – 0.375 метри на піксель. Це свідчить про те, що модель може ефективно адаптуватися до кращої якості вхідних даних без необхідності повторного навчання.

Зі зменшенням просторової роздільної здатності до 1.5 метра на піксель точність класифікації помітно знижується. Найбільше це вплинуло на класи, де для точного розпізнавання необхідна чітка структура, зокрема будівлі та рослинність. Це пояснюється тим, що при грубішому масштабі стає важче розрізнити дрібні об’єкти та деталі, що важливі для коректної сегментації.

Попри це, навіть при меншій роздільній здатності модель зберігає задовільні результати для класів, що мають великі або однорідні просторові структури, наприклад для водних об’єктів або доріг. Це дозволяє зробити

висновок, що модель частково узагальнює на нові умови, хоча її продуктивність знижується разом зі зменшенням якості вхідного зображення.

Загалом, результати свідчать про те, що модель може бути застосована в умовах змінної роздільної здатності супутникових зображень, однак для покращення її стійкості в таких випадках варто розглянути додаткові підходи на етапі навчання або попередньої обробки даних.

3.5 Напрями подальших досліджень

Розширення досліджень щодо сегментації супутникових зображень доцільно зосередити на кількох ключових напрямах. Збагачення навчальних масивів даних шляхом інтеграції супутникових знімків, отриманих за різних погодних умов, пір року та часу доби, дозволить підвищити стійкість моделі до варіативності вхідних даних. Такий підхід сприятиме створенню більш універсальних алгоритмів, здатних коректно сегментувати зображення навіть при значних змінах освітлення, наявності хмарності або інших атмосферних явищ, що часто призводять до зниження точності розпізнавання.

Впровадження гіперспектральних даних у процес сегментації відкриває доступ до суттєво ширшого спектрального діапазону порівняно з традиційними мультиспектральними знімками. Використання сотень спектральних каналів замість стандартних RGB або NIR забезпечує значно детальнішу спектральну сигнатуру об'єктів, що особливо цінно при розрізненні подібних типів рослинності, мінеральних поверхонь або антропогенних об'єктів. Аналіз специфічних спектральних характеристик на недоступних людському оку частотах може радикально покращити точність класифікації складних ландшафтів.

Архітектурні інновації в структурі енкодера становлять перспективний напрям удосконалення існуючої моделі. Заміна традиційних

згорткових блоків на механізми уваги та трансформерні структури дозволить ефективніше опрацьовувати контекстну інформацію та просторові залежності між різними частинами зображення. Трансформерні архітектури, що продемонстрували високу ефективність в задачах комп'ютерного зору, можуть забезпечити кращу генералізацію при навчанні на обмежених наборах даних та покращити розпізнавання об'єктів зі складною просторовою структурою.

Інтеграція тривимірних геопросторових даних, таких як цифрові моделі рельєфу, у процес сегментації супутникових зображень відкриває принципово нові можливості для точної класифікації. Врахування висотних характеристик місцевості та морфометричних параметрів рельєфу забезпечує додатковий контекст, що допомагає диференціювати об'єкти зі схожими спектральними характеристиками, але різною геоморфологічною позицією. Синергія спектральних та топографічних даних створює передумови для комплексного аналізу ландшафтів з урахуванням їх тривимірної структури.

ВИСНОВКИ

В результаті проведеної розробки системи для сегментації супутникових знімків була розроблена потужна модель з точністю сегмантації 92%. Модель була навчана на справжніх супутникових знімках, зроблених над територією ОАЕ.

У ході виконання кваліфікаційної роботи було повністю виконано поставлені завдання з розробки та дослідження системи для семантичної сегментації супутникових знімків в умовах обмеженої вибірки даних. Проведено всебічний аналіз існуючих методів семантичної сегментації та обґрунтовано доцільність використання глибокого навчання для вирішення даної задачі.

Використання архітектури U-Net з енкодером на основі InceptionResNetV2 дозволило досягти високої точності сегментації, що підтверджується експериментальними результатами. Модель продемонструвала здатність точно класифікувати пікселі навіть при обмеженій кількості навчальних даних, досягаючи показників точності понад 90% для всіх основних класів. Це стало можливим завдяки комбінації трансферного навчання, яке дозволило використати попередньо навчені ваги для виділення ознак, та механізму skip-зв'язків, що забезпечив збереження просторової інформації.

Важливим аспектом роботи стала ретельна підготовка даних, включаючи аугментацію, яка значно розширила різноманітність навчального набору. Застосування таких методів, як обертання, зміна яскравості та контрасту, дозволило покращити узагальнюючу здатність моделі. Крім того, використання зваженої функції втрат допомогло подолати проблему дисбалансу класів, що є характерним для супутникових зображень. Експерименти показали, що модель ефективно справляється із завданням сегментації навіть для дрібних об'єктів, таких як будівлі чи дороги, що є критично важливим для практичного застосування.

Дослідження виконано в рамках наукового напрямку університетського факультету комп'ютерних наук та має зв'язок з дослідженнями лабораторії штучного інтелекту університету. Отримані результати доповнюють існуючі розробки кафедри в галузі систем моніторингу та дистанційного зондування Землі.

Результати роботи мають значний потенціал для використання в реальних умовах, зокрема для моніторингу стану навколишнього середовища, міського планування та управління природними ресурсами. Модель може бути інтегрована в системи дистанційного зондування Землі, де висока точність сегментації є ключовою вимогою. Однак залишаються напрями для подальшого вдосконалення, такі як оптимізація обчислювальної складності для роботи в режимі реального часу та адаптація моделі до різних географічних регіонів з урахуванням їхніх особливостей.

У цілому, проведене дослідження підтвердило ефективність запропонованого підходу до сегментації супутникових зображень в умовах обмежених даних. Розроблена модель може слугувати основою для подальших досліджень у галузі комп'ютерного зору та дистанційного зондування, сприяючи розвитку автоматизованих систем аналізу великих обсягів супутникових даних.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Ng V., Hofmann D. Scalable feature extraction with aerial and satellite imagery. *Python in science conference*, Austin, Texas. 2018. URL: https://doi.org/10.25080/majora-4af1f417-015 (date of access: 22.03.2025).

2. A beginner's guide to segmentation in satellite images: walking through machine learning techniques for image segmentation and applying them to satellite imagery – GSI technology. *GSI Technology*. URL: https://gsitechnology.com/beginners-guide-to-segmentation-in-satellite-images/ (date of access: 22.03.2025).

3. Overview of sentinel-2 / F. Spoto et al. *IGARSS 2012 - 2012 IEEE international geoscience and remote sensing symposium*, Munich, Germany, 22 – 27 July 2012. 2012. URL: https://doi.org/10.1109/igarss.2012.6351195 (date of access: 22.03.2025).

4. Kumar M. S. Patch based approach for Satellite Aerial Image Segmentation using DeepLabV3+. *2024 3rd international conference for advancement in technology (ICONAT)*, GOA, India, 6–8 September 2024. 2024. P. 1–4. URL: https://doi.org/10.1109/iconat61936.2024.10774644 (date of access: 22.03.2025).

5. Порівняння landsat-9 та sentinel-2 – portalgis. *portalgis*. URL: https://portalgis.pro/kosmoznimki/porivnyannya-landsat-9-ta-sentinel-2/ (дата звернення: 22.03.2025).

6. Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation / R. Jiao et al. *Computers in biology and medicine*. 2023. P. 107840. URL: https://doi.org/10.1016/j.compbiomed.2023.107840 (date of access: 22.03.2025).

7. Ghaderpour E., Vujadinovic T. Change detection within remotely sensed satellite image time series via spectral analysis. *Remote sensing*. 2020. Vol. 12,

no. 23. P. 4001. URL: https://doi.org/10.3390/rs12234001 (date of access: 22.03.2025).

8. Real-Time high-performance semantic image segmentation of urban street scenes / G. Dong et al. *IEEE transactions on intelligent transportation systems*. 2020. P. 1–17. URL: https://doi.org/10.1109/tits.2020.2980426 (date of access: 22.03.2025).

9. Xie X. A k-nearest neighbor technique for brain tumor segmentation using minkowski distance. *Journal of medical imaging and health informatics*. 2018. Vol. 8, no. 2. P. 180–185. URL: https://doi.org/10.1166/jmihi.2018.2285 (date of access: 22.03.2025).

10. . R. G. V. Noise tolerant color image segmentation using support vector machine. *International journal of research in engineering and technology*. 2014. Vol. 03, no. 16. P. 29–34. URL: https://doi.org/10.15623/ijret.2014.0316006 (date of access: 22.03.2025).

11. Shrivastava K., Gupta N., Sharma N. Medical image segmentation using modified K means clustering. *International journal of computer applications*. 2014. Vol. 103, no. 16. P. 12–16. URL: https://doi.org/10.5120/18157-9341 (date of access: 22.03.2025).

12. Unsupervised image segmentation using hierarchical clustering / K. Ohkura et al. *Optical review*. 2000. Vol. 7, no. 3. P. 193–198. URL: https://doi.org/10.1007/s10043-000-0193-8 (date of access: 22.03.2025).

13. Zhou H., Wang X., Schaefer G. Mean shift and its application in image segmentation. *Innovations in intelligent image analysis*. Berlin, Heidelberg, 2011. P. 291–312. URL: https://doi.org/10.1007/978-3-642-17934-1_13 (date of access: 22.03.2025).

14. Shelhamer E., Long J., Darrell T. Fully convolutional networks for semantic segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 2017. Vol. 39, no. 4. P. 640–651. URL: https://doi.org/10.1109/tpami.2016.2572683 (date of access: 22.03.2025).

15. Fawzy M., Barsi A. A u-net model for urban land cover classification using VHR satellite images. *Periodica polytechnica civil engineering*. 2024. URL: https://doi.org/10.3311/ppci.37599 (date of access: 22.03.2025).

16. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs / L.-C. Chen et al. *IEEE transactions on pattern analysis and machine intelligence*. 2018. Vol. 40, no. 4. P. 834–848. URL: https://doi.org/10.1109/tpami.2017.2699184 (date of access: 22.03.2025).

17. Pyramid scene parsing network / H. Zhao et al. *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, 21–26 July 2017. 2017. URL: https://doi.org/10.1109/cvpr.2017.660 (date of access: 22.03.2025).

18. Inception-v4, inception-resnet and the impact of residual connections on learning / C. Szegedy et al. *Proceedings of the AAAI conference on artificial intelligence*. 2017. Vol. 31, no. 1. URL: https://doi.org/10.1609/aaai.v31i1.11231 (date of access: 11.05.2025).

19. Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. *Lecture notes in computer science*. Cham, 2015. P. 234–241. URL: https://doi.org/10.1007/978-3-319-24574-4_28 (date of access: 11.05.2025).

20. Going deeper with convolutions / C. Szegedy et al. *2015 IEEE conference on computer vision and pattern recognition (CVPR)*, Boston, MA, USA, 7–12 June 2015. 2015. URL: https://doi.org/10.1109/cvpr.2015.7298594 (date of access: 11.05.2025).

21. Data augmentation approaches for satellite image super-resolution / M. A. A. Ghaffar et al. *ISPRS annals of photogrammetry, remote sensing and spatial information sciences*. 2019. IV-2/W7. P. 47–54. URL: https://doi.org/10.5194/isprs-annals-iv-2-w7-47-2019 (date of access: 11.05.2025).

22. Kumar T. S. Enhancement of image classification through data augmentation using machine learning. *International journal of computer sciences*

and engineering. 2018. Vol. 6, no. 9. P. 220–224. URL: https://doi.org/10.26438/ijcse/v6i9.220224 (date of access: 11.05.2025).

23. CutMix: regularization strategy to train strong classifiers with localizable features / S. Yun et al. *2019 IEEE/CVF international conference on computer vision (ICCV)*, Seoul, Korea (South), 27 October – 2 November 2019. 2019. URL: https://doi.org/10.1109/iccv.2019.00612 (date of access: 11.05.2025).

24. Data augmentation method using generative adversarial networks / O. Chaikovskyi et al. *Technical sciences and technologies*. 2021. No. 2(24). P. 83–91. URL: https://doi.org/10.25140/2411-5363-2021-2(24)-83-91 (date of access: 11.05.2025).

25. A comprehensive survey on transfer learning / F. Zhuang et al. *Proceedings of the IEEE*. 2021. Vol. 109, no. 1. P. 43–76. URL: https://doi.org/10.1109/jproc.2020.3004555 (date of access: 11.05.2025).

26. Land cover segmentation in satellite images using transfer learning / K. S. Kalaivani et al. *2024 second international conference on intelligent cyber physical systems and internet of things (icoici)*, Coimbatore, India, 28–30 August 2024. 2024. P. 714–724. URL: https://doi.org/10.1109/icoici62503.2024.10695992 (date of access: 11.05.2025).

27. Demystifying learning rate policies for high accuracy training of deep neural networks / Y. Wu et al. *2019 IEEE international conference on big data (big data)*, Los Angeles, CA, USA, 9–12 December 2019. 2019. URL: https://doi.org/10.1109/bigdata47090.2019.9006104 (date of access: 11.05.2025).

28. Optimizers in deep learning: a comparative study and analysis / S. Bashetty et al. *International journal for research in applied science and engineering technology*. 2022. Vol. 10, no. 12. P. 1032–1039. URL: https://doi.org/10.22214/ijraset.2022.48050 (date of access: 11.05.2025).

29. Semantic segmentation dataset | humans in the loop. *Humans in the Loop | Continuously better models using a human-in-the-loop*. URL: https://humansintheloop.org/resources/datasets/semantic-segmentation-dataset-2/ (date of access: 11.05.2025).

30. Google colab. *Google Research – Explore Our Latest Research in Science and AI*. URL: https://research.google.com/colaboratory/faq.html (date of access: 11.05.2025).