

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ АВІАЦІЙНИЙ ІНСТИТУТ

*Кваліфікаційна наукова
праця на правах рукопису*

КОЗАЧУК ОЛЕКСАНДР ОЛЕКСАНДРОВИЧ

УДК 004.8:004.93

ДИСЕРТАЦІЯ

**«Інформаційна технологія моделювання та розпізнавання навчальних
наборів аерозйомки на основі архітектур нейромереж з автоенкодером»**

122 «Комп'ютерні науки»

12 «Інформаційні технології»

Подається на здобуття ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів
і текстів інших авторів супроводжується посиланнями на відповідні джерела.


_____ О.О. Козачук

Науковий керівник

Приставка Пилип Олександрович

доктор технічних наук, професор

АНОТАЦІЯ

Козачук О.О. Інформаційна технологія моделювання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії в галузі знань 12 «Інформаційні технології», за спеціальністю 122 «Комп’ютерні науки». – Київський авіаційний інститут, Київ, 2026.

Дисертаційна робота присвячена розробленню інформаційної технології моделювання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером. Робота спрямована на побудову, підготовку та розпізнавання геоприв’язаних наборів даних аерозйомки для задач зіставлення даних бортового спостереження з супутниковими сценами.

Дані дистанційного зондування зібрано з бортової зйомки та супутникових платформ і приведено до спільної системи координат для подальшого аналізу. На основі зібраних даних введено показник інформативності цифрового зображення. Окремо проаналізовано розподіли даних і сформульовано процедуру фільтрації вхідних зображень. Технологія розпізнавання ґрунтується на зіставленні аерознімків із відповідними супутниковими сценами з урахуванням параметрів камери, умов польоту за маршрутом і геоприв’язки.

Розроблено систему глибинного виділення ознак, в основу якої покладено нейромережеву модель, що навчається за принципом трійки зображень із використанням триплет-втрати. Запропоновано стратегію формування навчального набору з ефективним добором складних негативних пар зображень для кращого розділення істинних представлень у латентному просторі. Процес реалізовано у двох режимах. В офлайн-режимі виконується підготовка даних і наповнення векторної бази ознаками, виділеними із супутникових сцен. На онлайн стадії ознаки обчислюються для бортового кадру та зіставляються з елементами бази. Така організація дає змогу скоротити загальний час обробки та порівняння, оскільки база наповнюється перед тестуванням чи польотом.

Розроблено систему генерації тривимірних супутникових зображень на основі монокулярної оцінки глибини з використанням нейронної мережі. Запропонована технологія дає змогу отримати проєкцію карти, що відповідає ракурсу зйомки камери. Це забезпечує краще узгодження з польотними знімками та полегшує їх зіставлення. Перевагою методу є те, що запропонована процедура може бути використана для іншого домену вхідних даних, а за відсутності тривимірних даних на існуючих платформах методологія дає змогу самостійно синтезувати тривимірну модель об'єкта спостереження.

Ключові слова: БПЛА, автоматизовані системи, бортове обчислення, набір даних, виявлення аномалій, аналіз даних, машинне навчання, глибоке навчання, автоенкодер, трансформери, монокулярна оцінка глибини, абсолютна локалізація.

ABSTRACT

Kozachuk O.O. (2026). Information technology for modeling and recognizing training datasets of aerial imagery based on neural network architectures with autoencoders - Dissertation manuscript submitted for the degree of Doctor of Philosophy. Dissertation for the degree of Doctor of Philosophy in the field of knowledge 12 “Information Technologies”, specialty 122 “Computer Science”. Kyiv Aviation Institute, Kyiv, 2026.

The dissertation is devoted to the development of an information technology for modeling and recognizing aerial photography training datasets based on neural network architectures with autoencoders.

Remote sensing data were collected from airborne imagery and satellite platforms and transformed into a common coordinate system for further analysis. Based on the collected data, a digital image information index was introduced. Data distributions were analyzed separately, and a procedure for filtering input images was formulated. The recognition technology is based on matching aerial photographs with corresponding satellite scenes, taking into account camera parameters, flight conditions along the route, and georeferencing.

A deep feature extraction system has been developed, based on a neural network model trained using the triplet principle with triplet loss. A strategy for forming a training set with effective selection of complex negative image pairs is proposed to better separate true representations in latent space. The process is implemented in two modes. In offline mode, data preparation is performed, and the feature vector database is populated with features extracted from satellite scenes. In the online stage, features are computed for the onboard frame and matched against the database entries. This approach reduces the total processing and comparison time, as the database is populated prior to testing or flight.

A system for generating three-dimensional satellite images based on monocular depth estimation using a neural network has been developed. The proposed technology allows for obtaining a map projection that corresponds to the camera's shooting angle. This ensures better alignment with in-flight images and facilitates their comparison. An advantage of the method is that the proposed procedure can be applied to a different domain of input data, and in the absence of 3D data on existing platforms, the methodology allows for the independent synthesis of a 3D model of the observed object.

Keywords: UAVs, automated systems, onboard computing, datasets, anomaly detection, data analysis, machine learning, deep learning, autoencoders, transformers, monocular depth estimation, absolute localization.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

Праці, в яких опубліковані основні наукові результати дисертації:

Статті у наукових фахових журналах:

1. Козачук О.О. Порівняльний аналіз монокулярної оцінки глибини супутникових двовимірних зображень. Проблеми інформатизації та управління. Київський авіаційний інститут. Київ, 2025. Том 3 Вип. 83 С. 55-64.

Статті у виданнях, які включено до міжнародних наукометричних баз:

1. O. Cholyskhina. P. Prystavka. O. Kozachuk. V. Zivakin. Training set AERIAL SURVEY for data recognition systems from aerial surveillance cameras. IX

INTERNATIONAL CONFERENCE Information Technology and Implementation. Paper 79. 1 December, 2022, Kyiv.

Особистий внесок автора: тренування нейромереж та валідація отриманих результатів, проведення загального дослідження описаного в публікації.

2. P. Prystavka, S. Dolgikh, O. Kozachuk. Terrain Image Recognition with Unsupervised Generative Representations: the Effect of Anomalies. 12th International Conference on Advanced Computer Information Technologies (ACIT). Pages 485-488. October, 2022, Ruzomberok.

Особистий внесок автора: реалізація програмної частини роботи, проведення навчання нейромережі, дослідження розподілів даних і вилучення аномалій.

3. O. Kozachuk, P. Prystavka. Comparing UAV camera data with satellite imagery for global localization. 2024 IEEE 7th International Conference on Actual Problems of Unmanned Aerial Vehicles Development (APUAVD). Pages 197-200. November, 2024, Kyiv.

Особистий внесок автора: збір даних, описано та запропоновано метод порівняння супутникових знімків зі знімками БПЛА, тренування та валідація отриманих результатів, оформлення роботи.

4. P. Prystavka, S. Dolgikh, O. Cholyshkina, O. Kozachuk. Learned and Native Concepts in Latent Representations of Terrain Images. International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications. Pages 66-84. November, 2022, Kherson.

Особистий внесок автора: збір даних, побудова нейромережі, навчання та валідація отриманих результатів, побудова графіків функцій та проведення дослідження.

5. V. Zivakin, O. Kozachuk, P. Prystavka. Analysis of cross-sectional neural network recognition of satellite and aerial survey data. Proceedings of the Third International Conference on Cyber Hygiene & Conflict Management in Global

Information Networks (CH&CMiGIN 2024). Pages 106-119. January 2024, Kyiv.

Особистий внесок автора: проведення дослідження та аналіз наборів даних, тренування та валідація нейромереж і оформлення отриманих результатів.

Наукові праці, які додатково відображають наукові результати дисертації:

1. О. Козачук. Порівняльний аналіз Vision Transformer та ResNet-50 як автокодувальників для дослідження простору представлень у задачах розпізнавання даних аерофотозйомки. Тези доповідей XXV Міжнародної науково-практичної конференції здобувачів вищої освіти і молодих учених «ПОЛІТ». Сучасні інформаційні комунікаційні технології в авіації. С. 153-155. Квітень, 2025, Київ.
2. В. Зівакін, О. Козачук. Навчальний датасет для нейромережевої обробки даних повітряного спостереження. Proceedings The Tenth World Congress "Aviation in the 21-st Century" Safety in Aviation And Space Technologies Pages 2.5.27– 2.5.29, September, 2022 Kyiv.
3. В. Зівакін, О. Козачук, П. Приставка. Дослідження перехресного нейромережевого розпізнавання даних супутникової та повітряної зйомки. Матеріали XVI міжнародної науково-технічної конференції «ABIA». С. 15.78-15.81. Квітень, 2023, Київ.
4. О. Козачук. Визначення подібності зображень аерозйомки та супутникових даних на основі архітектури нейромережі автокодувальника. Тези доповідей XXIV Міжнародної науково-практичної конференції здобувачів вищої освіти і молодих учених «ПОЛІТ». Сучасні інформаційні комунікаційні технології в авіації. С. 15-16. Квітень, 2024, Київ.
5. О. Козачук. Порівняння розподілів синтетичних та несинтетичних даних аерозйомки з використанням автокодувальника. Тези доповідей XXIII Міжнародної науково-практичної конференції здобувачів вищої освіти і

молодих учених «ПОЛІТ». Сучасні інформаційні комунікаційні технології в авіації. С. 15-16, Квітень, 2023, Київ.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	11
ВСТУП.....	12
РОЗДІЛ 1. ГЛИБИННЕ МЕТРИЧНЕ НАВЧАННЯ ДЛЯ ЗАДАЧ АБСОЛЮТНОЇ ЛОКАЛІЗАЦІЇ	21
Підрозділ 1.1 Огляд методів метричного навчання	22
Підрозділ 1.1.1 Лінійні методи метричного навчання.....	22
Підрозділ 1.1.2 Нелінійні методи метричного навчання.....	25
Підрозділ 1.2 Огляд функцій втрат у глибинному метричному навчанні	27
Підрозділ 1.2.1 Контрастивна та триплетна функції втрат	27
Підрозділ 1.2.2 Проксі-орієнтовані функції втрат.....	29
Підрозділ 1.3 Глибинні нейронні мережі для формування ознак.....	32
Підрозділ 1.3.1 Автоенкодери та їх підвиди.....	32
Підрозділ 1.3.2 Згорткові та сіамські нейронні мережі	35
Підрозділ 1.3.3 Візуальні трансформери	36
Підрозділ 1.4 Візуальне розпізнавання місць.....	38
Підрозділ 1.5 Підсумковий аналіз і постановка задач.....	43
РОЗДІЛ 2. ТЕОРЕТИЧНІ ТА ІНФОРМАЦІЙНІ ЗАСАДИ ТЕХНОЛОГІЇ ЛОКАЛІЗАЦІЇ БПЛА ЗА ДАНИМИ АЕРОЗЙОМКИ	47
Підрозділ 2.1 Математична постановка задачі локалізації БПЛА за оптичним каналом	49
Підрозділ 2.2 Запропонована нейромережева модель	52
Підрозділ 2.3 Базовий енкодер та його параметроефективна адаптація	58
Підрозділ 2.4 Метод побудови навчальних, валідаційних і тестових наборів даних	60

Підрозділ 2.4.1 Дані аерозйомки та їх характеристики.....	60
Підрозділ 2.4.2 Дані супутникової зйомки та їх характеристики.....	64
Підрозділ 2.4.3 Формалізація навчальних наборів	68
Підрозділ 2.5 Метод видалення малоінформативних цифрових зображень і зменшення повторюваності кадрів	73
Висновки до Розділу 2	81
РОЗДІЛ 3. МОДУЛІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ МОДЕЛЮВАННЯ ТА РОЗПІЗНАВАННЯ НАВЧАЛЬНИХ НАБОРІВ АЕРОЗЙОМКИ.....	83
Підрозділ 3.1 Інформаційна та алгоритмічна структура технології розпізнавання даних аерозйомки	83
Підрозділ 3.2 Організація векторної бази представлень та алгоритми пошуку найближчих сусідів	88
Підрозділ 3.3 Програмна реалізація та інтеграція модулів інформаційної технології.....	91
Підрозділ 3.4 Експериментальний стенд	100
Висновки до Розділу 3	105
РОЗДІЛ 4. ДОСЛІДЖЕННЯ ТА ОЦІНКА ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ	107
Підрозділ 4.1 Дослідження інформативності зображень та аналіз розподілів у латентному просторі автоенкодера.....	107
Підрозділ 4.2 Оцінювання якості розпізнавання даних аерозйомки	126
Підрозділ 4.3 Порівняльний аналіз з існуючими методами.....	142
Підрозділ 4.4 Метод побудови тривимірного представлення супутникових сцен та оцінювання його впливу на якість розпізнавання	149
Висновки до Розділу 4	154
ЗАГАЛЬНІ ВИСНОВКИ.....	156

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	160
ДОДАТКИ	169

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

Скорочення, термін, позначення	Пояснення
БПЛА	Безпілотний літальний апарат
БпАК	Безпілотний авіаційний комплекс
GNSS	Global Navigation Satellite System
РЕБ	Радіоелектронна боротьба
НМ	Нейронна мережа
АЕ/АЕ	Автоенкодер/autoencoder
ЦЗ	Цифрове зображення
Офлайн/Offline	Процедура, що виконується поза реальним часом на попередньо зібраних даних
Онлайн/Online	Процедура, що виконується під час надходження даних у реальному часі
ЗАК	Знешумлювальний автоенкодер
Трансформер/Transformer	Архітектура нейронних мереж на основі механізму уваги
BPM/VPR	Візуальне розпізнавання місць/Visual place recognition
ОПМ/NLP	Обробка природної мови/Natural Language Processing
RTK	Кінематичне позиціонування в реальному часі/Real Time Kinematic
HNSW	Ієрархічний малий світ/Hierarchical Navigable Small World
FPS	Кадри на секунду/Frames Per Second

ВСТУП

Актуальність теми. В умовах повномасштабної збройної агресії Російської Федерації проти України одним із критичних завдань сектору безпеки і оборони є підвищення автономності, завадостійкості та живучості безпілотних літальних апаратів, здатних виконувати розвідку, пошук, спостереження та навігаційне забезпечення за втрати або пригнічення сигналів глобальних навігаційних супутникових систем. За таких умов особливої ваги набувають інформаційні технології та методи, що забезпечують локалізацію БПЛА за оптичним каналом, зменшують залежність від зовнішніх радіочастотних каналів і підвищують стійкість системи до навігаційних та інформаційних перешкод. Тематика дисертації відповідає чинним державним пріоритетам у сфері науки, інновацій та оборони, визначеним Законом України “Про пріоритетні напрями розвитку науки і техніки”, Законом України “Про пріоритетні напрями інноваційної діяльності в Україні”, постановою Кабінету Міністрів України від 30.04.2024 № 476 та постановою Кабінету Міністрів України від 05.07.2024 № 787, якими передбачено пріоритетність досліджень і розробок у галузі інформаційно-комунікаційних технологій, штучного інтелекту, робототехніки оборонного призначення, автоматичних систем розпізнавання цілей, інтегрованих систем інтелектуального управління безпілотними комплексами та перешкодозахищених навігаційних систем. Дисертаційне дослідження також узгоджується зі Стратегією воєнної безпеки України, Стратегією розвитку оборонно-промислового комплексу України та Стратегією цифрового розвитку інноваційної діяльності України на період до 2030 року, які визначають розвиток безпілотних, роботизованих та інтелектуальних систем як один із ключових напрямів зміцнення обороноздатності держави.

Науково-прикладна проблема, що розв’язується в дисертації, полягає у суперечності між практичною потребою у високоточній і швидкій абсолютній локалізації безпілотного літального апарата за оптичним каналом та обмеженістю наявних методів, які недостатньо враховують доменний зсув між аерознімками і супутниковими сценами, зміну масштабу та ракурсу зйомки, наявність малоінформативних і повторюваних кадрів, а також обмеження бортових

обчислювальних ресурсів. Розв'язання цієї проблеми потребує не окремого алгоритму, а цілісної інформаційної технології, що поєднує: метод моделювання та формування навчальних наборів, метод видалення малоінформативних цифрових зображень, нейромережеву модель, побудову векторної бази представлень.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційне дослідження виконано в Київському авіаційному інституті в межах наукового напрямку, пов'язаного з розробленням методів комп'ютерного зору, глибинного навчання, інтелектуальної обробки даних аерозйомки та інформаційних технологій для автономних роботизованих систем. Основні результати дисертації отримано під час виконання здобувачем як виконавцем таких держбюджетних науково-дослідних робіт:

1. НДР 479 – ДБ24 «Розроблення багатокритеріальної завадозахищеної навігаційної системи для безпілотних апаратів», державний реєстраційний номер 0124U000291;
2. НДР 528 – ДБ25 «Розвідувально-пошукова бортова підсистема для автономного безпілотного літального апарата», державний реєстраційний номер 0125U001589.

Мета і задачі дослідження. Метою дисертаційної роботи є розроблення інформаційної технології моделювання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером, що включає нейромережеву модель та методи моделювання, фільтрації й розпізнавання даних аерозйомки для забезпечення абсолютної локалізації безпілотного літального апарата за оптичним каналом в умовах втрати або пригнічення сигналів глобальних навігаційних супутникових систем, а також кількісне підтвердження працездатності запропонованих методів.

Під моделюванням у роботі розуміється побудова узгоджених із умовами бортового спостереження представлень супутникових сцен, зокрема формування квазітривимірної структури місцевості на основі монокулярної оцінки глибини та

синтезу ракурсів, наближених до геометрії аерозйомки. Під розпізнаванням у роботі розуміється процедура формування векторних ознак для кадру аерозйомки та встановлення його відповідності геоприв'язаному еталонному фрагменту місцевості за критерієм подібності з подальшим уточненням результату.

Для досягнення поставленої мети розв'язано такі задачі:

1. виконати аналіз сучасних методів оброблення даних бортових систем безпілотних літальних апаратів, підходів до візуального розпізнавання місцевості та глибинного метричного навчання;
2. сформулювати математичну постановку задачі абсолютної локалізації безпілотного літального апарата за оптичним каналом та визначити критерії оцінювання якості розпізнавання і локалізації;
3. розробити метод моделювання, анотування, геоприв'язки та формування навчальних, валідаційних і тестових наборів даних з урахуванням траєкторії польоту, параметрів камери та просторового розділення вибірок;
4. розробити метод видалення малоінформативних цифрових зображень і зменшення повторюваності кадрів у послідовностях;
5. запропонувати нейромережеву модель зіставлення аерознімків і супутникових сцен;
6. розробити метод, що забезпечує геометричне узгодження даних супутникової зйомки із реальними бортовими даними;
7. розробити структуру векторної бази і алгоритми пошуку, оптимізовані для роботи за обмежених бортових обчислювальних ресурсів;
8. . інтегрувати розроблені методи, нейромережеву модель, структуру векторної бази та програмні модулі в інформаційну технологію моделювання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером, провести її експериментальну перевірку та кількісно підтвердити працездатність запропонованих рішень у порівнянні з наявними підходами.

Об’єкт дослідження - процеси моделювання, формування, фільтрації, представлення та розпізнавання даних аерозйомки і супутникових зображень для абсолютної візуальної локалізації безпілотного літального апарата.

Предмет дослідження - методи, моделі, алгоритми та програмні засоби нейромережевого формування ознак, видалення малоінформативних цифрових зображень, побудови навчальних наборів типу «супутниковий знімок – аерознімок», побудови тривимірних представлень супутникових сцен і векторного пошуку в задачі абсолютної локалізації безпілотного літального апарата.

Методи дослідження. У роботі використано методи системного аналізу та порівняльного оцінювання; методи математичного моделювання, геометричних перетворень і проєктивної геометрії; методи статистичного аналізу даних, зокрема аналіз латентного простору, метод головних компонент, перевірку статистичних гіпотез; методи комп’ютерного зору, глибинного метричного навчання, автоенкодерного аналізу та монокулярної оцінки глибини; методи векторного пошуку найближчих сусідів; експериментальні методи дослідження з використанням натурних даних аерозйомки, супутникових еталонів і високоточних еталонних координат.

Інформаційна база дослідження. Інформаційну базу дослідження становили наукові праці з проблематики комп’ютерного зору, глибинного метричного навчання та візуального розпізнавання місцевості; власні записи аерозйомки, отримані під час польотних експериментів; супровідні телеметричні та високоточні навігаційні дані; геоприв’язані супутникові матеріали; програмні засоби для реалізації нейромережових моделей і векторного пошуку.

Наукова новизна одержаних результатів полягає в такому. **Уперше:**

1. запропоновано нейромережеву модель двоетапного зіставлення аерознімків і супутникових сцен, побудовану на базі параметроефективно адаптованого енкодера DINOv2, модуля порогового оцінювання впевненості, модуля відбору опорних дескрипторів і локального уточнення за взаємними

найближчими відповідниками, яка, на відміну від відомих підходів, забезпечує одночасне формування глобальних і локальних ознак за зменшених обчислювальних витрат і дає змогу досягати середнього часу оброблення одного запиту 0,064 секунди на бортовому комп'ютері NVIDIA Jetson Orin NX.

2. запропоновано метод видалення малоінформативних цифрових зображень із навчальних наборів аерозйомки, який, на відміну від евристичних методів фільтрації за окремими фотометричними ознаками, ґрунтується на аналізі латентних ознак автоенкодера, проєкції в простір головних компонент і квантильному відборі, що забезпечує спрямоване вилучення 30 % найменш інформативних кадрів і підвищення різноманітності вибірки у просторі ознак на 26,1 %.
3. розроблено інформаційну технологію моделювання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером, призначену для абсолютної локалізації безпілотного літального апарата за оптичним каналом в умовах втрати або пригнічення сигналів глобальних навігаційних супутникових систем, яка, на відміну від наявних рішень, реалізує двостадійне офлайн та онлайн-оброблення, попереднє формування і зберігання векторних представлень супутникових еталонів, просторово обмежений пошук, локальне геометричне уточнення та модуль тривимірного узгодження супутникових сцен, що забезпечило середню просторову похибку 4,45–4,57 метра для 85 % кадрів.

Удосконалено:

1. метод моделювання, формалізації та побудови навчальних, валідаційних і тестових наборів даних, який, на відміну від наявних підходів, враховує геоприв'язку аерознімка і супутникового фрагмента, траєкторію польоту, параметри зйомки, просторове розділення підвбірок і добір складних негативних прикладів, що дало змогу сформувати набір обсягом понад 413

000 коректних відповідностей та отримати 263 701 репрезентативний приклад після фільтрації;

2. метод геометричного узгодження супутникових еталонів з умовами бортового спостереження шляхом побудови тривимірного представлення супутникових сцен на основі монокулярної оцінки глибини та синтезу ракурсів, узгоджених із параметрами аерозйомки, що, на відміну від плаского двовимірного подання супутникових даних, підвищує якість розпізнавання в складних режимах зйомки, зокрема для горизонтального ракурсу збільшує $R@1$ з 25,0 % до 60,0 % на висоті 40 метрів і з 45,0 % до 80,0 % на висоті 60 метрів.

Набули подальшого розвитку:

1. методи зберігання та пошуку векторних представлень за рахунок розробки структури векторної бази, винесення локальних ознак у окреме сховище та використання пошуку найближчих сусідів з урахуванням просторових обмежень, що забезпечило придатність технології до бортового застосування в режимі, наближеному до реального часу;
2. методи статистичного аналізу латентного простору автоенкодера для оцінювання інформативності кадрів, повторюваності послідовностей і ступеня відокремлення позитивних і негативних пар, що дало змогу кількісно обґрунтувати процедури підготовки даних для подальшого метричного навчання.

Практичне значення одержаних результатів полягає в тому, що:

1. алгоритмізовано метод моделювання, анотування, геоприв'язки та формування навчальних, валідаційних і тестових наборів даних, метод видалення малоінформативних цифрових зображень, метод геометричного узгодження супутникових даних з реальними бортовими даними та нейромережеву модель двоетапного зіставлення, що забезпечило можливість їх подальшої програмної реалізації в складі інформаційної технології

модельовання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером;

2. створено навчальний набір обсягом понад 413 000 геоприв'язаних відповідностей, придатний для навчання, валідації та тестування нейромережових систем розпізнавання, а також відфільтровану підмножину обсягом 263 701 приклад для побудови репрезентативних навчальних вибірок;
3. на основі виконаної алгоритмізації розроблено та програмно реалізовано модулі інформаційної технології модельовання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером, зокрема модулі формування глобальних і локальних ознак, добору складних негативних прикладів, векторного пошуку, локального уточнення, геометричного перетворення та побудови тривимірного представлення супутникових сцен;
4. модуль побудови тривимірного представлення супутникових сцен, реалізований у складі інформаційної технології модельовання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером, може використовуватися для узгодження супутникових знімків за малих висот польоту, що розширює можливості застосування системи в умовах складної геометрії зйомки;
5. розроблені методи, алгоритми та програмні засоби можуть бути використані під час створення відомчих геоприв'язаних баз зображень місцевості, систем резервної навігації, систем візуальної прив'язки, розвідувально-пошукових бортових підсистем та інших інтелектуальних модулів для безпілотних літальних апаратів;
6. інформаційна технологія модельовання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером може бути використана як резервний канал навігації безпілотного літального апарата в умовах втрати або пригнічення сигналів глобальних навігаційних супутникових систем та за активної радіоелектронної протидії; на бортовому

комп'ютері NVIDIA Jetson Orin NX середній час оброблення одного запиту становить 0,064 секунди, що забезпечує придатність до роботи в режимі, наближеному до реального часу;

7. результати дисертаційної роботи можуть бути використані в наукових дослідженнях та в освітньому процесі під час підготовки фахівців зі спеціальності 122 «Комп'ютерні науки» та суміжних спеціальностей у межах дисциплін з глибинного навчання, комп'ютерного зору, оброблення зображень та інтелектуальних бортових систем;
8. результати дисертаційного дослідження впроваджено в освітній процес кафедри прикладної математики Національного університету «Київський авіаційний інститут» під час викладання навчальних дисциплін: «Супровід підсистем цільового навантаження БПЛА», «Машинне навчання та розпізнавання образів», «Мова Python для глибокого навчання»;
9. результати дисертаційного дослідження впроваджено у виробничий процес компанії «Рейндорф» під час створення безпілотного літального апарата для потреб виконання спеціальних місій в умовах дії військового стану за ТУ У 30.0-44764773-006:2024;
10. окремі результати дисертаційного дослідження відображено у патенті на корисну модель «Нейромережева 43-класова модель розпізнавання даних аерозйомки з камер безпілотних літальних апаратів».

Особистий внесок здобувача. Усі основні результати, викладені в дисертації, отримано здобувачем особисто. У працях, опублікованих у співавторстві, здобувачеві належать постановка задачі та її формалізація, збирання і підготовка даних аерозйомки, побудова та навчання нейромережевих моделей, розроблення програмних модулів, проведення обчислювальних і натурних експериментів, аналіз отриманих результатів та підготовка матеріалів до публікації.

Апробація результатів дисертації. Основні положення та результати дисертаційної роботи доповідалися й обговорювалися на наукових конференціях і форумах, зокрема: IX International Conference Information Technology and

Implementation (IT&I, 2022), 12th International Conference on Advanced Computer Information Technologies (ACIT, 2022), International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications (ICTERI, 2022), 2024 IEEE International Conference on Actual Problems of Unmanned Aerial Vehicles Development (APUAVD, 2024), International Conference on Cyber Hygiene & Conflict Management in Global Information Networks (CH&CMiGIN, 2024), а також на науково-практичних заходах «Aviation in the 21-st Century» (2022), «ABIA» (2023) та конференціях «ПОЛІТ» (2023–2025).

Публікації. Основні наукові результати дисертації опубліковано у 6 наукових працях, з яких 1 стаття - у фаховому виданні України, 5 публікацій - у матеріалах міжнародних конференцій/виданнях, індексованих у міжнародних наукометричних базах. Додатково результати відображено у 5 тезах доповідей.

Структура та обсяг дисертації. Дисертація складається з анотації, вступу, чотирьох розділів, висновків та списку використаних джерел. Загальний обсяг дисертації становить 173 сторінки, із них 142 основного тексту. Робота містить 58 рисунків і 42 таблиці. Список використаних джерел налічує 87 найменувань.

РОЗДІЛ 1. ГЛИБИННЕ МЕТРИЧНЕ НАВЧАННЯ ДЛЯ ЗАДАЧ АБСОЛЮТНОЇ ЛОКАЛІЗАЦІЇ

Традиційно метричне навчання використовувалося як метод створення оптимальної міри подібності, що враховує специфічні властивості та розподіл точок вхідних даних. У задачах метричного навчання подібність між об'єктами часто задають через функцію відстані у просторі представлень, де малі значення відстані відповідають подібним об'єктам, а великі - неподібним. З часом акцент у дослідженнях змістився від підбору самої метрики до отримання представлень, що оптимізовані під задану метрику. Методи глибинного навчання дають змогу моделювати складні нелінійні залежності в даних. Для задач метричного навчання ключовим стає проєктування функції втрат і стратегії формування вибірки, що безпосередньо впливають на структуру простору представлень. У цьому варіанті якість розпізнавання визначається вже не стільки вибором міри схожості, скільки властивостями відображення, яке створює стійкі та глобальні ознаки. Такий підхід отримав назву глибинного метричного навчання [47, 48, 51, 52].

Візуальне розпізнавання місць є окремим випадком метричного навчання над зображеннями. Метою є побудова перетворення, яке відображає зображення у латентному просторі так, щоб вектори ознак семантично подібних зображень були розташовані близько один до одного, а вектори ознак неподібних зображень - максимально віддалено. Для забезпечення цих характеристик часто застосовують методи глибинного навчання та комп'ютерного зору, які забезпечують формування інваріантних ознак і дають змогу моделювати нелінійні залежності в даних [7, 8, 11, 14].

Одним із варіантів дослідження розподілів представлень є використання згорткових автоенкодерів для формування компактних латентних представлень і подальшого аналізу розподілів даних. Такий підхід дає змогу кількісно зіставляти розподіли, швидко формулювати гіпотези щодо подальшого балансування навчальної вибірки та формалізувати критерії якості даних [43, 46, 34, 35].

У контексті БПЛА візуальне розпізнавання є методом абсолютної локалізації, який встановлює відповідність між аерознімками та геоприв'язаними зображеннями місцевості, відносячи їх до глобальної системи координат Землі. Основна методологічна складність полягає у виразному доменному зсуві між цими даними: джерела відрізняються ракурсами та висотами зйомки, просторовою і спектральною роздільною здатністю сенсорів, сезонністю, фотометричними умовами, наявністю тіней, оклюзій і паралаксу. Для навчання такої моделі необхідно побудувати якісний навчальний набір, що складається з записів бортових камер та сенсорів, а також, з урахуванням параметрів зйомки, сформувати відповідний геоприв'язаний набір супутникових зображень [7, 29, 56, 57, 58].

Підрозділ 1.1 Огляд методів метричного навчання

Виклад матеріалу у цьому розділі здійснюється в межах парадигми навчання з учителем, якщо не зазначено інше. Надалі припускаємо, що задано початковий набір даних $D = \{(X_i, y_i), i = 1, \dots, N\}$, де $X_i \in \mathbf{X}$ є входними прикладами, а $y_i \in \mathbf{Y}$ – відповідні мітки, де N - кількість елементів вибірки. Простір входних даних $\mathbf{X}$ визначається природою даних, коли вхід подано як d -вимірні числові вектори ознак, вважаємо $\mathbf{X} \subseteq \mathbb{R}^d$ - евклідовий простір розмірності d . Простір виходу $\mathbf{Y}$ залежить від конкретної постановки задачі навчання з учителем. Метою навчання є побудова відображення $f: \mathbf{X} \rightarrow \mathbf{Y}$, яке відносить входні приклади до їх міток [47, 48].

Підрозділ 1.1.1 Лінійні методи метричного навчання

Метричне навчання - це задача машинного навчання, що полягає у побудові функцій відстані на основі заданої вибірки. Отримані відстані надалі використовують для підвищення ефективності алгоритмів машинного навчання. Для довільної непорожньої множини $\mathbf{X}$ відстанню (метрикою) називають будь-яку функцію $d: \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$, яка задовольняє таким властивостям [80]:

1. Властивість невід'ємності: $d(X, Y) \geq 0$, для усіх $X, Y \in \mathbf{X}$.
2. Тотожність: $d(X, Y) = 0 \Leftrightarrow X = Y$, для усіх $X, Y \in \mathbf{X}$.

3. Симетричність: $d(X, Y) = d(Y, X)$, для усіх $X, Y \in \mathbf{X}$.

4. Нерівність трикутника: $d(X, Z) \leq d(X, Y) + d(Y, Z)$, для усіх $X, Y, Z \in \mathbf{X}$.

Якщо функція не задовольняє властивість тотожності, але задовольняє решті трьом, то її називають псевдометрикою або псевдовідстанню. Оскільки в більшості методів навчання метрик на практиці отримують саме такі функції, у цьому контексті їх зазвичай також називають метриками. Робота з неперервними даними в евклідових просторах є традиційним підходом, оскільки будь-який інший тип даних зазвичай можна перетворити на числовий за допомогою відповідних інструментів, а можливість визначення відстаней, придатних для оптимізації, у таких просторах вища, ніж у дискретних. Більшість підходів у метричному навчанні розглядають метрику вигляду:

$$d_M(X, Y) = \sqrt{(X - Y)^T M (X - Y)}, \quad (1.1)$$

що є відстанню Махаланобіса, де $M \in M_d(\mathbb{R})_0^+$ - симетрична додатно напівозначена матриця. Оскільки кожен можна подати у вигляді

$$M = L^T L, \quad L \in M_d(\mathbb{R}), \quad (1.2)$$

відстань Махаланобіса можна рівносильно записати у вигляді

$$d_M(X, Y) = \|L(X - Y)\|_2, \quad (1.3)$$

для усіх $X, Y \in \mathbb{R}^d$ [47, 48, 81]. Тобто як евклідову відстань після перетворення даних матрицею L . Таким чином, задачу навчання відстані Махаланобіса можна звести до задачі навчання лінійного перетворення $L \in M_d(\mathbb{R})$. Вибір між використанням параметризації через M або L здебільшого є питанням зручності й залежить від алгоритму. Використання M часто приводить до задач опуклої оптимізації та забезпечує гарантії оптимальності, тоді як параметризація через L зазвичай дає змогу виконувати зменшення розмірності, пришвидшувати обчислення й очищувати дані [47, 48].

Методи метричного навчання з учителем будують метрику, використовуючи додаткову інформацію про подібність та неподібність об'єктів, задану у вигляді обмежень. У загальному випадку найпоширенішими є:

1. Обмеження на подібні та неподібні пари.

Нехай

$$S, U \subset \{1, \dots, N\} \times \{1, \dots, N\}, \quad (1.4)$$

множини пар подібних та неподібних об'єктів відповідно, тобто $(i, j) \in S$, якщо $y_i = y_j$, і $(i, j) \in U$, якщо $y_i \neq y_j$. Для цих пар накладають умови

$$d_M(X_i, X_j) \leq u_{ij}, (i, j) \in S, \quad (1.5)$$

$$d_M(X_i, X_j) \geq l_{ij}, (i, j) \in U, \quad (1.6)$$

де $u_{ij}, l_{ij} \in \mathbb{R}$ задають, відповідно, верхні та нижні порогові значення для відстаней [47, 48]. На рисунку 1.1 наведено ілюстрацію методу обмеження подібності та неподібності пар об'єктів.

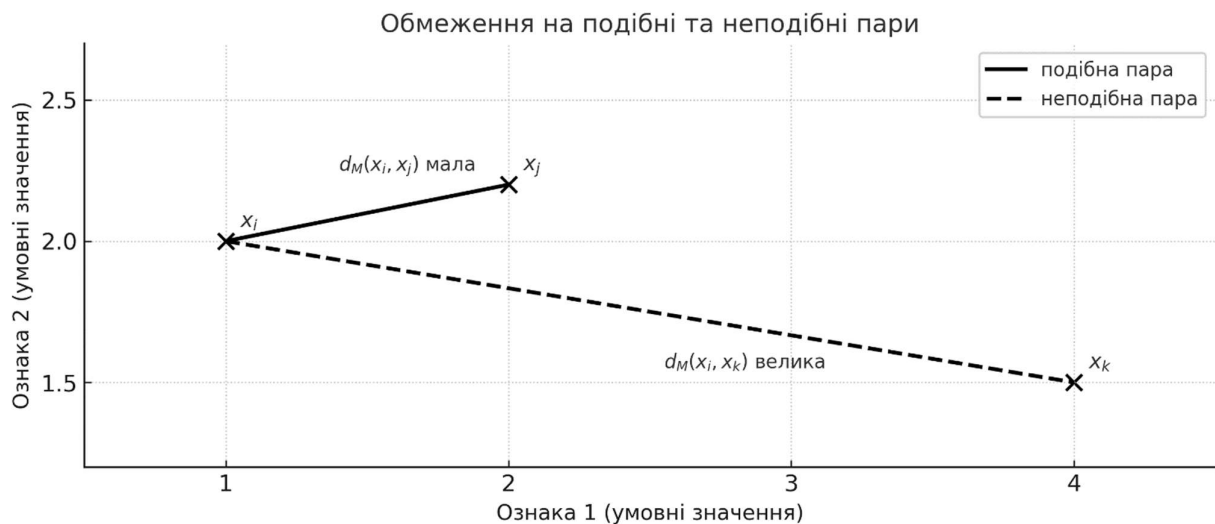


Рис. 1.1. Ілюстрація методу обмежень на подібні та неподібні пари

2. Відносні обмеження.

Нехай

$$R \subset \{1, \dots, N\} \times \{1, \dots, N\} \times \{1, \dots, N\}, \quad (1.7)$$

множина індексів трійок (i, j, k) , для яких об'єкт X_i має бути більш подібним до X_j , ніж до X_k . Для кожного $(i, j, k) \in R$ накладають умову

$$d_M(X_i, X_j) + m \leq d_M(X_i, X_k), \quad (1.8)$$

де $m > 0$ – параметр відступу, який зазвичай обирають рівним 1 [47, 48]. На рисунку 1.2 наведено ілюстрацію методу відносних обмежень.

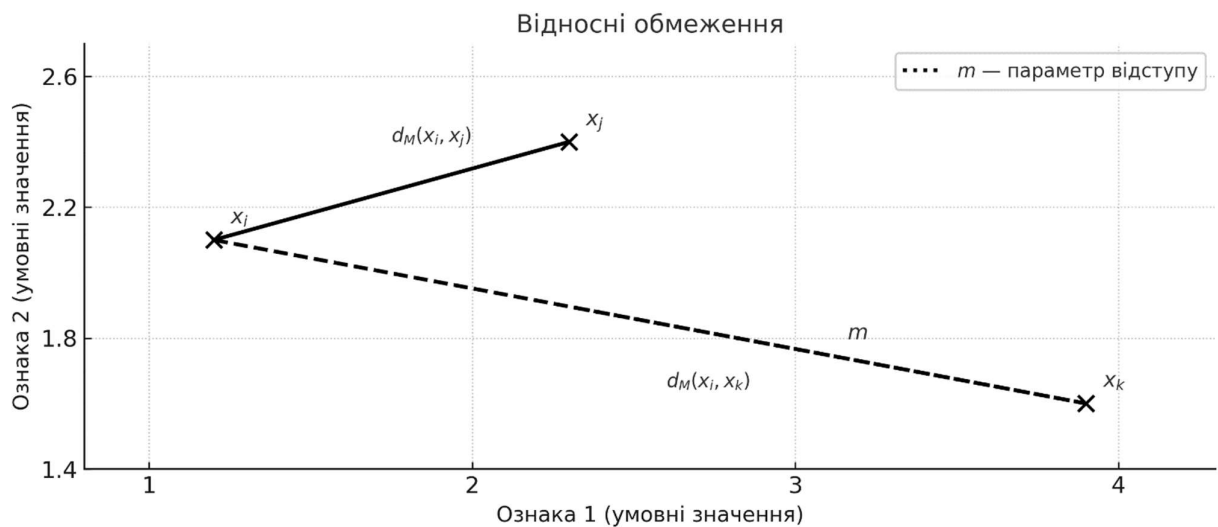


Рис. 1.2. Ілюстрація методу відносних обмежень

Описані вище методи ґрунтуються на лінійному перетворенні простору ознак. Однак у багатьох задачах відношення подібності є нелінійними, що обмежує застосування лінійних метрик. Це зумовлює перехід до нелінійних підходів метричного навчання, зокрема на основі метрики відстані ядра та глибинних методів.

Підрозділ 1.1.2 Нелінійні методи метричного навчання

Нелінійні методи метричного навчання поділяються на кілька основних груп [47, 48]:

1. Навчання метрики відстані ядра.

Під ядром $k: \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$ розумітимемо симетричну додатно напівозначену функцію, що задає скалярний добуток між образами у просторі представлень. У цьому підході вихідні дані неявно відображаються у простір вищої розмірності, де їх структура стає ближчою до лінійно відокремленої. Метрика навчається вже у цьому новому просторі. Для уникнення явного обчислення відображення застосовується ядровий прийом, аналогічно до методів опорних векторів, що дає змогу працювати лише з матрицею ядра.

2. Локальні методи метричного навчання.

У цьому підході замість однієї глобальної метрики використовуються кілька окремих метрик, кожна з яких визначається тільки в конкретній області простору ознак. Такий підхід використовується, коли розподіл даних неоднорідний і неможливо описати глобальну структуру за допомогою однієї лінійної метрики.

3. Глибинне метричне навчання.

У глибинних методах метричного навчання метрика визначається за допомогою нелінійного відображення, яке задається нейронною мережею. Вхідні об'єкти перетворюються у простір векторних представлень, де семантичну схожість об'єктів відображають евклідова або косинусна відстань. Цей підхід є одним із найпопулярніших напрямів досліджень останніх років і розглядається окремо від традиційних лінійних методів метричного навчання. Це пов'язано з розвитком нейронних мереж та відповідних апаратних рішень. Навчання представлень у глибинних нейронних мережах дає змогу визначати складні, нелінійні закономірності в даних, які традиційні лінійні методи метричного навчання не завжди здатні правильно описати. На відміну від підходів, що оптимізують лише точність класифікації, метричне навчання формує простір представлень, у якому близькість векторів узгоджується з подібністю об'єктів. Це підвищує роздільність між класами в задачах пошуку та зіставлення. Відмінність глибинного метричного навчання від методів глибинного навчання полягає у

способі навчання мережі та у виборі функції втрат, орієнтованої на відношення подібності та неподібності між об'єктами.

Підрозділ 1.2 Огляд функцій втрат у глибинному метричному навчанні

У глибинному метричному навчанні метою є побудова такого відображення $f_\theta : \mathbf{X} \rightarrow \mathbb{R}^d$, щоб у просторі представлень подібні об'єкти розташовувались близько один до одного, а неподібні – далі. Ця вимога формалізується через функцію втрат, що керується навчанням параметрів θ .

Більшість функцій втрат у глибинному метричному навчанні можна поділити на дві умовні групи [51-55]:

1. Функції втрат, що безпосередньо оперують подібністю між окремими зразками.
2. Проксі-орієнтовані функції втрат, у яких порівняння відбувається між зразками та спеціально введеними векторами-проксі.

Підрозділ 1.2.1 Контрастивна та триплетна функції втрат

Однією з найпростіших функцій втрат для метричного навчання є контрастивна функція втрат [51]. Нехай маємо два векторних представлення $f_\theta(X_i)$ та $f_\theta(X_j)$ об'єктів X_i та X_j , а двійкова міра подібності задається як:

$$y_{ij} = \begin{cases} 1, \text{ якщо } X_i \text{ та } X_j \text{ належать до одного класу,} \\ 0, \text{ якщо } X_i \text{ та } X_j \text{ належать до різних класів.} \end{cases} \quad (1.9)$$

Тоді контрастивна функція втрат має вигляд:

$$L_{ij} = \begin{cases} \|f_\theta(X_i) - f_\theta(X_j)\|^2, & \text{якщо } y_{ij} = 1; \\ \max(0, \alpha - \|f_\theta(X_i) - f_\theta(X_j)\|)^2, & \text{якщо } y_{ij} = 0; \end{cases} \quad (1.10)$$

де $\alpha > 0$ – параметр зсуву, який задає бажану мінімальну відстань між представленнями об'єктів різних класів. Недоліком (1.10) є те, що дана функція

враховує інформацію лише з парами зразків і не враховує відносні співвідношення між різними парами одночасно.

Щоб враховувати відносні співвідношення й забезпечити глобальну геометрію простору представлень, застосовують потрібну функцію втрат [52]. Нехай a, p, n – індекси опорного, позитивного та негативного зразків відповідно. Базова форма має вигляд:

$$L_{apn} = \max(0, \|f_{\theta}(X_a) - f_{\theta}(X_p)\|^2 - \|f_{\theta}(X_a) - f_{\theta}(X_n)\|^2 + \alpha). \quad (1.11)$$

Загальна втрата визначається як середнє за множиною інформативних трійок T :

$$L = \frac{1}{|T|} \sum_{apn \in T} L_{apn}. \quad (1.12)$$

При $L_{apn} > 0$ похідні мають вигляд:

$$\frac{\partial L_{apn}}{\partial f_{\theta}(X_a)} = 2(f_{\theta}(X_n) - f_{\theta}(X_p)), \quad (1.13)$$

$$\frac{\partial L_{apn}}{\partial f_{\theta}(X_p)} = 2(f_{\theta}(X_p) - f_{\theta}(X_a)), \quad (1.14)$$

$$\frac{\partial L_{apn}}{\partial f_{\theta}(X_n)} = 2(f_{\theta}(X_a) - f_{\theta}(X_n)). \quad (1.15)$$

Тобто опорний елемент вибірки відштовхується від негативного та притягує позитивний, позитивний рухається до опорного, а негативний від опорного. Водночас потрібна функція втрат не накладає прямої умови на відстань між позитивним і негативним зразками, тому їх рознесення відбувається через опорний елемент вибірки.

Зазвичай на практиці використовують $L2$ нормування представлень $\hat{f}(X) = \frac{f_\theta(X)}{\|f_\theta(X)\|}$ після чого застосовується косинусна подібність, що для потрійної функції втрат має вигляд [52]:

$$L_{apn}^{\cos} = \max(0, \hat{f}(X_a)^T \cdot \hat{f}(X_n) - \hat{f}(X_a)^T \cdot \hat{f}(X_p) + \alpha). \quad (1.16)$$

Параметр α задає мінімальне бажане рознесення $d(a, p) + \alpha \leq d(a, n)$, де d позначає обрану відстань у просторі представлень. Занадто мала α утворює слабку дискримінативність, а занадто велика ускладнює оптимізацію. Ефективність навчання суттєво залежить від інформативності трійок, тому існують підходи, що дозволяють якісно побудувати набір даних, зокрема: важкий відбір – для кожного зразка береться найвіддаленіший позитив і найближчий негатив, середній відбір – зазвичай негативні зразки, що знаходяться далі від позитивного, але ближче за заданий параметр α , зважений за відстанню відбір – відбір негативних зразків, який залежить від відстані до опорного зразка.

Підрозділ 1.2.2 Проксі-орієнтовані функції втрат

Як було зазначено у попередньому розділі, класичні функції втрат глибинного метричного навчання ґрунтуються на безпосередньому порівнянні елементів вибірки між собою. Для задачі з N елементами вибірки кількість можливих пар має порядок $O(N^2)$, а кількість можливих трійок - $O(N^3)$. Під час стохастичної оптимізації, коли на кожному кроці використовується мала підвибірка розміру B , модель тренується лише на $O(B^2)$ пар або $O(B^3)$ трійок. Щоб у процесі навчання фактично використати більшу частину інформативних трійок, потрібна значна кількість ітерацій, а збіжність моделі безпосередньо залежить від того, наскільки ефективно формуються ці пари та трійки. Це зумовлює характерну для пароорієнтованих методів проблему вузького місця під час добору інформативних прикладів [49, 53, 54, 55].

Для вирішення цієї проблеми були запропоновані проксі-орієнтовані методи. Їх ідея полягає в тому, що замість явного оперування всіма можливими парами або трійками вводиться окрема множина векторів-проксі, які є параметрами моделі та спільно використовуються усіма елементами вибірки. Кожен вхідний вектор X_i у процесі навчання апроксимується один або декількома проксі p_c , а функція втрат обчислюється за відстанями представлень між $f_\theta(X_i)$ та відповідними проксі. Оскільки вектори проксі охоплюють усю вибірку, відпадає потреба в явному доборі інформативних пар або ж трійок, натомість модель навчає компактну систему прототипів, до яких наближуються подібні приклади та від яких відштовхуються неподібні.

Одним із таких підходів є аналіз сусідніх компонент або ж NCA [49]. Нехай $f_\theta(X_i)$ та $f_\theta(X_j)$ представлення елементів вибірки X_i та X_j . Визначимо ймовірність вибору X_j як сусіднього для X_i як:

$$p_{ij} = \frac{\exp(-D_e(f_\theta(X_i), f_\theta(X_j)))}{\sum_{k \neq i} \exp(-D_e(f_\theta(X_i), f_\theta(X_k)))}, \quad (1.17)$$

де D_e – квадратична евклідова відстань у просторі представлень. Така ймовірність характеризує ступінь належності до околу, чим ближче $f_\theta(X_j)$ до $f_\theta(X_i)$, тим більшою є ймовірність p_{ij} . Метою NCA є максимізація сумарної ймовірності того, що елемент вибірки X_i має сусідів того самого класу. Відповідно функція втрат NCA для X_i матиме вигляд [49]:

$$L_{NCA} = -\log \left(\frac{\sum_{j: y_i = y_j, i \neq j} \exp(-D_e(f_\theta(X_i), f_\theta(X_j)))}{\sum_{k \neq i} \exp(-D_e(f_\theta(X_i), f_\theta(X_k)))} \right), \quad (1.18)$$

Пряма оптимізація (1.18) є обчислювально витратною, оскільки вимагає розгляду всіх пар i, k для кожного i . Для зменшення складності запропоновано метод ProxynCA [53], у якому елементи вибірки у знаменнику замінюються на

вектори проксі. Нехай задано множину проксі $P = \{p_c, c = 1, \dots, C\}$, де $p_c \in \mathbb{R}^d$, які інтерпретуються як прототипи класів. Для X_i з міткою y_i використовують відповідний проксі p_{y_i} . Позначимо нормовані проксі як:

$$\hat{f}_i = \frac{f_\theta(X_i)}{\|f_\theta(X_i)\|}, \hat{p}_c = \frac{p_c}{\|p_c\|}. \quad (1.19)$$

Тоді проксі орієнтована функція втрати для елемента вибірки X_i набуває вигляду:

$$L_{PNCA}^{(i)} = -\log \left(\frac{\exp(-D_e(\hat{f}_i, \hat{p}_{y_i}))}{\sum_{c=1} \exp(-D_e(\hat{f}_i, \hat{p}_c))} \right), \quad (1.20)$$

де C – кількість класів [53]. Функція (1.20) максимізує подібність між представленнями елемента вибірки і проксі його класу та мінімізує подібність до проксі інших класів, замінюючи обчислення над усіма сусідніми елементами на обчислення над обмеженою множиною проксі.

Подальший розвиток ProхуNCA реалізовано у методі ProхуNCA++ [54], де вводиться явне моделювання ймовірності призначення проксі із використанням нормованої експоненційної функції з масштабуванням. Нехай $T > 0$ параметр масштабу, ймовірність належності представлення $\hat{f}_i$ до проксі $\hat{p}_c$ виглядає як:

$$q_{ic} = \frac{\exp(-D_e(\hat{f}_i, \hat{p}_c) / T)}{\sum_{k=1} \exp(-D_e(\hat{f}_i, \hat{p}_k) / T)}. \quad (1.21)$$

Тоді функція втрат ProхуNCA++ можна записати у вигляді:

$$L_{PNCA++} = -\log q_{iy_i}, \quad (1.22)$$

тобто максимізується ймовірність, що представлення $\hat{f}_i$ буде правильно зіставлене з проксі свого класу [54, 55]. Параметр T визначає ступінь концентрації вихідного розподілу ймовірностей. Зі збільшенням T розподіл згладжується та наближається до рівномірного, тоді як зі зменшенням цього параметра він стає більш

концентрованим на найбільш імовірних класах. Це забезпечує чіткіше розмежування між класами. До цього класу належить також функція втрат Proxu Anchor Loss [55], у якій оптимізацію здійснюють відносно навчуваних проксі-якорів без явного перебирання трійок.

Таким чином, проксі-орієнтовані методи дозволяють зменшити обчислювальну складність порівняно з пароорієнтованими підходами та усунути проблему з добором інформативних пар чи трійок. Наявність спільної множини проксі, що оновлюється під час навчання, забезпечує більш стабільні градієнти та, як правило, швидшу збіжність глибинних метричних моделей на великих вибірках.

Підрозділ 1.3 Глибинні нейронні мережі для формування ознак

У попередніх розділах увагу було зосереджено на загальній постановці задачі глибинного метричного навчання та функціях втрат, що формують геометрію простору представлень. У цьому розділі розглядаються архітектури глибинних нейронних мереж, які реалізують відображення $f_{\theta} : \mathbf{X} \rightarrow \mathbb{R}^d$ для зображень. Проаналізовано архітектури нейромереж з автоенкодером та їх модифікації, згорткові й сіамські мережі, а також візуальні трансформери.

Підрозділ 1.3.1 Автоенкодери та їх підвиди

Автоенкодери належать до класу глибинних нейронних мереж, метою яких є навчання відновлювати вхідні дані в режимі навчання без учителя [43, 46]. Структурно така модель складається з двох частин: енкодера, що перетворює вхідний об'єкт $X \in \mathbf{X}$ у вектор представлення $f_{\theta}(X) \in \mathbb{R}^d$ та декодера, який за відповідним представленням відтворює наближення $\hat{X}$ до початкового X . Навчання здійснюється шляхом мінімізації реконструкції на основі функції втрати $L(X, \hat{X})$, внаслідок чого формується компактний простір представлень, у якому зберігаються найінформативніші ознаки. На рисунку 1.3 наведено типовий вигляд мережі автоенкодера.

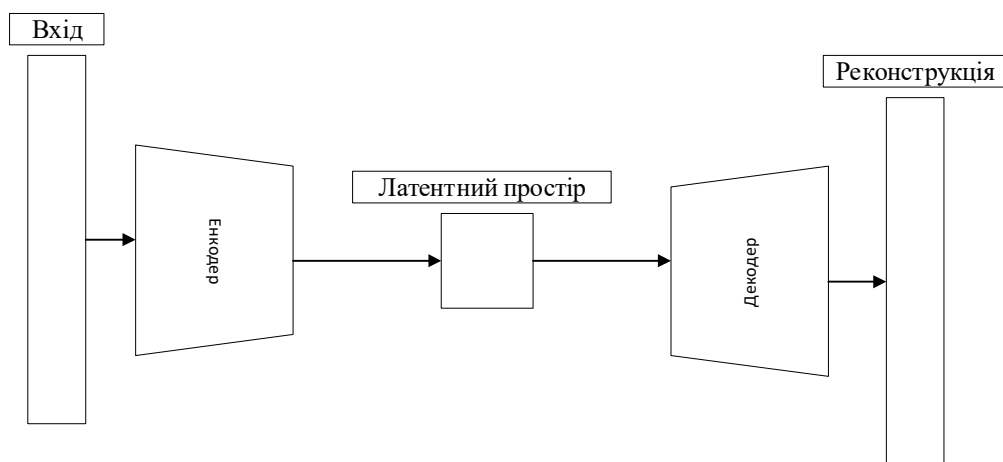


Рис. 1.3. Типова структура мережі автоенкодера

Найпростішим різновидом є автоенкодер зі зменшеною розмірністю латентного простору, у якому відповідно латентний шар має меншу розмірність, ніж вхідний простір [82]. Таке стиснення обмежує пропускну здатність моделі й перешкоджає простому відтворенню тотожного відображення. Щоб досягти малої похибки реконструкції, мережа повинна відкидати надлишкові деталі та зберігати лише ті компоненти, що є суттєвими для відновлення. Водночас за надмірної кількості шарів енкодера та декодера, навіть за наявності вузького латентного шару, модель усе одно здатна апроксимувати тотожність, що обмежує користь такого підходу й вимагає додаткового структурування латентного простору.

Цю проблему частково вирішують регуляризовані автоенкодери [82], у яких до реконструкційної втрати додаються спеціальні штрафні доданки. Вони не лише зменшують ризик перенавчання моделі, а й задають бажані властивості простору представлень. Однією з таких модифікацій є розріджені автоенкодери, у яких інформаційне обмеження реалізується не через зменшення кількості нейронів у прихованих шарах, а шляхом накладання обмеження на частку нейронів, що можуть бути активними одночасно. Це забезпечує збільшення ємності моделі, але змушує мережу вмикати лише ті нейрони, що відповідають суттєвим ознакам конкретного прикладу. Як регуляризатор використовують $L1$ норму або розбіжність Кульбака-Лейблера між бажаним і фактичним розподілом активацій.

Іншим різновидом регуляризованих архітектур є контрактні автоенкодер, у яких простір представлень роблять малочутливим до незначних збурень вхідних даних. Це досягається додаванням функції втрат штрафного доданка, що залежить від норми Якобіана відображення енкодера. Зазвичай використовують Фробеніусову норму Якобіана, яка характеризує середню величину зміни відображення за малих змін входу. Мінімізація такого доданка змушує енкодер приглушувати несуттєві варіації і концентруватись на стійких, структурних ознаках [82].

Знешумлювальні автоенкодері спрямовані на формування представлень стійких до шуму [44]. На відміну від стандартної схеми побудови АЕ, на вхід енкодера подаються спотворені $\tilde{X}$, отримані шляхом додавання шуму до оригінального X . Таким чином мережа навчається відновлювати інформативну структуру сигналу та ігнорувати випадкові шуми. Окрім зниження ризику перенавчання, це робить ЗАК ефективним інструментом попередньої обробки та очищення зображень [44].

Окремий клас становлять варіаційні автоенкодері, у яких латентний простір моделюється не як окремий вектор представлення, а як параметризований ймовірнісний розподіл [45]. На виході енкодера формуються два вектори параметрів латентної змінної $Z \in \mathbb{R}^m$: вектор середніх $\mu(X) \in \mathbb{R}^m$ та вектор стандартних відхилень $\sigma(X) \in \mathbb{R}^m$. Випадкова латентна змінна генерується за схемою:

$$Z = \mu(X) + D_{\sigma(X)} \cdot \varepsilon, \quad (1.23)$$

де $\varepsilon \sim N(0, I_m)$, а

$$D_{\sigma(X)} = \text{diag}(\sigma_1(X), \sigma_2(X), \dots, \sigma_m(X)),$$

діагональна матриця, побудована з компонент вектора стандартних відхилень $\sigma(X)$. Відповідний наближений апостеріорний розподіл має вигляд

$$q_{\theta}(Z | X) = N(\mu(X), \text{diag}(\sigma^2(X))).$$

Функцію втрат VAE можна записати як суму реконструкційного та регуляризаційного доданку [45]:

$$L_{VAE} = L_{rec}(X, \hat{X}) + D_{KL}(q_{\theta}(Z | X) || p(Z)), \quad (1.24)$$

де $L_{rec}(X, \hat{X})$ – реконструкційна складова, $q_{\theta}(Z | X)$ – апостеріорний розподіл у латентному просторі, що параметризується енкодером, $p(Z)$ – апріорний розподіл, D_{KL} – розбіжність Кульбака-Лейблера.

Різні модифікації автоенкодерів відрізняються способом накладання обмежень на латентні змінні та вибором регуляризації. Саме ці рішення визначають, наскільки стійкими й інформативними будуть отримані представлення.

Підрозділ 1.3.2 Згорткові та сіамські нейронні мережі

У задачах метричного навчання згорткові нейронні мережі розглядаються насамперед як енкодери, що реалізують відображення $f_{\theta}(X) \rightarrow \mathbb{R}^d$ для зображень. Замість використання вихідного класифікаційного шару, як у звичайних задачах розпізнавання, у метричному підході з мережі беруть вектор ознак з проміжного шару, що зазвичай є передостаннім. Такий вектор $f_{\theta}(X)$ надалі розглядається як точка у просторі представлень, на якій уже діє метрична функція втрат, розглянута у розділі 1.2. Поширеною практикою є попереднє навчання згорткового енкодера в режимі класифікації, після чого класифікаційний шар вилючають, а параметри θ донавчають уже з використанням метричних втрат [42].

Однак для метричного навчання недостатньо просто оптимізувати параметри θ , потрібно не лише розділити класи, а й задати бажану структуру відстаней між окремими елементами вибірки. Саме для цього запропоновані сіамські архітектури, у яких кілька екземплярів одного й того самого енкодера f_{θ} застосовуються до різних вхідних елементів [50, 51]. У найпростішому випадку попарного навчання на входи подаються зразки X_i і X_j , обчислюються відповідні представлення $f_{\theta}(X_i)$

та $f_{\theta}(X_j)$, а далі обчислюється відстань між ними, яка входить до контрастивної функції втрат. Схематичне подання такої моделі наведено на рис. 1.4.

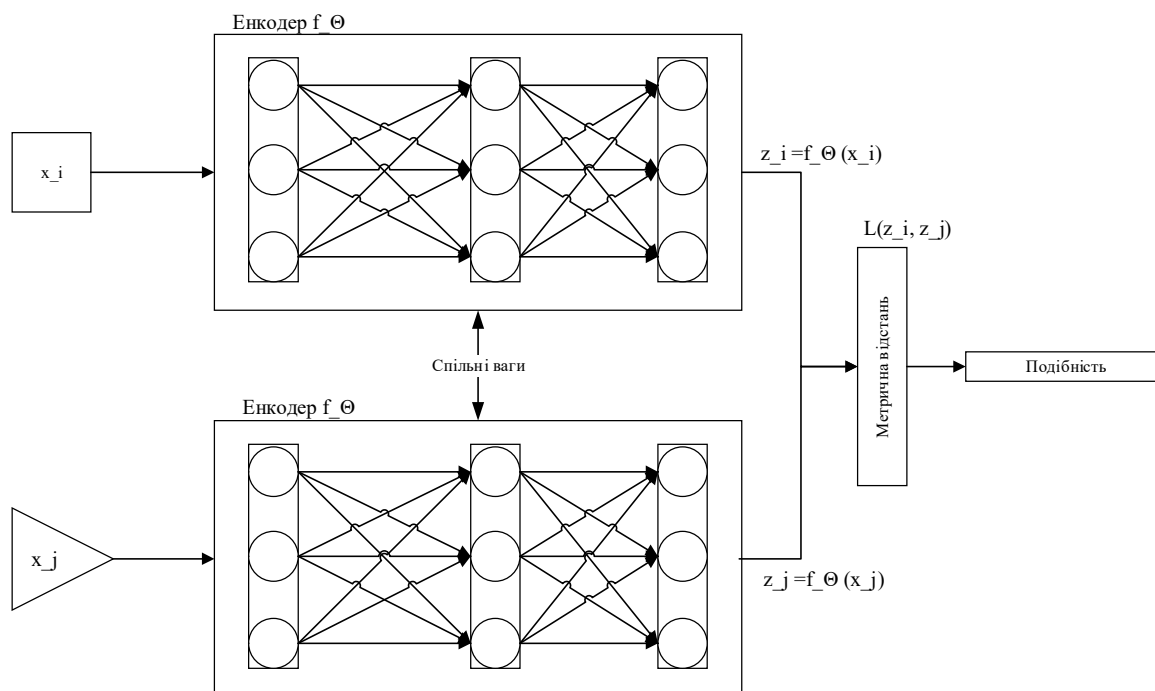


Рис. 1.4. Схема сіамської мережі для попарного навчання

У потрібних схемах використовують три незалежні застосування того ж енкодера для трійки, де опорний, позитивний і негативний зразки разом формують навчальний сигнал для потрібної функції втрат. Таким чином, згортковий енкодер забезпечує побудову простору представлень, а сіамська організація навчання із відповідними метричними втратами задає його геометрію з погляду подібності та неподібності елементів вибірки.

Підрозділ 1.3.3 Візуальні трансформери

Останніми роками суттєвого розвитку набули візуальні трансформери (ViT), у яких для побудови представлення зображення використовується трансформерна архітектура, запозичена з задач обробки природної мови (NLP) [41]. Ключова відмінність від згорткових підходів полягає в тому, що просторові залежності між фрагментами зображення моделюються не фіксованими згортками, а механізмом уваги, який дає змогу кожному фрагменту враховувати інформацію про будь-який

інший фрагмент зображення. Узагальнену схему архітектури ViT наведено на рис. 1.5.

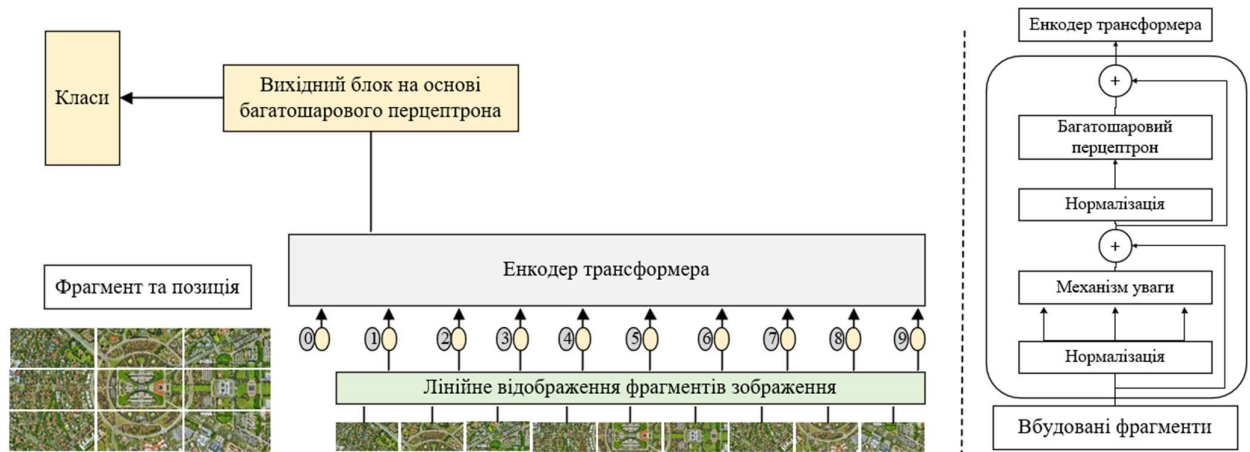


Рис. 1.5. Узагальнена архітектура ViT

Нехай $X \in \mathbf{X}$ - вхідне зображення розміру $H \times W \times C$, де H - висота зображення, W - ширина зображення, C - канал. На першому етапі ViT розбиває зображення на фрагменти без перекриття $P \times P$. Кожен фрагмент $X_k = \left\{ k = 1, \dots, N; N = \left(\frac{H}{P} \right) \cdot \left(\frac{W}{P} \right) \right\}$ розгортається у вектор довжини $P^2 \cdot C$ і через лінійне перетворення відображається в простір $\mathbb{R}^d$:

$$z_k = W_e \text{vec}(X_k) + b_e, z_k \in \mathbb{R}^d, \quad (1.25)$$

де W_e та b_e - параметри фрагментів енкодера, а d - розмірність простору представлень. Отримана послідовність векторів z_k інтерпретується як аналог послідовності опорних дескрипторів у трансформері для тексту.

Оскільки трансформер не враховує порядок елементів послідовності, до кожного вектора z_k додається вектор позиційного кодування $e_{pos,k}$. У більшості реалізацій ViT позиційні вектори є навчуваними параметрами моделі:

$$\tilde{z}_k = z_k + e_{pos,k}. \quad (1.26)$$

Крім того, до початку послідовності додають спеціальний вектор $\tilde{z}_{cls}$ який протягом проходження через усі шари трансформера акумулює інформацію про усе зображення. Послідовність $(\tilde{z}_{cls}, \tilde{z}_1, \dots, \tilde{z}_N)$ подається на вхід каскаду з L однакових енкодерних блоків трансформера. Кожен блок складається з двох основних частин: механізму уваги та двошарового повнозв'язного перцептрона, між якими застосовуються залишкові зв'язки й нормалізація шару. Механізм самоуваги забезпечує для кожного елемента послідовності $\tilde{z}_k$ зважене агрегування інформації з усіх інших елементів, таким чином модель може явно моделювати просторові залежності між усіма фрагментами зображення. Після проходження через L трансформерних блоків отримуються оновлені вектори $(\tilde{h}_{cls}, \tilde{h}_1, \dots, \tilde{h}_N)$. Вектор, що відповідає спеціальному дескриптору, розглядають як компактне представлення всього зображення й ототожнюють результатами енкодера [41]:

$$f_{\theta}(X) = \tilde{h}_{cls} \in \mathbb{R}^d. \quad (1.27)$$

У класичній постановці задачі класифікації даний вектор подають на багатошаровий перцептрон з подальшим застосуванням нормованої експоненційної функції. У задачах метричного навчання $f_{\theta}(X)$ безпосередньо використовується як точка в просторі представлень, до якої застосовують метричні функції втрат. Порівняно зі згортковими енкодерами трансформери не мають явної локальної та трансляційної інваріантності, а отже здатні більш гнучко пристосовуватися до структури даних, але водночас потребують більших обсягів навчальних даних або попереднього навчання на великих наборах даних. Натомість глобальний характер механізму уваги та гарна масштабованість з глибиною мережі роблять ViT перспективною архітектурою для задач, у яких критично важливо моделювати взаємодію віддалених фрагментів сцени.

Підрозділ 1.4 Візуальне розпізнавання місць

Візуальне розпізнавання місць (VPR) розглядається як задача визначення місцезнаходження за зображенням місцевості шляхом порівняння з еталонною

базою. Нехай задано базу геоприв'язаних зображень $D = \{(I_j, g_j), \text{ для } j = 1, \dots, M\}$, де I_j це еталонне зображення, а g_j - відповідні географічні координати, а M - кількість еталонних зображень у базі. Для вхідного зображення I_q візуальне розпізнавання полягає у знаходженні індексу еталона:

$$j^* = \arg \min_{j \in \{1, \dots, M\}} (d(f_\theta(I_q), f_\theta(I_j))), \quad (1.28)$$

де f_θ - енкодер простору представлень, d - функція відстані [7]. Отримані координати g_j інтерпретуються як оцінка положення, що відповідає вхідному зображенню.

Складність задачі VPR зумовлена тим, що вигляд однієї й тієї самої місцевості може істотно змінюватися залежно від умов освітлення та погоди, пори року, ракурсу й масштабу зйомки, а також через появу рухомих об'єктів. Додатковим ускладненням є візуальна неоднозначність: зображення різних місць можуть мати подібний вигляд і, відповідно, бути близькими у просторі представлень. Тому простір представлень $f_\theta(I)$ має бути інваріантним до типових змін умов спостереження та водночас достатньо розрізнявальним, щоб відокремлювати географічно різні, але візуально подібні ділянки.

Класичні підходи VPR ґрунтуються на виявленні ключових точок і формуванні локальних дескрипторів, які описують структурні елементи сцени [1-4]. Далі локальні ознаки використовують або для встановлення відповідностей між зображеннями, або агрегують у компактний глобальний вектор, придатний для швидкого пошуку в еталонній базі [5,6]. Узагальнену схему формування локальних ознак ключових точок, характерну для методу SIFT, наведено на рис. 1.6.

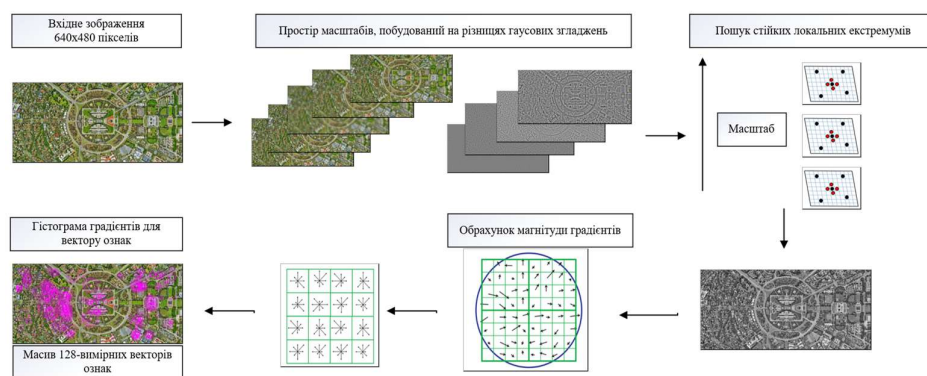


Рис. 1.6. Загальна архітектура класичних VPR підходів на основі ключових точок

Таке подання зазвичай забезпечує швидкий пошук і зберігає стійкість до змін ракурсу та часткового перекриття сцени. Водночас надмірне узагальнення локальної інформації зменшує внесок просторового розташування деталей, що підвищує ризик візуальної неоднозначності, коли різні місця мають подібний вигляд.

Методи на основі ключових точок мають практичну перевагу для прикладних систем, оскільки відповідності між локальними ознаками можна інтерпретувати та перевіряти шляхом оцінювання геометричної узгодженості. Разом із тим їх застосування обмежується залежністю якості зіставлення від кількості стійких ключових точок і від розміру вхідного зображення. Зі збільшенням кількості ключових точок зростають витрати часу на їх виявлення, обчислення локальних дескрипторів і подальше зіставлення, що критично для систем із малими допустимими затримками [1-4], особливо в режимі обробки на борту [74]. Під час зіставлення аерознімків із супутниковими еталонами ці обмеження додатково посилюються різницею в рівні деталізації та просторовій роздільній здатності між зображеннями. Супутникові знімки мають обмежену просторову роздільну здатність, тоді як аерозйомка з меншої висоти містить дрібні структурні елементи, відсутні на супутниковому рівні; унаслідок цього надійне встановлення локальних відповідностей ускладнюється. Висоту зйомки як окремих параметр доменної варіативності моделюють спеціалізовані набори даних для багатовисотних спостережень [29, 56, 57].

Подальший розвиток VPR пов'язаний із переходом до ознак, отриманих за допомогою глибоких нейронних мереж, де роль глобального дескриптора відіграє вектор представлення $f_{\theta}(I)$, сформований згортковим енкодером або візуальним трансформером. Важливою тенденцією є інтеграція операцій агрегування локальних ознак безпосередньо в архітектуру мережі та оптимізація цих операцій у межах навчання. Показовим прикладом є NetVLAD, у якому глобальний дескриптор формується параметризованим шаром агрегування типу VLAD і оптимізується за критеріями VPR [8]. Для задач пошуку зображень також застосовують узагальнене середнє об'єднання (GeM) як ефективний механізм формування глобального дескриптора. [9] На рис. 1.7 наведено візуальне порівняння поширених підходів до агрегації локальних ознак у глобальний дескриптор.

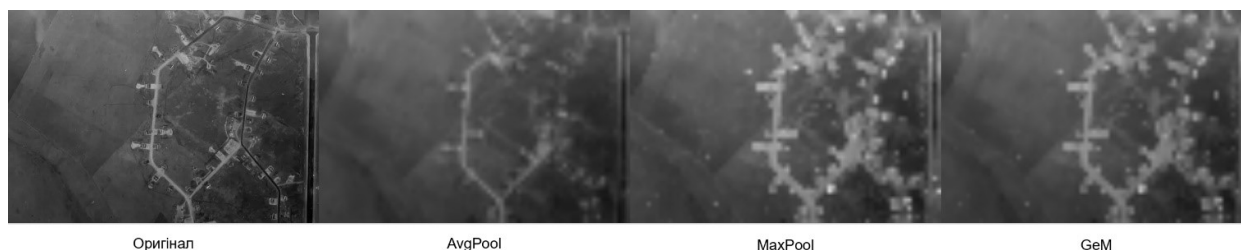


Рис. 1.7. Візуальне порівняння підходів до агрегації

Одноетапний пошук за глобальним вектором є технологічно простим і швидким, однак у задачах VPR такий підхід часто призводить до зростання кількості хибних зіставлень через візуальну подібність різних локацій. Це спостерігається у випадках, коли різні місця містять схожі структури (зокрема дороги, забудову, поля, лісосмуги), а внаслідок узагальнення локальної інформації значення міри подібності для різних місць стають близькими. У таких умовах алгоритм пошуку найближчих сусідів може повертати хибні зіставлення. Цей ефект посилюється, якщо під час агрегації втрачається інформація про взаємне розташування елементів сцени [7, 8, 9].

Для зменшення кількості хибних зіставлень широкого поширення набули двоетапні підходи, у яких пошук виконується ієрархічно. Спочатку за глобальними

ознаками відбирають обмежену множину кандидатів у всій базі, після чого уточнюють ранжування з використанням локальних ознак. Прикладом є Patch-NetVLAD: після швидкого відбору за глобальним дескриптором виконується уточнення шляхом порівняння локальних дескрипторів фрагментів зображення. Такий поділ поєднує швидкодію первинного відбору з точнішою локальною перевіркою [11]. У споріднених постановках запропоновано моделі, що поєднують глобальні й локальні ознаки в одній архітектурі, зокрема DELG [12]. Разом із тим практичним недоліком двоетапних схем є зростання обчислювальних витрат, зокрема через локальне зіставлення та можливе застосування геометричної перевірки відповідностей [11].

Окремий напрям пов'язаний із використанням універсальних попередньо навчених моделей, здатних формувати узагальнені та інформативні представлення [13]. Це відкриває можливість застосовувати щільні локальні ознаки для уточнення ранжування кандидатів із меншими часовими витратами. Разом із тим ефективне використання попередньо навчених моделей у задачах розпізнавання аерознімків потребує адаптації до доменного зсуву. Тому важливими є методи параметрично ефективною адаптації, які вводять невелику кількість додаткових параметрів і дають змогу коригувати представлення без повного донавчання базової моделі [15]. До цього напрямку належить підхід SelaVPR, у межах якого запропоновано механізм безперервної адаптації попередньо навчених моделей із використанням адаптерів, що забезпечують формування як глобальних дескрипторів, так і щільних локальних ознак. Додатково в SelaVPR використано локальний критерій, заснований на взаємності найближчих сусідів у просторі ознак; завдяки цьому знижується потреба в геометричній перевірці узгодженості під час уточнення результатів [14]. На рис. 1.8 подано узагальнену часову схему еволюції підходів до візуального розпізнавання місцевості із зазначенням ключових алгоритмів і типів методів.

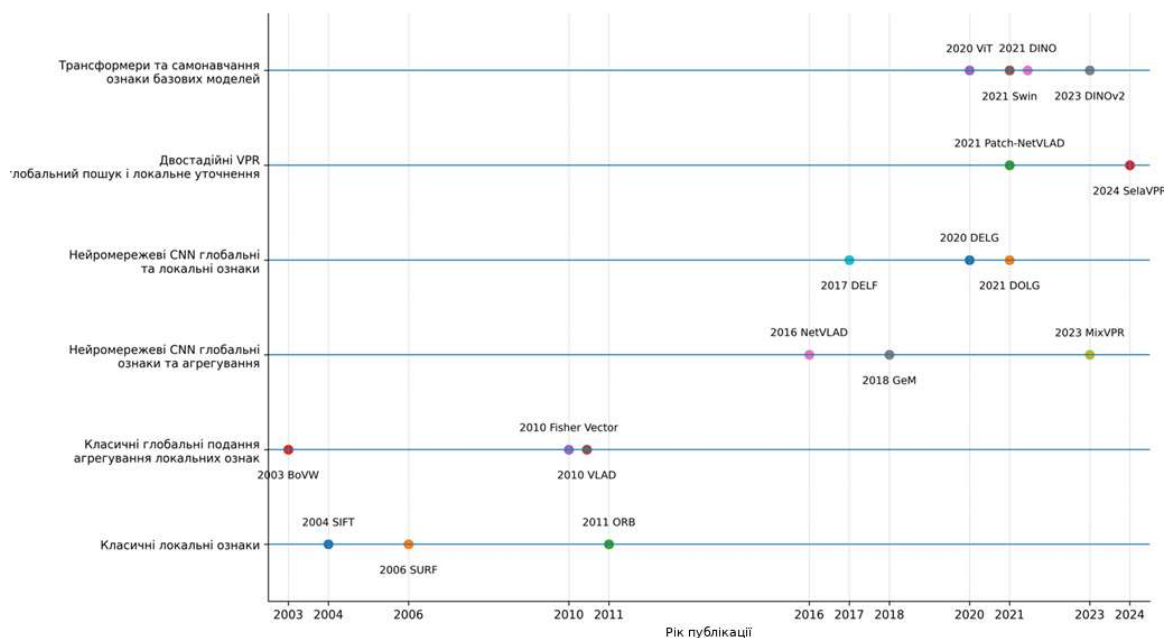


Рис. 1.8. Еволюція підходів до задачі розпізнавання місць

Узагальнення розглянутих підходів показує, що сучасний розвиток VPR спрямований у бік двоетапних схем, які поєднують швидкодію первинного відбору з підвищеною надійністю остаточного вибору за рахунок локального уточнення. Визначальним чинником ефективності першої стадії є спосіб формування глобальних ознак, оскільки саме він задає якість відбору множини кандидатів [7, 11, 14]. Отже, ключовим практичним завданням є вибір збалансованого рішення, у якому модуль глобальних ознак забезпечує достатню точність первинного відбору, а локальне уточнення зменшує кількість хибних зіставлень. З огляду на наведені міркування далі розглядаються методи формування глобальних та локальних ознак.

Підрозділ 1.5 Підсумковий аналіз і постановка задач

Ключовою проблемою абсолютної локалізації БПЛА за оптичним каналом є виразна невідповідність між зображеннями аерозйомки та супутниковими знімками, що безпосередньо впливає з різної природи їх формування. Бортова камера реєструє спостереження з висоти, яка змінюється, за довільних кутів орієнтації та в умовах паралаксу, тіней і часткових перекриттів. Натомість супутникові знімки представлені ортофотопланами або сценами з наближено вертикальною проєкцією, для яких характерні фіксована та обмежена просторова

роздільна здатність. Унаслідок цього за зменшення висоти польоту аерознімки містять дрібні деталі й високочастотні структури, відсутні на супутниковому рівні, а навіть для однієї й тієї самої географічної ділянки порівнюються зображення різної деталізації та з різною геометрією відображення. Додатково невідповідність посилюється відмінностями сенсорів і їх спектральних характеристик, змінами освітлення та сезонними ефектами, появою рухомих об'єктів, а також фактичними змінами місцевості між часом супутникової зйомки та виконанням польоту. Ця невідповідність проявляється не лише як зниження точності розпізнавання, а і як ускладнення навчання моделей. Польотні відеодані містять значну частку кадрів, які не забезпечують надійного зіставлення через розмиття від руху, слабку текстурованість, засвічення, однорідні або повторювані фрагменти сцени. Додавання таких кадрів до навчальних і тестових наборів призводить до зміщення оцінок якості та ускладнює побудову коректних пар відповідностей [29, 56-58].

Аналіз сучасних підходів метричного навчання та обробки зображень показав, що перехід від лінійних метрик до глибинного метричного навчання дає змогу знаходити зв'язки у даних, у яких наявна складна, нелінійна структура. Для зображень, зокрема аерознімків та супутникових сцен, глобальні лінійні перетворення є надто обмежувальними й не дають змоги якісно описати геометрію розподілу даних. Натомість у латентному просторі глибинних моделей відстані між векторами можна налаштувати так, щоб вони відображали семантичну подібність об'єктів, а функції втрат задавали потрібну геометрію цього простору [47, 48, 51, 52].

Автоенкодери розглядаються як універсальний засіб побудови представлень із можливістю структуризації латентного простору та аналізу розподілів даних. Вони поєднують задачі реконструкції, виявлення аномалій і дослідження статистичних властивостей вибірки. Згорткові й сіамські нейронні мережі забезпечують формування ознак, безпосередньо узгоджених із метричними функціями втрат, тоді як трансформерні архітектури завдяки механізмам уваги дають змогу моделювати глобальні просторові залежності між фрагментами зображення. Водночас

трансформери потребують значних обсягів навчальних даних або спеціальних стратегій попереднього навчання, а класичні згорткові архітектури залишаються чутливими до доменного зсуву між різними типами зображень [41-46, 50, 82].

У контексті локалізації БПЛА особливо важливими є питання організації та обробки масштабної бази супутникових зображень, що використовується як еталонна. Значні обсяги даних, потреба в регулярному оновленні, вимоги до точності геоприв'язки та до швидкого доступу за обмежених обчислювальних ресурсів бортових систем ускладнюють побудову такої бази й реалізацію ефективного пошуку [7, 59, 60].

Враховуючи викладений аналіз, для подальшого дослідження поставлено такі завдання:

1. Сформулювати математичну постановку задачі абсолютної локалізації безпілотного літального апарата за оптичним каналом та визначити критерії оцінювання якості розпізнавання і локалізації.
2. Розробити нейромережеву модель двоетапного зіставлення аерознімків і супутникових сцен, придатну до роботи за обмежених бортових обчислювальних ресурсів.
3. Розробити метод моделювання, анотування, геоприв'язки та формування навчальних, валідаційних і тестових наборів даних типу «супутниковий знімок – аерознімок» з урахуванням траєкторії польоту, параметрів камери та просторового розділення вибірок.
4. Розробити метод видалення малоінформативних цифрових зображень і зменшення повторюваності кадрів у послідовностях на основі аналізу латентних ознак автоенкодера.
5. Розробити стратегію добору складних негативних прикладів для підвищення роздільної здатності простору ознак у задачі метричного навчання.

6. Розробити метод побудови тривимірного представлення супутникових сцен на основі монокулярної оцінки глибини та забезпечити геометричне узгодження еталонів із реальними ракурсами бортової зйомки.
7. Розробити структуру векторної бази представлень і алгоритми пошуку найближчих сусідів, оптимізовані для роботи за обмежених бортових обчислювальних ресурсів.
8. Інтегрувати розроблені методи у єдину інформаційну технологію та провести підтвердження працездатності запропонованих методів і порівняння з наявними підходами.

РОЗДІЛ 2. ТЕОРЕТИЧНІ ТА ІНФОРМАЦІЙНІ ЗАСАДИ ТЕХНОЛОГІЇ ЛОКАЛІЗАЦІЇ БПЛА ЗА ДАНИМИ АЕРОЗЙОМКИ

У цьому розділі здійснюється перехід від аналітичного огляду сучасних підходів до формалізації власних методів і моделей, що становлять наукову основу дисертаційного дослідження. Метою розділу є побудова теоретичних та інформаційних засад запропонованої інформаційної технології абсолютної локалізації безпілотного літального апарата за оптичним каналом. Для досягнення цієї мети у розділі сформульовано математичну постановку задачі локалізації, запропоновано нейромережеву архітектуру двоетапного зіставлення аерознімків і супутникових сцен, розроблено метод моделювання та побудови навчальних, валідаційних і тестових наборів даних, а також розроблено метод видалення малоінформативних цифрових зображень і зменшення повторюваності кадрів на основі аналізу латентних ознак автоенкодера. Отримані в цьому розділі результати формують науково-методичну основу для подальшої реалізації інформаційної технології в розділі 3 та її кількісного експериментального підтвердження в розділі 4.

Оскільки подальші дослідження спрямовані на побудову інформаційної технології розпізнавання, перейдемо від загальної постановки задач глибинного метричного навчання та візуального розпізнавання місць до формального опису саме процесу локалізації за даними аерозйомки [7, 29, 56-58]. На цьому етапі необхідно сформулювати математичну модель, що пов'язує простір цифрових зображень з простором геопросторових станів БПЛА, а також описати інформаційні структури навчальних і тестових вибірок, на основі яких реалізуються методи глибинного метричного навчання, оцінка розподілів даних на основі автоенкодера та векторний пошук.

Під об'єктом навчального набору розуміється пара (X, s) , де X - цифрове зображення, s - його геопросторова прив'язка. Нехай X_a - простір ЦЗ аерозйомки, отриманих з бортових оптичних систем БПЛА, X_s - простір геоприв'язаних

супутникових зображень місцевості, а Υ - простір векторів стану. Спостереження навігаційної системи можна подати у вигляді множини:

$$D_s = \left\{ (X_s^j, s_j), j = 1, \dots, N_s \right\}, \quad (2.1)$$

де $X_s^j \in \mathbf{X}_s$ - супутникові знімки, а $s_j \in \Upsilon$ його геопросторова інформація. Відповідно задача розпізнавання це пошук відповідного s для нового X_a через зіставлення з базою D_s . Задача абсолютної локалізації формулюється як побудова відображення:

$$F : \mathbf{X}_a \times D_s \rightarrow \Upsilon, \quad (2.2)$$

яке для будь-якого нового бортового зображення $X_a \in \mathbf{X}_a$ відновлює оцінку поточного стану $s \approx F(X_a)$, узгоджену з геоприв'язаною базою супутникових ЦЗ.

У межах розроблюваної інформаційної технології відображення (2.2) розглядається як композиція кількох операторів. По-перше, будується спільний метричний простір представлень Z за допомогою глибинного відображення:

$$f_\theta : \mathbf{X}_a \cup \mathbf{X}_s \rightarrow Z \quad (2.3)$$

де f_θ реалізується нейромережевою моделлю, узгодженою з функціями втрат глибинного метричного навчання. По-друге, формується векторна база представлень супутникових сцен:

$$B = \left\{ (f_\theta(X_s^j), s_j), j = 1, \dots, N_s \right\} \quad (2.4)$$

у якій здійснюється пошук метрично близьких векторів для $f_\theta(X_a)$, отриманих з бортових ЦЗ. Оператор пошуку у B породжує оцінку геопросторового стану БПЛА та реалізує функцію візуального розпізнавання місць в умовах доменного зсуву між $\mathbf{X}_a$ та $\mathbf{X}_s$.

Коректність і стійкість локалізації залежать не лише від вибору архітектури НМ та метричних функцій, а й від інформаційної організації даних. Необхідно

визначити, які саме елементи множин D_a та D_s включаються до навчальних і тестових вибірок, за якими принципами будується розбиття на підмножини, як формалізується інформативність ЦЗ для задачі локалізації та які критерії відбору й фільтрації застосовуються до малоінформативних, повторюваних прикладів. Формування такої інформаційної структури, разом із суворою математичною постановкою задачі локалізації БПЛА за оптичним каналом, становить зміст теоретичних та інформаційних засад технології, що розглядаються у цьому розділі.

Підрозділ 2.1 Математична постановка задачі локалізації БПЛА за оптичним каналом

Надалі X_a та X_s позначають простори аерознімків і супутникових зображень відповідно, а X_a , X_s , X_a^l , X_s^j та X - окремі зображення як елементи цих просторів. Малі літери x, y використовуються лише для позначення координат, компонент та інших скалярних величин. Розглянемо обмежену область місцевості $\Omega \subset \mathbb{R}^2$ задану в деякій фіксованій картографічній проєкції та допустимий діапазон висот польоту $H \subset \mathbb{R}_+$. Стан БПЛА у глобальній системі координат описується вектором:

$$s = (p, h) \in \Upsilon, \quad \Upsilon = \Omega \times H, \quad (2.5)$$

де $p = (p_x, p_y)^T \in \Omega$ - проєкція положення БПЛА на площину карти, $h \in H$ - висота польоту над поверхнею [58, 75].

Орієнтація борта в момент отримання зображення задається одиничним кватерніоном:

$$q = (q_0, q_1, q_2, q_3)^T \in \mathbb{R}^4, \quad \|q\|_2 = 1, \quad (2.6)$$

який визначає матрицю обертання $R(q) \in SO(3)$ між глобальною системою координат та системою, пов'язаною з корпусом БПЛА [83]. Разом з іншими параметрами польоту та камери орієнтація входить до вектора бортових параметрів:

$$\eta = (q, v, w, u, \theta_{cam}) \in E, \quad (2.7)$$

де v - лінійна швидкість, w - кутова швидкість, u - керуючі сигнали, θ_{cam} - параметри камери, E - простір параметрів телеметрії.

Вважаємо, що:

$$\mathbf{X}_a \subset \mathbb{R}^{H_a \times W_a \times C_h}, \quad \mathbf{X}_s \subset \mathbb{R}^{H_s \times W_s \times C_h}, \quad (2.8)$$

де H_a, W_a, H_s, W_s - розміри зображень, C_h - кількість каналів.

Нехай відомою є вибірка отримана в процесі збору даних аерозйомки:

$$D_a = \left\{ (X_a^l, s_l, \eta_l), l = 1, \dots, N_a \right\}, \quad (2.9)$$

де $X_a^l \in \mathbf{X}_a$ - зображення аерозйомки, $s_l = (p_l, h_l) \in \Upsilon$ - стан БПЛА, отриманий бортовими сенсорами, $\eta_l \in E$ - відповідні бортові параметри.

Відповідно до (2.9) та (2.1) D_a та D_s розглядаються як скінченні вибірки реалізацій випадкових векторів (X_a, S, η) та (X_s, S) з відповідними розподілами P_a та P_s :

$$(X_a, S, \eta) \sim P_a, \quad (X_s, S) \sim P_s, \quad (2.10)$$

де $P_a \neq P_s$, що формалізує доменний зсув між зображеннями аерозйомки та супутниковими сценами [58, 75].

Для подальшого використання методів глибинного метричного навчання вводиться простір представлень:

$$Z = \mathbb{R}^d, \quad (2.11)$$

та відображення (2.3), яке реалізується нейромережевою моделлю і переводить як аерознімки, так і супутникові сцени у спільний латентний простір. На Z задається метрика

$$d : Z \times Z \rightarrow \mathbb{R}_+, \quad (2.12)$$

яка відображає невідповідність між представленнями зображень. Бажаною є узгодженість між геометрією простору станів S та геометрією простору представлень Z , якщо стани s_i, s_j близькі, то відстань між їх представленнями має бути малою, і навпаки, для географічно віддалених станів представлення мають бути рознесені.

Алгоритм локалізації за оптичним каналом формалізується як оператор (2.2) який кожному новому бортовому зображенню $X_a \in \mathbf{X}_a$ за фіксованої еталонної бази D_s ставить у відповідність оцінку стану [7, 47, 48]

$$\hat{s} = F(X_a, D_s) = (\hat{p}, \hat{h}) \in \Upsilon. \quad (2.13)$$

Оператор F включає відображення зображень у простір представлень за допомогою f_θ , пошук найближчих сусідів у Z відносно метрики d та перетворення знайдених відповідностей у оцінку стану $\hat{s}$.

Якість локалізації оцінюється метрикою на Υ :

$$\rho((p, h), (\hat{p}, \hat{h})) = \sqrt{\|p - \hat{p}\|_2^2 + \alpha^2 (h - \hat{h})^2}, \quad (2.14)$$

де $\alpha > 0$ - коефіцієнт, що узгоджує внесок похибки за висотою з похибкою по площині. Для навчальної вибірки D_a емпіричний функціонал похибки локалізації має вигляд:

$$J(F) = \frac{1}{N_a} \sum_{l=1}^{N_a} \rho(s_l, F(X_a^l, D_s)), \quad (2.15)$$

а відповідний функціонал ризику

$$R(F) = E_{(X^a, S)} [\rho(S, F(X_a, D_s))]. \quad (2.16)$$

Математична постановка задачі локалізації БПЛА за оптичним каналом у межах даної роботи полягає в побудові оператора

$$F^* \in F, \quad (2.17)$$

де F - допустимий клас алгоритмів локалізації, який мінімізує середню похибку локалізації за введеною метрикою [47, 48]:

$$F^* = \arg \min_F R(F) \approx \arg \min_F J(F). \quad (2.18)$$

Таким чином, задано простір станів БПЛА Υ , простори зображень аерозйомки та супутникових сцен X_a , X_s , структуру навчальних вибірок D_a , D_s , простір представлень Z , метрику похибки ρ та критерій оптимальності, відносно якого в подальших розділах розробляються інформаційні та алгоритмічні засади технології.

Підрозділ 2.2 Запропонована неймережева модель

Запропонована неймережева архітектура забезпечує формування глобальних і локальних ознак для двоетапного зіставлення кадрів аерозйомки із супутниковими еталонами [13, 14, 15]. На відміну від підходів, у яких локальне уточнення виконується для всіх відібраних кандидатів і базується на щільних картах локальних ознак, у цій роботі до архітектури введено два механізми, спрямовані на зниження обчислювальних витрат і потреби в пам'яті під час онлайн обробки. Перший механізм є пороговим модулем оцінювання впевненості за глобальними оцінками подібності, який забезпечує раннє завершення зіставлення в однозначних випадках і не використовує локальне уточнення, що наведено на рисунку 2.1.

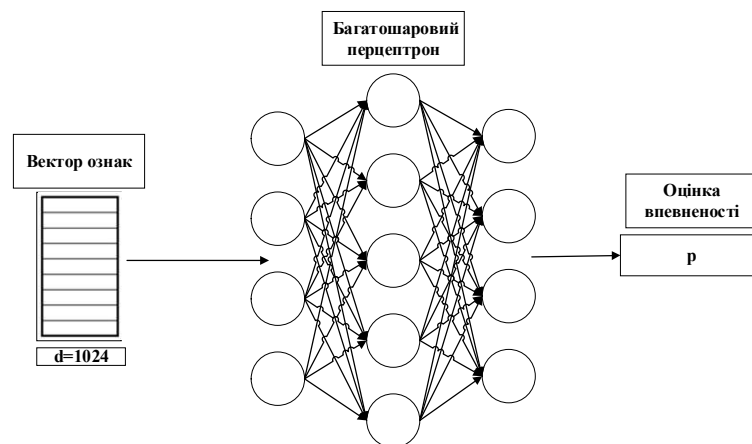


Рис. 2.1. Механізм порогової оцінки впевненості за глобальними ознаками

Другий механізм реалізовано у вигляді модуля відбору опорних дескрипторів, який замінює щільне локальне представлення компактною множиною фіксованої кількості найінформативніших ознак. Це зменшує обчислювальну складність локального зіставлення кандидатів. Механізм відбору опорних дескрипторів наведено на рисунку 2.2.

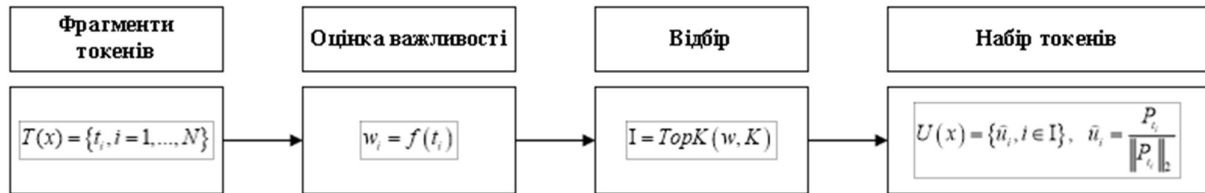


Рис. 2.2. Механізм відбору опорних дескрипторів

Для узгодження ознак моделі із прикладною задачею застосовано параметроефективну адаптацію, за якої параметри базової моделі фіксуються, а додаткові параметри зосереджуються в адаптаційних модулях. Нехай $H^{(l-1)} \in \mathbb{R}^{(N+1) \times D}$ є послідовністю векторів після $(l-1)$ блоку, де N - кількість фрагментів (патчів), D - розмірність ознак. Спочатку обчислюється вихід багатоголової уваги, після чого застосовується послідовний адаптаційний модуль

$$H^{(l)} = H^{(l-1)} + A_1^{(l)} \left(MHA \left(LN \left(H^{(l-1)} \right) \right) \right),$$

де LN - нормування шару, MHA - механізм багатоголової уваги, $A_1^{(l)}$ - адаптаційний модуль. Далі вихід блока формується з урахуванням багатозарового перцептрона та паралельного адаптаційного модуля

$$H^{(l)} = H^{(l)} + MLP \left(LN \left(H^{(l)} \right) \right) + \gamma \cdot A_2^{(l)} \left(LN \left(H^{(l)} \right) \right),$$

де $\gamma > 0$ - коефіцієнт масштабування, A_2 - адаптаційний модуль, доданий паралельно до MLP [15]. Кожен адаптаційний модуль задається як залишкове перетворення зі звуженням і відновленням розмірності

$$A(z) = z + W_{\uparrow} \sigma(W_{\downarrow} z),$$

де $W_{\downarrow} \in \mathbb{R}^{d \times D}$, $W_{\uparrow} \in \mathbb{R}^{D \times d}$, $d < D$, σ - нелінійне перетворення. У реалізації використано відношення $d / D = 0.5$ та $\gamma = 0.2$, що забезпечує компроміс між силою адаптації та стабільністю представлення [15].

Глобальне представлення формується на основі фрагментних векторів останнього блока без використання узагальнювального вектора. Вектори фрагментів відновлюються у просторову карту ознак $f_m(X) \in \mathbb{R}^{h \times w \times D}$, після чого застосовується узагальнене середнє об'єднання GeM та нормування за евклідовою нормою [9, 14]

$$f_{\theta}(X) = L2(GeM(f_m(X))).$$

Пороговий модуль оцінювання впевненості використовує глобальні оцінки подібності для k найкращих кандидатів і вирішує, чи потрібне локальне уточнення. Нехай після глобального пошуку отримано впорядкований список індексів $\{j_{t,1}, \dots, j_{t,k}\}$ та вектор глобальних оцінок $s_t \in \mathbb{R}^k$, елементами якого є, наприклад, значення подібності між $f_{\theta}(X_a^t)$ та $f_{\theta}(X_s^j)$. Тоді скалярна оцінка впевненості визначається як

$$p_t = \text{sigm}(MLP(s_t)),$$

де sigm є сигмоїдним перетворенням, що обмежує p_t на відрізок $[0,1]$. За заданого порога $\tau \in (0,1)$ застосовується правило: якщо $p_t > \tau$, зіставлення завершується на глобальній стадії, а підсумковим вибором вважається найкращий кандидат, якщо $p_t \leq \tau$, то виконується локальне уточнення порядку кандидатів. Ознаки фрагментів, сформовані базовою моделлю на виході трансформера, задані на регулярній сітці $h \times w$, просторова роздільність якої визначається процедурою поділу вхідного зображення на фрагменти та параметрами фрагментації. Після глобально

адаптованої базової моделі введено модуль локальної адаптації на основі двох транспонованих згорток зі збільшенням просторової роздільності

$$f_m^l(X) = \text{intraL2}\left(\text{UpConv}_2\left(\text{ReLU}\left(\text{UpConv}_1\left(f_m(X)\right)\right)\right)\right),$$

де UpConv_1 , UpConv_2 - транспоновані згортки, ReLU є нелінійністю, intraL2 є нормуванням кожного локального вектора у каналному вимірі до одиничної $L2$ -норми [14]. У реалізації використано згортки 3×3 зі кроком 2 та доповненням 1, а кількість каналів після першої і другої операції становить відповідно 256 та 128. У результаті карта ознак $16 \times 16 \times 1024$ перетворюється на щільну сітку локальних ознак $61 \times 61 \times 128$, тобто формується $N = 61 \cdot 61$ локальних векторів розмірності 128 [14].

Модуль відбору опорних дескрипторів застосовується для зменшення обсягу локального подання, яке використовується при локальному зіставленні. Нехай множина локальних векторів, отриманих з $f_m^l(X)$, позначається як $T(X) = \{t_i(X), i = 1, \dots, N\}$, де $t_i(X) \in \mathbb{R}^{128}$. Для кожного локального вектора обчислюється вага важливості

$$w_i(X) = f(t_i(X)),$$

де $f: \mathbb{R}^{128} \rightarrow \mathbb{R}$ є скалярною функцією оцінювання інформативності. Далі формується множина індексів

$$L(X) = \text{TopK}(w(X), K),$$

що містить K індексів із найбільшими значеннями $w(X)$, де $K \ll N^l$. Вихід модуля задається множиною опорних дескрипторів

$$U(X) = \{\tilde{u}_i(X), i \in L(X)\}, \quad \tilde{u}_i(X) = \frac{P_i(X)}{\|P_i(X)\|_2},$$

де $P \in \mathbb{R}^{128 \times 128}$ є лінійним перетворенням малої складності [14].

На етапі використання спочатку виконується пошук у просторі глобальних ознак: для кадру X_a^t та еталонів бази обчислюються $f_\theta(X_a^t)$ і $f_\theta(X_s^j)$, після чого формується множина k найкращих кандидатів за глобальною мірою близькості. Якщо $p_t \leq \tau$, то далі порядок цих k еталонів уточнюється на основі локальних ознак, причому локальне подання кожного зображення задається множиною $U(X)$. Нехай $U(X_a^t) = \{\tilde{u}_r^a, r = 1, \dots, K\}$ та $U(X_s^j) = \{\tilde{u}_q^s, q = 1, \dots, K\}$. Оскільки локальні вектори нормовані, подібність визначається скалярним добутком

$$s_{t,j}(r,q) = (\tilde{u}_r^a)^T \tilde{u}_q^s.$$

Далі формується множина взаємних найближчих відповідників [14]

$$M(t,j) = \left\{ (r,q) : q = \arg \max_v s_{t,j}(r,v), r = \arg \max_\mu s_{t,j}(\mu,q) \right\}.$$

Оцінка локальної узгодженості визначається кількістю взаємних відповідників $S_{loc}(t,j) = |M(t,j)|$, після чого підсумковий кандидат при $p_t \leq \tau$ задається правилом

$$j_t^* = \arg \max_{j \in \{j_{t,1}, \dots, j_{t,k}\}} S_{loc}(t,j)$$

На відміну від підходів, де після отримання локальних відповідностей виконують геометричну перевірку узгодженості, подібність оцінюється кількістю взаємних відповідників $|M(t,j)|$, що зменшує часові витрати уточнення. Узагальнену архітектуру запропонованої моделі, а також взаємодію її основних блоків у двоетапній схемі зіставлення наведено на рис. 2.3.

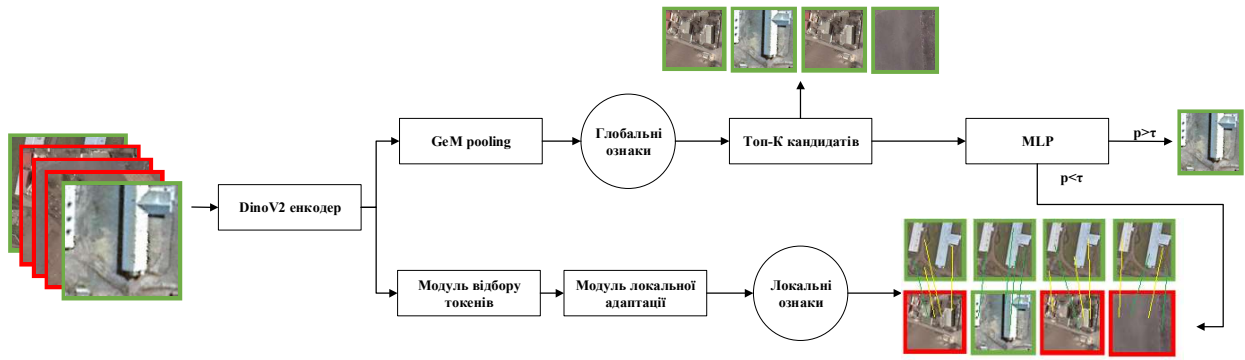


Рис. 2.3. Узагальнена архітектура моделі

Навчання архітектури спрямоване на узгодження одночасно глобальних та локальних ознак. Для глобальних ознак використано трійковий критерій з шарнірною функцією [52]

$$L_g = \sum_j \max \left(\left\| f(X_a) - f(X_s^+) \right\|_2 + m - \left\| f(X_a) - f(X_s^{n_j}) \right\|_2, 0 \right), \quad m > 0.$$

Оскільки оцінка $|M(t, j)|$ є дискретною, локальне навчання спрямоване на збільшення середньої подібності у множині взаємних відповідників для позитивної пари та на зменшення її для негативних пар. Позначимо $M^+ = M(X_a, X_s^+)$ та $M_j^- = M(X_a, X_s^{n_j})$. Тоді локальний критерій може бути записаний як [14, 52]

$$L_l = \sum_j \max \left(-\frac{1}{|M^+| + \varepsilon} \sum_{(r, q) \in M^+} s_{a, +}(r, q) + \frac{1}{|M_j^-|} \sum_{(r, q) \in M_j^-} s_{a, n_j}(r, q), 0 \right).$$

Підсумковий критерій має вигляд

$$L = L_g + \lambda L_l,$$

де λ - ваговий коефіцієнт.

Узгоджене застосування параметрно ошадної адаптації базової моделі, модуля локальної адаптації, порогового модуля оцінювання впевненості та модуля відбору опорних дескрипторів забезпечує практично важливий компроміс між точністю двоетапного зіставлення та витратами часу.

Отже, запропонована архітектура визначає спосіб побудови глобальних і локальних дескрипторів та порядок їх використання у двоетапному зіставленні. Базовий енкодер, на основі якого реалізовано цю схему, і механізм його адаптації до прикладної задачі розглянуто в підрозділі 2.3

Підрозділ 2.3 Базовий енкодер та його параметроефективна адаптація

У межах запропонованої архітектури формування ознак реалізовано на основі попередньо навченої моделі DINOv2 [13, 14], що є візуальним трансформером та використовується як базовий енкодер для подальшого отримання глобальних і локальних представлень. Суттєвою перевагою даної архітектури є те, що модель формує інформативні локальні ознаки, пов'язані з окремими ділянками зображення. Такі локальні вектори дозволяють зіставляти структурно подібні елементи сцени навіть у випадках, коли загальний вигляд кадрів відрізняється через зміну масштабу, ракурсу, роздільної здатності або фотометричних умов. На відміну від підходів, де локальні дескриптори потрібно конструювати окремо за допомогою спеціалізованих процедур [16], у DINOv2 локальні представлення отримуються безпосередньо з виходів енкодера та можуть бути використані для локального порівняння між зображеннями.

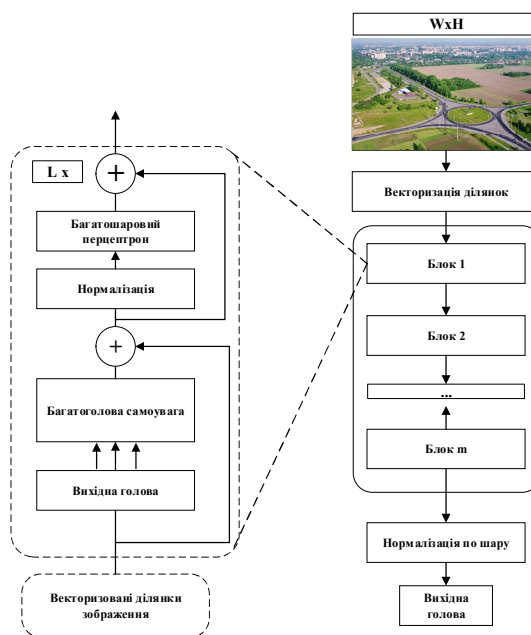


Рис. 2.4. Узагальнена архітектура DINOv2

Це дає змогу впорядковувати відібраних кандидатів швидше, ніж підходи, що потребують геометричної перевірки просторової узгодженості локальних ознак. [14]. У роботі використано DINOv2 на архітектурі ViT L/14 [13, 41], яка при розмірі вхідного зображення 224×224 формує сітку фрагментів 16×16 і забезпечує розмірність вихідних ознак розміру 384 (див. рис. 2.5).

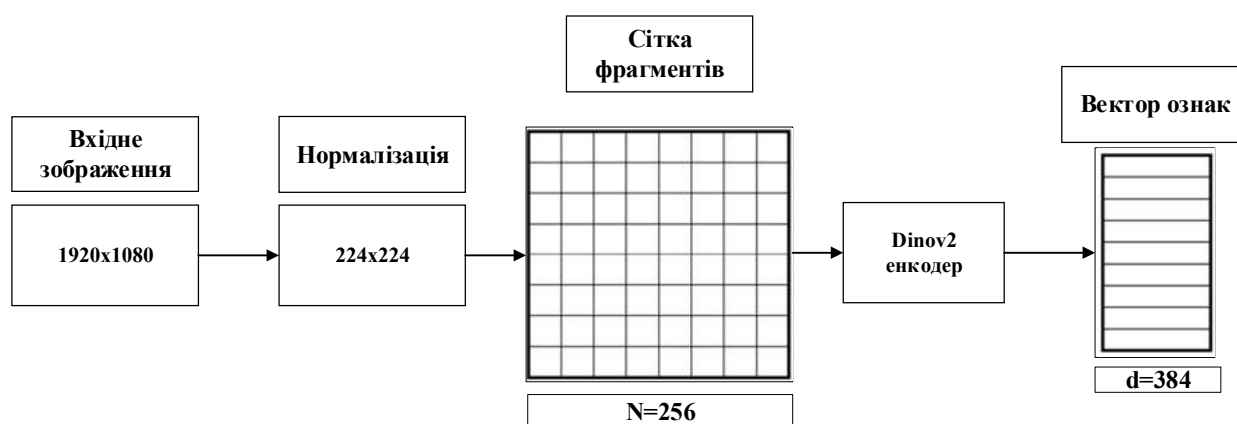


Рис. 2.5. Схема перетворення вхідного зображення на глобальний вектор ознак

Обрана конфігурація забезпечує формування глобального дескриптора сталої розмірності, придатного для ефективного індексування у векторній базі даних, і водночас зберігає достатню просторову деталізацію локальних подань, необхідну для подальшого отримання щільних локальних ознак у модулі локальної адаптації. Така узгодженість глобального та локального рівнів подання є принциповою для стабільного уточнення відповідності на множині кандидатів.

Повне донавчання енкодера великого масштабу потребує значних обчислювальних ресурсів і підвищує ризик погіршення узагальнювальних властивостей, сформованих на етапі попереднього навчання. З цієї причини у роботі застосовано параметроефективну адаптацію, за якої параметри базової моделі фіксуються, а навчаються лише компактні додаткові модулі. Такий підхід зменшує потребу в ресурсах для навчання, полегшує відтворення експериментів і забезпечує вищу стабільність репрезентацій під час застосування моделі до даних, отриманих на інших ділянках та за інших умов спостереження [15].

Отже, DINOv2 обрано базовим енкодером, оскільки він забезпечує стійкі представлення в умовах доменного зсуву, формує локальні дескриптори, придатні до зіставлення структурних елементів сцени, та узгоджується з подальшими етапами побудови глобальних і щільних локальних ознак у запропонованій архітектурі.

Підрозділ 2.4 Метод побудови навчальних, валідаційних і тестових наборів даних

У цьому розділі розглянуто створення навчальних, валідаційних і тестових вибірок на основі експериментальних польотів БПЛА та супутникових даних, а також формалізацію відповідних множин D_a та D_s . Описано джерела та характеристики даних аерозйомки, отриманих з бортової RGB-камери БПЛА, принципи побудови еталонної бази супутникових сцен та узгодження пар, що надалі використовуються для глибокого метричного навчання в задачі розпізнавання місць за даними аерозйомки.

Загальні принципи побудови авторських наборів аерозйомки та їх початкові описові версії наведено в попередніх публікаціях про датасет «AERIAL SURVEY» [37, 38].

Підрозділ 2.4.1 Дані аерозйомки та їх характеристики

Дані аерозйомки в роботі було зібрано за допомогою квадрокоптера Autel EVO II V2 Dual Rugged Bundle (640T) [84], зовнішній вигляд якого подано на рис. 2.6.



Рис. 2.6. Зовнішній вигляд Autel EVO II V2 Dual Rugged Bundle (640T)

Дані аерозйомки були отримані за допомогою вбудованої RGB-камери з матрицею типу 1/2" CMOS, ефективною роздільною здатністю 48 Мпкс. Для побудови вибірки D_a використовувався відеорежим Full HD із роздільною здатністю 1920×1080 пікселів, трьома каналами (RGB) та частотою зйомки 30 кадрів за секунду. Таким чином, у термінах (2.8) для аерознімків маємо:

$$H_a = 1080, W_a = 1920, C_h = 3, \quad (2.19)$$

тобто $\mathbf{X}_a \subset \mathbb{R}^{1080 \times 1920 \times 3}$.

Польоти виконувалися над територією України на дискретних висотах

$$h \in H_{\text{exp}} = \{100, 150, 200, 250\} \text{ м}, \quad (2.20)$$

що є підмножиною загального діапазону висот H , введеного у (2.5). Для кожної висоти фіксувалися середня швидкість руху, що становила 7–10 м/с, залежно від погодних умов. Згідно з паспортними характеристиками камери, горизонтальний кут поля зору становить приблизно $\theta \approx 79^\circ$. Геометрія зйомки організовувалася таким чином, щоб оптична вісь камери була орієнтована близько до вертикального напрямку. Така схема зйомки дає змогу розглядати кадри як наближену ортогональну проєкцію локальної ділянки на площину Ω , що узгоджується з двовимірною ортогональною проєкцією супутникових карт. Відхилення оптичної осі від вертикалі, а також оклюзії, зумовлені рельєфом і об'єктами на місцевості, у подальшому враховуються як складова доменного зсуву між аерознімками та супутниковими сценами, формалізованого у (2.10).

Паралельно з відеопотоком реєструвалися телеметричні дані польотного контролера, часові мітки яких попередньо узгоджені з часовими мітками відеокадрів. Телеметричний запис містить географічні координати центра БПЛА, висоту, орієнтацію борта, кути гіростабілізованого підвісу камери, часову мітку та службову інформацію. Завдяки наявній часовій прив'язці додаткова синхронізація

відео й телеметрії під час обробки не виконувалася. Відповідно, формувалися трійки (X_a^l, s_l, η_l) , що відповідають елементам множини D_a .

Для висоти польоту h ширина ділянки місцевості, що потрапляє в кадр у проєкції на площину Ω , оцінюється формулою:

$$L(h) = 2h \tan \frac{\theta}{2}. \quad (2.21)$$

Для відповідних висот H_{exp} маємо:

$$L(100) \approx 165 \text{ м}, L(150) \approx 247 \text{ м}, L(200) \approx 330 \text{ м}, L(250) \approx 412 \text{ м}, \quad (2.22)$$

що відповідає просторовому розрізненню на рівні приблизно 0,09–0,21 метрів на піксель при ширині кадру 1920 пікселів. Така деталізація є достатньою для подальшого зіставлення з супутниковими сценами та реалізації абсолютної локалізації. Відповідно до параметрів камери та швидкості руху це відповідає зміщенню між сусідніми кадрами приблизно

$$0,23 \text{ м} \leq \Delta s_{30} \leq 0,33 \text{ м}. \quad (2.23)$$

Оскільки телеметричні дані реєструвалися з частотою 10 Гц, для підмножини кадрів з унікальною прив'язкою до телеметрії крок уздовж траєкторії становить

$$0,7 \text{ м} \leq \Delta s_{10} \leq 1,0 \text{ м}, \quad (2.24)$$

що забезпечує значний перекрив зображень. Узагальнені параметри виконаних польотів наведено в табл. 2.1, а загальна таблиця наведена у Додатку В.

Таблиця 2.1. Узагальнена інформація щодо виконаних польотів

Номер польоту	Дата польоту	Погодні умови	Знята відстань
1	21.02.2024	Хмарно	9,3 км
2	23.02.2024	Мінлива хмарність	4 км
3	27.02.2024	Мінлива хмарність	13,2 км
4	28.02.2024	Мінлива хмарність	17,4 км

Продовження таблиці 2.1

Номер польоту	Дата польоту	Погодні умови	Знята відстань
5	01.03.2024	Хмарно	7,1 км
...			
61	13.06.2024	Сонячно	1,8 км
62	21.06.2024	Сонячно	35,2 км
Загальна кількість польотів = 62		Загальна відстань = 546 км	

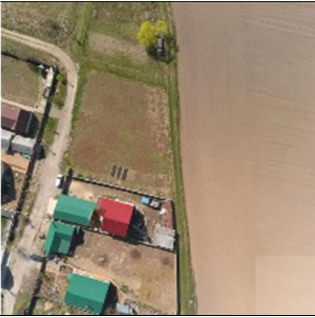
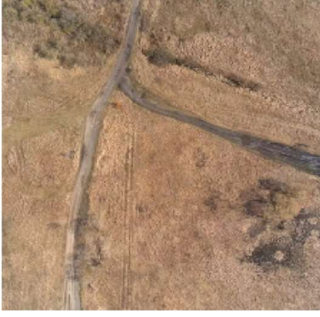
Підсумовуючи наведені дані у таблиці 2.1, сумарна протяжність маршрутів, за якими було зібрано дані, становить 546 км, що ілюструється на графіку обсягу відзнятого матеріалу на рис. 2.7.



Рис. 2.7. Загальний об'єм польотів для збору даних

Початковий масив аерознімків містив усі відеокадри, експортовані з відібраних відрізків польотів для подальшої обробки. З урахуванням тривалості зйомки та кроку дискретизації траєкторій обсяг сформованої вибірки становить $N_a^{raw} = 1\,927\,059$ кадрів. Водночас дані телеметрії реєструвалися з частотою 10 Гц, тому кількість кадрів, що мають унікальну прив'язку до телеметрії становить $N_a^{tel} = 642\,353$. У таблиці 2.2 наведено приклад зображень, що були зняті у межах побудови вибірки.

Таблиця 2.2. Приклади зображень аерозйомки

№	Аерознімок	Роздільна здатність	Висота зйомки
1.		1920x1080	150 метрів
2.		1920x1080	100 метрів
3.		1920x1080	200 метрів
4.		1920x1080	150 метрів

Підрозділ 2.4.2 Дані супутникової зйомки та їх характеристики

Еталонні растрові зображення місцевості X_s^j , що входять до множини D_s , введеної у (2.1), сформовано шляхом завантаження супутникових фрагментів із сервісу Google Maps у картографічній проєкції Web Mercator (EPSG:3857) [78, 79]. Фрагменти зображень, що надаються сервісом, мають фіксований розмір:

$$T \times T, T = 256, \quad (2.25)$$

організованих за рівнями масштабування $z \in \mathbb{N}$. Для кожного рівня z поверхня Землі в обраній картографічній проєкції розбивається на регулярну сітку фрагментів, ідентифікованих індексами (z, i, j) . Орієнтація такого зображення є сталою, верхня сторона кожного фрагменту відповідає напрямку на північ.

Нехай R_E - радіус сферичної моделі Землі, $R_E \approx 6,378 \cdot 10^6$ м. Довжина екватора в цій моделі дорівнює:

$$C_E = 2\pi R_E. \quad (2.26)$$

На рівні масштабування z кількість фрагментів вздовж екватора становить 2^z , отже один фрагмент покриває відрізок довжиною:

$$S_{tile}(z) = \frac{C_E}{2^z}. \quad (2.27)$$

Тоді довжина ділянки місцевості, що припадає на один піксель дорівнює:

$$s_{eq}(z) = \frac{S_{tile}(z)}{T} = \frac{C_E}{T \cdot 2^z}. \quad (2.28)$$

У проєкції типу Web Mercator розрізнення на широті φ зменшується пропорційно до $\cos\varphi$:

$$S_{pix}(z, \varphi) = s_{eq}(z) \cos \varphi. \quad (2.29)$$

Для характерних широт України $\varphi \in [48^\circ; 50^\circ]$ та рівнів масштабування, що використовувалися в роботі, отримуємо оцінки [78, 79]

$$S_{pix}(19, \varphi) \approx 0,20 \text{ м/пікс}, S_{pix}(20, \varphi) \approx 0,10 \text{ м/пікс}. \quad (2.30)$$

Ширина та висота ділянки місцевості, що потрапляє в кадр у проєкції на площину Ω для висоти польоту h_l , використовуючи (2.21), дорівнює:

$$L_x(h_l) = 2h_l \tan \frac{\theta}{2}, \quad L_y(h_l) = L_x(h_l) \frac{H_a}{W_a}. \quad (2.31)$$

У локальній системі координат, пов'язаній із центром кадру, проєкція поля зору має половину довжин сторін

$$a_x(h_l) = \frac{L_x(h_l)}{2}, \quad a_y(h_l) = \frac{L_y(h_l)}{2}. \quad (2.32)$$

Для побудови супутникового фрагмента, узгодженого з кадром аерозйомки, розглянемо область покриття в площині Ω , заданій у декартових координатах. Нехай $p_l = (p_{l,x}, p_{l,y})^T$ - координати центру кадру. Тоді прямокутна область місцевості, яку необхідно покрити супутниковими фрагментами, визначається як:

$$\Omega_l^{(0)} = \left\{ p \in \Omega : |p_x - p_{l,x}| \leq a_x(h_l), \quad |p_y - p_{l,y}| \leq a_y(h_l) \right\}. \quad (2.33)$$

Для широти, що відповідає точці p_l , обчислюється географічна широта φ_l за оберненим перетворенням:

$$\varphi_l = 2 \arctan \left(\exp \left(\frac{p_{l,y}}{R_E} \right) \right) - \frac{\pi}{2}. \quad (2.34)$$

Отримане значення φ_l використовується для визначення просторового розрізнення супутникового шару $S_{pix}(z, \varphi_l)$, необхідного для узгодження масштабу супутникового фрагмента з масштабом аерознімка. Просторове розрізнення аерознімка на висоті h_l оцінюється як відношення ширини покриття місцевості до ширини кадру:

$$S_a(h_l) = \frac{L_x(h_l)}{W_a}. \quad (2.35)$$

Після вибору рівня масштабування z_l визначаються індекси фрагментів, що перекривають область $\Omega_l^{(0)}$. Координати кутів прямокутника в площині Ω задаються як:

$$p_l^- = (p_{l,x} - a_x(h_l), p_{l,y} - a_y(h_l))^T, \quad p_l^+ = (p_{l,x} + a_x(h_l), p_{l,y} + a_y(h_l))^T. \quad (2.36)$$

Відповідні діапазони індексів (z_l, i, j) обчислюються за формулами:

$$i_{\min} = \left\lfloor \frac{p_{l,x}^- + \pi R_E}{S_{tile}(z_l)} \right\rfloor, \quad i_{\max} = \left\lfloor \frac{p_{l,x}^+ + \pi R_E}{S_{tile}(z_l)} \right\rfloor, \quad (2.37)$$

$$j_{\min} = \left\lfloor \frac{\pi R_E - p_{l,y}^+}{S_{tile}(z_l)} \right\rfloor, \quad j_{\max} = \left\lfloor \frac{\pi R_E - p_{l,y}^-}{S_{tile}(z_l)} \right\rfloor, \quad (2.38)$$

Для всіх $(i, j) \in [i_{\min}, i_{\max}] \times [j_{\min}, j_{\max}]$ завантажуються відповідні растрові фрагменти супутникового шару

$$X_s^{(z_l, i, j)} \in \mathbb{R}^{T \times T \times C}. \quad (2.39)$$

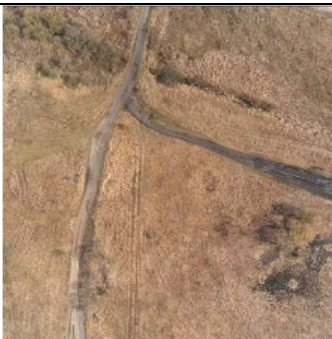
Завантажені фрагменти об'єднуються за їхніми індексами i та j у мозаїку $\tilde{X}_s^l$, після чого з мозаїки виконується обрізання до області $\Omega_l^{(0)}$. Це забезпечує покриття ділянки місцевості, відповідної полю зору кадру аерозйомки на висоті h_l .

Для узгодження орієнтації супутникового фрагмента з кадром аерозйомки використовується азимут ψ_l отриманий з телеметрії. Оскільки зйомка ведеться в наближеній вертикальній проєкції на площину та камера стабілізована, узгодження орієнтації виконується поворотом супутникового фрагмента в площині:

$$X_s^l = R_{\psi_l} \left(C_t \left(\tilde{X}_s^l; \Omega_l^{(0)} \right) \right), \quad (2.40)$$

де C_t - оператор обрізання до прямокутної області, а R_{ψ_l} - оператор повороту зображення на кут ψ_l навколо його центру. У результаті для кожного кадру X_a^l формується узгоджений супутниковий фрагмент X_s^l . Приклади сформованих растрових фрагментів місцевості X_s^l узгоджених із кадрами аерозйомки, наведено в таблиці 2.3.

Таблиця 2.3. Приклади растрових фрагментів місцевості

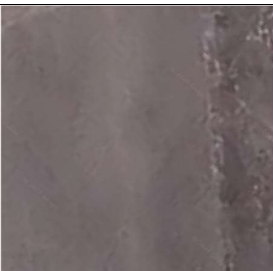
№	Супутниковий знімок	Відповідний аерознімок
1.		
2.		
3.		
4.		

Підрозділ 2.4.3 Формалізація навчальних наборів

Після формування узгоджених супутникових фрагментів для кадрів аерозйомки та виконання первинного очищення пар отримано множину коректних відповідностей між зображеннями. Первинне очищення включає вилучення

прикладів із некоректною відповідністю між X_a^l і X_s^l та усунення надмірно повторних спостережень. Приклади пар, вилучених під час фільтрації подано у таблиці 2.4.

Таблиця 2.4. Приклади вилучених пар зображень

№	Аерознімок	Супутниковий знімок	Причина вилучення
1.			Невідповідність у масштабі та повороті між зображеннями.
2.			Невідповідність у масштабі між зображеннями.
3.			Мала кількість ознак для відрізнєння.
4.			Невідповідність у масштабі між зображеннями.

Після фільтрації сформовано набір даних:

$$\tilde{D}_{as} = \left\{ (X_a^l, X_s^l, s_l), l = 1, \dots, N \right\}, N = 413\,000. \quad (2.41)$$

Отриманий набір даних $\tilde{D}_{as}$ поділено на дві частини відповідно до просторової належності даних:

$$\tilde{D}_{as} = \tilde{D}_{as}^{gen} \cup \tilde{D}_{as}^{te}, \quad \tilde{D}_{as}^{gen} \cap \tilde{D}_{as}^{te} = \emptyset, \quad (2.42)$$

де $\tilde{D}_{as}^{gen}$ - генеральний набір даних, що використовується для тренування нейромереж, $\tilde{D}_{as}^{te}$ - дані отримані з інших локацій, що не використовуються під час навчання та валідації. Тестова зйомка виконувалася без зміни апаратної конфігурації, що забезпечує зіставність умов отримання зображень.

Оскільки елементи $\tilde{D}_{as}^{gen}$ сформовано з відеопотоку аерозйомки зі значним перекривом суміжних кадрів, покадрове випадкове розбиття може спричинити потрапляння корельованих спостережень одночасно до тренувальної й валідаційної підвбірок. Це спричиняє залежність між тренувальною та валідаційною підвбірками через наявність корельованих спостережень у обох частинах розбиття, що веде до систематично завищеної оцінки узагальнювальної здатності моделі. З огляду на це, розбиття $\tilde{D}_{as}^{gen}$ на $\tilde{D}_{as}^{tr}$ та $\tilde{D}_{as}^{val}$ здійснювали за груповим принципом, за яким усі кадри, отримані під час одного польоту, відносили виключно до однієї з підвбірок. Групи відбирались випадково до досягнення співвідношення 70% та 30% між $\tilde{D}_{as}^{tr}$ та $\tilde{D}_{as}^{val}$:

$$\tilde{D}_{as}^{gen} = \tilde{D}_{as}^{tr} \cup \tilde{D}_{as}^{val}, \quad \tilde{D}_{as}^{tr} \cap \tilde{D}_{as}^{val} = \emptyset. \quad (2.43)$$

Оскільки під час побудови $\tilde{D}_{as}$ кожному зображенню аерозйомки X_a^l поставлено у відповідність один узгоджений супутниковий фрагмент X_s^l . За такої структури даних застосування проксі орієнтованих методів метричного навчання є обмежено доцільним, оскільки вони вимагають введення дискретних міток класів $c = g(s)$ та навчання проксі-векторів для кожного класу [53-55]. У цьому випадку при дрібній дискретизації простору Ω виникає режим із малою кількістю прикладів на, що ускладнює стабільне оцінювання проксі-векторів. З огляду на це, для

формування тренувальної, валідаційної та тестової вибірок використовується подання у вигляді трійок:

$$\tilde{D}_v^{(3)} = \left\{ (X_a^l, X_s^l, X_s^{n(l)}) : (X_a^l, X_s^l, s_l) \in \tilde{D}_{as}^v \right\}, \quad v \in \{train, val, test\}, \quad (2.44)$$

де $X_s^{n(l)}$ - негативний супутниковий фрагмент, що відповідає іншій ділянці місцевості. Негативний індекс $n(l)$ обирається з тієї ж підвибірки $\tilde{D}_{as}$ за умовою просторової віддаленості:

$$n(l) \neq l, \quad \|p_l - p_{n(l)}\|_2 \geq r_{neg}, \quad r_{neg} > 0, \quad (2.45)$$

де $p_l \in \Omega$ - планові координати стану s_l , а r_{neg} - поріг, що виключає вибір негативів із тієї локальної області, що й позитивний відповідник [52]. Відповідно до величини порога негативні приклади умовно поділяють на три категорії: r_3 легкі, r_2 середні та r_1 важкі, які визначаються радіусом їх віддаленості від позитивного прикладу. Де r_1, r_2, r_3 обираються на основі параметру покриття кадру $L(h)$. Приклад зображено на рисунку 2.8.

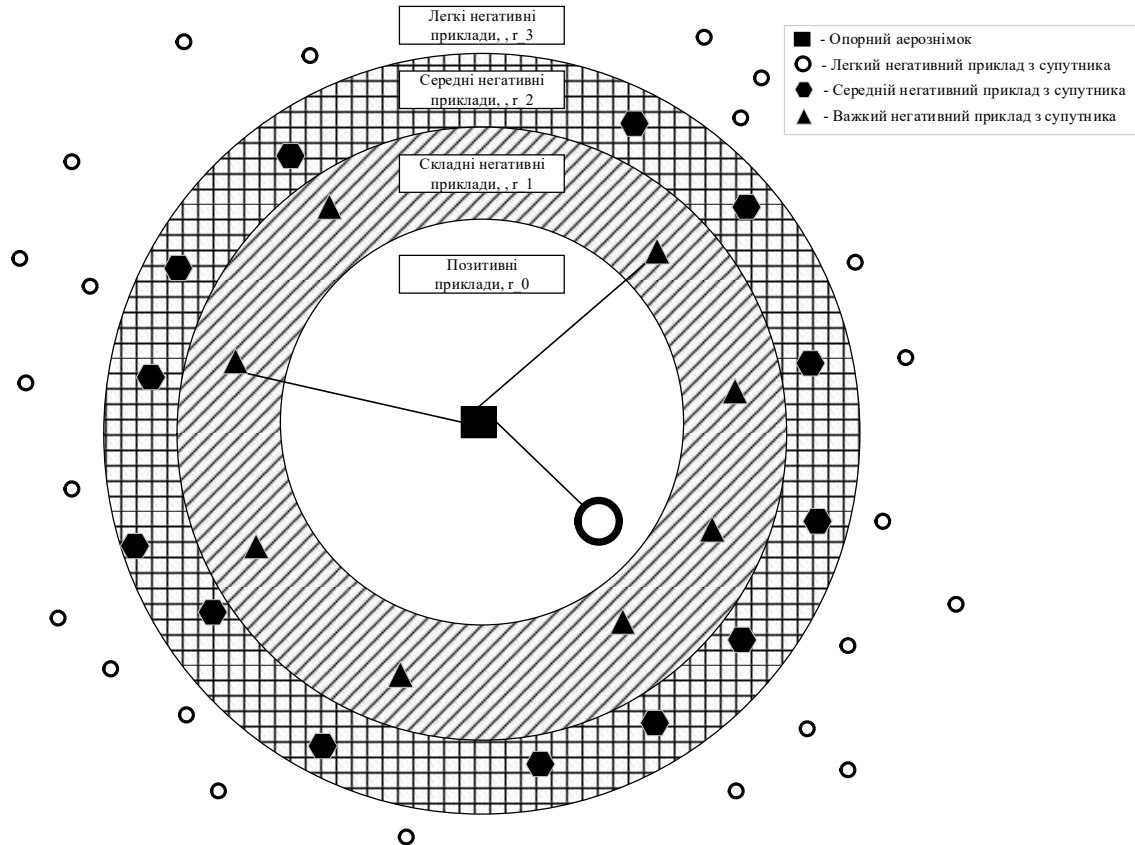


Рис. 2.8. Приклад відбору негативних пар зображень

Умова (2.45) задає множину допустимих індексів негативних прикладів. Водночас для підвищення ефективності метричного навчання негativi доцільно добирати за інформативністю, тобто обирати такі, що є просторово віддаленими, але водночас найбільш близькими до опорного зображення у просторі ознак за візуальним змістом. З цією метою негативний індекс $n(l)$ визначають як елемент допустимої множини, для якого відстань у просторі ознак між опорним зображенням та супутниковим фрагментом є мінімальною [14, 52]:

$$n(l) = \arg \operatorname{sort}_{j \in N_K(l)} d(\varphi(X_a^l), \varphi(X_s^j)), \quad (2.46)$$

$$N_K(l) = \left\{ j : (X_a^j, X_s^j, s_j) \in \tilde{D}_{as}^u, j \neq l, \|p_l - p_j\| \geq r_{neg} \right\}. \quad (2.47)$$

Для фіксованого опорного прикладу з індексом l відбирають K негативних індексів $\{n_k(l), k=1, \dots, K\}$. За фіксованих X_a^l і X_s^l кожен індекс $n_k(l)$ визначає окрему навчальну трійку $(X_a^l, X_s^l, X_s^{n_k(l)})$, тому для одного й того самого l формують K трійок. Повний перебір кандидатів під час знаходження мінімізатора в (2.46) на множині $\tilde{D}_{as}$ є обчислювально витратним. Тому відбір виконують у просторі попередньо обчислених векторних представлень $f_\theta(X)$, які зберігають у сховищі та періодично актуалізують у процесі навчання. Для кожного l формують скінченну множину кандидатів шляхом випадкового відбору, після чого з неї вилучають просторово близькі фрагменти відповідно до обмежень на допустимі негativi, заданих множиною $N_K(l)$. Далі індекси $n_k(l)$ визначають за критерієм (2.46). Приклади сформованих трійок наведено в табл. 2.5.

Таблиця 2.5. Приклади сформованих трійок

№	Опорний аерознімок	Позитивний супутниковий знімок	Негативний супутниковий знімок
1.			
2.			
3.			

Підрозділ 2.5 Метод видалення малоінформативних цифрових зображень і зменшення повторюваності кадрів

Під інформативністю цифрового зображення розуміється ступінь наявності в ньому стійких структурних ознак, які забезпечують відокремленість спостережень і зменшують їх неоднозначність під час зіставлення даних аерозйомки із супутниковими зображеннями. Для даних аерозйомки типовими є: малоінформативні кадри з низькою контрастністю та однорідністю; надмірно повторні кадри, що виникають через високу частоту зйомки та значний перекрив суміжних зображень. З огляду на це вводиться процедура автоматизованої фільтрації.

Для побудови простору ознак використано попередньо навчений автоенкодер на наборі даних «Аерозйомка» [37]. Архітектуру автоенкодера адаптовано [40] до розміру вхідного зображення $224 \times 224 \times 3$ наведено на рисунку 2.9 та 2.10. Порівняльне оцінювання Vision Transformer і ResNet-50 як автокодувальників для цієї задачі наведено в окремому дослідженні [36]. У подальшому використовується лише перша частина моделі, без використання шару декодера.

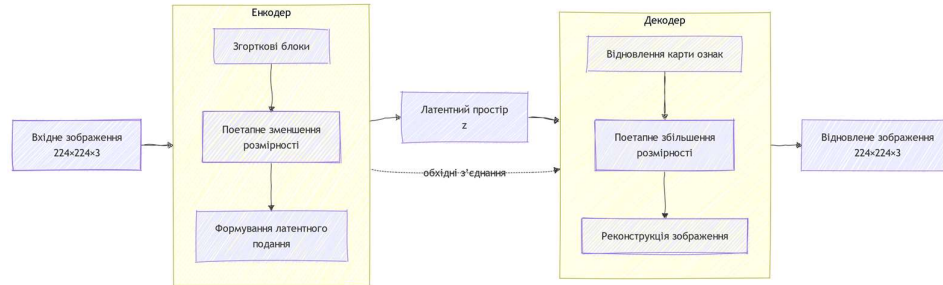


Рис. 2.9. Архітектура адаптованого автоенкодера

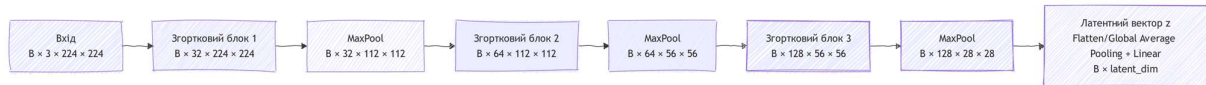


Рис. 2.10. Схема енкодера, архітектури автоенкодера

Для кожного зображення X_a обчислюється вектор прихованих ознак фіксованої розмірності $d_{AE} = 256$

$$z(X) = G(X) \in \mathbb{R}^{d_{AE}}, \quad (2.48)$$

де G - енкодер ознак. Вектори ознак нормуються за евклідовою нормою:

$$\tilde{z}(X) = \frac{z(X)}{(\|z(X)\|_2 + \varepsilon_0)}, \quad \varepsilon_0 > 0, \quad (2.49)$$

що забезпечує стабільність порівняння та усуває вплив масштабних варіацій векторів ознак, де ε_0 - запобігає діленню на нуль [40, 43, 46].

Для множини зображень $\{X_a^i, i=1, \dots, N\}$ застосовано метод головних компонент [71-73] у просторі $\tilde{z}$ зі зменшенням розмірності до $m = 32$ компонент.

Нехай μ - вибіркове середнє, а $U_m \in \mathbb{R}^{d_{AE} \times m}$ - матриця перших m компонент. Тоді проєкція в простір головних компонент визначається як

$$y(X) = U_m^T (\tilde{z}(X) - \mu) \in \mathbb{R}^m, \quad (2.50)$$

а реконструкція у просторі ознак задається виразом [71-73]

$$z_{rec}(X) = U_m y(X) + \mu. \quad (2.51)$$

Показник інформативності визначає нормовану залишкову похибку апроксимації $\tilde{z}(X)$ підпростором перших m головних компонент у латентному просторі автоенкодера та виражається як:

$$I(X) = \frac{1}{d_{AE}} \|\tilde{z}(X) - z_{rec}(X)\|_2^2. \quad (2.52)$$

Малі значення $I(X)$ відповідають зображенням, латентні ознаки яких добре апроксимуються підпростором перших m головних компонент, тобто мають малу залишкову складову поза цим підпростором. Для даних аерозйомки до цієї групи належать також малоінформативні зображення з обмеженою кількістю стійких ознак. Вилучення таких прикладів здійснюється за квантильним правилом:

$$X \text{ залишається у вибірці} \Leftrightarrow I(X) \geq Q_{0.30}(I), \quad (2.53)$$

тобто вилучаються 30% прикладів із найменшими значеннями $I(X)$ [71-73, 34, 35].

Усунення надмірного повторювання кадрів виконується локально в межах кожної короткої послідовності зображень, отриманої під час безперервної зйомки на одному фрагменті маршруту. Для впорядкованої послідовності $D_a = \{X_a^t, t=1, \dots, N\}$ вводиться міра міжкадрової зміни у просторі головних компонент як косинусна відстань між представленнями $y(X_a^t)$ та $y(X_a^{t-1})$:

$$\delta_t = 1 - \frac{y(X_a^t)^T y(X_a^{t-1})}{\|y(X_a^t)\|_2 \|y(X_a^{t-1})\|_2 + \varepsilon_0}, \quad t=2, \dots, T. \quad (2.54)$$

Правило вилучення повторюваних зображень задається квантильним порогом у межах послідовності:

$$X_a^t \text{ залишається у вибірці} \Leftrightarrow \delta_t > Q_{0.10}(\delta | D_a), \quad (2.55)$$

де $Q_{0.10}(\delta | D_a)$ - емпіричний 10% квантиль множини $\{\delta_t, t = 2, \dots, T\}$ обчислений для даної послідовності.

Для підтвердження ефективності запропонованих методів підвищення інформативності навчальної вибірки сформувано гіпотези:

1. Гіпотеза Н1. Фільтрація за $I(X)$ вилучає підмножину малоінформативних зображень.

Підтвердження. За результатами фільтрації вилучено 123 900 кадрів за критерієм $I(X)$, що становить 30% від N . Розподіл $I(X)$ і поріг наведено на рис. 2.11. Візуальні приклади вилучених зображень з мінімальною $I(X)$ на рис. 2.12 (а) демонструють наявність малокоонтрастних спостережень, тоді як приклади з великими значеннями $I(X)$ містять виразні структурні елементи сцени, що представлено на рисунку 2.12. (б). Це узгоджується з порогом $Q_{0.30}(I)$ та підтверджує керованість процедури. Таким чином, гіпотеза Н1 підтверджується.

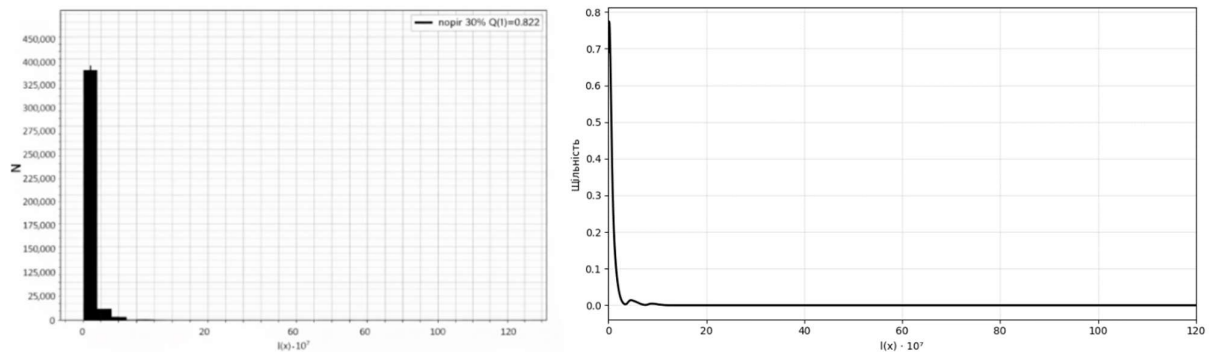


Рис. 2.11. Ліворуч - розподіл похибки відновлення. Праворуч – щільність розподілу

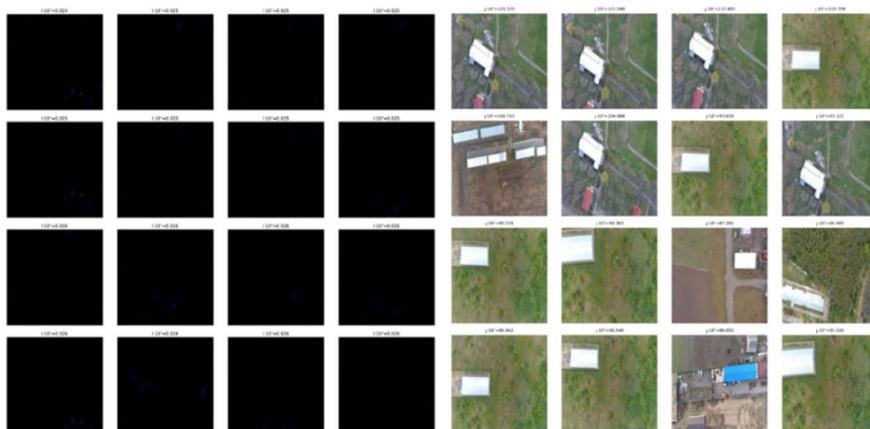


Рис. 2.12. (а) Приклади вилучених (б) Приклад виразних структурних зображень

2. Гіпотеза Н2. Фільтрація за δ_i зменшує повторюваність кадрів у межах послідовностей.

Підтвердження. Вилучено 40 928 кадрів за критерієм δ_i , що становить 9,91%. Відхилення від 10% пояснюється тим, що перший елемент кожної послідовності не має визначеної δ_i та не може бути вилучений за критерієм (2.55). Розподіл δ_i наведено на рис. 2.13 і має виражену кількість прикладів біля нуля, що відповідає значній кількості майже однакових сусідніх зображень. На рис. 2.14 наведено часовий ряд δ_i для найдовшої послідовності з порогом $Q_{0.10}(\delta|D_a)$ та позначенням вилучених зображень, які концентруються в локальних зонах мінімальних значень δ_i . Таким чином, гіпотеза Н2 підтверджується.

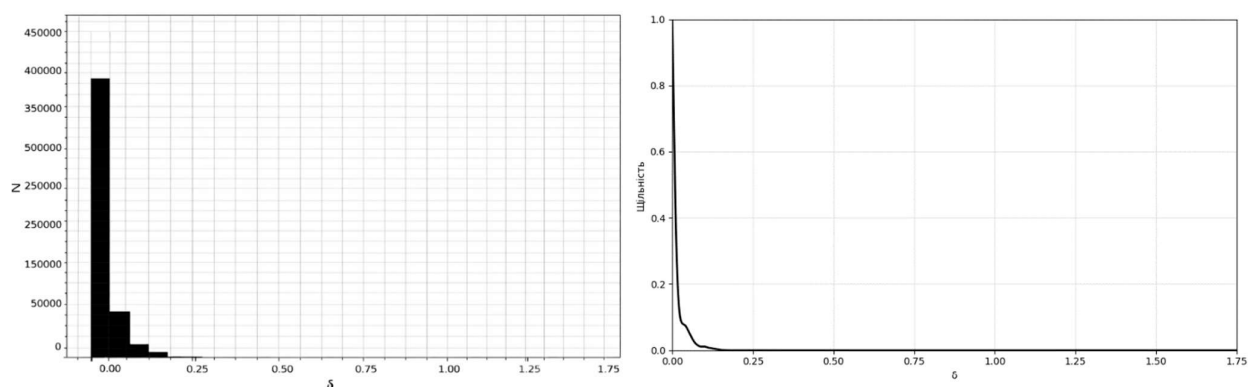


Рис. 2.13. Ліворуч - розподіл відстані δ_i між сусідніми кадрами. Праворуч – щільність розподілу

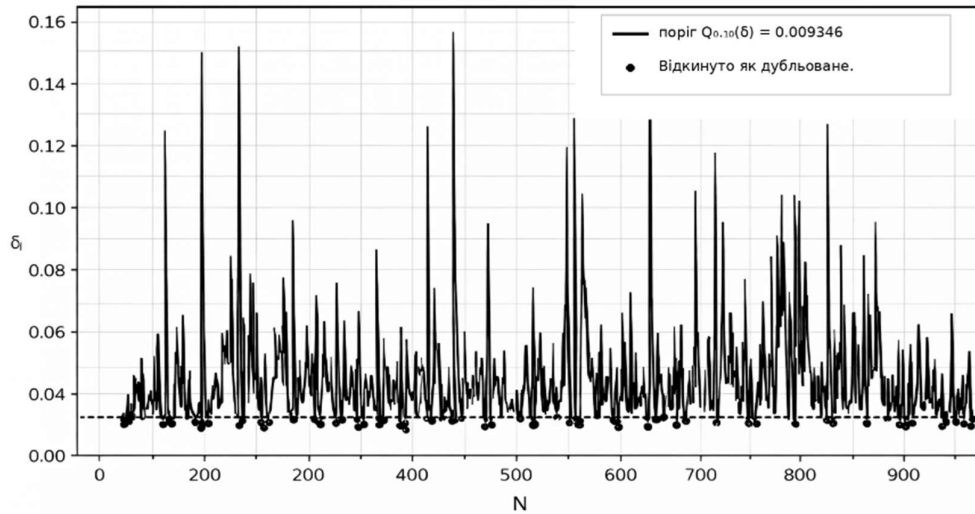


Рис. 2.14. Динаміка δ_i з порогом $Q_{0.10}$ та позначенням вилучених кадрів

3. Гіпотеза НЗ. Сукупна дія критеріїв $I(X)$ та δ_i підвищує різноманітність вибірки у просторі ознак.

Підтвердження. Для кількісного вимірювання різноманітності використано міру середньої косинусної відстані до $k=10$ найближчих сусідів у просторі головних компонент:

$$d(i) = \frac{1}{k} \sum_{j \in N_k(i)} \left(1 - \frac{y_i^T y_j}{\|y_i\|_2 \|y_j\|_2 + \varepsilon_0} \right), \quad (2.56)$$

де $y_i = y(X_i)$, а $N_k(i)$ - множина k найближчих сусідів [71-73]. Середнє значення $d(i)$ зросло з 0,023 до 0,029, що відповідає відносному зростанню середньої косинусної відстані на 26,1%. Розподіли $d(i)$ до та після фільтрації (рис. 2.15) демонструють зсув у бік більших значень, що свідчить про зменшення локальної щільності вибірки у просторі ознак і, відповідно, про зниження частки взаємно подібних кадрів. Отже, гіпотеза НЗ підтверджується.

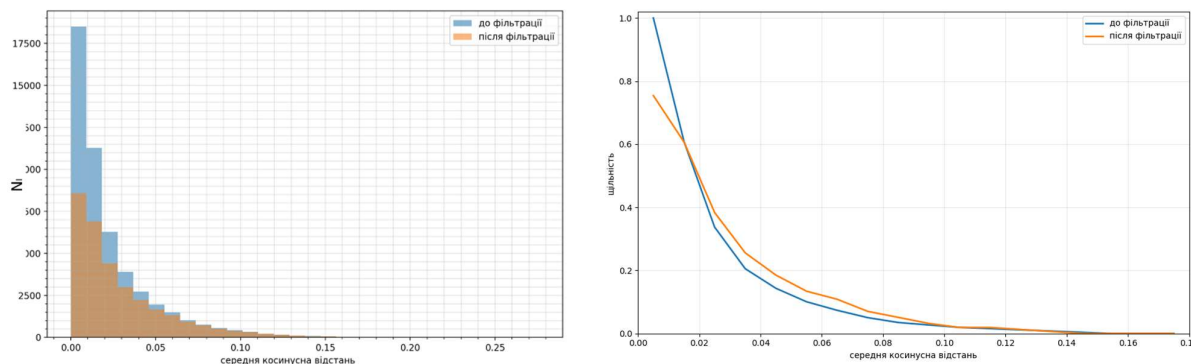


Рис. 2.15. Ліворуч – розподіл міри різноманітності до та після фільтрації.

Праворуч – щільність розподілу

4. Гіпотеза Н4. Відбір за показником $I(X)$ не є випадковим і спрямовано вилучає кадри з низькою інформаційною насиченістю.

Оцінювання. Як незалежний показник використано ентропію яскравості:

$$H(X) = -\sum_b p_b(X) \ln p_b(X), \quad (2.57)$$

де $p_b(X)$ - нормована гістограма рівнів яскравості, тобто оцінка ймовірності потрапляння пікселя до b -го інтервалу [85]. На рис. 2.16 наведено емпіричні розподіли $H(X)$ до та після фільтрації для підвибірки. Після фільтрації спостерігається зменшення частки зображень у області малих значень ентропії $H(X)$, що відповідає однорідним і малоконтрастним зображенням. Це узгоджується з прикладами вилучених зображень на рис. 2.12 (а). Отже, зміна розподілу має спрямований характер і підтримує гіпотезу Н4.

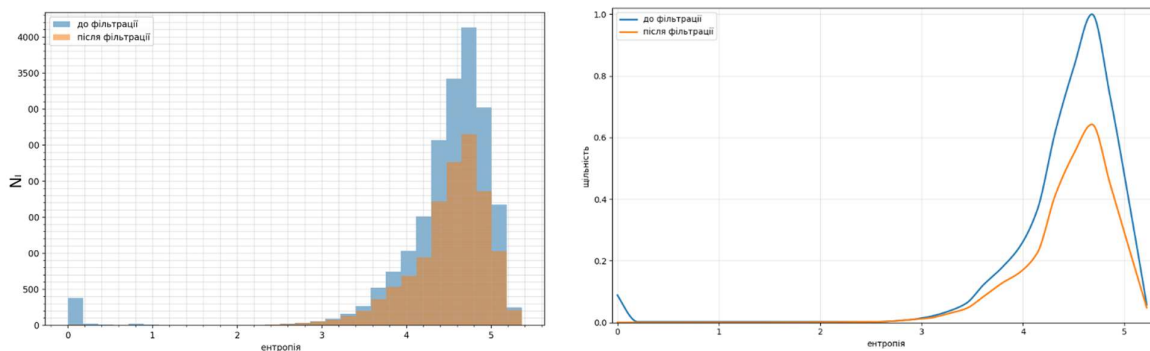


Рис. 2.16. Ліворуч – розподіл ентропії яскравості до та після фільтрації. Праворуч – щільність розподілу

Для узагальненої інтерпретації взаємодії критеріїв $I(X)$ і δ_i проаналізовано перетин множин прикладів, вилучених кожним із них. У цьому експерименті частка зображень, що одночасно задовольнили обидва критерії, становить 3,76% і дорівнює 15 529 прикладам. Це свідчить про наявність спільної області дії, однак загалом критерії виконують різні функції. Критерій δ_i спрямований на усунення дубльованих кадрів у межах послідовностей, тобто зображень із малою міжкадровою зміною ознак. Натомість критерій $I(X)$ вилучає кадри з низькою виразністю ознак, які добре апроксимуються низьковимірною моделлю у просторі ознак, і тим самим сприяє зростанню різноманітності вибірки. У результаті застосування обох критеріїв із початкової множини збережено 263 701 прикладів, що становить 63,85% від початкового обсягу.

Для візуальної ілюстрації структури вибірки у просторі головних компонент наведено проєкцію в площину перших двох компонент з позначенням причин вилучення (рис. 2.17). Ця проєкція має ілюстративний характер, оскільки кількість компонент для вилучення було взято як $m = 32$, однак дозволяє якісно показати, що вилучені приклади концентруються в компактних областях, тоді як збережені займають ширшу область простору.

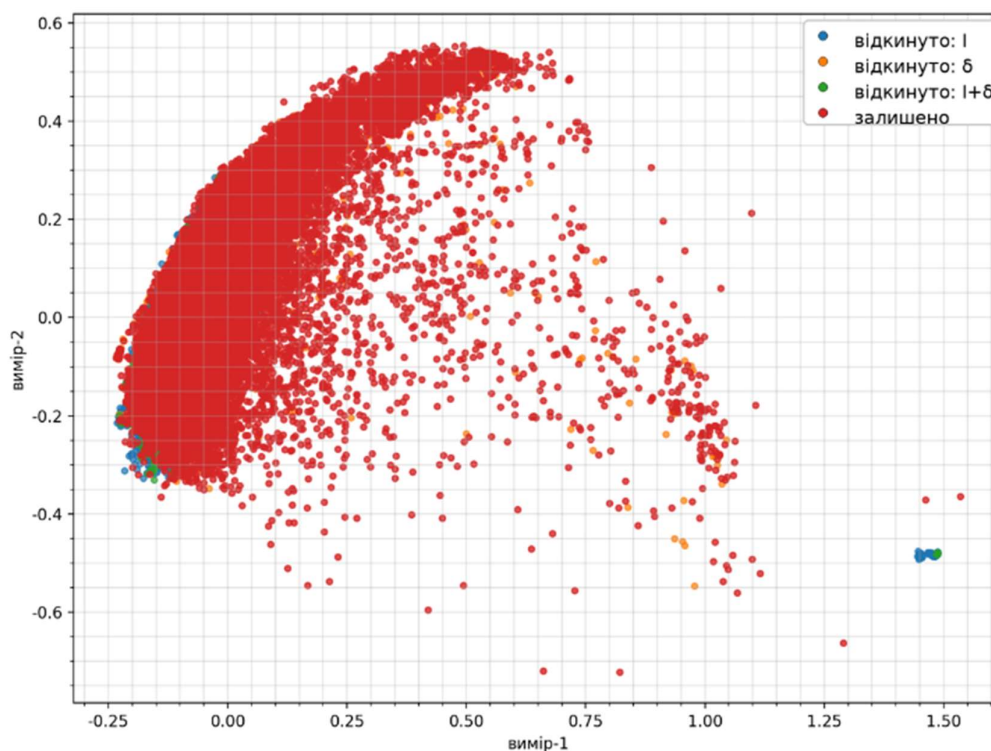


Рис. 2.17. Проекція в площину перших двох головних компонент із позначенням причин вилучення

Таким чином, запропонована процедура фільтрації забезпечує вилучення малоінформативних прикладів за критерієм $I(X)$, усунення дублювання у межах послідовностей за критерієм δ_i , а також підвищення різноманітності вибірки у просторі ознак, що підтверджується зростанням середньої відстані до найближчих сусідів.

Висновки до Розділу 2

У розділі розроблено теоретичні та інформаційні засади запропонованої інформаційної технології абсолютної локалізації безпілотного літального апарата за оптичним каналом. Отримано такі результати:

1. Сформульовано математичну постановку задачі абсолютної локалізації безпілотного літального апарата за оптичним каналом, визначено простір геопросторових станів, простори аерознімків і супутникових сцен, простір

телеметричних параметрів, метрику похибки локалізації та критерій оптимальності.

2. Запропоновано нейромережеву модель двоетапного зіставлення аерознімків і супутникових сцен, що поєднує глобальне ранжування кандидатів, порогове оцінювання впевненості, модуль відбору опорних дескрипторів і локальне уточнення на основі взаємних найближчих відповідників.
3. Розроблено метод моделювання, геоприв'язки та побудови навчальних, валідаційних і тестових наборів даних типу «супутниковий знімок – аерознімок», який забезпечив формування набору обсягом понад 413 000 коректних відповідностей і створив основу для подальшого метричного навчання моделі.
4. Розроблено метод видалення малоінформативних цифрових зображень і зменшення повторюваності кадрів на основі латентних ознак автоенкодера, проєкції в простір головних компонент і квантильного відбору. Його застосування забезпечило вилучення 30 % найменш інформативних кадрів, 9,91 % надмірно повторних кадрів та підвищення різноманітності вибірки у просторі ознак на 26,1 %.
5. Отримані в розділі результати є науково-методичною основою запропонованої інформаційної технології та безпосередньо використовуються у подальших розділах для програмної реалізації, побудови векторної бази представлень та кількісного підтвердження працездатності запропонованих методів.

РОЗДІЛ 3. МОДУЛІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ МОДЕЛЮВАННЯ ТА РОЗПІЗНАВАННЯ НАВЧАЛЬНИХ НАБОРІВ АЕРОЗЙОМКИ

У цьому розділі розробляється програмно-алгоритмічна реалізація запропонованої інформаційної технології моделювання та розпізнавання навчальних наборів аерозйомки для абсолютної локалізації безпілотної літального апарата за оптичним каналом. Розділ безпосередньо спирається на результати розділу 2, у якому було сформульовано математичну постановку задачі, запропоновано нейромережеву архітектуру, розроблено метод побудови навчальних наборів даних та метод видалення малоінформативних цифрових зображень. Метою розділу є інтеграція розроблених методів у єдину інформаційну технологію, придатну до практичного використання за обмежених бортових обчислювальних ресурсів. Для цього у розділі розглянуто інформаційну та алгоритмічну структуру технології, організацію векторної бази представлень і алгоритми пошуку найближчих сусідів, програмну реалізацію основних модулів, а також побудову експериментального апаратно-програмного стенда для подальшої кількісної перевірки працездатності запропонованих методів.

Діаграми відображають структуру системи, функціональний поділ між її компонентами, взаємозв'язки між сутностями та послідовність обміну даними під час виконання ключових операцій.

Підрозділ 3.1 Інформаційна та алгоритмічна структура технології розпізнавання даних аерозйомки

Інформаційна технологія розпізнавання даних аерозйомки реалізується у двох стадіях: офлайн та онлайн. Офлайн-стадія виконується на етапі передпольотної підготовки й охоплює побудову місії, формування еталонного набору супутникових фрагментів, обчислення їх векторних представлень та збереження отриманих ознак у векторній базі B . Онлайн стадія здійснюється під час польоту та полягає в обробленні поточних кадрів аерозйомки, обчисленні їх векторних представлень і пошуку найближчих кандидатів у векторній базі з подальшим отриманням оцінки геопросторового стану. Такий поділ переносить ресурсоемні обчислення на етап

підготовки та зменшує часову затримку локалізації в режимі реального часу. На рис. 3.1 наведено узагальнену схему розпізнавання даних аерозйомки за попередньо сформованою базою супутникових зображень та точок маршруту.

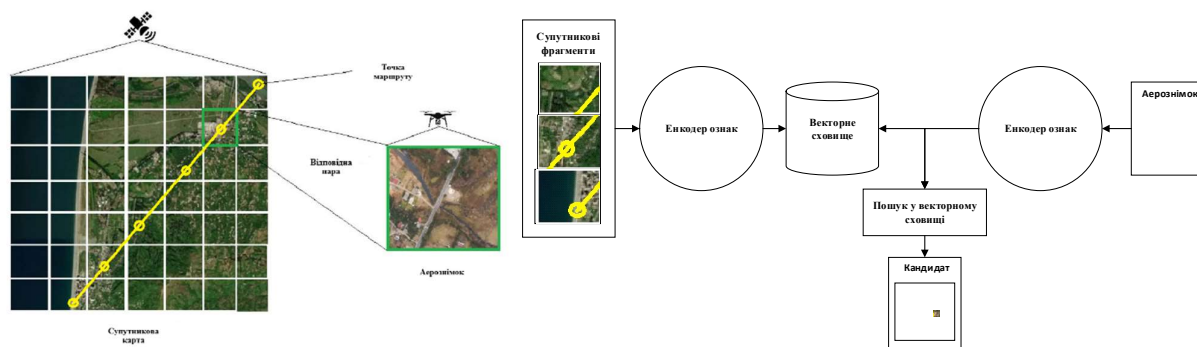


Рис. 3.1. Схема розпізнавання даних аерозйомки

Ключовою операцією онлайн-стадії є зіставлення поточного кадру аерозйомки X_a^t з еталонними супутниковими зображеннями $\{X_s^j, j = 1, \dots, N\}$ у просторі векторних представлень та відбір найближчих кандидатів. Обчислення ознак супутникових зображень у режимі реального часу для кожного вхідного кадру передбачало б багаторазове застосування моделі до значної кількості еталонних фрагментів у межах області пошуку. Це підвищує обчислювальні витрати та робить неможливим дотримання вимог щодо затримки за високої частоти надходження кадрів і обмежених обчислювальних ресурсів бортового комп'ютера. На відміну від аерозйомки, супутникові дані протягом виконання місії є практично незмінними, тому їхні векторні представлення можуть бути обчислені один раз на офлайн-стадії та повторно використані під час виконання локалізації.

У підготовленій векторній базі B кожному супутниковому фрагменту ставиться у відповідність його векторне представлення та сукупність супровідних даних, що забезпечують геоприв'язку й відтворюваність. Структура B визначається співвідношенням (2.4). Супровідні дані включають, зокрема, параметри геопросторового стану s^j , а також атрибути його побудови, які дозволяють встановити відповідність між векторним представленням і конкретною ділянкою місцевості. Офлайн стадія формування B охоплює послідовність

операцій: визначення області інтересу місії, формування множини супутникових фрагментів із фіксацією геоприв'язки, приведення фрагментів до уніфікованого вигляду представлень, обчислення векторних представлень за допомогою f_θ та занесення отриманих векторів ознак разом із супровідними даними до B . Для забезпечення швидкого доступу до еталонів під час подальшого використання бази формується пошуковий індекс, який підтримує пошук найближчих сусідів у просторі представлень. Блок-схему офлайн-процедури побудови та наповнення векторної бази B супутниковими представленнями й метаданими наведено на рис. 3.2.

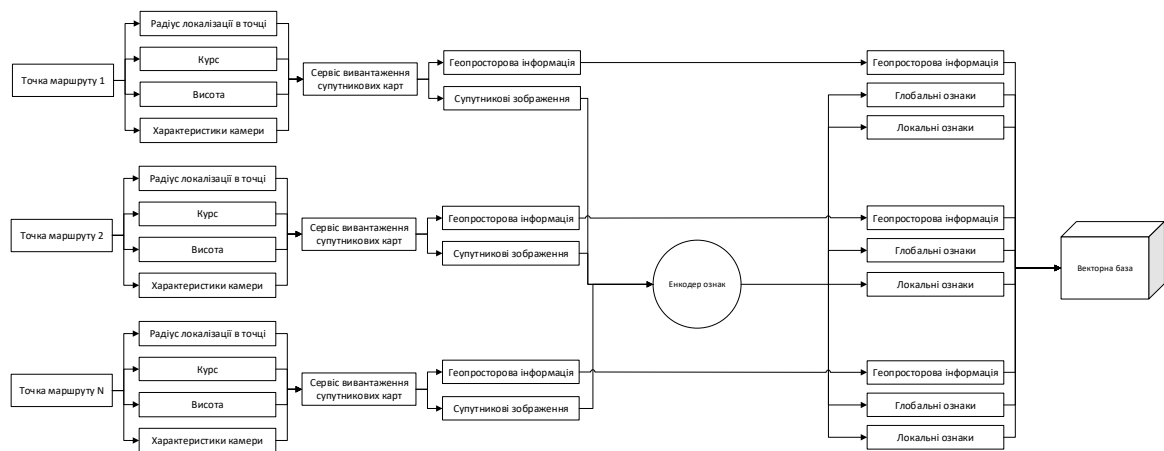


Рис. 3.2. Загальна схема побудови векторної бази

За базової конфігурації зйомки, коли оптична вісь камери зорієнтована близько до вертикального напрямку, масштаб аерознімка узгоджується із пласким поданням супутникової карти. У цьому випадку ширина ділянки місцевості, що потрапляє в кадр, визначається (2.31), а просторове розрізнення аерознімка співвідношенням (2.35). За малих висот польоту h значення $S_a(h)$ (2.35) зменшується, що відповідає підвищеній деталізації аерознімка. Водночас просторове розрізнення супутникового шару обмежується значенням $S_{pix}(z, \varphi)$ (2.29), тому за малих висот може виникати суттєва невідповідність масштабів між аерознімком і супутниковим фрагментом.

Після завершення офлайн-стадії сформована векторна база B разом із пошуковим індексом і супровідними метаданими переноситься на борт БПЛА та використовується під час виконання польотного завдання. У процесі польоту формується потік синхронізованих даних аерозйомки та телеметрії; для кожної часової мітки t доступні поточний кадр аерозйомки X_a^t і відповідні бортові параметри η^t а також апріорна оцінка планового положення, отримана від інерційної системи навігації. На основі кадру X_a^t за допомогою нейромережевого кодувальника f_θ обчислюється вектор глобальних ознак $Z_a^t = f_\theta(X_a^t)$, який надалі використовується як запит до векторної бази. Пошук відповідності у B виконується багатостадійно з метою зменшення обчислювальних витрат і підвищення стійкості локалізації. На першій стадії реалізується просторове обмеження області пошуку за апріорною координатою $\tilde{p}^t$, отриманою з інерційної системи навігації. Нехай для поточного стану задано радіус локалізації r_t , що відбиває очікувану похибку апріорної навігаційної оцінки та обмеження місії. Тоді формується множина кандидатів

$$C_t = \left\{ j \in \{1, \dots, N\} : \|p_s^j - \tilde{p}^t\|_2 \leq r_t \right\}, \quad (3.1)$$

де p_s^j - планові координати супутникового фрагмента X_s^j , що зберігаються у метаданих елемента бази B . Таким чином, подальші порівняння виконуються не з усією множиною еталонів, а лише в межах C_t .

На другій стадії здійснюється ранжування відібраних кандидатів у просторі глобальних представлень за косинусною відстанню між вектором Z_a^t і векторними представленнями супутникових зображень, збережених у B . Для кожного $j \in C_t$ обчислюється

$$d_{\cos}(Z_a^t, Z_s^j) = 1 - \frac{\langle Z_a^t, Z_s^j \rangle}{\|Z_a^t\|_2 \|Z_s^j\|_2 + \varepsilon_0}, \quad (3.2)$$

де $Z_s^j = f_\theta(X_s^j)$ [7, 14]. За результатами ранжування формується підмножина $K_t \subseteq C_t$ фіксованої кількості елементів $|K_t|=10$, яка містить 10 кандидатів із найменшими значеннями $d_{\cos}$.

Третя стадія призначена для усунення неоднозначностей, що виникають через повторювані структури місцевості та доменний зсув. Для кожного кандидата $j \in K_t$ виконується порівняння за локальними ознаками між поточним кадром X_a^t та відповідним супутниковим фрагментом X_s^j . Локальні ознаки отримуються нейромережевим енкодером у вигляді множин карт локальних ознак, сформованих з його проміжних представлень, після чого обчислюється оцінка узгодженості $S_{loc}(X_a^t, X_s^j)$, яка характеризує якість локального зіставлення. Підсумковий кандидат визначається правилом [14]

$$j^*(t) = \arg \max_{j \in K_t} S_{loc}(X_a^t, X_s^j). \quad (3.3)$$

Для наочності онлайн-стадію локалізації подано у вигляді узагальненої схеми (рис. 3.3), яка відображає послідовність оброблення кадру X_a^t , залучені вхідні дані $(\eta_t, \tilde{p}_t, r_t)$ та формування оцінки положення за результатами вибору еталона $j^*(t)$ згідно з (3.1)–(3.3).

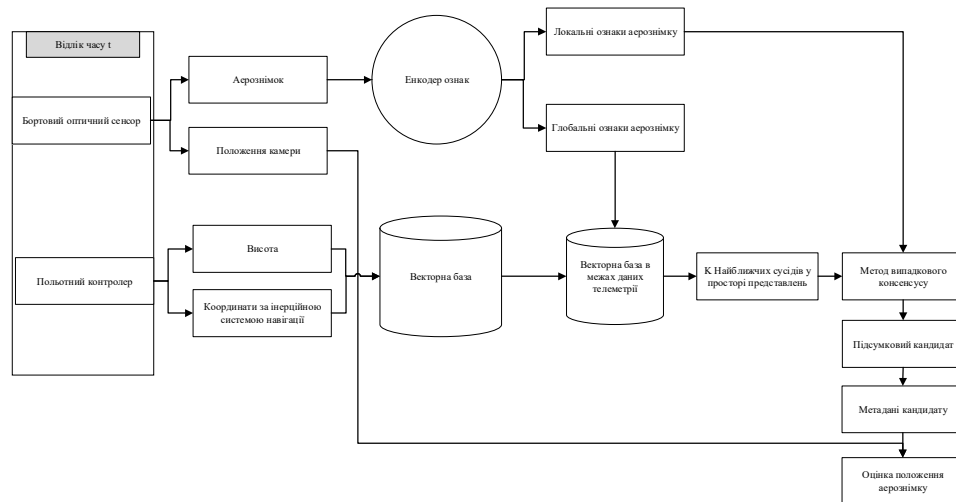


Рис. 3.3 Узагальнена схема онлайн-стадії розпізнавання

Оцінка геопросторового стану на часовій мітці t формується за метаданими вибраного еталона $X_s^{j^*(t)}$. При цьому планові координати беруться з геоприв'язки супутникового фрагмента, а висота та орієнтація - з бортової телеметрії η' , що забезпечує узгодження з моделлю стану, введеною у розділі 2.1. Запропонована онлайн-процедура поєднує просторове обмеження області пошуку, швидке ранжування за глобальними ознаками та верифікацію за локальними ознаками. Це дає змогу забезпечити прийнятну часову затримку при обмежених обчислювальних ресурсах бортового комп'ютера та зберегти стійкість локалізації на ділянках із повторюваними або слабоінформативними структурами.

Підрозділ 3.2 Організація векторної бази представлень та алгоритми пошуку найближчих сусідів

Векторна база представлень B , визначена співвідношенням (2.4), реалізує етап глобального відбору кандидатів у двоетапній архітектурі розпізнавання, описаній у підрозділах 2.2 і 2.3. На цьому етапі необхідно оперативно ранжувати велику кількість супутникових фрагментів за глобальними дескрипторами, одночасно враховуючи просторові обмеження, задані співвідношенням (3.1). З огляду на ці вимоги, для реалізації векторної бази в роботі обрано систему Qdrant [64]. Такий вибір зумовлений кількома чинниками. По-перше, ця система підтримує наближений пошук найближчих сусідів на основі HNSW [87], що забезпечує швидке глобальне ранжування без повного перебору еталонів. По-друге, на рівні колекції в ній може бути зафіксована косинусна метрика, узгоджена з правилом ранжування, заданим співвідношенням (3.2). По-третє, підтримується фільтрація за супровідними полями, завдяки чому просторове обмеження множини кандидатів реалізується без винесення відповідної логіки в окреме зовнішнє сховище. По-четверте, система придатна до локального розгортання, що є важливим для експериментальної програмної реалізації в умовах обмежених обчислювальних ресурсів.

У колекції Qdrant використовується косинусна метрика, узгоджена з математичною постановкою пошуку в підрозділі 2.2 та зі співвідношенням (3.2). Фіксація метрики на рівні сховища усуває неоднозначність ранжування та забезпечує узгодженість процедури пошуку з її математичною постановкою. Використання геопросторових умов відбору за супровідними полями забезпечує просторову локалізацію області пошуку, завдяки чому мінімізуються витрати обчислювальних ресурсів на опрацювання завідомо нерелевантних еталонів. Зазначене є особливо важливим в умовах обмеженості ресурсів. На рис. 3.4 узагальнено логіку повернення K найближчих результатів у просторі ознак.

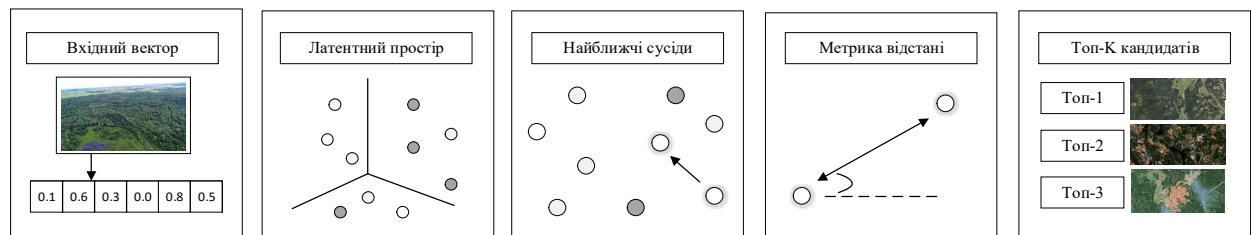


Рис. 3.4. Пошук за подібністю у Qdrant та повернення K найближчих результатів

На рис. 3.5 подано структуру даних Qdrant на рівні колекції, де кожен запис містить вектор ознак і супровідні поля, а індекси створюються як для векторного пошуку, так і для відбору за метаданими.

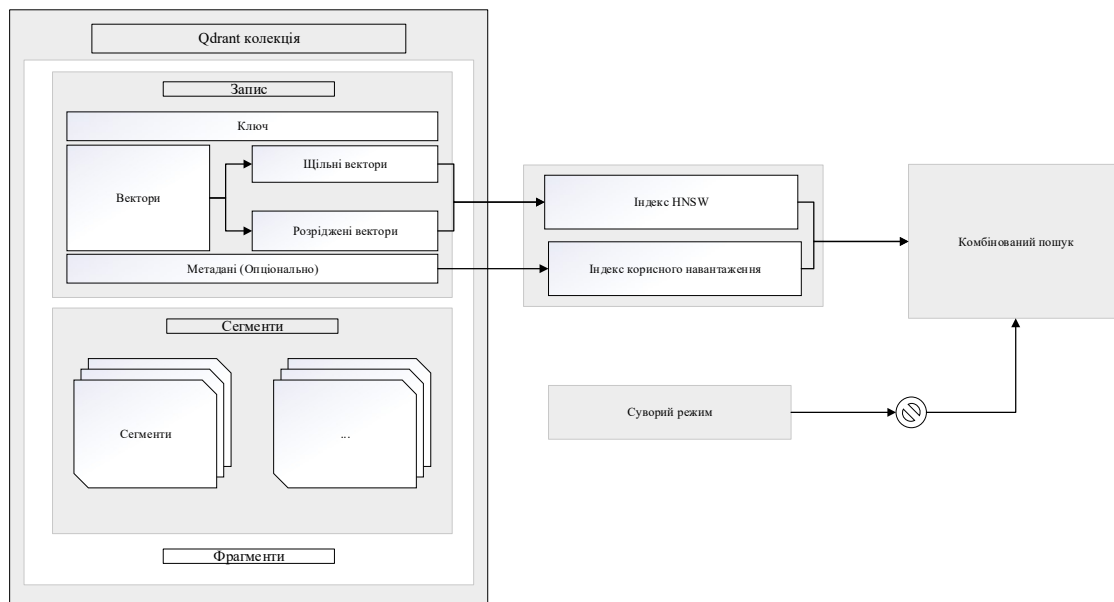


Рис. 3.5. Архітектура та структура даних Qdrant

Наповнення колекції виконується на підготовчій стадії як послідовність операцій, узгоджених із формуванням еталонної бази супутникових фрагментів. Для кожної навігаційної точки маршруту формується набір супутникових зображень із заданим кроком уздовж траєкторії та з урахуванням вікна можливих значень курсу. Така побудова забезпечує наявність еталонів, які відповідають різним орієнтаціям камери відносно сцени, що є суттєвим для коректного пошуку за подібністю. З огляду на значний обсяг локальних ознак, їх зберігання безпосередньо в записі бази є нераціональним через зростання розміру сховища та часу обміну даними. Тому локальні ознаки зберігаються окремо у файловому сховищі, а в колекції фіксується лише посилання на відповідний файл. У такий спосіб у векторній базі зберігається компактний глобальний опис для швидкого пошуку, тоді як детальні дані, потрібні для уточнення, завантажуються лише для обмеженої кількості відібраних кандидатів. Кожен запис колекції має структуру

$$(id_i, f_{\theta,i}(X_s), v_i),$$

де id_i - унікальний ідентифікатор фрагмента, $f_{\theta,i}(X_s)$ - глобальні ознаки, v_i супровідні поля. У супровідних полях зберігаються координати чотирьох кутів фрагмента (φ, λ) , що однозначно задає просторове положення й область покриття, а також дає змогу застосовувати просторові обмеження в запитах. Додатково фіксуються значення курсу ψ_i , висоти h_i та індекс навігаційної точки маршруту, з якою пов'язаний фрагмент. Наявність координат центра фрагмента забезпечує реалізацію просторового обмеження множини кандидатів відповідно до (3.1), тоді як координати кутів використовуються для геопросторової прив'язки та подальшого геометричного уточнення результату розпізнавання. Для підвищення продуктивності завантаження виконується пакетно, а після додавання контролюється стан готовності колекції до пошуку, зокрема завершеність побудови індексів. Діаграму наповнення векторної бази наведено на рис. 3.6.

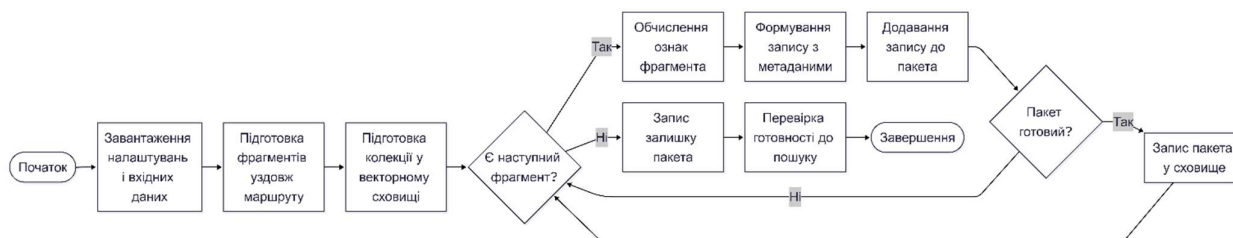


Рис. 3.6. Діаграма наповнення векторної бази представлень

Підрозділ 3.3 Програмна реалізація та інтеграція модулів інформаційної технології

Програмну реалізацію інформаційної технології розпізнавання даних аерозйомки виконано у вигляді модульної системи з можливістю адаптивного налаштування. Така організація забезпечує можливість заміни джерела відеопотоку та каналу телеметрії без зміни логіки основного циклу оброблення. Векторна база представлень у запропонованій реалізації побудована на Qdrant відповідно до вимог, обґрунтованих у підрозділі 3.2. Вибір реалізацій здійснюється через конфігураційні файли та параметри середовища, а створення об'єктів виконано за фабричним принципом [19], що зменшує зв'язаність компонентів і спрощує експериментальне порівняння альтернативних налаштувань. Узагальнену структуру програмної системи та зв'язки між ключовими модулями подано на рис. 3.7-3.9.

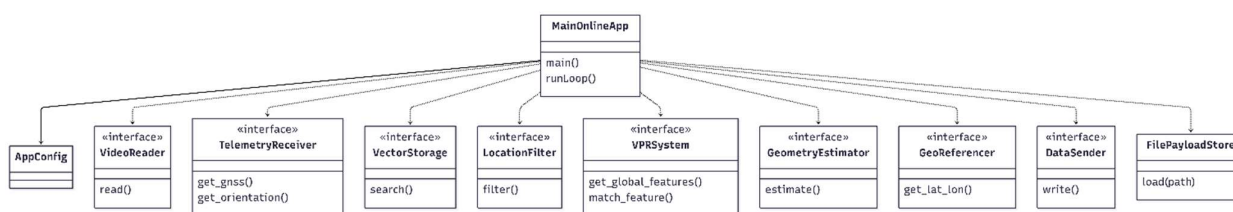


Рис. 3.7. UML-діаграма загальної архітектури застосунку

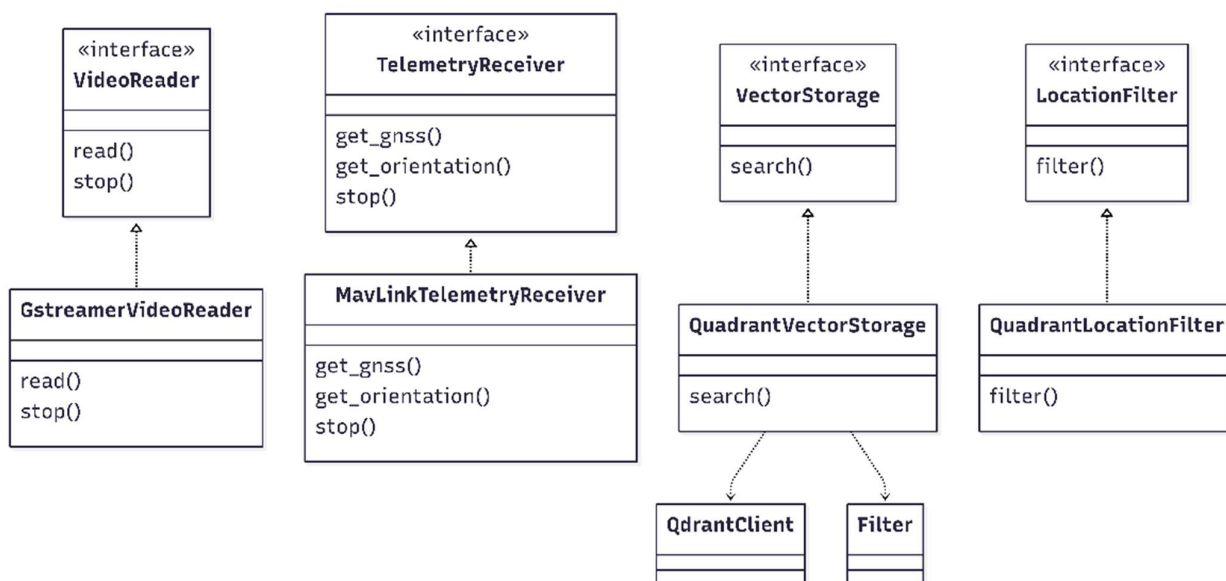


Рис. 3.8. UML-діаграма інфраструктурних адаптерів

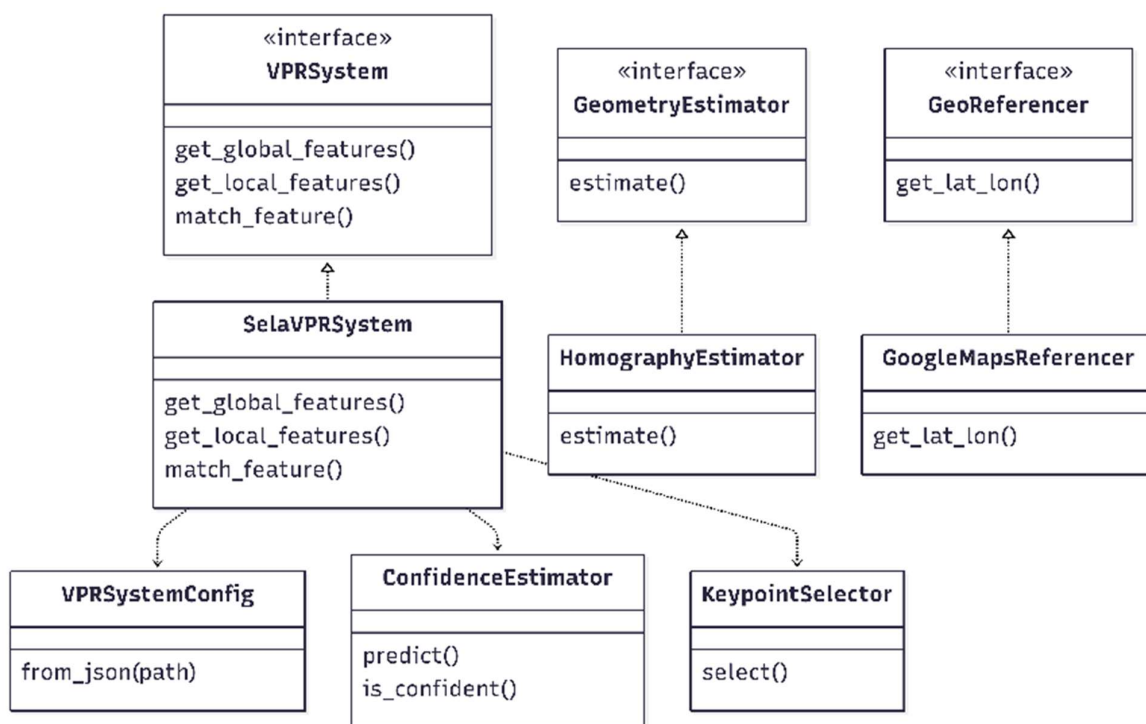


Рис. 3.9. UML-діаграма модуля розпізнавання та зіставлення

Надходження навігаційної інформації та даних орієнтації організовано на основі протоколу MAVLink [20], який є типовим стандартом обміну між польотним контролером, бортовим комп'ютером і наземною станцією. Отримання даних телеметрії реалізовано як окремий модуль, що декодує повідомлення MAVLink,

актуалізує останні валідні значення та надає уніфікований інтерфейс доступу до даних. Перед запуском обчислень для кожного кадру виконується перевірка наявності актуальної телеметрії, що запобігає формуванню запитів до еталонної бази за відсутності початкових даних.

Відеопотік подається до системи через модуль захоплення, побудований на GStreamer [21], із налаштуваннями мінімізації затримки. Структуру підсистем телеметрії та відеозчитування наведено на рис. 3.10-3.11. Узагальнену схему синхронізації потоків даних наведено на рис. 3.24.

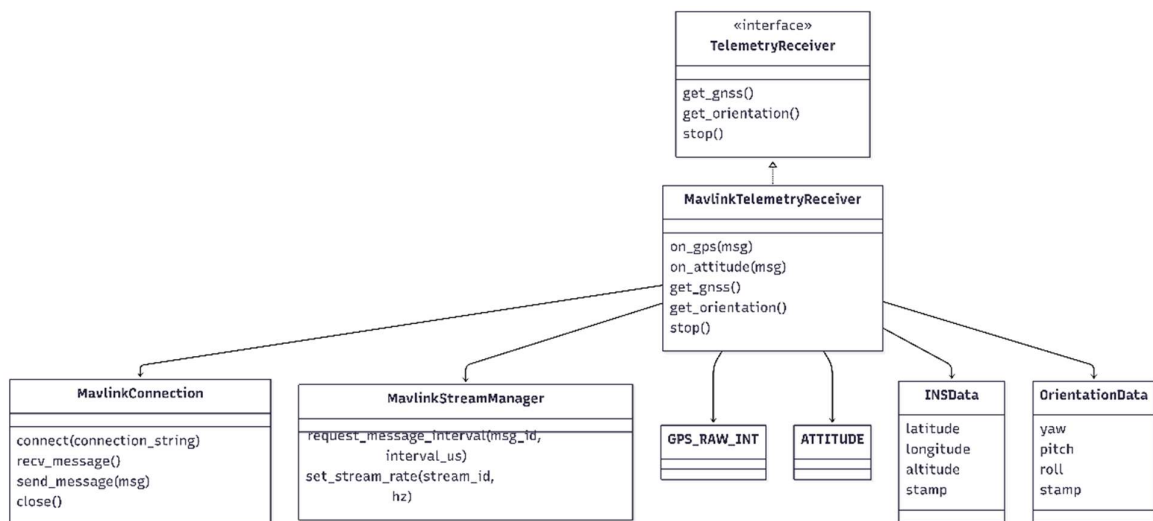


Рис. 3.10. UML-діаграма підсистеми телеметрії

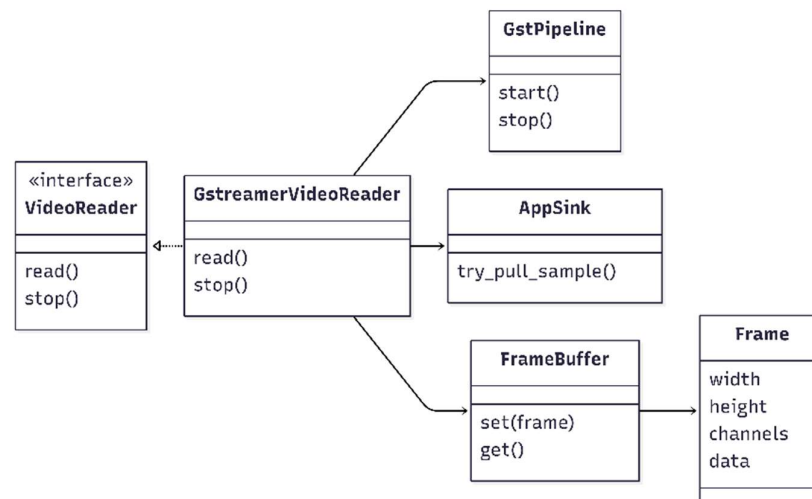


Рис. 3.11. UML-діаграма підсистеми відеозчитування

Обробка в режимі реального часу виконується у вигляді безперервного циклу, де для кожного кадру послідовно: визначаються межі пошуку; обчислюється глобальний дескриптор кадру; надсилається запит до векторного сховища; уточнюється положення. Схему керування етапами обробки кадру та взаємодію модулів, що забезпечують процес локалізації, наведено на рис. 3.12-3.13.



Рис. 3.12. UML-діаграма послідовності обробки кадру

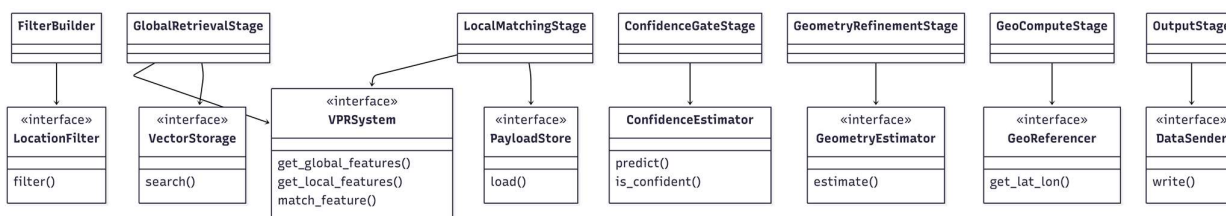


Рис. 3.13. UML-діаграма модулів та залежностей

Первинну стадію пошуку реалізовано як пошук у просторі глобальних ознак, сформованих модулем VPR. Для зберігання векторів і виконання наближеного пошуку найближчих сусідів використано Qdrant, який підтримує одночасне застосування векторної метрики та фільтрів за метаданими, включно з геопросторовими умовами та обмеженнями за орієнтацією. Результатом запиту є впорядкований список кандидатів, до якого входять оцінки подібності та метадані. Для зменшення обсягу передачі даних і прискорення доступу до сховища глобальні вектори зберігаються у векторній базі разом із компактними атрибутами, тоді як об'ємні локальні представлення винесено у файлове сховище з посиланням у метаданих. Завантаження локальних ознак виконується лише для кандидатів, відібраних на стадії глобального ранжування, і тільки в тих випадках, коли модуль оцінювання впевненості не дає підстав завершити формування результату на глобальній стадії. Структуру класів, що реалізують доступ до Qdrant і формування фільтрів пошуку, подано на рис. 3.14.

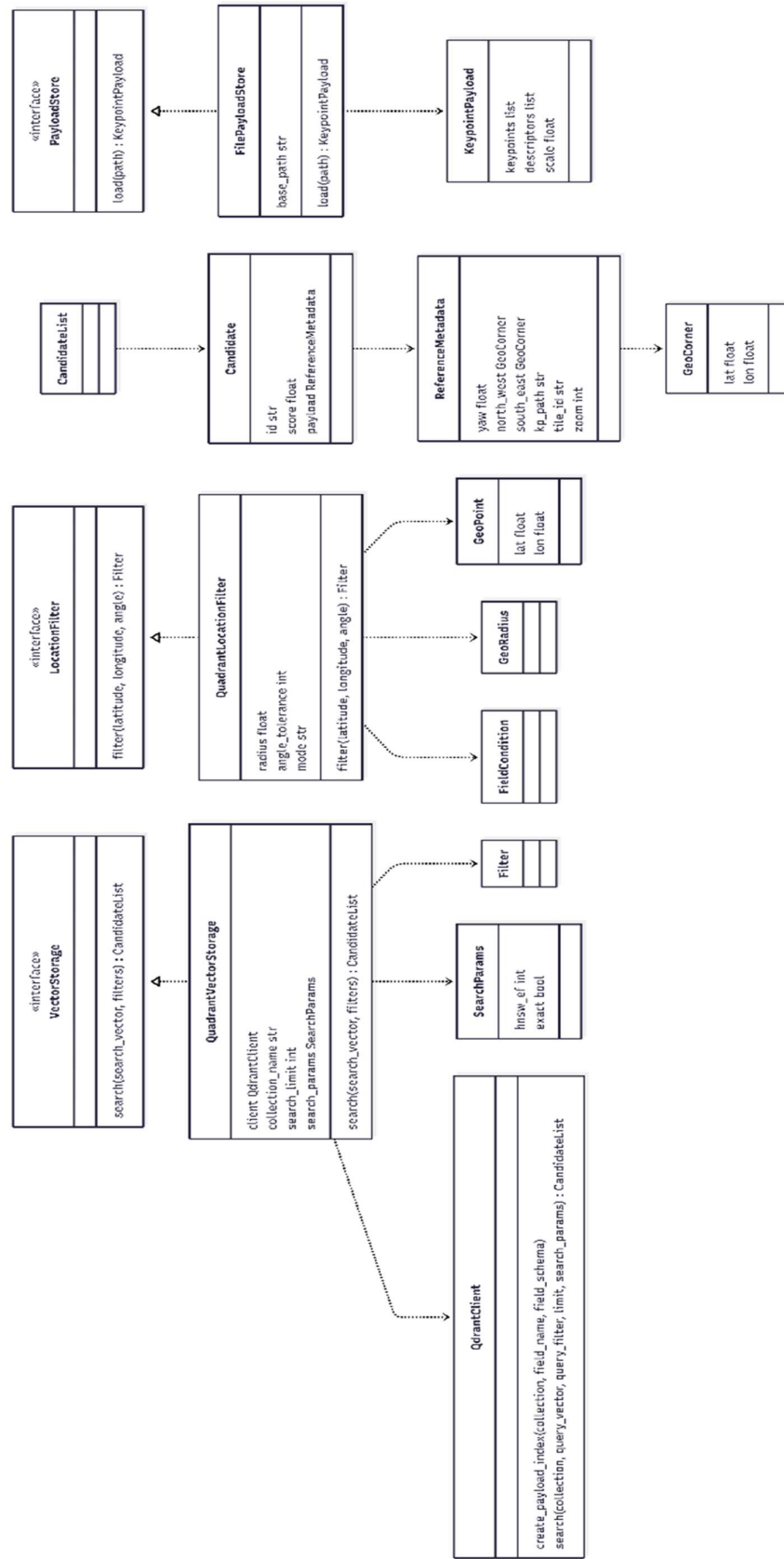


Рис. 3.14. UML-діаграма класів векторного сховища Qdrant

Модуль візуального зіставлення реалізовано на основі нейромережі, описаної у розділі 2.2 та 2.3, а процес обробки доповнено двома модифікаціями, орієнтованими на зменшення затримки без погіршення точності. Перша модифікація полягає у введенні модуля оцінювання впевненості після глобального ранжування. Він аналізує розподіл подібності серед найкращих кандидатів і визначає, чи є рішення однозначним на глобальному рівні. За високої впевненості локальне уточнення пропускається, а результат формується на основі обраного еталону, що зменшує обчислювальні витрати. Друга модифікація реалізує відбір опорних локальних дескрипторів, коли щільне локальне подання стискається до підмножини найбільш інформативних елементів. Це зменшує обсяг даних, які потрібно зберігати й передавати, і скорочує час локального зіставлення. Локальна перевірка відповідностей здійснюється за принципом взаємно найближчих сусідів у просторі ознак, після чого вводиться порогове значення за кількістю узгоджених відповідностей, щоб відбракувати нестійкі випадки ще до геометричного уточнення. У ролі альтернативних схем локального зіставлення в суміжних задачах застосовують SuperGlue, LoFTR та LightGlue, однак вони є обчислювально дорогими для обраного бортового сценарію [61-63]. Структуру класів модуля VPR та місце зазначених модифікацій наведено на рис. 3.15-3.16.

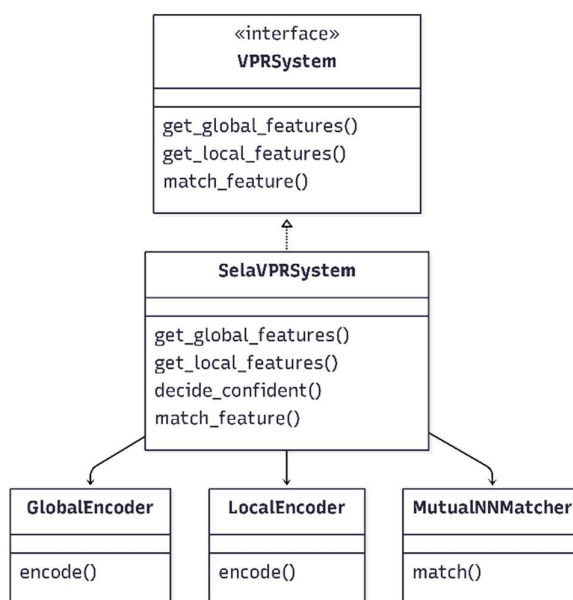


Рис. 3.15. UML-діаграма класів модуля VPR

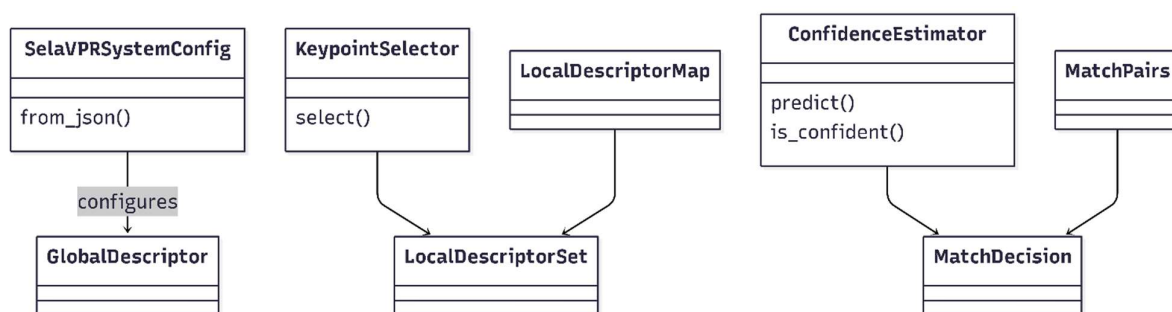


Рис. 3.16. UML-діаграма допоміжних компонентів і класів

Після підтвердження відповідності здійснюється геометричне уточнення положення кадру відносно супутникового фрагмента. Для цього оцінюється гомографічне перетворення за узгодженими парами ключових точок із використанням робастної процедури RANSAC [22]. Далі визначається положення реперної точки кадру, зокрема центра нормованого зображення, у координатах супутникового фрагмента. Піксельні координати точки перетворюють у географічні координати за координатами кутових точок еталонного фрагмента та співвідношеннями картографічної проєкції. Взаємодію модулів геометричного уточнення та геоприв'язки показано на рис. 3.17.

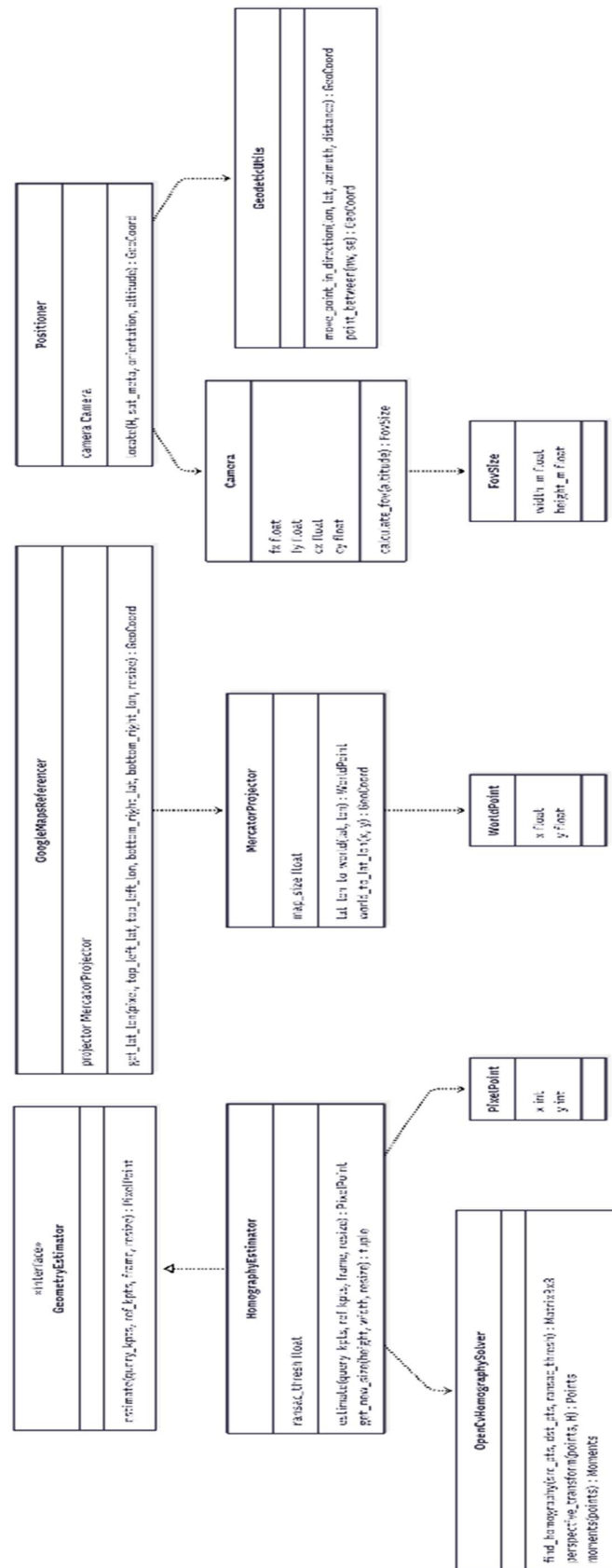


Рис. 3.17. UML-діаграма класів геометричного уточнення

Результат розпізнавання та просторової прив'язки передається у зовнішні підсистеми через модуль виводу. Для контролю функціонування системи передбачено журналювання подій, фіксацію проміжних рішень та збереження допоміжних даних, включно із записом кадрів і телеметрії. Структуру класів, що відповідають за передачу результату розпізнавання та просторової прив'язки, наведено на рис. 3.18.

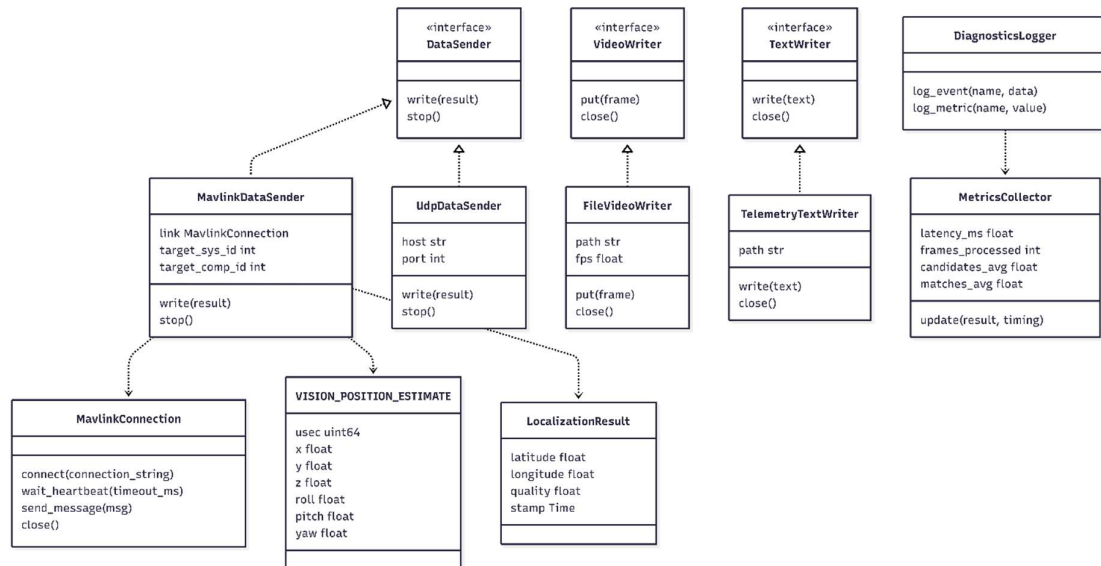


Рис. 3.18. UML-діаграма класів видачі результату

Для експериментальної перевірки реалізовано режим відтворення польоту, у якому використовуються ті самі модулі пошуку та зіставлення, однак джерело телеметрії замінюється на читання з попередньо підготовленого файлу, синхронізованого з відеорядом. Схему класів, що забезпечують режим відтворення польоту, наведено на рис. 3.19.

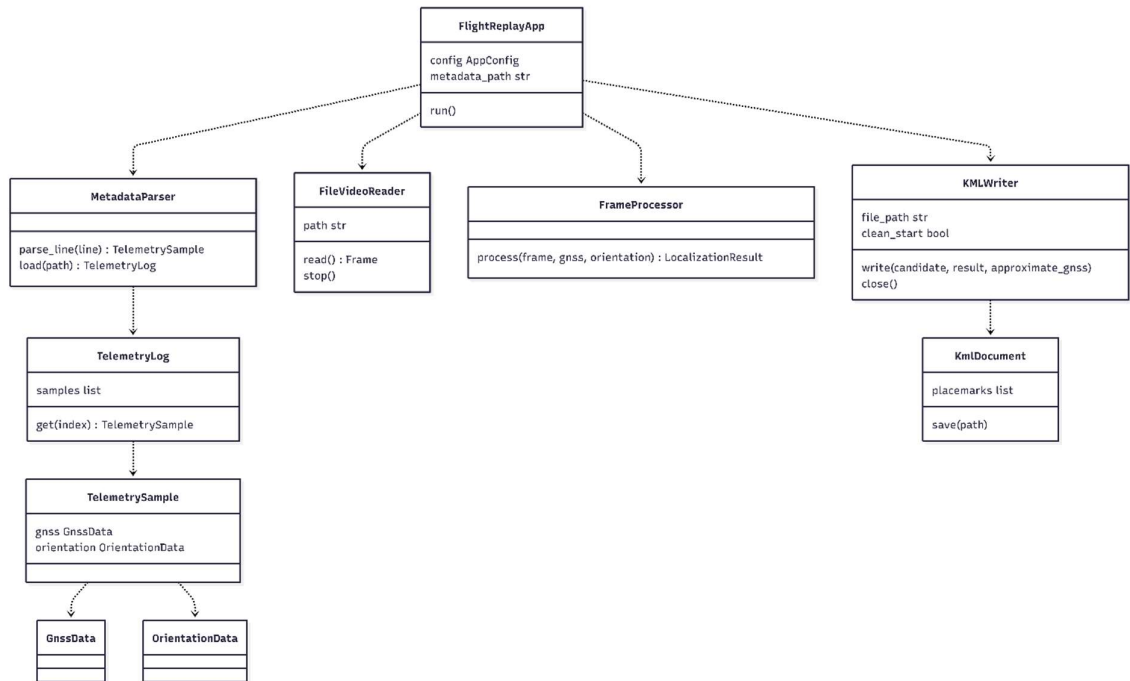


Рис. 3.19. UML-діаграма класів режиму відтворення польоту та формування вихідного KML

Підрозділ 3.4 Експериментальний стенд

Для експериментальної перевірки працездатності запропонованої інформаційної технології моделювання та розпізнавання навчальних наборів аерозйомки сформовано апаратно-програмний стенд, до складу якого входять літальна платформа, високоточний навігаційний модуль, бортовий обчислювальний засіб та оптичний сенсор. Конфігурацію стенду підібрано таким чином, щоб забезпечити обробку відеопотоку на борту в режимі, наближеному до реального часу, а також отримання опорних навігаційних вимірювань для кількісного оцінювання точності розпізнавання та просторової прив'язки [75].

Як носій сенсорного та обчислювального навантаження використано гексакоптер на рамі HEXA_600 зі злітною масою 4,8 кг. Загальний вигляд експериментальної літальної платформи наведено на рис. 3.20.



Рис. 3.20. Експериментальна літальна платформа

Польотний контролер реалізовано на базі CUAV X7+ з установленою прошивкою ArduCopter 4.5.7 [65, 66]. Він забезпечує стабілізацію та навігаційне керування польотом, формування та передачу бортової телеметрії, а також інтеграцію даних від навігаційних і інерціальних датчиків. Зовнішній вигляд польотного контролера наведено на рис. 3.21. Канал радіокерування реалізовано за протоколом ExpressLRS (ELRS) 2,4 ГГц із передавачем TX16S [67]. Силова установка містить двигуни Emax Eco Series II 2814 KV830, а живлення забезпечується акумуляторною батареєю Li-Po 6S ємністю 16 000 мА·год. За типового застосування тривалість польоту становить орієнтовно 30 хв, чого достатньо для виконання серій польових експериментів із реєстрацією телеметрії та відеоданих. Для бортового зв'язку та передавання даних на наземну станцію використано модуль SIYI HM30 [27].

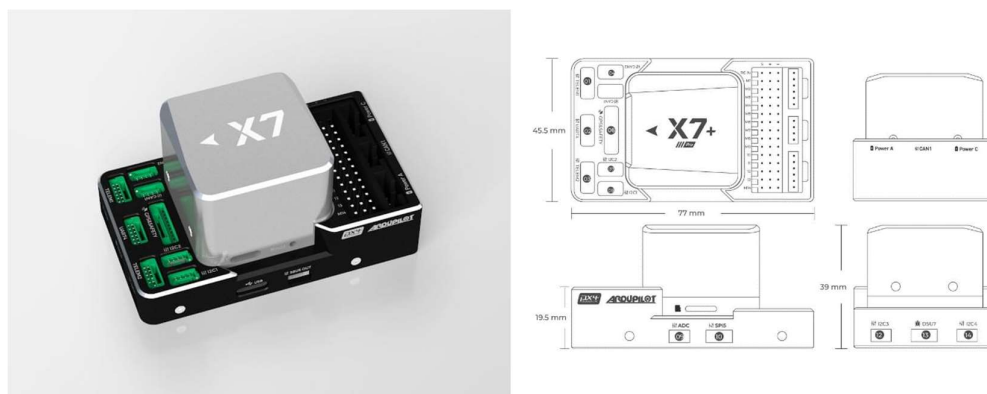


Рис. 3.21. Польотний контролер CUAV X7+

Для формування опорної оцінки положення апарата застосовано модуль супутникової навігації HolyBro H-RTK F9P з інтегрованим компасом. У режимі кінематичного позиціювання в реальному часі (RTK) з диференціальними поправками модуль забезпечує сантиметровий рівень точності визначення координат; за внутрішніми даними точність становить $1 \text{ см} + 1 \text{ ppm}$ за сприятливих умов прийому та за наявності RTK-поправок [26]. Отримані RTK-координати використовуються як еталонна траєкторія під час розрахунку похибок локалізації, отриманої за оптичним каналом, що дає змогу коректно інтерпретувати внесок алгоритмічних рішень у загальну точність системи [26].

Обробку відеопотоку та виконання алгоритмів комп'ютерного зору реалізовано на бортовому комп'ютері NVIDIA Jetson Orin NX 16GB. Зовнішній вигляд бортового комп'ютера наведено на рис. 3.22. Обчислювальна платформа містить багатоядерний ARM-процесор і графічний процесор архітектури NVIDIA Ampere з апаратною підтримкою прискорення матричних обчислень у середовищі CUDA, що є важливим для задач нейромережевого обрахунку та високопродуктивної обробки кадрів безпосередньо на борту [24]. На Jetson Orin NX встановлено Ubuntu 22.04 у складі програмного комплексу JetPack 6.1. Програмну частину реалізовано мовою Python 3.10, а нейромережеві модулі із застосуванням PyTorch та Torchvision з підтримкою CUDA, що забезпечує виконання обчислень на GPU та зменшення затримок під час обробки відеоданих [23, 25, 68–70].

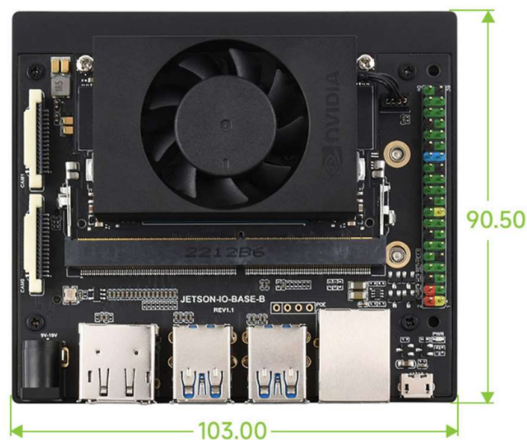


Рис. 3.22. Бортовий комп'ютер NVIDIA Jetson Orin NX 16GB

Як оптичний сенсор використано Raspberry Pi Global Shutter Camera з об'єктивом фокусної відстані 6 мм, зовнішній вигляд якої наведено на рис. 3.23. Під час експериментів камера формує відеопотік із частотою 30 кадрів/с та розміром кадру $1440 \times 1080 \times 3$, що відповідає режиму обробки, прийнятому в програмній реалізації алгоритмів на борту. Використання камери типу global shutter є обґрунтованим для зйомки з рухомої платформи, оскільки експозиція всього кадру відбувається одночасно. На відміну від сенсорів із рухомих затвором (rolling shutter), це зменшує геометричні викривлення та знижує вплив вібрацій на локальні структури зображення, що підвищує відтворюваність ознак і стабільність процедур зіставлення кадрів у задачах локалізації [28, 32].



Рис. 3.23. Оптичний сенсор Raspberry Pi Global Shutter Camera з об'єктивом 6 мм

Кут встановлення камери відносно борта задається механічно, а його зміна може здійснюватися електромеханічним приводом під програмним керуванням в діапазоні $0-90^\circ$ відносно горизонту, що дає змогу адаптувати геометрію зйомки до умов експерименту. Це дозволяє в межах єдиного стенду досліджувати вплив ракурсу, перспективних спотворень і структури сцени на стабільність виділення ознак, точність розпізнавання та якість просторової прив'язки [58].

Обмін даними між польотним контролером і бортовим комп'ютером організовано через послідовний інтерфейс UART із використанням протоколу MAVLink [20]. На польотному контролері задіяно порт TELEM1 зі швидкістю обміну 115200 бод, а на Jetson Orin NX використано апаратний UART, виведений на

контакти GPIO. Частоту передавання телеметричних повідомлень у стенді обмежено значенням 10 Гц, що враховується під час часової прив'язки даних і подальшої інтеграції навігаційних вимірювань із відеопотоком [76]. Для систем, що функціонують у режимі реального часу, узгодження часових міток між вимірюваннями є критичною умовою працездатності. За відсутності синхронізації кадр може бути зіставлений із некоректним станом апарата, що призводить до систематичних помилок часової прив'язки, погіршення точності локалізації та загальної стійкості алгоритму; у граничних випадках це спричиняє накопичення затримки та деградацію роботи всієї бортової системи [76]. Узагальнену схему синхронізації потоків даних наведено на рис. 3.24.

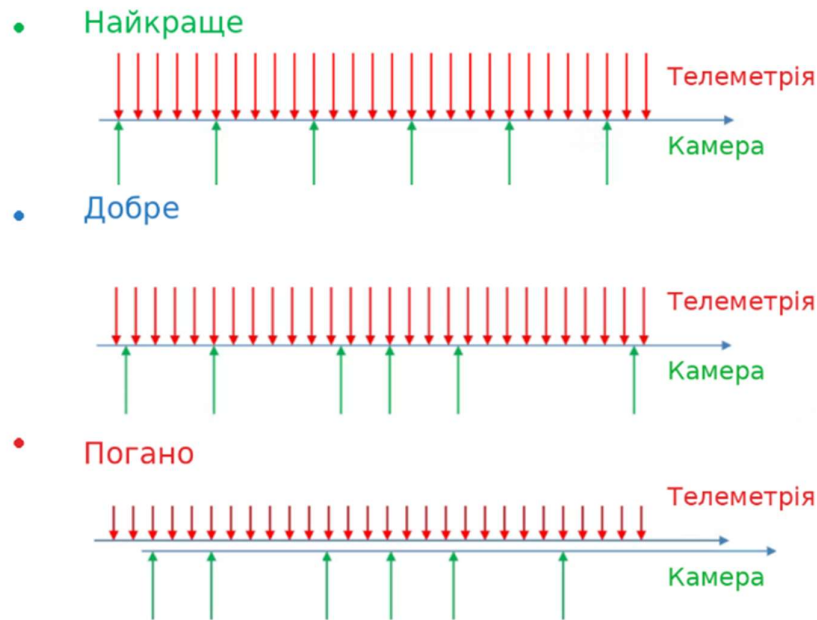


Рис. 3.24. Схема синхронізації потоків даних

Синхронізацію моменту формування кадру та бортових вимірювань реалізовано апаратно. Для цього бортовий комп'ютер формує синхронізаційний імпульс запуску експозиції, який подається на вхід синхронізації матриці камери через сигнальну лінію GPIO. Така схема забезпечує ініціювання експозиції у визначений момент часу та зменшує невизначеність часової прив'язки між кадрами і телеметрією, що є важливим як під час роботи алгоритму локалізації у режимі реального часу, так і під час післяпольотного аналізу результатів [76].

Узгоджена робота компонентів стенду забезпечує відтворювану експериментальну перевірку запропонованої інформаційної технології. Відеопотік обробляється безпосередньо на борту із застосуванням GPU-прискорення. Телеметричні дані надходять із контрольованою частотою 10 Гц і використовуються для часової прив'язки навігаційного стану до кадрів. Високоточні RTK-вимірювання формують опорні координати, які застосовуються як еталон під час кількісного оцінювання похибки локалізації в умовах, наближених до практичного застосування.

Висновки до Розділу 3

У розділі розроблено програмно-алгоритмічні засади реалізації запропонованої інформаційної технології моделювання та розпізнавання навчальних наборів аерозйомки. Отримано такі результати:

1. Розроблено інформаційну та алгоритмічну структуру інформаційної технології, що інтегрує запропоновані в роботі метод побудови навчальних наборів даних, метод видалення малоінформативних цифрових зображень, нейромережеву архітектуру двоетапного зіставлення, метод побудови тривимірного представлення та модуль тривимірного узгодження супутникових сцен, а також модулі формування глобальних і локальних ознак, локального уточнення та геоприв'язки.
2. Розроблено структуру векторної бази представлень еталонних супутникових сцен і алгоритми пошуку найближчих сусідів, придатні для роботи в режимі, наближеному до реального часу, за обмежених бортових обчислювальних ресурсів.
3. Реалізовано програмні модулі запропонованої інформаційної технології, зокрема модулі формування глобальних і локальних ознак, оцінювання впевненості, відбору опорних дескрипторів, локального уточнення, векторного пошуку та просторової прив'язки.

4. Інтегровано розроблені модулі у єдину програмну систему з офлайн- та онлайн-стадіями функціонування, що забезпечує попередню підготовку еталонної бази та оперативне опрацювання бортових кадрів під час польоту.
5. Сформовано експериментальний апаратно-програмний стенд для натурної перевірки працездатності запропонованої інформаційної технології та для подальшого кількісного підтвердження працездатності запропонованих методів у розділі 4.

РОЗДІЛ 4. ДОСЛІДЖЕННЯ ТА ОЦІНКА ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ

У цьому розділі наведено результати експериментального дослідження та кількісного підтвердження працездатності запропонованих у дисертації методів, моделей і засобів. Метою розділу є експериментальна перевірка працездатності запропонованої інформаційної технології, нейромережевої архітектури двоетапного зіставлення, методу побудови навчальних наборів даних типу «супутниковий знімок – аерознімок», методу видалення малоінформативних цифрових зображень, а також методу побудови тривимірного представлення супутникових сцен.

У підрозділі 4.1 досліджено інформативність зображень та проаналізовано розподіли у латентному просторі автоенкодера для кількісного підтвердження працездатності методу видалення малоінформативних цифрових зображень. У підрозділі 4.2 оцінено якість запропонованої нейромережевої архітектури та інформаційної технології в задачі розпізнавання даних аерозйомки. У підрозділі 4.3 виконано порівняльний аналіз із наявними методами. У підрозділі 4.4 наведено результати побудови тривимірного представлення супутникових сцен та кількісно оцінено вплив цього методу на якість розпізнавання в складних геометричних режимах зйомки. Отримані результати дають змогу встановити межі застосовності запропонованих методів і підтвердити практичну придатність розробленої інформаційної технології для задач резервної навігації без супутникової підтримки.

Підрозділ 4.1 Дослідження інформативності зображень та аналіз розподілів у латентному просторі автоенкодера

У цьому підрозділі наведено результати експериментального аналізу латентних ознак, сформованих енкодером попередньо навченої моделі автоенкодера, та обґрунтовано, як властивості отриманого простору представлень впливають на інформативність кадрів і на стійкість подальшого зіставлення аерознімків із супутниковими фрагментами. Дослідження виконано на одному з польотів за довільно обраним маршрутом №62, що наведено в таблиці 2.1.

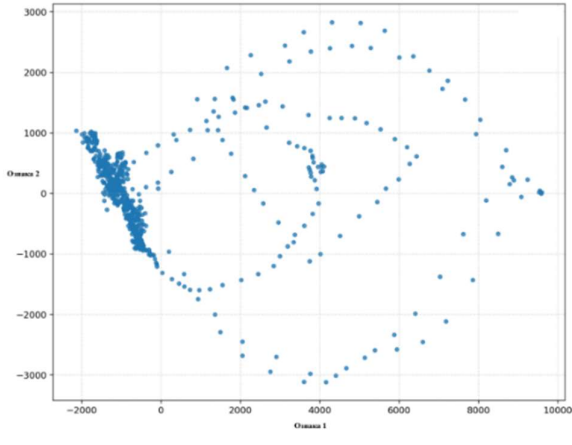
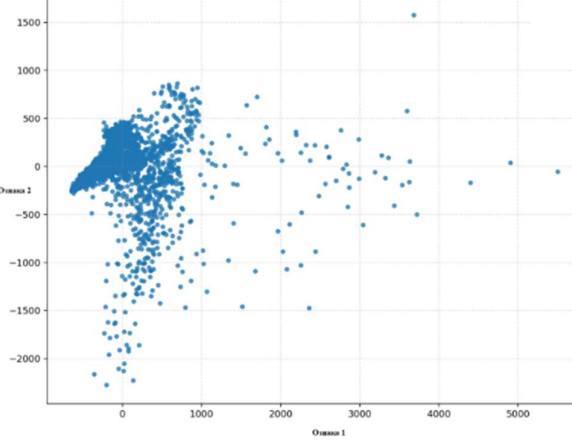
Отримані результати узагальнено для вибірки, сформованої на висотах 100 м., 150 м. та 200 м. Такий підхід дає змогу оцінити вплив масштабу сцени та геометрії зйомки на структуру розподілів ознак. Також він дає змогу оцінити вплив цих факторів на розділення позитивних і негативних пар. Підходи до аналізу аномалій у латентному просторі та інтерпретації набутих представлень раніше апробовано в авторських роботах [34, 35].

Досліджено глобальну геометрію розподілів латентних ознак для повного маршруту №62. Для кожної висоти виконано двовимірне відображення методом головних компонент, для трьох множин ознак: аерознімків, позитивних супутникових і негативних супутникових фрагментів. Відповідні візуалізації наведено у табл. 4.1 для висоти 100 м., у табл. 4.2 для висоти 150 м. та у табл. 4.3 для висоти 200 м. Таке подання дозволяє оцінити, наскільки близькими є розподіли ознак аерознімків і еталонних супутникових фрагментів у спільному просторі, а також наскільки відокремлюються негативні супутникові приклади, що безпосередньо пов’язано зі стійкістю пошуку.

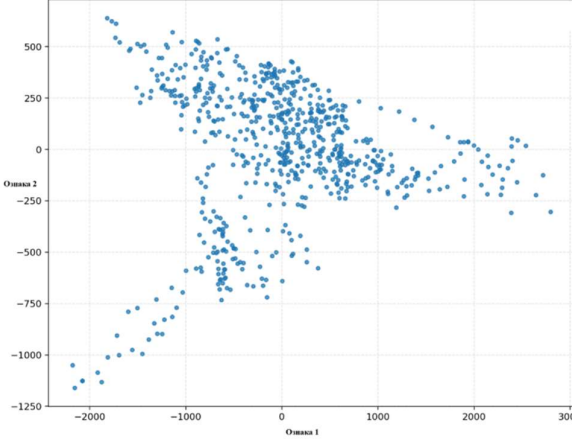
Таблиця 4.1. Двовимірна проєкція латентного простору маршруту №62 методом головних компонент для висоти 100 м.

Підвибірка	Рисунок
Дані аерозйомки	

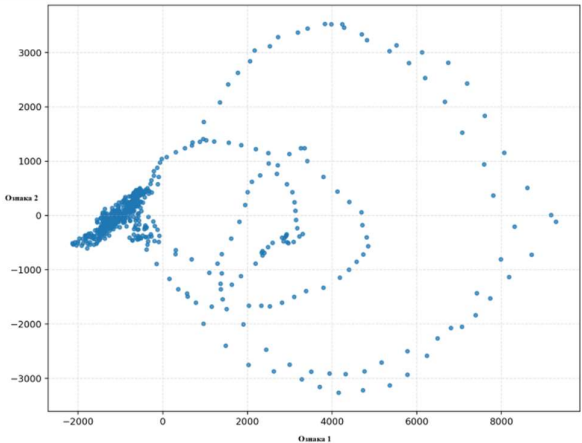
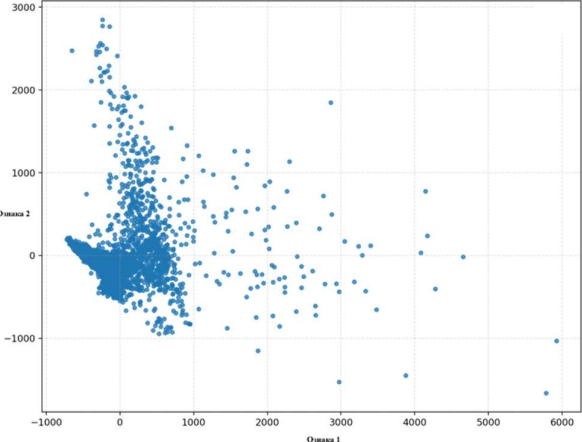
Продовження таблиці 4.1

Підвибірка	Рисунок
Відповідні супутникові знімки	
Негативні супутникові знімки	

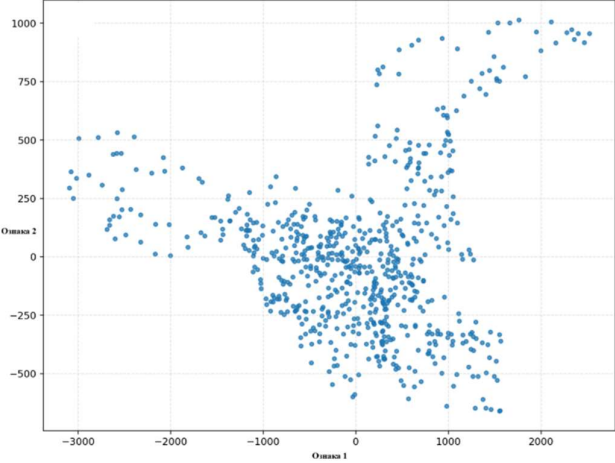
Таблиця 4.2. Двовимірна проєкція латентного простору маршруту №62 методом головних компонент для висоти 150 м.

Підвибірка	Рисунок
Дані аерозйомки	

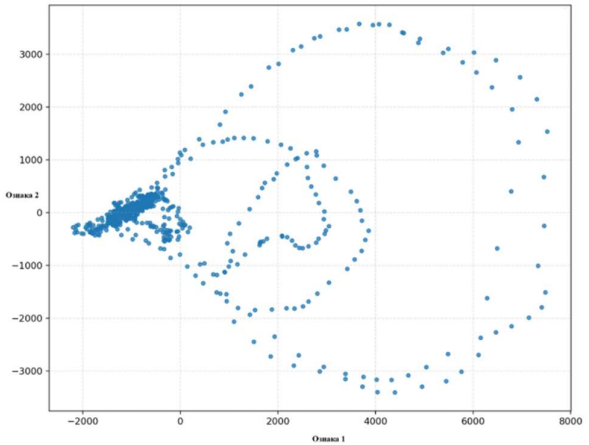
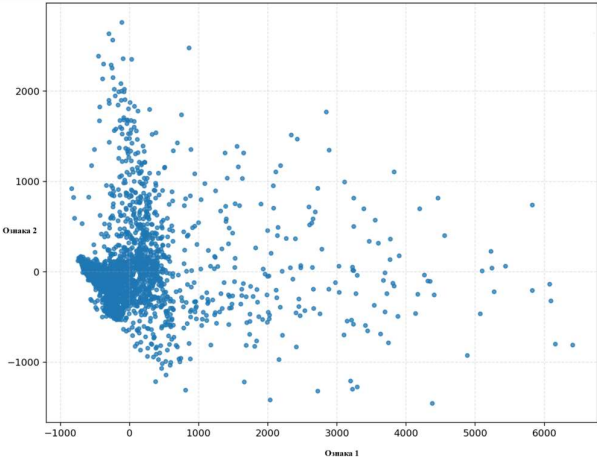
Продовження таблиці 4.2

Підвибірка	Рисунок
Відповідні супутникові знімки	
Негативні супутникові знімки	

Таблиця 4.3. Двовимірна проєкція латентного простору маршруту №62 методом головних компонент для висоти 200 м.

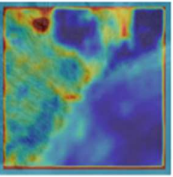
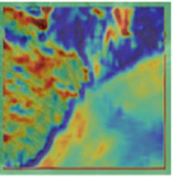
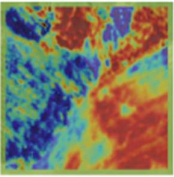
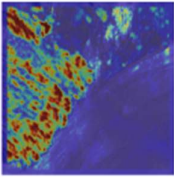
Підвибірка	Рисунок
Дані аерозйомки	

Продовження таблиці 4.3

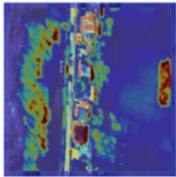
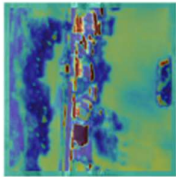
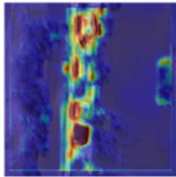
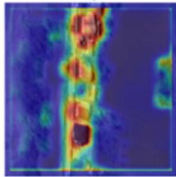
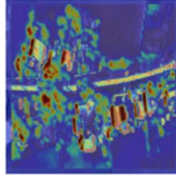
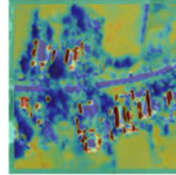
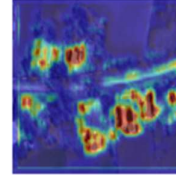
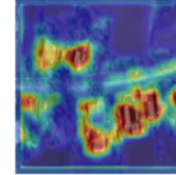
Підвибірка	Рисунок
Відповідні супутникові знімки	
Негативні супутникові знімки	

Для локальної інтерпретації сформованих представлень у табл. 4.4 наведено три репрезентативні позитивні пари та візуалізації проміжних карт ознак енкодера для супутникових зображень і відповідних даних аерозйомки.

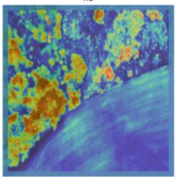
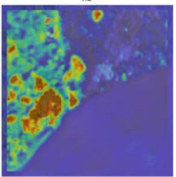
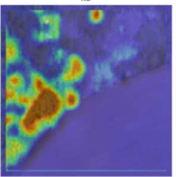
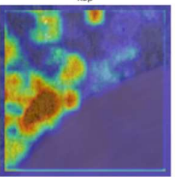
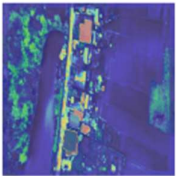
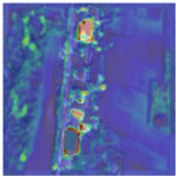
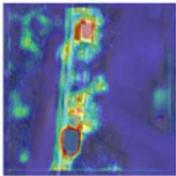
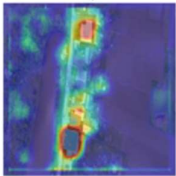
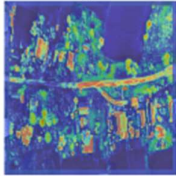
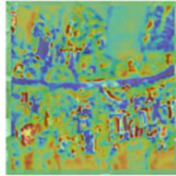
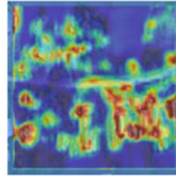
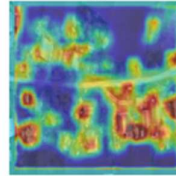
Таблиця 4.4. (а) Візуалізація проміжних карт ознак. Супутникові зображення

№	Рисунок
1.	

Продовження таблиці 4.4. (а)

№	Рисунок
2.	    
3.	    

Таблиця 4.4. (б) Візуалізація проміжних карт ознак. Дані аерозйомки

№	Рисунок
1.	    
2.	    
3.	    

Візуальний аналіз табл. 4.4 показує закономірний перехід від локальних контурів і дрібних текстур на ранніх шарах до більш узагальнених структурних компонент на глибших шарах. На глибших рівнях активації концентруються на стабільних елементах місцевості, зокрема на протяжних лінійних об'єктах і межах однорідних ділянок, що зберігають відтворюваність між аерознімком та відповідним супутниковим фрагментом навіть за відмінностей у кольоровій передачі, деталізації та умовах спостереження. Практичний висновок полягає в тому, що ознаки глибших

шарів є більш придатними для побудови порівняльного простору, оскільки вони менш чутливі до випадкових текстурних варіацій і краще відображають структурну організацію сцени, яка є ключовою для задачі зіставлення аерознімка із супутниковим фрагментом.

Для кількісного аналізу рознесення у латентному просторі для кожного кадру маршруту обчислювали нормовану евклідову відстань між латентним представленням аерознімка та латентними представленнями відповідного позитивного і негативного супутникових фрагментів. Нехай $z_a^{(l)} \in \mathbb{R}^{d_{AE}}$ - латентний вектор l -го аерознімка, $z_{s,+}^{(l)} \in \mathbb{R}^{d_{AE}}$ - латентний вектор відповідного позитивного супутникового фрагмента, а $z_{s,-}^{(l)} \in \mathbb{R}^{d_{AE}}$ - латентний вектор негативного супутникового фрагмента. Тоді відстані визначалися як

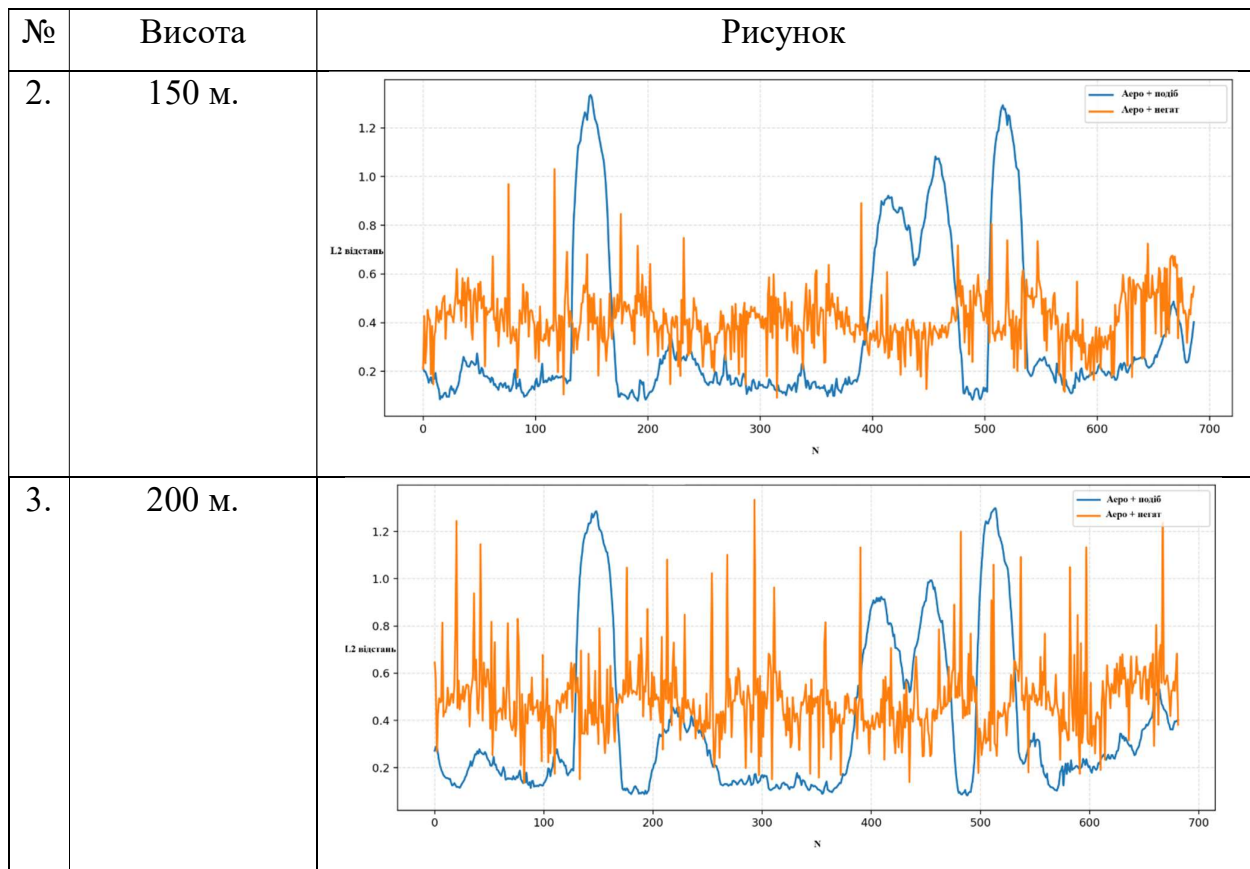
$$d_+^{(l)} = \frac{1}{\sqrt{d_{AE}}} \|z_a^{(l)} - z_{s,+}^{(l)}\|_2, \quad d_-^{(l)} = \frac{1}{\sqrt{d_{AE}}} \|z_a^{(l)} - z_{s,-}^{(l)}\|_2.$$

На основі послідовностей $d_+^{(l)}$ та $d_-^{(l)}$ вздовж маршруту побудовано криві, наведені у табл. 4.5.

Таблиця 4.5. Криві евклідової відстані між аерознімками та відповідними позитивними і негативними супутниковими фрагментами вздовж маршруту

№	Висота	Рисунок
1.	100 м.	

Продовження таблиці 4.5



Аналіз табл. 4.5 показує, що для більшості ділянок маршруту відстань до позитивного супутникового фрагмента залишається нижчою за відстань до негативного. Однак на окремих сегментах спостерігаються локальні піки, у яких ця різниця різко зменшується або зникає. Такі сегменти доцільно трактувати як малоінформативні або неоднозначні кадри, які повинні враховуватися під час подальшого відбору зображень. У прикладному аспекті такі криві є зручним інструментом для налаштування відбору зображень, оскільки дозволяють виявляти проблемні сегменти маршруту та ухвалювати рішення про видалення або пріоритизацію кадрів до виконання повного циклу пошуку відповідника та подальшого зіставлення.

Для статистичної перевірки відмінності між розподілами нормованих евклідових відстаней для позитивних і негативних пар на кожній висоті застосовано двовибірковий критерій Колмогорова–Смирнова [86]. Нехай $F_+^{(h)}(t)$ - емпірична

функція розподілу відстаней для позитивних пар на висоті h , а $F_-^{(h)}(t)$ - емпірична функція розподілу відстаней для негативних пар. Перевірялася нульова гіпотеза

$$H_0^{(h)} : F_+^{(h)}(t) = F_-^{(h)}(t), \quad \forall t,$$

проти альтернативної гіпотези

$$H_1^{(h)} : F_+^{(h)}(t) \neq F_-^{(h)}(t) \quad \text{для деяких } t.$$

Статистика критерію обчислювалась за формулою

$$D_h = \sup_t |F_+^{(h)}(t) - F_-^{(h)}(t)|.$$

Нульову гіпотезу відхиляли, якщо $p_h < \alpha$, де α - рівень значущості. Додатково результат перевіряли шляхом порівняння обчисленої статистики з критичним значенням

$$D_{crit,h}(\alpha) = c_\alpha \sqrt{\frac{n_{+,h} + n_{-,h}}{n_{+,h} n_{-,h}}},$$

де $n_{+,h}$, $n_{-,h}$ - обсяги вибірок позитивних і негативних пар відповідно, $c_{0,05} = 1,36$, $c_{0,01} = 1,63$. У табл. 4.6 наведено описові характеристики розподілів, які використано для інтерпретації зсуву між вибірками.

Таблиця 4.6. Статистичні характеристики розподілів позитивних і негативних пар

№	Висота	Середнє значення	Медіана	Мінімум	Максимум
1 - Позитив	100м	0,3473	0,1913	0,0466	1,3517
1 - Негатив	100м	0,4587	0,4528	0,1311	1,3357
2 - Позитив	150м	0,3476	0,1930	0,0790	1,3353
2 - Негатив	150м	0,4116	0,4030	0,0918	1,0304

Продовження таблиці 4.6

№	Висота	Середнє значення	Медіана	Мінімум	Максимум
3 - Позитив	200	0,3789	0,2365	0,0822	1,2981
3 - Негатив	200	0,4862	0,4730	0,1365	1,3339

Для висоти 100 м отримано $D_{100} \approx 0,31, p_{100} \approx 3 \cdot 10^{-28}, D_{crit,100}(0.05) \approx 0,074$. Для висоти 150м $D_{150} \approx 0,24, p_{150} \approx 4 \cdot 10^{-18}, D_{crit,150}(0.05) \approx 0,073$. Для висоти 200м $D_{200} \approx 0,29, p_{200} \approx 2 \cdot 10^{-25}, D_{crit,200}(0.05) \approx 0,072$. Таким чином, для всіх розглянутих висот виконується нерівність $D_h > D_{crit,h}(0,05)$ та $p_h < 0,01$, тому нульову гіпотезу про однаковість розподілів позитивних і негативних пар відхилено на рівні значущості 1 %. Додатково наведено гістограми розподілів евклідових відстаней для позитивних і негативних пар, що дозволяє наочно оцінити зміщення розподілів і характер їх перекриття, візуалізацію подано на рисунку 4.1.

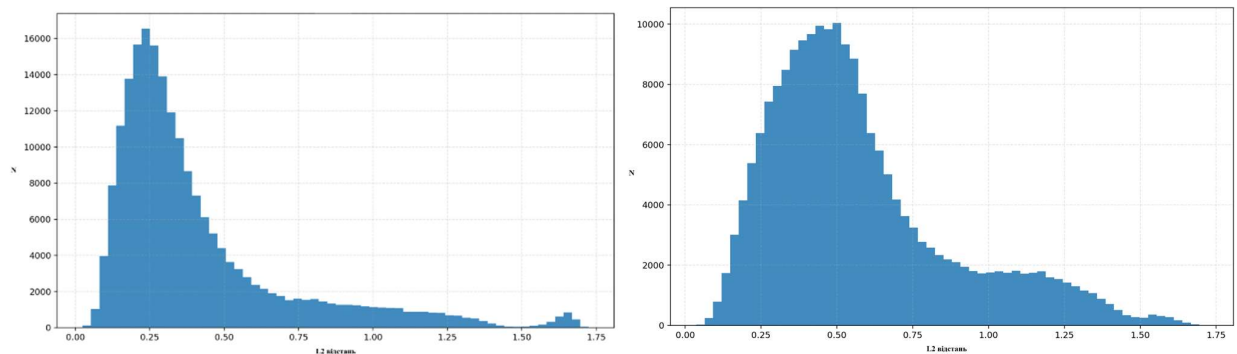


Рис. 4.1. Гістограми евклідових відстаней для усього набору даних. Ліворуч - пари аерознімок та відповідний супутниковий фрагмент. Праворуч аерознімок - негативний приклад

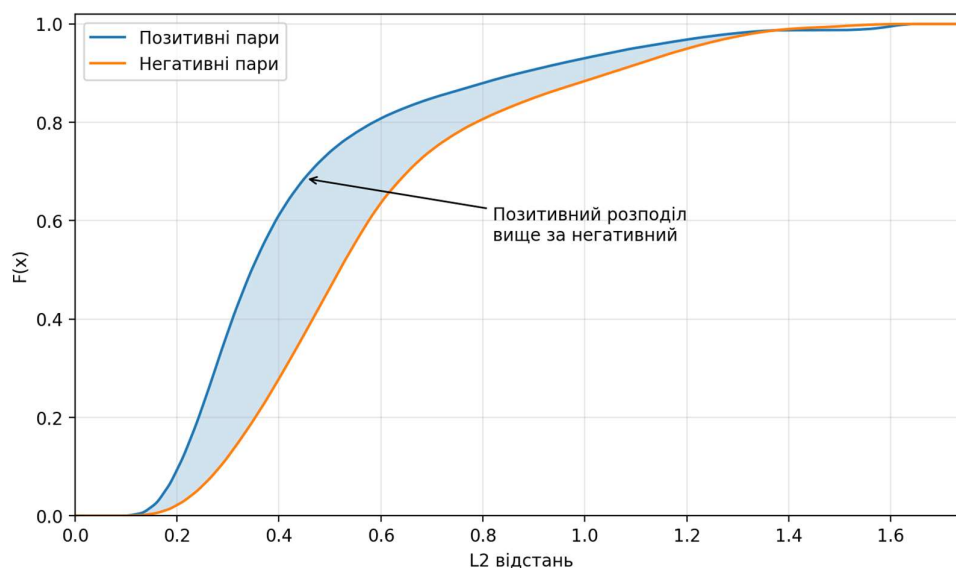
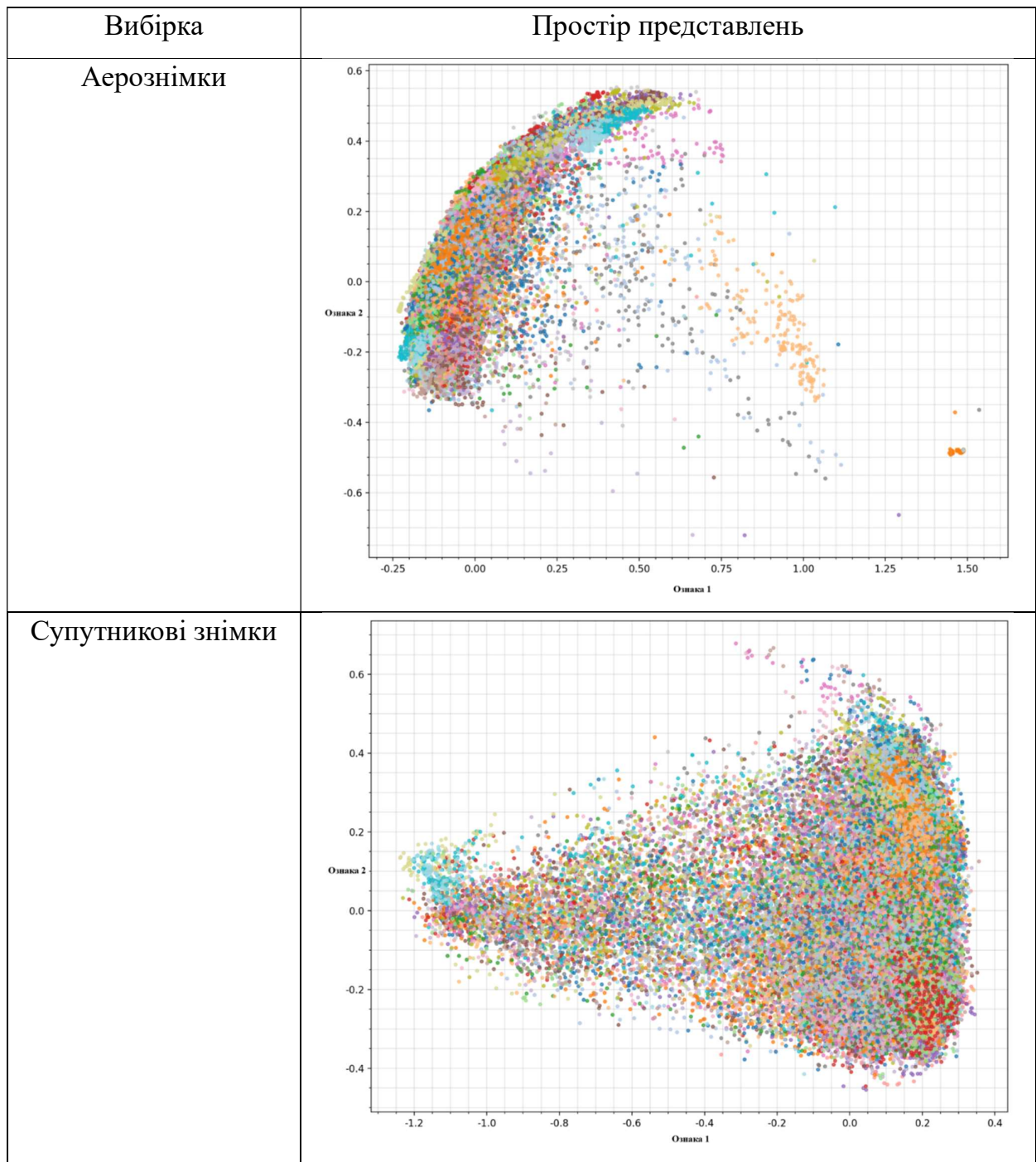


Рис. 4.2. Емпіричні функції розподілу нормованих евклідових відстаней для позитивних і негативних пар

Порівняння емпіричних функцій розподілу на рис. 4.2 показує, що крива $F_+^{(h)}(t)$ для позитивних пар у діапазоні малих і середніх значень t розташована вище за криву $F_-^{(h)}(t)$ для негативних пар. Це означає, що частка позитивних пар із відстанню, не більшою за поріг t , є більшою, ніж для негативних пар, тобто позитивні пари стохастично домінують негативні. Перекриття хвостів розподілів узгоджується з висновками табл. 4.5–4.6 про наявність проблемних кадрів і підтверджує доцільність процедур відбору інформативних зображень як складової загальної технології моделювання та розпізнавання навчальних наборів аерозйомки.

Для оцінювання відмінності між даними аерозйомки та супутниковими прикладами виконано візуалізацію простору представлень для всієї вибірки аерознімків і супутникових знімків методом головних компонент. Результати наведено у табл. 4.7.

Таблиця 4.7. Графік щільності розподілів на всій вибірці



Матеріали табл. 4.7 підтверджують відмінність геометрії розподілів ознак, що відображає різну природу формування зображень, зокрема різницю в масштабі, характері деталізації та умовах спостереження. Подібний підхід до порівняння розподілів різних доменів аерозйомки на основі автоенкодера розглянуто в окремому авторському дослідженні [31]. Разом із тим сам факт наявності

стабільного статистичного розділення позитивних і негативних пар, підтверджений табл. 4.6, узгоджується з тим, що енкодер формує ознаки, придатні для метричного зіставлення та пошуку відповідників між аерознімками і супутниковими фрагментами.

Для переходу від якісного аналізу геометрії розподілів до інтерпретованих кількісних показників на тому самому маршруті використано ті самі три позитивні пари, що наведені у табл. 4.1, до яких додано по одному негативному супутниковому фрагменту, що формує трійки прикладів. Для кожної висоти виконано двовимірне відображення методом головних компонент і обчислено евклідові відстані у координатах МГК-проекції від опорного аерознімка до позитивного та негативного супутникових фрагментів.

Таблиця 4.8. Двовимірна візуалізація трійок та евклідових відстаней для висоти 100 м.

№	Евклідова відстань	Рисунок
1.	До позитиву: 1307 До негативу: 1518	

Продовження таблиці 4.8

№	Евклідова відстань	Рисунок
2.	До позитиву: 1368 До негативу: 1740	
3.	До позитиву: 1000 До негативу: 1074	

Таблиця 4.9. Двовимірна візуалізація трійок та евклідових відстаней для висоти 150 м.

№	Евклідова відстань	Рисунок
1.	До позитиву: 909 До негативу: 1246	

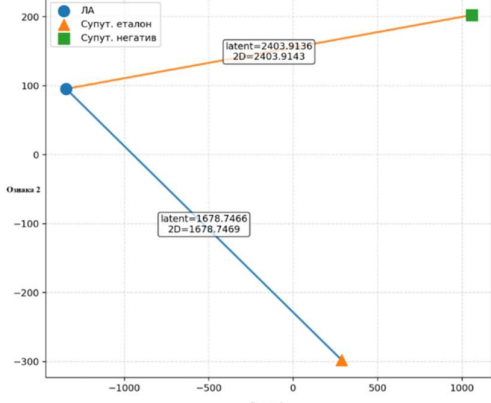
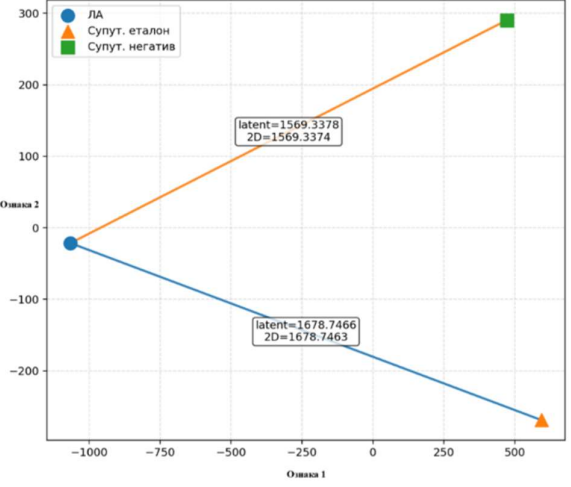
Продовження таблиці 4.9

№	Евклідова відстань	Рисунок
2.	До позитиву: 881 До негативу: 1383	
3.	До позитиву: 764 До негативу: 1260	

Таблиця 4.10. Двовимірна візуалізація трійок та евклідових відстаней для висоти 200 м.

№	Евклідова відстань	Рисунок
1.	До позитиву: 1971 До негативу: 2356	

Продовження таблиці 4.10

№	Евклідова відстань	Рисунок
2.	До позитиву: 1676 До негативу: 2403	
3.	До позитиву: 1678 До негативу: 1569	

За даними табл. 4.8 і табл. 4.9 для висот 100 м. і 150 м. у наведених прикладах виконується очікувана нерівність, тобто позитивний супутниковий фрагмент розташований ближче до відповідного аерознімка, ніж негативний. Це створює запас розділення, який є необхідним для коректного ранжування кандидатів у процедурі пошуку відповідників. Для висоти 200 м. у табл. 4.10 зафіксовано приклад, у якому відстань до негативного супутникового фрагмента є меншою за відстань до еталонного, що означає локальну неоднозначність у просторі ознак. Така поведінка може бути зумовлена повторюваними структурними мотивами місцевості або зменшенням виразності ознак через зміну масштабу сцени. Практичний висновок полягає в тому, що зі зростанням висоти польоту ймовірність

складних випадків зростає, а тому зростає роль відбору інформативних кадрів і коректного формування негативних прикладів.

Окремо досліджено чутливість латентного простору ознак до зміни орієнтації аерознімка. Для висот 100 м., 150 м. та 200 м. аерознімки було повернуто на 45°, 90° та 180°, після чого повторно оцінено відстані.

Таблиця 4.11. Вплив повороту аерознімка на рознесення у латентному просторі, висота 100 м.

№	Евклідова відстань	Кут повороту	Рисунок
1.	До позитиву: 6402 До негативу: 6862	45°	
2.	До позитиву: 7112 До негативу: 7592	90°	

Продовження таблиці 4.11

№	Евклідова відстань	Кут повороту	Рисунок
3.	До позитиву: 1571 До негативу: 1876	180°	

Таблиця 4.12. Вплив повороту аерознімка на рознесення у латентному просторі, висота 150 м.

№	Евклідова відстань	Кут повороту	Рисунок
1.	До позитиву: 6094 До негативу: 5837	45°	
2.	До позитиву: 7439 До негативу: 7194	90°	

Продовження таблиці 4.12

№	Евклідова відстань	Кут повороту	Рисунок
3	До позитиву: 1257 До негативу: 1754	180°	

Таблиця 4.13. Вплив повороту аерознімка на рознесення у латентному просторі, висота 200 м.

№	Евклідова відстань	Кут повороту	Рисунок
1.	До позитиву: 6513 До негативу: 6513	45°	
2.	До позитиву: 7617 До негативу: 7285	90°	

Продовження таблиці 4.13

№	Евклідова відстань	Кут повороту	Рисунок
3.	До позитиву: 950 До негативу: 986	180°	

Результати табл. 4.11–4.13 демонструють, що повороти на 45° та 90° призводять до різкого збільшення відстаней і в окремих випадках погіршують розділення позитивного та негативного супутникових фрагментів настільки, що негативний приклад стає ближчим. Для повороту на 180° у розглянутих прикладах ефект є менш критичним, що узгоджується з частковою симетрією деяких просторових структур сцени. Практичний висновок цього експерименту полягає в тому, що без спеціальних заходів узгодження орієнтації кадрів або без доповнення навчальних даних перетвореннями повороту латентні ознаки не забезпечують стійкої інваріантності до зміни курсу, а це безпосередньо впливає на стабільність зіставлення та якість подальшого просторового визначення положення.

Підрозділ 4.2 Оцінювання якості розпізнавання даних аерозйомки

Експерименти проведено на відкладеній тестовій підвибірці, сформованій згідно з правилами розділення даних, а також під час натурних випробувань на експериментальному стенді, описаному в розділі 3.4. Метою оцінювання є підтвердження того, що навчання моделі забезпечує стійке розділення позитивних і негативних відповідників у просторі ознак та що досягнута якість розпізнавання зберігається за зміни геометрії зйомки, зокрема при варіюванні висоти польоту та кута встановлення оптичного сенсора. Як зазначено в розділі 2.4.3, після

узгодження та фільтрації відповідностей дані подано у вигляді трійок, а розбиття на навчальну, валідаційну та тестову підвибірки виконано груповим принципом за польотами. Такий підхід унеможлиблює потрапляння корельованих кадрів одного польоту до різних частин даних і запобігає завищенню оцінок узагальнювальної здатності моделі.

Для навчання моделі використано 263 701 трійок. Позитивні супутникові відповідники визначалися за геоприв'язкою як фрагменти бази в межах 10 м. у планових координатах. Для уникнення помилкового формування негативних пар усі фрагменти в межах 25 м. від опорної позиції вилучалися з множини кандидатів у негативи, що виключало потрапляння просторово близьких сцен у негативний клас. Під час формування трійок застосовано стратегію добору складних прикладів. Для кожного опорного аерознімка позитивний приклад обирався як найближчий серед допустимих позитивів у просторі глобальних ознак поточної моделі. Негативні приклади добиралися за принципом складних негативів: зі множини просторово допустимих кандидатів формувалася обмежена підмножина випадково відібраних супутникових фрагментів, після чого як негативи обиралися ті, що мають найменшу відстань до опорного зображення в просторі ознак. Добір трійок періодично актуалізувався відповідно до зміни простору представлень у процесі навчання, тоді як валідаційна й тестова підвибірки залишалися фіксованими для коректного контролю узагальнювальної здатності.

Навчання моделі проводилося на нормованих зображеннях розміру 224×224 та виконувалося з використанням оптимізатора Адама [77] з коефіцієнтом навчання $1 \cdot 10^{-5}$ та із параметром відступу $\alpha = 0,1$ у триплетній функції втрат (1.11). Як основу мережі використано попередньо навчену модель DINOv2, параметри якої фіксувалися, а навчання виконувалося переважно для шарів адаптерів. Вихідна розмірність глобального вектора ознак становила 384, а розмір карти локальних ознак дорівнював $61 \times 61 \times 128$.

Оцінювання якості розпізнавання в задачі доцільно формалізувати як задачу ранжування. Нехай множина запитів $Q = \{q_i, i = 1, \dots, N\}$ утворена з кадрів аерозйомки тестової підвибірki, а база еталонів $D = \{d_j\}$ із супутникових фрагментів. Для кожного запиту обчислюється глобальний дескриптор $z(q_i)$ та виконується ранжування еталонів за відстанню $p(z(q_i), z(d_j))$, де p - вибрана метрика подібності у просторі ознак. Позитивним вважається еталон, що задовольняє умову просторової близькості, тобто розташований у межах 10 м. від опорної позиції. Еталони, розташовані на відстані від 10 м. до 25 м., виключаються з оцінювання як неоднозначні. Негативними вважаються еталони, розташовані поза зоною 25 м. Основним показником якості ранжування є повнота пошуку, що визначається як частка запитів, для яких хоча б один позитивний еталон потрапив до перших k позицій ранжованого списку:

$$R@k = \frac{1}{N} \sum_{i=1}^N I\{rank(q_i) \leq k\},$$

де $rank(q_i)$ - позиція першого позитивного еталона для запиту q_i , а I - індикатор події. Додатково під час навчання на фіксованій валідаційній множині пар відстежувалися двійкові метрики, що дозволяють контролювати загальну динаміку розділення позитивних і негативних пар у просторі ознак. Для обчислення точності, повноти та F1-міри використовувався поріг τ_d на значенні відстані, визначений за валідаційною підвибіркою. Відповідні графіки наведено у табл. 4.14.

Таблиця 4.14. Метрики тренування моделі

Метрика	Рисунок
Двійкова повнота	
F1-міра	
Двійкова точність	
Двійковий AUROC	

Для дострокового припинення навчання застосовано критерій відсутності покращення показника $R@5$ на валідаційній підвибірці протягом трьох послідовних епох. У виконаному експерименті навчання завершено після 29 епох, а найкраще значення повноти на валідації досягало 98,9%. Історію середньої втрати

за епоху наведено на рис. 4.3, динаміку $R@1, R@5, R@10, R@20$ на валідації після переранжування на рис. 4.4, а зіставлення найкращого валідаційного стану моделі з підсумковими значеннями на тестовій підвибірці на рис. 4.5.

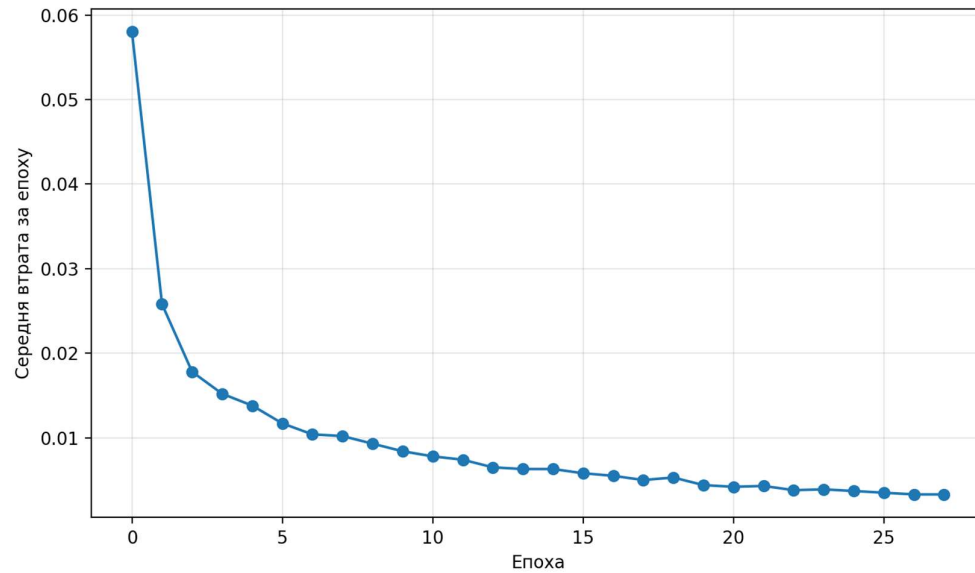


Рис. 4.3. Тренування. Середня втрата за епоху

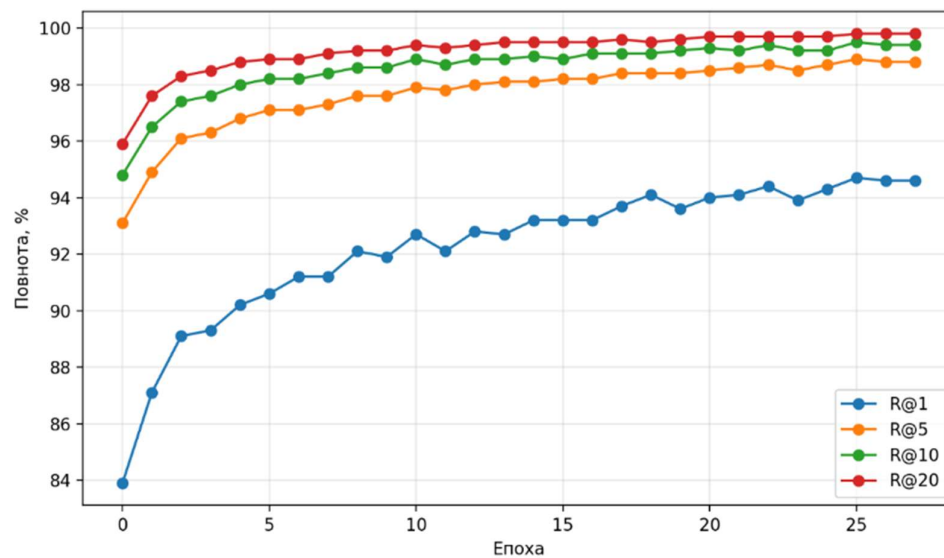


Рис. 4.4. Валідація. Повнота пошуку після переранжування

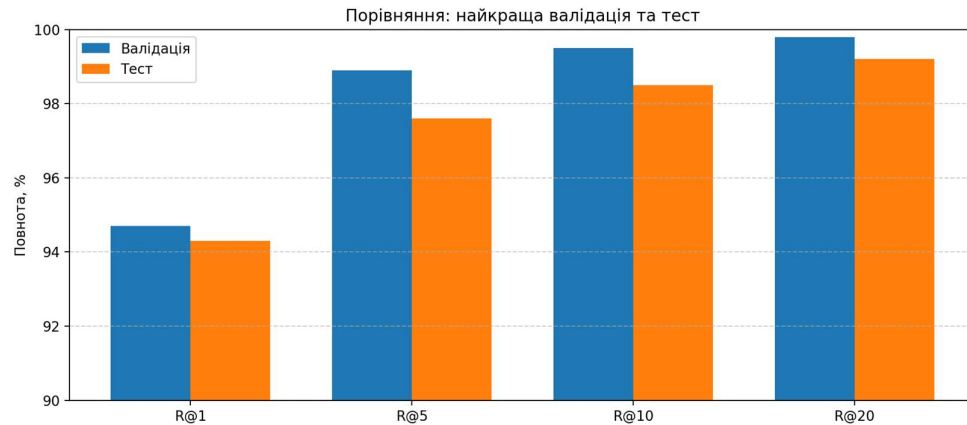


Рис. 4.5. Тестування та валідація

Аналіз кривих навчання свідчить про збіжність оптимізації: середнє значення функції втрат монотонно зменшується на початкових епохах і надалі переходить у режим насичення. Обмежена різниця між найкращими валідаційними значеннями та тестовими значеннями вказує на відсутність вираженого перенавчання та підтверджує здатність моделі узагальнювати на відкладені дані.

Для інтерпретованого аналізу геометрії розділення в просторі ознак додатково розглянуто репрезентативні трійки зображень для висот 100 м., 150 м. і 200 м. Для кожної трійки сформовано глобальні вектори ознак опорного аерознімка, позитивного та негативного супутникових фрагментів нейронною мережею, архітектуру якої описано в розділах 2.2 і 2.3, а програмну реалізацію наведено в розділі 3.3. Для наочної інтерпретації виконано двовимірну проєкцію векторів методом головних компонент і обчислено евклідові відстані між проєкціями опорного аерознімка та його позитивного/негативного супутникових відповідників. Результати наведено у табл. 4.15–4.17.

Таблиця 4.15. Двовимірна візуалізація трійок та евклідових відстаней для висоти 100 м.

№	Евклідова відстань	Рисунок
1.	До позитиву: 0,50 До негативу: 0,87	Scatter plot for height 100m. The x-axis is labeled 'Оznака 1' and ranges from -0.2 to 0.6. The y-axis is labeled 'Оznака 2' and ranges from -0.2 to 0.2. Three points are plotted: LA (blue circle) at approximately (-0.2, 0.25), Sumu. stalov (orange triangle) at approximately (-0.2, -0.25), and Sumu. negativ (green square) at approximately (0.5, 0.0). Lines connect the points, and labels indicate the global and 2D distances: global=0.5038, 2D=0.5038 for the LA-Sumu. stalov pair, and global=0.8769, 2D=0.8769 for the LA-Sumu. negativ pair.
2.	До позитиву: 0,36 До негативу: 0,77	Scatter plot for height 100m. The x-axis is labeled 'Оznака 1' and ranges from -0.2 to 0.5. The y-axis is labeled 'Оznака 2' and ranges from -0.15 to 0.15. Three points are plotted: LA (blue circle) at approximately (-0.2, 0.18), Sumu. stalov (orange triangle) at approximately (-0.2, -0.18), and Sumu. negativ (green square) at approximately (0.5, 0.0). Lines connect the points, and labels indicate the global and 2D distances: global=0.3610, 2D=0.3610 for the LA-Sumu. stalov pair, and global=0.7745, 2D=0.7745 for the LA-Sumu. negativ pair.
3.	До позитиву: 0,36 До негативу: 0,67	Scatter plot for height 100m. The x-axis is labeled 'Оznака 1' and ranges from -0.2 to 0.4. The y-axis is labeled 'Оznака 2' and ranges from -0.15 to 0.20. Three points are plotted: LA (blue circle) at approximately (-0.2, 0.19), Sumu. stalov (orange triangle) at approximately (-0.2, -0.19), and Sumu. negativ (green square) at approximately (0.4, 0.0). Lines connect the points, and labels indicate the global and 2D distances: global=0.3663, 2D=0.3663 for the LA-Sumu. stalov pair, and global=0.6717, 2D=0.6717 for the LA-Sumu. negativ pair.

Таблиця 4.16. Двовимірна візуалізація трійок та евклідових відстаней для висоти 150 м.

№	Евклідова відстань	Рисунок
1.	До позитиву: 0,53 До негативу: 0,88	
2.	До позитиву: 0,34 До негативу: 0,79	
3.	До позитиву: 0,53 До негативу: 0,84	

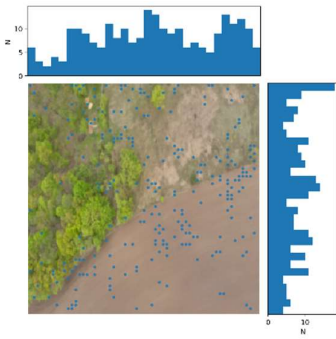
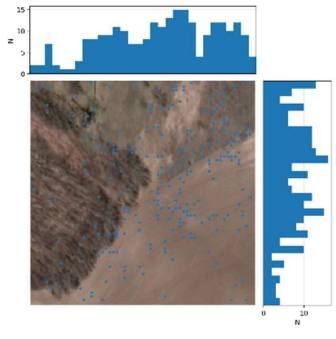
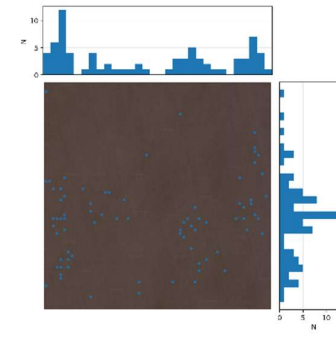
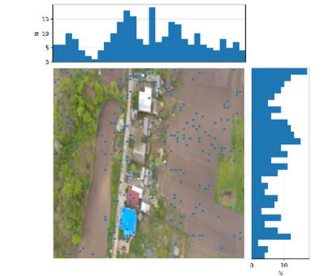
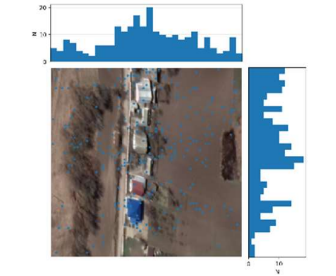
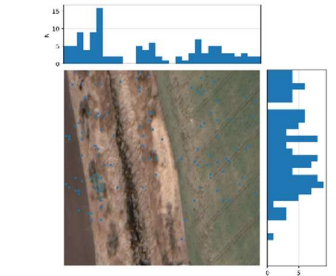
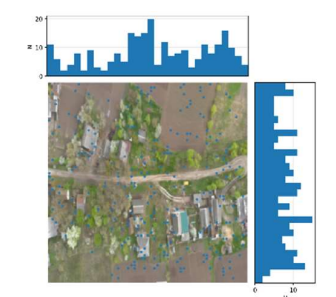
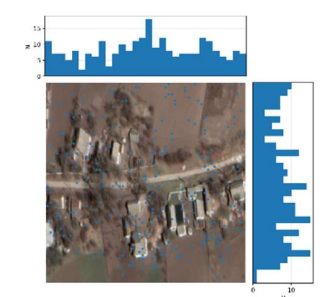
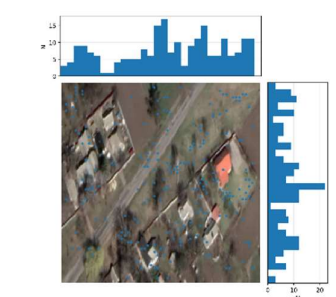
Таблиця 4.17. Двовимірна візуалізація трійок та евклідових відстаней для висоти 200 м.

№	Евклідова відстань	Рисунок
1.	До позитиву: 0,41 До негативу: 0,92	
2.	До позитиву: 0,42 До негативу: 0,82	
3.	До позитиву: 0,45 До негативу: 0,75	

З отриманих значень випливає, що в усіх прикладах виконується очікувана нерівність: позитивний супутниковий знімок розташований ближче до опорного аерознімка, ніж негативний. У середньому для висоти 100 м. відстань позитивної пари становить близько 0,41, тоді як відстань до негативу близько 0,77; для висоти 150 м. близько 0,47 та 0,84 відповідно; для висоти 200 м. близько 0,43 та 0,8. Це означає, що за зміни висоти польоту зберігається узгоджена структура рознесення, а сформовані глобальні ознаки забезпечують коректне ранжування кандидатів і зменшують імовірність хибного зближення з негативним супутниковим фрагментом.

Для інтерпретації дворівневого подання додатково проаналізовано локальні ознаки, сформовані мережею для тих самих трійок, що використовувалися під час аналізу глобальних представлень. Локальні ознаки подано у вигляді множини ключових точок, що відповідають найбільш інформативним ділянкам сцени та застосовуються для уточнення відповідності між аерознімком і кандидатними супутниковими фрагментами. Візуалізацію виконано окремо для пари «аерознімок - позитивний супутниковий фрагмент» та для пари «аерознімок - негативний супутниковий фрагмент» на трьох висотах польоту. Приклади наведено у табл. 4.18–4.20. Також для кількісного аналізу локальних ознак використано маргінальні розподіли нормованих координат ключових точок за горизонталлю $\bar{u} = u / W$ та вертикаллю $\bar{v} = v / H$. Маси розподілів обчислено як відносні частоти $m_b^u = N_b^u / n_x$, де N_b^u та N_b^v задають кількість ключових точок у відповідному інтервалі гістограми, а n_x є загальною кількістю локальних ознак на зображенні. Перевірено нульову гіпотезу H_0 про збіг емпіричних розподілів координат локальних ознак для аерознімка та супутникового фрагмента. Альтернативна гіпотеза H_1 полягала у відмінності цих розподілів. Для перевірки використано двовибірковий критерій Колмогорова-Смирнова [86] окремо для горизонтальної та вертикальної координат.

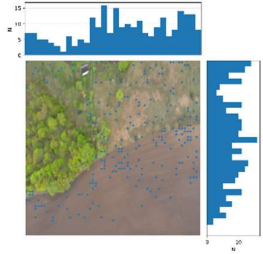
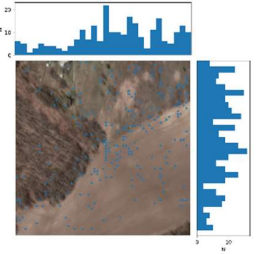
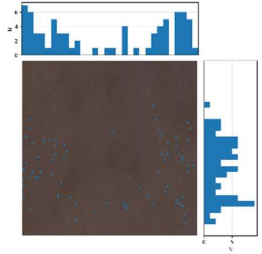
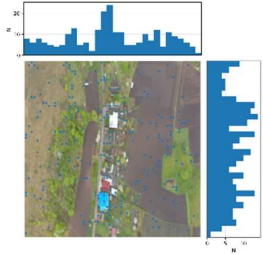
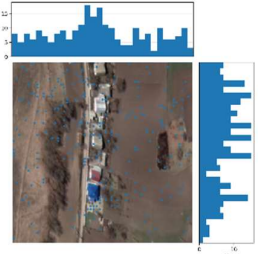
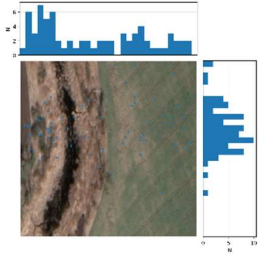
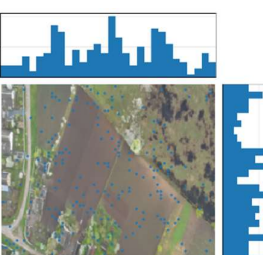
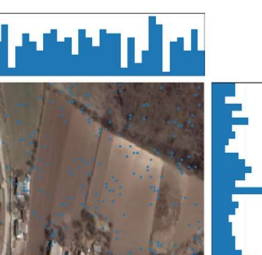
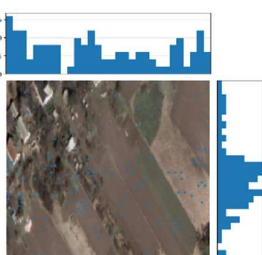
Таблиця 4.18. Візуалізація локальних ознак зображень 100 м.

№	Аерознімок	Позитивний супутниковий фрагмент	Негативний супутниковий фрагмент
1.			
2.			
3.			

Для табл. 4.18, що відповідає висоті 100 м, для позитивних пар отримано $D_u^+ \approx 0,11, p_u^+ \approx 0,10, D_v^+ \approx 0,09, p_v^+ \approx 0,18$. Нульову гіпотезу не відхилено на рівні значущості 0,05. Для негативних пар отримано $D_u^- \approx 0,31, p_u^- < 10^{-3}, D_v^- \approx 0,34, p_v^- < 10^{-3}$. Нульову гіпотезу відхилено на рівні значущості 0,01. Для позитивних пар гістограми мають близькі широкі розподіли без різкого зміщення максимумів. Для негативних пар спостерігаються асиметрія, крайові піки за

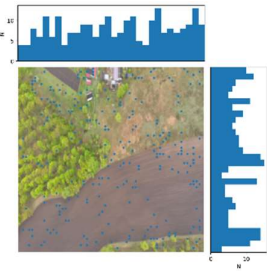
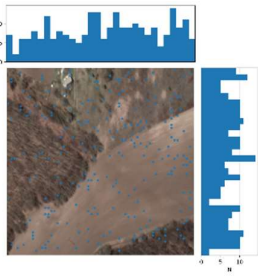
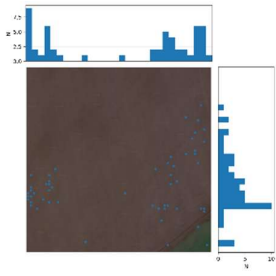
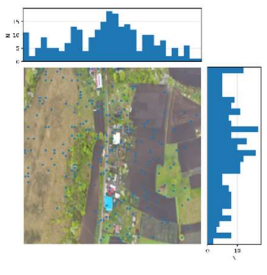
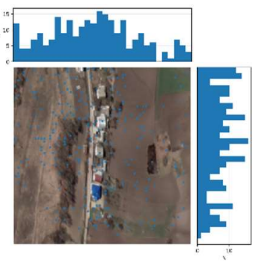
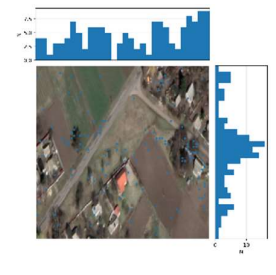
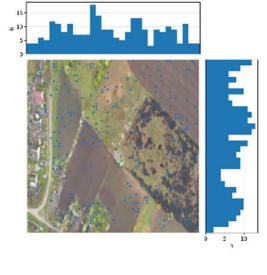
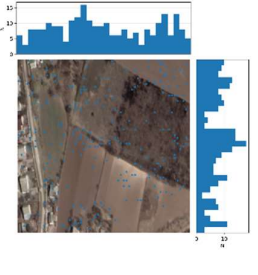
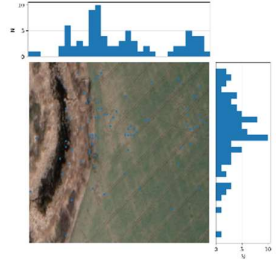
горизонтальною координатою та зміщення основної маси за вертикальною координатою.

Таблиця 4.19. Візуалізація локальних ознак зображень 150 м.

№	Аерознімок	Позитивний супутниковий фрагмент	Негативний супутниковий фрагмент
1.			
2.			
3.			

Для табл. 4.19, що відповідає висоті 150 м, для позитивних пар отримано $D_u^+ \approx 0,12, p_u^+ \approx 0,08, D_v^+ \approx 0,10, p_v^+ \approx 0,13$. Нульову гіпотезу не відхилено на рівні значущості 0,05. Для негативних пар отримано $D_u^- \approx 0,28, p_u^- \approx 0,002, D_v^- \approx 0,26, p_v^- \approx 0,004$. Нульову гіпотезу відхилено на рівні значущості 0,01. Позитивні розподіли залишаються узгодженими та більш гладкими, тоді як негативні є багатомодальними і мають розсинхронізовані положення максимумів.

Таблиця 4.20. Візуалізація локальних ознак зображень 200 м.

№	Аерознімок	Позитивний супутниковий фрагмент	Негативний супутниковий фрагмент
1.			
2.			
3.			

Для табл. 4.20, що відповідає висоті 200 м, для позитивних пар отримано $D_u^+ \approx 0,13, p_u^+ \approx 0,07, D_v^+ \approx 0,11, p_v^+ \approx 0,12$. Нульову гіпотезу не відхилено на рівні значущості 0,05. Для негативних пар отримано $D_u^- \approx 0,24, p_u^- \approx 0,006, D_v^- \approx 0,22, p_v^- \approx 0,009$. Нульову гіпотезу відхилено на рівні значущості 0,01. На цій висоті позитивні розподіли дещо вирівнюються, але зберігають узгодженість, тоді як негативні пари залишаються статистично відмінними від аерознімків.

Отже, для всіх трьох висот статистично підтверджено узгодженість маргінальних розподілів координат локальних ознак у позитивних парах та статистично значущу

відмінність цих розподілів у негативних парах. Це підтверджує, що локальний рівень моделі зберігає просторову структуру правильного відповідника і забезпечує відсікання хибних супутникових фрагментів.

Натурні випробування виконуються на експериментальному стенді, що описаний у розділі 3.4 із використанням високоточного RTK-позиціювання як еталона. У цьому режимі якість локалізації доцільно характеризувати просторовою похибкою між оціненою позицією $\hat{p}_i$ та еталонною позицією p_i для кожного моменту часу. Підсумкова статистика похибок розпізнавання та локалізації наведено у таблицях 4.21-4.23 та на рисунках 4.6-4.8.

Таблиця 4.21. Статистика просторової похибки для проведеного польоту №1

Набір	К-сть	Частка	Середня похибка,м.	СКП,м.	СВ,м.	P95,м.	Мін.,м.	Макс,м.
Усі точки	213	100,0	8,4	12,67	9,51	33,87	2,04	33,97
Без аномалій	181	84,98	4,57	4,78	1,4	6,79	2,04	6,98
Аномалії	32	15,02	30,05	30,64	6,08	37,81	20,09	33,97

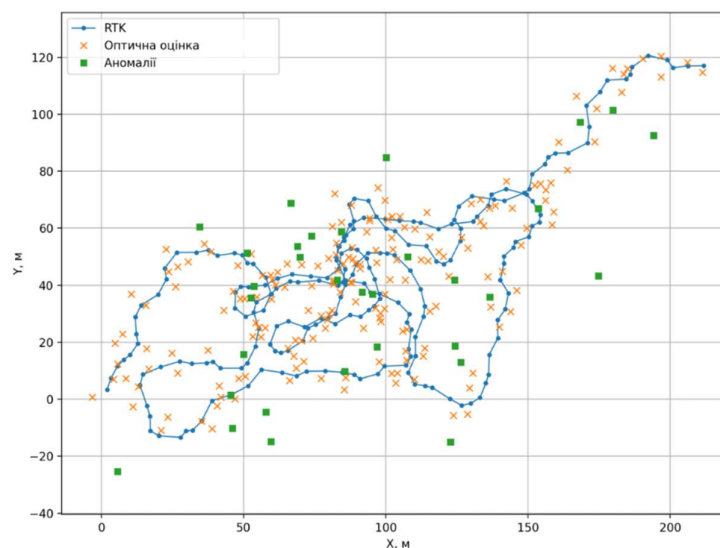


Рис. 4.6. Ілюстрація проведеного польоту №1

За результатами польоту № 1 встановлено, що оцінки положення, отримані за оптичним каналом, у більшості випадків узгоджуються з еталонною RTK-траєкторією та формують компактну хмару відхилень відносно опорної лінії руху. Для всієї сукупності із 213 точок середня просторова похибка становить 8,4 м. при СКП 12,67 м., що вказує на наявність важкого хвоста розподілу похибок. Отже, після відсіювання аномальних оцінок базова якість локалізації в штатних умовах відповідає середній просторовій похибці на рівні близько 4,5–4,6 м., що узгоджується з даними рядка “Без аномалій” у табл. 4.21–4.23., що на рис. 4.6 проявляються як просторово віддалені викиди.

Таблиця 4.22. Статистика просторової похибки для проведеного польоту №2

Набір	К-сть	Частка	Середня похибка,м.	СКП,м.	СВ,м.	P95,м.	Мін.,м.	Макс,м.
Усі точки	224	100,0	8,3	12,58	9,47	32,98	2,0	38,97
Без аномалій	190	84,83	4,45	4,66	1,38	6,7	2,0	7,0
Аномалії	34	15,17	29,82	30,36	5,78	38,25	21,0	38,97

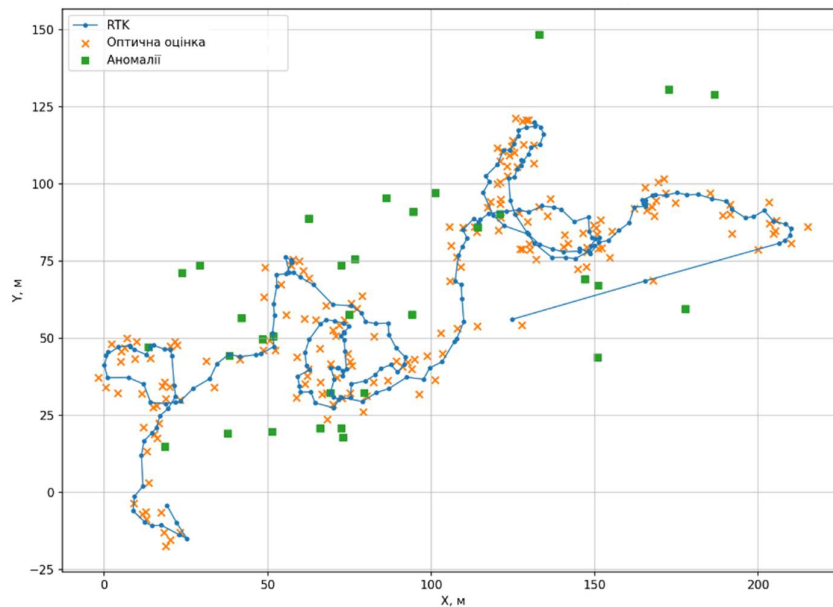


Рис. 4.7. Ілюстрація проведеного польоту №2

За результатами польоту № 2 отримано чисельно близькі результати, що підтверджує відтворюваність якості локалізації за зміни траєкторії та умов зйомки. Для всієї вибірки з 224 точок середня похибка становить 8,3 м, тоді як підмножина без аномалій характеризується малою дисперсією похибки та значеннями 4,45 м., а 95-й перцентиль не перевищує 6,7 м. Таким чином, у переважній частині часу система забезпечує метричну точність, достатню для задач грубої абсолютної прив'язки.

Таблиця 4.23. Статистика просторової похибки для проведеного польоту №3

Набір	К-сть	Частка	Середня похибка,м.	СКП,м.	СВ,м.	P95,м.	Мін.,м.	Макс,м.
Усі точки	247	100,0	8,27	12,53	9,34	33,87	2,0	38,9
Без аномалій	210	85,02	4,47	4,67	1,34	6,56	2,0	6,92
Аномалії	37	14,98	29,81	30,4	6,05	38,0	20,0	38,9

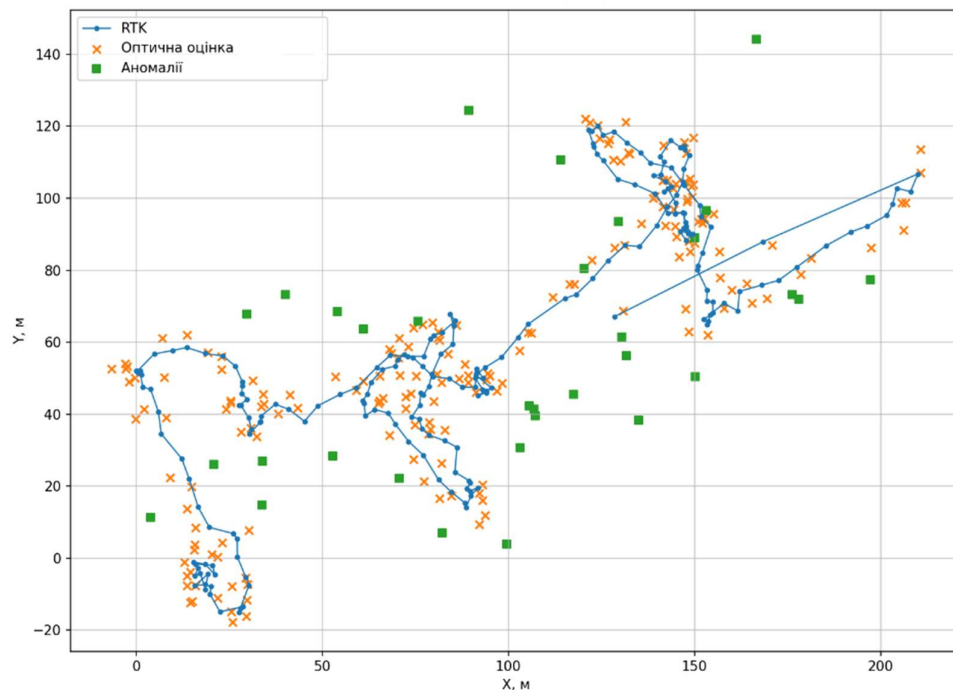


Рис. 4.8. Ілюстрація проведеного польоту №3

За результатами польоту № 3 підтверджено сталість розподілу похибок: для 247 точок середня похибка становить 8,27 м. при СКП 12,53 м. Частка аномалій становить 14,98% із характерними відхиленнями на рівні 30 м. та максимальним значенням до 38,9 м; на рис. 4.8 такі оцінки візуально виокремлюються як викиди, що не відповідають геометрії основного маршруту. Узагальнюючи результати трьох натурних експериментів, можна зробити висновок, що запропонований підхід оптичної локалізації забезпечує стабільну просторову точність порядку 4-5 м. для 85% кадрів і демонструє відтворюваність характеристик між польотами.

Наведені в підрозділі 4.2 результати характеризують робочий діапазон висот 100–200 м., для якого сформовано основну тестову вибірку та виконано натурні випробування. Для малих висот 20–60 м., за яких зменшується огляд сцени та посилюються перспективні спотворення, якість ранжування та просторового визначення положення помітно погіршується. Відповідне порівняння для малих висот наведено в підрозділі 4.3.

Підрозділ 4.3 Порівняльний аналіз з існуючими методами

Порівняльний аналіз запропонованого підходу виконано на власній тестовій підвибірці, сформованій із кадрів аерозйомки та відповідних супутникових фрагментів. Оскільки безпосереднє зіставлення з опублікованими результатами на зовнішніх наборах даних є методично некоректним за відмінних умов зйомки, висот, роздільної здатності та структури еталонної бази, у цьому розділі всі методи оцінювалися в однакових умовах на єдиному тестовому наборі. Як основний показник точності використано повноту пошуку $R@k$, визначену у розділі 4.2, що характеризує частку запитів, для яких правильний еталон потрапляє до перших k позицій ранжованого списку. Результати порівняння за показником повноти пошуку наведено в таблиці 4.24. До порівняння залучено сучасні моделі глобального VPR, зокрема MixVPR, AnyLoc, SALAD, NetVLAD та варіанти SelaVPR [8, 10, 14, 59, 60]. Попередні авторські етапи розвитку підходу до перехресного розпізнавання супутникових і аероданих та до використання багаторівневих ознак наведено в окремих публікаціях [39].

Таблиця 4.24. Порівняння точності на власній вибірці за метрикою повноти пошуку $R@k$

Метод	Чи навчена модель на даних аерозйомки	$R@1$, %	$R@5$, %	$R@10$, %	$R@20$, %
SelaVPR+	Так	98,4	99,5	99,7	99,9
SelaVPR	Так	97,3	99,0	99,4	99,7
SelaVPR	Ні	90,1	95,8	97,1	98,0
AnyLoc	Ні	88,0	94,5	96,0	97,2
MixVPR	Ні	86,5	93,2	95,0	96,5
SALAD	Ні	83,0	91,0	93,6	95,2
NetVLAD	Ні	80,5	90,2	93,0	95,0

Для запропонованої моделі, додатково навченої на власних даних аерозйомки, зафіксовано найвищі значення: для варіанта SelaVPR повнота $R@1$ становить 97,3 %, тоді як для варіанта без додаткового навчання - 90,1 %. Отже, приріст $R@1$ дорівнює 7,2 відсоткового пункту, що підтверджує визначальний вплив адаптації моделі до конкретних умов спостереження. Модифікація SelaVPR+ демонструє подальше покращення: $R@1$ зростає до 98,4 %, а також зберігається високий рівень $R@5-R@20$, що зменшує ймовірність потреби у поглиблених процедурах переранжування. Методи без адаптації до власних даних демонструють нижчі значення $R@1$, 80,5–88,0 %, що вказує на наявність зсуву розподілів між даними, на яких формувалися їхні ознаки, та цільовими даними аерозйомки і супутникових еталонів у цьому дослідженні.

Окрім точності, для практичного бортового застосування визначальним є час оброблення одного запиту, який включає формування глобального дескриптора та виконання пошуку у векторному сховищі. Порівняльні результати за часовими витратами наведено в таблиці 4.25 для одноплатних платформ Orange Pi 5, Raspberry Pi 5 та Jetson Orin NX.

Таблиця 4.25. Середній час обробки одного запиту на одноплатних комп'ютерах

Метод	Чи навчена модель на даних аерозйомки	Orange PI5, с	Raspberry PI 5, с	Jetson Orin NX, с
SelaVPR+	Так	0,28	0,35	0,06
SelaVPR	Так	0,30	0,38	0,07
SelaVPR	Hi	0,30	0,38	0,07
MixVPR	Hi	0,36	0,45	0,08
SALAD	Hi	0,40	0,5	0,09
NetVLAD	Hi	0,75	0,95	0,18
NetVLAD + SuperPoint + SuperGlue	Hi	2,6	3,4	0,53

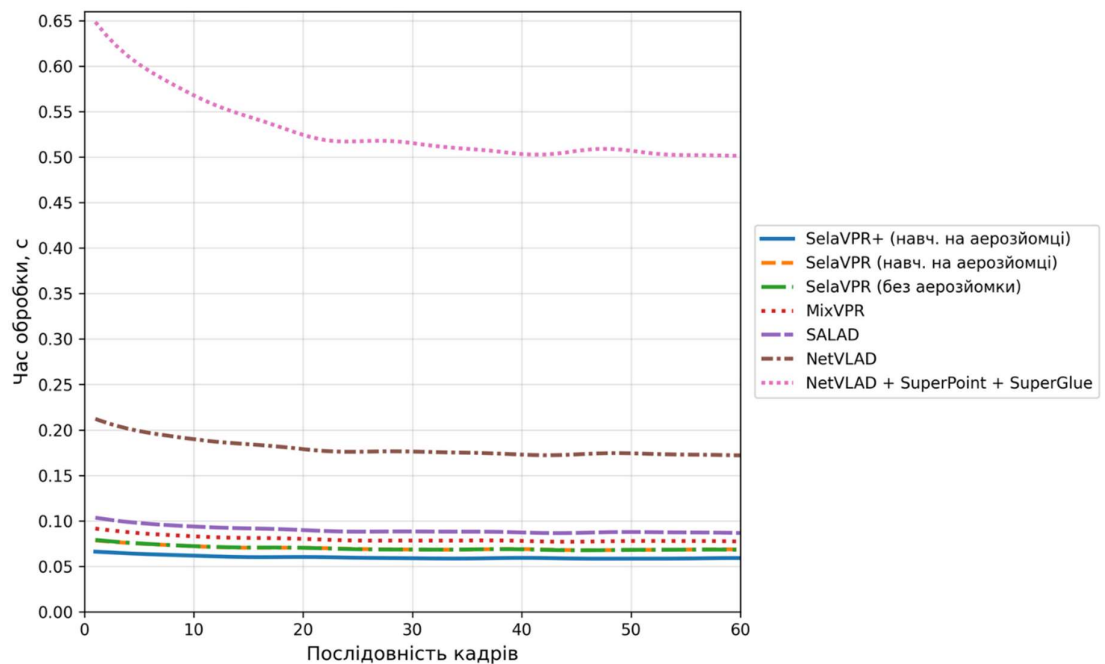


Рис. 4.9. Час обробки послідовності на Jetson Orin NX для наведених моделей

Запропоновані варіанти SelaVPR і SelaVPR+ забезпечують малу затримку оброблення. На Jetson Orin NX середній час запиту становить близько 0,07 с. та 0,06 с. відповідно, що відповідає частоті оброблення приблизно 14–16 кадрів за секунду й є достатнім для режиму роботи, наближеного до реального часу. На Orange Pi 5 та Raspberry Pi 5 середній час запиту перебуває в межах 0,28–0,38 с, що еквівалентно приблизно 2–4 кадрам за секунду; така швидкодія залишається практично прийнятною для навігаційної підтримки за умови зменшення частоти оброблення кадрів. Водночас комбіновані схеми з локальною перевіркою відповідностей істотно збільшують обчислювальні витрати: на одноплатних платформах час запиту зростає до 2,6–3,4 с, що фактично переводить систему в режим менш як один кадр за секунду та обмежує її використання без потужного апаратного прискорювача.

Додаткове узагальнення переваг удосконаленого варіанта SelaVPR+ подано в таблиці 4.26. Порівняно з базовим SelaVPR, модифікація забезпечує одночасне зростання точності та скорочення часу оброблення на Jetson Orin NX: $R@1$ підвищується з 97,3 % до 98,4 %, а час запиту зменшується з 0,070 с до 0,064 с в середньому, що підвищує частоту оброблення з 14,3 до 15,6 кадрів за секунду. Такий ефект є практично значущим, оскільки збільшує частку випадків коректного визначення відповідника з першої позиції та водночас зменшує затримку в навігаційному контурі.

Таблиця 4.26. Порівняння SelaVPR та SelaVPR+ на власній вибірці та на Jetson Orin NX

Показник	SelaVPR	SelaVPR+	Різниця
Точність $R@1$	97,3%	98,4%	1,1%
Точність $R@5$	99,0%	99,5%	0,5%

Продовження таблиці 4.26

Показник	SelaVPR	SelaVPR+	Різниця
Середній час обробки на Jetson Orin NX	0,070 с.	0,064 с.	-0,006 с.
FPS	14,3	15,6	1,3

Висновок до таблиці 4.26. SelaVPR+ забезпечує одночасно покращення точності та зменшення часу обробки. Приріст $R@1$ на власній вибірці зменшує потребу у глибокому переранжуванні та збільшує частку випадків коректного розпізнавання.

Додатково оцінено вплив висоти польоту на якість локалізації за метрикою повноти пошуку $R@k$ на власній тестовій вибірці. Розглянуто три висоти 20 м., 40 м. і 60 м. та два режими орієнтації камери. У режимі, коли камера орієнтована перпендикулярно до земної поверхні і дивиться в горизонт. Зі збільшенням висоти зростає інформативність сцени за рахунок більшого охоплення місцевості в кадрі, що приводить до підвищення $R@k$. При зменшенні висоти до 20 м. відбувається різке падіння точності через втрату контексту, зростання неоднозначності текстур і суттєвіший вплив локальних змін на зображенні, унаслідок чого $R@1$ для універсальних моделей може знижуватися до одиниць або десятків відсотків. У режимі, коли камера дивиться вздовж горизонту, відповідність із двовимірною супутниковою картою погіршується ще сильніше через перспективні спотворення. У цьому режимі для висоти 20 м. значення $R@1$ для більшості мереж знижуються до діапазону близько 1–10%.

Таблиця 4.27. Повнота пошуку для камери, спрямованої вниз, на висоті 20 м.

Метод	Чи навчена модель на даних аерозйомки	$R@1$, %	$R@5$, %	$R@10$, %	$R@20$, %
SelaVPR+	Так	25,0	50,0	58,0	68,0

Продовження таблиці 4.27

Метод	Чи навчена модель на даних аерозйомки	R@1, %	R@5, %	R@10, %	R@20, %
SelaVPR	Hi	12,0	25,0	32,0	40,0
AnyLoc	Hi	10,0	22,0	28,0	35,0
MixVPR	Hi	8,0	18,0	24,0	31,0
SALAD	Hi	6,0	14,0	19,0	25,0
NetVLAD	Hi	5,0	12,0	16,0	22,0

Таблиця 4.28. Повнота пошуку для камери, спрямованої вниз, на висоті 40 м.

Метод	Чи навчена модель на даних аерозйомки	R@1, %	R@5, %	R@10, %	R@20, %
SelaVPR+	Так	93,0	97,8	98,6	99,2
SelaVPR	Hi	60,0	78,0	84,0	90,0
AnyLoc	Hi	55,0	74,0	81,0	88,0
MixVPR	Hi	52,0	71,0	79,0	86,0
SALAD	Hi	45,0	64,0	73,0	81,0
NetVLAD	Hi	40,0	58,0	67,0	76,0

Таблиця 4.29. Повнота пошуку для камери, спрямованої вниз, на висоті 60 м.

Метод	Чи навчена модель на даних аерозйомки	R@1, %	R@5, %	R@10, %	R@20, %
SelaVPR+	Так	97,0	98,5	99,1	99,5
SelaVPR	Hi	88,0	95,0	96,8	97,5
AnyLoc	Hi	85,0	94,0	96,5	97,7
MixVPR	Hi	82,0	93,0	96,0	97,5
SALAD	Hi	78,0	90,0	94,0	96,5
NetVLAD	Hi	75,0	88,0	92,0	95,0

Для навченої на власних даних моделі SelaVPR зберігається високий рівень точності на 60 м., тоді як на 20 м. відбувається різке падіння $R@1$ через дефіцит контексту в кадрі.

Таблиця 4.30. Повнота пошуку для камери, спрямованої в горизонтальному напрямку, на висоті 20 м.

Метод	Чи навчена модель на даних аерозйомки	$R@1$, %	$R@5$, %	$R@10$, %	$R@20$, %
SelaVPR+	Так	8,0	18,0	25,0	35,0
SelaVPR	Ні	3,0	7,0	10,0	15,0
AnyLoc	Ні	3,0	7,0	10,0	14,0
MixVPR	Ні	2,0	5,0	8,0	12,0
SALAD	Ні	1,0	3,0	5,0	9,0
NetVLAD	Ні	1,0	3,0	5,0	8,0

Таблиця 4.31. Повнота пошуку для камери, спрямованої в горизонтальному напрямку, на висоті 40 м.

Метод	Чи навчена модель на даних аерозйомки	$R@1$, %	$R@5$, %	$R@10$, %	$R@20$, %
SelaVPR+	Так	25,0	45,0	55,0	65,0
SelaVPR	Ні	10,0	20,0	28,0	38,0
AnyLoc	Ні	9,0	18,0	25,0	34,0
MixVPR	Ні	8,0	16,0	23,0	31,0
SALAD	Ні	6,0	12,0	18,0	25,0
NetVLAD	Ні	5,0	10,0	15,0	20,0

Таблиця 4.32. Повнота пошуку для камери, спрямованої в горизонтальному напрямку, на висоті 60 м.

Метод	Чи навчена модель на даних аерозйомки	R@1, %	R@5, %	R@10, %	R@20, %
SelaVPR+	Так	45,0	70,0	78,0	85,0
SelaVPR	Ні	20,0	40,0	50,0	62,0
AnyLoc	Ні	18,0	36,0	46,0	58,0
MixVPR	Ні	16,0	33,0	43,0	55,0
SALAD	Ні	12,0	26,0	36,0	48,0
NetVLAD	Ні	10,0	22,0	30,0	40,0

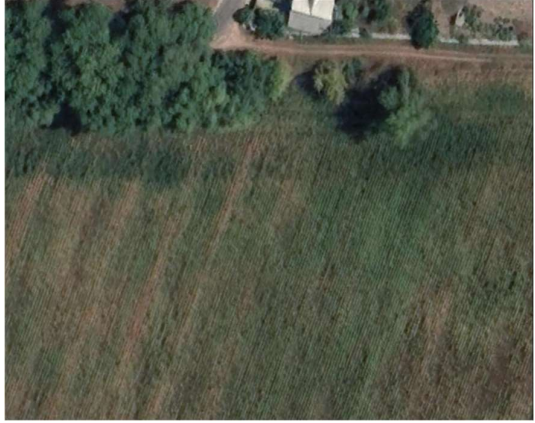
Для камери, що дивиться в горизонт, умови зіставлення з супутниковим еталоном є найбільш несприятливими, а зменшення висоти до 20 м. призводить до критичного падіння точності. Для більшості моделей R@1 знаходиться в межах 1–10%, що означає практичну неможливість стабільної локалізації без додаткової геометричної компенсації. На основі отриманих результатів запропоновано покращення якості розпізнавання на низьких висотах, що описано у наступному розділі.

Підрозділ 4.4 Метод побудови тривимірного представлення супутникових сцен та оцінювання його впливу на якість розпізнавання

Побудова тривимірного представлення супутникових сцен у межах запропонованої технології використовується для практичного формування висотної поверхні з текстурою на основі двовимірних супутникових знімків. Результати цього розділу спираються на апробовані рішення, викладені у публікації [30] та [33]. Вхідними даними модуля є впорядкована послідовність супутникових фрагментів уздовж маршруту $\{X_s^j, j = 1, \dots, N\}$, де $X_s^j \in \mathbf{X}_s \subset \mathbb{R}^{H \times W \times 3}$ відповідає ортофото ділянки місцевості. Для кожного фрагмента задано метадані центру та орієнтації, які надалі використовуються для початкового просторового узгодження:

географічні координати (φ_j, λ_j) , висота h_j і курс ψ_j . Ілюстративні приклади проміжних результатів наведено у табл. 4.33.

Таблиця 4.33. Приклади отриманих тривимірних зображень

№	Двовимірний супутниковий знімок	Тривимірна проєкція
1.		
2.		
3.		

Продовження таблиці 4.33

№	Двовимірний супутниковий знімок	Тривимірна проєкція
4.		

Ключовою операцією є монокулярна оцінка глибини супутникових зображень. Нехай Φ є нейромережовим оператором оцінювання глибини на супутниковому домені. Тоді для кожного фрагмента обчислюється карта глибини

$$D_j = \Phi(X_s^j), \quad D_j \in \mathbb{R}^{H \times W}.$$

Оскільки більшість картографічних сервісів не надає еталонних висотних карт для валідації, вибір архітектури Φ та цільова адаптація до супутникових сцен виконувалися на основі експериментального порівняння сучасних моделей [17, 18, 33]. Це забезпечує відтворюваність отримання глибинних карт для ортофото та стабільність контурів і локальної структури сцени у межах маршруту.

Вихід D_j не слід інтерпретувати як повністю метричну цифрову модель рельєфу, оскільки монокулярне відновлення глибини містить невизначеність масштабу та зсуву. Для узгодження динамічного діапазону між різними сценами вводиться фіксоване нормування N , яке застосовується однаково для всіх фрагментів одного джерела даних, після чого формується висотна поверхня

$$Z_j = N(D_j), \quad Z_j \in \mathbb{R}^{H \times W}.$$

У такому поданні карта Z_j задає відносну вертикальну структуру ділянки, а значення x_s^j використовуються як текстурні атрибути для візуалізації та подальшого аналізу.

Для переходу від матричного подання Z_j до тривимірної геометрії кожному пікселю (u, v) ставиться у відповідність точка локальної поверхні. Планові координати переводяться в метричну шкалу за величиною розрізнення супутникового шару. Нехай $g_j = S_{pix}(z, \varphi_j)$ є розмір пікселя в метрах на широті φ_j для обраного рівня масштабування. Тоді тривимірне подання пікселя можна задати як

$$p_j(u, v) = (g_j(u - u_0), g_j(v - v_0), Z_j(u, v)),$$

де $(u - u_0)$ відповідає центру зображення. Це визначає висотну поверхню з однозначною прив'язкою до текстури та з відтворюваною плановою шкалою.

Окремо враховується узгодженість вертикальної складової. Підхід забезпечує практично необхідну однорідність висот у межах маршруту за рахунок використання наявних висотних значень h_j як опорної інформації для корекції відносних поверхонь Z_j . Така корекція не перетворює результат на повністю метричну цифрову модель місцевості в загальному випадку, однак робить тривимірне представлення придатним для оцінювання висоти над поверхнею в прикладних сценаріях, що узгоджується з висновками [33].

Для компенсації геометричного розриву застосовано варіант SelaVPR++, який порівнює 3D-проекцію двовимірної супутникової карти із зображенням аерозйомки, що забезпечує суттєво вищі значення $R@k$ у режимі камери вздовж горизонту на всіх висотах.

Таблиця 4.34. Повнота пошуку для камери, спрямованої вниз, на висоті 20 м.

Метод	Чи навчена модель на даних аерозйомки	R@1, %	R@5, %	R@10, %	R@20, %
SelaVPR++	Так	30,0	55,0	63,0	72,0
SelaVPR+	Так	25,0	50,0	58,0	68,0
SelaVPR	Ні	12,0	25,0	32,0	40,0

Таблиця 4.35. Повнота пошуку для камери, спрямованої вниз, на висоті 60 м.

Метод	Чи навчена модель на даних аерозйомки	R@1, %	R@5, %	R@10, %	R@20, %
SelaVPR++	Так	98,1	99,1	99,6	99,7
SelaVPR+	Так	97,0	98,5	99,1	99,5
SelaVPR	Ні	88,0	95,0	96,8	97,5

У режимі камери вниз збільшення висоти з 20 м. до 60 м. приводить до різкого покращення якості ранжування для всіх методів.

Таблиця 4.36. Повнота пошуку для камери, спрямованої в горизонтальному напрямку, на висоті 40 м.

Метод	Чи навчена модель на даних аерозйомки	R@1, %	R@5, %	R@10, %	R@20, %
SelaVPR++	Так	60,0	82,0	88,0	92,0
SelaVPR+	Так	25,0	45,0	55,0	65,0
SelaVPR	Ні	10,0	20,0	28,0	38,0

Таблиця 4.37. Повнота пошуку для камери, спрямованої в горизонтальному напрямку, на висоті 60 м.

Метод	Чи навчена модель на даних аерозйомки	R@1, %	R@5, %	R@10, %	R@20, %
SelaVPR++	Так	80,0	92,0	95,0	97,0

Продовження таблиці 4.37

Метод	Чи навчена модель на даних аерозйомки	R@1, %	R@5, %	R@10, %	R@20, %
SelaVPR+	Так	45,0	70,0	78,0	85,0
SelaVPR	Ні	20,0	40,0	50,0	62,0

Використання SelaVPR++ суттєво змінює ситуацію. Для камери, спрямованої в горизонтальному напрямку, на висоті 40 м значення R@1 зростає до 60,0 %, а на висоті 60 м - до 80,0 %, що підтверджує ефективність введення 3D-проекції супутникової карти саме для косих ракурсів. У режимі камери вниз приріст також є позитивним: на висоті 20 м значення R@1 зростає з 25,0 % до 30,0 %, а на висоті 60 м - з 97,0 % до 98,1 %. Зростання R@k для більших k у всіх методів свідчить, що частина правильних відповідників усе ж потрапляє до короткого списку кандидатів, однак без SelaVPR++ якість першої позиції залишається недостатньою для автономного застосування.

Висновки до Розділу 4

У розділі проведено підтвердження працездатності запропонованих у дисертації методів, моделей і засобів. Отримано такі результати:

1. Кількісно підтверджено працездатність методу видалення малоінформативних цифрових зображень та зменшення повторюваності кадрів. Показано, що фільтрація забезпечує вилучення 30 % найменш інформативних кадрів і 9,91 % надмірно повторних кадрів, а також підвищує різноманітність вибірки у просторі ознак на 26,1 %.
2. Кількісно підтверджено працездатність запропонованої нейромережевої архітектури (моделі) двоетапного зіставлення аерознімків і супутникових сцен. На власній тестовій підвибірці досягнуто повноти пошуку $R@1 = 98,4 \%$, $R@5 = 99,5 \%$, $R@10 = 99,7 \%$ та $R@20 = 99,9 \%$.
3. Кількісно підтверджено працездатність розробленої структури векторної бази представлень та алгоритмів пошуку найближчих сусідів. На бортовому

комп'ютері NVIDIA Jetson Orin NX середній час оброблення одного запиту, що включає формування глобального дескриптора та пошук у векторному сховищі, становить 0,064 секунди, що відповідає частоті близько 15,6 кадрів за секунду і забезпечує придатність до роботи в режимі, наближеному до реального часу.

4. Підтверджено працездатність запропонованої інформаційної технології в умовах бортового застосування. За результатами натурних випробувань із використанням високоточного позиціювання у режимі реального часу встановлено, що для приблизно 85 % кадрів середня просторова похибка локалізації становить 4,45–4,57 метра, що підтверджує практичну придатність технології для задач грубої абсолютної прив'язки.
5. Порівняльний аналіз із наявними методами показав перевагу доменно адаптованого варіанта запропонованої архітектури за сукупністю показників точності та швидкодії на власній тестовій вибірці.
6. Кількісно підтверджено працездатність методу побудови тривимірного представлення супутникових сцен. Показано, що використання тривимірного узгодження супутникового еталона підвищує якість розпізнавання в складних геометричних режимах: значення $R@1$ зростає з 25,0 % до 60,0 % для горизонтального ракурсу на висоті 40 метрів та з 45,0 % до 80,0 % на висоті 60 метрів.
7. Отримані результати підтверджують, що запропоновані методи, нейромережева архітектура та розроблена інформаційна технологія є працездатними, узгодженими між собою та придатними до практичного використання у задачі абсолютної локалізації безпілотного літального апарата за оптичним каналом.

ЗАГАЛЬНІ ВИСНОВКИ

Дисертаційна робота є завершеним науковим дослідженням, у якому розв'язано актуальне науково-прикладне завдання розроблення інформаційної технології моделювання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером, призначеної для забезпечення абсолютної локалізації безпілотного літального апарата за оптичним каналом в умовах втрати або пригнічення сигналів глобальних навігаційних супутникових систем за обмежених бортових обчислювальних ресурсів.

У результаті виконаного дослідження отримано такі основні результати:

1. Проведено аналіз сучасних методів глибинного метричного навчання, візуального розпізнавання місць, а також підходів до абсолютної локалізації безпілотного літального апарата за оптичним каналом. Сформульовано математичну постановку задачі та визначено критерії оцінювання якості розпізнавання і локалізації.
2. Запропоновано метод видалення малоінформативних цифрових зображень із навчальних наборів аерозйомки, який, на відміну від евристичних методів фільтрації за окремими фотометричними ознаками, ґрунтується на аналізі латентних ознак автоенкодера, проєкції в простір головних компонент і квантильному відборі. Кількісно підтверджено, що його застосування забезпечує спрямоване вилучення 30 % найменш інформативних кадрів і підвищення різноманітності вибірки у просторі ознак на 26,1 %.
3. Удосконалено метод моделювання, формалізації та побудови навчальних, валідаційних і тестових наборів даних, який, на відміну від наявних підходів, враховує геоприв'язку аерознімки і супутникового фрагмента, траєкторію польоту, параметри зйомки, просторове розділення підвбірок і добір складних негативних прикладів. На основі цього методу створено навчальний набір обсягом понад 413 000 геоприв'язаних відповідностей, придатний для навчання, валідації та тестування нейромережових систем розпізнавання, а

після фільтрації отримано 263 701 репрезентативний приклад для побудови навчальних вибірок.

4. Набули подальшого розвитку методи статистичного аналізу латентного простору автоенкодера для оцінювання інформативності кадрів і ступеня відокремлення позитивних і негативних пар, що дало змогу обґрунтувати процедури підготовки даних для подальшого метричного навчання та підтвердити працездатність запропонованого методу фільтрації.
5. Запропоновано нейромережеву модель двоетапного зіставлення аерознімків і супутникових сцен. На відміну від відомих підходів, модель забезпечує одночасне формування глобальних і локальних ознак за зменшених обчислювальних витрат; на власній тестовій підвибірці досягнуто повноти пошуку $R@1 = 98,4 \%$, а середній час оброблення одного запиту на бортовому комп'ютері NVIDIA Jetson Orin NX становить 0,064 секунди.
6. Удосконалено метод геометричного узгодження супутникових еталонів з умовами бортового спостереження шляхом побудови тривимірного представлення супутникових сцен на основі монокулярної оцінки глибини та синтезу ракурсів, узгоджених із параметрами аерозйомки. На відміну від плаского двовимірного подання супутникових даних, запропонований підхід підвищує якість розпізнавання в складних режимах зйомки: для горизонтального ракурсу значення $R@1$ зростає з 25,0 % до 60,0 % на висоті 40 метрів і з 45,0 % до 80,0 % на висоті 60 метрів. Реалізований у роботі модуль побудови тривимірного представлення супутникових сцен може використовуватися для узгодження супутникових знімків за малих висот польоту, що розширює можливості застосування системи в умовах складної геометрії зйомки.
7. Набули подальшого розвитку методи зберігання та пошуку векторних представлень за рахунок розроблення структури векторної бази, винесення локальних ознак в окреме сховище та використання пошуку найближчих сусідів з урахуванням просторових обмежень. Це забезпечило придатність

- технології до бортового застосування в режимі, наближеному до реального часу.
8. Виконано алгоритмізацію методу моделювання, анотування, геоприв'язки та формування навчальних, валідаційних і тестових наборів даних, методу видалення малоінформативних цифрових зображень, методу геометричного узгодження супутникових даних з реальними бортовими даними та нейромережевої моделі двоетапного зіставлення, що забезпечило можливість їх подальшої програмної реалізації в складі інформаційної технології моделювання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером.
 9. На основі виконаної алгоритмізації розроблено та програмно реалізовано модулі інформаційної технології моделювання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером, зокрема модулі формування глобальних і локальних ознак, добору складних негативних прикладів, векторного пошуку, локального уточнення, геометричного перетворення та побудови тривимірного представлення супутникових сцен.
 10. Розроблено інформаційну технологію моделювання та розпізнавання навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером, призначену для абсолютної локалізації безпілотного літального апарата за оптичним каналом в умовах втрати або пригнічення сигналів глобальних навігаційних супутникових систем. На відміну від наявних рішень, вона реалізує двостадійне офлайн- та онлайн-оброблення, попереднє формування і зберігання векторних представлень супутникових еталонів, просторово обмежений пошук, локальне геометричне уточнення та модуль тривимірного узгодження супутникових сцен.
 11. За результатами натурних випробувань із використанням модуля високоточного позиціювання підтверджено працездатність розробленої інформаційної технології. Для наведених 85 % кадрів середня просторова похибка локалізації становить 4,45–4,57 метра.

12. Розроблені методи, алгоритми та програмні засоби можуть бути використані під час створення відомчих геоприв'язаних баз зображень місцевості, систем резервної навігації, систем візуальної прив'язки, розвідувально-пошукових бортових підсистем та інших інтелектуальних модулів для безпілотних літальних апаратів.
13. Результати дисертаційної роботи можуть бути використані в наукових дослідженнях та в освітньому процесі під час підготовки фахівців зі спеціальності 122 «Комп'ютерні науки» та суміжних спеціальностей у межах дисциплін з глибинного навчання, комп'ютерного зору, оброблення зображень та інтелектуальних бортових систем.
14. результати дисертаційного дослідження впроваджено в освітній процес кафедри прикладної математики Національного університету «Київський авіаційний інститут» під час викладання навчальних дисциплін: «Супровід підсистем цільового навантаження БПЛА», «Машинне навчання та розпізнавання образів», «Мова Python для глибокого навчання».
15. Результати дисертаційного дослідження впроваджено у виробничий процес компанії «Рейндорф» під час створення безпілотного літального апарата для потреб виконання спеціальних місій в умовах дії військового стану за ТУ У 30.0-44764773-006:2024.
16. Окремі результати дисертаційного дослідження відображено у патенті на корисну модель «Нейромережева 43-класова модель розпізнавання даних аерозйомки з камер безпілотних літальних апаратів».

Отже, поставлене в дисертаційній роботі науково-прикладне завдання розв'язано, а одержані результати мають наукову новизну, практичну цінність і підтверджують доцільність застосування розробленої інформаційної технології в задачах розпізнавання даних аерозйомки та резервної візуальної локалізації безпілотних літальних апаратів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
2. H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.
3. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2011, pp. 2564–2571.
4. J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2003, pp. 1470–1477.
5. H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2010, pp. 3304–3311.
6. F. Perronnin, J. Sánchez, and T. Mensink, "Improving the Fisher Kernel for large-scale image classification," in *Computer Vision – ECCV 2010 (Lecture Notes in Computer Science, vol. 6314)*. Berlin, Germany: Springer, 2010, pp. 143–156.
7. S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual place recognition: A survey," *IEEE Trans. Robot.*, vol. 32, no. 1, pp. 1–19, 2016.
8. R. Arandjelović, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 5297–5307.
9. F. Radenović, G. Tolias, and O. Chum, "Fine-tuning CNN image retrieval with no human annotation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1655–1668, 2019.
10. A. Ali-Bey, B. Chaib-Draa, and P. Giguère, "MixVPR: Feature mixing for visual place recognition," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, 2023, pp. 2998–3007.
11. S. Hausler, M. Xu, M. Milford, and T. Fischer, "Patch-NetVLAD: Multi-scale fusion of local and global descriptors for place recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 14141–14152.

12. B. Cao, A. Araujo, and J. Sim, "Unifying deep local and global features for image search," in *Computer Vision – ECCV 2020 (Lecture Notes in Computer Science*, vol. 12365). Cham, Switzerland: Springer, 2020, pp. 726–743.
13. M. Oquab et al., "DINOv2: Learning robust visual features without supervision," arXiv preprint arXiv:2304.07193, 2023. [Online]. Available: <https://arxiv.org/abs/2304.07193>
14. F. Lu, L. Zhang, X. Lan, S. Dong, Y. Wang, and C. Yuan, "Towards seamless adaptation of pre-trained models for visual place recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2024. [Online]. Available: <https://openreview.net/forum?id=EHROn1rQJ8>
15. S. Chen, C. Ge, X. Tong *et al.*, "AdaptFormer: Adapting vision transformers for scalable visual recognition," in *Adv. Neural Inf. Process. Syst.*, vol. 35, 2022
16. D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2018, pp. 224–236
17. R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, "Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1623–1637, 2022
18. R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 12179–12188.
19. E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*. Reading, MA, USA: Addison-Wesley, 1994.
20. MAVLink, "MAVLink Developer Guide," MAVLink Protocol Specification. [Online]. Available: <https://mavlink.io/en/>
21. GStreamer, "GStreamer Documentation: Application Development Manual." [Online]. Available: <https://gstreamer.freedesktop.org/documentation/application-development/>
22. M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
23. A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 8024–8035.

24. NVIDIA, "Jetson Orin NX Series Data Sheet." [Online]. Available: <https://developer.nvidia.com/downloads/jetson-orin-nx-series-data-sheet>
25. NVIDIA, "JetPack 6.1 Release Notes." [Online]. Available: <https://docs.nvidia.com/jetson/archives/jetpack-archived/jetpack-61/release-notes/index.html>
26. Holybro, "H-RTK F9P (u-blox ZED-F9P): Specification and RTK Accuracy." [Online]. Available: <https://docs.holybro.com/gps-and-rtk-system/h-rtk-f9p-gnss-series/h-rtk-f9p>
27. SIYI Technology (Shenzhen) Co., Ltd., "HM30 Full HD Transmission System: User Manual (EN) v1.6," 2025. [Online]. Available: https://siyi.biz/siyi_file/HM30/HM30%20User%20Manual%20EN%20v1.6.pdf
28. Raspberry Pi Ltd., "Raspberry Pi Global Shutter Camera: Product Brief," 2023. [Online]. Available: <https://datasheets.raspberrypi.com/gs-camera/gs-camera-product-brief.pdf>
29. W. Xu, Y. Yao, J. Cao et al., "UAV-VisLoc: A large-scale dataset for UAV visual localization," arXiv preprint arXiv:2405.11936, 2024. [Online]. Available: <https://arxiv.org/abs/2405.11936>
30. O. Kozachuk and P. Prystavka, "Comparing UAV camera data with satellite imagery for global localization," in Proc. 2024 IEEE 7th Int. Conf. Actual Problems Unmanned Aerial Vehicles Development (APUAVD), Kyiv, Ukraine, Nov. 2024, pp. 197–200.
31. Козачук О. О. Порівняння розподілів синтетичних та несинтетичних даних аерозйомки з використанням автокодувальника. Тези доповідей XXIII Міжнародної науково-практичної конференції здобувачів вищої освіти і молодих учених «ПОЛІТ. Сучасні проблеми науки». Сучасні інформаційні комунікаційні технології в авіації. Київ, 2023. С. 15–16.
32. M. Meingast, C. Geyer, and S. Sastry, "Geometric models of rolling-shutter cameras," arXiv preprint arXiv:cs/0503076, 2005. [Online]. Available: <https://arxiv.org/abs/cs/0503076>
33. O. O. Kozachuk, "Comparative analysis of monocular depth estimation of satellite two-dimensional images," Problemy Informatyzatsii ta Upravlinnia, vol. 3, no. 83, pp. 55–64, 2025. (in Ukrainian)

34. P. Prystavka, S. Dolgikh, and O. Kozachuk, "Terrain image recognition with unsupervised generative representations: The effect of anomalies," in Proc. 12th Int. Conf. Advanced Computer Information Technologies (ACIT), Ruzomberok, Slovakia, Oct. 2022, pp. 485–488.
35. Prystavka P., Dolgikh S., Cholyshkina O., Kozachuk O. Learned and Native Concepts in Latent Representations of Terrain Images. In: Information and Communication Technologies in Education, Research, and Industrial Applications. ICTERI 2021. Communications in Computer and Information Science, vol. 1698. Springer, Cham, 2022. Pp. 64–84. DOI: 10.1007/978-3-031-20834-8_4.
36. O. Kozachuk, "Comparative analysis of Vision Transformer and ResNet-50 as autoencoders for studying the representation space in aerial photography data recognition tasks," in Proc. XXV Int. Sci.-Practical Conf. POLIT. Modern Information Communication Technologies in Aviation, Kyiv, Ukraine, Apr. 2025, pp. 153–155. (in Ukrainian)
37. Zivakin V., Kozachuk O., Prystavka P., Cholyshkina O. Training set AERIAL SURVEY for Data Recognition Systems From Aerial Surveillance Cameras. CEUR Workshop Proceedings, Vol. 3347, IT&I-2022, 2022. Pp. 246–255.
38. V. Zivakin and O. Kozachuk, "Training dataset for neural network processing of aerial surveillance data," in Proc. 10th World Congress Aviation in the 21st Century: Safety in Aviation and Space Technologies, Kyiv, Ukraine, Sep. 2022, pp. 2.5.27–2.5.29. (in Ukrainian)
39. V. Zivakin, O. Kozachuk, and P. Prystavka, "Study of cross-domain neural network recognition of satellite and aerial survey data," in Proc. XVI Int. Sci.-Technical Conf. AVIA, Kyiv, Ukraine, Apr. 2023, pp. 15.78–15.81. (in Ukrainian)
40. O. Kozachuk, "Determining similarity of aerial survey images and satellite data based on autoencoder neural network architecture," in Proc. XXIV Int. Sci.-Practical Conf. POLIT. Modern Information Communication Technologies in Aviation, Kyiv, Ukraine, Apr. 2024, pp. 15–16. (in Ukrainian)

41. A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Represent. (ICLR), 2021. [Online]. Available: <https://openreview.net/forum?id=YicbFdNTTy>
42. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.
43. G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
44. P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in Proc. 25th Int. Conf. Machine Learning (ICML), 2008, pp. 1096–1103.
45. D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in Proc. Int. Conf. Learn. Represent. (ICLR), 2014. [Online]. Available: <https://arxiv.org/abs/1312.6114>
46. J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Artificial Neural Networks and Machine Learning – ICANN 2011 (Lecture Notes in Computer Science, vol. 6791)*. Berlin, Germany: Springer, 2011, pp. 52–59.
47. K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, 2009.
48. J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in Proc. 24th Int. Conf. Machine Learning (ICML), 2007, pp. 209–216.
49. J. Goldberger, G. E. Hinton, S. T. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," in *Adv. Neural Inf. Process. Syst.*, vol. 17, 2004, pp. 513–520.
50. J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a 'Siamese' time delay neural network," in *Adv. Neural Inf. Process. Syst.*, vol. 6, 1993, pp. 737–744.
51. R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 2, 2006, pp. 1735–1742.

52. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 815–823.
53. Y. Movshovitz-Attias, A. Toshev, T. K. Leung, S. Ioffe, and S. Singh, "No fuss distance metric learning using proxies," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 360–368.
54. E. W. Teh, T. DeVries, and G. W. Taylor, "ProxyNCA++: Revisiting and revitalizing Proxy Neighborhood Component Analysis," in *Computer Vision – ECCV 2020 (Lecture Notes in Computer Science, vol. 12369)*. Cham, Switzerland: Springer, 2020, pp. 448–464.
55. S. Kim, D. Kim, M. Cho, and S. Kwak, "Proxy Anchor Loss for deep metric learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 3238–3247.
56. Z. Zheng, Y. Wei, and Y. Yang, "University-1652: A multi-view multi-source benchmark for drone-based geo-localization," in *Proc. 28th ACM Int. Conf. Multimedia (MM '20)*, 2020, pp. 1395–1403.
57. R. Zhu, L. Yin, M. Yang, F. Wu, Y. Yang, and W. Hu, "SUES-200: A multi-height multi-scene cross-view image benchmark across drone and satellite," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 9, pp. 4825–4839, 2023.
58. M. Dai, E. Zheng, Z. Feng, L. Qi, J. Zhuang, and W. Yang, "Vision-based UAV self-positioning in low-altitude urban environments," *IEEE Trans. Image Process.*, vol. 33, pp. 493–508, 2024.
59. N. Keetha, A. Mishra, J. Karhade, K. M. Jatavallabhula, S. Scherer, M. Krishna, and S. Garg, "AnyLoc: Towards universal visual place recognition," *IEEE Robot. Autom. Lett.*, vol. 9, no. 2, pp. 1286–1293, 2024.
60. S. Izquierdo and J. Civera, "Optimal transport aggregation for visual place recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2024, pp. 17658–17668.
61. P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 4938–4947.

62. J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, "LoFTR: Detector-free local feature matching with transformers," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021, pp. 8922–8931.
63. P. Lindenberger, P.-E. Sarlin, and M. Pollefeys, "LightGlue: Local feature matching at light speed," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2023, pp. 17627–17638.
64. Qdrant, "Qdrant Documentation." [Online]. Available: <https://qdrant.tech/documentation/>
65. ArduPilot, "ArduPilot Copter Documentation." [Online]. Available: <https://ardupilot.org/copter/>
66. CUAV, "CUAV X7+ Controller Documentation." [Online]. Available: <https://doc.cuav.net/controller/x7/en/x7-plus.html>
67. ExpressLRS, "ExpressLRS Documentation." [Online]. Available: <https://www.expresslrs.org/>
68. PyTorch, "Torchvision Documentation." [Online]. Available: <https://docs.pytorch.org/vision/main/index.html>
69. NVIDIA, "CUDA Toolkit Documentation." [Online]. Available: <https://docs.nvidia.com/cuda/>
70. Python Software Foundation, "Python 3.10 Documentation." [Online]. Available: <https://docs.python.org/3.10/>
71. I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer, 2002.
72. K. Pearson, "On lines and planes of closest fit to systems of points in space," *London Edinburgh Dublin Philos. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901.
73. H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933; no. 7, pp. 498–520, 1933.
74. M. Qasaimeh, K. Denolf, A. Khodamoradi, M. Blott, J. Lo, L. Halder, K. Vissers, J. Zambreno, and P. H. Jones, "Benchmarking vision kernels and neural network inference accelerators on embedded platforms," *J. Syst. Archit.*, vol. 113, art. 101896, 2021.

75. M. Sanfourche, J. Delaune, G. Le Besnerais, H. de Plinval, J. Israel, P. Cornic, A. Treil, Y. Watanabe, and A. Plyer, "Perception for UAV: Vision-based navigation and environment modeling," *Aerospace Lab*, no. 4, pp. 1–19, 2012.
76. E. R. Jellum, T. H. Bryne, T. A. Johansen, and M. Orlandić, "The Syncline model: Analyzing the impact of time synchronization in sensor fusion," *arXiv preprint arXiv:2209.01136*, 2022. [Online]. Available: <https://arxiv.org/abs/2209.01136>
77. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015. [Online]. Available: <https://arxiv.org/abs/1412.6980>
78. J. P. Snyder, *Map Projections – A Working Manual*. Washington, DC, USA: U.S. Geological Survey, 1987.
79. Google, "Map and Tile Coordinates, Maps JavaScript API." [Online]. Available: <https://developers.google.com/maps/documentation/javascript/coordinates>
80. M. M. Deza and E. Deza, *Encyclopedia of Distances*, 4th ed. Cham, Switzerland: Springer, 2016.
81. P. C. Mahalanobis, "On the generalized distance in statistics," *Proc. Nat. Inst. Sci. India*, vol. 2, no. 1, pp. 49–55, 1936.
82. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
83. J. B. Kuipers, *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace, and Virtual Reality*. Princeton, NJ, USA: Princeton Univ. Press, 1999.
84. Autel Robotics, "EVO II Dual 640T Quick Guide V2," 2021. [Online]. Available: <https://www.autelrobotics.com/wp-content/themes/autel/userfiles/files/2021/10/15/EVO%20II%20Dual%20640T%20Quick%20Guide%20V2%20-%20EN.pdf>
85. C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948; no. 4, pp. 623–656, 1948.
86. F. J. Massey Jr., "The Kolmogorov–Smirnov test for goodness of fit," *J. Amer. Statist. Assoc.*, vol. 46, no. 253, pp. 68–78, 1951.

87. Y. A. Malkov and D. A. Yashunin, "Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 824–836, 2020.

ДОДАТКИ

Додаток А. Акт про впровадження у навчальний процес кафедри прикладної математики Київського авіаційного інституту

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНЕ НЕКОМЕРЦІЙНЕ ПІДПРИЄМСТВО
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «КИЇВСЬКИЙ АВІАЦІЙНИЙ ІНСТИТУТ»

АКТ

Про впровадження результатів дисертаційного дослідження у навчальний процес здобувача наукового ступеня доктора філософії кафедри «Інформаційних технологій» державного університету «Київський авіаційний інститут»

Козачука Олександра Олександровича

Тема дисертації: «Інформаційна технологія моделювання та розпізнавання об'єктів навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером»

Шифр і назва спеціальності: «122 Комп'ютерні науки»

Етап виконання: дисертаційне дослідження виконується з 2022 р.

Науковий керівний:

Пилип ПРИСТАВКА
(ім'я та ПРІЗВИЩЕ)

Комісія в складі:

Голова комісії: В.О. зав. кафедри Т.М.

Метена Олександр
(ім'я та ПРІЗВИЩЕ)

Члени комісії: Доцент кафедри Т.М.

Валерій Зісковий
(ім'я та ПРІЗВИЩЕ)

Парант О.П.
(посада)

Валерій Дребен
(ім'я та ПРІЗВИЩЕ)

Тупко Катерина, гарант О.П.
(посада)

Катерина Тупко
(ім'я та ПРІЗВИЩЕ)

Встановила, що результати дисертаційного дослідження (проміжні) використовуються в освітньому процесі за освітніми програмами «Прикладна математика» та «Прикладна програма забезпечення»

(назва програми)

освітнього ступеня Бакалавр, магістр (бакалавр, магістр, доктор філософії)

спеціальності 43 Прикладна математика (назва спеціальності)

шляхом реалізації наступного: основною змістом освітніх програм в термі виконання дисертації є моделювання та розпізнавання об'єктів зображень аерозйомки. Запропоновано новий підхід до моделювання та розпізнавання об'єктів зображень аерозйомки, що дає можливість використовувати результати дослідження для розробки нових програм та методів розпізнавання об'єктів зображень аерозйомки. Також запропоновано новий підхід до розпізнавання об'єктів зображень аерозйомки, що дає можливість використовувати результати дослідження для розробки нових програм та методів розпізнавання об'єктів зображень аерозйомки.

«28» серпня 2025 р.

Голова комісії:

Т.М.
(підпис)

Метена Олександр
(ім'я та ПРІЗВИЩЕ)

Члени комісії:

В.О.
(підпис)

Валерій Зісковий
(ім'я та ПРІЗВИЩЕ)

О.П.
(підпис)

Валерій Дребен
(ім'я та ПРІЗВИЩЕ)

К.Т.
(підпис)

Тупко Катерина
(ім'я та ПРІЗВИЩЕ)

Додаток Б. Акт впровадження у виробничий процес компанії «Рейндорф». Результати роботи використовуються у кодифікованому (ТУ У 30.0-44764773-006:2024) літальному апараті



АКТ

Про впровадження результатів дисертаційного дослідження здобувача наукового ступеня доктора філософії кафедри «Інформаційних технологій» державного університета «Київський авіаційний інститут»

Козачука Олександра Олександровича

Чинний акт складено на підставі застосування результатів дисертаційної роботи «Інформаційна технологія моделювання та розпізнавання об'єктів навчальних наборів аерозйомки на основі архітектур нейромереж з автоенкодером», які пройшли процедуру кодифікації ТУ У 30.0-44764773-006:2024 та використані у наведеному рішенні. А саме, досягнуто наступних результатів:

- Сформовано навчальні набори аерозйомки (схеми розмітки, метадані, процедури контролю якості).
- Реалізовано засоби обробки зображень на основі нейромереж, інтегровані в програмне рішення через стандартизовані інтерфейси.
- Розроблено методику моделювання об'єктів і умов зйомки та сценарного випробування алгоритмів.

Директор
ТОВ «Рейндорф»

Головний інженер,
к.т.н.

А. ПАВЛЕНКО

В. ЧУЙКО

Додаток В. Повна інформація щодо виконаних польотів для отриманої вибірки даних

Номер польоту	Дата польоту	Погодні умови	Знята відстань
1	21.02.2024	Хмарно	9,3 км
2	23.02.2024	Мінлива хмарність	4 км
3	27.02.2024	Мінлива хмарність	13,2 км
4	28.02.2024	Мінлива хмарність	17,4 км
5	01.03.2024	Хмарно	7,1 км
6	03.03.2024	Хмарно	8,0 км
7	05.03.2024	Хмарно	6,1 км
8	07.03.2024	Хмарно	9,4 км
9	09.03.2024	Хмарно	4,8 км
10	11.03.2024	Хмарно	5,1 км
11	13.03.2024	Хмарно	11,2 км
12	14.03.2024	Хмарно	5,4 км
13	16.03.2024	Хмарно	9,0 км
14	18.03.2024	Хмарно	11,9 км
15	20.03.2024	Хмарно	4,9 км
16	22.03.2024	Мінлива хмарність	10,8 км
17	24.03.2024	Хмарно	7,0 км
18	26.03.2024	Хмарно	4,6 км
19	27.03.2024	Мінлива хмарність	5,3 км
20	29.03.2024	Мінлива хмарність	9,9 км
21	31.03.2024	Сонячно	9,7 км
22	02.04.2024	Сонячно	5,0 км
23	04.04.2024	Мінлива хмарність	7,3 км
24	06.04.2024	Мінлива хмарність	5,3 км
25	08.04.2024	Сонячно	11,5 км
26	09.04.2024	Сонячно	9,8 км

Продовження таблиці

Номер польоту	Дата польоту	Погодні умови	Знята відстань
27	11.04.2024	Мінлива хмарність	4,9 км
28	13.04.2024	Мінлива хмарність	11,7 км
29	15.04.2024	Мінлива хмарність	5,7 км
30	17.04.2024	Хмарно	7,1 км
31	19.04.2024	Хмарно	12,5 км
32	21.04.2024	Мінлива хмарність	12,5 км
33	23.04.2024	Мінлива хмарність	11,9 км
34	24.04.2024	Мінлива хмарність	4,9 км
35	26.04.2024	Мінлива хмарність	11,8 км
36	28.04.2024	Мінлива хмарність	11,9 км
37	30.04.2024	Мінлива хмарність	9,4 км
38	02.05.2024	Сонячно	4,8 км
39	04.05.2024	Мінлива хмарність	7,1 км
40	06.05.2024	Сонячно	4,7 км
41	08.05.2024	Мінлива хмарність	11,6 км
42	10.05.2024	Мінлива хмарність	5,9 км
43	11.05.2024	Мінлива хмарність	8,0 км
44	13.05.2024	Хмарно	9,7 км
45	15.05.2024	Мінлива хмарність	6,0 км
46	17.05.2024	Сонячно	11,4 км
47	19.05.2024	Сонячно	5,7 км
48	21.05.2024	Сонячно	11,8 км
49	22.05.2024	Сонячно	8,2 км
50	24.05.2024	Сонячно	11,6 км
51	26.05.2024	Сонячно	6,6 км
52	28.05.2024	Сонячно	5,5 км
53	30.05.2024	Мінлива хмарність	11,9 км

Продовження таблиці

Номер польоту	Дата польоту	Погодні умови	Знята відстань
54	31.05.2024	Сонячно	11,8 км
55	02.06.2024	Сонячно	6,7 км
56	04.06.2024	Сонячно	9,1 км
57	06.06.2024	Сонячно	5,4 км
58	08.06.2024	Сонячно	11,5 км
59	10.06.2024	Сонячно	5,0 км
60	12.06.2024	Мінлива хмарність	11,7 км
61	13.06.2024	Сонячно	1,8 км
62	21.06.2024	Сонячно	35,2 км
Загальна кількість польотів = 62		Загальна відстань = 546 км	