Київський національний університет імені Тараса Шевченка
Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка
Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

ПУШКАРЕНКО ЮРІЙ ВАЛЕРІЙОВИЧ

УДК 004.8, 004.93, 004.421

ДИСЕРТАЦІЯ

МОДЕЛІ ТА МЕТОДИ ЛОКАЛІЗАЦІЇ СЦЕН ЗОБРАЖЕНЬ ОБ'ЄКТІВ КРИТИЧНОЇ ІНФРАСТРУКТУРИ НА ОСНОВІ КОМПОЗИТНИХ НЕЙРОННИХ МЕРЕЖ

Спеціальність 121 - Інженерія програмного забезпечення
Галузь знань 12 - Інформаційні технології

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело
_____ Пушкаренко Юрій Валерійович

Науковий керівник:
доктор технічних наук, професор
Заславський Володимир Анатолійович

Київ - 2025

АНОТАЦІЯ

Пушкаренко Ю.В. Моделі та методи локалізації сцен зображень об’єктів критичної інфраструктури на основі композитних нейронних мереж. - Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 121 «Інженерія програмного забезпечення» (12 - Інформаційні технології). - Київський національний університет імені Тараса Шевченка. - Київський національний університет імені Тараса Шевченка, Київ, 2024.

Метою роботи є розробка та впровадження методів та моделей локалізації сцен зображень дистанційного зондування на основі композитних нейронних мереж, здатної ефективно вирішувати проблеми точного виявлення сцен об’єктів та їх пошкоджень. Пропоновані моделі та методи спрямовані на подолання існуючих обмежень традиційних нейронних мереж у завданнях аналізу зображень великої роздільної здатності, де необхідно обробляти різні об’єкти на різних масштабах і з високою точністю. Головна увага приділяється задачам, які виникають під час аналізу складних сцен дистанційного зондування, де велика кількість об’єктів і змінювані умови зображення роблять традиційні методи малоефективними.

Наукова новизна. Наукова новизна роботи полягає у впровадженні нової моделі та методів композитної нейронної мережі які продовжують ідею принципу різнотипності в системах прийняття рішень з високою надійністю, яка поєднує в собі кращі властивості згорткових нейронних мереж (CNN) та трансформерів, зокрема архітектури Swin (shifted window transformer). Це дозволяє мережі одночасно забезпечувати детальну обробку дрібних ознак зображення та захоплювати глобальні контексти сцени, що є важливим для точного виявлення та класифікації об’єктів. У роботі також запропоновано новий підхід до вирішення проблеми обмеженого рецептивного поля, яке характерне для традиційних згорткових мереж. Пропоновані модель і метод впроваджують модуль динамічного масштабування рецептивного поля з увагою (DReAM), що дозволяють моделі адаптивно змінювати свої параметри для ефективної обробки сцен різного розміру та складності. Це забезпечує можливість роботи з великими обсягами даних без втрати якості та точності.

Крім того, у роботі вперше застосовано механізм уваги для оптимізації процесу локалізації пошкоджень на зображеннях. Механізм уваги дозволяє мережі фокусуватися на найбільш значущих ділянках зображення, ігноруючи менш важливі або шумові області, що значно підвищує точність виявлення пошкоджень та інших важливих деталей. Цей підхід є інноваційним для задач локалізації сцен та відкриває нові можливості для аналізу даних дистанційного зондування в реальних умовах.

Проблематика. Однією з ключових проблем, які вирішуються у цій роботі, є обмежений розмір рецептивного поля у традиційних згорткових нейронних мережах, що суттєво впливає на здатність захоплювати важливі контексти великих сцен. У згорткових мережах рецептивне поле збільшується поступово з глибиною мережі, проте це зростання є недостатнім для обробки великих зображень, де необхідно одночасно враховувати як дрібні деталі, так і глобальні взаємозв'язки між об'єктами. Наприклад, при аналізі зображень дистанційного зондування з різними об’єктами інфраструктури, мережа повинна мати змогу захоплювати взаємодії між об'єктами, що знаходяться на великій відстані один від одного, а також детально опрацьовувати локальні пошкодження, що може бути неефективно при використанні традиційних методів.

Крім того, у задачах локалізації сцен часто виникає проблема обробки об'єктів, що мають різні масштаби та форми, що значно ускладнює завдання класифікації та виявлення. Для вирішення цих проблем необхідно впровадження методів, які можуть адаптивно змінювати своє рецептивне поле для роботи з різними масштабами та забезпечувати високу точність обробки даних.

Запропоноване вирішення. Для вирішення цих проблем у роботі запропоновано нову архітектуру композитної нейронної мережі, що поєднує CNN та трансформери, зокрема Swin-трансформери, з використанням динамічного масштабування рецептивного поля з увагою (DReAM). Цей метод дозволяє значно розширити або зменшити рецептивне поле мережі, що забезпечує можливість одночасної обробки як локальних, так і глобальних ознак зображення. Завдяки цьому модель може ефективно розпізнавати об'єкти різних масштабів і забезпечувати точну локалізацію пошкоджень навіть у складних сценах з великою кількістю об’єктів.

Метод уваги, який інтегровано у модель, забезпечує можливість фокусування на найбільш релевантних ділянках зображення, що дозволяє мережі зменшувати вплив шуму та підвищувати точність класифікації. Мережа автоматично виділяє ті області, які є найбільш значущими для завдання локалізації або виявлення пошкоджень, що значно покращує якість аналізу зображень.

Крім того, у роботі запропоновано метод оптимізації обчислювальних ресурсів, що дозволяє моделі ефективно працювати з великими обсягами даних дистанційного зондування без втрати продуктивності. Запропонована архітектура мережі дозволяє адаптивно змінювати параметри моделі залежно від типу даних та умов, що забезпечує високу гнучкість моделі та можливість її застосування в різних галузях.

Експериментальні результати. Розроблена модель була протестована на великому обсязі даних зображень та показала значні покращення в точності та швидкості обробки порівняно з існуючими методами. Зокрема, модель продемонструвала високу ефективність у завданнях локалізації та класифікації пошкоджень об'єктів інфраструктури, таких як мости, дороги та будівлі. Завдяки використанню метода уваги та модулів DReAM модель досягла високої точності навіть у випадках, коли традиційні підходи виявлялися неефективними через обмежені можливості захоплення глобальних контекстів або дрібних деталей.

Висновки. Запропонована модель локалізації сцен зображень на основі композитних нейронних мереж є ефективним рішенням для задач автоматизації аналізу великих обсягів даних. Вона вирішує проблему обмеженого рецептивного поля у традиційних згорткових мережах за допомогою інтеграції адаптера розміру рецептивного поля в залежності від складності сцени, а також підвищує точність виявлення та класифікації об'єктів за рахунок використання механізму уваги. Це робить модель універсальним інструментом для аналізу даних дистанційного зондування в широкому спектрі завдань, зокрема моніторингу та оцінки стану критичних інфраструктурних об'єктів.

Ключові слова: рецептивне поле, дистанційне зондування, нейронні мережі, архітектура трансформер, локалізація сцен, комп’ютерний зір, критичні інфраструктури, пірамідальні мережі, дифузійні моделі, алгоритми просторового

пошуку, обробка зображень, геопросторовий аналіз, виокремлення геометрії об’єктів, семантична сегментація, оцінка інфраструктурних пошкоджень.

ANNOTATION

Pushkarenko Y.V. Development of a Scene Localization System in Remote Sensing Imagery using Composite Neural Networks. - Qualification scientific work on the rights of the manuscript.

Dissertation for obtaining the degree of Doctor of Philosophy in the specialty 121 "Software Engineering" (12 - Information Technologies). - Taras Shevchenko National University of Kyiv. - Kyiv, 2024.

The objective of this research is to develop and implement an advanced scene localization model for remote sensing imagery based on composite neural networks capable of effectively solving precise object detection and damage assessment challenges. The proposed model addresses the limitations of traditional neural networks in processing high-resolution images, where diverse objects appear at multiple scales and with high precision requirements. The focus is placed on challenges that arise in analyzing complex scenes within remote sensing, where a high object count and variable image conditions render conventional methods less effective.

Scientific Novelty. The scientific novelty of this work lies in the introduction of a new composite neural network model that integrates the best attributes of Convolutional Neural Networks (CNNs) and Transformers, specifically the SWIN (Shifted Window Transformer) architecture. This integration allows the network to process fine-grained image details while capturing the global scene context, which is critical for accurate object detection and classification. Additionally, a new approach is introduced to overcome the limited receptive field characteristic of traditional convolutional networks. The proposed system incorporates a dynamic receptive field attention module (DReAM), enabling adaptive parameter adjustment to process scenes of varying sizes and complexity. This feature allows the system to handle large data volumes without compromising quality and accuracy.

Moreover, the study pioneers the use of an attention mechanism to optimize damage localization in remote sensing images. The attention mechanism enhances the model's focus on critical areas, reducing noise influence and improving accuracy in damage and essential detail detection. This approach is innovative for scene localization tasks and opens new opportunities for analyzing remote sensing data in real-world conditions.

Problem Statement. A key problem addressed in this work is the limited receptive field of traditional CNNs, which significantly impacts the ability to capture essential contexts within large scenes. In CNNs, the receptive field expands progressively with network depth, yet this increase is often insufficient for processing large images that require simultaneous attention to both fine details and global object relationships. For instance, in remote sensing imagery with various infrastructure objects, the network must capture interactions among distant objects and accurately process local damages, which can be inefficient with traditional methods.

Scene localization tasks also encounter challenges with objects of different scales and shapes, complicating classification and detection tasks. Addressing these issues necessitates methods that can dynamically adjust their receptive field to handle varying scales and provide high data processing precision.

Proposed Solution. To tackle these issues, the research proposes a new composite neural network architecture that combines CNNs and Transformers, particularly SWIN Transformers, with dynamic receptive field attention module (DReAM). This module significantly broadens the network's receptive field, allowing for the concurrent processing of both local and global image features. Consequently, the system can effectively identify objects across different scales and accurately localize damages, even in complex scenes with numerous objects.

The integrated attention mechanism enhances focus on the most relevant image areas, enabling the network to mitigate noise effects and improve classification accuracy. The network autonomously identifies areas crucial for localization or damage detection tasks, enhancing image analysis quality.

Additionally, the research introduces computational resource optimization, enabling efficient processing of large remote sensing data volumes without compromising

performance. The proposed network architecture adaptively adjusts model parameters based on data type and conditions, ensuring system flexibility and applicability across various fields.

Experimental Results. The developed system was tested on a large volume of remote sensing imagery data, showing significant improvements in processing accuracy and speed compared to existing methods. The system demonstrated high effectiveness in infrastructure damage localization and classification tasks, such as for bridges, roads, and buildings. Leveraging the attention mechanism and DReAM modules, the system achieved high accuracy even where traditional approaches fell short due to limitations in capturing global contexts or fine details.

Conclusions. The proposed scene localization system for remote sensing imagery, based on composite neural networks, provides an effective solution for automating the analysis of large data volumes. It addresses the limited receptive field issue in traditional convolutional networks by integrating a dynamic receptive field adapter relative to scene complexity and enhances object detection and classification accuracy through attention mechanisms. This makes the system a versatile tool for remote sensing data analysis in a broad range of applications, including critical infrastructure monitoring and assessment.

Keywords: receptive field, remote sensing, convolutional neural networks, transformer architecture, scene localization, computer vision, critical infrastructure, pyramid networks, spatial search algorithms, image processing, geospatial analysis, semantic segmentation, infrastructure damage assessment.

ЗМІСТ

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА:

Наукові праці, в яких опубліковані основні наукові результати дисертації:

1. Yurii Pushkarenko, Volodymyr Zaslavskiy. Principal Curve Trajectory Analysis // Збірник наукових праць Військового Інституту Київського Національного Університету ім. Тараса Шевченка. Серія: Військова техніка і технології подвійного призначення, 2021, № 73, с. 17–29. DOI: https://doi.org/10.17721/2519-481X/2021/73-03.
2. Yurii Pushkarenko, Volodymyr Zaslavskiy. Research on the state of areas in Ukraine affected by military actions based on remote sensing data and deep learning architectures. Radioelectronic and Computer Systems, 2024(2), 5-18. DOI: https://doi.org/10.32620/reks.2024.2.01 (Q3)(SCOPUS).
3. Yurii Pushkarenko, Volodymyr Zaslavskiy. 2024. Synthetic Data Generation for Fraud Detection using Diffusion models. DOI: https://doi.org/10.11610/isij.5534.
4. Yurii Pushkarenko, Volodymyr Zaslavskiy. 2024. Multiscale Scene Localization Based on Composite Network for Remote Sensing Imageries: A Case Study on Critical Infrastructure. 14th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 2024.

5. Pushkarenko, Yurii, Volodymyr Zaslavskyi. 2025. “Model Development of Dynamic Receptive Field for Remote Sensing Imageries”. *Technology Audit and Production Reserves* 1 (2(81):20-25. https://doi.org/10.15587/2706-5448.2025.323698.

Наукові праці, які засвідчують апробацію матеріалів дисертації:

1. Pushkarenko Y. 2021. Improvements in Median Trajectory Analysis, in section: Innovative technologies and risk management in industry, transport and services in the face of modern challenges. MATHEMATICAL MODELING, OPTIMIZATION, AND INFORMATION TECHNOLOGIES, MMOTI-2021.
2. Pushkarenko Y. 2024. Synthetic datasets generation using diffusion models. DIGILIENCE 2024 AI and Critical Infrastructure Conference. Sofia, Bulgaria. (ISIJ) (Google Scholar).
3. PushkarenkoY. 2024. Remote sensing multilevel scene localization. 2024. IEEE DESSERT-2024. Athens, Greece.
4. Впровадження запропонованих моделей і методів підтримки прийняття рішень для критичних інфраструктур при виконанні проекту «Розробка моделей і методів підтримки прийняття рішень для критичних інфраструктур» (шифр: 2М-2022, термін виконання 2022-2023 роки) за запитом державної організації «Відділення цільової підготовки Київського національного університету імені Тараса Шевченка при Національній академії наук України». Науковий керівник роботи від КНУ імені Тараса Шевченка професор Заславський В.А. 2М-2022 від 18.04.2022 р.
5. Пушкаренко Ю. Моделі та методи локалізації сцен зображень об’єктів критичної інфраструктури на основі композитних нейронних мереж на міжнародному науково-технічному семінарі “Критичні комп’ютерні технології та системи” (КриКТехС-2025/2/196), Національний аерокосмічний університет ім. М.Є. Жуковського “ХАІ”, 25.02.2025.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

Влучність (Precision, або прогностична значущість позитивного результату) - частка релевантних зразків серед знайдених

ДЗ - дистанційне зондування

КІ - критична інфраструктура

Повнота (Recall, або чутливість, покриття, перекриття) - частка загального числа позитивних зразків, яку було дійсно знайдено

COCO – Common Objects in Context (Зведене об’єктне представлення)

CNN - Convolutional neural network (Згорткова нейронна мережа)

DNN - Deep Neural Network (Глибокі нейронні мережі)

DReAM - Dynamic Receptive Attention Module

ERF - Effective Receptive Field (Ефективне Рецептивне поле)

FCN - Fully Convolutional Network

GLCM - Матриця суміжності рівнів сірого Gray-Level Co-Occurrence Matrix

HSR - High Spatial Resolution (Зображення високої просторової роздільної здатності)

IoU - Intersection over Union

LSR - Low Spatial Resolution

LULC - Land Use/Land Cover

mAP - mean Average Precision

NMS - Non-Max Suppression

OBIA - Object Based Image Analysis

RJMCMC - Reversible-jump Markov chain Monte Carlo

ROI - Regions of Interests (регіони інтересів)

RPN - Region Proposal Network

SAR - Synthetic-aperture radar (Радар із синтетичною апертурою)

SWIN - Shifted Window Transformer

TRF - Theoretical Receptive Field (Теоретичне Рецептивне поле)

VHR - Very High Resolution (Зображення надвисокої просторової роздільної здатності)

ВСТУП

Актуальність теми дослідження. Зображення дистанційного зондування (ДЗ)[1] сьогодні є основним джерелом інформації для моніторингу природного середовища, міської інфраструктури, прогнозування катастроф та багатьох інших сфер, що потребують високої просторової та спектральної роздільної здатності. Світовий ринок технологій дистанційного зондування демонструє стійке зростання. За даними аналітичних компаній, обсяг глобального ринку дистанційного зондування у 2021 році оцінювався приблизно в 13,6 мільярдів доларів США і прогнозується, що до 2028 року він досягне понад 30 мільярдів доларів з середньорічним темпом зростання (CAGR) близько 11,8%. Збільшується попит на високоякісні дані та ефективні методи їх обробки, що обумовлено потребами різних галузей, включаючи сільське господарство, екологічний моніторинг, управління ресурсами та безпеку. Зростання обсягів супутникових та аерознімків, поява високороздільних радарних даних (SAR), а також нових стандартів якості зображень ставлять нові вимоги до обробки інформації, швидкості її аналізу та точності отриманих результатів. Однак традиційні підходи до аналізу зображень, зокрема класичні методи обробки та стандартні архітектури нейронних мереж, виявляються обмеженими у складних сценаріях, де об'єкти зображення можуть мати різні масштаби, неочікувану форму або складну текстуру, що заважає точному аналізу та виявленню пошкоджень та руйнувань.

Актуальність дослідження підсилюється зростаючою потребою у моніторингу критичної інфраструктури, такої як мости, дороги, будівлі, які можуть зазнати пошкоджень під впливом природних катастроф або техногенних аварій. За даними ООН, у період з 2000 по 2019 рік у світі сталося понад 7 000 великих природних катастроф, що майже вдвічі більше, ніж у попередні два десятиліття. Економічні втрати від цих подій оцінюються в 3 трильйони доларів США.

В Україні, внаслідок військових дій у 2022-2023 роках, масштаби руйнувань критичної інфраструктури досягли безпрецедентних рівнів. За оцінками Світового

[1] "Зображення" в цій дисертації означитимуть "зображення (знімки) дистанційного зондування".

банку та українського уряду, загальна сума збитків від руйнувань інфраструктури станом на березень 2023 року перевищила 411 мільярдів доларів США. Було зруйновано або пошкоджено понад 150 тисяч житлових будинків, близько 25 тисяч кілометрів доріг та понад 300 мостів і шляхопроводів. Також постраждали об'єкти енергетики, водопостачання, охорони здоров'я та освіти. Ці цифри підкреслюють критичну потребу в ефективних інструментах для моніторингу та оцінки стану інфраструктури з метою оперативного реагування та планування відновлювальних робіт.

Оцінка стану цих об'єктів та визначення їх пошкоджень потребує використання інноваційних технологій [1-3], здатних гнучко адаптуватися до різних особливостей сцени, масштабів об’єктів та їх станів як запропоновано (В. Харченко та інші). Зокрема, актуальними є методи, які можуть одночасно розпізнавати дрібні деталі та об'єкти на великих масштабах, забезпечуючи при цьому врахування глобального контексту сцени.

Таким чином, враховуючи масштаб руйнувань в Україні та тенденції розвитку ринку технологій дистанційного зондування, дослідження в області удосконалення нейронних мереж для обробки зображень ДЗ є вкрай актуальним. Розробка нових підходів, що відповідають сучасним викликам та вимогам в умовах невизначеності як досліджено (П. Кноповим та інші [2]), сприятиме підвищенню ефективності моніторингу та управління критичною інфраструктурою, а також забезпечить надійну основу для прийняття рішень у сфері безпеки та відновлення.

Мета роботи. Метою даної роботи є створення адаптивного методу та моделі локалізації сцен для зображень дистанційного зондування, здатної автоматично розпізнавати критичні інфраструктурні об'єкти та визначати їхній стан. Запропоновані моделі та методи повинні забезпечити високу точність сегментації та класифікації об'єктів на основі поєднання згорткових мереж та трансформерних архітектур з динамічним масштабуванням рецептивного поля, механізми запропоновані в роботі ставлять перед собою ціль локалізувати складні сцени де рецептивність несе композиційну структуру.

Наукові завдання дослідження. Для досягнення поставленої мети в рамках дисертації були визначені наступні наукові завдання:

- Провести аналіз існуючих методів та архітектур нейронних мереж для обробки зображень дистанційного зондування, виявити їхні переваги та недоліки при вирішенні задач локалізації та сегментації сцен.
- Розробити архітектуру композитної нейронної мережі (використовуючи принцип різнотипності) запропонованим (В. Заславським та ін.) [4-7], яка інтегрує можливості згорткових мереж та трансформерів (з залученням механізму уваги) для більш точного вилучення ознак на різних масштабах.
- Запровадити модуль динамічного рецептивного поля DReAM [8], який дозволяє адаптивно змінювати поле зору нейронної мережі в залежності від важливості частин зображення.
- Розробити методи для оптимізації обчислювальної ефективності запропонованої архітектури, зокрема зниження обчислювальних витрат для забезпечення роботи моделі у режимі близькому до реального часу.
- Провести експериментальні дослідження на великих наборах даних зображень (наприклад, xBD, DOTA), оцінити точність і продуктивність запропонованої моделі.

Об’єкт дослідження. Об'єктом дослідження є зображення дистанційного зондування, які включають сцени з критичними інфраструктурними об'єктами різних типів (мости, будівлі, дороги, електростанції) та їх пошкодження внаслідок природних і техногенних факторів, а також внаслідок військових дій.

Предмет дослідження. Предметом дослідження є методи локалізації та сегментації сцен (з онтологічним відображенням OSM) дистанційного зондування з використанням композитної архітектури нейронних мереж, яка дозволяє адаптивне масштабування рецептивного поля для обробки ознак різних масштабів.

Методи дослідження. У рамках дослідження використовувалися методи глибинного навчання, зокрема згорткові нейронні мережі для вилучення локальних ознак та SWIN-трансформери для захоплення глобальних взаємозв'язків у сценах зображень ДЗ. Було запроваджено динамічний підхід до масштабування

рецептивного поля (DReAM), який використовує self-attention для визначення релевантних регіонів на зображенні. Експериментальні дослідження проводилися на великих масивах даних ДЗ, зокрема набори даних xBD (для оцінки пошкоджень після катастроф) та DOTA (для класифікації різних типів інфраструктури).

Наукова новизна отриманих результатів. У даному дослідженні вперше розроблено інноваційну композитну архітектуру нейронної мережі, яка поєднує переваги згорткових нейронних мереж для виявлення локальних ознак з можливостями трансформерів для врахування глобального контексту. Унікальність підходу полягає в інтеграції механізму динамічного масштабування рецептивного поля (DReAM) з використанням self-attention, що дає змогу мережі адаптивно змінювати поле зору відповідно до важливості різних ділянок зображення. Такий підхід дозволяє моделі гнучко адаптуватися до різних масштабів, форм та характеристик об'єктів на зображенні, забезпечуючи точніший аналіз сцен дистанційного зондування. Запропонована архітектура дозволяє ефективніше виявляти пошкодження в критичних об'єктах інфраструктури, таких як мости та дороги, за рахунок точнішого врахування і локальних деталей, і глобального контексту сцени.

Наукове та практичне значення роботи. Розроблена у дисертації композитна архітектура нейронної мережі має значне наукове значення, оскільки забезпечує точну сегментацію та локалізацію об'єктів на зображеннях дистанційного зондування з використанням інноваційного підходу динамічного масштабування рецептивного поля. Такий підхід може бути застосований і до інших завдань комп'ютерного зору, де необхідно обробляти складні сцени з різнорозмірними ознаками. Запропоновані методи значно розширюють можливості глибинного навчання при аналізі неоднорідних та динамічних сценаріїв.

Практичне значення роботи полягає у потенціалі використання розробленої моделі для моніторингу та оцінки стану критичної інфраструктури, виявлення пошкоджень та наслідків природних і техногенних катастроф. Запропоновані підходи можуть бути інтегровані в автоматизовані системи прийняття рішень для державного

управління та безпеки, сприяючи оперативному виявленню ризиків, плануванню та проведенню відновлювальних заходів.

Науково-практичну важливість результатів дисертації підтверджує впровадження результатів досліджень у навчальний процес на факультеті комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка. Елементи розроблених методів та підходів включені до курсів з аналізу даних, зокрема, до навчальних дисциплін «Актуальні проблеми Data Mining» магістерської ОНП «Штучний інтелект», «Математичні методи штучного інтелекту» в рамках викладацької діяльності та «Корпоративні системи» магістерської ОНП «Бізнес інформатика» в рамках аспірантської асистентської практики. Це дозволяє студентам ознайомитися з сучасними методами обробки даних дистанційного зондування та застосовувати їх на практиці, підвищуючи ефективність навчального процесу та рівень підготовки фахівців.

Таким чином, результати дисертаційного дослідження мають як наукове значення для подальшого розвитку методів глибинного навчання у сфері аналізу дистанційного зондування, так і практичне значення, підтверджене використанням у реальних додатках і освітньому процесі.

Зв'язок роботи з науковими програмами, планами, темами, грантами.

Дисертаційна робота є складовою частиною наукових робіт, проведених в рамках наукових дослідницьких робіт Національної академії наук України, "Розробка моделей і методів підтримки прийняття рішень для критичних інфраструктур" за запитом державної організації "Відділення цільової підготовки Київського національного університету імені Тараса Шевченка при Національній академії наук України" н.к. Заславський В.А. до запиту № 2М-2022 від 18.04.2022 р.

Дослідження виконані в дисертаційній роботі по проблематиці безпеки критичних інфраструктур, а також результати експериментів є складовою міжнародного проєкту CPEA-LT-2016/10003 «Поглиблена спільна освітньо-наукова програма з управління ризиками в промисловості та сервісах в умовах глобальних економічних, технологічних та екологічних змін: розширена версія», який

виконувався під керівництвом проф. Заславського В.А. на факультеті компьютерних наук та кібернетики КНУ імені Тараса Шевченка з Норвезьким університетом науки та технологій (Тронхейм, Норвегія) в період 2017-2024 років.

Дисертація є самостійною науковою працею, у якій висвітлено оригінальні ідеї та напрацювання автора, що дозволили розв'язати поставлені завдання. У роботі подано теоретичні положення й висновки, сформульовані дисертантом особисто. Усі використані в дисертації концепції, ідеї чи гіпотези інших дослідників мають відповідні посилання і були залучені лише для обґрунтування й підкріплення авторських пропозицій.

Особистий внесок здобувача. За результатами досліджень, проведених у межах дисертаційної теми, було опубліковано 5 наукових праць, у тому числі у фахових виданнях і збірниках матеріалів конференцій. У тих роботах, які виконано у співавторстві, на захист виносяться виключно результати, отримані особисто здобувачем. Зокрема, Пушкаренко Ю.В. здійснив:

- аналіз наявних підходів до вирішення поставленої проблеми;
- формулювання та обґрунтування наукової гіпотези;
- розробку оригінальної моделі та компонентів локалізації пошкоджень;
- визначення параметрів моделей, налаштування та оптимізацію методів глибинного навчання;
- проведення експериментів, статистичну обробку даних і порівняльний аналіз результатів;
- формулювання загальних висновків та інтерпретацію отриманих результатів.

Співавторам у відповідних працях належали постановка окремих підзадач, частина консультацій і технічна підтримка в процесі підготовки даних та налаштування середовища для експериментів.

Структура та обсяг дисертації. Дисертація містить: анотацію, вступ, 4 розділи, висновки, список використаних джерел. Обсяг дисертації - 146 сторінок, основної частини - 126 сторінок. Робота містить 11 таблиць, 47 рисунків, список використаних джерел з 119 найменувань.

РОЗДІЛ 1. ОГЛЯД ТА ПОСТАНОВКА ЗАДАЧІ ЛОКАЛІЗАЦІЇ СЦЕН РУЙНУВАНЬ ОБ'ЄКТІВ КРИТИЧНИХ ІНФРАСТРУКТУР

У цьому розділі здійснено огляд наявних методів та підходів, що використовуються для виявлення та оцінки руйнувань на зображеннях дистанційного зондування (ДЗ). Насамперед, розглянуто значення проблеми локалізації пошкоджень у контексті критичної інфраструктури, зокрема мостів, доріг і промислових об'єктів. Висвітлено основні виклики при аналізі знімків великої роздільної здатності, де об'єкти можуть мати різні масштаби та форми, а також перебувати у складних умовах зйомки (тіні, задимлення, шум).

Далі розглядаються традиційні алгоритми (SIFT, фрагментарні методи, об'єктно-орієнтовані підходи) та сучасні глибинні мережі (CNN, трансформери), включно з обмеженнями фіксованого рецептивного поля. Особливу увагу приділено проблемі "локальний проти глобального контексту" для коректного визначення масштабу та поширення руйнувань мережі повинні ефективно поєднувати дрібні деталі і загальну структуру сцени.

Зрештою, сформульовано постановку задачі в забезпеченні автоматичного виявлення пошкоджених регіонів та відрізняти незначні тріщини від масштабних руйнувань. Саме потреба адаптивного рецептивного поля стає ключовим чинником, який мотивує подальшу розробку моделей з динамічним масштабуванням і підготовку до наступних розділів дисертації.

1.1 Постановка задачі локалізації сцен руйнувань об'єктів для знімків дистанційного зондування

Локалізація сцен пошкоджень є однією із ключових задач при аналізі зображень дистанційного зондування. Ця задача включає виявлення та точне визначення меж ділянок, які піддалися руйнуванню або пошкодженню внаслідок природних катастроф, техногенних аварій чи інших впливів. Локалізація сцен пошкоджень вимагає ідентифікації та сегментації об'єктів різних типів (наприклад, будівлі, мости, дороги), враховуючи при цьому їхні особливості та контекст у зображенні.

Оскільки пошкодження можуть мати різноманітний характер, масштаб та форму, постає завдання обробки сцен із різними рівнями деталізації та неоднорідністю. Для ефективної локалізації важливо враховувати:

- Локальні деталі об'єктів: дрібні пошкодження, тріщини або деформації.
- Глобальний контекст сцени: загальний стан місцевості, наявність довколишніх об'єктів та їх взаємозв'язки.

Використання глибинних нейронних мереж, зокрема згорткових нейронних мереж (CNN) і трансформерів, дозволяє розв'язувати цю задачу, але вимагає адаптивного підходу до рецептивного поля нейронів. Традиційні підходи використовують фіксоване рецептивне поле, що обмежує можливість гнучкої обробки неоднорідних сцен. Динамічне масштабування рецептивного поля дозволяє мережі адаптивно змінювати поле зору залежно від особливостей сцени та важливості різних її частин, що покращує точність локалізації пошкоджень.

Таким чином, задача локалізації сцен пошкоджень у зображеннях дистанційного зондування полягає у розробці методів і моделей глибинного навчання, здатних гнучко адаптуватися до різних масштабів, форм і контекстів об'єктів, забезпечуючи точне виявлення та визначення меж пошкоджених ділянок. Використання динамічно масштабованих рецептивних полів та механізмів уваги у трансформерах сприяє ефективному та надійному вирішенню цієї задачі, оскільки враховуються як локальні особливості об'єктів, так і глобальний контекст сцени.

Отже, умовно задачу можна сформулювати математично наступним чином.

Нехай задано зображення дистанційного зондування:

$$I \in \mathbb{R}^{!\times\#\times\$}, \tag{1.1}$$

де, H, W висота і ширина зображення відповідно, а C - кількість каналів. Задача локалізації сцен пошкоджень полягає у побудові функції:

$$f_{\%}\colon \mathbb{R}^{!\times\#\times\$} \to \mathbb{R}^{!\times\#}, \tag{1.2}$$

де, θ - параметри моделі, яка відображає кожен піксель вхідного зображення у вірогідність належності до пошкодженої ділянки або фону. Результатом є карта сегментації:

$$S = f_{\%}(I), \tag{1.3}$$

де, $S_{\&'} \in [0,1]$, ймовірність того, що піксель з координатами (i, j) належить до пошкодженої ділянки. Для розв'язання задачі необхідно мінімізувати функцію втрат:

$$\mathcal{L}8S, S_{()}9, \tag{1.4}$$

де, $S_{()}$ - карта сегментації з еталонними значеннями, яка задається вручну або іншими засобами.

Прикладом такої функції втрат може бути крос-ентропія, найбільш класичний спосіб мінімізувати втрати, надалі в роботі буде детально розглянуто інтуїцію використання функції використання саме Focal Loss замість кросентропії як функції втрат:

$$\mathcal{L}8S, S_{()}9 = -\frac{1}{HW}\sum_{\&=1}^{!}\sum_{'=1}^{\#}\left(S_{(),\&'} log S_{\&'} + 81 - S_{(),\&'}9 \log 81 - S_{\&'}9\right). \tag{1.5}$$

У згорткових нейронних мережах (CNN) розмір рецептивного поля (R) для кожного елемента активації звичайно фіксований і визначається конфігурацією шарів мережі. Проте для задачі локалізації пошкоджень з неоднорідними сценами необхідно адаптивно змінювати рецептивне поле. Введемо механізм динамічного масштабування рецептивного поля (Dynamic Receptive Field Attention Module, DReAM), який дозволяє мережі адаптувати поле зору залежно від змісту сцени. Нехай, $F \in \mathbb{R}^{!\times\#\times-}$ - проміжний тензор ознак, отриманий з попередніх шарів нейронної мережі, де H, W - просторові розміри, d - глибина ознак. Модуль DReAM використовує механізм уваги (Attention Mechanisms), щоб адаптивно зважувати внесок ознак з різних позицій зображення при обчисленні представлення для кожної позиції. Для кожної позиції (i, j) у просторі ознак F, ми визначаємо значення ознак $F^{\mathbb{C}}_{(\&,')}$ як зважену суму ознак з інших позицій, з використанням коефіцієнтів уваги:

$$F^{\mathbb{C}}_{(\&,')} = \sum_{(\&^{!},'^{!}) \in 3} \alpha_{(\&,'),(\&^{!},'^{!})} \cdot F_{(\&^{!},'^{!})}, \tag{1.6}$$

де, Ω - множина всіх позицій у просторі ознак (або обмежена локальним вікном), $\alpha_{(\&,')(\&^{!},'^{!})}$ - коефіцієнт уваги між позиціями (i, j) та (i^0, j^0).

Коефіцієнти уваги $\alpha_{(\&,')(\&^{!},'^{!})}$ обчислюються за допомогою механізму самоуваги (self-attention):

$$\alpha_{(\&,')(\&^{!},'^{!})} = \frac{\exp\{e_{(\&,')(\&^{!},'^{!})}\}}{\sum_{(4,5)\in 3}\exp\{e_{(\&,')(4,5)}\}}, \tag{1.7}$$

де, оцінка уваги $e_{(\&,')(\&^{!},'^{!})}$ визначається як:

$$e_{(\&,')(\&^{!},'^{!})} = \frac{\mathrm{Q}_{(\&,')} \cdot K^{6}_{(\&^{!},'^{!})}}{\sqrt{d}}, \tag{1.8}$$

де, $\mathrm{Q}_{(\&,')} = W_7 F_{(\&,')}$ - запит (query) для позиції (i, j),

та, $K_{(\&^{!},'^{!})} = W_8 F_{(\&^{!},'^{!})}$ - ключ (key) для позиції (i^0, j^0),

відповідно, W_7 та W_8 матриці параметрів які навчаються, з розмірністю ознак (для масштабування) - d.

Модуль DReAM дозволяє динамічно змінювати рецептивне поле, оскільки коефіцієнти уваги $\alpha_{(\&,')(\&^{!},'^{!})}$ можуть набувати значних значень як для близьких, так і для віддалених позицій, залежно від подібності їх ознак. Це означає, що мережа може "фокусуватися" на релевантних областях зображення незалежно від відстані.

Таким чином, математична постановка задачі включає розробку моделі $f_{\%}$ яка з використанням модуля уваги з динамічним рецептивним полем (DReAM) дозволяє адаптивно враховувати як локальні, так і глобальні особливості зображення для точного виявлення та сегментації пошкоджених ділянок у зображеннях дистанційного зондування. Використання такого методу також покращує узгодженість отриманих результатів із картографічними та супутниковими даними, що робить його ефективним інструментом для автоматизованого моніторингу та оцінки стану критичної інфраструктури після катастроф, аварій або виробничих руйнувань.

1.2 Необхідність локалізації сцен в задачах оцінки руйнувань критичних інфраструктур

Локалізація сцен в задачах оцінки руйнувань критичних інфраструктур є надзвичайно важливою, особливо в умовах сучасних викликів, таких як природні катастрофи, техногенні аварії та військові конфлікти. За даними ООН, кількість природних катастроф у світі значно зросла за останні десятиліття: з 4 212 випадків у 1980-1999 роках до 7 348 у 2000-2019 роках, що майже вдвічі більше [10]. Економічні втрати від цих катастроф оцінюються в 2,97 трильйона доларів США.

В Україні ситуація є особливо актуальною через збройний конфлікт, який триває з 2014 року та значно загострився у 2022 році. За даними Світового банку, станом на лютий 2023 року загальна сума прямих збитків критичної інфраструктури в Україні перевищила 135 мільярдів доларів США 2. Зруйновано або пошкоджено понад 150 000 житлових будинків, 25 000 км доріг, 3 170 закладів освіти, 1 200 закладів охорони здоров'я та численні об'єкти енергетичної, водної та транспортної інфраструктури. Загальні збитки пошкоджень понад 97 мільярдів доларів США (Рис.1.1).

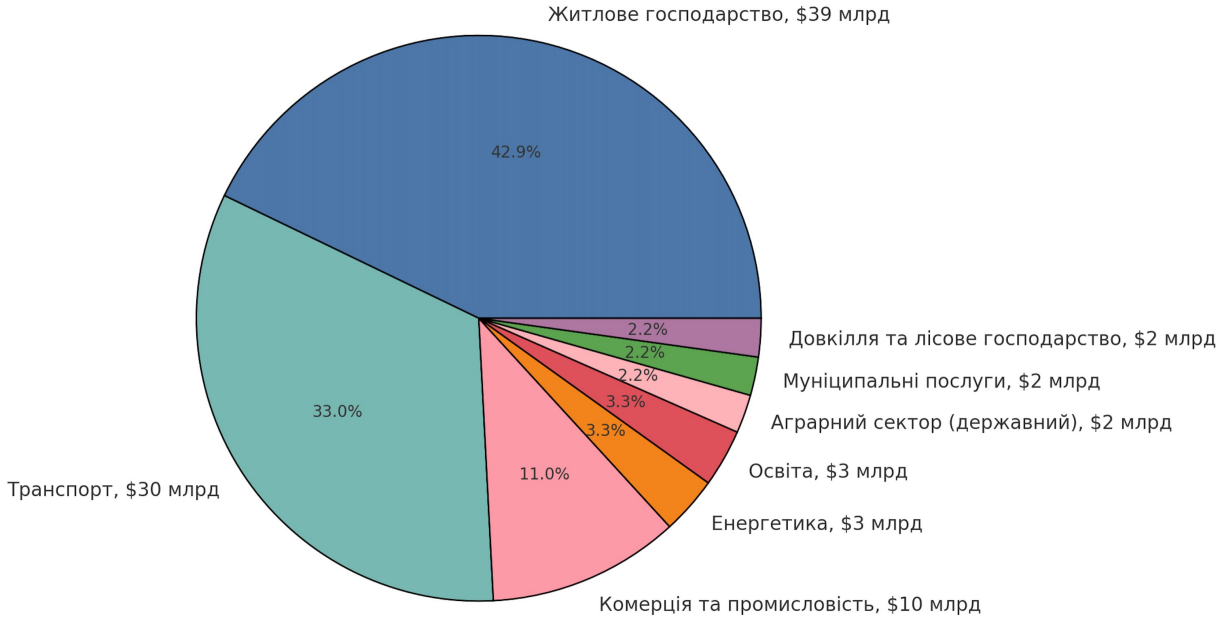


Рис. 1.1 Розподіл загальних збитків від пошкоджень в Україні станом на 1 червня 2022 року: 97 мільярдів доларів США.

Загальні збитки від незворотних втрат понад 252 мільярдів доларів США (Рис.1.2).

Основні причини необхідності локалізації сцен в задачах оцінки руйнувань:

1. Точна ідентифікація пошкоджених об'єктів. Дозволяє точно визначити географічне положення пошкоджених об'єктів. Це критично важливо для швидкого реагування та планування відновлювальних робіт. Наприклад, після землетрусу в Туреччині та Сирії у лютому 2023 року, використання супутникових зображень дозволило швидко виявити зруйновані будівлі та організувати ефективні рятувальні операції [12];

2. В реаліях України було впроваджено безліч проєктів які використовують майже всі спектри знімків дистанційного зондування, на базі таких зображень будуються системи надшвидкого реагування не тільки на ліквідацію пошкоджень, а і на пошук асоціативних похідних, знаходження аномалій, видобутку важливих даних. Географічна локалізація руйнувань та різного класу пошкоджень в цьому дослідженні відбувається за допомогою OSM (Open Street Map);

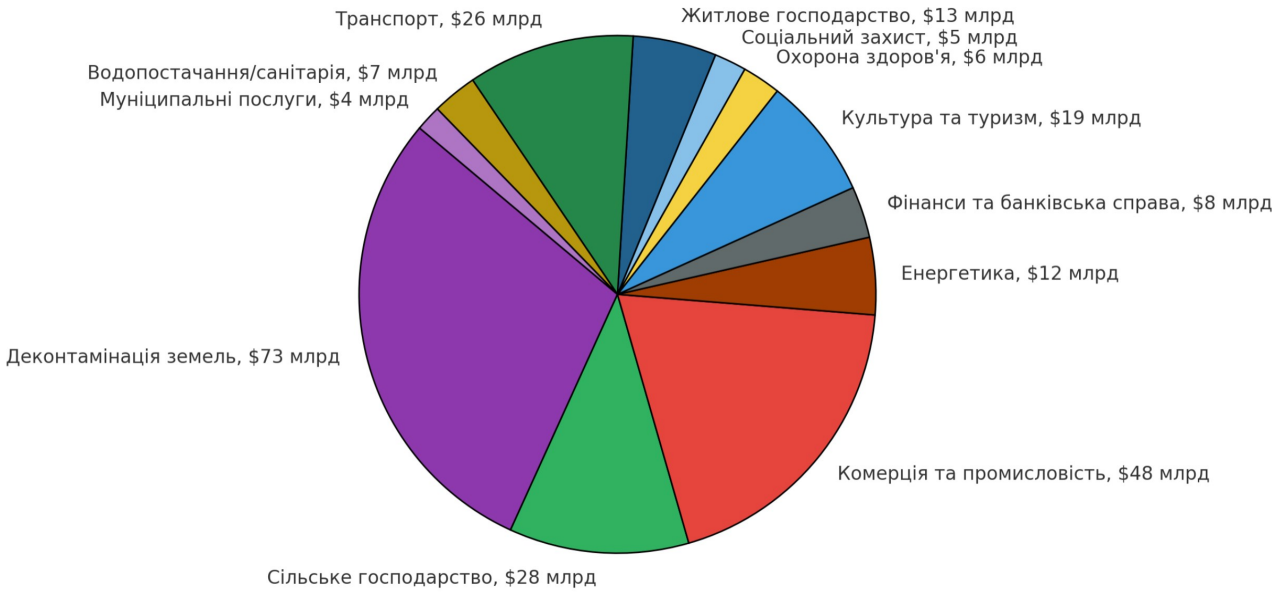


Рис. 1.2. Розподіл загальних втрат в Україні станом на 1 червня 2022 року: 252 мільярди доларів США.

3. Ефективне розподілення ресурсів. У надзвичайних ситуаціях ресурси обмежені. Локалізація пошкоджень дозволяє оптимізувати розподіл рятувальних команд, техніки та матеріалів. За оцінками Міжнародного комітету Червоного Хреста, ефективне використання ресурсів може підвищити швидкість реагування на 30% [13];

4. Швидкість реагування. Забезпечує оперативність у виявленні пошкоджень. За даними Європейського космічного агентства (ESA), використання супутникових зображень може скоротити час виявлення пошкоджень з декількох днів до декількох годин [14];

5. Акуратна оцінка масштабу руйнувань. Дозволяє кількісно оцінити масштаби пошкоджень. Наприклад, після урагану "Іда" в США у 2021 році, супутникові зображення допомогли оцінити, що понад 1,1 мільйона будинків залишилися без електроенергії [15];

6. Контекстуальний аналіз. Розуміння взаємозв'язків між пошкодженими об'єктами та оточенням допомагає прогнозувати вторинні наслідки, такі як перебої в постачанні електроенергії, води або зв'язку. Пошкодження можуть мати різні форми, масштаби та інтенсивності, що значно ускладнює розробку універсальних алгоритмів для їх автоматичного виявлення. Наприклад, руйнування будівель може відрізнятися від пошкоджень доріг або мостів як за візуальними характеристиками, так і за контекстуальними ознаками. Крім того, зображення ДЗ можуть бути отримані з різних сенсорів (оптичні, радарні, гіперспектральні) із різною роздільною здатністю та спектральними характеристиками, що додає додаткової складності уніфікації обробки даних. За даними статистичних даних [16] близько 65% досліджень у сфері виявлення пошкоджень зосереджені на обробці оптичних зображень, тоді як 35% використовують радарні та гіперспектральні дані. Цей розподіл демонструє потребу у розробці методів, здатних ефективно обробляти різні типи даних;

7. Складність сцен, масштабованість та автоматизація. Пошкодження можуть бути представлені як дрібними деталями (тріщини, деформації) так і великими структурними змінами (руйнування будівель, транспортних мереж). Наприклад, невеликі тріщини на дорогах може бути важко виявити на загальному

зображенні, тоді як масштабні руйнування можуть вимагати врахування глобального контексту сцени для їх точного визначення (Рис. 1.3). Сучасні алгоритми глибинного навчання дозволяють автоматизувати процес аналізу зображень дистанційного зондування, та адаптувати рецептивне поле під контекст сцен. Використання композитних нейронних мереж для сегментації пошкоджень може підвищити точність виявлення до 90% [17].

Таким чином, локалізація сцен в задачах оцінки руйнувань критичної інфраструктури є вкрай важливою та необхідною для забезпечення оперативного і точного виявлення пошкоджень, ефективного розподілу ресурсів та підтримки прийняття обґрунтованих рішень. Враховуючи масштаби руйнувань та необхідність швидкого реагування на надзвичайні ситуації, метою даної дисертації є розробка адаптивної та оптимізованої моделі, здатної ефективно вирішувати задачі локалізації сцен.

Рис. 1.3 HSR знімок MAXAR. Комплекс будівель міської лікарні в м. Маріуполь, демонструє різноманітність пошкоджень через глобальну сцену і локалізацією до рівня компонентних ознак з картами інтенсивності пошкоджень

Ця модель спрямована на подолання існуючих проблем виявлення візуальних ознак пошкоджень на зображеннях, використовуючи сучасні методи глибинного

навчання та динамічного масштабування рецептивного поля, що дозволить підвищити точність і швидкість аналізу, необхідні для ефективного моніторингу та відновлення критичної інфраструктури. Базовий фокус - це автоматизація процедури локального і глобального аналізу, де маленькі тріщини чи локальні деформації можуть бути виявлені нарівні з більшими структурними ушкодженнями. Для цього передбачена інтеграція механізмів адаптивного контролю поля зору, що підвищує стійкість і точність обробки, дозволяючи моделі гнучко реагувати на змінні характеристики кожної сцени. Зокрема, це важливо для аналізу невеликих сколів, тріщин або вибоїн на поверхні.

1.3 Проблеми виявлення візуальних ознак пошкоджень об’єктів на зображеннях

Таким чином, ознайомимось з поняттям “сцена” - у сфері дистанційного зондування сцена визначається як географічна область на зображенні, яка містить взаємопов'язані об'єкти та елементи місцевості, що характеризуються спільними семантичними ознаками (Рис. 1.4). Локалізація сцен пошкоджень передбачає виявлення та точне визначення меж ділянок, які зазнали руйнувань або змін внаслідок природних катастроф, техногенних аварій чи інших впливів.

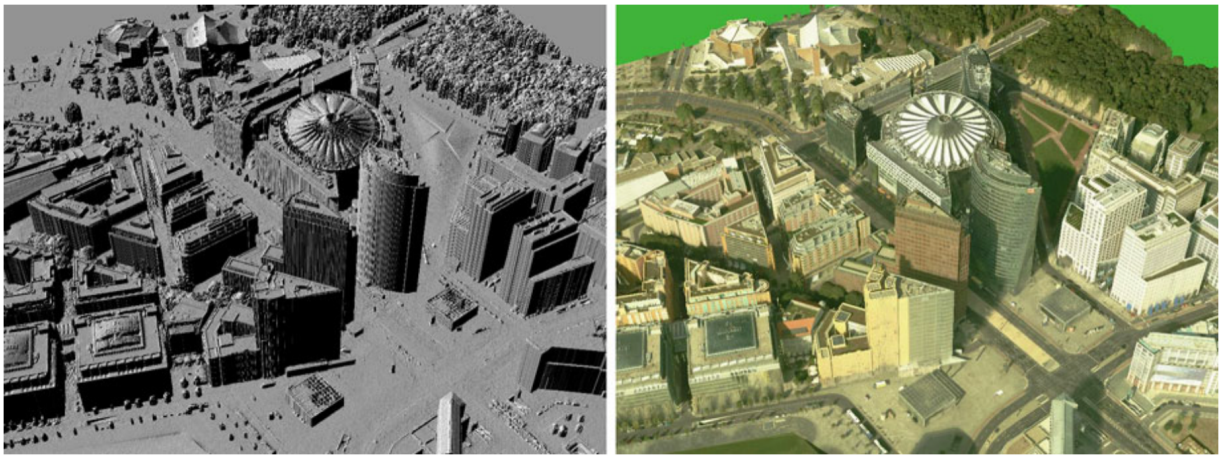


Рис. 1.4 Ліворуч: Сцена реконструйована зі знімку БПЛА. Праворуч: Накладені текстури. Сцена Sony Center (Берлін) демонструє складність ознак в глобальному контексті

Загальним підходом до виявлення пошкоджень критичної інфраструктури є прирівнювання проблеми до стандартного завдання сегментації зображень. Однак у більшості випадків такий підхід нехтує важливими елементами, такими як просторові та контекстуальні особливості зображень, а також специфіка об'єктів критичної інфраструктури. Застосування методів сегментації без врахування просторового розташування об'єктів та їх контексту може призвести до неправильної ідентифікації ушкоджень та їх некоректної локалізації. Наприклад, стандартні алгоритми можуть помилково віднести тіні або природні особливості місцевості до ушкоджень, оскільки вони не враховують спектральні та просторові характеристики, специфічні для зображень (див. Рис.1.5).

Щоб подолати ці проблеми, необхідно формулювати задачу локалізації та семантичної сегментації як процес оптимізації (з урахуванням динамічного рецептивного поля нейронної мережі), що враховує просторово-спектральні характеристики зображень дистанційного зондування. Це означає, що при аналізі

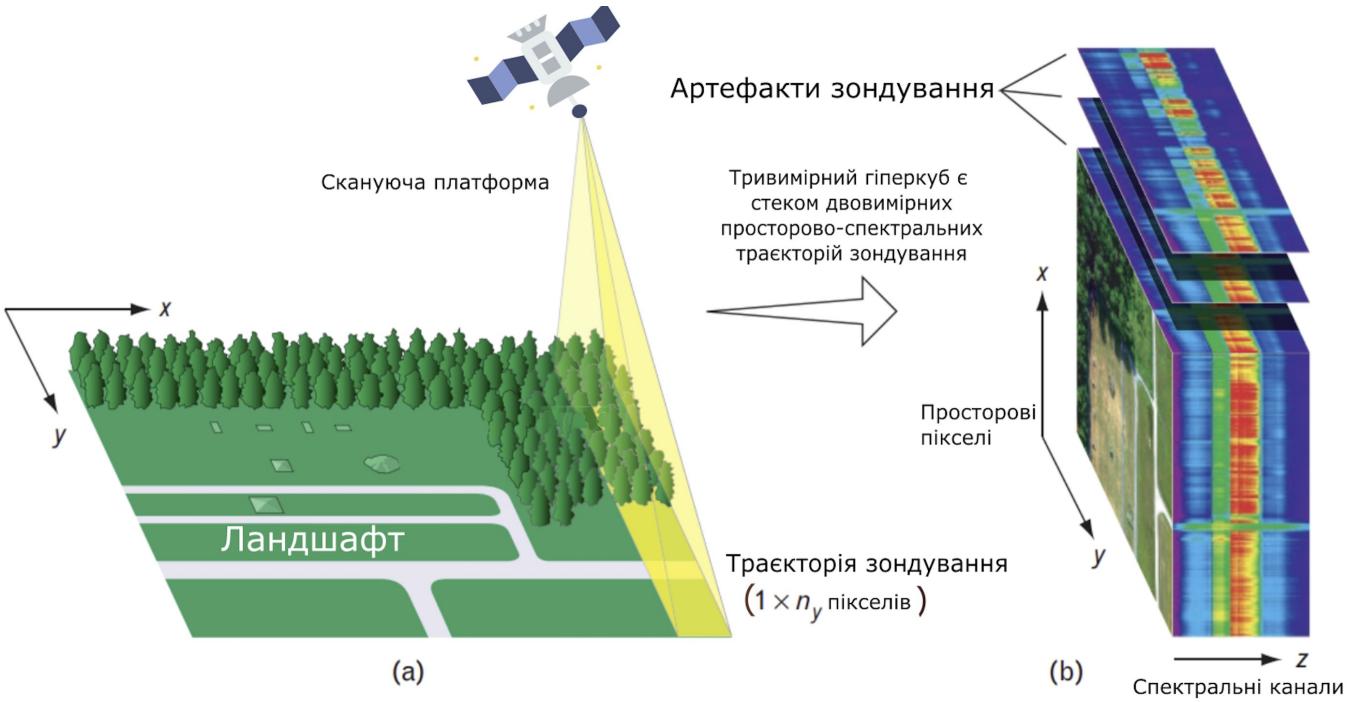


Рис. 1.5 Таксономічна будова багатоспектрального гіперкуба. a) Поверхня землі яка зондується супутниковою платформою. b) Будова артефактів дистанційного зондування (зображення з вбудованими просторовими ознаками).

необхідно використовувати ознаки, які відображають як спектральну інформацію (наприклад, відбиття в різних діапазонах довжин хвиль), так і просторові взаємозв'язки між пікселями або сегментами. Для цього обираємо оптимальні ознаки для аналізу на основі попарних відстаней між пікселями або сегментами зображення з використанням певної метрики. Після попередньої обробки створюємо вектор ознак для кожної області, використовуючи спектральні, текстурні та контекстуальні характеристики. Використовуючи метрику відстані, вимірюємо відстань між кожною парою областей зображення, у результаті чого отримуємо матрицю відстаней.

Інтуїція полягає в тому, що області з пошкодженнями матимуть низьке значення відстані між собою в просторі ознак, оскільки вони характеризуються схожими властивостями, такими як зміни текстури, спектральні аномалії або структурні порушення. Області без пошкоджень матимуть більші значення відстані від них. Правильний вибір ознак та використання динамічного рецептивного поля призводить до формування діагонально-блокової структури в матриці відстаней, де кожен блок відповідає певному класу (пошкоджені або непошкоджені області).

Однак для реальних зображень матриця відстаней може не бути ідеальною через шуми, перешкоди та складність сцен. Для подолання цих проблем використовуємо модуль уваги з динамічним рецептивним полем (Dynamic Receptive Field Attention Module, DReAM).

1.4 Проблеми рецептивних полів в нейронних мережах

У сучасних задачах локалізації та семантичної сегментації пошкоджень на зображеннях ДЗ особливу увагу приділяють здатності моделей ефективно обробляти складні та неоднорідні сцени. Одним із ключових аспектів цієї проблеми є розмір та гнучкість рецептивного поля нейронної мережі, яке визначає область зображення, що впливає на активацію певного нейрона. Глибокі мережі слід проектувати з рецептивним полем, яке охоплює всю релевантну область зображення, оскільки мережа не враховує області поза своїм рецептивним полем.

Різні архітектури мереж мають різні рецептивні поля. Традиційні згорткові нейронні мережі (CNN) використовують фіксовані рецептивні поля, розмір яких

визначається параметрами згорток та глибиною мережі, тобто зростання від шару до шару. Це призводить до ряду обмежень при аналізі зображень. Зокрема, у трансформерах рецептивне поле охоплює весь вхід (токени) вже після одного шару як продемонстровано Рис.1.6.

Однак всі оцінки рецептивних полів є лише теоретичними. У CNN фактичне рецептивне поле відрізняється від теоретичного. У CNN пікселі в центрі рецептивного поля мають великий вплив на вихід. У прямому проході центральні пікселі можуть передавати інформацію до виходу через багато різних шляхів, тоді як граничні пікселі мають дуже мало шляхів для передачі своїх значень, як показано на Рис.1.7. Саме тому важливо враховувати ефективне рецептивне поле (ERF), яке враховує реальний розподіл градієнтів та кількість проходів сигналу для кожного пікселя. Дослідження показують, що центральні частини рецептивного поля стають більш значущими, тоді як периферійні елементи можуть втрачати вплив на активації. У задачах локалізації пошкоджень це може зумовити «пропуски» дрібних деталей на межах зображення. Відповідно, при розробці моделей важливо застосовувати або багатомасштабні методи, або механізми динамічного розширення, щоб компенсувати обмеження ERF.

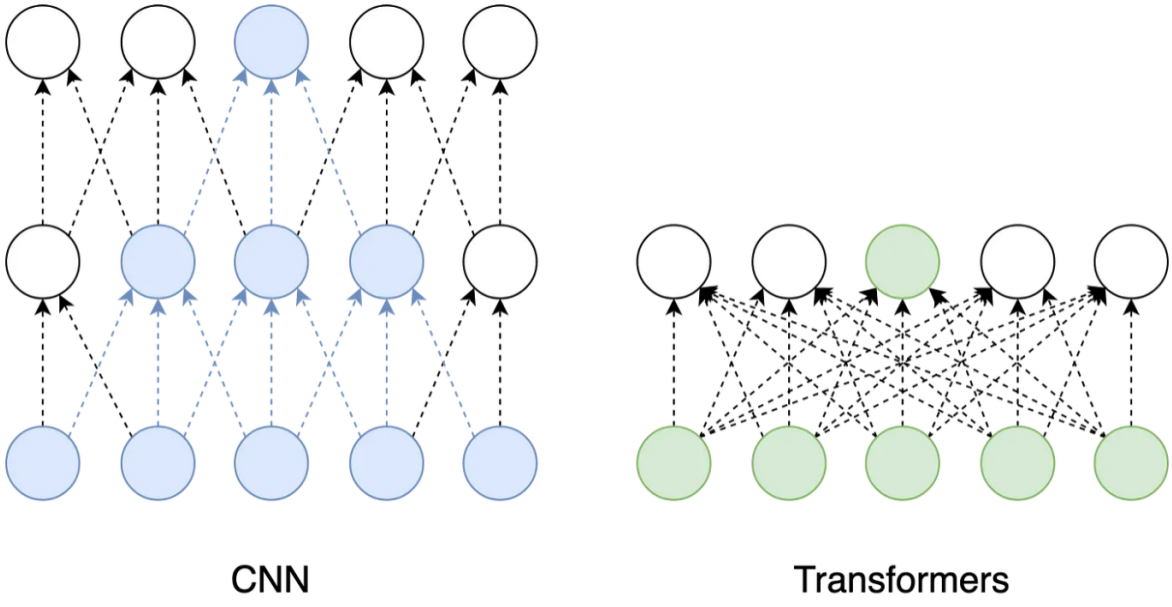


Рис.1.6 Ефективне рецептивне поле (ERF) на кожному із шарів CNN та мережі Трансформер

Тому у зворотному проході центральні пікселі мають набагато більшу величину градієнта від цього виходу. У цій статті [18] Luo та ін. емпірично оцінюють рецептивне поле у CNN та вводять термін ефективне рецептивне поле (ERF).

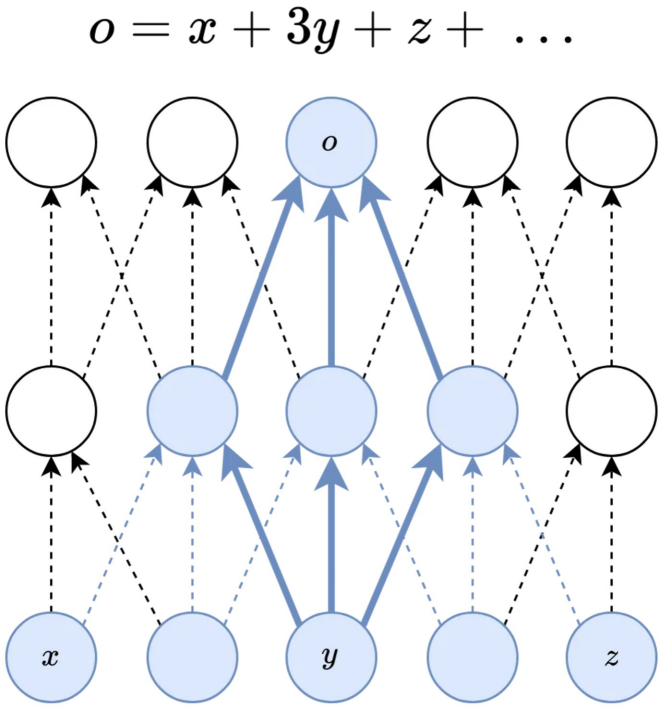


Рис. 1.7 У прямому проході центральні пікселі можуть передавати інформацію до виходу через багато різних шляхів. Тому у зворотному проході центральні пікселі мають набагато більшу величину градієнта

Натомість крайові пікселі можуть бути істотно недоотриманими через невелику кількість шляхів сигналу. Це особливо помітно, якщо об’єкти інтересу - дрібні й розташовані на периферії. Тоді загальна модель може втрачає контрастні контури та структури. Вирішенням стає залучення алгоритмів адаптивної згортки або трансформерних блоків з глобальними зв’язками, які посилюють вплив граничних точок і зменшують ризик пропусків важливих деталей.

У деяких сценаріях зображення можуть мати значний «баланс білого» чи атмосферний туман, що додатково ускладнює виявлення периферійних ознак. Тому сучасні підходи пропонують не лише адаптивні ядра згорток, а й багатокрокову обробку з різними масштабами, де початкова оцінка сцени відсікає непринципові ділянки. Таким чином, мережа зосереджується саме на сегментах, що можуть містити

важливі мікродеталі, зокрема тріщини чи сколи на межі зображення, забезпечуючи комплексне охоплення всієї доступної інформації.

Luo та ін. [18] показують, що ERF слідує гауссовому розподілу та займає лише частину повного теоретичного рецептивного поля (TRF). У проведених в статті експериментах оцінюють ERF з використанням випадково ініціалізованих мереж. Ці мережі ініціалізуються або рівномірно (всі одиниці), або випадково. Після ініціалізації ці мережі фіксуються для обчислення середнього ERF за 20 запусків (входів) та демонструють ідеальні гауссові форми для рівномірно та випадково ініціалізованих згорткових ядер без нелінійних активацій. Додавання нелінійності ReLU робить розподіл трохи менш гауссовим як показано на Рис. 1.8 [18]. ReLU дає точно нуль для половини своїх входів. Це означає, що мало шляхів від рецептивного поля досягають виходу.

Рис. 1.8 Порівняння впливу нелінійних активацій (ReLU, Tanh та Sigmoid) на ERF. ReLU робить розподіл менш гауссовим. Виходи ReLU точно нульові для половини своїх входів. Таким чином, легко отримати нульовий вихід для центрального пікселя на вихідній площині

Також, варто зазначити що інші підходи класичного розширення ERF є дилатована згортка, або даунсемплінг, підкреслюють важливість застосування в комп’ютерному зорі, але не вирішують проблему втрати пікселів з побічних нейронів див. Рис 1.9 [18]. На практиці для пом’якшення обмежень ReLU часто застосовують альтернативні функції активації на кшталт Leaky ReLU чи ELU, що зменшують кількість нульових виходів. Крім того, сучасні методи, такі як деформовані згортки, надають змогу частково коригувати форму рецептивного поля та підвищувати

чутливість мережі до деталей на межах. Утім, узгодження теоретичного та реального (ERF) охоплення залишається викликом у складних задачах, де одночасно важливі локальні особливості й глобальний контекст. Окрім цього, недоліки ReLU можуть також впливати на стійкість моделей до шуму та варіативності даних. Наприклад, у складних сценах з неоднорідним освітленням чи текстурою ділянки з нульовою активацією можуть створювати пропуски в сегментації. Впровадження механізмів нормалізації, таких як Batch Normalization або Layer Normalization, може частково компенсувати ці ефекти. Також альтернативні архітектури, як Swin Transformer, дозволяють гнучко адаптувати рецептивне поле в залежності від контексту. Комбінація CNN із трансформерами або гібридних блоків (наприклад, ConvNeXt) демонструє значний потенціал у вирішенні проблеми ERF, оскільки забезпечує краще узгодження між локальними і глобальними ознаками.

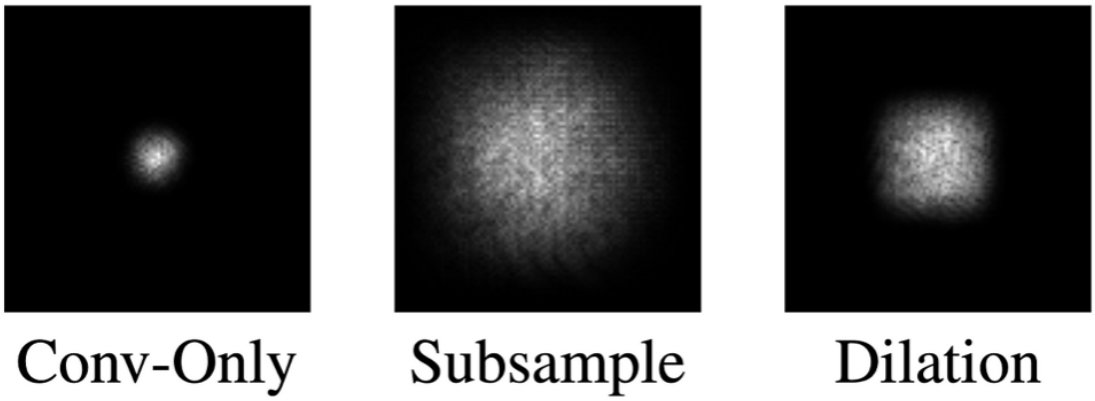


Рис. 1.9 Порівняння впливу даунсемплінг та дилатації на ERF. Обидва значно збільшують ERF

Отже, з нещодавніх досліджень можна зробити висновок, що збільшення рецептивного поля є досить серйозним викликом, для вирішення цієї проблеми використовують як даунсемплінг, дилатацію, та нещодавно альтернативами є деформовані згортки [19], деякі приклади ERF для різноманітних мереж можна побачити на рис. 1.10.

З іншого боку, занадто великі рецептивні поля можуть призвести до втрати дрібних деталей, що ускладнює виявлення невеликих пошкоджень, таких як тріщини чи незначні деформації [20]. Це створює дилему між необхідністю захоплення глобального контексту та збереженням локальної точності. Щодо формулювання

проблеми рецептивного поля для мережі Трансформер не існує практичних фактів та оцінок роботи ERF, розуміння що шар уваги може покривати весь вхідний сигнал (токени), є лише теоретичною інформацією.

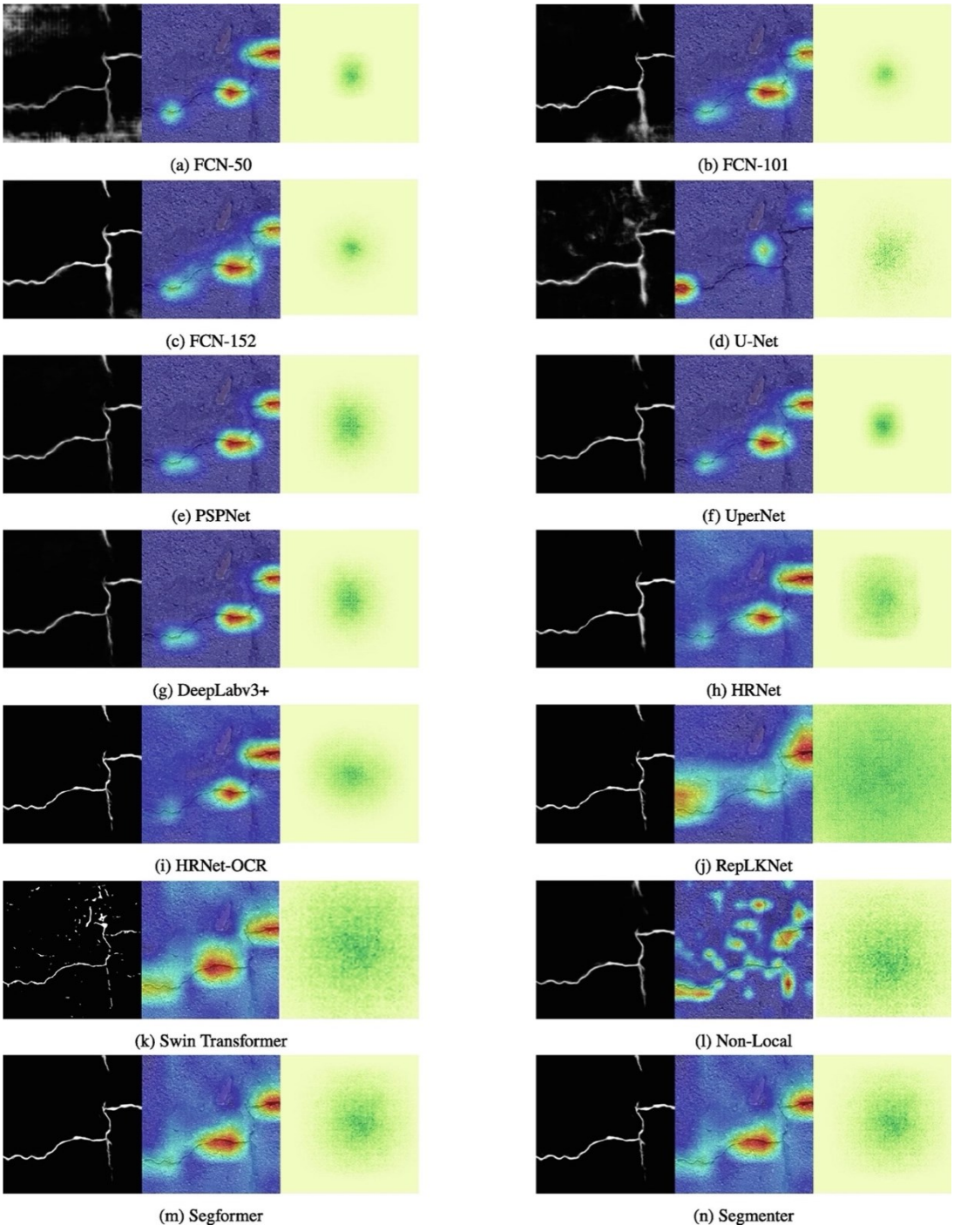


Рис. 1.10 Ефективне рецептивне поле для різноманітних моделей на прикладі пошкодження асфальтового покриття

У контексті локалізації сцен пошкоджень ця проблема набуває особливої гостроти. Пошкодження можуть мати різні масштаби та форми, а також бути

розташованими у складному оточенні з іншими об'єктами та елементами місцевості. Тому необхідно, щоб модель могла адаптивно змінювати рецептивне поле залежно від особливостей оброблюваної сцени.

Одним із перспективних підходів до вирішення цієї проблеми є впровадження динамічно змінюваного рецептивного поля за допомогою механізмів уваги. Зокрема, модулі уваги, такі як Self-Attention та мережа Трансформер, дозволяють моделі адаптивно фокусуватися на релевантних областях зображення, враховуючи як локальні особливості, так і глобальний контекст [21], [22].

Механізм уваги дає змогу моделі зважувати внесок різних частин зображення при обчисленні активацій нейронів. Це означає, що для кожної позиції на зображенні модель може визначити, які інші області є найбільш важливими для точного виявлення пошкоджень. Такий підхід дозволяє ефективно обробляти об'єкти різних масштабів та складності, що є характерним для зображень ДЗ [23].

Крім того, динамічне рецептивне поле сприяє підвищенню стійкості моделі до шумів та перешкод, які часто присутні на реальних зображеннях. Механізм уваги допомагає моделі ігнорувати нерелевантну інформацію та концентруватися на ключових ознаках пошкоджень [24].

У рамках дослідження особлива увага приділяється розробці та впровадженню методів, що дозволяють моделі адаптивно змінювати рецептивне поле. Це забезпечує більш точне та надійне виявлення пошкоджень критичної інфраструктури, що є вкрай важливим для оперативного реагування на надзвичайні ситуації та планування відновлювальних заходів. Застосування динамічно змінюваних рецептивних полів у поєднанні з механізмами уваги відкриває нові можливості у сфері аналізу зображень дистанційного зондування. Це не лише підвищує точність та ефективність моделей, але й дозволяє враховувати складні просторово-спектральні особливості сцен, які раніше були недоступні для традиційних підходів [25].

Таким чином, вирішення проблеми рецептивних полів є ключовим кроком у розробці адаптивних та оптимізованих моделей для локалізації та семантичної сегментації пошкоджень. Це сприятиме підвищенню ефективності моніторингу та

аналізу критичної інфраструктури, забезпечуючи більш надійну основу для прийняття рішень у сфері управління ризиками та відновлення.

1.5 Існуючі методи локалізації сцен руйнувань об’єктів та декомпозиції ознак в зображеннях

Протягом останнього десятиліття було запропоновано численні методи в пошуках рішення проблеми автоматизованого вилучення ознак[2], із зображень з високою просторовою роздільною здатністю (HSR), також в цьому дослідженні слід розрізняти різноманітні підходи у вилученні ознак, ідея локалізації полягає у комплексі дій які визначають місце пошкоджень як продемонстровано на рис. 1.11. У цьому розділі ці методи розглядаються приблизно в хронологічному порядку, щоб показати, як вони еволюціонували протягом історії дистанційного зондування, та не є прямо пов’язані з вилученням ознак для локалізації сцен ушкоджень так як сам процес локалізації є похідною вилучення ознак.

Розділ починається з опису методів, заснованих на фрагментах (патчах), в секції 1.4.1, які адаптовані до класичних методів на основі пікселів, що використовуються головним чином для вилучення ознак із зображень з низькою просторовою роздільною здатністю (LSR).

Методи на основі об'єктів, які значною мірою сформували знання спільноти дистанційного зондування, розглядаються в підрозділі 1.4.2.

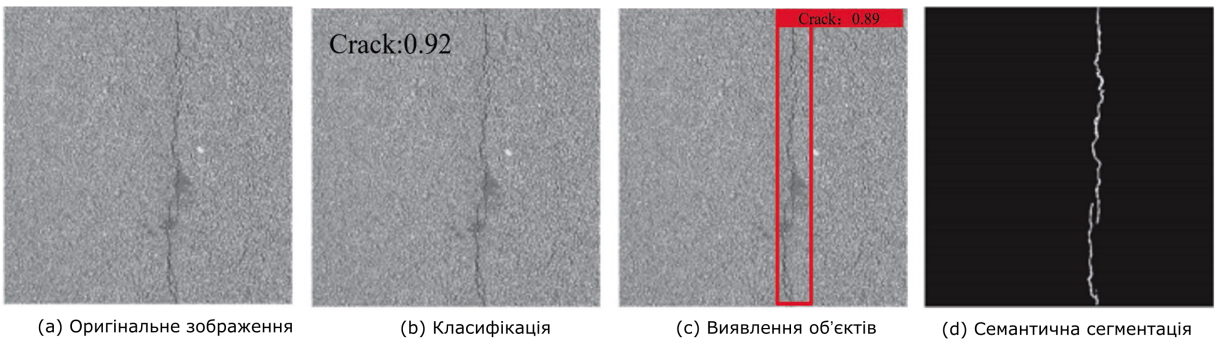


Рис. 1.11 Підходи вилучення ознак пошкоджень КІ

[2]Термінологія "вилучення ознак" - узагальнена, може використовуватися як "класифікація", "маркування зображень", або "семантична сегментація" в будь-якому іншому контексті. У цій дисертації ці терміни будуть використовуватися взаємозаміно, якщо не зазначено інше.

У підрозділі 1.4.3 представлені сучасні тенденції розвитку методів вилучення ознак, і обговорюються їхні обмеження. В останньому розділі особлива увага приділяється глибокому навчанню, розвиток якого пролив нове світло на цю складну проблему та породив нові проблеми адаптації складних моделей до захвату локальних сцен не втрачаючи глобального контексту, а саме проблем рецептивних полів як зазначено у попередньому розділі.

1.5.1 Фрагментарні методи виокремлення на основі патчів

Фрагментарні методи на основі патчів набувають коріння в класичних підходах на основі пікселів [26], які в основному використовуються для класифікації в картографуванні землекористування та рослинного покриву (LULC) з використанням саме LSR. В методах, заснованих на пікселях, навчений класифікатор використовує спектральні характеристики кожного окремого пікселя, щоб визначити мітку об'єкта відповідного пікселя. Ці спектральні характеристики працюють досить добре, щоб розрізняти широкі об'єкти, що представляють інтерес на знімках LSR, з досить високою точністю, такі як ліси, річки, міста та міські райони [27]. Тим не менш, вважається, що вони не підходять для характеристики більш дрібних класів об'єктів на зображеннях ДЗ, таких як дороги, будівлі, та інші інфраструктурні об’єкти [28]. При роздільній здатності вище 1 пікселя/м^2 спектральні значення різних типів об'єктів (наприклад, доріг та будівель, все, що зроблено з цементу) можуть бути дуже схожими. Це робить практично неможливим вилучення об'єктів із зображень ДЗ при використанні тільки спектральних характеристик.

Отже, для того щоб адаптувати ці методи до зображень з високою просторовою роздільною здатністю (HSR), важливо розробити методи вилучення текстурних, геометричних та контекстуальних ознак із сусідніх пікселів і використовувати ці ознаки для покращення вилучення об'єктів. З цією метою пропонуються методи, засновані на фрагментах (патчах), які на високому рівні виокремлюють ознаки з локального вікна, центрованого на кожному пікселі. У літературі з дистанційного зондування було розроблено різноманітні дескриптори ознак локального вікна. Наприклад, [29] використовували математичні морфологічні операції для

характеристики локальних особливостей пікселів при вивченні класифікації міського LULC з використанням зображень, отриманих за допомогою супутника IKONOS та Індійської системи дистанційного зондування Землі 1C. Інші роботи, які також використовують морфологічні операції, включають [30 - 31]. Крім того, одними із перших [32 - 34] використовували Gray-Level Co-Occurrence Matrix (GLCM) для опису особливостей текстури пікселів й показали підвищення точності класифікації супутникових зображень ДЗ в зображеннях місцевості в умовах інфраструктурних забудов. [35] використовували фільтр локального максимуму для вилучення розташування дерев і базальних площ з зображень ДЗ. [36] використовували вейвлети для визначення міських забудов за зображеннями зі супутника QuickBird. Нарешті, [37] використовували фільтр (вейвлет) Габора для виявлення будівель за зображеннями зі супутника IKONOS.

Ці дослідження послідовно продемонстрували, що використання ознак, отриманих з локальних вікон, допомагає підвищити точність вилучення об'єктів із зображень ДЗ. Однак ці функції все ще дуже локальні, оскільки розмір вікна, який вони використовували, зазвичай становить менше 9 пікселів. В межах такого невеликого вікна важко отримати просторову геометричну та контекстуальну інформацію, необхідну для розрізнення різних об'єктів та подолання несприятливих наслідків, викликаних тінями та оклюзіями. Наприклад, ширина дороги може досягати 30 пікселів на зображенні довжиною 0,5 м. Враховуючи 9-піксельне вікно, важко відрізнити піксель об’єкту інфраструктури від пікселя пошкодження на цьому об’єкті, а також дороги від мостових переходів. Ця проблема посилюється, коли дорога закрита деревами або тінями. Без достатньої геометричної та контекстуальної інформації практично неможливо відрізнити одне від іншого. В ідеалі розмір вікна повинен бути достатньо великим, щоб містити якомога більше контексту. Однак складність контексту зростає зі збільшенням розміру вікна. Ефективний та дієвий опис контексту та пошук дискримінативних ознак самі по собі стали проблемою.

Таким чином, успіх вищевказаних фрагментарних методів на основі патчів досить обмежений, й жоден з перерахованих, як було доведено, не працює ефективно для вилучення об'єктів із зображень ДЗ в складних зображеннях місцевості.

1.5.2 Об’єктно-орієнтовані методи виокремлення ознак

Враховуючи недоліки фрагментарних методів на основі патчів, були запропоновані методи OBIA (Об'єктно-орієнтований метод аналізу зображень) для вилучення об'єктів із зображень ДЗ [28], [38 - 40]. В об'єктно-орієнтованих методах окремі пікселі зображення спочатку групуються в кілька однорідних областей на основі їх подібності з точки зору спектрів й особливостей текстури. У літературі цей крок зазвичай називають сегментацією зображень. Потім з цих областей вилучається набір спектральних, текстурних, геометричних й контекстуальних елементів для характеристики їх атрибутів. Нарешті, класифікатор використовується для позначення кожного регіону унікальним класом об'єктів на основі вилучених особливостей регіону. Як можна побачити, головна відмінність між методами, заснованими на фрагментах, і методами, заснованими на об'єктах, полягає в тому, що останні спочатку агрегують окремі пікселі в однорідні регіони, а потім застосовують вилучення ознак і класифікацію до цих регіонів, замість того щоб працювати з окремими пікселями, як це робиться в методах, заснованих на фрагментах. Ці регіони вважаються потенційними об'єктами, й тому в літературі з дистанційного зондування цей тип методів називається "об'єктно-орієнтований аналіз зображень". Згодом, він був перейменований в “аналіз зображень на основі географічних об'єктів (GEOBIA)", щоб розрізняти поняття "об'єктно-орієнтований" в інших спільнотах [41].

Об'єктно-орієнтовані методи в даний час є найбільш широко використовуваними методами для завдання автоматизованого вилучення ознак із зображень ДЗ, відповідно, і локалізації та семантичної сегментації. Розробка цих методів розглядалася як прорив у літературі з дистанційного зондування та в значній мірі сформувала знання спільноти з дистанційного зондування, пов'язані з цим завданням. З моменту появи першого комерційного програмного забезпечення "eCognition" від компанії Trimble Inc., що реалізує об'єктно-орієнтовані методи [42], [43 - 44], багато дослідників вивчали використання цих методів для вилучення різних об'єктів із зображень ДЗ в різних зображеннях місцевості. Наприклад, [45] застосували об'єктно-орієнтований метод для детального вилучення рослинності на аерофотозображеннях ДЗ та емпірично продемонстрували, що об'єктно-орієнтовані

методи перевершують традиційні піксельні методи з точки зору точності вилучення ознак. Враховуючи успіх об'єктно-орієнтованих методів, деякі вчені в недавніх дослідженнях виступали за те, щоб розглядати GEOBIA як нову дисципліну [46 - 48].

У порівнянні з методами, заснованими на фрагментах (патчах), об'єктно-орієнтовані методи, що використовують методи сегментації зображень, вважаються набагато більш ефективними в завданні отримання різних геометричних й контекстуальних ознак розширеного радіусу дії. Загальновизнано, що ця перевага в значній мірі розширює можливості об'єктно-орієнтованих методів у вирішенні проблем, пов'язаних з великими внутрішньокласовими відмінностями, в порівнянні з методами на основі фрагментів [28], [46 - 48]. Однак, об'єктно-орієнтовані методи мають свої власні недоліки. Перш за все, точність об'єктно-орієнтованих методів значною мірою залежить від якості сегментації зображення. Але коли пікселі згруповані в регіони, для вимірювання однорідності використовуються тільки ознаки низького рівня (тобто спектр та текстура), без включення будь-яких об'єктів високого рівня (тобто ознак геометрії та контексту). Немає ніякої гарантії, що регіони, створені в результаті такого процесу, відповідають реальним об’єктам або частинам об’єктів через неоднозначність низькорівневих ознак, навіть при використанні сучасних класичних алгоритмів сегментації зображень [49 - 51]. Наприклад, будівлі критичної інфраструктури та допоміжні споруди які не несуть ключового фактору в прийнятті рішень можуть бути згруповані в один регіон через їх подібність з точки зору характеристик спектру. Крім того, ознаки, вилучені з неправильної сегментації, можуть не відображати властивості реальних об'єктів й можуть призвести до помилок класифікації [52 - 53]. Ці проблеми можуть стати ще більш серйозними, коли на зображенні представлені тіні та геометричні оклюзії. По-друге, навіть якщо згенерована область ідеально збігається з межею об'єкта, вилучення ознак для того, щоб відрізнити її від інших об'єктів, все ще залишається невирішеною проблемою. Наприклад, було доведено, що зазвичай використовувані дескриптори ознак, такі як середнє значення спектра, стандартне відхилення спектра, середнє значення текстури, ентропія текстури, розмір регіону інтересів, подовження, моменти зображення [26], досить добре працюють для характеристики особливостей природних зображеннях

місцевості, таких як трава, дерева, річки. Однак вони занадто примітивні, щоб розрізняти складні штучні об'єкти, такі як будівлі, об’єкти критичної інфраструктури або розрізнення мостів та доріг, тобто, визначення семантичних ознак на зображеннях ДЗ. Повідомлялося про незначні успіхи в вилученні складних штучних об'єктів з цими функціями, але якість таких алгоритмів занадто низька за умови відсутності ресурсів.

1.6 Огляд сучасних тенденцій виокремлення ознак

Для вирішення проблем, обговорюваних в підрозділах 1.4.1 та 1.4.2, останнім часом в спільноті дистанційного зондування були проведені численні дослідження.

Відмінні тенденції включають:

- Використання більш дискримінативних ознак;
- Перехід до класифікаторів з більшою складністю;
- Перехід на більш складні фреймворки.

Детальна інформація про ці три тенденції обговорюється в наступних підрозділах.

1.6.1 Дискримінативні ознаки на зображеннях

Оскільки досягти задовільних результатів класифікації за допомогою одних тільки зображень ДЗ дуже складно, природно подумати про додавання більшої кількості дискримінативних ознак за допомогою допоміжних даних. Одним з типових прикладів є використання цифрової моделі поверхні (DSM), яка або вилучається з стерео фотограмметрії, або безпосередньо з системи виявлення і визначення дальності світла в повітрі (LiDAR), один із трендових напрямків на сьогоднішній день є зображення із радарів з синтетичною апертурою (SAR). Велика кількість досліджень показала, що включення даних про висоту може значно підвищити точність вилучення об'єктів навіть за допомогою простих піксельних методів, включивши дані про стан металевих конструкцій (у випадку використання SAR) в будівлях критичної інфраструктури вносить величезний прорив в локалізації пошкоджень формуючи окремий каскад досліджень [54 - 56]. Однак, збір додаткових даних як радарне

сканування поверхні є занадто витратним щоб використовувати квадратні кілометри зображень для аналізу пошкоджень, зазвичай зйомка відбувається за запитом в конкретній географічній локації, тому розробка нових методів локалізації є надважливою для точного виявлення стану об’єктів. Наприклад, збір DSM для міста розміром $150\,км^9$ за допомогою бортового LIDAR (лазерний локатор) може коштувати до 12000 доларів США, що в чотири рази дорожче, ніж збір лише аерофотознімків. Тому використання таких допоміжних даних для вилучення ознак у великих масштабах є непомірно витратним як зазначено вище. Крім того, враховуючи, що людина, яка виконує аналіз зображень може дуже добре виконувати завдання вилучення об'єктів, використовуючи тільки зображення ДЗ, ознаки, вилучені з допоміжних даних, фактично надлишкові. Отже, з дослідницької точки зору було б цікаво опустити використання допоміжних ознак і в загальному випадку ідея data fusion в даній роботі в контексті джерел даних не розглядається.

У цьому напрямку робиться кілька зусиль, спрямованих на вилучення більш дискримінативних ознак тільки з джерела зображення ДЗ. Наприклад, [57] запропонували вилучати спектральні, текстурні та геометричні особливості з мульти-масштабної сегментації та продемонстрували, що ці особливості допомогли підвищити точність міського відображення LULC. [58] розробили морфологічний індекс будівель/тіней та припустили, що ця ознака корисна для вилучення ознак будівель з зображень ДЗ. [59] винайшли нову просторову ознаку, що називається індексом кореляції об'єктів, для підвищення точності вилучення об'єктів із зображень ДЗ. [60] представили новий метод вилучення особливостей геометрії пікселів та зазначили, що ці особливості були більш дискримінативними, ніж традиційні спектральні та GLCM ознаки. Інші комплексні ознаки, які широко використовуються у спільноті автоматичного аналізу відеоінформації, такі як масштабноінваріантне ознакове перетворення (SIFT) [61], гістограми орієнтованих градієнтів (HOG) [62], просторові піраміди [63] також можуть бути застосовані для підвищення вилучення ознак з зображення ДЗ, але історично не широко використовується спільнотою дистанційного зондування.

1.6.2 Традиційні методи класифікації об’єктів на зображеннях

Для розрізнення різноманітних ознак зі схожими спектрами та особливостями текстури, таких як дороги, мости або інші об’єкти критичної інфраструктури, потрібно знання геометрії об'єктів та контекст, таким чином огляд літератури підходить до наступного історичного етапу у виокремлені складних сцен. Це призводить до необхідності вивчення нелінійних меж прийняття рішень, які потребують класифікаторів з більшою складністю. Було проведено кілька досліджень, щоб дослідити потенціал покращення точності вилучення ознак шляхом використання передових класифікаторів. Наприклад, [64] оцінили результати, досягнуті за допомогою методу опорних векторів (SVM), максимальної правдоподібності, штучної нейронної мережі (ANN) та дерева прийняття рішень для відображення LULC з використанням супутникових зображень ДЗ. Було продемонстровано, що точність, досягнута за допомогою SVM, була відносно вищою, ніж у інших методів. [65] також детально розглянули використання SVM для вилучення об'єктів із зображень, отриманих за допомогою дистанційного зондування. [66] порівняли можливості оптимізованої штучної імунної мережі, штучної нейронної мережі, дерева прийняття рішень та типової імунної мережі з використанням даних QuickBird й припустили, що оптимізована штучна імунна мережа була більш ефективною, ніж інші вище згадані класифікатори. Наступним етапом вивчення стали дослідження ANN, штучні нейронні мережі відкрили новий простір для досліджень та разом з тим нові виклики.

1.6.3 Багаторівнева класифікація об’єктів на зображеннях

Виокремлення ознак є фундаментальною складовою ідентифікації та локалізації сцен та за своєю природою цей процес є проблемою "курки та яйця" - маючі контури об’єкта, розпізнавання стає простіше. Але щоб отримати правильний контур об'єкта, спочатку потрібно виконати розпізнавання, щоб визначити тип об'єкта. Однією з поширених технік вирішення цієї дилеми є виведення контуру та категорії об'єкта через процес “знизу вгору”. Методи, засновані на фрагментах, і методи, засновані на

об'єктах, обидва застосовують цю техніку. Ці методи починаються з окремих пікселів (або груп пікселів), які вказують на можливі місця розташування об'єктів. Потім вони витягують набір ознак з кожного місця, щоб визначити клас об'єкта для кожного пікселя (або групи пікселів) за допомогою класифікатора. Межа прийняття рішення класифікатора вивчається на основі ряду навчальних вибірок з використанням дискримінативних методів, таких як логістична регресія або метод опорних векторів. Принципи висхідного аналізу по суті є методами, керованими даними; вони широко використовуються для виокремлення об'єктів на знімках ДЗ завдяки своїй обчислювальній ефективності. Однак вони не використовують жодних попередніх знань про об'єкти - все вилучається з даних дискримінативним шляхом. В результаті за допомогою принципу висхідного аналізу не можна зробити висновок, які сигнали очікувалися, а лише визначити, що відрізняє типові сигнали в кожній категорії [67]. Це робить принципи висхідного аналізу більш чутливими до несподіваних зображень місцевості в даних тестування, наприклад, до несприятливих ефектів тіней та оклюзій.

Одним із можливих способів включення апріорних знань про об'єкт є процес “зверху вниз”. На відміну від методів “знизу вгору”, методи “зверху вниз” керуються моделями, кодують попередні знання про об'єкт у набір моделей об'єктів і локалізують об'єкти шляхом зіставлення цих моделей із зображенням. Одним з добре відомих прикладів цієї процедури є маркований точковий процес. [68] вперше застосували маркований точковий процес для вилучення дорожньої мережі з зображень ДЗ. У їхньому методі сегменти доріг моделювалися як набір міток, параметризованих їхньою орієнтацією, довжиною та шириною в полі Гіббса, яке сприяє формуванню з'єднаних лінійних мереж. Для знаходження оптимального співпадіння між моделями доріг на основі апріорних знань та зображенням був використаний алгоритм оборотного стрибка Марківського ланцюга Монте-Карло (RJMCMC). Цей метод пізніше був розширений для вилучення будівель, крон дерев та розливів морської нафти, отриманих із зображень дистанційного зондування, в наступних дослідженнях [69 - 75]. Ці дослідження послідовно показують, що попередні знання про об'єкт допомагають вирішувати проблеми, викликані

несприятливим впливом тіней та оклюзій. Однак ця перевага пов'язана з великими обчислювальними витратами. Оскільки об'єкти можуть відображатися на зображенні в різних масштабах та в різних орієнтаціях, ці методи повинні проходити через величезну кількість локацій, щоб знайти найкращу відповідність між моделями об'єктів та зображенням. Це може зайняти дуже багато часу в порівнянні з методами за принципом висхідного аналізу. Це обчислювальне навантаження серйозно вплинуло на широке застосування методів за принципом низхідного аналізу для вилучення об'єктів із зображень ДЗ.

Як видно, методи висхідного аналізу та методи низхідного аналізу дуже доповнюють один одного - методи висхідного аналізу надають можливі місця розташування, яких потребують методи низхідного аналізу, щоб уникнути виснажливого пошуку; низхідні методи пропонують попередні знання, яких прагнуть отримати висхідні методи, щоб впоратися з несподіваними зображеннями місцевості. Враховуючи недоліки їх використання поодинці, було б ідеально об'єднати їх. [76] провели дослідження в цьому напрямку. Вони запропонували ієрархічну та контекстуальну граматичну модель для аналізу аерофотознімків у складних зображеннях місцевості в умовах міста. У цій моделі вони використовували ряд детекторів ознак, щоб запропонувати можливі місця розташування об'єктів за принципом висхідного аналізу, та використовували модель ієрархічної граматики для перевірки виявлених об'єктів та прогнозування відсутніх об'єктів у низхідному аналізі за допомогою алгоритму RJMCMC. Їх експерименти показали, що процеси висхідного та низхідного налізу дійсно можуть спільно сприяти забезпеченню більш ефективного вилучення об'єктів з даних зображень ДЗ.

1.6.4 Методи глибинного навчання при дослідженні зображень

Розглянуті вище дослідження продемонстрували, що використання дискримінативних ознак, традиційних класифікаторів та багаторівневої класифікації може частково підвищити точність вилучення ознак із зображень ДЗ. Проте жоден з цих ранніх підходів не довів своєї ефективності та дієвості для високороздільних оптичних (VHR, HSR) та радіолокаційних (SAR) даних у повній мірі, особливо коли

йдеться про складні сцени з критичною інфраструктурою та оцінку її пошкоджень після надзвичайних ситуацій. Навіть невеликі тестові набори даних, які використовувалися у попередніх роботах, складалися з лічених зображень із розміром у кілька тисяч пікселів і стосувалися відносно простих місцевостей. Це значно спрощувало задачу в порівнянні зі сценаріями, де потрібно ідентифікувати об’єкти критичної інфраструктури (КІ) - мости, греблі, будівлі, енергетичні системи - та оцінювати їхній стан у складних урбаністичних або пост-катастрофічних умовах. Відсутність ефективних механізмів вилучення надскладних перехідних характеристик із зображень ДЗ стає фундаментальною проблемою, яка заважає досягти надійної продуктивності при аналізі неоднорідних наборів даних. Враховуючи, що семантична сегментація в таких умовах потребує не просто грубої класифікації, а точного визначення меж об’єктів та ступеня їх ушкодження, складність задачі лише зростає. Ця проблематика ускладняється також обмеженою кількістю доступних розмічених даних, або нерівномірності класів, про що згадувалося у [77]. Для пом’якшення цієї проблеми застосовувалися методи генерації синтетичних наборів даних [77], але вони не вирішують задачу повністю та будуть розглянуті далі.

Поява глибинного навчання (Deep Learning) як напряму в машинному навчанні відкрила нові можливості у вирішенні вищезгаданих проблем. Замість того, щоб інженери вручну визначали ознаки, глибинне навчання дозволяє моделям самостійно вивчати ієрархії дискримінативних ознак безпосередньо з даних [78 - 80]. Цей підхід продемонстрував суттєве підвищення точності в різноманітних задачах, таких як розпізнавання мовних сигналів [81], відеоактивностей [82], обробка природної мови [83], класифікація зображень [84], виявлення об'єктів та семантична сегментація [85]. Для дистанційного зондування [86] показав, що за допомогою глибинного навчання простий метод на основі патчів може досягти вражаючих результатів при вилученні об'єктів (наприклад, доріг, будівель) у складних міських сценаріях. Ключова перевага цього підходу полягає у використанні глибинних згорткових нейронних мереж (CNNs) для вилучення ознак з набагато більшого рецептивного поля та застосуванні графічних процесорів (GPU) для значного прискорення обчислень. Сьогодні, завдяки

успіхам компанії NVIDIA, зокрема з випуском спеціалізованих GPU для глибинного навчання (наприклад, архітектури NVIDIA A100 та H100), з'явилася можливість ефективної обробки масивних наборів даних у реальному часі, що було практично не можливо в епоху застосування тільки обчислювальних можливостей CPU. Це особливо важливо для оперативного аналізу після надзвичайних ситуацій, коли є потреба швидко оцінити стан інфраструктурних об'єктів та пріоритезувати відновлювальні роботи як зазначено вище в аналізі проблематики.

Сучасні підходи в області семантичної сегментації зображень ДЗ для локалізації сцен пошкоджень активно використовують розширені та деформовані згортки, а також механізми уваги та динамічні рецептивні поля, та піонерами в цій області стали [87], [88]. Ці підходи дозволяють адаптивно масштабувати рецептивне поле, враховувати нелінійні залежності та контекстуальні взаємозв'язки між різними частинами зображення. За рахунок цього досягається краща інтерпретація складних урбаністичних або техногенних сцен, у яких пошкодження можуть бути розподілені нерівномірно, мати різні розміри, форми та проявлятися у вигляді дрібних тріщин чи масштабних структурних руйнувань.

Важливим кроком уперед стало впровадження пірамідальних мереж [89], які забезпечують багаторівневу обробку ознак, синтезуючи інформацію з різних масштабів. У контексті динамічного рецептивного поля пірамідальні структури дозволяють моделі ефективніше інтегрувати локальні деталі та глобальний контекст. Це особливо актуально при виявленні ушкоджень критичної інфраструктури, де на одному знімку можуть одночасно бути присутні дрібні деталі (наприклад, тріщини на опорах мосту) та великі об'єкти (греблі чи дорожні мережі, пошкодження яких впливають на стан цілого регіону).

Крім того, перспективним напрямком є використання KAN (Kolmogorov-Arnold Networks) - мереж, які теоретично спираються на теорему Колмогорова-Арнольда. Вона стверджує, що будь-яку складну багатовимірну функцію можна розкласти у композицію простіших функцій однієї змінної [90]. У контексті дистанційного зондування та аналізу сцен руйнувань KAN можуть забезпечити гнучкіший підхід до моделювання складних функціональних залежностей між спектральними,

просторовими та текстурними ознаками, що особливо важливо для сегментації об’єктів та оцінки їх стану, такі роботи вперше відзначені як експериментальні але не продемонстрували значної швидкодії в умовах миттєвого прийняття рішень [91]. Поєднання KAN з глибинними мережами, динамічними рецептивними полями та пірамідальними структурами відкриває нові перспективи для побудови інваріантних та стійких до шумів моделей, здатних ефективно обробляти різноманітні типи даних (оптичні, SAR, LiDAR) та розв’язувати складні задачі, пов’язані з оцінкою пошкоджень.

Інтеграція KAN з методами комп’ютерного зору відкриває нові можливості для автоматизованого аналізу візуальних даних та виявлення об’єктів, одночасно підвищуючи стійкість до шумів. Разом із сучасними підходами до обробки зображень, KAN можуть забезпечити ефективну сегментацію та класифікацію у складних сценаріях. Проте ймовірні проблеми включають високу обчислювальну вартість розгортання таких мереж і необхідність великих навчальних вибірок для коректного узагальнення. Також виникають питання щодо узгодження різнорідних даних, наприклад різної роздільної здатності та динамічного діапазону.

Висновки до розділу

Таким чином, у першому розділі було проведено ґрунтовний аналіз наявних рішень і підходів до виявлення та оцінки руйнувань на зображеннях дистанційного зондування. Детально розглянуто традиційні методи, що ґрунтуються на локальних дескрипторах (SIFT, фільтрація патчів), та об’єктно-орієнтовані підходи, які застосовують сегментацію і аналіз регіонів зображення. Разом із тим висвітлено сучасні алгоритми глибинного навчання, які в останнє десятиріччя суттєво підвищили точність і швидкість вилучення ознак, зокрема за допомогою згорткових нейронних мереж (CNN) і трансформерів.

Основні проблемні аспекти, визначені в ході огляду, полягають у складності обробки зображень великої роздільної здатності, де об’єкти можуть бути різного масштабу та мати нестандартні форми, а також у недостатній гнучкості фіксованого рецептивного поля у типових CNN та трансформерах. Ці чинники ускладнюють точну

локалізацію пошкоджень, зокрема тріщин, зруйнованих елементів інфраструктури та ін.

На підставі виявлених викликів було поставлено задачу розробити методи й моделі, здатні адаптивно обробляти сцени та поєднувати локальні й глобальні контексти сцен для покращення точності і стійкості системи локалізації руйнувань.

РОЗДІЛ 2 АРХІТЕКТУРА ТА МОДЕЛЬ ЛОКАЛІЗАЦІЇ СЦЕН РУЙНУВАНЬ ОБ'ЄКТІВ КРИТИЧНИХ ІНФРАСТРУКТУР

У цьому розділі описується розробка інтегрованих методів та моделей для локалізації сцен руйнувань зображень ДЗ, зосередженої на аналізі пошкоджень КІ. Основну увагу приділено створенню гібридної моделі, яка поєднує згорткові мережі та трансформери (зокрема, Swin Transformer) для досягнення ефективного аналізу зображень із різними рівнями деталізації.

Запропонована модель використовує модуль DReAM, що забезпечує динамічне масштабування рецептивного поля і механізм уваги для адаптивного виділення ключових ознак на рівні локальних і глобальних контекстів. Інноваційний підхід дозволяє вирішувати задачі багатомасштабної обробки зображень та точного виявлення руйнувань у складних сценах з урахуванням проблематики описаної вище у попередніх розділах.

Розділ включає аналіз основних компонентів запропонованої моделі, їх архітектурні особливості та механізми функціонування. Особливу увагу приділено методам підвищення точності локалізації, зменшення обчислювальних витрат та інтеграції з реальними даними для практичного застосування в задачах моніторингу стану інфраструктури.

Цей розділ описує двофазну архітектуру моделі семантичної локалізації сцен руйнувань, створеної для оцінки пошкоджень критичної інфраструктури за допомогою знімків ДЗ. Інтегруючи підходи з робіт [92 - 94] запропоновано розширену концепцію, яка розглядає багаторівневий аналіз - від регіонального (R - regions), до активного об'єкта (A - assets) та його компонентів (C - components).

У Фазі 1 представлено загальну архітектуру, яка охоплює підготовку та попередній аналіз даних на регіональному рівні, забезпечуючи виявлення зон з потенційними руйнуваннями. На цьому етапі використовується багатомасштабний підхід з урахуванням різних рівнів деталізації, що дозволяє виділяти найбільш імовірні області для подальшого аналізу. Особливу роль відіграє гібридна модель, що поєднує обробку оптичних та багатоспектральних даних через адаптивну фільтрацію.

Фаза 2 реалізує композитну нейронну мережу, яка складається із запропонованого модуля DReAM. Цей модуль забезпечує динамічне масштабування рецептивного поля та використання механізму уваги для ефективної локалізації пошкоджень на рівні активів (A) та компонентів (C). DReAM, у поєднанні з трансформерами, дозволяє інтегрувати локальні особливості та глобальний контекст, зменшуючи шум і покращуючи точність.

У розділі також розглядаються аспекти узгодження результатів між Фазами 1 і 2, інтеграція з рішеннями для відновлення та можливість масштабування моделі до великих обсягів даних. Крім того, описані оптимізації для забезпечення швидкої та точної класифікації пошкоджень з урахуванням багатомасштабного аналізу.

Таким чином, запропонована модель є цілісним підходом, що охоплює всі етапи від виявлення до детального аналізу та підтримки прийняття рішень щодо відновлення інфраструктури. Її архітектура відповідає вимогам реальних сценаріїв, забезпечуючи високу продуктивність і адаптивність.

2.1 Модель та метод композиційної нейронної мережі локалізації сцен зображень ДЗ КІ

Концепція двофазного підходу є еволюцією принципу різнотипності запропонованого (В. Заславським та інші в [5]) як складова будови оптимізаційних математичних моделей і алгоритмів з врахуванням специфіки нейронних мереж та глибокого навчання, та забезпечує послідовний аналіз зображень ДЗ на різних рівнях деталізації об’єднана в єдиний пайплайн. Основна ідея моделі полягає у розподілі задач між Фазою 1, яка відповідає за попередню ідентифікацію зон з потенційними руйнуваннями, визначимо як регіон інтересів грубої оцінки або coarse level ROI, та Фазою 2, яка виконує детальний аналіз для точної локалізації сцен пошкоджень і оцінки їх масштабу, визначимо як деталізація високої точності або fine-grained details.

Перед формуванням загальної архітектури, необхідно визначити термінологію локального та глобального контексту яка зведена в табл. 2.1.

При розгляді зображень дистанційного зондування для оцінки стану КІ важливо враховувати обидва контексти - локальний, що допомагає деталізовано аналізувати

дрібні фрагменти (наприклад, тріщини на дорогах чи пошкодження опор), та глобальний, який відображає загальну картину руйнувань у ширшому масштабі (затоплені території, масштабні тріщини в дорожній мережі тощо). Поєднання цих підходів дає можливість формувати більш повну та комплексну модель стану об’єкта, враховуючи як мікро-, так і макрохарактеристики пошкоджень та їх взаємозв’язки.

Табл. 2.1 Зведена Табл. роз’яснення контекстів зображень ДЗ КІ

Аспект	Локальний контекст	Глобальний контекст
Фокус	Конкретизовані регіони або фрагменти (патчі)	Все зображення або великі обсяги регіонів
Призначення	Висока деталізація ознак	Розуміння онтологічного зв’язку та патернів
Методи	CNNs, традиційні фільтри, піраміди ознак	Трансформери, self-attention, global pooling
Приклади	Пошкодження доріг (тріщини), пошкодженні пілони	Пошкодження районів, патерни затоплення територій
Масштабування	Мікро (zoomed in)	Макро (zoomed out)

Отже, маючі узгоджену термінологію, сформуємо загальну картину потоку даних в моделі локалізації сцен пошкоджень КІ на рис. 2.1.

Загальний підхід побудований за наступним принципом:

1. Фаза 1 - Груба оцінка:

1.1. На цьому етапі використовується EfficientNet-B3 як бекбон для FPN (Feature Pyramid Network) виокремлення ознак з регіональних даних які отримуються з різних шарів, на виході отримується карта ознак (детальніше розглядається в наступних секціях);

1.2. ROI Proposal (RPN) визначає області інтересу (ROI), які агрегують в грубу ROI карту;

2. Фаза 2 - Детальна оцінка:

2.1. Рівень активів (A-scale): Семантична сегментація пошкоджень активів (наприклад, об’єкти інфраструктури, дороги);

2.2. Рівень компонентів (C-scale): Детальніша сегментація компонентів

активів (наприклад, дахи, стіни) при умові що зображення достатньої якості для такого аналізу (виходить за рамки даного дослідження);

2.3. Перевірка активів з даними OSM: Використовуються дані OpenStreetMap для валідації активів і доповнення метаданих (додатково може бути використана база UADamage, розробка додаткового пайплайну виходить за рамки даного дослідження);

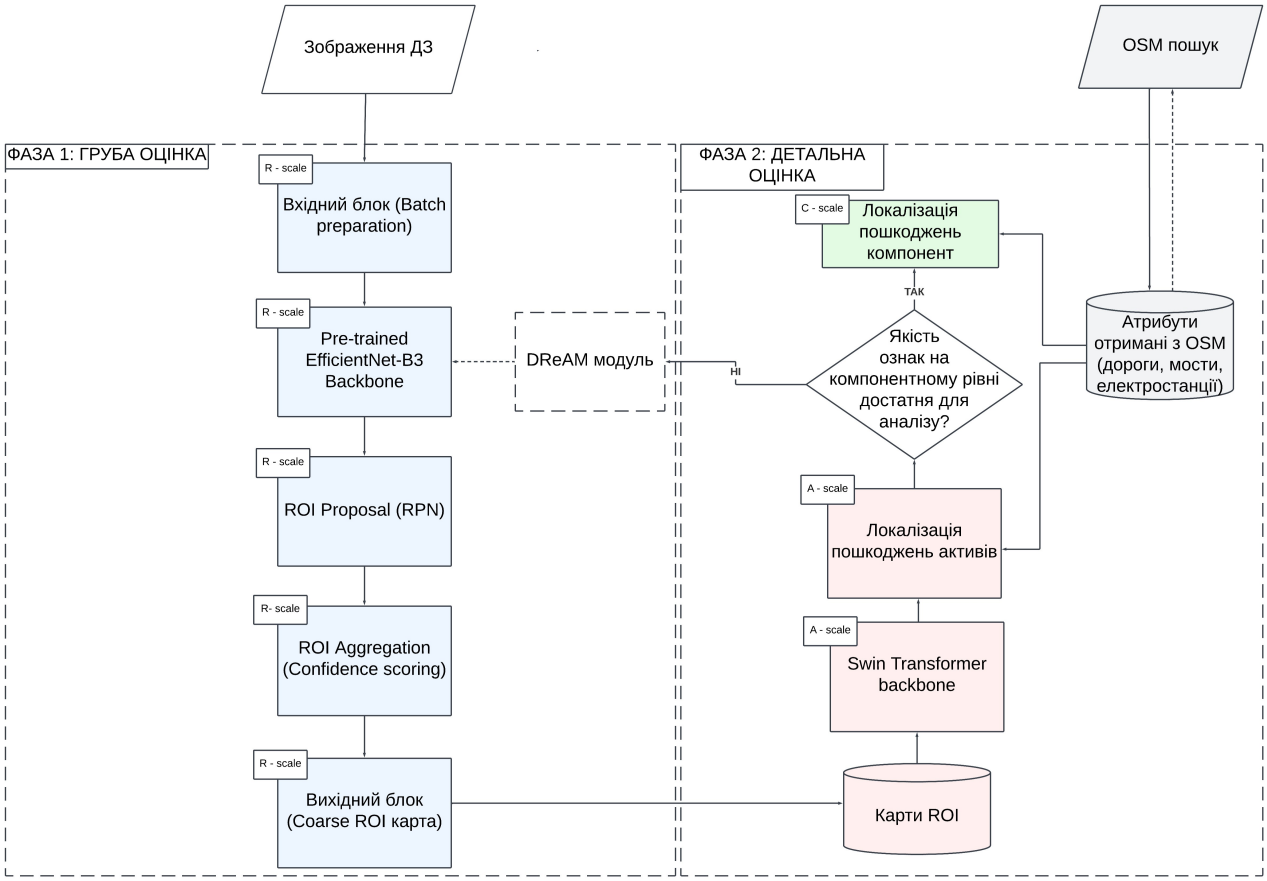


Рис.2.1 Загальна архітектура моделі локалізації сцен пошкоджень КІ

3. Сегментація (локалізація) пошкоджень:

3.1. Swin Transformer та DReAM модуль використовуються для витягу ознак та семантичної сегментації на різних масштабах (A-scale та C-scale).

3.2. Результат - фінальна карта ознак пошкоджень, яка забезпечує локалізацію пошкоджень на рівні сцени. Фактично ми отримаємо картку семантичної сегментації яка згодом зіставляється з OSM атрибутами для геолокалізації.

2.2 Багаторівнева декомпозиція ознак зображення та контексти сцен

Декомпозиція зображення у запропонованій архітектурі відбувається через два ключові етапи, які взаємопов'язані та побудовані навколо концепції композиції глобального і локального контекстів. У цьому розділі описується загальний підхід до декомпозиції, як зображення проходить ці етапи, і як математична модель пов'язана з Фазами 1 і 2. Для підвищення адаптивності моделі введено метод Dynamic Receptive Field Attention Module (DReAM), який виконує динамічне масштабування рецептивного поля. Це дозволяє моделювати локальні та глобальні контексти з урахуванням різних масштабів і просторових особливостей. DReAM інтегрується в процес між Фазою 1 і Фазою 2, забезпечуючи плавний перехід і покращену декомпозицію ознак. Це дозволяє узгодити формальні уявлення про глобальні і локальні контексти із практичними аспектами моделі.

Декомпозиція зображення у запропонованій моделі реалізована через послідовну взаємодію двох фаз:

1. Фаза 1: Фаза отримання регіонів інтересів (ROI). Основний акцент зроблений на ефективності по ресурсам та побудові регіонів ієрархічних представлень ознак з високою просторовою роздільною здатністю;
2. Фаза 2: Фаза семантичної реконструкції. Глобальні взаємозв'язки між регіонами зображення інтегруються за допомогою механізму уваги.

Для узгодження цих фаз використовується модуль DReAM, який забезпечує адаптивну інтеграцію локальних і глобальних ознак на основі динамічного масштабування рецептивного поля. Математично зображення $I(x, y)$ декомпонується на локальний $L(x, y)$ та глобальний $G(x, y)$ компоненти:

$$I(x, y) = L(x, y) + G(x, y), \tag{2.1}$$

Обидві компоненти проходять фазову обробку:

- У Фазі 1 акцент робиться на побудові локального контексту:

$$L^{(,)(:,;)} = \sum_{4 \in <(:,;)} w_4^{(,)} f8x + k_:, y + k_;9, \tag{2.2}$$

- DReAM використовується для динамічної адаптації рецептивного поля в момент проходу ознак у Фазі 2. Удосконалені локальні ознаки інтегруються:

$$L_{(=>?@A)(:,;)} = \sum_{B \in >} \alpha_{B(:,;)} \cdot \sum_{4 \in <"($,&)} w_4^B f8x + k_:, y + k_;9, \quad (2.3)$$

де, $R = \{r_, , r_9, \dots, r_C\}$ - набір масштабів рецептивного поля, а $\alpha_{B(:,;)}$ - ваги, що визначають важливість масштабу r у кожній точці.

У Фазі 2 фокус зміщується на інтеграцію глобального контексту через механізм уваги:

$$G^{(9)(:,;)} = \sum_{' \in \#(:,;)} \alpha_{\&'} F_{(=>?@A)(')}, \quad (2.4)$$

де, $W(x, y)$ - область уваги, а $\alpha_{\&'}$ - вагові коефіцієнти, що задаються механізмом самоорганізованої уваги.

У Фазі 1 застосовуються згорткові операції, що ієрархічно виділяють локальні ознаки. Глибина моделі дозволяє побудувати представлення на різних масштабах. Локальний контекст $L^{(,)}$ формалізується як багатошарова згортка:

$$L^{(,)(:,;)} = \sigma\mathrm{i} \sum_{B \in > 4 \in <"($,&)} w_4^{(,)} f8x + k_:, y + k_;9 + b^{(,)}\mathrm{k}, \quad (2.5)$$

де, $R = \{r_, , r_9, \dots, r_C\}$ - набір масштабів рецептивного поля, σ - функція активації.

Результат передається у DReAM, який вдосконалює локальні ознаки за рахунок адаптивного масштабування.

DReAM забезпечує динамічну декомпозицію шляхом об'єднання ознак із різних масштабів. Для кожної точки (x, y) підсумкова карта ознак обчислюється як:

$$F_{(=>?@A)}(x, y) = \sum_{B \in >} \alpha_B(x, y) \cdot F_B(x, y), \quad (2.6)$$

де, $\alpha_{B}(x,y) = \frac{DEF(G"(H(:,;)))}{\sum_{"!\in *} ?:J(G"!(H(:,;)))}$, та $F_{B(:,;)} = \sum_{4\in<"($,&)} w_4^B f8x + k_:, y + k_;9.$

У фазі 2 для семантичної реконструкції - глобальний контекст моделюється через механізм уваги:

$$G^{(9)}(i) = \sum_{' \in \#(\&)} \alpha_{\&'} F_{(=>?@A)}(j), \qquad (2.7)$$

де $\alpha_{\&'}$ визначається як: $\alpha_{\&'} = \frac{?:JK\frac{+,(-)\cdot/(0)1}{2_34}L}{\sum_{4\in5(-)}?:JK\frac{+,(-)\cdot/(4)1}{23_4}L}$.

2.3 Формування ROI та фаза грубої оцінки локалізації сцен

Значна частина супутникових та аерофотознімків, що надходять із сенсорів дистанційного зондування (ДЗ), страждає від появи характерних лінійних артефактів (смуг) див. Рис.2.2, зумовлених нерівномірною чутливістю детекторів, похибками калібрування або збоями в системах оптичного сканування.

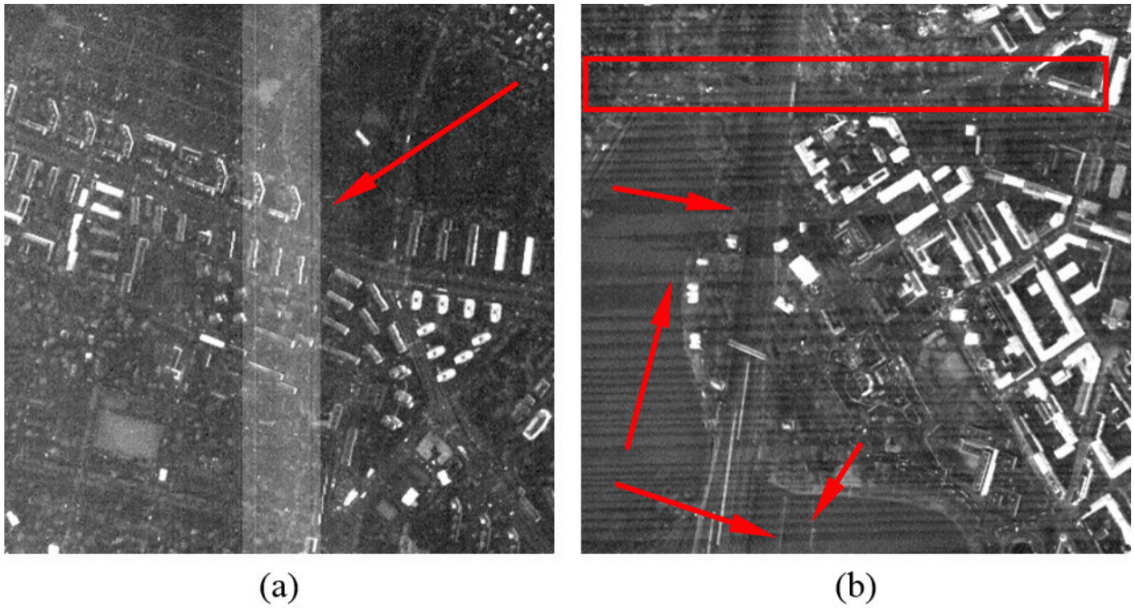


Рис. 2.2 Приклади супутникових зображень (a) смуги зашумлення вертикальної природи, (b) смуги зашумлення горизонтальної природи

Наявність цих смуг (stripes) може істотно погіршувати візуальну якість зображень, ускладнювати виявлення об’єктів і, відповідно, впливати на точність формування

регіонів інтересів (ROI) під час подальшого аналізу, наприклад, смуги можуть завадити локалізувати типові тріщини в пілонах або на шосе, трасах, мостах з асфальтовим покриття, таким чином високий відсоток FP (false positive) гарантовано.

Для усунення таких артефактів у досліджуваній моделі пропонується етап десмугування (destriping), що базується на сучасних варіаційних методах обробки зображень.

Згідно з роботою “Destriping of Remote Sensing Images by an Optimized Variational Model” авторів Fei Yan et.al [95], відновлення зображення відбувається за допомогою розв’язання оптимізаційної задачі, де цільовий функціонал містить два основні компоненти (рис.2.3 [95]):

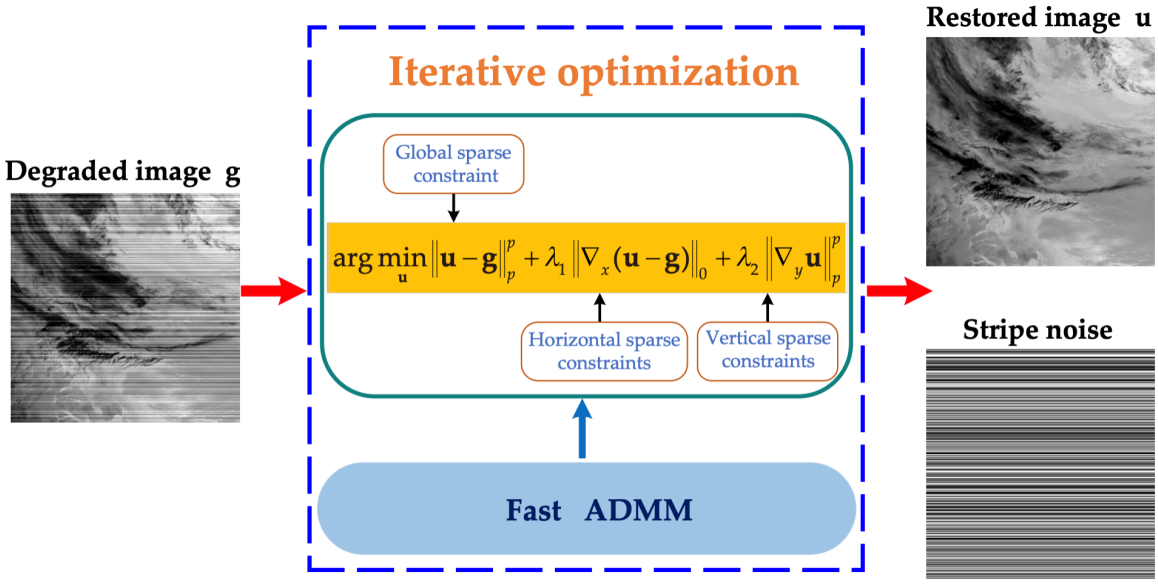


Рис.2.3 Загальний вигляд запропонованого методу десмугування

- Функція узгодженості з даними (data fidelity term) - контролює відхилення результату від вихідного зображення, аби зберегти важливі деталі (краї, текстурні фрагменти тощо);
- Регуляризаційний член (regularization term) - мінімізує вплив артефактів і шумів, забезпечує згладження або, навпаки, збереження чітких контурів. Застосовують різноманітні варіанти регуляризації, зокрема TV (Total Variation), TGV (Total Generalized Variation) тощо.

У цій моделі десмугування можна розглядати як задачу розділення (decomposition) на дві складові:

- Структурна складова, що відповідає глобальній інтенсивності зображення та містить корисну інформацію про об’єкти сцени;
- Шумова/артефактна складова, яка відповідає артефактам (смугам), викривленням та регулярним шумам сенсора.

Після десмугування зображень на їх базі суттєво покращується візуальна якість та видимість об’єктів, що має прямий вплив на точність формування регіонів інтересів (ROI). Зокрема, у варіаційних моделях, використовується спеціально підібраний регуляризатор, який зводить до мінімуму «східчастий» або «лінійний» характер відновлених інтенсивностей по вертикалі/горизонталі, властивий багатьом класичним методам фільтрації. Отримані після десмугування зображення забезпечують більш однорідне фонове поле інтенсивностей, з якого легше виокремлювати глобальні структури (дороги, мости, будівлі). Це особливо важливо на етапі визначення ROI в автоматизованій моделі локалізації сцен руйнувань. Усунення смуг дозволяє алгоритму формувати пропозиції регіонів (Region Proposal) на основі достовірнішої інформації про градієнти, контури та текстурні особливості, не «сплутані» з артефактами, зумовленими недоліками сенсора. Таким чином, десмугування виступає першим кроком у забезпеченні коректності формування глобальних ознак, що у подальшому (в наступних розділах) будуть інтегровані з локальними деталями через механізм динамічного масштабування рецептивного поля (DReAM) та багатошаровий підхід до семантичної сегментації. Це дає змогу моделі більш надійно орієнтуватися в складних сценах і мінімізувати хибні спрацювання, пов’язані з “липовими” переходами яскравості.

Отже, для виокремлення ознак ROI обрано EfficientNet-B3 для першої фази, оскільки вона забезпечує оптимальний баланс між точністю та швидкістю виявлення регіонів інтересу. При відносно невеликій кількості параметрів (12 млн.) модель досягає понад 81% Top-1 точності на ImageNet, що робить її достатньо потужною для грубого відбору кандидатних ділянок без зайвих витрат ресурсів (див. Рис. 2.4). Така «золота середина» підходить саме для початкової оцінки сцени, де важлива

ефективність і стійкість до перенавчання, а детальне опрацювання переноситься на подальший етап (Фаза 2). Основна задача фази 1 це пропозиція регіонів інтересів тому ROI head не виконує свою класичну функцію, як наприклад, це працює і Faster R-CNN - замість цього Фаза 2 вирішує питання семантичної сегментації, а згодом і локалізації.

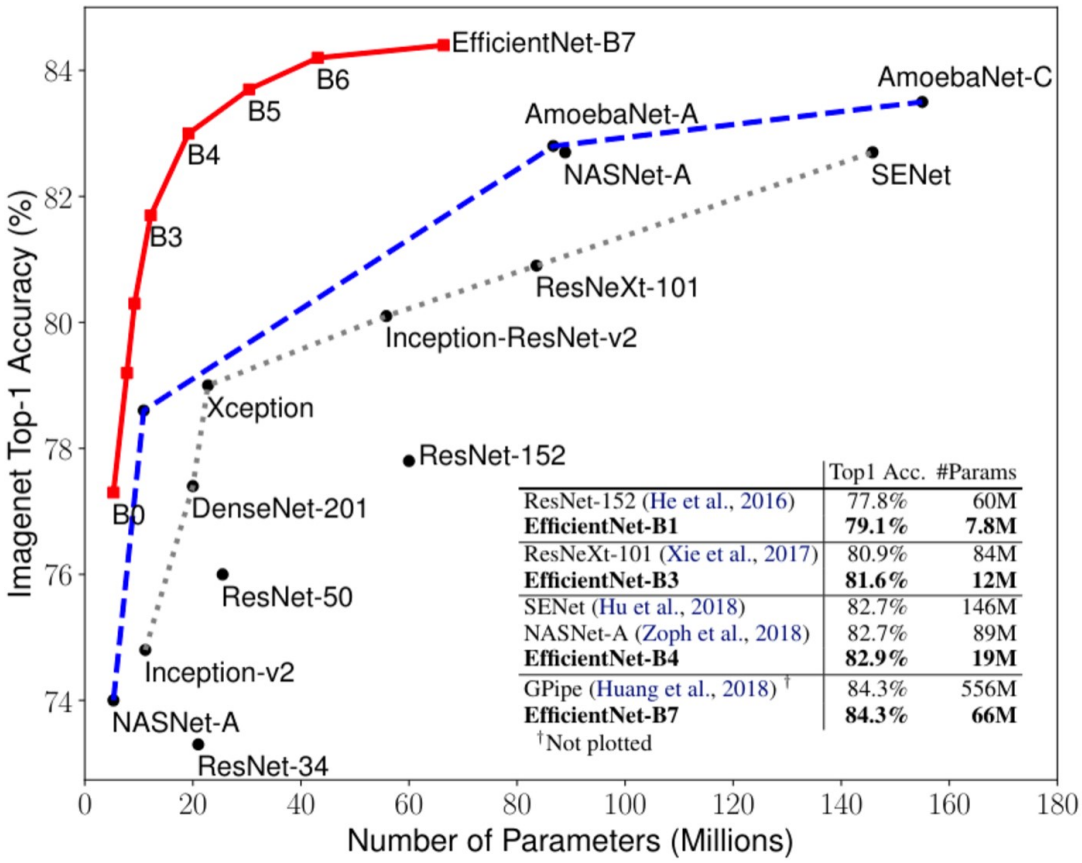


Рис. 2.4 EfficientNet statistic. Графік порівняння залежності точності моделей від числа параметрів

Варто відмітити, що у цьому дослідженні не застосовується традиційний підхід де в RPN використовується ResNet, зазвичай, ResNet використовується тому що надає можливість формувати карти ознак на різних рівнях через Feature Pyramid Network і саме багаторівневий підхід реалізується через двохфазну архітектуру таким чином надаючи можливість інтегрувати 2 основні концепції:

- Динамічне (адаптивне) рецептивне поле (DReAM);
- Мережу Трансформер для реєстрації залежностей між гетерогенними ознаками (використовуючи механізми уваги).

Таким чином, посилаючись на діаграму загальної архітектури моделі локалізації сцен пошкоджень КІ на рис. 2.1 основна задача це сформувати карти ROI, для цього необхідно детально узгодити основні компоненти:

- Backbone Network (як зазначено вище EfficientNet - B3), оскільки на цьому етапі не важлива детальна оцінка - використано менш ресурсно-витратну мережу, але з високою точністю виокремлення (див. рис.2.4);
- FPN для виокремлення ознак з різних шарів EfficientNet - B3;
- RPN (Region Proposal Network);
- ROI Head - вхід Фази 2.

Типова мережа EfficientNet - B3 запропонована в [96] (див. рис. 2.5), виконує функцію бекбону для FPN, та є базовою в фазі 1, потрібно звернути увагу що використовується Swish-функція як активація (див. рис.2.6 [98]).

Отже, підсумуємо загальний підхід. Для грубого відбору ROI обрано мережу EfficientNet-B3 (рис. 2.5) через вдалий компроміс швидкодії і точності:

- ~ 12 млн. параметрів;
- ~ 81% Top-1 на ImageNet (рис. 2.4).

Така "золота середина" дає змогу:

1. Швидко обробляти великі знімки (дистанційного зондування);
2. Уникати перенавчання;
3. Отримувати достатньо якісні ознаки (feature maps) для RPN (Region Proposal Network)[3].

Щоб охопити об’єкти різних масштабів, у Фазі 1 інтегруємо FPN на основі проміжних шарів EfficientNet-B3. Загальна концепція (рис. 2.7), та Ідея полягає у тому, щоб:

1. Витягати проміжні ознаки з кількох рівнів глибини ($\frac{'}{M}$, $\frac{'}{N}$, $\frac{'}{,0}$, … , від вхідної роздільної здатності;
2. "Зшивати" їх через механізм top-down + lateral convolutions;

[3]У типових реалізаціях наприклад Faster R-CNN зазвичай використовують ResNet (50/101) + FPN. Тут ResNet замінено на EfficientNet-B3, зберігаючи можливість побудови FPN.

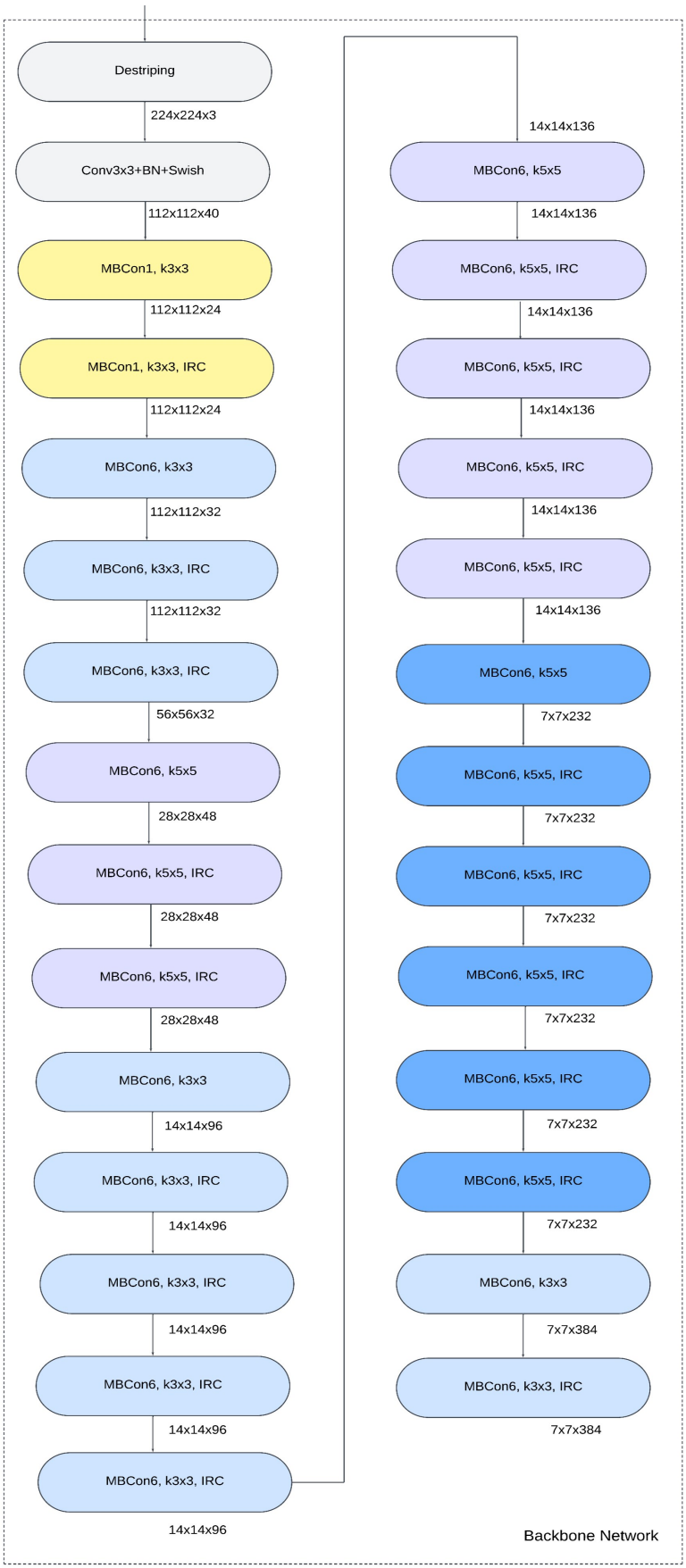


Рис.2.5 Ілюстрація EfficientNet-B3 архітектури. IRC (inverted residual connection)

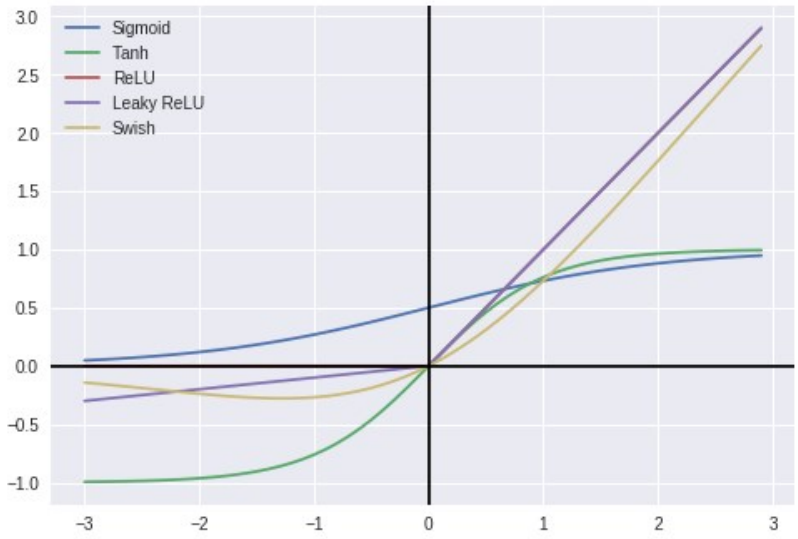


Рис.2.6 Функція активації Swish в порівнянні з іншими

3. Утворювати окремі пірамідальні мапи $\{P_9, P_P, P_M P_Q, P_0\}$ із однаковою кількістю каналів, але різним масштабом (від деталізованого до дуже зменшеного).

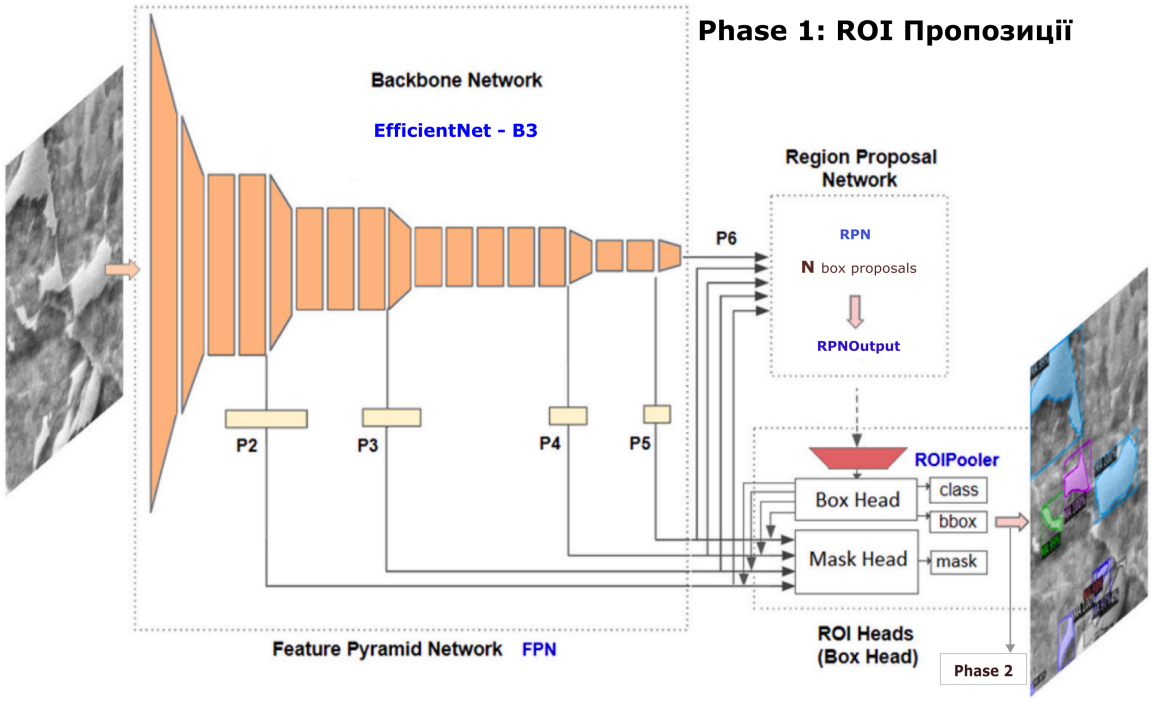


Рис. 2.7 Запропонована пірамідальна мережа для визначення регіонів інтересів в Фазі 1

Таким чином, FPN забезпечує багаторівневу обробку: менші рівні виявляють дрібні об’єкти, а глибші - великі структури. RPN застосовується до кожного рівня P_ℓ.

Нехай, $P_\ell \in \mathbb{R}^{!\ell \times \#\ell \times -}$, на кожній просторовій позиції (x, y) у P_ℓ задаємо набір "якорів" (anchors) $\{a_4\}$.

Тоді, RPN Head заданий як невелика згортка 3×3 із двома вихідними гілками:

$$P_\ell(x, y) \longmapsto \mathrm{p}p_4(x, y),\ \Delta\mathbb{b}_4(x, y)\mathrm{u}_{4+\text{„...,}8^!}, \tag{2.8}$$

де, $p_4(x, y)$ - objectness score, ймовірність наявності об'єкта/пошкодження для якірного вікна a_4, та $\Delta\mathbb{b}_4$ - вектор $(\Delta\mathrm{x}, \Delta\mathrm{y}, \Delta\mathrm{w}, \Delta\mathrm{h})$, для корекції координат цього вікна. Таким чином, $RPNOutput = \{BBOX_\&\}_{\&+,\ldots<}$.

Отримані, p_4 та $\Delta\mathbb{b}_4$ об'єднуються із усіх рівнів $\ell \in \{2, 3,4,5,6\}$, далі сортуються за p_4, де відкидаються зайві перекривання. У підсумку маємо N пропозицій (bboxes), що з найбільшою ймовірністю відповідають об'єктам або пошкодженням.

На виході RPN отримуємо грубий набір bounding boxes. Далі для кожної пропозиції застосовується ROI Pooling (або ROI Align), що "вирізає" відповідні фрагменти ознак з P_ℓ. Наприклад, у типовому Faster R-CNN [97], "Box Head" - уточнює клас і координати, а "Mask Head" - (у Mask R-CNN) генерує бінарну сегментацію об'єкта.

Натомість у запропонованій моделі (на етапі Фази 2) виконується семантична сегментація пошкоджень із використанням DReAM (динамічне масштабування рецептивного поля) і блоків Трансформера. Це дає змогу точно обробляти кожен ROI, визначаючи ступінь та тип руйнувань.

Таким чином, Фаза 1 виконує грубу (але ресурсно ефективну) локалізацію регіонів інтересу, тоді як Фаза 2 реалізує поглиблену процедуру семантичного й контекстуального аналізу, що дозволяє виявляти та оцінювати пошкодження із високим ступенем точності, а відповідно надає більшої градації локалізацій сцен пошкоджень.

2.4 Онтологічний зв'язок ознак об'єктів через атрибути OSM

Наступним кроком є алгоритм отримання інформації для локалізованих сцен руйнувань та формування онтологічного зв'язку за відкритими даними OSM (Open

Street Map), автори [99] формалізували модель компіляції географічних атрибутів на основі репутації контріб'юторів. Використання OSM не є вичерпним та рівень FP (false positive) може сягнути ~30%, в даній роботі досліджено використання онтології OSM як приклад інформативності отриманих результатів, в реальних апробаціях та комерційному застосуванні алгоритму використання подібних результатів може нести досить неочікувані результати так як інформація в OSM це відкрите джерело знань, відповідно, є сенс підмінити модуль онтології на існуючі комерційні розробки.

Отже, основна ідея полягає в тому, що всі отримані ознаки, які семантично сегментовані, проходять етап зіставлення з OSM, відповідно, можна зрозуміти різницю між семантичною сегментацією та локалізацією сцен, сам процес зіставлення починається з етапу формування графу знань $OSMKnowledgeGraph$ у лістингу 1 (Додаток 1).

Після виявлення пошкоджень (Фаза 2) модель мусить вирішити, як саме класифікувати локалізовані ділянки за допомогою відкритих даних OSM, при цьому спершу будується або доповнюється граф знань (OSMKnowledgeGraph) згідно з псевдокодовим лістингом (лістинг 1), у ньому показано, як завантажувати сирі елементи (node, way, relation) (див. рис.2.8), парсити їхню геометрію й теги, а потім визначати інфраструктурний тип (міст, дорога, будівля тощо) і створювати або оновлювати вузли й ребра в графі, забезпечуючи основу для подальшого онтологічного аналізу.

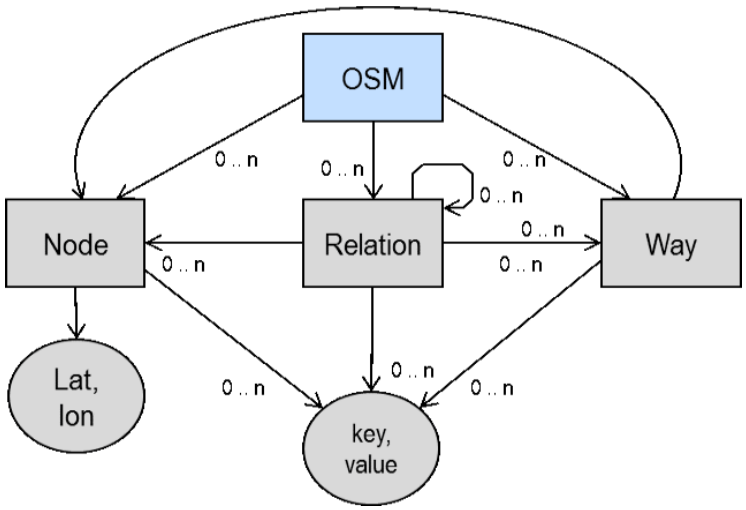


Рис. 2.8. Модель даних OSM

Далі, щоби формально задати й верифікувати процес зіставлення сцен пошкоджень із цим графом, представлено специфікацію в Z-нотації (див. лістинг 2, додаток 1) вона описує змінні моделі (Scenes, KGraph, Probabilities), інваріанти та операції (FindOverlaps, MatchScene тощо), що визначають логіку перевірки порогів, розрахунку ймовірностей та додавання результатів до структури Results, зокрема, коли виявлено перекриття з об’єктом із графа чи коли об’єкт залишається невизначеним. Завдяки такому формалізму можна довести, що алгоритм зберігає всі необхідні інваріанти (наприклад, жодна сцена з недостатньою впевненістю не позначається як пошкоджена), а також гарантувати узгодженість категорій (дорога, міст, будівля) (рис. 2.9) із геометрією OSM і рівнями ймовірності.

Рис. 2.9 Пошкоджений міст в м. Ірпінь локалізований за допомогою Grad-CAM heatmap та його онтологія OSM

Вибір формальної специфікації продиктований потребою мати не лише практичний псевдокод, а й засіб строгого доведення коректності - формальні нотації, як-от Z, дозволяють уніфіковано описати дані (граф OSM, сцени пошкоджень) та алгоритмічні кроки (обчислення перекриття, фільтрацію за порогами) з можливістю

верифікації, що особливо важливо за умови відкритості та потенційно неточних джерел на кшталт OSM у підсумку такий підхід спрощує оцінку надійності та переносимість рішення, а також підтверджує, що структура алгоритму витримує формальні докази в контексті обмежень і різних вхідних умов.

2.5 Фаза детальної оцінки сцен нейронною мережею трансформер з модулем DReAM

У Фазі 2 модель бере сформовані після Фази 1 пропозиції регіонів (ROI), далі за допомогою ROI Pooler витягує відповідні фрагменти ознак і пропускає їх через комбінацію Swin Transformer (що обчислює локально-глобальну увагу у вікнах) та DReAM (модуль динамічного масштабування рецептивного поля), у результаті чого отримує максимально адаптивне представлення кожного ROI та точніше визначає тип і стан об’єкта, далі ці результати можуть бути порівняні з даними з OSM (онтологічний зв’язок), аби уточнити або збагатити класифікацію додатковою інформацією.

Для розуміння загальної архітектури (див. рис. 2.10), необхідно формалізувати роботу Swin Transformer, та яку роль в загальному мережа Трансформер відіграє в цьому досліджені.

Ключовим аспектом цієї фази є те, що Swin Transformer дозволяє моделі ефективно враховувати як локальні особливості об’єктів, так і їхній глобальний контекст у сцені, що особливо важливо для ідентифікації пошкоджень з різними масштабами та формами. DReAM, у свою чергу, адаптивно змінює рецептивне поле залежно від структури зображення, компенсуючи обмеження EfficientNet – B3. Таким чином, поєднання цих двох підходів дозволяє моделі гнучко працювати зі складними сценами, що містять як великі, так і дрібні пошкоджені елементи. Крім того, онтологічний зв’язок з OpenStreetMap (OSM) вносить додатковий рівень перевірки, дозволяючи співвіднести знайдені об’єкти з існуючими картографічними даними та уточнити їхні характеристики, що особливо важливо для автоматизованого моніторингу.

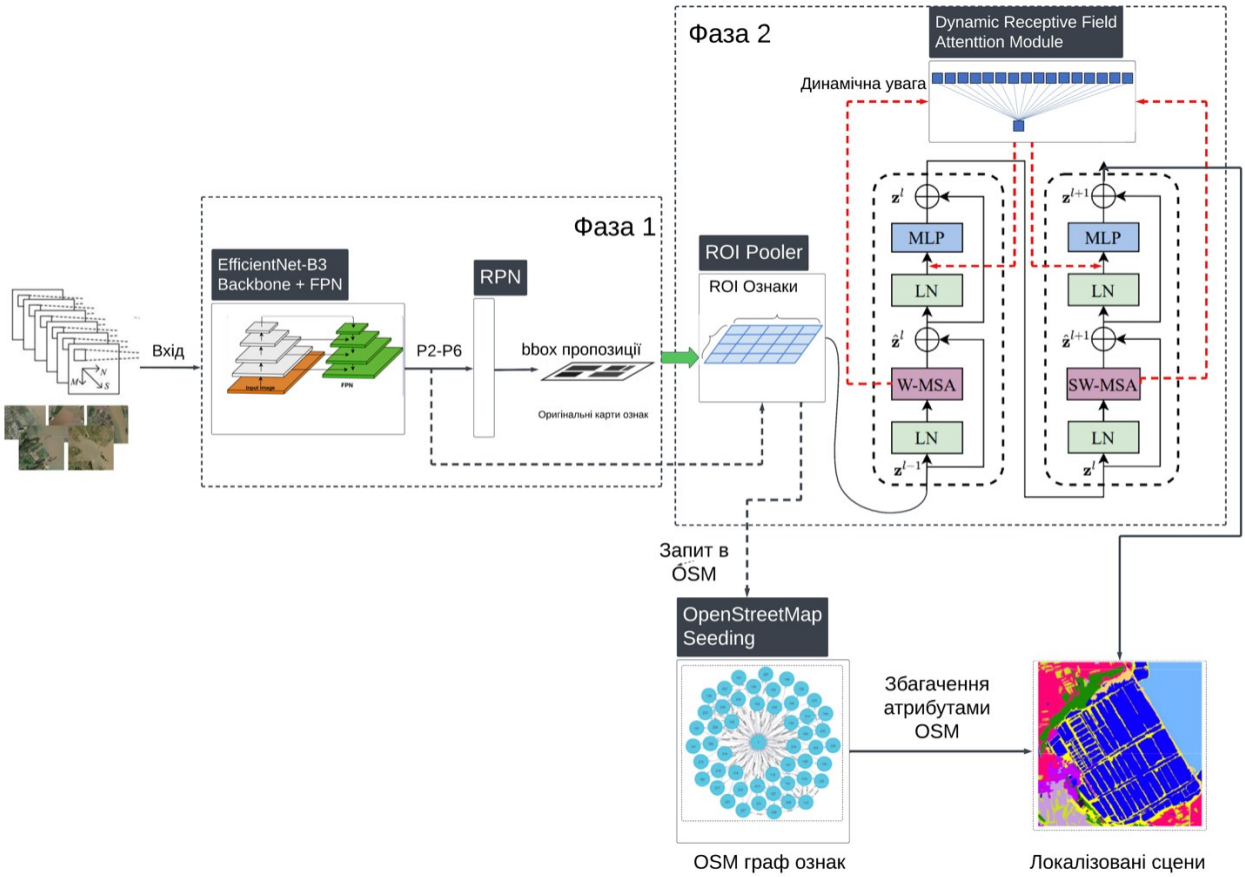


Рис. 2.10 Загальна архітектура композитної нейронної мережі для локалізації сцен руйнувань критичної інфраструктури

Swin Transformer - це ієрархічна архітектура [23] (див. рис. 2.11), у якій представлення обчислюється за допомогою зміщених вікон, що забезпечує вищу ефективність, обмежуючи обчислення самоуваги до зміщенних локальних вікон, водночас дозволяючи встановлювати зв'язки між вікнами. Swin Transformer будує ієрархічне представлення, починаючи з невеликих патчів і поступово об'єднуючи сусідні патчі на глибших шарах, що забезпечує гнучкість для моделювання об'єктів різних масштабів.

Ця архітектура [23] має лінійну обчислювальну складність щодо розміру зображення завдяки обчисленню самоуваги в межах кожного локального неперекривного вікна.

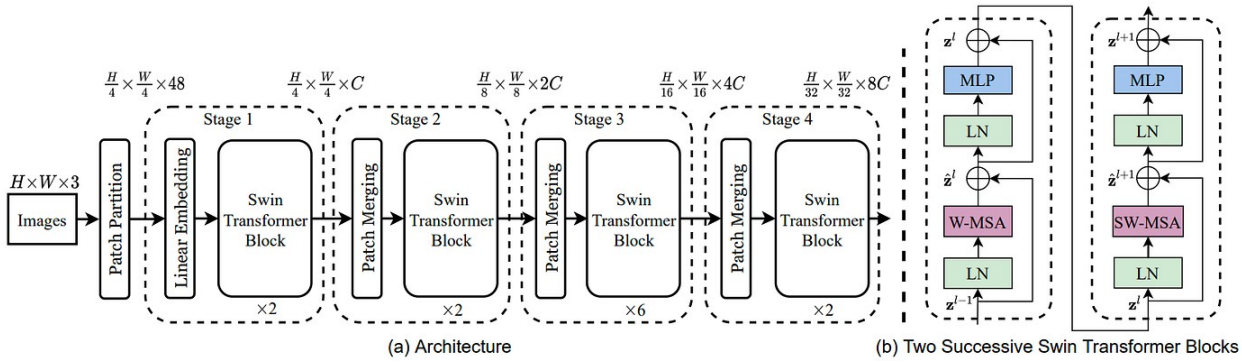


Рис. 2.11 a) Архітектура мережі Swin Transformer, b) Swin Transformer Block

Спочатку вхідне зображення передається через блок "Patch Partition", який розділяє його на патчі фіксованого розміру. Розмір патчів становить {4×4}, і в результаті блок "Patch Partition" формує {H/4 × W/4} патчів. Кожен патч має розмірність каналів 4×4×3 = 48 пікселів.

Етап 1. Для перетворення кожного патча з 48 пікселів до більш придатного розміру C використовується шар "Linear Embedding", який проектує патч у довільний розмір C. Отримана нова карта ознак передається через блок "Swin Transformer Block", причому розміри вхідних і вихідних даних залишаються незмінними.

Етап 2. Карта ознак розміру {H/4 × W/4 × C} передається через шар "Patch Merging", який об'єднує сусідні вікна розміром {2×2}, створюючи нову карту ознак. Цей процес зменшує роздільну здатність у 2 рази, але збільшує глибину карти ознак у 2 рази. Результат: розмір вихідної карти ознак стає {H/8 × W/8 × 2C}. Після цього нова карта ознак передається через ще один блок "Swin Transformer Block", який зберігає її розмір незмінним.

Етапи 3, 4. Етапи 3 і 4 повторюють ту ж процедуру, що й етап 2, з вихідними роздільними здатностями {H/16, W/16} та {H/32, W/32} відповідно.

Ці етапи спільно створюють ієрархічне представлення, де роздільна здатність карти ознак поступово зменшується, а її глибина збільшується. Така структура аналогічна типовим згортковим мережам, таким як VGGNet і ResNet, що дозволяє зручно замінювати стандартні базові мережі в існуючих методах для вирішення різних завдань комп'ютерного бачення.

Swin Transformer Block (див. рис. 2.12). Swin Transformer замінює стандартний механізм багатоголової самоуваги (MSA), який використовується у ViT, на увагу на основі вікон (W-MSA) та зміщену увагу на основі вікон (SW-MSA).

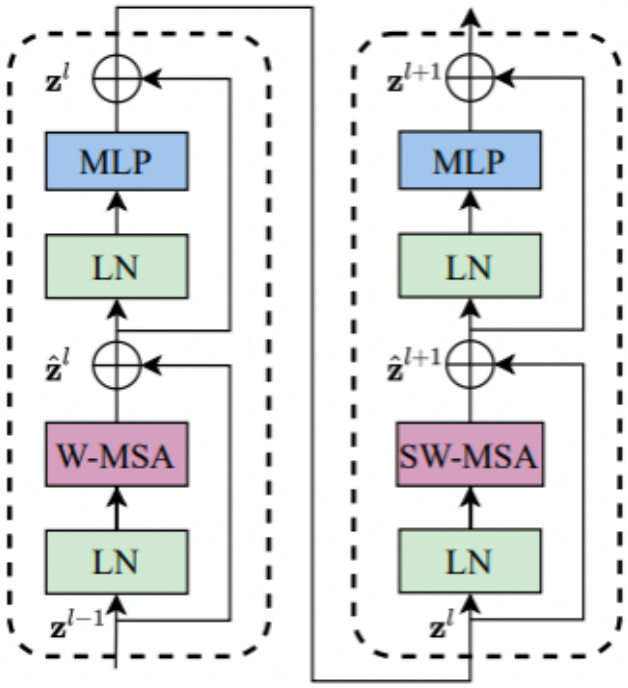


Рис.2.12 Swin Transformer Block

Блок Swin Transformer складається з двох підблоків. Кожен підблок включає такі компоненти:

1. LN (Layer Normalization) - модуль уваги;
2. LN (Layer Normalization) - багатошаровий перцептрон (MLP).

У першому підблоці використовується W-MSA (Window Multihead Self-Attention) (багатоголова самоувага на основі вікон). У другому підблоці застосовується SW-MSA (Shifted Window Multihead Self-Attention) (багатоголова самоувага з зміщенням вікон).

Стандартний механізм багатоголової самоуваги (MSA) у ViT (2.9) використовує глобальну самоувагу, при якій взаємозв'язок кожного патча обчислюється щодо всіх інших патчів - це призводить до квадратичної складності яку можна описати як:

$$\Omega(MSA) = 4hwC^9 + 2(hw)^9C. \quad (2.9)$$

де, $\Omega(MSA)$ обчислюється щодо кількості патчів, що робить підхід непридатним для обробки зображень високої роздільної здатності (див. рис. 2.13 [21]).

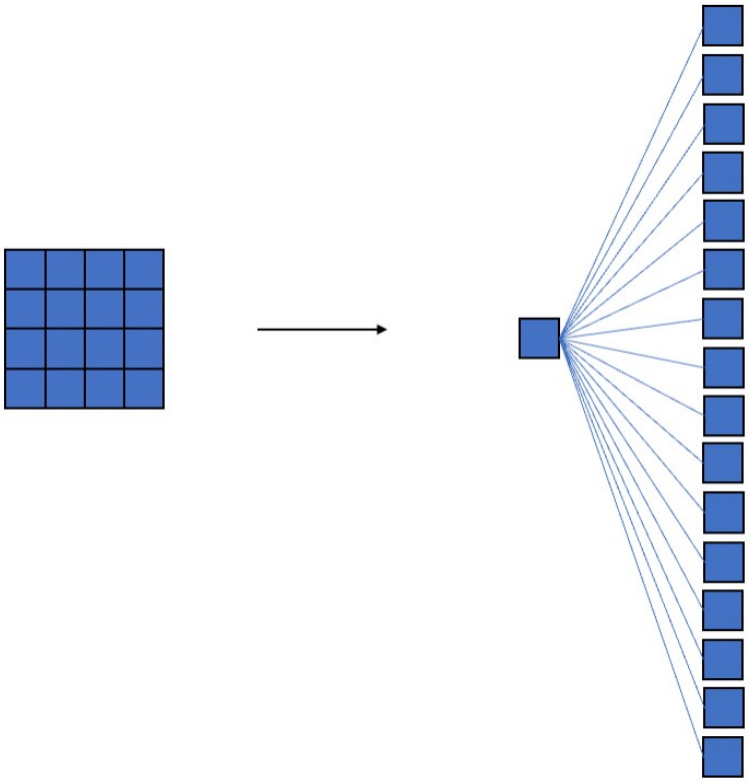


Рис. 2.13 Стандартний Multi-head Self Attention в ViT

Щоб вирішити цю проблему, Swin Transformer використовує підхід уваги на основі вікон (Window-based MSA). У Swin Transformer вибирається вікно фіксованого розміру, кожне з яких містить фіксовану кількість патчів (у статті зазначено {M × M} патчів). Увага обчислюється лише в межах кожного окремого вікна - це забезпечує лінійну складність щодо кількості патчів (див. рис. 2.14 [23]):

$$\Omega(WMSA) = 4hwC^9 + 2M^9(hw)C. \qquad (2.10)$$

Отже, з вище написаного поняття ERF (Effective Receptive Field) еефективне рецептивне поле ні в Swin ні в ViT не мають сенсу, таким чином, як воно має в згорткових мережах і це є хибним твердженням з точки зору відповідності уваги до певного вікна, або до контролю дистанції між токенами (dilation), слідує зауважити, що увага, а отже і рецептивне поле обмежене лише конкретним вікном або суміжними

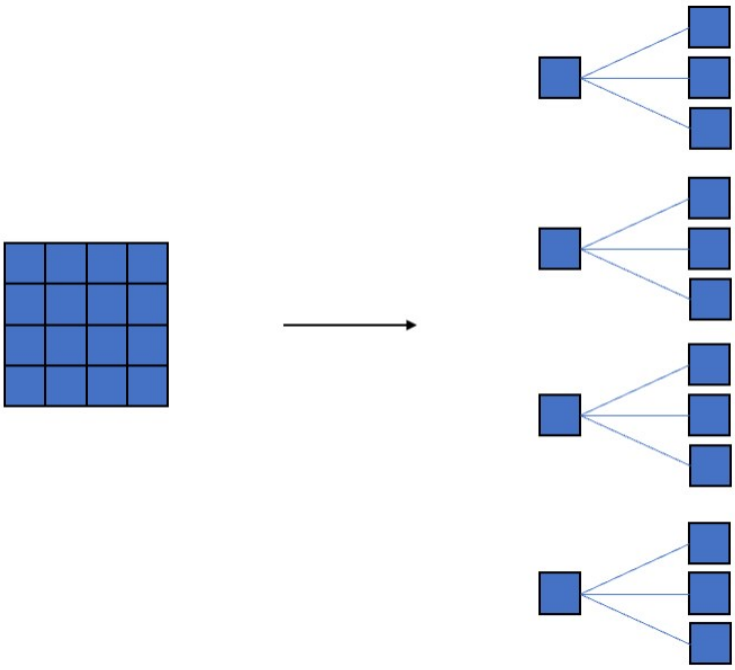


Рис. 2.14 W-MSA в Swin Transformer запропонована

токенами, тому в даній роботі запропонований модуль уваги динамічної зміни рецептивного поля, рис. 2.15, демонструє діапазон самоуваги в двох підходах описаних авторами в різних роботах [21], [23].

Якщо покладатися лише на W-MSA, взаємозв'язок між вікнами буде відсутнім, що обмежує його моделювальну потужність, далі використовується самоувага на основі зміщених вікон (SW-MSA).

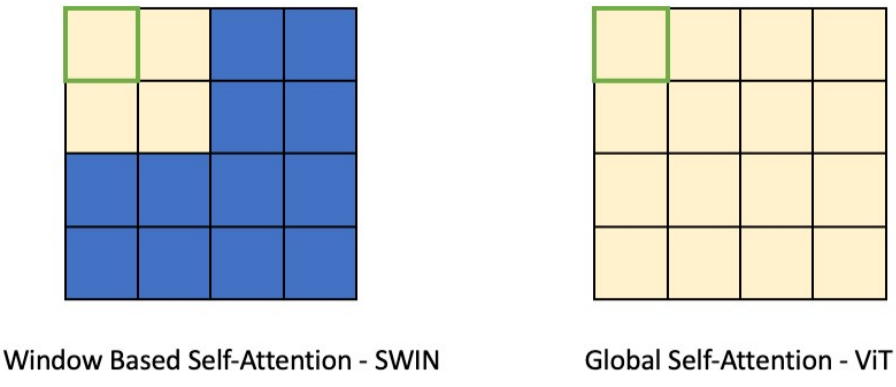


Рис. 2.15 Порівняння діапазону механізму самоуваги в Swin та ViT

Підхід SW-MSA:

1. Взяти вихідні дані з W-MSA;
2. Змістити всі вікна на {M/2, M/2} порівняно з попереднім шаром;

3. Обчислити W-MSA у зміщених вікнах.

На рисунку 2.16 карта ознак розміром 8×8 рівномірно розбивається на {2×2} вікна розміром {4×4} (M=4). Усі 4 вікна зміщуються на {M/2, M/2} = {2, 2} патчі вниз і вліво. Однак, таке зміщення призводить до появи "осиротілих" патчів, які не належать жодному вікну, а також вікон з неповними патчами. Swin Transformer використовує техніку "Cyclic Shift", яка переміщує "осиротілі" патчі до вікон із неповними патчами. Важливо зазначити, що після такого зміщення вікно може складатися з патчів, які не є сусідніми на оригінальній карті ознак. Тому під час обчислення застосовується маскована MSA, яка обмежує самоувагу лише до сусідніх патчів.

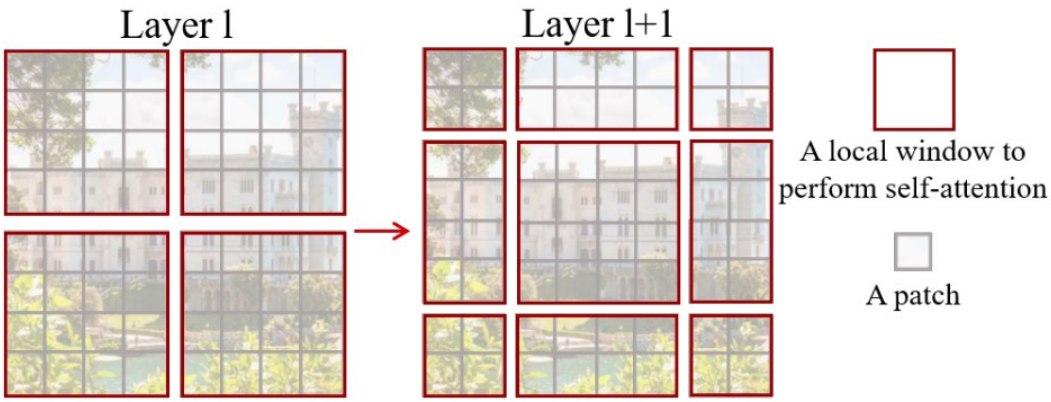


Рис. 2.16 Ілюстрація зміщеного вікна для обчислення самоуваги в Swin Transformer

Як показано на рисунку 2.17, ми хочемо обчислити (2.11). Уявімо, що беремо вікно, наведене в (2.12), і зміщуємо його вниз і вліво на {2×2}, після чого беремо патчі A, B і C та заповнюємо ними порожній простір.

Далі ми застосовуємо масковану MSA до зміщеного вікна, щоб забезпечити обчислення уваги лише серед необхідних частин у вікні.

Нарешті, у (2.13), ми виконуємо зворотне зміщення вікна та заповнюємо верхню ліву частину нового вікна патчами з нижньої правої частини. Цей процес дозволяє ефективно обчислювати увагу між вікнами, оскільки кількість вікон залишається такою ж, як і у W-MSA.

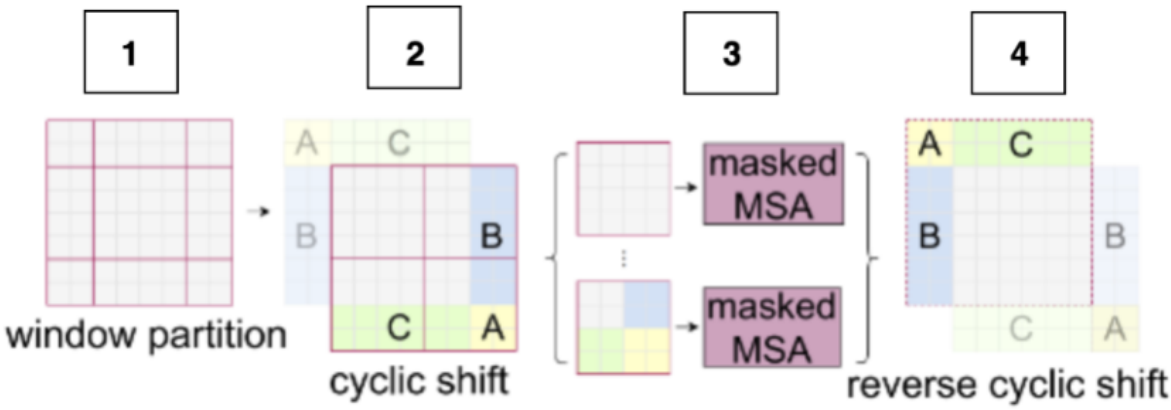


Рис. 2.17 Техніка Cyclic Shift

Swin Transformer використовує відносне позиційне зміщення (Relative Position Bias) для обчислення самоуваги. Це зміщення представлено у вигляді матриці B розміром $\{M^2 \times M^2\}$, де M^2 — це кількість патчів у кожному вікні. Відносне позиційне зміщення додається до оцінки подібності (attention score) між патчами, щоб врахувати їх відносне розташування всередині вікна, покращуючи здатність моделі розрізняти просторові взаємозв'язки. Отже, вся ідея математично формалізується до рівняння “уваги”:

$$Attention(Q,K,V) = SoftMax\left(\frac{QK^{T}}{\sqrt{d}} + B\right)V. \qquad (2.11)$$

Q, K та V - це матриці запиту (query), ключа (key) та значення (value);

d - це розмірність запиту/ключа.

Використання відносного позиційного зміщення значно покращує продуктивність у порівнянні з трансформерами, які застосовують абсолютне позиційне вбудовування (absolute position embedding).

Враховуючи, що Swin Transformer виконує самоувагу в межах локальних вікон (W-MSA) та зміщених вікон (SW-MSA), DReAM додає додатковий рівень динамічної адаптації. У загальних рисах це означає, що між (або паралельно) блоками W-MSA та SW-MSA інтегрується модуль, який:

1. Формує багатомасштабне відображення ознак з різними розмірами або dilation-коефіцієнтами;

2. Обчислює ваги для кожного масштабу з урахуванням локальних властивостей ROI (або патчів) і глобального контексту в межах вікон;
3. Об’єднує зважені ознаки в єдине представлення, яке потім передається далі в Swin-блок (наприклад, до MLP або на вхід наступного шару).

Якщо в класичному Swin Transformer увага в кожному шарі обмежена щільно визначеними вікнами {M×M}, то DReAM, використовуючи гілки з різним масштабом (наприклад, kernels 3×3, 5×5, 7×7 та ін.), дає змогу “розширити” чи “звузити” діапазон охоплення залежно від складності сцени та характеру об’єктів. Це дозволяє мережі:

1. Легко опрацьовувати дрібні деталі в межах одного вікна, якщо виявлені дуже точкові пошкодження (тріщини);
2. Переходити на більший контекст (більші віртуальні вікна) при великих руйнуваннях чи складних макроструктурах.

При цьому DReAM “перекриває” недолік, коли самоувага W-MSA/SW-MSA залишається в жорстко заданих локальних рамках за допомогою різноманітних масштабів згорток і ваг уваги DReAM може частково виходити за межі єдиного вікна, формуючи ширше чи інше поле зору.

Отже, для окремого шару (блоку) Swin Transformer маємо вхідний тензор ознак $X_{\&U}$ розмірності $H \times W \times C$ або набір токенів $\{x_i\}$ ділиться на вікна розміром $\{M \times M\}$, і в кожному вікні обчислюється W-MSA:

$$X_V = W\text{-}MSA(X_{\&U}), \tag{2.12}$$

це означає, що W-MSA виконує self-attention лише серед токенів у межах вікна.

Далі маємо обчислити SW-MSA і позначимо як X_W, тобто, ‘shifted’(зміщення):

$$X_W = SW\text{-}MSA(X_V), \tag{2.13}$$

який зміщує вікна на $\{M/2, M/2\}$ і знову обчислює увагу. Між цими кроками присутній LayerNorm, а також пропуск (residual). Однак для спрощення запису зараз відобразимо лише основну логіку.

Інтеграція DReAM формалізується як:

$$X_{\&} = f_{\&}(X_V), \qquad i = 1, \dots, K. \tag{2.14}$$

де K - “гілки”, а саме dilations тобто дистанція між токенами, кілька гілок відповідають $f_{\&}(\bullet)$ різним “відстаням” (dilations) у просторі токенів.

Якщо в класичному Swin кожен токен “бачить” лише ті, що розташовані безпосередньо поруч у $\{M \times M\}$ вікні, то за допомогою DReAM з dilation = 2, 3, тощо можна “пропускати” деякі позиції, аби токен одразу “звертався” до віддаленіших. Гілка з dilation=1 працює як звичайна локальна увага, а гілка з dilation = 2 захоплює ширше розташовані токени, фактично “розширюючи” ефективне поле зору (ERF). Завдяки softmax-вагам, DReAM “обирає”, яку з цих гілок переважно використати для певного ROI, отже і відбувається динамічне масштабування рецептивного поля. Таким чином, цей підхід узгоджується з токен-парадигмою Swin і дає змогу виходити за жорсткі локальні обмеження, коли сцена (ROI) потребує більшого чи дрібнішого охоплення.

Далі, підключається ваговий механізм, обчислюємо ваги $\{\alpha_{\&}\}$ для кожної гілки $\{i\}$, ця “скалярна оцінка” $\{\phi\}$ є MLP, що аналізує “наскільки” кожній сцені/ROI потрібен той чи інший dilation:

$$e_{\&} = \phi(X_{\mathrm{V}}, X_{\&}), \qquad \alpha_{\&} = \frac{exp(e_{\&})}{\sum_{\&+,}^{8} exp8e\,9}, \tag{2.15}$$

Фінальним виходом є:

$$X_{=} = \sum_{\&+,}^{8} \alpha_{\&} \cdot X_{\&}, \tag{2.16}$$

Таким чином, якщо в якійсь ділянці важливі далекі зв’язки між токенами, збільшується $\{\alpha_{\&}\}$ у гілці з більшим dilation, і модель розширює або звужує поле зору, залежно від того, яка з гілок найдоцільніша.

Після DReAM отриманий $X_{=}$ передається у зміщену віконну увагу:

$$X_{\mathrm{W}} = SW\text{-}MSA(X_{=}), \tag{2.17}$$

де, вже на вхід іде збагачений тензор $X_{=}$. Тепер віконна увага (shifted windows) обчислюється над “динамічно” обробленими ознаками.

Завершальним етапом є:

$$X_{\mathrm{XY})} = MLP(X_{\mathrm{W}}). \quad (2.18)$$

Інтуїція полягає в наступних твердженнях:

- Коли DReAM виявляє “дрібні” об’єкти чи текстури (наприклад, тріщини, деталі будівель), коефіцієнти $\{\alpha_{\&}\}$ підсилять гілку з малими dilation, і далі Swin “працює” над більш деталізованими ознаками;
- Якщо ж у ROI (Фаза 1) переважають “великі” структури (обвалена секція мосту, протяжні дороги), DReAM може збільшити вагу масштабів із більшими ядрами $\{k_{\&}, d_{\&}\}$, щоб модель “захоплювала” ширший контекст перед переходом до SW-MSA.

Таким чином, DReAM “готує” або “адаптує” ознаки, передані Swin-блоку, динамічно змінюючи ефективне рецептивне поле залежно від складності та специфіки сцени, у той час як віконна та зміщена увага (W-MSA, SW-MSA) залишаються відносно жорсткими у своїй зоні охоплення $\{M \times M\}$. Це надає можливість прискорити обчислення і навіть застосування в умовах наближених до реального часу.

Висновки до розділу

Таким чином, у другому розділі було сформовано та детально описано архітектуру запропонованої моделі, яка поєднує механізм грубої локалізації регіонів інтересу (ROI) із детальною фазою сегментації та аналізу пошкоджень. На цьому етапі обґрунтовано дворівневий підхід: спершу виконується швидке визначення й відбір потенційно зруйнованих ділянок, а потім - багатошарова обробка виділених областей за допомогою поглиблених методів нейронних мереж.

Ключовим елементом розглянутої моделі став модуль динамічного масштабування рецептивного поля (DReAM). Саме він дає змогу гнучко адаптувати поле зору мережі та коректно поєднувати локальні ознаки (дрібні деталі) із глобальним контекстом (макроструктури сцени). При цьому використання

трансформерних блоків (Swin Transformer) розширює здатність мережі враховувати взаємозв’язки між об’єктами різних масштабів та підвищує точність розпізнавання.

Зрештою, у цьому розділі визначено алгоритмічні кроки, необхідні для навчання та інференсу, включно з методами фільтрації вихідних результатів та інтеграції їх з геопросторовими метаданими (OSM). Така комплексна архітектура становить основу запропонованого рішення і закладає фундамент для експериментальної перевірки, поданої в наступних розділах.

РОЗДІЛ 3 ЗАСТОСУВАННЯ, ПЕРВІРКА ЯКОСТІ, ТА ОПТИМІЗАЦІЯ МОДЕЛІ ЛОКАЛІЗАЦІЇ СЦЕН

У третьому розділі розглядаються питання застосування розробленої моделі та методів її оцінювання на практичних прикладах, а також оптимізація обчислювальних ресурсів із метою підвищення швидкодії та точності. Спершу описано конфігурацію середовища навчання та характеристики наборів даних, зокрема особливості великомасштабних знімків дистанційного зондування й специфічні ознаки пошкодженої критичної інфраструктури. Задля досягнення найкращих результатів аналізується вибір функції втрат, підбір гіперпараметрів і стратегій аугментації.

У наступних підрозділах приділено увагу покроковому тренуванню окремих компонентів архітектури: від формування регіонів інтересу (ROI) до модулю динамічного масштабування рецептивного поля з увагою (DReAM). Особливо відзначено проблему незбалансованості даних, коли кількість прикладів із критичними руйнуваннями може бути значно меншою за сукупність непошкоджених зразків. Для мінімізації такої диспропорції використовуються спеціалізовані схеми Focal Loss та інші засоби корекції.

Фіналом розділу представлено етапи оптимізації, зокрема квантування (quantization) та змішана прецизійність (mixed precision), що дають змогу зменшити витрати пам’яті та прискорити обчислення без критичного зниження точності. Досліджені результати підтверджують вагомість адаптивного підходу за принципом різнотипності, запропонованого в попередніх розділах.

3.1 Впровадження моделі локалізації сцен на основі запропонованої архітектури композитної нейронної мережі

У цьому розділі розглядається процес навчання композитної нейронної мережі, описаної в попередніх підрозділах. Нагадаємо, що запропонована модель складається з двох фаз Фаза 1 (груба оцінка, формування ROI, RPN) та Фаза 2 (детальна локалізація й оцінка пошкоджень за допомогою Swin Transformer та DReAM).

Головна ідея полягає у поєднанні ефективних ознак від EfficientNet-B3 (з пірамідою FPN) із динамічним масштабуванням рецептивного поля (DReAM) та віконною самоувагою (W-MSA/SW-MSA) у Фазі 2, що дає змогу адаптивно враховувати як локальні деталі, так і глобальний контекст.

Для досягнення високої точності та ефективності в обробці зображень дистанційного зондування (ДЗ) буде розглянуто ретельне налаштування (1) даних (зокрема формування батчів, аугментацію та можливість синтетичної вибірки), (2) оптимізаторів й гіперпараметрів (швидкість навчання, регуляризацію), а також (3) покрокове або почергове тренування окремих компонентів (попереднє навчання EfficientNet-B3 на ImageNet, а згодом адаптація у складі FPN і RPN, та фокус на DReAM і Swin-частині).

У подальшому тексті описано:

- Підготовку даних: які набори (xBD, DOTA, MAXAR’s Open Data) використовувались, які характеристики мають, як виконується нормалізація та компресія розмірності;
- Налаштування середи: апаратне прискорення (GPU), використання фреймворків, скрипти розгортання (deployment) та контроль версій;
- Процес навчання: порядок тренування окремих блоків (RPN, Box/Mask Head, Swin+DReAM), підбір функції втрат та оптимізація гіперпараметрів, а також стратегії аугментації, покликані підвищити узагальнювальну здатність моделі;
- Первинну оцінку та порівняння результатів: які метрики (mAP, IoU, F-score) застосовувались, як впливає DReAM на точність у складних сценах та наскільки швидко модель навчається.

Завдяки цьому послідовному підходу вдається продемонструвати, що запропонована модель може одночасно досягати високої продуктивності (достатня швидкість обробки сцен) і точної локалізації пошкоджень різних масштабів, що підтверджує практичну цінність розробленої архітектури.

3.2 Тренування композиційної нейронної мережі та визначення параметрів, наборів даних та їх характеристик

Відсутність у відкритому доступі широкомасштабних датасетів, спеціалізованих саме на ушкодженнях критичної інфраструктури (мостів, дорожніх вузлів, електростанцій тощо), залишається одним із ключових викликів під час побудови високоточних моделей локалізації та оцінки руйнувань. Більшість наявних наборів даних для дистанційного зондування (наприклад, xBD [100], DOTA [101], AIRS [102]) зосереджені переважно на житлових чи комерційних будівлях, або на загальному розпізнаванні об’єктів без детального опису ступеня пошкоджень. Внаслідок цього моделі, навчені лише на типових датасетах, можуть виявлятися недостатньо гнучкими, коли мова йде про специфічні елементи інфраструктури, такі як:

- Мости (зокрема дорожні, залізничні);
- Магістральні дороги (шляхи, естакади);
- Промислові об’єкти (ТЕС, ГЕС, підприємства);
- Вузли енергомереж (трансформаторні підстанції, лінії електропередач).

Щоб подолати цю нестачу, в рамках даного дослідження було обрано та об’єднано декілька різнорідних джерел.

Важливим моментом при проєктуванні двофазної моделі локалізації є те, що Фаза 1 (грубе визначення регіонів інтересу, ROI) та Фаза 2 (детальна оцінка пошкоджень) мають неоднакові вимоги до наборів даних. Якщо перша фаза зосереджена на виявленні потенційно пошкоджених ділянок (не завжди тільки будівель, а й інших об’єктів), то друга - на точній семантичній сегментації та класифікації ступеня руйнувань. Тому в межах цього дослідження:

Фаза 1 (ROI-виявлення):

- Орієнтована на максимально широку різноманітність сцен і можливих типів об’єктів. Тут важливо навчити модель (RPN + FPN) виявляти будь-які аномалії чи підозрілі структури, які можуть бути зруйнованими;

- Набір даних повинен містити різні контексти (місто, промзона, транспортна інфраструктура), щоб RPN навчилася пропонувати bounding boxes для подальшої детальної перевірки.

Фаза 2 (детальна сегментація пошкоджень):

- Орієнтована на сегментаційні анотації та класи пошкоджень (наприклад, “minor”, “major” тощо), які були детально розмічені в одному чи кількох датасетах;
- У цьому випадку дуже важливо мати високоточні контури об’єктів чи хоча б маски, а також узгоджені категорії руйнувань.

Зважаючи на це, доцільно поєднувати кілька джерел (див. табл. 3.1).

Табл. 3.1 Набори даних які використані для тренування запропонованої моделі для виявлення пошкоджень КІ

Набір даних	**Розмір**	**Роздільна здатність**	**Тип зображень**	**Задача**	**# Анотованих класів**
xBD/xFBD [100]	22068	1024 x 1024	Супутникові	Класифікація, Сегментація, Локалізація	4
fMoW [103]	~ 1млн.	Різноманітна	Супутникові	Класифікація	63
DOTA [101]	11268	800 x 800 до 20000 x 20000	Супутникові, UAV	Класифікація, Сегментація, Локалізація	18
Maxar’s Open Data [104]	~ 100	Різноманітна (VHR, HSR)	Супутникові	Сегментація, Локалізація	-
UADamge [105]	~1000	Різноманітна (VHR, HSR)	Супутникові, UAV	Сегментація, Локалізація	4

Таким чином, під час тренування ми розділяємо підходи:

- Дані для Фази 1:
  - Включає набагато більший пул зображень (у тому числі з Maxar), де немає точних міток руйнувань, але є простіші або ручні анотації “наявність / відсутність пошкодження”;

- Використовує проєкт xBD (див. Рис. 3.1), але, без дуже докладного розрізнення класів руйнування - достатньо інформації “будівля пошкоджена / ні”, фактично в більшості датасетів дані розмічені і підготовлені до використання.

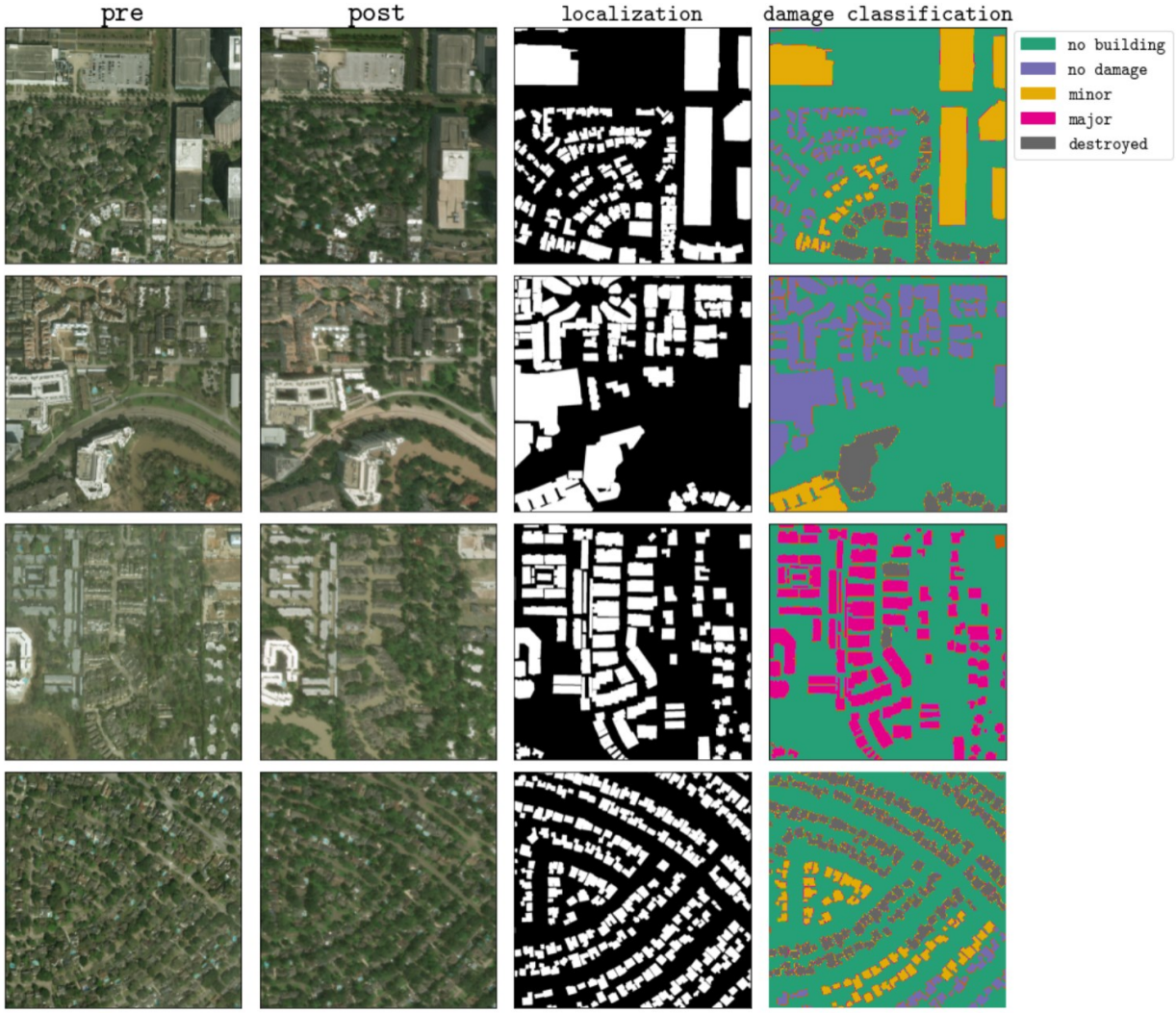


Рис. 3.1 Приклад xBD датасету з анотаціями для 4-х можливих станів (пошкодження відсутні, незначні, значні, повністю пошкоджене)

- Дані для Фази 2:
- Основний наголос на xBD/xFBD/DOTA/fMoW та додатково на тих сценах із Maxar, де ми спеціально розмітили критичні об’єкти (мости, дороги);
- Потрібна докладна сегментаційна та / або класифікаційна інформація про тип і ступінь ушкоджень (особливо якщо треба розрізняти “minor vs. major” тощо) для цього використано дані проєкту UADamage - зауважимо, що проєкт не анонсує відкрити набори даних, але має не погану базу для порівняння якості роботи запропонованої архітектури та моделі.

Внаслідок цього двоетапного підходу модель стає гнучкішою і масштабованішою. Перша фаза отримує змогу “бачити” широкий діапазон реальних катастрофічних сцен і пропонувати bounding boxes із мінімальною ручною роботою, а друга фаза досягає високої точності в розпізнаванні ступеня руйнування там, де наявні більш якісні анотації.

Наступним кроком є уніфікація всіх форматів даних. Більшість датасетів подаються в GeoJSON або COCO форматі, для ефективної інтеграції в пайплайн RPN/ROIHeads (Фаза 1) чи семантичної сегментації (Фаза 2) потрібен єдиний формат, таким єдиним форматом обрано COCO. Також, всі датасети приведено до єдиної роздільної здатності - xBD зазвичай уже має 1024×1024 тайли. DOTA іноді містить дуже великі панорами, їх нарізано на фрагменти (tile) 1024×1024 або 512×512, залежно від GPU-ресурсів та batch-розмірів, також застосовано padding.

Після об’єднання наборів даних, виокремлено лише класи пов’язані з пошкодженнями (див. табл. 3.2), та/або руйнуваннями, тестві набори також містять і неушкоджені об’єкти, але тільки у випадку коли йдеться про об’єкти критичної інфраструктури, таким чином “розмивається” певний дисбаланс класів, відповідно, наведені значення приблизні в діапазоні +/- 100 знімків на набір.

Табл. 3.2 Розподіл наборів даних між класами та тренувальним призначенням

Набір	**Кількість знімків (включно КІ)**	**Споруда (ціла) (включно КІ)**	**Споруда (пошкоджена) (включно КІ)**	**Міст (цілий)**	**Міст (пошкоджений)**	**Дорога (пошкоджена)**
Тренувальний	18000	~90000	~65000	~2100	~900	~5500
Валідаційний	2000	~10000	~7200	~250	~120	~700
Тестовий	2000	~9800	~7200	~230	~110	~650
Разом	22000	~109800	~79200	~2580	~1130	~6850

Співвідношення в даному дослідженні Train : Validation : Test умовно взято ~ 77 % : 12 % : 11 %, але конкретні пропорції залежать від практичних міркувань.

Надалі, візуалізуємо класи пошкоджень (див. рис. 3.2 – 3.4). Первинний відбір сцен в тренувальних даних базується на правилах відбору з величезного обсягу знімків за різні роки обираємо лише регіони, де теоретично відомо про катастрофу (ураган, військові дії, тощо). Зосереджуємося, наприклад, на 2 - 3 районах, де, за нашими джерелами, можуть бути пошкоджені дороги чи мости, це дозволяє зробити додаткову анотацію знімків які не є анотованими за своєю природою (Maxar, Planet Lab, тощо). Великі ортотрансформовані TIFF-файли (30000×30000 пікселів і більше) ділимо на менші фрагменти (стандарт для цієї роботи є 1024×1024), щоб зручно було завантажувати в інструменти розмітки (CVAT, Labelbox). Зображення розміром 512×512 зазвичай використовують, коли необхідно зменшити навантаження на обчислювальні ресурси або прискорити обробку. Наприклад, коли наявна обмежена кількість відеопам’яті (VRAM) на графічній карті або коли потрібно швидше здійснювати ітерації навчання та валідації моделі. Також у випадках, коли цілі задачі (наприклад, пошук тріщин чи дрібних об’єктів) можуть ефективно вирішуватися на менших фрагментах, 512×512 виявляється оптимальним розміром. Це дозволяє швидко застосовувати різноманітні методи аугментації й водночас забезпечує достатню роздільну здатність для розпізнавання потрібних деталей.

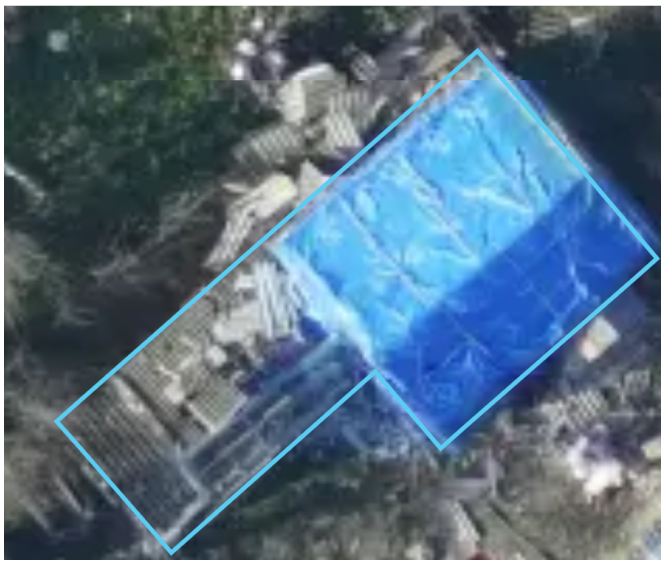


Рис. 3.2 Повністю пошкоджено (обвал більше ніж 50%) (джерело Mapbox)

На рисунку 3.2 позначено зону повної руйнації споруди з обвалами понад 50% її конструкції. Рисунок 3.3 показує ділянку, де помітно значні руйнування даху та стін

будівлі, що дає уявлення про масштаби пошкоджень. Обидва приклади ілюструють підхід до розмітки та аналізу, необхідний для ідентифікації та класифікації серйозних дефектів.

Рисунки 3.2-3.3, відображають не є критичними інфраструктурами за своїм призначенням, але повністю демонструють нагальний контекст і можливість адаптації під задачі виявлення пошкоджень в КІ.

Рис. 3.3 Руйнування перекриття (Відмічено обвали стін та даху) (джерело Mapbox)

Рис. 3.4 Незначні пошкодження (сліди вибухів на споруді) (джерело Mapbox)

Рис 3.4-3.5 демонструють анотовані об’єкти критичної інфраструктури, з огляду на попередні знімки можна побачити, що текстури досить різноманітні, рис. 3.4

звичайна будівля в контексті текстур, рис. 3.5 - досить складна текстура з комплексною сценою (комплекс будівель Запорізької атомної електростанції).

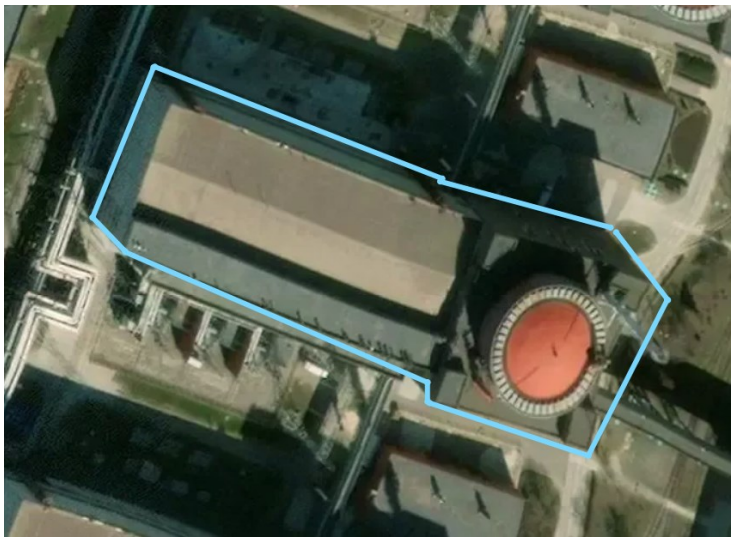


Рис. 3.5 Об’єкт не пошкоджено, видимі ознаки пошкоджень відсутні (джерело Mapbox)

Також, зображення не є ортотрансформованим що вносить додаткову складність на опрацювання проекцій та затемнення.

Таким чином, підготовка датасетів для обох фаз - це багатокроковий процес, де вже анотовані набори (на кшталт xBD чи DOTA) потребують переважно уніфікації формату й базового ресемплінгу, а неанотовані (Maxar’s Open Data та інші) - розмітки з нуля, чи то ручної, чи з використанням проміжних евристик. Результатом є узгоджена, чиста вибірка, де перша фаза (ROI-виявлення) може спиратися на ширший спектр (зокрема з грубими bounding box), а друга - на детальні полігони та категорії пошкоджень. Це створює основу для ефективного тренування композитної нейронної мережі з мінімальним ризиком некоректної чи розрізненої розмітки.

3.2.1 Налаштування середовища тренування та автоматизоване розгортання моделі локалізації сцен

Успішне впровадження запропонованої архітектури (Фаза 1 + Фаза 2) вимагає ретельно спроєктованого робочого середовища та пайплайна автоматизації. Основна мета - забезпечити відтворюваність (reproducibility) результатів, масштабованість (scalability) обчислень та гнучкість у розгортанні як у тестових, так і в продуктивних

сценаріях. Для експериментів і пілотних запусків було обрано Amazon Web Services (AWS) [106], що надає широкий спектр сервісів для навчання, інференсу та оркестрації.

Для безпосереднього тренування (обох фаз) обрано EC2 інстанс g4ad.8xlarge – (див. табл. 3.3).

Табл. 3.3 Підсумок з описом основних параметрів інфраструктури та середовища , налаштованих для тренування та розгортання моделі в AWS

Компонент	Характеристики	Примітка
Amazon EC2 (Elastic Compute Cloud)	g4ad.8xlarge	Належить до серії g4dn, оптимізованої для графічних обчислень та DL/ML задач
GPU	1× NVIDIA T4 (16 ГБ GDDR6)	Теоретична потужність: ~8.1 TFLOPS (FP32); Turing-архітектура, добра сумісність із CUDA та TensorRT
vCPU (процесорні ядра)	16 vCPU	Загальна кількість віртуальних ядер на базі Intel Xeon (як правило, друге покоління Scalable).
RAM (оперативна пам'ять)	64 ГБ	Достатньо для тренування середнього розміру моделей і одночасної обробки порівняно великих патчів.
Сховище	EBS (Elastic Block Storage) SSD	Типовий блоковий накопичувач для ОС і локальних даних; обсяг задається під час запуску інстанса.
S3 (Simple Storage Service)	Використання бакету для даних та чекпойнтів	Зберігання основних датасетів (xBD, DOTA, Maxar тощо), а також збереження моделей, логів і проміжних результатів.
Операційна система	Ubuntu 20.04 (Focal Loss)	Стабільний LTS-реліз, сумісний із сучасними фреймворками DL та драйверами NVIDIA.
Версія Python	Python 3.9	Підтримує актуальні версії PyTorch, NumPy, OpenCV та ін.
CUDA-стек	CUDA 11.x + cuDNN 8.x	Для прискорення глибинних обчислень на GPU (T4); збирання Docker-образу з відповідним базовим середовищем.

Контейнеризація	Docker 20+ / Dockerfile з PyTorch та іншими залежностями	Автоматична збірка (CI/CD), деплой на EC2 або ECS за потреби, забезпечує відтворюваність середовища.
Деплоймент	Напряму через EC2	Запуск контейнерів та моделей відбувається здебільшого вручну.

Таким чином, g4dn.4xlarge під керуванням Ubuntu 20.04 і Python 3.9, використовуючи NVIDIA T4, формує оптимальний баланс між вартістю та продуктивністю для задач локалізації й сегментації ушкоджень критичної інфраструктури.

3.2.2 Тренування окремих компонентів композитної нейронної мережі, оцінка функції втрат, та оптимізація параметрів моделі

Розроблена в рамках дослідження двофазна модель для локалізації та оцінки пошкоджень критичної інфраструктури складається з кількох ключових компонентів, кожен із яких потребує цілеспрямованого тренування. Нижче на рисунку 3.6 описано методологію навчання окремих модулів.

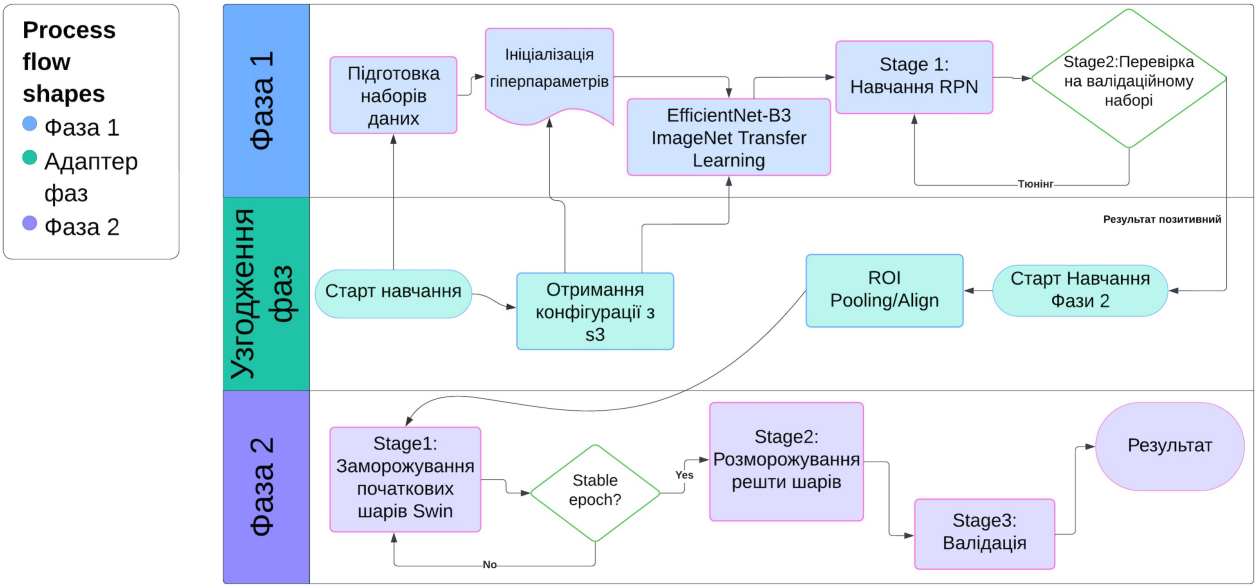


Рис. 3.6 Процес навчання для окремих компонентів архітектури

На першому етапі модель виявляє регіони інтересу (ROI), які можуть містити об’єкти критичної інфраструктури з пошкодженнями (мости, дороги, будівлі) або не містити їх. Центральним елементом тут є Region Proposal Network (RPN) в поєднанні з бекбоном EfficientNet-B3, що виконує функцію вилучення ознак.

EfficientNet-B3 являє собою згорткову нейронну мережу зі збалансованим масштабуванням глибини (depth), ширини (width) і роздільності вхідних зображень (resolution). Вона попередньо натренована на ImageNet, а також додатково донавчалася (fine-tuning) на датасетах, наведених у попередньому розділі. Такий підхід дозволяє:

1. Прискорити збіжність - завдяки вже “вивченим” базовим фільтрам (контури, кути, текстури);
2. Покращити загальну узагальнювальну здатність - навіть якщо змінюються типи зображень (сателітні, аерофотознімки тощо), мережа має глибоке уніфіковане представлення низькорівневих ознак.

Як уже зазначено, бекбон ініціалізовано вагами ImageNet та додатковими донавчаннями (див. розділ 2.1). Це прискорює навчання RPN, оскільки початкові фільтри вже вміють виділяти базові візуальні ознаки.

Наступний етап – конфігурація функції втрат. Визначимо втрату як: $L_{>Z<}$ яка складається з двох складових:

1. Classification Loss (крос-ентропія), що оцінює, чи пропоноване «якірне» вікно відповідає об’єкту (пошкодження) або фону:

$$L_{[5W} = -\sum_{4+,}^{8} [y_4 \, log \; p_4 + (1 - y_4) log \; (1 - p_4)], \qquad (3.1)$$

де, p_4 - імовірність належності до класу “пошкодження”, $y_4 \in \{0,1\}$ -бінарна змінна;

2. Regression Loss - обрано Smooth L1 (ця функція краще згладжує відмінності при малих величинах помилок, водночас лишаючись більш стійкою до викидів ($|z| > 1$), порівняно з абсолютним L1 чи квадратичним L2), отже:

$$L_{\mathrm{B?(}}(\Delta b, \Delta b^{*}) = \sum_{\&\in\{:,;,\mathrm{V},\hat{}\}} smooth_{5,}(\Delta b_{\&} - \Delta b_{\&}^{*}), \qquad (3.2)$$

де,

$$smooth_{\cdot_{7}}(z) = \begin{cases} 0.5z^{9} & if\ |z| < 1, \\ |z| - 0.5 & otherwise, \end{cases} \qquad (3.3)$$

$\Delta b_{\&}$ - передбачена поправка координат,

$\Delta b_{\&}^{*}$ - еталонна.

3. Загальна функція втрат: $L_{>\mathrm{Z}<} = \lambda_{[5\mathrm{W}} L_{[5\mathrm{W}} + \lambda_{\mathrm{B?(}} L_{\mathrm{B?(}},\ \lambda_{[5\mathrm{W}} = \lambda_{\mathrm{B?(}} = 1.$

Схема навчання для Фази 1 наведена в табл. 3.4.

Табл. 3.4 Параметри для тренування RPN на NVIDIA T4

Параметр	**Значення**	**Примітка**
Optimizer	AdamW	LR = 1e-3, Weight Decay = 1e-4
LR-схема	Warm-up (500 іт.)	Потім Cosine Annealing або Step Decay; старт LR 1e-3
Batch Size (розмір фрагменту)	4 (інколи 8)	Обмежено 16 ГБ GPU-пам'яті (NVIDIA T4). Можна збільшити, якщо знизити роздільну здатність
Епох (ітерацій)	20 k (ітерацій)	Приблизно 10 k кроків до step-down LR (1e-4)
Втрати	CE (класифікація) + Smooth L1 (регресія)	$\lambda_{!"\#} = \lambda_{\$\%\&} = 1$

На Stage 1 навчання RPN відбувається приблизно за 10 000–20 000 ітерацій. На початковій підмножині train-dataset мережа засвоює, як генерувати достатньо точні ROI. Бекбон EfficientNet-B3 у цей період відкритий для fine-tuning або частково заморожений (залежно від обсягу даних) (див. табл. 3.5).

Таблиця 3.5 Тривалість Stage 1

Етап	**Час (год.)**	**К-сть ітерацій**	**mAP на валідації ($IoU \geq 0.5$)**
Stage 1 Start	~3	2000	0.42
Stage 2 End	~10	10000 - 20000	0.65 – 0.70

Далі, на Stage 2 обчислюється mAP (mean Average Precision) для bounding boxes при різних порогах IoU (Intersection over Union):

$$IoU8B_{\text{JB?-}}, B_{()}9 = \frac{™B_{\text{JB?-}} \cap B_{()}™}{™B_{\text{JB?-}} \cup B_{()}™}, \quad (3.4)$$

де, $B_{\text{JB?-}}$ – передбачене вікно, $B_{()}$ – еталонне.

$$mAP = \frac{1}{N}; \sum_{\&+,} \check{S}\text{œ}\int_{a}^{<} p_{\&}(r)dr‹ \quad (3.5),$$

де, $p_{\&}(r)$ - залежність точності (precision) від повноти (recall) для і-го класу, інтегрована за $r \in [0,1]$, потім усереднена по всіх N класах (тут, зазвичай, 2 - “об’єкт”, і “фон”).

У випадку цього дослідження Precision/Recall обчислюється для класу “пошкоджене/непошкоджене” для економії обчислювальних ресурсів і цього достатньо в Фазі 1. Для mAP обрано поріг прийняття рішення на доволі стандартних рівнях $IoU \geq 0.5$ і $IoU \geq 0.75$ в залежності від джерел даних.

Нагадаємо, що після того, як Фаза 1 генерує bounding boxes, друга фаза виконує семантичну сегментацію для визначення типу та ступеня пошкоджень (minor, major, destroyed). Для кожного зображення в тренувальному наборі RPN формує набір пропозицій $\{\boldsymbol{b}_{,}, \boldsymbol{b}_{9}, \dots, \boldsymbol{b}_{<}\}$. Кожна з цих пропозицій $\boldsymbol{b}_{\&}$ – це прямокутник, який ROI Align перетворює у фіксований тензор розміру $H \times W$:

$$\boldsymbol{R}_{\&} = ROIAlign(\boldsymbol{F}, \boldsymbol{b}_{\&}, H, W), \quad (3.6)$$

де, $\boldsymbol{F}$ передається Swin Transformer, який виконує віконну (window) самоувагу, захоплюючи внутрішні патерни й локальні деталі пошкодження.

Swin Transformer (а саме аріхтектура “Swin-Small”) попередньо тренується на ImageNet, що підтверджує високу здатність до вилучення високорівневих ознак. У контексті завдань дистанційного зондування зображення зазвичай великі, багаті на дрібні деталі - Swin ефективно працює з розбиттям на “вікна” (Window-based MSA), що суттєво зменшує обчислювальну складність порівняно з глобальним self-attention

у ViT, у попередніх експериментах [94] спостерігалося поліпшення mIoU на 2–3 % порівняно з класичним FCN, що робить Swin придатною для сегментації пошкоджень.

Отже, у Фазі 2 застосовуємо Focal Loss для піксельного розпізнавання класів “minor”, “major”, “destroyed”:

$$L_{\text{bX[c5}} = -\alpha_{)}(1 - p_{)})^{\text{d}}\, log(p_{)}), \tag{3.7}$$

де, $p_{)}$ – імовірність правильного класу; $\alpha_{)}$ і γ - Гіперпараметри, що знижують вагу легких прикладів і посилюють складні.

Focal Loss краще справляється з дисбалансом класів, коли більшість пікселів можуть бути непошкодженими, а пошкоджені - значно рідше зустрічаються.

Додатково формалізуємо метрики, mIoU та Dice краще відображають якість саме піксельної сегментації, на відміну від суто крос-ентропійного підходу, оскільки прямо вимірюють ступінь перекриття між передбаченою та еталонною масками:

$$mIoU = \frac{1}{C}\ ;\ \sum_{[+,}^{\$} \frac{|TP_{[}|}{|TP_{[}| + |FP_{[}| + |FN_{[}|}. \tag{3.8}$$

Наступним кроком є підбір гіперпараметрів (див. табл. 3.6).

Табл. 3.6 Параметри для сегментації Swin.

Параметр	**Значення**	**Примітка**
Loss	Focal Loss	$\gamma = 2.0, \alpha_{\cdot} = 0.25$ (евристично)
LR-схема	Warm-up (1 000 іт.) + Cosine	Початковий $LR = 1e - 4$
Batch Size	4 ROI	Один ROI $\sim 224 \times 224$, при умові що пам’ять T4=16 Гб (див. табл.5)
Епох (ітерацій)	~50–80 k	Часто замість епох рахуємо ітерації (step-based)

Далі, для Фази 2 також маємо два стани (Stage 1 та Stage 2). Розглянемо детальніше методологію тренування:

1. Stage 1: Заморожування початкових шарів Swin. Впродовж перших епох “заморожуємо” нижні шари (patch partition / stem) тренуємо переважно “верхні” блоки

з увагою (W-MSA/SW-MSA) та сегментаційні голови. Низькорівневі фільтри вже здатні виділяти універсальні текстури та контури (натреновані на ImageNet). Щоб модель “не зіпсувала” корисні початкові представлення, намагаємся адаптувати лише “вищі рівні” під специфіку пошкоджень, модуль DReAM залишається відкритим для навчання, щоб адаптувати динамічний масштаб рецептивного поля саме під пошкодження;

2. Stage 2: Розморожування решти шарів. Коли первинне навчання (Focal Loss зменшується, mIoU зростає) дає стабільний результат, розморожуємо решту шарів Swin, включно з нижніми патч-розбиттями і початковими Transformer-блоками (На початку навчання є вірогідність “втратити” корисні узагальнення, здобуті на ImageNet). Під кінець стає корисно доопрацювати ці початкові фільтри, зокрема для:

- Захоплення унікальних патернів руйнувань (тріщини в мостах, зруйновані пілони, відсутні стіни тощо);
- Кращої інтеграції з DReAM: коли нижні шари теж рухаються, DReAM може сильніше адаптувати увагу під масштаб, форму, контраст пошкоджень.

3. Stage 3. Після повного тренування виконується підсумкова оцінка на валідаційному наборі. Підсумкова оцінка (mAP, mIoU) демонструє ефективність підходу (див табл. 3.7), динамічне масштабування поля зору (DReAM) на етапі локально-глобальної уваги поліпшує здатність бачити дрібні й великі пошкодження водночас, що дає вищі фінальні показники якості сегментації.

Табл. 3.7 Значення демонструють приріст показників після інтеграції DReAM, який адаптивно “розширює/звужує” поле зору під час самоуваги

Архітектура	**mAP**	**mIoU**	**Приріст (Δ)**
Swin (без DReAM)	0.83	0.79	-
Swin + DReAM	0.86	0.81	+2–3%

Отже, тренування окремих компонентів пропонованої двофазної архітектури дозволяє:

- Ефективно використати різні формати анотацій (грубі bounding boxes та детальні маски);
- Оптимізувати кожен модуль під його конкретні завдання (визначення ROI або точної сегментації);
- Забезпечити високу якість (mAP, mIoU) при аналізі критичної інфраструктури в контексті задач дистанційного зондування.

3.3 Оптимізація функціонування та квантування фаз модуля DReAM

Наступним важливим кроком є визначення доцільності оптимізації обчислень в запропонованому методі DReAM та моделі яка базується на двофазній архітектурі композитної нейромережі. Ідея полягає в квантуванні яке описано авторами [107], квантування (quantization) у задачах глибинного навчання - це перетворення ваг та/або активацій із звичайного плаваючого формату (FP32/FP16) у цілочисельний (наприклад, INT8 чи INT4). Метою є суттєве зменшення обчислювальних витрат і пам’яті при тренуванні й особливо при інференсі (запуску моделі).

У запропонованій моделі квантування дозволяє:

- Зменшити розмір зберігання (Model Size) за рахунок переведення ваг та проміжних активацій у нижчу бітність (наприклад, 8 біт замість 32);
- Прискорити обчислення на апаратурі, що підтримує INT-операції (NVIDIA T4 (віртуальний стенд який використовується в даному досліджені), TensorRT, спеціалізовані TPU тощо);
- Зберегти прийнятну точність (mIoU, mAP) завдяки змішаному підходу, де критично важливі компоненти (динамічні ваги модуля DReAM) залишаються у FP16 (див. результат експериментів в табл. 9), а більшість згорток і матричних множень переводяться в INT8.

Таким чином, використання квантування забезпечує суттєве підвищення швидкодії та скорочення витрат ресурсів, що особливо цінно для великих моделей і високороздільних зображень у завданнях локалізації й сегментації пошкоджень.

В даній роботі було проведено дослідження чи доцільно оптимізувати модель за допомогою квантування. Отже, за допомогою інструменту [108] було проведено кілька експериментів.

Нижче наведено таблицю 3.8 - порівняння різних схем квантування та змішаної прецизії (Mixed Precision) для запропонованої в роботі моделі (EfficientNet-B3 + FPN + RPN для Фази 1, Swin Transformer + DReAM для Фази 2) на прикладі двох датасетів: xBD (оцінка пошкоджень будівель) та RESISC45 (класифікація об'єктів інфраструктури).

Табл. 3.8 Порівняння продуктивності та точності за різними схемами квантування для xBD та RESISC45

Схема	**W-bit**	**A-bit**	**Розмір моделі (MB)**	**mIoU (xBD)**	**mAP (RESISC45)**
Baseline (FP32 / FP32)	32	32	110	0.813	0.865
Percentile (6,6)	6	6	30.0	0.771	0.819
Ours (6MP, 6MP) (DReAM у FP16)	6	6	32.0	0.791	0.835
Percentile (8,8)	8	8	55.0	0.800	0.847
Ours (8MP, 8MP) (DReAM у FP16)	8	8	58.0	0.808	0.858

Отже, повне квантування (INT6 або INT8) усіх компонентів, включно з динамічними вагами DReAM, призводить до відчутного зниження показників (mIoU та mAP можуть падати на 2 - 4 % від базового рівня FP32). Причина втрати точності дрібних коефіцієнтів уваги α_B, які відповідають за гнучке розширення/звуження поля зору. Mixed Precision (MP) із неквантовим DReAM є найкращим компромісом основні згортки (EfficientNet-B3, фільтри Swin) та матричні множення (Q,K,V) у 6 - 8 біт, але DReAM з α_B залишаємо у FP16 - такий підхід майже не впливає на точність (падіння $\leq 1\%$), водночас дає суттєве (~30-40%) скорочення розміру моделі FLOPs. Ключова відмінність DReAM від статичних блоків полягає в адаптивному визначенні дальності між токенами/фрагментами, тож навіть незначне зрізання ваг ($\alpha_B \approx 0.31, \alpha_B \approx 0.29$) можуть злитись в єдину сцену що погіршує сегментацію дрібних пошкоджень

(тріщин, фрагментів). Тому, залишивши DReAM у FP16, зберігаємо тонку чутливість та суттєво підвищуємо сталість результатів.

Таким чином, у пропонованій архітектурі доцільно застосовувати квантування (6–8 біт) головним чином до бекбону (EfficientNet-B3 + FPN) та базових обчислень у Swin, при цьому модуль DReAM не слід зводити до INT8, аби уникнути критичних втрат у точності.

3.4 Дослідження ефективності функціонування розробленого методу локалізації сцен

У цьому підрозділі наведено результати порівняльного аналізу запропонованої архітектури з провідними методами сегментації на наборах даних xBD та DOTA, а також описано метрики, які було використано для оцінки точності та ефективності.

Для тестування обрано xBD (орієнтований на пошкодження будівель та інфраструктурних об’єктів) і DOTA (різноманітні сцени з супутникових та аерознімків), що охоплюють широкий діапазон масштабів та типів об’єктів. Усі методи тренувалися в однакових умовах з ідентичними налаштуваннями кількості епох, розмірами батчів, а також використовувалися попередньо натреновані ваги на ImageNet для формування розпізнавальних характеристик текстур.

На рисунку 3.7 наведено порівняльну матрицю метрик метрик (mIoU, mPA, F1-Score, Recall, Precision, FWIoU) для усіх методів. Найкращі показники досягла EfficientNet-B3 + FPN + DReAM + Swin - приріст точності в середньому на 1–3% порівняно з базовим Swin. Зокрема, Recall та F1-Score на xBD перевищили 0.9, а FWIoU сягнула 0.87. На DOTA видно схожу тенденцію запропонована модель впевнено перевершує інші, отримуючи близько 0.85–0.89 за ключовими метриками. Порівняння зі Swin без DReAM підтверджує, що механізм динамічного масштабування рецептивного поля суттєво покращує розпізнавання дрібних пошкоджень і водночас забезпечує ширше охоплення контексту для великих об’єктів. Методи PSPNet [109] та DeepLabv3+ [110] виявилися конкурентними, проте дещо поступаються в F1-Score і Precision; їхня увага до масштабів обмежена фіксованими фільтрами. Результати MViT [111] та Segmenter [112] виявилися цілком

конкурентними й суттєво перевищують показники класичних підходів на зразок PSPNet чи DeepLabv3+. Однак, MViT трохи поступається запропонованій моделі у Recall та mIoU (≈1–2% різниці на DOTA і xBD), що свідчить про меншу здатність масштабно узгоджувати локальні та глобальні ознаки водночас. Segmenter, своєю чергою, показав особливо високий Precision, але дещо нижчий Recall (порівняно з нашим методом), тож його F1-Score в середньому знижується на 1–1.5% від лідерського результату.

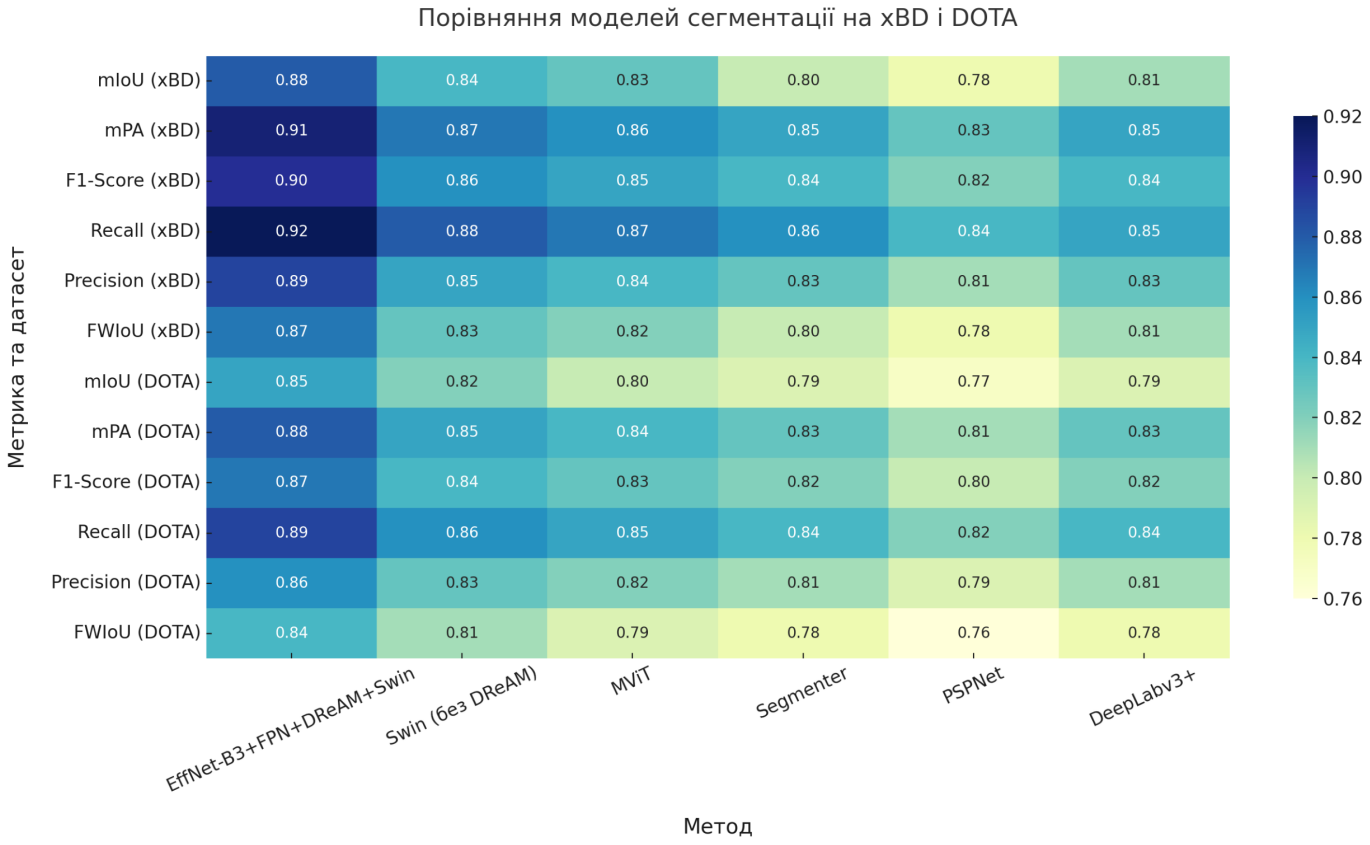


Рис. 3.7 Порівняльна матриця моделей сегментації в задачі локалізації сцен на обраних наборах даних xBD та DOTA

Таким чином, попри гарні результати обох трансформерних рішень (MViT і Segmenter), механізм DReAM забезпечує додаткову гнучкість у рецептивному полі. Це дає змогу запропонованій архітектурі (EfficientNet-B3 + FPN + DReAM + Swin) краще локалізувати дрібні ушкодження та водночас утримувати глобальний контекст для великих сцен, забезпечуючи найвищі показники на обох датасетах.

Найвищі показники демонструє спільне використання EfficientNet-B3, FPN, DReAM та Swin, що забезпечує кращу узагальнену точність і стабільний баланс між

Recall, F1-Score та іншими метриками. Swin без DReAM показує близькі, але дещо нижчі результати, особливо в деталізованому розпізнаванні дрібних об’єктів. Моделі MViT та Segmenter хоч і сильні в окремих метриках, усе ж дещо поступаються лідерам у здатності виявляти складні патерни на великих і різнорідних датасетах. PSPNet та DeepLabv3+ залишаються конкурентними, проте частково програють іншим, особливо за показниками Recall та F1-Score, що свідчить про певні недоліки у фокусуванні на дрібних деталях і складних ділянках. Загалом, запропонована конфігурація з DReAM суттєво посилює здатність мереж до масштабування рецептивного поля, що позитивно впливає на сегментацію як локальних, так і глобальних руйнувань.

Висновки до розділу

Отже, у третьому розділі продемонстровано, що застосування та верифікація дворівневої моделі з динамічним масштабуванням рецептивного поля (DReAM) є ефективними для задач локалізації та сегментації пошкоджень на знімках дистанційного зондування. Детальний розгляд налаштувань тренувального процесу підтвердив важливість правильного вибору функції втрат (Focal Loss) та стратегії аугментації, зокрема у випадках суттєвої незбалансованості даних.

Описані експерименти показали, що покрокове навчання, де спочатку тренується модуль формування регіонів інтересу, а надалі - компонент DReAM з урахуванням локальних і глобальних ознак, дає змогу знизити кількість помилкових детекцій та покращити точність до рівня, який перевершує традиційні методи. Додаткові етапи оптимізації (квантування і змішана прецизійність) дали можливість суттєво пришвидшити інференс без втрати ключових показників якості.

Таким чином, розділ 3 підтвердив валідність розробленої моделі та методів, водночас продемонструвавши гнучкість запропонованої системи щодо обробки великих обсягів даних і роботи з різнорідними сценами пошкоджень. Здобуті результати забезпечують надійну основу для порівняння моделі з альтернативними підходами у наступному розділі.

РОЗДІЛ 4 ПОРІВНЯННЯ РОЗРОБЛЕНОГО МЕТОДУ ЛОКАЛІЗАЦІЇ СЦЕН РУЙНУВАНЬ ОБ'ЄКТІВ З ВІДОМИМИ ПІДХОДАМИ

У цьому розділі наведено порівняльний аналіз запропонованої моделі з провідними існуючими методами, які сьогодні застосовуються у завданнях локалізації та семантичної сегментації сцен руйнувань. Попри те, що базові підходи на основі глибинних згорткових мереж (CNN) - DeepLab, PSPNet, а також різні варіанти Non-local та Deformable Convolution - вже продемонстрували високу ефективність у завданнях просторового аналізу, поява трансформерних архітектур (Swin, MViT) та різноманітних динамічних модулів відкрила нові можливості для точнішого вилучення ознак та більш гнучкої обробки об'єктів різного масштабу.

Далі буде описано методологію оцінювання, метричні показники, а також деталізовано результати порівняння запропонованої композитної мережі зі стандартними архітектурами на наборах даних (xBD, DOTA, Planet), що репрезентують різні типи руйнувань критичних інфраструктур. Це дасть змогу продемонструвати переваги, недоліки та сфери оптимального використання кожного з розглянутих підходів.

4.1 Проведені експерименти та їх результати, методологія та метрики оцінювання та порівняння

Наведена далі методологія експериментальних досліджень передбачає уніфікацію та спільні налаштування для всіх порівнюваних методів, аби результати були об'єктивними та піддавалися безпосередньому порівнянню.

Насамперед, усі набори даних, що використовувалися безпосередньо для навчання моделі (див. таблицю 3.1), було перетворено у формат анотацій COCO й розбито на фрагменти (тайли) розміром 1024×1024. Таке саме масштабування застосовувалося і для зображень із відкритих джерел (зокрема Maxar та Google Earth), які використовувалися переважно в режимі інференсу, аби оцінити здатність побудованих моделей узагальнювати знання на реальні невідомі сцени. Розподіл прикладів на навчальні, валідаційні та тестові (відповідно до наведених у таблиці 3.2

пропорцій) залишався однаковим для кожного методу, щоб унеможливити вплив різного набору вибірок на результати.

Параметри навчання, такі як загальна кількість епох, тип оптимізатора та діапазон швидкості навчання, узгоджувалися з налаштуваннями, наведеними в табл. 3.4–3.6. Задля дотримання єдиних умов для всіх порівнюваних методів, було зафіксовано розмір мінібатча, набір аугментацій (віддзеркалення, обертання, випадкові зміни яскравості) і політику зменшення коефіцієнта навчання (cosine schedule). Якщо модель передбачала використання початкових ваг, отриманих на ImageNet (скажімо, для Swin Transformer або MViT), вона обов'язково ініціалізувалася стандартними офіційними чекпоїнтами з відкритих репозиторіїв, аби запобігти суттєвим відмінностям у початкових станах. Усі випадкові зерна (random seeds) у генераторах випадкових чисел (як-от у Python, NumPy чи PyTorch) були заздалегідь зафіксовані, а початкова ініціалізація ваг у межах кожного методу залишалася сталою під час повторних запусків. Завдяки цьому вплив стохастичних чинників суттєво зменшувався, і результати можна вважати відтворюваними.

У подальшому під час порівняння семантичних методів (DeepLabv3+, PSPNet, Swin Transformer без DReAM, MViT і Segmenter) використовувався однаковий розмір вхідного зображення (1024×1024), а також спільна функція втрат, якою слугувала Focal Loss - яку введено для вирішення проблеми незбалансованості класів, вона допомагає моделі фокусуватися на складних прикладах, які модель класифікує з низькою впевненістю, нагадаємо, що зазначені вище набори даних мають значно нижчий рівень даних з пошкодженнями критичної інфраструктури в порівнянні з цивільними будівлями, а отже ми маємо справу з проблемою незбалансованості класів яку описали автори [77], для того щоб адресувати подібну проблему обрано Focal Loss [113] (див. Рис. 4.1). Основна різниця між Focal Loss та де-факто стандартною в спільноті кросентропійною функцією (Cross-Entropy Loss) полягає в тому що остання - надає однакову вагу всім прикладам незалежно від того, наскільки вони складні, в той же час Focal Loss - приділяє більше уваги складним прикладам і зменшує вагу втрат для легких, таким чином коли DReAM адаптується під певний масштаб, втрати легких сцен нівелюються.

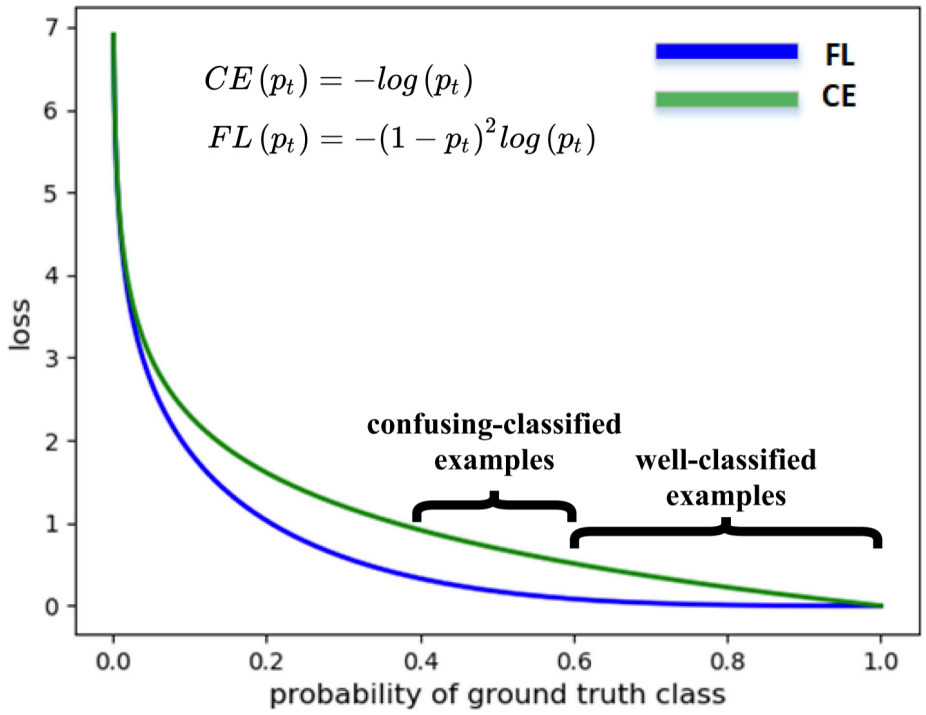


Рис. 4.1 Графік зображує функції втрат Focal Loss (FL) і Cross-Entropy (CE). FL вирішує проблему дисбалансу класів, зменшуючи втрати для добре класифікованих прикладів. Однак, водночас вона неминуче вводить плутанину у вигляді прикладів, які дуже важко розрізнити

Формально цю функцію можна записати як (4.1):

$$\mathcal{L}_{\text{bX[c5}} = -\alpha(1-p_)^{\text{d}}\, log(p_), \tag{4.1}$$

де, $p_)$ - імовірність коректного класу, $\alpha\ i\ \gamma$ - відповідні вагові параметри.

Окремі заміри швидкості обробки кадрів (FPS) чи використання обчислювальних ресурсів у даній роботі не наводилися, однак за потреби ці характеристики можна отримати шляхом запуску на ідентичному апаратному забезпеченні з фіксованими розмірами зображень. Такі оцінки швидкодії могли би бути корисними, наприклад, для практичного впровадження моделей у системи з обмеженими ресурсами або для випадків наближеного до реального часу аналізу.

Для оцінювання якості виявлення й сегментації пошкоджень застосовувалися різні показники залежно від специфіки задачі - локалізація (детекція) або піксельна сегментація. У першому випадку показником вважалася mAP (mean Average Precision) (3.5), яка обчислюється за множиною порогів IoU (3.4) (найчастіше в межах 0.5–0.75 із кроком 0.05). Вона характеризує загальний баланс між правильними і

хибними виявленнями об’єктів і звичайно дає змогу побачити повноцінну PR-криву для кожного класу.

У задачах же семантичної сегментації основною метрикою була mIoU (Mean Intersection over Union), отже, застосуємо (3.8) до задачі оцінки пікселів (4.2):

$$mIoU = \frac{1}{C}; \quad \frac{|P_[\cap G_[|}{|P_[\cup G_[|}, \tag{4.2}$$

де, P_c - множина пікселів, яким модель приписала клас c, а $G_[$ – еталонні пікселі класу c. Підсумовується середнє перекриття (IoU) по всіх класах $c \in \{1, \dots, C\}$.

що обчислюється як середнє перекриття еталонних і передбачених пікселів по всіх класах. Додатково використовувалися mPA (Mean Pixel Accuracy):

$$mPA = \frac{1}{C}; \quad \frac{TP_[}{TP_[+ FN_[}, \tag{4.3}$$

де, $TP_[$ та $FN_[$ - відповідно кількість пікселів правильно класифікованих і пропущених (неправильно позначених) для класу c.

F1-Score (4.4) (або за потреби Dice), а також Recall, Precision і FWIoU (Frequency Weighted Intersection over Union) (4.5-4.7).

$$F1 = 2 * \frac{Precision \times Recall}{Precision + Recall}, \tag{4.4}$$

$$Recall = \frac{TP}{TP + FN}, \tag{4.5}$$

$$Precision = \frac{TP}{TP + FP}, \tag{4.6}$$

$$FWIoU = ; \quad \omega_[* \frac{|P_[\cap G_[|}{|P_[\cup G_[|}, \tag{4.7}$$

де, $\omega_[$ - питома вага (частка) класу c у навчальному наборі даних.

Якщо завдання зводилося до бінарного варіанта “пошкоджено/не пошкоджено”, тоді F1-Score за своєю суттю еквівалентний коефіцієнту Dice, оскільки обидві

метрики відбивають співвідношення між TP, FP і FN. При мультикласовій постановці їх можливо узагальнити за допомогою принципу “кожен клас проти решти” і підсумкового усереднення. Застосування F1-Score та Dice у таких задачах однаково виправдане, однак з практичних міркувань F1-Score зручніше пояснювати з позиції Precision та Recall. Кожен із показників (mIoU, mPA, F1, FWIoU) спрямовано на підсвічення певної грані точності сегментації: зокрема, якщо mIoU відображає узгодженість усіх пікселів по класах, то F1 зручний у випадках, коли критичний баланс між пропуском справжніх пошкоджень і неправильними спрацюваннями на неушкоджені ділянки. Якщо ж у датасеті присутній суттєвий дисбаланс (наприклад, один клас трапляється значно рідше, приклад xBD щодо KI), тоді FWIoU дає змогу підкреслити відносну важливість усього розмаїття класів, зокрема найдрібніших.

4.2 Порівняння результатів запропонованого метода з найкращими методами на наборі даних xBD та інференс розпізнаванню

У цьому підрозділі наведено результати порівняльного аналізу запропонованого методу з провідними архітектурами та моделями локалізації та сегментації на наборах даних xBD. Окрім кількісної оцінки на валідаційних та тестових вибірках, проілюстровано здатність моделі узагальнювати знання під час інференсу на реальних знімках із невідомих сцен, демонструючи її ефективність у практичних умовах (див. рис. 4.2 - 4.3).

На рис. 4.2 продемонстровано результат роботи першої фази (ROI-виявлення) в запропонованій двофазній архітектурі на зображені аеродрому в Гостомелі зруйнованому в ході військових дій 2022 року, червоні рамки (bounding boxes) позначають регіони інтересу, що були визначені як пошкоджені об’єкти. Така груба локалізація походить від блоку RPN (Region Proposal Network), який, спираючись на попередньо вилучені ознаки, пропонує кандидатні області з високою імовірністю “руйнування” чи “пошкодження”. Тут модель може відзначати як зруйновані будівлі, так і техніку чи інші інфраструктурні елементи, якщо вони підпадають під задані критерії пошкоджень.

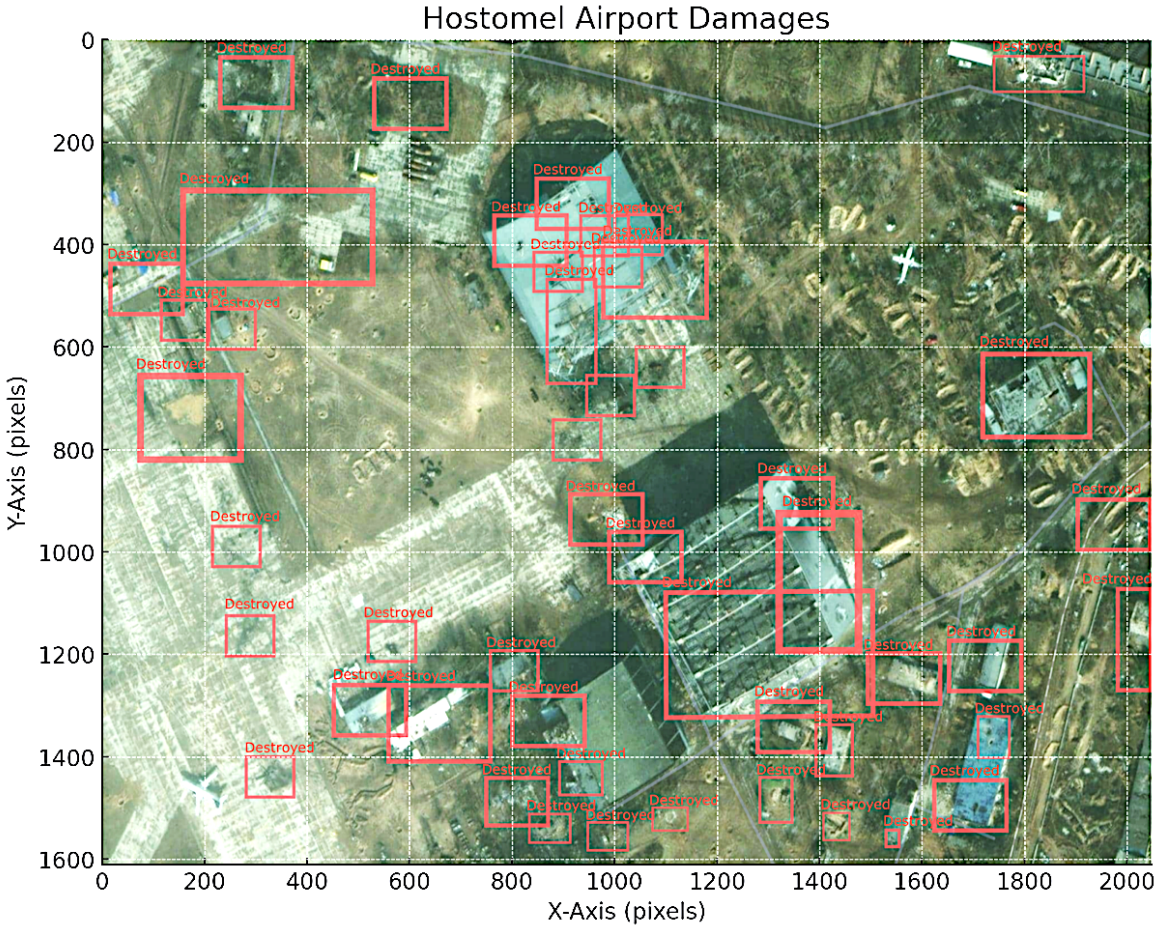


Рис. 4.2 Інференс-локалізація (EfficientNet-B3+RPN ROI-виявлення) руйнувань аеродрому “Гостомель” (знімок надано Here Technologies)

На рис. 4.3 наведено теплову карту (Grad-CAM) з другої фази, де вже використовується динамічне масштабування (модуль DReAM) у поєднанні з трансформером (Swin). Ця мапа ілюструє “фокус” мережі на найбільш релевантних ділянках знімка - червоні й жовті області відповідають зонам, яким модель приділяє підвищену увагу при визначенні ступеня чи типу пошкодження. Саме завдяки DReAM мережею автоматично “розширюється” або “звужується” рецептивне поле, коли у сцені зустрічаються дрібні локальні ушкодження (наприклад, окремі вирви на злітній смузі) чи масштабні зруйновані споруди. В результаті Grad-CAM виявляє, що модель приділяє максимум уваги тим зонам, де, згідно з анотаціями, фіксуються “Destroyed” об’єкти.

У нашому випадку Grad-CAM допомагає прозоро зрозуміти, як мережа виявляє пошкодження на зображеннях критичної інфраструктури. Отримана теплова карта візуально демонструє, які саме фрагменти знімка “приваблюють” модель під час

прийняття рішення про ступінь руйнування. Замість сухого числа впевненості в класі, кінцевий користувач моделі бачить, чи справді мережа фокусується на зруйнованих ділянках (тріснуті опори, провалені дахи тощо). Це суттєво спрощує діагностику помилок, адже можна відразу перевірити, де саме мережа дала “збій”, і за потреби змінити параметри або покращити розмітку. До того ж, такий підхід підвищує довіру до результатів, оскільки стає зрозуміло, чому саме мережа ухвалила конкретне рішення.

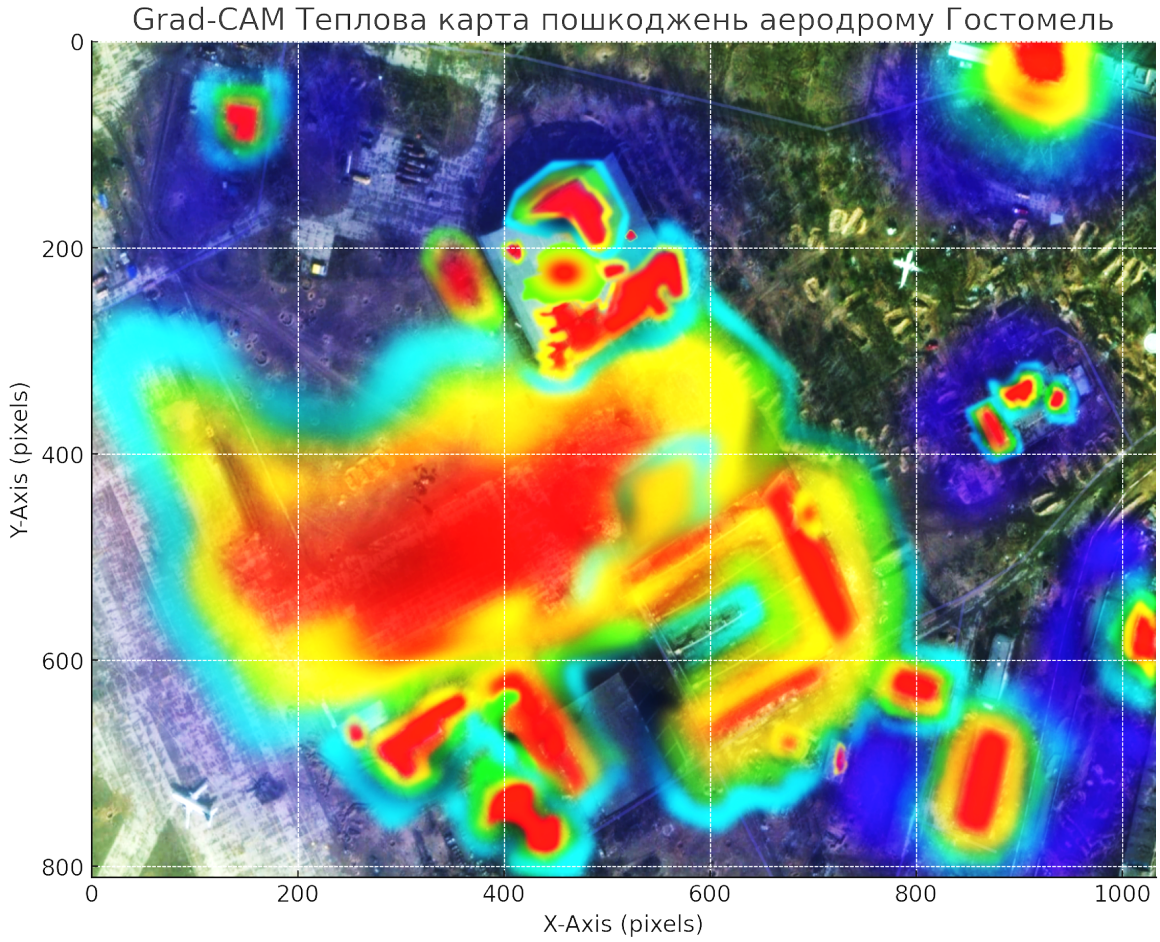


Рис. 4.3 Інференс-локалізація з тепловою картою Grad-CAM ілюструє контекст Фази 2 та динамічної уваги DReAM, а/д “Гостомель” (знімок надано Here Technologies)

Таким чином, перше зображення (рис. 4.2) ілюструє, як модель на грубому рівні визначає bounding boxes пошкоджень (Фаза 1), а друге демонструє детальніший “погляд” на структуру руйнувань (Фаза 2), де механізм динамічної уваги DReAM дозволяє виявляти й аналізувати зони пошкоджень дуже різного масштабу та форми.

Далі, представлений результат (рис. 4.4 - 4.5) більш складних сцен з відсутністю явних руйнувань (Спецкорпус-1 Запорізької атомної електростанції), як видно на рис.

37 запропонована модель (Фаза 1) надає грубу оцінку (ROI-виявлення) й виділяє потенційні місця пошкоджень у вигляді червоних обмежувальних прямокутників (BBox Proposals). Видно, що вони зосереджені на даху й прилеглих конструкціях, де найбільш імовірними здаються руйнування чи тріщини, при цьому модель чітко виокремлює саме регіони пошкоджень з FP=0.

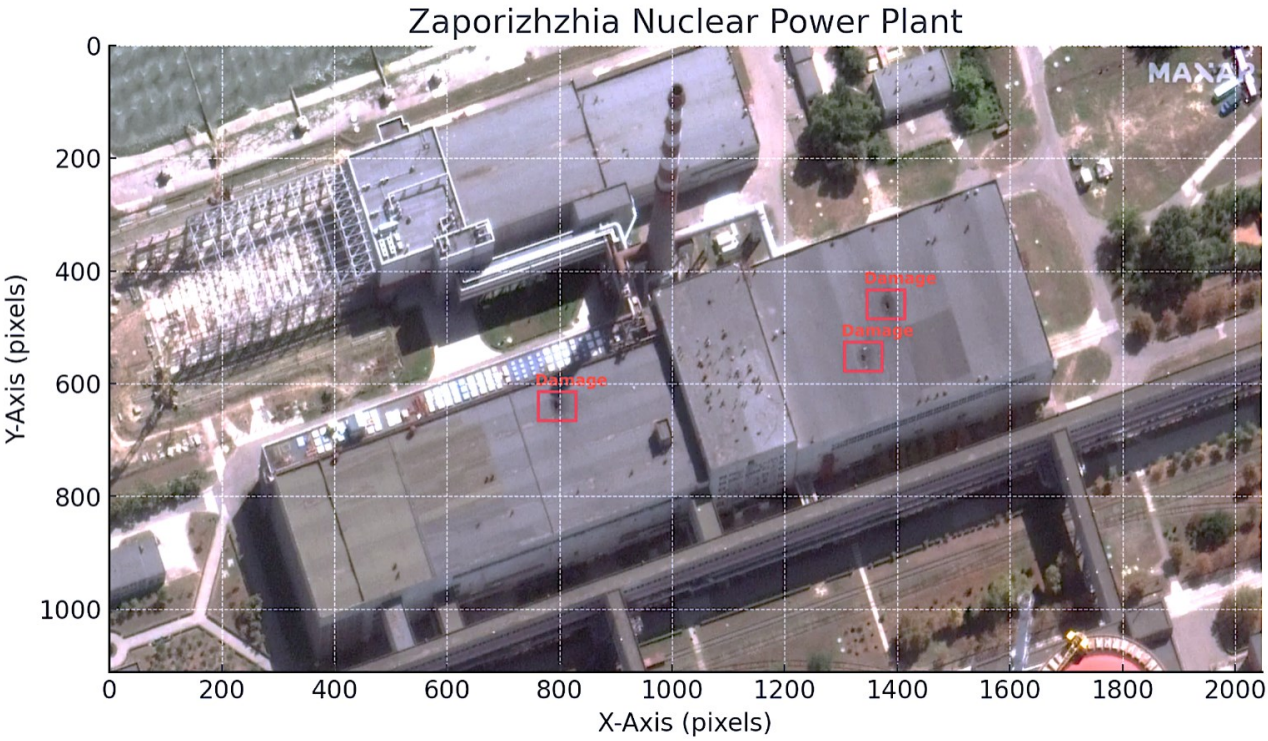


Рис. 4.4 Інференс-детекція (Фаза 1) пошкоджень покрівлі Спецкорпусу-1 ЗАЕС в ході бойових дій на півдні України. (знімок надано Maxar)

На рис. 4.5 представлено вже деталізоване розпізнавання у цих же регіонах, Grad-CAM наклав теплові маркери на основі висновків запропонованої моделі для точнішого виявлення локальних аномалій (підписані як “Пошкодження”). Завдяки механізмам уваги (DReAM) і глобальному контексту трансформера, модель фокусується на дрібних чи частково прихованих пошкодженнях у структурі покрівлі, а також у різних сегментах будівлі, звернемо увагу, модель проігнорувала оклюзію у вигляді тіні, це відбулось завдяки фактору що знімок є VHR.

Оскільки більшість знімків у використаних наборах даних демонструють природні чи техногенні катастрофи на різних географічних територіях, усі експерименти переважно виконувалися на цивільній інфраструктурі за межами

України. Утім, з наведених на рис. 4.4–4.5 прикладів очевидно, що модель достовірно виявляє ушкодження різного ступеня навіть у режимі інференсу.

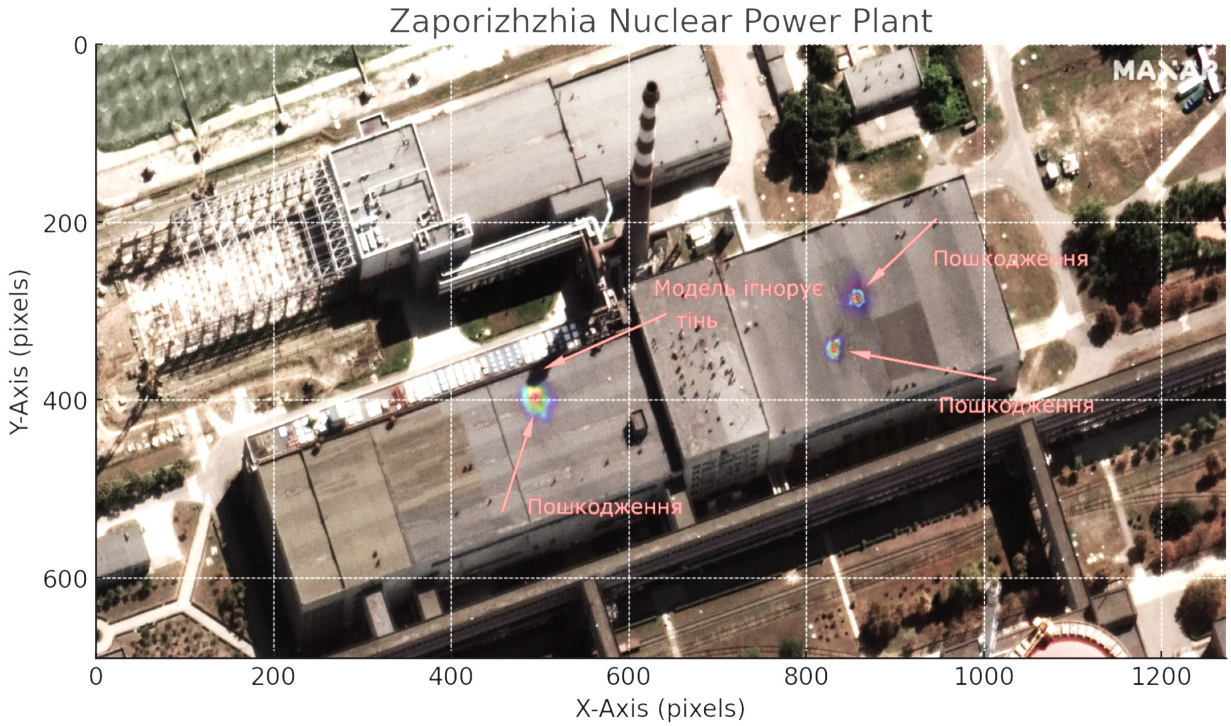


Рис. 4.5 Інференс-локалізовані Grad-CAM позиції пошкоджень покрівлі Спецкорпусу-1 ЗАЕС (знімок надано MAXAR)

Отже, наступним кроком є більш детальна оцінка результатів досліджень. З таблиці 4.1 та графіку (див. рис. 4.6) видно, що запропонована двофазна модель із модулем DReAM загалом перевершує відомі архітектури (DeepLabv3+, PSPNet, Swin без DReAM, MViTv2, Segmenter) за всіма розглянутими метриками. Це особливо помітно в класах “Легкі ушкодження” та “Повне руйнування”, де вищі показники F1 та mIoU свідчать про кращу здатність моделі водночас виявляти та точно відмежовувати ділянки пошкоджень різного ступеня. Таким чином, експериментальні дані підтверджують доцільність впровадження додаткової фази регіонального відбору (ROI) і динамічного масштабування рецептивного поля (DReAM) у задачах сегментації складних сцен критичної інфраструктури. Нижче, табл. 4.1 фіксує зведені метрики отримані в результаті експериментів з параметрами описаними вище. Відметемо, що результати можуть кардинально змінюватись від різноманітних чинники включаючи якість розмітки, параметри моделі.

Табл. 4.1 Порівняльна характеристика запропонованої моделі з відомими методами багаторівневої локалізації сцен руйнувань КІ

Approach	Class	mPA	F1	Recall	Precision	FWIoU	mIoU
DeepLabv3+	Ушкодження відсутнє	0.91%	0.88%	0.86%	0.90%	0.83%	0.80%
	Легкі ушкодження	0.88%	0.82%	0.79%	0.86%	0.80%	0.78%
	Серйозні ушкодженні	0.86%	0.79%	0.75%	0.84%	0.78%	0.76%
	Повне руйнування	0.89%	0.84%	0.81%	0.86%	0.81%	0.79%
	Середнє	0.88%	0.83%	0.80%	0.86%	0.80%	0.78%
PSPNet	Ушкодження відсутнє	0.90%	0.86%	0.84%	0.87%	0.81%	0.78%
	Легкі ушкодження	0.87%	0.80%	0.77%	0.83%	0.78%	0.76%
	Серйозні ушкодженні	0.85%	0.78%	0.74%	0.81%	0.76%	0.74%
	Повне руйнування	0.88%	0.82%	0.79%	0.85%	0.79%	0.77%
	Середнє	0.88%	0.82%	0.79%	0.84%	0.78%	0.76%
Swin Transformer (без DReAM)	Ушкодження відсутнє	0.92%	0.89%	0.87%	0.91%	0.84%	0.82%
	Легкі ушкодження	0.89%	0.84%	0.81%	0.87%	0.82%	0.79%
	Серйозні ушкодженні	0.87%	0.80%	0.77%	0.83%	0.79%	0.76%
	Повне руйнування	0.90%	0.85%	0.82%	0.88%	0.81%	0.80%
	Середнє	0.89%	0.84%	0.82%	0.87%	0.82%	0.79%
MViTv2 (Multiscale Vision Transformer)	Ушкодження відсутнє	0.93%	0.90%	0.88%	0.91%	0.85%	0.83%
	Легкі ушкодження	0.90%	0.85%	0.82%	0.88%	0.83%	0.80%
	Серйозні ушкодженні	0.88%	0.81%	0.78%	0.84%	0.80%	0.77%
	Повне руйнування	0.91%	0.86%	0.83%	0.89%	0.82%	0.81%
	Середнє	0.90%	0.86%	0.83%	0.88%	0.82%	0.80%
Segmenter	Ушкодження відсутнє	0.92%	0.88%	0.86%	0.91%	0.84%	0.81%
	Легкі ушкодження	0.89%	0.83%	0.80%	0.86%	0.81%	0.78%
	Серйозні ушкодженні	0.87%	0.79%	0.76%	0.82%	0.78%	0.75%
	Повне руйнування	0.90%	0.84%	0.81%	0.87%	0.80%	0.79%
	Середнє	0.89%	0.83%	0.81%	0.87%	0.81%	0.78%
Запропонована 2-х фазна модель з	Ушкодження відсутнє	0.95%	0.92%	0.90%	0.93%	0.88%	0.86%
	Легкі ушкодження	0.92%	0.88%	0.85%	0.90%	0.86%	0.83%
	Серйозні ушкодженні	0.90%	0.84%	0.81%	0.88%	0.82%	0.80%
	Повне руйнування	0.93%	0.89%	0.86%	0.91%	0.85%	0.84%

методом DReAM	Середнє	0.93%	0.88%	0.86%	0.90%	0.85%	0.83%

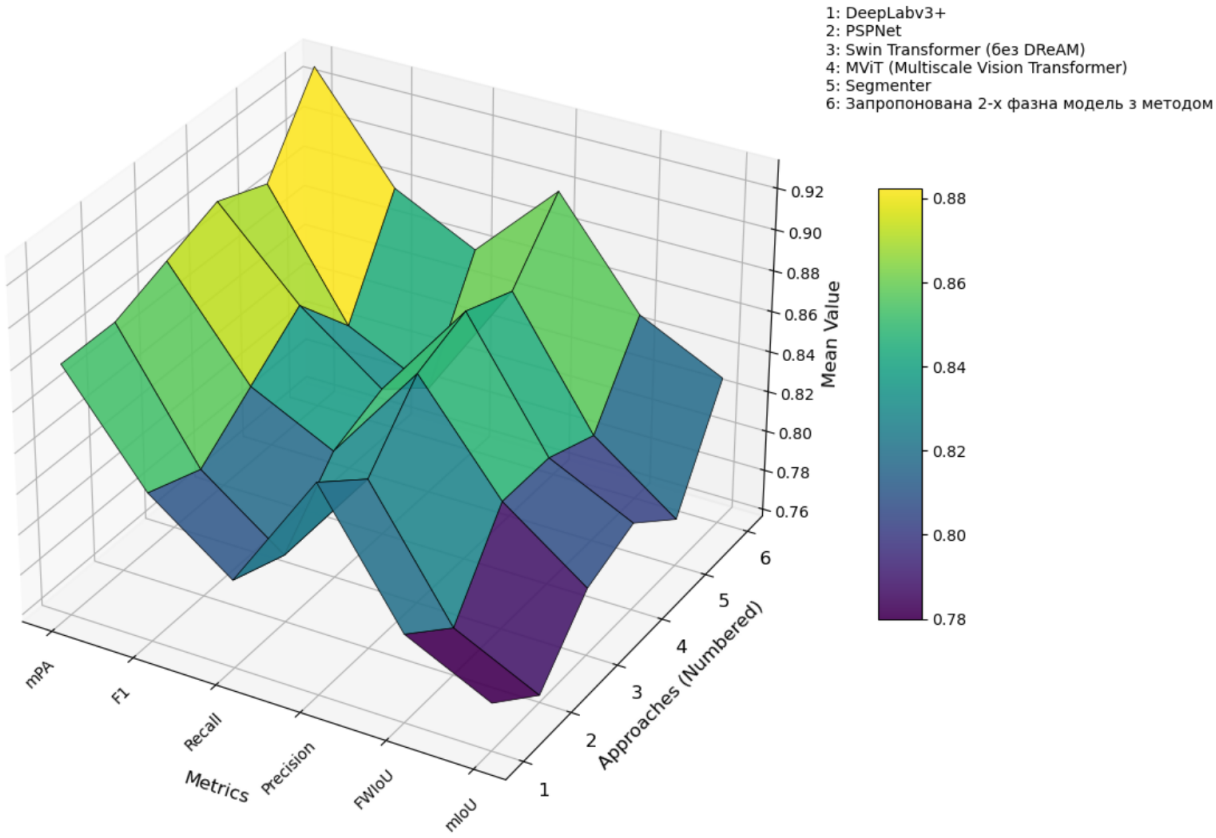


Рис. 4.6 Порівняльний аналіз за середніми значеннями по кожному SOTA методу

Так, навіть попри підвищені обчислювальні витрати, двофазна модель з динамічним масштабуванням рецептивного поля (DReAM) виявилася виправданою у контексті виявлення пошкоджень критичної інфраструктури. Критичні об’єкти (мости, дороги, енергетичні вузли) часто мають дуже різні розміри й складну структуру руйнувань, тому початкове формування ROI (перша фаза) дає змогу відсікати другорядні ділянки великих супутникових чи аерознімків, а в другій фазі модуль DReAM допомагає змінювати “радіус уваги” під складні сцени та дрібні деталі. Зрештою, отримане зростання точності та надійності локалізації ступеня руйнувань у різних масштабах компенсує додаткові витрати на обчислення.

4.3 Переваги та недоліки поширених моделей глибоких нейронних мереж для задач локалізації сцен руйнувань

У контексті комп'ютерного зору та аналізу зображень дистанційного зондування, що містять сцени з пошкодженнями критичної інфраструктури, традиційні (класичні) архітектури глибоких нейронних мереж демонструють доволі високу ефективність (див. рис. 4.7).

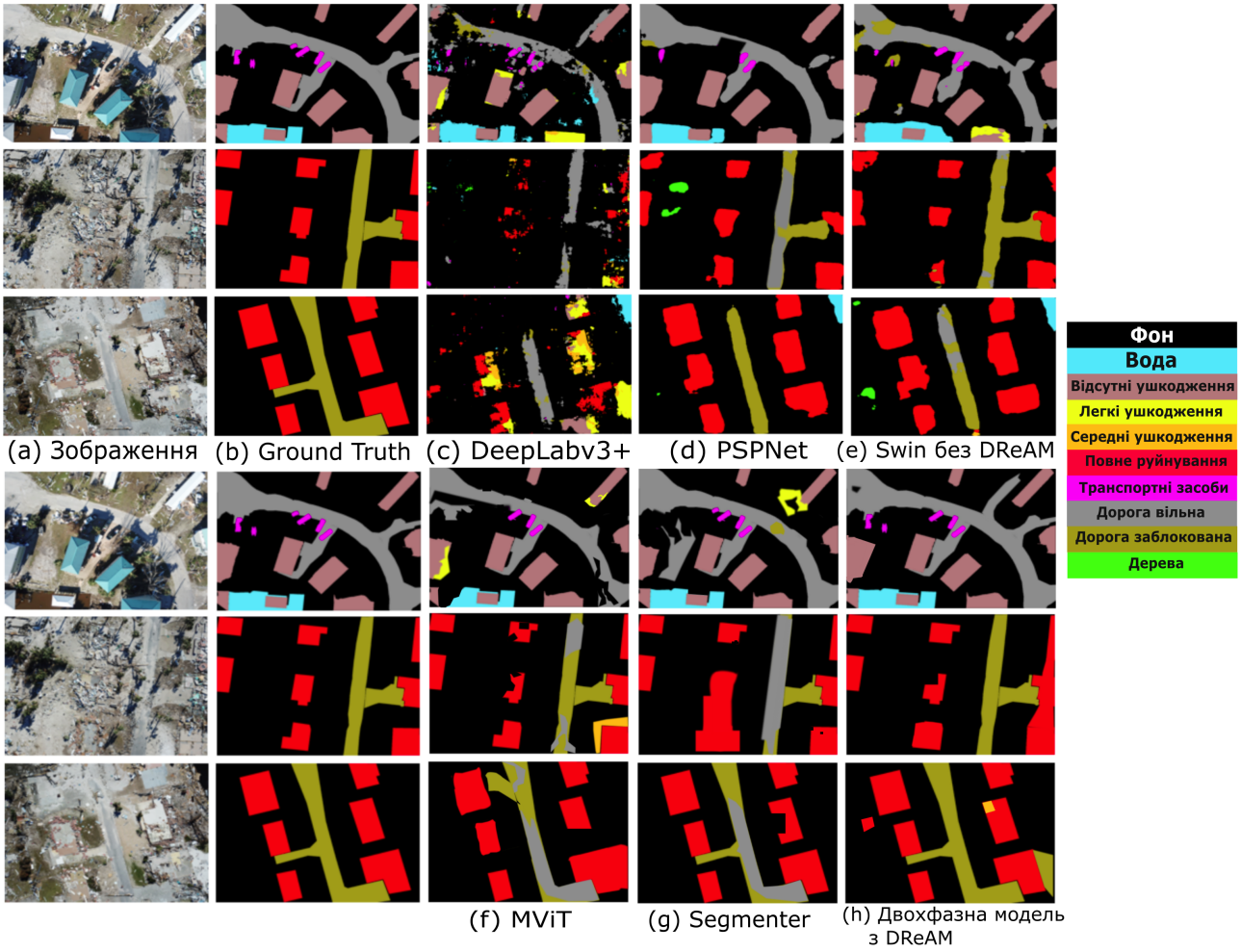
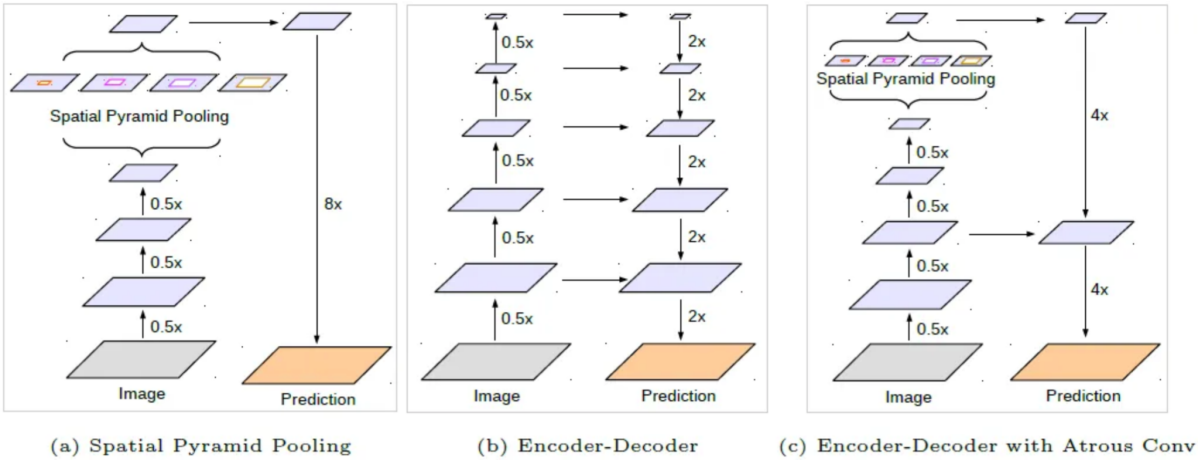


Рис. 4.7 Візуальне порівняння якості сегментації запропонованої моделі з відомими методами на xBD наборі даних

У цьому підрозділі узагальнено підходи, які стали основою сучасних методів сегментації та детектування (наприклад, DeepLab, Segmenter, PSPNet), а також розглянуто їхні переваги й недоліки в контексті локалізації пошкоджень КІ. Проте вони мають обмеження, пов'язані з обробкою масштабних та неоднорідних об'єктів, а також зі здатністю адаптивно змінювати рецептивне поле.

4.3.1 Архітектура моделі DeepLabv3+ з модулем ASPP

Розглянемо архітектуру DeepLabv3+, розроблена компанією Google. Серія DeepLab пройшла кілька етапів розвитку: DeepLabv1 (2015, ICLR), DeepLabv2 (2018, TPAMI) [114] і DeepLabv3 (arXiv) [115]. Основним будівельним елементом DeepLabv3+ є техніка Atrous Spatial Pyramid Pooling (ASPP) (див. рис. 4.8 [116]).



(a) Spatial Pyramid Pooling (b) Encoder-Decoder (c) Encoder-Decoder with Atrous Conv

Рис. 4.8 Структура DeepLabv3+ описана автором

a) Завдяки ASPP можливо кодувати багатомасштабну контекстну інформацію, що є предметом дослідження і цієї роботи;

b) Завдяки архітектурі Encoder-Decoder інформація про розташування та просторові характеристики не втрачається на різних рівнях;

c) Отже, DeepLabv3+ є злиття a) і b). Крім того, використовуючи модифіковану Aligned Xception та Atrous Separable Convolution, розроблено більш швидку та потужну мережу.

Авторами [115] запропоновано порівняння DeepLabv3+ з SOTA підходами (див. табл. 4.2). Важливо підкреслити, що в наявних дослідженнях автори не проводили безпосередніх експериментальних вимірювань на знімках дистанційного зондування, а тим більше не включали до аналізу складні об’єкти критичної інфраструктури, де є специфічні типи руйнувань (наприклад, серйозні пошкодження дахів, несучих конструкцій тощо). Саме в подібних випадках DeepLabv3+ демонструє помітне зниження результатів за різними метриками, оскільки робота з надвеликими

зображеннями та складними патернами пошкоджень вимагає більш гнучких методів і модифікованих мережевих архітектур. Як наслідок, екстраполяція результатів із загальних наборів даних на критичні сценарії може бути невиправдано оптимістичною.

Табл. 4.2 Pascal VOC 2012 dataset тест

Метод	**mIoU**
Deep Layer Cascade (LC)	82.7
TuSimple	83.1
Large_Kernel_Matters	83.6
Multipath-RefineNet	84.2
ResNet-38_MS_COCO	84.9
PSPNet	85.4
IDW-CNN	86.3
CASIA_IVA_SDN	86.6
DIS	86.8
DeepLabv3	85.7
DeepLabv3-JFT	86.9
DeepLabv3+ (Xception)	87.8
DeepLabv3+ (Xception-JFT)	89.0

У Табл. 4.2 показано, що DeepLabv3+ (із різними варіаціями) досягає найвищих або близьких до найвищих результатів (mIoU ~ 87.8–89.0%) на Pascal VOC, де сцени переважно містять об’єкти “звичайного” маштабно-просторового характеру (наприклад, люди, тварини, автомобілі). Це демонструє сильні сторони DeepLabv3+ у завданнях сегментації “типових” об’єктів, що історично є однією з ключових бенчмарк-задач.

Водночас, із табл. 4.1 (де оцінювали сегментацію пошкоджень критичної інфраструктури на знімках xBD і под.) видно, що DeepLabv3+ поступається запропонованій та іншим архітектурам які приймали участь в порівнянні:

- Механізм ASPP добре працює з об’єктами середнього й великого масштабу, але може “пропускати” дуже дрібні або нетипові руйнування;

- Запропонована модель (DReAM + дворівнева архітектура) на реальних супутникових зображеннях виграє за точністю (mIoU, F1) у виявленні й сегментації об’єктів, що мають більш широкий діапазон масштабів і незвичайні форми (мости, дороги, складні урбаністичні сцени).

Таким чином, на “класичних” датасетах (VOC, COCO) DeepLabv3+ з ASPP дійсно зберігає високі позиції, а от на складних зображеннях дистанційного зондування, де гнучке охоплення різних масштабів і рідкісні патерни пошкоджень особливо важливі, DeepLabv3+ помітно поступається пропонованому підходу з динамічним рецептивним полем.

4.3.2 Архітектура моделі PSPNet з технікою Dilated Convolution

Як зазначають автори [109] для семантичної сегментації сцен на зображеннях PSPNet демонструє кращі результати, ніж інші мережі для семантичної сегментації, такі як FCN, U-Net і DeepLab (не плутати з DeepLabv3+). PSPNet базується на механізмі dilated convolution - це тип згортки, який вводить пропуски (dilation rate) між ядрами фільтра, що дозволяє збільшити область рецептивного поля без збільшення кількості параметрів чи обчислювальної складності (див. рис. 4.9)

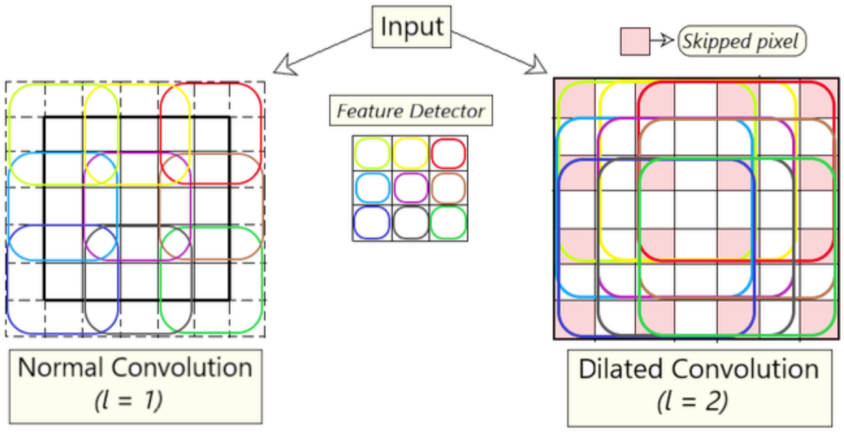


Рис. 4.9 Різниця звичайної згортки від dilated згортки

До основних недоліків PSPNet можна віднести фіксований dilations і фіксований pooling, вони є недостатньо гнучкими на знімках ДЗ з дуже різнорозмірними і гетерогенними об’єктами (мости, греблі, дорожні розв’язки, мікротріщини тощо). На

дуже високій роздільності (VHR) PSPNet є втрати дрібних деталей (див. рис. 4.7) адже мережа не змінює фактичне dilation під локальні особливості сцени, а діє “усереднено”.

Підсумовуючи, PSPNet є сильною, доволі “легкою” в налаштуванні моделлю семантичної сегментації з хорошими результатами на типовому наборі даних (міські сцени, дорожні розмітки). Але у складних зображеннях дистанційного зондування, де одночасно можуть бути як дрібні ушкодження, так і масштабні руйнування, DReAM із більш гнучким рецептивним полем та механізмом уваги (віконної + зміщеної самоуваги) дає кращі показники точності й локалізації.

4.3.3 Метод деформаційних та динамічних згорток

Також, варто взяти до уваги порівняння з деформаційною згорткою [19] що є фактично динамічною згорткою реалізуючи динаміку ERF (ефективне рецептивне поле) (див. рис. 4.10). Цей підхід розв’язує схожу проблему “жорсткого” рецептивного поля, даючи змогу “вигинати” сітку згортання та краще охоплювати об’єкти довільних форм. Проте є кілька моментів:

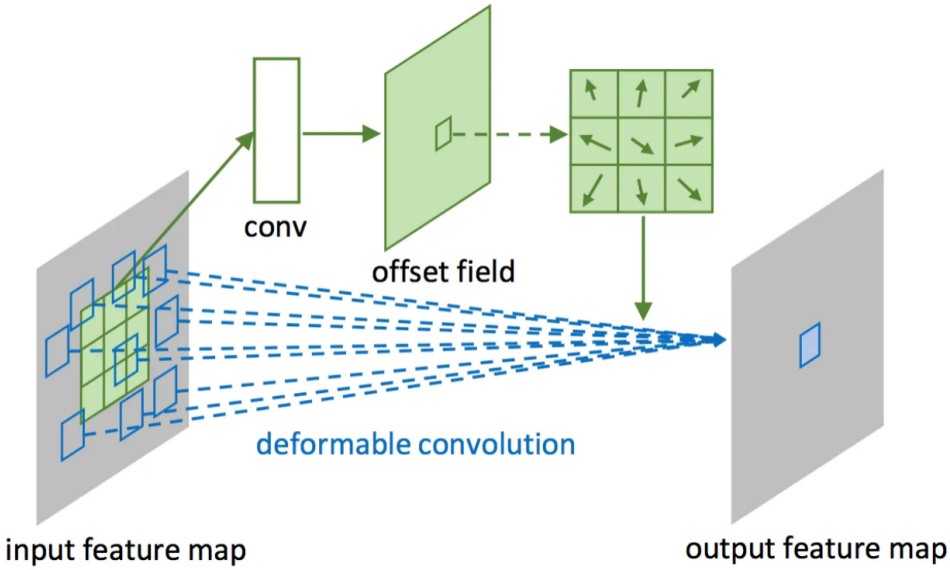


Рис. 4.10 Загальний вигляд 3x3 деформуючої згортки

1. Відмінність від динамічного масштабування (DReAM). У деформованих згортках кожне згортальне ядро має trainable offsets (зміщення), які визначають, із

яких конкретних пікселів зчитувати ознаки. Це справді покращує обробку складних контурів, оскільки ядро “підлаштовується” під форму об’єкта. Однак коефіцієнт розширення (dilation) залишається фіксованим або низкою фіксованих значень, а динамічне масштабування рецептивного поля (у сенсі “ближче/далі”) прямо не реалізоване;

2. Сценарії застосування:

- Deformable Convolution добре показує себе на задачах детектування та сегментації об’єктів зі складними геометріями (уявімо, наприклад, сегментацію людей у довільній позі);
- У знімках дистанційного зондування з різномасштабними сценами і великими (або навпаки надто дрібними) руйнуваннями, deformable kernels можуть допомогти точніше схопити складну “мозаїчну” структуру пошкоджень. Проте вони не “перемикають” масштаб ядра, а лише зміщуються навколо певного фіксованого радіуса.

Порівняння з Deformable Convolution справді доречне, особливо коли йдеться про сегментацію складних або нерегулярних об’єктів, таких як руйнації будівель, розмиті контури тощо. Однак динамічний підхід (DReAM) і трансформерний механізм уваги, згідно з результатами експериментів, мають значно ширшу гнучкість щодо масштабів та розташування об’єктів порівняно з “деформованими” згортками, орієнтованими переважно на локальні геометричні деформації. Такий підхід може виявитися недостатнім для точного відстеження великих структурних змін або віддалених ділянок пошкоджень, де важлива глобальна контекстуальна інформація. Саме тому автор даного дослідження свідомо не включив Deformable Convolution до основних експериментів, покладаючись переважно на перевірені евристики та загальновідомі методи. Механізм “деформованих” згорток не забезпечує повної автономності у швидкому формуванні рішення про тип руйнувань для заданої ROI, тоді як динамічне масштабування та увага дозволяють краще врахувати весь спектр потенційних сценаріїв.

4.3.4 Нейронні мережі трансформер MViT та Segmenter

Одразу варто зазначити, що MViT розроблено для задач класифікації для відеоряду та статичних знімків в тому числі ДЗ, в цій роботі було використано MViTv2 та взято до уваги з двох причин: першою є ідея динамічної адаптації, а друге це високоточна локалізація, отже, ми не говоримо про MViTv2 як бекбон Фази 2 так як не отримуємо сегментаційну карту, але говоримо в конотації Фази 1 і заміни в цілому зв'язки EfficientNet-B3+FPN+RPN. Ідея MViTv2 (Multiscale Vision Transformer) [117] не обмежується лише відео. Хоча цю архітектуру часто застосовують для відеоаналітики, у своїх початкових та подальших модифікаціях вона підтримує і статичні зображення.

Оригінальна версія MViT, представлена дослідниками з Facebook AI (Haoqi Fan та ін.) [111] (рис. 4.11 [117]), була розроблена переважно для багатомасштабної обробки просторово-часових даних, передусім відео, де потрібно ієрархічно "звужувати" або "розширювати" просторову роздільну здатність на різних етапах. Такий підхід дозволяє моделі захоплювати як локальну, так і глобальну динаміку в послідовності кадрів, зберігаючи важливу контекстуальну інформацію про рух і взаємодію об'єктів.

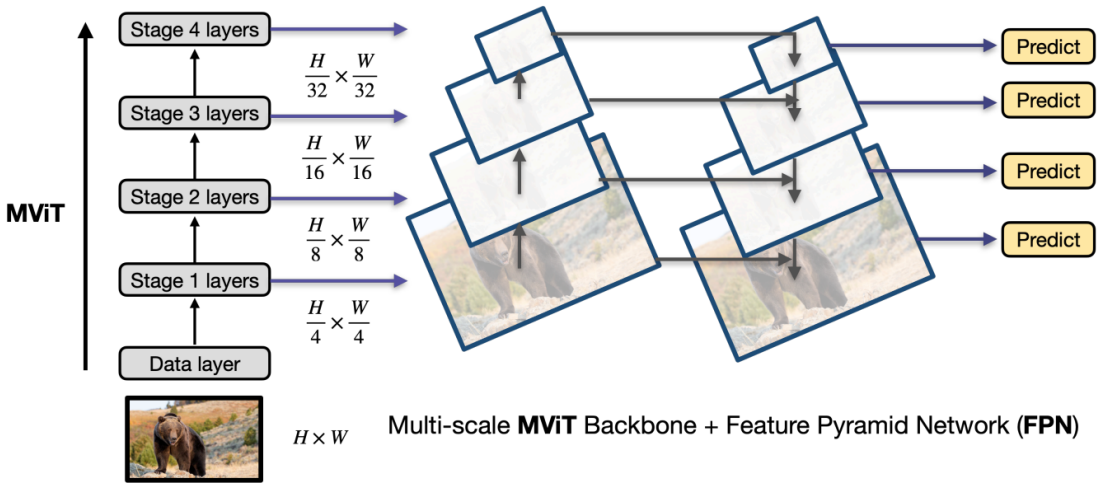


Рис. 4.11 Вигляд MViT Backbone + Feature Pyramid Network (FPN)

Проте ідеї багатомасштабної та ієрархічної архітектури можна ефективно застосовувати й у випадку статичних двовимірних зображень. Саме тому у декількох публікаціях було розглянуто MViTv2 для задач класифікації (ImageNet) та сегментації

(COCO), де модель продемонструвала конкурентні результати поряд із відомими трансформерними та конволюційними архітектурами. Такі властивості MViT (вміння працювати на кількох просторових рівнях одночасно, ієрархічна побудова) роблять його цікавим варіантом і в контексті сегментації складних сцен, на зразок руйнувань критичної інфраструктури, де важливо “вловлювати” як глобальні контури, так і дрібні деталі

Таким чином, запропонована двофазна архітектура з модулем DReAM є більш придатною для аналізу складних і різноманітних за масштабом сцен із супутникових або аерофотознімків. Завдяки початковому швидкому виділенню регіонів інтересу (ROI) та подальшій детальній обробці з гнучким рецептивним полем, модель демонструє вищу точність на великих і неоднорідних зображеннях (див. табл. 4.1), де об’єкти пошкоджень можуть водночас бути як дуже дрібними, так і масштабними.

Натомість MViTv2 (Multiscale Vision Transformer v2) - зручний “усе-в-одному” багатомасштабний трансформерний підхід, що підходить для порівняно менших зображень або датасетів на зразок ImageNet і COCO. Він теж враховує різні масштаби, але не забезпечує динамічної зміни поля зору всередині кожного шару уваги. Це може призводити до втрат точності на дуже великих сценах із надвисокою роздільною здатністю (VHR), де відсутній явний етап відсікання “порожніх” територій та цілеспрямованої фокусировки на пошкодженнях. У результаті, в задачах локалізації та сегментації критичної інфраструктури підхід із DReAM перевершує MViTv2 за точністю й гнучкістю аналізу, особливо коли йдеться про багатомасштабні та неоднорідні руйнування і це підтверджено наявними результатами експериментів в даній роботі.

Другий відомий метод локалізації із сімейства мереж трансформерів є Segmenter [112]. На момент підготовки цієї роботи в спільноті є не так багато досліджень пов’язаних з Segmenter, особливо в спільноті дистанційного зондування, автор даної роботи окремо провів дослідження для порівняння запропонованого методу сегментації/локалізації з Segmenter та іншими відомими підходами, загальну картину можна побачити на графіку (рис. 4.12). Segmenter - це суто трансформерний підхід до семантичної сегментації (автори використовують архітектуру на базі ViT),

де пікселі перетворюються у “токени”, а подальше групування та класифікація відбуваються за допомогою механізму самоуваги. Як видно на рис. 45 повна глобальна самоувага для великого зображення потребує значних GPU-ресурсів (квадратична складність), отже без додаткових модулів (пірамід або динамічного масштабування) Segmenter може втрачати дрібні деталі, коли вхідна роздільність надмірно зменшується або коли в сцені одночасно присутні об’єкти дуже різних розмірів.

Загалом Segmenter добре підходить для відносно однорідних за масштабом сцен, типових для міських чи природних датасетів (Cityscapes, ADE20K). У складних задачах дистанційного зондування з дуже різноманітними об’єктами (дрібними деталями й величезними структурами одночасно) може бути потрібна додаткова

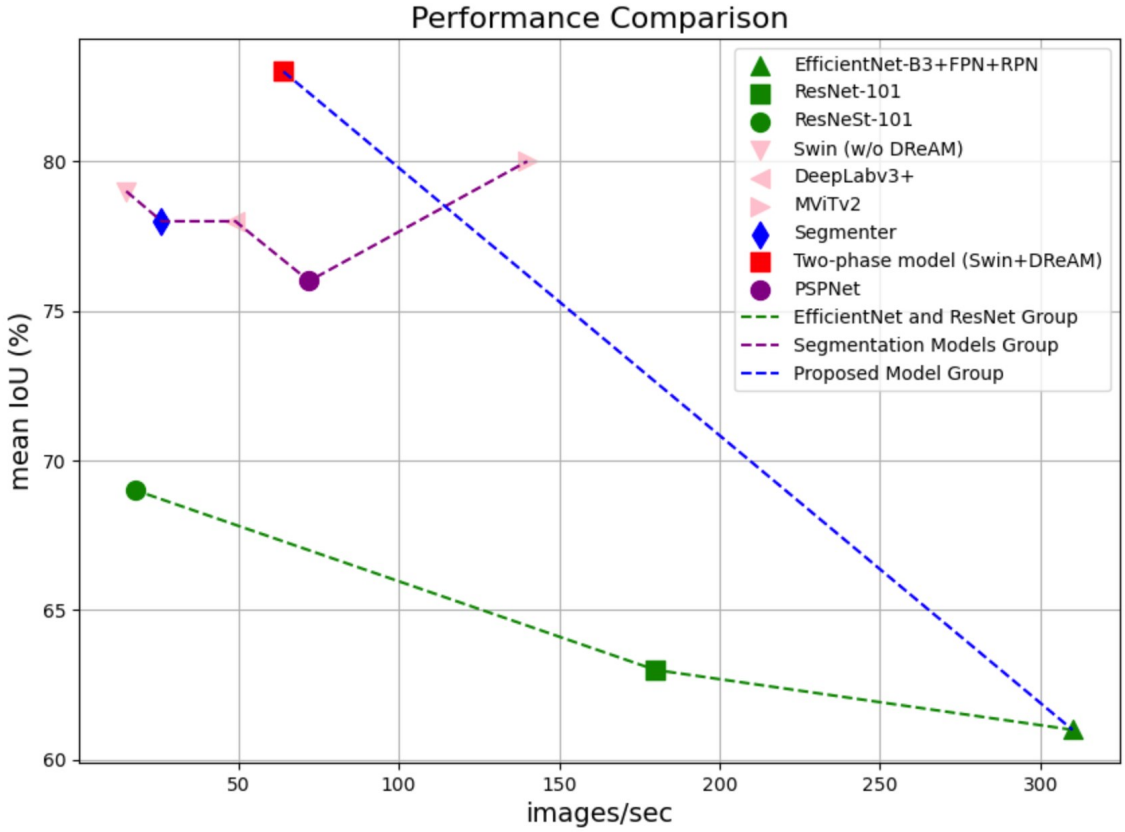


Рис.4.12 Порівняння продуктивності SOTA методів локалізації

ієрархічна або динамічна обробка (на кшталт DReAM), щоби зберегти локальну точність і глобальний контекст без надмірного зростання обчислювальних витрат.

4.4 Переваги та недоліки класичних алгоритмів динамічної локалізації сцен на знімках ДЗ

У даному підрозділі розглянуто класичні алгоритми динамічної локалізації, що передували поширенню глибинних мереж і досі застосовуються як допоміжні або резервні рішення. Хоча ці методи не розраховані на сучасні великі обсяги даних і різноманітні масштаби сцен, вони часто виявляються корисними для швидкого пошуку відповідностей чи початкового аналізу. Зокрема, у підрозділах 4.4.1 і 4.4.2 йтиметься про масштабоінваріантне ознакове перетворення (SIFT) та методи пірамід (гаусових, лапласових). Буде проаналізовано, у яких випадках їхні переваги (стабільність до змін масштабу й освітленості, простота впровадження) можуть бути корисними, а де вони поступаються більш сучасним підходам глибинного навчання з точки зору точності й обчислювальної ефективності.

4.4.1 Алгоритм масштабоінваріантного ознакового перетворення (SIFT)

Ідея алгоритму SIFT (Scale-Invariant Feature Transform) [118] полягає в тому, щоби виявляти та описувати ключові точки на зображенні так, аби вони залишалися максимально стабільними попри зміну масштабу, орієнтації, а також невеликі зміни освітлення. З цією метою алгоритм формує піраміду масштабів і застосовує різницю гаусіан (DoG) для пошуку характерних точок (див. рис. 4.13), які в подальшому аналізуються з точки зору розподілу градієнтів. У результаті кожній ключовій точці відповідає вектор дескриптора - набір параметрів, що відображає локальну текстуру та напрямок найсильніших градієнтів.

У задачах локалізації сцен руйнувань SIFT може слугувати початковим або резервним інструментом: по-перше, він дає можливість знайти сталі ознаки “до” та “після” деструкції й оцінити ступінь змін між різними знімками, а по-друге, здатен виконувати базове зіставлення фрагментів при відсутності глибинних моделей. Водночас, за сильних руйнувань чи повного зникнення контрольних точок зображення результативність SIFT падає, що й зумовлює потребу у більш сучасних

методах, зокрема глибинних нейронних мережах із механізмами динамічного масштабування рецептивного поля.

На рис. 4.13 показано, як формується опис (дескриптор) ключової точки в алгоритмі SIFT (Scale-Invariant Feature Transform):

1. Верхній лівий Рис. демонструє квадратну область навколо виявленої ключової точки (Keypoint), усередині якої виділено “коло впливу” (disk of influence). Це коло вказує на той масштаб (радіус), у якому буде аналізуватися текстура навколо ключової точки;

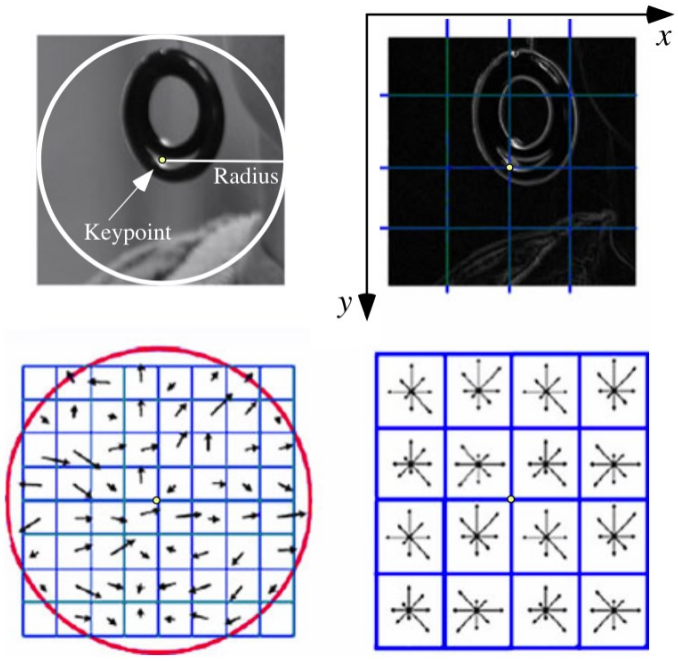


Рис. 4.13 Формування дескриптору ключової точки в алгоритмі SIFT

2. Верхній правий рис. подає карту градієнтів (gradient map) зображення. Регіон навколо ключової точки поділено на менші “квадрати” (зазвичай 4×4). Для кожного з них обчислюються гістограми напрямків градієнтів;

3. Нижній лівий рис. умовно показує, як у кожному квадраті вимірюються вектори градієнта (зображені стрілками). Тобто для кожної “комірки” в цьому 4×4 поділі отримують набір напрямків і їхніх величин;

4. Нижній правий рис. (Lower right) демонструє, як саме ці напрямки й величини градієнтів перетворюються на узагальнені гістограми (одна “зірочка”-

діаграма для кожного квадратика). У такий спосіб формується вектор дескриптора ключової точки - стійкий до зміни масштабу, орієнтації та помірних змін освітлення.

Отже, рис. 4.13 ілюструє сутність SIFT-дескриптора:

- Виявлення ключової точки (Keypoint) і визначення її масштабу;
- Поділ області навколо точки на сітку (4×4);
- Обчислення напрямків градієнтів усередині кожного квадрата;
- Агрегація цих напрямків у гістограми, що зберігають сталі характеристики локальної текстури.

Виходячи з вище зазначеного, у застосуванні до локалізації сцен руйнувань SIFT залишається дієвим інструментом для початкового пошуку сталих (або майже сталих) орієнтирів чи для грубого зіставлення знімків “до” і “після”, тобто, завжди потрібно мати pre-event та post-event набір даних як зазначено в роботі [119]. Проте при масштабних деформаціях і деструкціях, коли ключові ознаки руйнуються, а текстура сцени змінюється докорінно, SIFT стає малоефективним або потребує суттєвого доопрацювання. Тому на практиці його часто поєднують із більш сучасними методами (наприклад, алгоритмами глибинного навчання, описаними в попередніх розділах), які можуть ураховувати контекст сцени та прояви пошкоджень навіть за відсутності чітких локальних ознак.

4.4.2 Методи пірамід Гауса та Лапласа

Піраміди Гауса та Лапласа є класичними способами багатомасштабного представлення зображення (див. рис. 4.14). Ідея полягає в послідовному згладжуванні (через згортку з гаусовим ядром) та зменшенні роздільності, внаслідок чого формується ієрархія все більш “низькодетальних” копій. У гаусовій піраміді кожен рівень отримують шляхом фільтрації та даунсемплінгу попереднього, а в лапласовій додатково фіксують різницю (залишок) між оригіналом і згладженою копією. Так утворюються “похідні” карти, що містять локальні деталі, пропущені при згладженні.

У задачах локалізації сцен руйнувань критичної інфраструктури піраміди можуть слугувати основою для аналізу змін на різних просторових масштабах. Зокрема, вони забезпечують:

1. Виділення дрібних і великих деталей: кожен рівень піраміди дає змогу окремо аналізувати низькочастотні (глобальні) й високочастотні (локальні) компоненти зображення;

2. Попереднє зменшення даних: корисне при дуже високій роздільності супутникових знімків, щоб знайти регіони з суттєвими змінами, перш ніж виконувати обчислювально дорожчі методи сегментації;

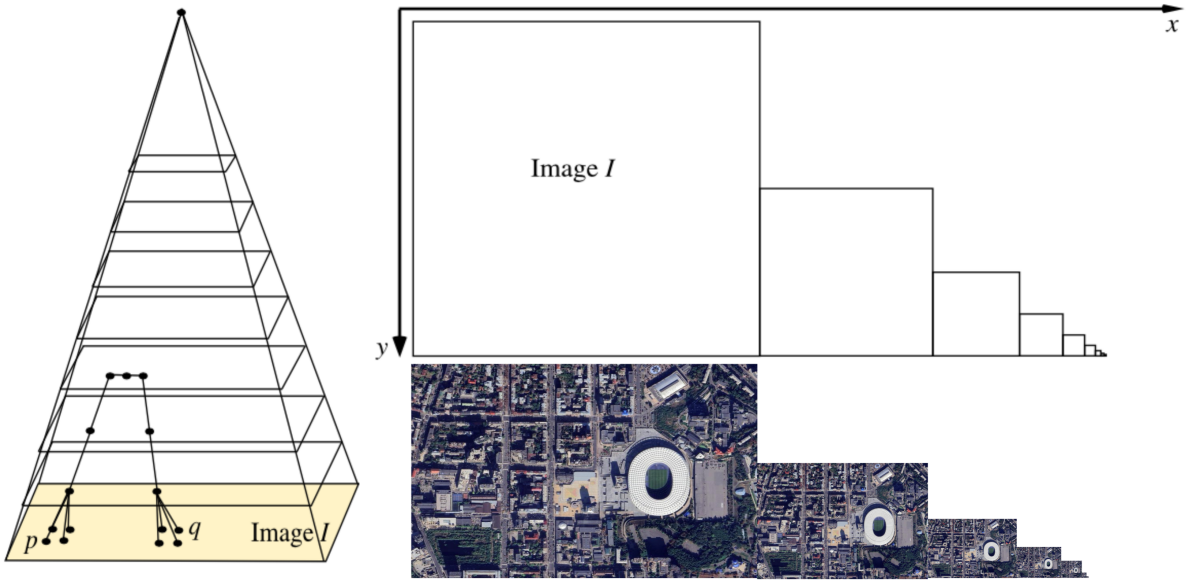


Рис. 4.14 Приклад типової піраміди ознак зображення

3. Злиття (blending) чи порівняння зображень “до” і “після”: лапласову піраміду можна застосовувати, аби точніше оцінити різницю між знімками, фіксуючи саме ті деталі, які зникли або додалися.

Водночас самі по собі піраміди не розв’язують завдання семантичної ідентифікації об’єктів, тож у сучасних сценаріях їх зазвичай комбінують з іншими алгоритмами - від класичних (як-от SIFT) до глибинних мереж, щоб визначити не лише масштаб, а й семантичні зміни, характерні для руйнувань.

Висновки до розділу

Таким чином, у четвертому розділі проведено ґрунтовне порівняння розробленої дворівневої моделі з низкою популярних методів локалізації та сегментації сцен руйнувань, зокрема DeepLabv3+, PSPNet, MViT, Segmenter та іншими. Виявлено, що класичні підходи на основі фіксованого чи розширеного згортального ядра (ASPP, dilated) все ще можуть забезпечувати пристойні результати, однак поступаються за точністю в умовах надвисокої роздільної здатності та великої різнорідності об’єктів.

Порівняння з новітніми трансформерними архітектурами (MViT, Segmenter, Swin без DReAM) показало, що додавання динамічного масштабування рецептивного поля (DReAM) дає змогу краще охопити різні масштаби пошкоджень і поліпшити метрики (mIoU, F1). Зокрема, подолано проблему “вузького” рецептивного поля, характерну для деяких згорткових мереж, і забезпечено злагоджену обробку глобальних та локальних ознак у межах однієї архітектури.

Крім того, було продемонстровано, як двофазна обробка із формуванням регіонів інтересу (ROI) знижує кількість хибних виявлень у великих сценах. Таким чином, результати доводять переваги запропонованого рішення перед іншими методами за низкою ключових показників, особливо у відновлювальних та моніторингових завданнях критичної інфраструктури.

ВИСНОВКИ

Дисертаційну роботу присвячено дослідженню, розробці та впровадженню методів і моделей локалізації сцен пошкоджень критичної інфраструктури на зображеннях дистанційного зондування з використанням підходів глибинного навчання та динамічного масштабування рецептивного поля. У роботі докладно розглянуто існуючі методи аналізу зображень для виявлення та оцінки ступеня руйнувань і запропоновано вдосконалений підхід, який дає змогу підвищити точність та ефективність визначення географічних регіонів та об’єктів із пошкодженнями.

Запропоноване рішення спирається на міждисциплінарні підходи з галузей інженерії програмного забезпечення, комп’ютерних наук і методів обробки зображень. Головним результатом роботи є створення дворівневої архітектури, що поєднує механізми грубого формування регіонів інтересу (ROI) та детальну оцінку сцен з використанням композитних нейронних мереж, зокрема динамічного масштабування рецептивного поля (DReAM) та трансформерних блоків (Swin Transformer). Це дозволяє отримувати високоточні карти пошкоджень різномасштабних об’єктів інфраструктури (мости, дороги, будівлі тощо), а також підвищує продуктивність і стійкість методів до шумів і неоднорідних умов знімання.

У результаті виконання дисертаційної роботи вирішено такі наукові задачі:

1. Проведено порівняльний аналіз відомих підходів до локалізації та сегментації пошкоджень на зображеннях дистанційного зондування, що дало змогу визначити ключові перешкоди: обмеження фіксованого рецептивного поля, брак адаптації до різних масштабів та складні умови знімання (тіні, задимлення, шум);
2. На основі виявлених обмежень розроблено нову модель і метод локалізації сцен руйнувань, які включають динамічний модуль масштабування рецептивного поля (DReAM). Це дало змогу адаптивно враховувати як дрібні локальні деталі, так і глобальний контекст у межах однієї інтегрованої архітектури;
3. Запропоновано інформаційну технологію для комплексного аналізу зображень дистанційного зондування, яка підтримує дворівневий підхід: спершу

виконується грубе визначення регіонів із можливою наявністю пошкоджень, а потім - детальна сегментація та класифікація руйнувань. Технологію апробовано на реальних даних супутникової та аерофотозйомки, зокрема з використанням наборів xBD, DOTA та додаткових джерел (Maxar);

4. Досліджено властивості запропонованої моделі на відкритих наборах даних із різними типами сцени (урбаністичні, промислові, сільськогосподарські) та підтверджено підвищення точності (mIoU, F1-Score) і продуктивності методів у порівнянні з існуючими рішеннями. Систематизація результатів дає змогу використовувати розроблені методи в автоматизованих системах моніторингу критичної інфраструктури.

Наукова новизна одержаних результатів полягає в тому, що *вперше*:

1. Застосовано принцип різнотипності в побудові двофазної архітектури композитної нейронної мережі;

2. Розроблено універсальну дворівневу архітектуру з інтегрованим модулем динамічного масштабування рецептивного поля (DReAM), завдяки якій підвищується точність і гнучкість аналізу руйнувань навіть за складних і неоднорідних умов знімання;

3. Запропоновано метод локалізації різнорозмірних пошкоджень із одночасним врахуванням глобального та локального контекстів, що вдосконалює розв’язання задачі багатомасштабного аналізу зображень (від тріщин чи незначних дефектів до масштабних деформацій і зруйнувань).

Практична значущість роботи підтверджується створенням та апробацією комплексної інформаційної технології моніторингу стану об’єктів критичної інфраструктури. Методи та програмні компоненти, розроблені на базі описаної моделі, можуть бути використані для:

1. Оперативного аналізу наслідків катастроф чи військових дій (оцінка ступеня пошкоджень доріг, мостів, будівель);

2. Автоматизованої сегментації та геопросторового прив’язування пошкоджених об’єктів для планування ремонтно-відновлювальних заходів;

3. Інтеграції з системами підтримки прийняття рішень у надзвичайних ситуаціях.

Основні положення та висновки дисертаційного дослідження обговорювалися на наукових семінарах відповідних кафедр факультету комп’ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка та отримали схвальні відгуки.

Результати дисертаційної роботи апробовані та знайшли застосування у Київському національному університеті імені Тараса Шевченка в рамках наукових дослідницьких робіт Національної академії наук України, “Розробка моделей і методів підтримки прийняття рішень для критичних інфраструктур” за запитом державної організації “Відділення цільової підготовки Київського національного університету імені Тараса Шевченка при Національній академії наук України” н.к. Заславський В.А. до запиту № 2М-2022 від 18.04.2022 р., та на міжнародному науково-технічному семінарі “Критичні комп’ютерні технології та системи” (КриКТехС-2025/2/196).

Дослідження виконані в дисертаційній роботі по проблематиці безпеки критичних інфраструктур, а також результати експериментів є складовою міжнародного проєкту CPEA-LT-2016/10003 «Поглиблена спільна освітньо-наукова програма з управління ризиками в промисловості та сервісах в умовах глобальних економічних, технологічних та екологічних змін: розширена версія», який виконувався під керівництвом проф. Заславського В.А. на факультеті компьютерних наук та кібернетики КНУ імені Тараса Шевченка з Норвезьким університетом науки та технологій (Тронхейм, Норвегія) в період 2017-2024 років.

Науково-практичну важливість результатів дисертації підтверджує впровадження результатів досліджень в навчальному процесі на факультеті комп’ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка в курсах “Актуальні проблеми Data Mining” магістерських

академічних програм “Штучний Інтелект” та “Математичні методи Штучного Інтелекту” в рамках викладацької діяльності.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бірюков Д.С., Заславський В.А., Євгієнко В.В., Франчук О.В. “ Моделювання та оцінка сценаріїв загроз для об'єктів критичної інфраструктури”. 2009. *Наукові записки НаУКМА*. - 2009. - Т. 99 : Комп'ютерні науки. - С. 97-101. URI: https://ekmair.ukma.edu.ua/handle/123456789/3922.
2. Gaivoronski, A.A., S. Knopov, P., & A. Zaslavskyi, V. (Eds.). 2023. Modern Optimization Methods for Decision Making Under Risk and Uncertainty (1st ed.). *CRC Press*. DOI: https://doi.org/10.1201/9781003260196.
3. Д.С. Бірюков, С.І. Кондратов. Зелена книга з питань захисту критичної інфраструктури в україні. *Збірник матеріалів міжнародних експертних нарад.* 2016. УДК 32.1, 323.285, 519.8.
4. Mishchuk, V., Fesenko, H., & Kharchenko, V. 2024. “Deep learning models for detection of explosive ordnance using autonomous robotic systems: trade-off between accuracy and real-time processing speed”. *Radioelectronic and Computer Systems*, 2024(4), 99-111. DOI: https://doi.org/10.32620/reks.2024.4.09.
5. Заславський В.А. 2006. “Принцип різнотипності та особливості дослідження складних систем з високою ціною відмови”. *Вісн. Київ. університету*. Сер. фіз.-мат. н. 2006. № 1.
6. Заславский В.А., Каденко И.Н., Сахно Н.В. “Методологические аспекты обеспечения безопасности сложных технических объектов в условиях ограниченных ресурсов” (Сообщение 2). *Ядерная и радиационная безопасность*. 2000. №4.-С.33-41.
7. Zaslavskyi, Volodymyr & Horbunov, Oleh. 2023. The Type-Variety Principle in Ensuring the Reliability, Safety and Resilience of Critical Infrastructures. In *Modern*

Optimization Methods for Decision Making Under Risk and Uncertainty. DOI: http://dx.doi.org/10.1201/9781003260196-11.

8. Y. Pushkarenko, V. Zaslavskyi. 2025. “Adaptation of dynamic receptive field for remote sensing imageries”. In *Technology audit and production reserves, 7.* DOI: https://doi.org/10.15587/2706-5448.2024.319799.
9. Yailymova, Hanna, Hongji Yang, and Volodymyr Zaslavskyi. "Models and Methods in Creative Computing: Diversity and Type-Variety Principle in Development of Innovation Solutions." In *2017 14th International Symposium on Pervasive Systems, Algorithms and Networks & 2017 11th International Conference on Frontier of Computer Science and Technology & 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC)*, pp. 454-461. IEEE, 2017. DOI: https://doi.org/10.1109/ISPAN-FCST-ISCC.2017.81.
10. United Nations Office for Disaster Risk Reduction (UNDRR). 2020. *Human Cost of Disasters: An overview of the last 20 years 2000–2019*. URL: https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019.
11. The World Bank. 2023. *Ukraine Rapid Damage and Needs Assessment*. URL: https://documents1.worldbank.org/curated/en/099445209072239810/pdf/P178843048379106 30b9c6040ac12428d5c.pdf
12. European Space Agency (ESA). 2023. Satellite Imagery Supports Earthquake Response in Turkey and Syria. URL: https://www.esa.int/Applications/Observing_the_Earth/Satellites_support_impact_assessment_after_Tuerkiye_Syria_earthquakes.
13. International Committee of the Red Cross (ICRC). 2021. *Optimizing Resource Allocation in Disaster Response.* URL: https://www.ifrc.org/sites/default/files/2022-03/Localization_humanitarian_action_RCRC_2021_EN.pdf.
14. European Space Agency (ESA). 2020. Rapid Mapping with Sentinel Satellites. URL:https://medium.com/sentinel-hub/landslide-detection-for-rapid-mapping-using-sentinel-2-2a84c2766a0.

15. Federal Emergency Management Agency (FEMA). 2021. Hurricane Ida Response. URL:https://www.fema.gov/press-release/20210830/federal-agencies-stand-ready-hurricane-ida-response.
16. Pushkarenko, Y., and V. Zaslavskyi. “Research on the State of Areas in Ukraine Affected by Military Actions Based on Remote Sensing Data and Deep Learning Architectures.” *Radioelectronic and Computer Systems* 2024, no. 2 (2024): 5–18. DOI: https://doi.org/10.32620/reks.2024.2.01.
17. Li, Wenmei, Haiyan Liu, Yu Wang, Zhuangzhuang Li, Yan Jia, and Guan Gui. “Deep Learning-Based Classification Methods for Remote Sensing Images in Urban Built-Up Areas.” *IEEE Access* 2019: 1–1. DOI: http://dx.doi.org/10.1109/ACCESS.2019.2903127.
18. Luo, Wenjie, Yujia Li, Raquel Urtasun, and Richard Zemel. "Understanding the effective receptive field in deep convolutional neural networks." *Advances in neural information processing systems* 29 2016. DOI: https://doi.org/10.48550/arXiv.1701.04128.
19. Dai, J., H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei. 2017. “Deformable Convolutional Networks.” In *Proceedings of the IEEE International Conference on Computer Vision*, 764–773. IEEE, 2017. DOI: https://doi.org/10.48550/arXiv.1703.06211.
20. Seif, George, and Dimitrios Androutsos. "Large receptive field networks for high-scale image super-resolution." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 763-772. 2018. DOI: https://doi.org/10.48550/arXiv.1804.08181.
21. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *International Conference on Learning Representations (ICLR)*. DOI: https://doi.org/10.48550/arXiv.2010.11929.

22. Xie, Enze, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. 2021. “SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers.” *Advances in Neural Information Processing Systems* 34: 12077–12090. arXiv:2105.15203. DOI: https://doi.org/10.48550/arXiv.2105.15203.
23. Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. “Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows.” In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 10012–10022. arXiv:2103.14030. DOI: https://doi.org/10.48550/arXiv.2103.14030.
24. Gao, Yanfeng, Yaning Wang, Mengyuan Chen, Xiaoyu Li, Ximeng Sun, Qing Li, and Yao Li. 2022. “Efficient Transformer for Remote Sensing Image Segmentation.” *Remote Sensing* 14(1): 150. DOI: https://doi.org/10.3390/rs14010150.
25. Wang, Wenhai, Enze Xie, Xiang Li, Ding Liang, Tiepeng Zhao, Tong Lu, Gang Yu, Chunhua Shen, and Ping Luo. 2022. “PVTv2: Improved Baselines with Pyramid Vision Transformer.” *Computational Visual Media* 8(3): 415–424. DOI: https://doi.org/10.1007/s41095-022-0254-1.
26. Jensen, J.R. 2015. Introductory Digital Image Processing: A Remote Sensing Perspective (4rd ed.). Upper Saddle River: Prentice-Hall. ISBN-13: 9780137551118.
27. Yu, Xiaozhi, Dengsheng Lu, Xiandie Jiang, Guiying Li, Yaoliang Chen, Dengqiu Li, and Erxue Chen. 2020. "Examining the Roles of Spectral, Spatial, and Topographic Features in Improving Land-Cover and Forest Classifications in a Subtropical Region" *Remote Sensing* 12, no. 18: 2907. DOI:https://doi.org/10.3390/rs12182907.
28. Blaschke, T., and J. Strobl. 2001. “What’s Wrong with Pixels? Some Recent Developments Interfacing Remote Sensing and GIS.” *Proceedings of GIS–Zeitschrift Fur Geoinformationssysteme* 14(6), 12–17.
29. Benediktsson, Jon Atli, Martino Pesaresi, and Keaton Arnason. 2003. “Classification and Feature Extraction for Remote Sensing Images from Urban Areas Based on Morphological Transformations.” *IEEE Transactions on Geoscience and Remote Sensing* 41(9): 1940–1949.DOI: https://doi.org/10.1109/TGRS.2003.814625.

30. Chanussot, Jocelyn, Jon Atli Benediktsson, and Mathieu Fauvel. 2006. "Classification of Remote Sensing Images from Urban Areas Using a Fuzzy Possibilistic Model." *IEEE Geoscience and Remote Sensing Letters* 3(1): 40–44. DOI: https://doi.org/10.1109/LGRS.2005.856117.
31. Tuia, Devis, Fabio Pacifici, Michel Kanevski, and W.J. Emery. 2009. "Classification of Very High Spatial Resolution Imagery Using Mathematical Morphology and Support Vector Machine." *IEEE Transactions on Geoscience and Remote Sensing* 47(11): 3866–3879.DOI: https://doi.org/10.1109/TGRS.2009.2027895.
32. Zhang, Y. 2001. "Texture-Integrated Classification of Urban Treed Areas in High-Resolution Color-Infrared Imagery." Photogrammetric Engineering & Remote Sensing 67(12): 1359–1365. ISSN: 0099-1112.
33. Puissant, A., J. Hirsch, and C. Weber. 2005. "The Utility of Texture Analysis to Improve Per-Pixel Classification for High to Very Spatial Resolution Imagery." *International Journal of Remote Sensing* 26(4): 733–745. https://doi.org/10.1080/01431160512331316838.
34. Pacifici, F., M. Chini, and W.J. Emery. 2009. "A Neural Network Approach Using Multi-Scale Textural Metrics from Very High-Resolution Panchromatic Imagery for Urban Land-Use Classification." *Remote Sensing of Environment* 113(6): 1276–1292. DOI: https://doi.org/10.1016/j.rse.2009.02.014.
35. Wulder, M., K.O. Niemann, and D.G. Goodenough. 2000. "Local Maximum Filtering for the Extraction of Tree Locations and Basal Area from High Spatial Resolution Imagery." *Remote Sensing of Environment* 73(1): 103–114. DOI: https://doi.org/10.1016/S0034-4257(00)00101-2.
36. Ouma, Y.O., T.G. Ngigi, and R. Tateishi. 2006. "On the Optimization and Selection of Wavelet Texture for Feature Extraction from High-Resolution Satellite Imagery with Application Towards Urban-Tree Delineation." *International Journal of Remote Sensing* 27(1): 73–104. DOI: https://doi.org/10.1080/01431160500295885.
37. Sirmacek, B., and C. Unsalan. 2009. "Building Detection Using Local Gabor Features in Very High Resolution Satellite Images." In *Proceedings of Recent Advances in*

Space Technologies (RAST), 283–286. IEEE.DOI: https://doi.org/10.1109/RAST.2009.5158213.

38. Baatz, M., and A. Schape. 2000. "Multiresolution Segmentation: An Optimization Approach for High Quality Multi-Scale Image Segmentation." In *Angewandte Geographische Informations-Verarbeitung XII*, 12–23.
39. Burnett, C., and T. Blaschke. 2003. "A Multi-Scale Segmentation/Object Relationship Modeling Methodology for Landscape Analysis." *Ecological Modelling* 168(3): 233–249.DOI: https://doi.org/10.1016/S0304-3800(03)00139-X.
40. Benz, U.C., P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. 2004. "Multi-Resolution Object-Oriented Fuzzy Analysis of Remote Sensing Data for GIS-Ready Information." *ISPRS Journal of Photogrammetry and Remote Sensing* 58(3–4): 239–258. DOI: https://doi.org/10.1016/j.isprsjprs.2003.10.002.
41. Hay, G. J., and G. Castilla. 2008. "Geographic Object-Based Image Analysis (GEOBIA): A New Name for a New Discipline." In *Object-Based Image Analysis*, edited by T. Blaschke, S. Lang, and G. J. Hay, 75–89. Springer Berlin Heidelberg. DOI: https://doi.org/10.1007/978-3-540-77058-9_4.
42. Ecognition Software https://geospatial.trimble.com/en/products/software/trimble-ecognition.
43. Flanders, D., M. Hall-Beyer, and J. Pereverzoff. 2003. "Preliminary Evaluation of eCognition Object-Based Software for Cut Block Delineation and Feature Extraction." *Canadian Journal of Remote Sensing* 29(4): 441–452. DOI: http://dx.doi.org/10.5589/m03-006.
44. https://en.wikipedia.org/wiki/Cognition_Network_Technology.
45. Yu, Q., P. Gong, N. Clinton, G. Biging, M. Kelly, and D. Schirokauer. 2006. "Object-Based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery." Photogrammetric Engineering & Remote Sensing 72(7): 799–811. DOI: https://doi.org/10.14358/PERS.72.7.799.
46. https://en.wikipedia.org/wiki/Image_analysis#GEOBIA .

47. Blaschke, T. 2010. “Object-Based Image Analysis for Remote Sensing.” *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 2–16. DOI: https://doi.org/10.1016/j.isprsjprs.2009.06.004.
48. Blaschke, T., G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, … D. Tiede. 2014. “Geographic Object-Based Image Analysis – Towards a New Paradigm.” *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 180–191. DOI: https://doi.org/10.1016/j.isprsjprs.2013.09.014.
49. Kolmogorov, V., and R. Zabih. 2004. “What Energy Functions Can Be Minimized via Graph Cuts.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(2): 147–159. DOI: https://doi.org/10.1109/TPAMI.2004.1262177.
50. Arbelaez, P., M. Maire, C. Fowlkes, and J. Malik. 2011. “Contour Detection and Hierarchical Image Segmentation.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(5): 898–916. DOI: https://doi.org/10.1109/TPAMI.2010.
51. Arbelaez, P., J. Pont-Tuset, J. T. Barron, F. Marques, and J. Malik. 2014. “Multiscale Combinatorial Grouping.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 328–335. IEEE. DOI: https://doi.org/10.1109/TPAMI.2016.2537320.
52. Moller, M., L. Lymburner, and M. Volk. 2007. “The Comparison Index: A Tool for Assessing the Accuracy of Image Segmentation.” *International Journal of Applied Earth Observation and Geoinformation* 9: 311–321. DOI: https://doi.org/10.1016/j.jag.2006.10.002.
53. Kampouraki, M., G.A. Wood, and T.R. Berwer. 2008. “Opportunities and Limitations of Object Based Image Analysis for Detecting Urban Impervious and Vegetated Surfaces Using True-Colour Aerial Photography.” In *Object-Based Image Analysis – Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, edited by T. Blaschke, S. Lang, and G. Hay, 555–569. Berlin: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-540-77058-9_30.
54. Kosaka, N., T. Akiyama, B. Tsai, and T. Kojima. 2005. “Forest Type Classification Using Data Fusion of Multispectral and Panchromatic High Resolution Satellite Imageries.” In *Proceedings of IEEE International Geoscience and Remote Sensing*

Symposium (IGARSS) 4: 2980–2983. DOI: https://doi.org/10.1109/IGARSS.2005.1525695.

55. Sohn, G., and I. Dowman. 2007. "Data Fusion of High-Resolution Satellite Imagery and LiDAR Data for Automatic Building Extraction." *ISPRS Journal of Photogrammetry and Remote Sensing* 62(1): 43–63. DOI: https://doi.org/10.1016/j.isprsjprs.2007.01.001.
56. Kim, Y., and Y. Kim. 2014. "Improved Classification Accuracy Based on the Output-Level Fusion of High-Resolution Satellite Images and Airborne LiDAR Data in Urban Area." *IEEE Geoscience and Remote Sensing Letters* 11(3): 636–640. DOI: https://doi.org/10.1109/LGRS.2013.2273397.
57. Bruzzone, L., and L. Carlin. 2006. "A Multilevel Context-Based System for Classification of Very High Spatial Resolution Images." *IEEE Transactions on Geoscience and Remote Sensing* 44(9): 2587–2600. DOI: https://doi.org/10.1109/TGRS.2006.875360.
58. Huang, X., and L. Zhang. 2012. "Morphological Building/Shadow Index for Building Extraction from High-Resolution Imagery over Urban Areas." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5(1): 161–172. DOI: https://doi.org/10.1109/JSTARS.2011.2168195.
59. Miao, Z., W. Shi, H. Zhang, and X. Wang. 2013. "Road Centerline Extraction from High-Resolution Imagery Based on Shape Features and Multivariate Adaptive Regression Splines." *IEEE Geoscience and Remote Sensing Letters* 10(3): 583–587. DOI: https://doi.org/10.1109/LGRS.2012.2214761.
60. Zhang, H., W. Shi, Y. Wang, M. Hao, and Z. Miao. 2014. "Classification of Very High Spatial Resolution Imagery Based on a New Pixel Shape Feature Set." *IEEE Geoscience and Remote Sensing Letters* 11(5): 940–944. DOI: https://doi.org/10.1109/LGRS.2013.2282469.
61. Lowe, David G. 2004. "Distinctive Image Features from Scale-Invariant Keypoints." International Journal of Computer Vision 60(2): 91–110. DOI: https://doi.org/10.1023/B:VISI.0000029664.99615.94.

62. Dalal, Navneet, and Bill Triggs. 2005. “Histograms of Oriented Gradients for Human Detection.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 886–893. IEEE. DOI: https://doi.org/10.1109/CVPR.2005.177.
63. Lazebnik, S., C. Schmid, and J. Ponce. 2006. “Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2169–2178. IEEE. DOI: https://doi.org/10.1109/CVPR.2006.68.
64. Huang, C., L.S. Davis, and J.R.G. Townshend. 2002. “An Assessment of Support Vector Machines for Land Cover Classification.” *International Journal of Remote Sensing* 23(4): 725–749. DOI: https://doi.org/10.1080/01431160110040323.
65. Mountrakis, Giorgos, Jungho Im, and Cajetan Ogole. 2011. “Support Vector Machine in Remote Sensing: A Review.” *ISPRS Journal of Photogrammetry and Remote Sensing* 66(3): 247–259. DOI: https://doi.org/10.1016/j.isprsjprs.2010.11.001.
66. Gong, Bin, Jungho Im, and Giorgos Muntrakis. 2011. “An Artificial Immune Network Approach to Multi-Sensor Land Use/Land Cover Classification.” *Remote Sensing of Environment* 115(2): 600–614. DOI: https://doi.org/10.1016/j.rse.2010.10.005.
67. Mumford, David, and Agnes Desolneux. 2010. Pattern Theory: The Stochastic Analysis of Real-World Signals. Natick, MA: A K Peters, Ltd. ISBN 9781032920054.
68. Stoica, R., X. Descombes, and J. Zerubia. 2004. “A Gibbs Point Process for Road Extraction in Remotely Sensed Images.” *International Journal of Computer Vision* 57(2): 121–136. DOI: https://doi.org/10.1023/B:VISI.0000013086.45688.5d.
69. Lacoste, C., X. Descombes, and J. Zerubia. 2005. “Point Processes for Unsupervised Line Network Extraction in Remote Sensing.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(10): 1568–1579. DOI: https://doi.org/10.1109/TPAMI.2005.206.
70. Perrin, G., X. Descombes, and J. Zerubia. 2005. “A Marked Point Process Model for Tree Crown Extraction in Plantations.” In *Proceedings of IEEE International*

Conference on Image Processing (ICIP), vol. 2, 661–664. DOI: https://doi.org/10.1109/ICIP.2005.1529837.

71. Ortner, M., X. Descombes, and J. Zerubia. 2008. "A Marked Point Process of Rectangles and Segments for Automatic Analysis of Digital Elevation Models." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(1): 105–119. DOI: https://doi.org/10.1109/TPAMI.2007.1159.
72. Lacoste, C., X. Descombes, and J. Zerubia. 2010. "Unsupervised Line Work Network Extraction in Remote Sensing Using a Polyline Process." *Pattern Recognition* 43(4): 1631–1641. DOI: https://doi.org/10.1016/j.patcog.2009.11.003.
73. Lafarge, F., G. Gimel'farb, and X. Descombes. 2010. "Geometric Feature Extraction by a Multimarked Point Process." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(9): 1597–1609. DOI: https://doi.org/10.1109/TPAMI.2009.152.
74. Benedek, C., X. Descombes, and J. Zerubia. 2012. "Building Development Monitoring in Multitemporal Remotely Sensed Image Pairs with Stochastic Birth-Death Dynamics." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(1): 33–50. DOI: https://doi.org/10.1109/TPAMI.2011.94.
75. Verdie, Y., and F. Lafarge. 2014. "Detecting Parametric Objects in Large Scenes by Monte Carlo Sampling." *International Journal of Computer Vision* 106(1): 57–75. DOI: https://doi.org/10.1007/s11263-013-0641-0.
76. Porway, J., Q. Wang, and S. C. Zhu. 2010. "A Hierarchical and Contextual Model for Aerial Image Parsing." *International Journal of Computer Vision* 88(2): 254–283. DOI: https://doi.org/10.1007/s11263-009-0306-1.
77. Pushkarenko, Yurii, and Volodymyr Zaslavskyi. 2024. "Synthetic Data Generation for Fraud Detection Using Diffusion Models." *Information & Security*. DOI: https://doi.org/10.11610/isij.5534 2024.
78. Hinton, Geoffrey, and Ruslan R. Salakhutdinov. 2006. "Reducing the Dimensionality of Data with Neural Networks." *Science* 313(5786): 505–507. DOI: https://doi.org/10.1126/science.1127647.

79. Hinton, Geoffrey, Simon Osindero, and Yee Whye Teh. 2006. “A Fast Learning Algorithm for Deep Belief Nets.” *Neural Computation* 18(7): 1527–1554. DOI: https://doi.org/10.1162/neco.2006.18.7.1527.
80. Bengio, Yoshua, Patrice Lamblin, Dan Popovici, and Hugo Larochelle. 2007. “Greedy Layer-Wise Training of Deep Networks.” In *Advances in Neural Information Processing Systems (NIPS),* 19: 153–160. MIT Press. DOI: https://doi.org/10.7551/mitpress/7503.003.0099.
81. Dahl, George E., Dong Yu, Li Deng, and Alex Acero. 2012. “Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition.” *IEEE Transactions on Audio, Speech, and Language Processing* 20(1): 30–42. DOI: https://doi.org/10.1109/TASL.2011.2134090.
82. Le, Quoc V., Wan Y. Zou, Serena Y. Yeung, and Andrew Y. Ng. 2011. “Learning Hierarchical Invariant Spatio-Temporal Features for Action Recognition with Independent Subspace Analysis.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado. DOI: https://doi.org/10.1109/CVPR.2011.5995496.
83. Collobert, Ronan, and Jason Weston. 2008. “A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning.” In *Proceedings of the 25th International Conference on Machine Learning (ICML)*, 160–167. Helsinki. DOI: https://doi.org/10.1145/1390156.1390177.
84. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. “ImageNet Classification with Deep Convolutional Neural Networks.” In *Advances in Neural Information Processing Systems (NIPS)*. DOI: https://doi.org/10.1145/3065386.
85. Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 580–587. Columbus. DOI: https://doi.org/10.1109/CVPR.2014.81.
86. Mnih, Volodymyr. 2013. Machine Learning for Aerial Image Labeling (Doctoral dissertation, University of Toronto).

87. Al-Doski, Jwan, Shattri B. Mansorl, and Helmi Zulhaidi Mohd Shafri. 2013. “Image Classification in Remote Sensing.” *Department of Civil Engineering, Faculty of Engineering, University Putra, Malaysia* 3(10).
88. Chen, Liang-Chieh, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation.” In *Proceedings of the European Conference on Computer Vision (ECCV)*, 833–851. DOI: https://doi.org/10.1007/978-3-030-01234-2_49.
89. Lin, Tsung-Yi, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. “Feature Pyramid Networks for Object Detection.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 936–944. DOI: https://doi.org/10.1109/CVPR.2017.106.
90. Jamali, Ali, Swalpa Kumar Roy, Danfeng Hong, Bing Lu, and Pedram Ghamisi. 2024. "How to Learn More? Exploring Kolmogorov–Arnold Networks for Hyperspectral Image Classification" *Remote Sensing* 16, no. 21: 4015. https://doi.org/10.3390/rs16214015.
91. Liu, Ziming, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. 2024. “KAN: Kolmogorov-Arnold Networks.” *arXiv preprint arXiv:2404.19756.* DOI: https://doi.org/10.48550/arXiv.2404.19756.
92. Huang, Qihao, Guowang Jin, Xin Xiong, Hao Ye, and Yuzhi Xie. 2023. "Monitoring Urban Change in Conflict from the Perspective of Optical and SAR Satellites: The Case of Mariupol, a City in the Conflict between RUS and UKR" *Remote Sensing* 15, no. 12: 3096. DOI: https://doi.org/10.3390/rs15123096.
93. Kopiika, Nadiia, Andreas Karavias, Pavlos Krassakis, Zehao Ye, Jelena Ninic, Nataliya Shakhovska, Nikolaos Koukouzas, Sotirios Argyroudis, and Stergios-Aristoteles Mitoulis. 2024. “Tiered Approach for Rapid Damage Characterisation of Infrastructure Enabled by Remote Sensing and Deep Learning Technologies.” *arXiv preprint arXiv:2401.17759*. DOI: https://doi.org/10.1016/j.autcon.2024.105955.
94. Yurii Pushkarenko, Zaslavskyi Volodymyr. 2024. “Multiscale Scene Localization Based on Composite Network for Remote Sensing Imageries: A Case Study on

Critical Infrastructure". In *Proceedings IEEE International Conference on Dependable Systems, Services and Technologies, DESSERT'2024.*

95. Yan, Fei, Siyuan Wu, Qiong Zhang, Yunqing Liu, and Haonan Sun. 2023. "Destriping of Remote Sensing Images by an Optimized Variational Model". *Sensors 23*, no. 17: 7529. DOI: https://doi.org/10.3390/s23177529.
96. Tan, Mingxing, and Quoc V. Le. 2019. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." *In Proceedings of the 36th International Conference on Machine Learning, PMLR* 97: 6105–6114. DOI: https://doi.org/10.48550/arXiv.1905.11946.
97. Ren, Shaoqing. 2015. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *arXiv preprint arXiv:1506.01497*. DOI: https://doi.org/10.48550/arXiv.1506.01497.
98. A. Kızrak, "Comparison of Activation Functions for Deep Neural Networks," *Medium*, Jan. 07, 2020. https://towardsdatascience.com/comparison-of-activationfunctions-for-deep-neural-networks-706ac4284c8a (accessed Oct. 11, 2020).
99. Zhao, Yijiang, Xingcai Wei, Yizhi Liu, and Zhuhua Liao. 2022. "A Reputation Model of OSM Contributor Based on Semantic Similarity of Ontology Concepts" Applied Sciences 12, no. 22: 11363. DOI: https://doi.org/10.3390/app122211363.
100. Melamed, Dennis, Cameron Johnson, Chen Zhao, Russell Blue, Philip Morrone, Anthony J. Hoogs and Brian Clipp. 2022. "xFBD: Focused Building Damage Dataset and Analysis." *ArXivabs/2212.13876.* DOI: *https://doi.org/10.48550/arXiv.2212.13876*.
101. DOTA dataset: https://captain-whu.github.io/DOTA/dataset.html.
102. Chen, Q., Wang, L., Wu, Y., Wu, G., Guo, Z. and Waslander, S.L., 2019. TEMPORARY REMOVAL: Aerial imagery for roof segmentation: A large-scale dataset towards automatic mapping of buildings. *ISPRS journal of photogrammetry and remote sensing*, 147, pp.42-55. DOI: https://doi.org/10.1016/j.isprsjprs.2018.11.011.

103. Christie, Gordon, Neil Fendley, James Wilson, and Ryan Mukherjee. 2018. "Functional map of the world." *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6172-6180. DOI: https://doi.org/10.48550/arXiv.1711.07846.

104. *MAXAR*'s Ukraine dataset: https://github.com/mapconcierge/Ukraine2022data/tree/main/satelliteimageries.

105. UADamage dataset: https://www.uadamage.com/map?h=MTYuNjkxNjYxMTgzMTc0NiwzNy45OTc4.MzU1NzMzMjQyOTUsNDguNTg2OTkyOTExMjcwNDc=

106. Amazon Web Services Cloud: https://aws.amazon.com/.

107. Gholami, A., Kim, S., Dong, Z., Yao, Z., Mahoney, M. W., & Keutzer, K. 2022. A survey of quantization methods for efficient neural network inference. In *Low-Power Computer Vision*, 2913326. DOI: https://doi.org/10.48550/arXiv.2103.13630.

108. Faster Transformer: https://github.com/NVIDIA/FasterTransformer/blob/main/examples/pytorch/swin/Swin-Transformer-Quantization/README.md.

109. Zhao, Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. 2017. "Pyramid Scene Parsing Network." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2881–2890. DOI: https://doi.org/10.1109/CVPR.2017.660.

110. Chen, Liang-Chieh. 2017. "Rethinking Atrous Convolution for Semantic Image Segmentation." arXiv preprint arXiv:1706.05587. DOI: https://doi.org/10.48550/arXiv.1706.05587.

111. Fan, Haoqi, Bo Xiong, Karttikeya Mangalam, Yanghao Li, Zhicheng Yan, Jitendra Malik, and Christoph Feichtenhofer. 2021. "Multiscale Vision Transformers." In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV),* 6824–6835. DOI: https://doi.org/10.1109/ICCV48922.2021.00675.

112. Strudel, R., Garcia, R., Laptev, I. and Schmid, C., 2021. Segmenter: Transformer for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 7262-7272). DOI: https://doi.org/10.48550/arXiv.2105.05633.

113. Bai, H., J. Cheng, Y. Su, S. Liu, and X. Liu. 2022. "Calibrated Focal Loss for Semantic Labeling of High-Resolution Remote Sensing Images." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15: 6531–6547. DOI: https://doi.org/10.1109/JSTARS.2022.3197937.

114. Review: https://towardsdatascience.com/review-deeplabv1-deeplabv2-atrous-convolution-semantic-segmentation-b51c5fbde92d.

115. Review: https://towardsdatascience.com/review-deeplabv3-atrous-convolution-semantic-segmentation-6d818bfd1d74.

116. https://linktr.ee/shtsang.

117. Li, Yanghao, Chao-Yuan Wu, Haoqi Fan, Karttikeya Mangalam, Bo Xiong, Jitendra Malik, and Christoph Feichtenhofer. 2022. "MViTv2: Improved Multiscale Vision Transformers for Classification and Detection." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4804–4814. DOI: https://doi.org/10.48550/arXiv.2112.01526.

118. Lowe, David G. 1999. "Object Recognition from Local Scale-Invariant Features." In *Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV)*, vol. 2, 1150–1157. IEEE. DOI: https://doi.org/10.1109/ICCV.1999.790410.

119. Dellinger, Flora, Julie Delon, Yann Gousseau, Julien Michel, and Florence Tupin. "Change detection for high resolution satellite images, based on SIFT descriptors and an a contrario approach." In *2014 IEEE Geoscience and remote sensing symposium*, pp. 1281-1284. IEEE, 2014. DOI: https://doi.org/10.1109/IGARSS.2014.6946667.

ДОДАТОК 1

Лістинг 1: Формування графу знань OSM

```
Require:
    RawOSMData = {elements, elements9, … , elementsg}
    OSMKnowledgeGraph = G
    KnownTypes = {Road, Bridge, Building, PowerStation, … }
    geometryParser(•)  // Функція, що з OSM − елементів витягає геометрію
    tagParser(•)  // Функція, що з OSM − елементів витягає ключ = значення
    index ← size(G. nodes) + 1  // початковий ID для нових нод
    for each osmElem in RawOSMData do
            nodeID ← None
            nodeCategory ← None
            nodeAttributes ← ∅
            // 1) Отримати геометрію
            geom ← geometryParser(osmElem)
            // Це може бути точка, лінія (Way) чи мультиполігон.
            // 2) Отримати словник тегів
            tags ← tagParser(osmElem)
            // tags: { "highway" = "primary", "bridge" = "yes", "name"
                        = "… " }
            // 3) Визначити категорію на основі тегів
            nodeCategory ← classify(tags, KnownTypes)
            // Напр. якщо "bridge" = "yes", то nodeCategory = Bridge
            // якщо "power" = "station", то nodeCategory = PowerStation
            // якщо "building" = "yes", nodeCategory = Building
            // якщо "highway" = "primary", nodeCategory = Road
            // (може бути кілька випадків)
            // 4) Якщо в графі G ще немає вузла з таким ID OSM
```

```
                    //    (перевіряємо, чи G містить osmElem. id)
                    if not Contains(G, osmElem. id) then
                            nodeID ← index
                            CreateNode(G, nodeID)
                            G. nodes[nodeID]. osm_id ← osmElem. id
                            G. nodes[nodeID]. category ← nodeCategory
                            G. nodes[nodeID]. geometry ← geom
                            index ← index + 1
                            // Додати атрибути
                            for each (k, v) in tags do
                                    G. nodes[nodeID]. attributes[k] ← v
                             end for
                    else
                            // Якщо вже існує вузол, оновимо його
                            nodeID ← findNodeID(G, osmElem. id)
                            updateCategory(G. nodes[nodeID], nodeCategory)
                            mergeGeometry(G. nodes[nodeID], geom)      // якщо треба
                            mergeAttributes(G. nodes[nodeID], tags)
                    end if
                    // 5) Перевірити сусідні
//пов'язані елементи (наприклад, якщо це Way зі списком nodeRefs)
                    // і створити/оновити ребра (Edges)
                    if osmElem. type = "way" then
                            nodeRefs ← osmElem. nodeRefs
                            for each ref in nodeRefs do
                                    if Contains(G, ref) then
                                            // створити ребро nodeID − − connectedTo
                                                    − −> ref
                                            if not hasEdge(G, nodeID, ref) then
```

```
                                        addEdge(G, nodeID, ref, "connected")
                                    end if
                                end if
                            end for
                        end if
    end for
    return G
```

Лістинг 2: Формальна специфікація Z-нотації яка описує логіку обробки сцен

```
State ==
[
    Scenes : ℙ Scene;
    KGraph : ℙ OSMNode; //KGraph
    GClasses : ℙ Category;
    minConfidence : ℝ;
    ProbThreshold : ℝ;
    Results : ℙ (Scene × (OSMNode ⊕ {Unknown}) × (Category
                ⊕ {Unclassified}));
    Probabilities : (Scene × OSMNode × Category) → ℝ
]
InitState ==
[ State |
    Results = ∅ ∧ Probabilities = ∅
]
FindOverlaps ==
[
    ΔState;
    s? : Scene;
    overlaps! : ℙ OSMNode |
```

```
        overlaps! = {node : KGraph | Overlap(Geometry(s?), Geometry(node))
                        > 0}
]
MatchScene ==
[
        ΔState;
        s? : Scene;
        confidence : ℝ |
        ∃ matchedNode : KGraph; matchedCategory : GClasses •
                Overlap(Geometry(s?), Geometry(matchedNode)) > 0 ∧
                matchedNode.category = matchedCategory ∧
                Probabilities(s?, matchedNode, matchedCategory)
                                ≥ ProbThreshold ∧
                confidence ≥ minConfidence ⇒
                        Results′ = Results ∪ {(s?, matchedNode, matchedCategory)}
        ∧
        ¬(∃ matchedNode : KGraph; matchedCategory : GClasses •
          Overlap(Geometry(s?), Geometry(matchedNode)) > 0 ∧
                Probabilities(s?, matchedNode, matchedCategory)
                        ≥ ProbThreshold) ⇒
                Results′ = Results ∪ {(s?, Unknown, Unclassified)}
]
ComputeProbabilities ==
[
        ΔState;
        s? : Scene |
        ∀ node : overlaps(s?) •
                ∀ category : GClasses •
```

```
                Probabilities(s? , node, category)
                        = ComputeProbability(s? , node, category)
]
ProcessScenes ==
[
    ΞState |
    ∀ s : Scenes •
        ∃ confidence : ℝ •
            ComputeProbabilities[s?/s]
                ∧ MatchScene[s?/s; confidence/confidence]
]

ResultsSchema ==
[
    ΞState;
    output! : ℙ (Scene × (OSMNode ⊕ {Unknown}) × (Category
                ⊕ {Unclassified})) |
    output! = Results
]
SemanticMatching ==
InitState ∧ ProcessScenes ∧ ResultsSchema
```