

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

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

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти _____ перший (бакалаврський)

_____ Інтелектуальна система автоматизованого детектування малорозмірних
_____ повітряних цілей у відеопотоках
(тема)

Виконав:
здобувач _____ четвертого _____ року навчання,
групи _____ ІТШ-22-1

_____ Йоханнес МАШМАЙЕР
(власне ім'я, прізвище)

Спеціальність 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми _____ освітньо-професійна

Освітня програма _____ Штучний інтелект

(повна назва освітньої програми)

Керівник _____ ас. Сергій ГРИНЬОВ
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ШІ

(підпис)

_____ Лариса ЧАЛА
(власне ім'я, прізвище)

2026 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____

Кафедра _____ Штучного інтелекту _____

Рівень вищої освіти _____ перший (бакалаврський) _____

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

Тип програми _____ освітньо-професійна _____

Освітня програма _____ Штучний інтелект _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____

(підпис)

« _____ » _____ 20 ____ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _____ Машмайеру Йоханнесу _____
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Інтелектуальна система автоматизованого детектування малорозмірних повітряних цілей у відеопотоках _____

затверджена наказом університету від 5 червня 2026 р. № 488Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 9 липня 2026 р.

3. Вихідні дані до роботи Науково-технічні публікації з тематики комп'ютерного зору, глибинного навчання, методів виявлення малорозмірних об'єктів та безпілотних літальних апаратів (БПЛА); дані Інтернет-джерел (офіційна технічна документація Ultralytics YOLO, аналітичні звіти щодо ринку та засобів Anti-Drone моніторингу); документація бібліотек та фреймворків мови Python (PyTorch, OpenCV, NumPy); документація інструментів апаратної оптимізації та прискорення обчислювальних графів (NVIDIA TensorRT, відкритий формат ONNX); реальні та симуляційні відеопослідовності динамічного повітряного простору з малорозмірними цілями на складному завадовому фоні для верифікації ефективності запропонованих алгоритмічних рішень

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Аналіз предметної галузі _____

2) Обґрунтування вибору технологій та методів виявлення повітряних цілей _____

3) Архітектура системи та програмна реалізація _____

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / термін виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	08.06.2026	виконано
2	Аналіз предметної галузі	11.06.2026	виконано
3	Огляд існуючих рішень та технологій	14.06.2026	виконано
4	Аналіз підходів до генерації подкастів	17.06.2026	виконано
5	Проектування архітектури програмної системи	20.06.2026	виконано
6	Написання пояснювальної записки	26.06.2026	виконано
7	Перевірка на академічний плагіат	03.07.2026	виконано
8	Нормоконтроль	03.07.2026	виконано
9	Підготовка презентації та доповіді	05.07.2026	виконано
10	Попередній захист	06.07.2026	виконано
11	Рецензування	07.07.2026	виконано
12	Захист перед ЕК	09.07.2026	

Дата видачі завдання 8 червня 20 26 р.

Здобувач 
(підпис)

Керівник роботи 
(підпис)

ас. Сергій ГРИНЬОВ
(посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка: 66 с., 5 рис., 1 табл., 1 дод., 20 джерел.

БПЛА, ВІДЕОПОТІК, ГЛИБОКЕ НАВЧАННЯ, ДЕТЕКТУВАННЯ,
ЗГОРТКОВА НЕЙРОННА МЕРЕЖА, КООРДИНАТНА УВАГА,
МАЛОРОЗМІРНІ ЦІЛІ.

Об'єкт дослідження – процес автоматизованого детектування та супроводження малорозмірних об'єктів у динамічних відеопотоках за умов складного навколишнього середовища.

Предмет дослідження – моделі, алгоритми та програмні архітектури семантичного вилучення ознак малих об'єктів, когнітивного аналізу та прецизійної цифрової стабілізації відеопотоків із використанням одноетапних згорткових нейронних мереж та механізмів координатної уваги.

Мета роботи – розробити архітектурно-алгоритмічну концепцію інтелектуальної системи автоматизованого моніторингу повітряного простору та виявлення об'єктів малого розміру в режимі реального часу на основі модифікованої архітектури YOLOv8 із впровадженням підсистеми координатної уваги та цифрової стабілізації зображення.

Методи дослідження – машинне навчання та одноетапні згорткові архітектури, механізми просторово-структурної селекції ознак, комп'ютерний зір та цифрова стабілізація зображень, фільтрація та згладжування траєкторій.

Запроєктована концепція інтелектуальної системи та змодельовані алгоритмічні рішення забезпечують теоретичні й інженерні засади для автоматизації виявлення малорозмірних повітряних цілей у потокових відеоданих за допомогою новітнього підходу чотириголової мікродетекції.

ABSTRACT

Bachelor's thesis contains: 66 pp., 5 fig., 1 tabl., 1 ann., 20 references.

COMPUTER VISION, CONVOLUTIONAL NEURAL NETWORK, COORDINATE ATTENTION, DEEP LEARNING, DETECTION, SMALL OBJECT DETECTION, UAV, VIDEO STREAM, YOLOv8.

The object of the study – the process of automated detection and tracking of small objects in dynamic video streams.

The subject of the study – models, algorithms, and architectures for semantic feature extraction of ultra-small objects, spatio-temporal anomaly analysis, and digital video stabilization using convolutional neural networks and coordinate attention.

The purpose of the work – to develop an architectural and algorithmic concept of an automated airspace monitoring system for real-time detection of ultra-small objects based on a modified YOLOv8 model.

Research methods – deep learning, convolutional architectures (YOLOv8), feature selection (Coordinate Attention), digital image stabilization (ORB, RANSAC), trajectory filtering (Kalman filter).

The designed conceptual architecture and its algorithmic solutions provide the engineering framework for automating small aerial target detection in streaming video data using an advanced four-head micro-detection approach based on YOLOv8 with coordinate attention and ego-motion compensation.

ЗМІСТ

Вступ.....	8
1 Аналіз предметної галузі.....	12
1.1 Аналіз існуючих програмно-технологічних рішень для виявлення повітряних цілей.....	12
1.1.1 Програмні комплекси на базі класичних бібліотек комп'ютерного зору (OpenCV, MATLAB)	12
1.1.2 Програмні рішення на основі двоетапних (Two-stage) нейромережевих фреймворків (Faster R-CNN, Cascade R-CNN)	13
1.1.3 Програмні інструменти на основі одноетапних (Single-stage) детекторів реального часу (Ultralytics YOLO, SSD).....	13
1.1.4 Спеціалізовані комерційні C-UAS (Anti-Drone) програмні платформи інтегрованого моніторингу	14
1.2 Специфіка та деструктивні фактори предметної галузі.....	16
1.2.1 Характеристики малорозмірних повітряних об'єктів у відеопотоках	16
1.2.2 Вплив завадового динамічного фону та умов зовнішнього середовища.....	18
1.3 Актуальність роботи та обґрунтування розробки	20
1.4 Мета роботи та постановка задачі.....	21
2 Обґрунтування вибору технологій та методів виявлення повітряних цілей	24
2.1 Проблема еґо-руху та вібрації сенсорів спостереження.....	24
2.2 Аналіз класичних методів комп'ютерного зору для виявлення рухомих об'єктів.....	26
2.2.1 Методи віднімання фону (Background Subtraction).....	26
2.2.2 Методи аналізу оптичного потоку (Optical Flow)	28
2.2.3 Обмеження традиційних підходів у задачах мікродетекції	30
2.3 Огляд сучасних підходів на основі глибинного навчання.....	31

2.3.1 Одноетапні та двоетапні згорткові нейронні мережі	31
2.3.2 Особливості застосування механізмів уваги.....	32
2.3.3 Моделі аналізу просторово-часових ознак у відеопослідовностях	33
2.4 Метрики оцінювання якості та ефективності інтелектуальних систем детектування	33
2.5 Обґрунтування вибору базової моделі глибинного навчання та інструментальних засобів розробки.....	38
2.6 Концепція та математичний опис механізму координатної уваги в задачах мікродетекції.....	39
2.7 Модифікація архітектури згорткової нейронної мережі та проектування шарів мікродетекції	41
3 Архітектура системи та програмна реалізація	44
3.1 Архітектурна структура та модульний розподіл програмного комплексу	44
3.1.1 Підсистема інгестії та стабілізації відеопотоку	45
3.1.2 Підсистема нейромережевого аналізу та мікродетекції	45
3.1.3 Підсистема просторово-часової пост-обробки та фільтрації	46
3.1.4 Підсистема користувацького інтерфейсу та зовнішньої інтеграції	46
3.2 Організація асинхронного конвеєра обробки поточкових відеоданих	47
3.3 Алгоритмічна реалізація та високорівневі лістинги ключових компонентів системи.....	49
3.4 Проектування інтерфейсу користувача та системи сповіщень	53
Висновки	59
Перелік джерел посилання	63
Додаток А Відомість кваліфікаційної роботи	66

ВСТУП

Актуальність роботи та підстави для її виконання. Сучасний етап розвитку систем безпеки та військових технологій характеризується стрімким розширенням сфери застосування безпілотних літальних апаратів (БПЛА), а також інших малорозмірних повітряних засобів. Забезпечення надійного захисту об'єктів критичної інфраструктури, державних кордонів, військових позицій та цивільних зон потребує створення систем безперервного автоматизованого моніторингу повітряного простору в режимі реального часу.

Традиційні радіолокаційні системи (РЛС), які протягом тривалого часу залишалися основою протиповітряної оборони, демонструють низьку ефективність під час виявлення сучасних БПЛА [1]. Це зумовлено малою ефективною площею розсіювання таких цілей, їхньою здатністю рухатися на наднизьких висотах, з малими швидкостями або навіть зависати на місці, повністю зливаючись із радіовідбиттям від підстильної поверхні (дерев, пагорбів, будівель).

У таких умовах оптико-електронні системи спостереження та методи комп'ютерного зору стають безальтернативним або критично важливим комплементарним каналом виявлення інформації. Проте ручне (візуальне) детектування повітряних цілей оператором у відеопотоці є неефективним через високе психофізіологічне навантаження, втому та обмеженість людського сприйняття, особливо під час тривалого моніторингу. Таким чином, виникає гостра потреба в автоматизації цього процесу шляхом розробки та впровадження інтелектуальних систем автоматизованого детектування малорозмірних повітряних цілей, що й визначає актуальність цієї кваліфікаційної роботи.

Оцінка сучасного стану проблеми: на сьогодні у сфері комп'ютерного зору досягнуто значних успіхів завдяки розвитку технологій глибинного навчання та згорткових нейронних мереж (CNN). Архітектури сімейства

YOLO, SSD, Faster R демонструють високу точність та швидкість роботи під час розпізнавання стандартних об'єктів (людей, автомобілів, міського транспорту) за стабільних умов спостереження.

Разом з тим, пряме застосування базових моделей глибинного навчання для виявлення малорозмірних повітряних об'єктів у реальних відеопотоках стикається із суттєвим зниженням метрик точності та високим рівнем хибних спрацьовувань. Більшість існуючих комерційних та відкритих рішень орієнтовані на роботу з об'єктами, які займають значну площу кадру і мають виражені текстурні ознаки, що робить проблему детектування цілей на великих відстанях усе ще відкритою для наукових та інженерних пошуків.

Науково-технічна проблематика дослідження: незадовільна якість роботи стандартних інтелектуальних систем під час виявлення малорозмірних повітряних цілей зумовлена низкою специфічних деструктивних факторів та фундаментальних технічних протиріч предметної галузі.

Ефект «розчинення» ознак у глибоких шарах нейромереж. Класичні згорткові нейромережі використовують послідовні операції субдискретизації та згортки з великим кроком для зменшення просторової розмірності кадру та вилучення високорівневих семантичних ознак. Малорозмірна повітряна ціль на початкових кадрах може мати лінійні розміри від 3×3 до 16×16 пікселів, займаючи менше 0.01% від загальної площі зображення. У результаті проходження через перші кілька блоків мережі просторова інформація про такий об'єкт повністю втрачається (яскравість пікселів цілі усереднюється з фоном), і фінальні шари детектора виявляються нездатними згенерувати правильний обмежувальний контур.

Вкрай низька інформативність та семантична порожнеча цілі. На відміну від стандартних класів об'єктів, малорозмірна ціль на великій відстані позбавлена чітких геометричних контурів, унікальних текстурних

градієнтів чи характерного забарвлення. Об'єкт відображається у відеопотоці як монохромна або слабкоконтрастна група пікселів із розмитими межами, що робить його майже ідентичним до випадкового цифрового шуму матриці камери.

Складний, динамічний та нестационарний фон. Повітряний простір, що виступає фоном спостереження, характеризується високим рівнем природних завад. До них належать рухомі хмари з мінливими межами, сонячні відблиски, туман, смог, а також птахи. Поведінка, візуальні розміри та траєкторії руху птахів часто є повністю ідентичними до мікро-БПЛА, що призводить до критичного зростання частоти хибних спрацьовувань інтелектуальної системи.

Его-рух та вібрації сенсора. Оптико-електронні комплекси моніторингу часто встановлюються на мобільних платформах або високих щоглах, які піддаються вітровим навантаженням та механічним вібраціям. Це викликає постійне зсунення всього кадру (ефект тремтіння камери). За таких умов класичні та швидкі методи комп'ютерного зору, такі як віднімання фону або аналіз оптичного потоку, стають непрацездатними, оскільки вони сприймають рух самого фону як рух сотень потенційних цілей.

Обмеження обчислювальної складності (проблема Real-time обробки). Для забезпечення своєчасного реагування на повітряні загрози система повинна обробляти відеопотік без затримок із частотою не менше 30 кадрів на секунду. Важкі та глибокі нейромережеві архітектури (наприклад, трансформери бачення), які здатні враховувати глобальний контекст кадру, потребують значних обчислювальних ресурсів (потужних серверних GPU). Водночас реальні системи моніторингу мають розгортатися на периферійних пристроях із обмеженою потужністю. Це створює жорстке інженерне протиріччя між математичною складністю алгоритму та його швидкодією.

Таким чином, наукова проблема полягає в необхідності розробки такого методу інтелектуального аналізу відеопотоку, який дозволить підвищити чутливість системи до малорозмірних об'єктів без пропорційного збільшення кількості хибних спрацьовувань на заводовому фоні в умовах обмежених обчислювальних ресурсів периферійних систем.

Об'єкт дослідження – процес автоматизованого детектування малорозмірних об'єктів у динамічних відеопотоках за умов складного навколишнього середовища.

Предмет дослідження – моделі, методи та алгоритми глибинного навчання, архітектури згорткових нейронних мереж, механізми просторово-часової уваги та методи міжкадрової обробки відеоданих для виявлення малорозмірних повітряних цілей.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Аналіз існуючих програмно-технологічних рішень для виявлення повітряних цілей

Для вирішення задачі автоматизованого виявлення та супроводження повітряних об'єктів у сучасній практиці комп'ютерного зору сформувалося кілька основних програмно-технологічних підходів. Вони відрізняються за математичною основою, вимогами до обчислювальної потужності апаратного забезпечення та здатністю працювати в складних заводових умовах. Проведемо аналіз переваг та недоліків основних категорій існуючих програмних рішень.

1.1.1 Програмні комплекси на базі класичних бібліотек комп'ютерного зору (OpenCV, MATLAB)

Цей технологічний підхід базується на використанні детермінованих алгоритмів обробки зображень (віднімання фону, аналіз оптичного потоку, контурний аналіз) без застосування нейромережевих структур.

Переваги – мінімальна обчислювальна складність. Алгоритми ефективно виконуються на центральних процесорах (CPU) низької потужності, що дозволяє розгортати їх на бюджетних мікрокомп'ютерах. Відсутня потреба у зборі, розмітці та попередньому навчанні на великих масивах даних (датасетах). Час обробки одного кадру є стабільним і мінімальним, що забезпечує високу частоту кадрів (FPS).

Недоліки – критична чутливість до его-руху. Будь-яке тремтіння камери або зміна ракурсу призводить до руйнування моделі фону та масової генерації помилкових маркерів. Програми не здатні класифікувати об'єкт, тобто вони фіксують будь-який рух, не відрізняючи БПЛА від птаха,

рухомої хмари чи листя дерев. Амплітудні пороги бінаризації неефективні за умов низького просторового контрасту цілі відносно неба.

1.1.2 Програмні рішення на основі двоетапних (Two-stage) нейромережевих фреймворків (Faster R-CNN, Cascade R-CNN)

Цей клас програмного забезпечення використовує архітектури глибинного навчання, які спочатку виділяють регіони потенційного знаходження об'єктів (Region Proposal Network), а потім виконують їх попиксельну класифікацію та регресію обмежувальних контурів.

Переваги – висока точність локалізації. Двоетапний аналіз забезпечує точне визначення координат об'єктів навіть за умов складного фону. Ефективність на середніх відстанях. Завдяки детальному аналізу виділених регіонів, архітектури здатні розпізнавати об'єкти з мінімальними текстурними ознаками.

Недоліки – низька швидкодія. Велика математична складність робить ці рішення непридатними для обробки потокового відео в реальному часі (FPS часто падає нижче 10–15 кадрів/сек). Апаратна залежність – програми вимагають потужних дискретних графічних процесорів (GPU) з великим об'ємом відеопам'яті (VRAM), що унеможливорює периферійні обчислення.

1.1.3 Програмні інструменти на основі одноетапних (Single-stage) детекторів реального часу (Ultralytics YOLO, SSD)

Найбільш поширений сучасний підхід в інженерії комп'ютерного зору, де прогнозування координат обмежувальних рамок та визначення класів об'єктів відбувається за один наскрізний прохід (one forward pass) через нейронну мережу [2].

Переваги – високий FPS в реальному часі. Оптимізовані конвеєри обробки дозволяють досягати швидкодії у 30–100+ кадрів за секунду. Гнучкість оптимізації – програмні моделі легко конвертуються у специфічні формати (TensorRT, ONNX, OpenVINO) для прискорення на мобільних чи спеціалізованих Edge-пристроях. Розвинена екосистема – наявність готових відкритих інструментів (наприклад, бібліотеки Ultralytics [3]) суттєво спрощує процес навчання та розгортання моделей.

Недоліки – проблема мікродетекції. Стандартні архітектури YOLO оптимізовані для великих та середніх класів. Об'єкти розміром менше ніж 16×16 пікселів втрачаються на шарах субдискретизації. Хибні спрацьовування на схожі класи. Без додаткової модифікації базові моделі часто плутають віддалені дрони із птахами через брак семантичної інформації на поодинокому кадрі.

1.1.4 Спеціалізовані комерційні C-UAS (Anti-Drone) програмні платформи інтегрованого моніторингу

Проприєтарне програмне забезпечення закритих екосистем, розроблене спеціалізованими підприємствами для комплексного захисту об'єктів.

Переваги – мультисенсорна інтеграція. Здатність програмно об'єднувати дані з різних каналів (оптичного, тепловізійного, радіолокаційного та радіочастотного). Готові алгоритми фільтрації. Наявність відпрацьованих фільтрів промислових завод та адаптований інтерфейс оператора.

Недоліки – закритий вихідний код. Повна неможливість модифікації архітектури алгоритмів під специфічні інженерні задачі (наприклад, підключення автомобільних відеореєстраторів). Висока вартість – це колосальні витрати на ліцензування програмного комплексу та жорстка прив'язка до коштовного фірмового апаратного забезпечення.

Порівняльні дані програмно-технологічних рішень наведено в таблиці 1.1.

Таблиця 1.1 – Порівняльна таблиця програмно-технологічних рішень

Підхід/Клас програмних рішень	Основні переваги	Основні недоліки
Класичні детерміновані алгоритми (OpenCV, MATLAB)	Ефективна робота на CPU. Без потреби у зборі датасетів. Висока частота кадрів (FPS).	Чутливість до тремтіння камери. Відсутність класифікації об'єктів. Помилки при низькому контрасті цілі.
Двоетапні нейромережеві фреймворки (Faster R-CNN)	Висока точність локалізації координат Ефективність на середніх відстанях; Робота з мінімальною текстурою цілі.	Низька швидкодія (менше 15 FPS); Непридатність для реального часу; Залежність від потужних серверних GPU.
Одноетапні детектори реального часу (Ultralytics YOLO)	Високий FPS в реальному часі (30–100+); Доступність відкритих інструментів.	Втрата мікрооб'єктів на шарах згортки; Хибні спрацьовування на птахів
Спеціалізовані комерційні C-UAS платформи	Мультисенсорна інтеграція (РЛС, оптика); Готові фільтри промислових завод; Адаптований інтерфейс.	Повністю закритий вихідний код; Неможливість кастомної модифікації; Висока вартість та прив'язка до заліза.

Для наочного зіставлення проаналізованих підходів сформовано узагальнену таблицю за ключовими критеріями ефективності.

Висновок порівняльного аналізу. Жодне з існуючих базових рішень у своєму первинному вигляді не задовольняє інженерним вимогам розроблюваної системи (робота в реальному часі на периферійному залізі з високою точністю виявлення мікрооб'єктів). Це обґрунтовує необхідність

взяття за основу одноетапного детектора сімейства YOLO та його подальшої науково-технічної модернізації за рахунок інтеграції механізмів координатної уваги та оптимізації шарів вилучення мікроознак.

1.2 Специфіка та деструктивні фактори предметної галузі

1.2.1 Характеристики малорозмірних повітряних об'єктів у відеопотоках

У задачах комп'ютерного зору та автоматизованого аналізу відеопотоків важливе місце посідає точне визначення поняття «малорозмірний об'єкт» (small object). Загальноприйняті критерії класифікації об'єктів за їхніми розмірами поділяють на абсолютні та відносні. Відповідно до стандартів Міжнародного товариства оптичної інженерії (SPIE), малорозмірною вважається ціль, чия площа на кадрі не перевищує 80 пікселів, що орієнтовно відповідає лінійним розмірам менше ніж 9×9 пікселів. У сучасних загальноцільових наборах даних, таких як MS COCO, використовується фіксований абсолютний критерій, де до категорії малих відносять об'єкти з площею менше ніж 32×32 пікселі [4]. Проте для систем моніторингу повітряного простору з високою роздільною здатністю сучасних оптичних сенсорів більш доцільним є використання відносного критерію. Згідно з ним, об'єкт класифікується як малорозмірний, якщо його площа становить менше ніж 0,1% або навіть 0,01% від загальної площі кадру. У контексті детектування віддалених повітряних цілей площа об'єкта на початкових етапах спостереження часто варіюється в межах від декількох одиниць до розміру 16×16 пікселів, що накладає жорсткі обмеження на архітектуру засобів обробки даних.

Аналіз геометричних властивостей малорозмірних повітряних цілей свідчить про майже повну відсутність просторової структури об'єкта. Якщо великі та середні об'єкти мають виражені геометричні примітиви, такі як

чіткі лінії, кути чи характерні контури крил та фюзеляжу, то малорозмірна ціль на великій відстані втрачає свій первісний вигляд через обмеження роздільної здатності оптичної системи та явища дифракції. Внаслідок дії функції розсіювання точки оптичного сенсора зображення віддаленого БПЛА розмивається і набуває форми нечіткої просторової плями із симетричним або асиметричним розподілом енергії. Додатковим деструктивним фактором є нестабільність геометричних пропорцій під час руху: зміна ракурсу польоту чи кута спостереження призводить до того, що відношення сторін обмежувального контуру змінюється хаотично. Це унеможливорює ефективне використання класичних геометричних шаблонів, дескрипторів форми або контурного аналізу для ідентифікації та супроводження об'єкта.

Текстурні властивості таких об'єктів характеризуються високою однорідністю внутрішньої інтенсивності та семантичною порожнечою. Через мінімальну кількість пікселів, що відображають ціль, всередині її силуету повністю відсутні внутрішні градієнти, локальні бінарні шаблони або текстурні візерунки. Малорозмірний об'єкт фактично не має власної текстури, а його візуальне представлення на кадрі повністю інтегрується з колірними характеристиками навколишнього фону. Більше того, під час стиснення відеопотоку сучасними кодексами (зокрема H.264 або H.265) інформація у високочастотних областях кадру піддається значному згладжуванню. Внаслідок цього малорозмірні об'єкти часто потрапляють у межі одного макроблоку стиснення і розмиваються, втрачаючи будь-які текстурні відмінності від сусідніх ділянок неба чи хмар, що різко знижує ймовірність їх правильної сегментації.

Яскравісні властивості малорозмірних повітряних цілей є найбільш критичними для їхнього успішного детектування, але водночас і найменш стабільними. Головною характеристикою в даному випадку виступає низький просторовий контраст об'єкта відносно локального фону. Велика відстань до цілі призводить до розсіювання та поглинання світла в

атмосферних шарах, що суттєво нівелює різницю між яскравістю об'єкта та фону. За умов похмурої погоди або під час руху об'єкта на фоні неоднорідних хмар значення яскравості пікселів цілі стають близькими до значень яскравості перешкод, що зумовлює низьке відношення сигналу до шуму. Навпаки, за ясної погоди під впливом прямого сонячного світла яскравість об'єкта може різко коливатися від глибокої тіні до інтенсивного відблиску. Мала кількість пікселів також робить яскравісний профіль об'єкта вкрай чутливим до випадкового цифрового шуму матриці камери (теплого або фотонного шуму), оскільки флуктуація яскравості навіть одного пікселя завади здатна повністю маніпулювати сигналом малорозмірної цілі та спровокувати помилкове спрацьовування системи.

1.2.2 Вплив завадового динамічного фону та умов зовнішнього середовища

Ефективність функціонування інтелектуальних систем автоматизованого детектування малорозмірних об'єктів безпосередньо залежить від характеристик оптичного фону, на якому відбувається спостереження. Повітряний простір є нестаціонарним середовищем, що постійно змінює свої фізико-оптичні властивості під впливом природних факторів. Одним із головних джерел завад при цьому виступає хмарність. Рух хмарних мас під дією вітру, постійна трансформація їхньої геометричної форми та внутрішньої текстури створюють складний ефект динамічного фону. На межах хмар виникають значні просторові градієнти яскравості, які детектуються першими згортковими шарами нейронних мереж як потенційні контури цілей. Крім того, мінлива хмарність спричиняє різкі локальні зміни освітленості кадру, що призводить до динамічних коливань яскравості та контрастності всього зображення або його окремих фрагментів. Сонячні відблиски, які виникають при прямому потраплянні променів на оптичні елементи камери або при віддзеркаленні від крапель

води в атмосфері, створюють зони локального оптичного засвічення. У таких зонах інформація про колір та корисну яскравість повністю втрачається, що унеможлиблює виявлення малорозмірної цілі стандартними амплітудними методами.

Серйозним викликом для інтелектуальних відеосистем є наявність у кадрі рухомих біологічних об'єктів, насамперед птахів. Фізичні розміри птахів на середніх відстанях та характер їхньої кінематики польоту мають високий ступінь подібності до геометричних та динамічних характеристик малорозмірних безпілотних літальних апаратів. На великій відстані, коли об'єкт спостереження займає площу лише у кілька пікселів, згортовка нейронна мережа втрачає здатність виділяти унікальні семантичні ознаки для надійної диференціації класу малорозмірних БПЛА від представників орнітофауни. Це призводить до критичного зростання частоти хибних спрацьовувань системи. Слід зазначити, що задача детектування та просторово-часового аналізу поведінки мікрооб'єктів у щільних динамічних потоках є алгоритмічно тотожною задачі виявлення малорозмірних повітряних цілей на складному фоні та їхньої диференціації від біологічних завад, що підтверджує фундаментальну цінність розробки концепцій нейро-детермінованого відстеження [5]. Додаткові динамічні завади виникають при моніторингу повітряного простору поблизу лінії горизонту, де в поле зору оптичного сенсора потрапляють крони дерев або інші елементи рослинності. Їхнє коливання під дією вітру генерує високочастотний візуальний шум, який класичні алгоритми сегментації руху та віднімання фону помилково інтерпретують як переміщення цілі на низькій висоті.

Метеорологічні умови та атмосферні явища чинять комплексний деструктивний вплив на процес формування зображення в оптико-електронних системах. Такі фактори, як туман, серпанок, дощ, сніг або промисловий смог, призводять до інтенсивного розсіювання та поглинання світлових хвиль в шарах атмосфери. Внаслідок цього суттєво знижується дальність прямої видимості та падає загальний просторовий контраст

зображення. Малорозмірний об'єкт, який спочатку має низьку інформативність, за умов туману або серпанку практично повністю зливається з фоном, оскільки різниця між яскравістю цілі та навколишнього простору стає меншою за поріг чутливості матриці камери. Атмосферні опади у вигляді дощу чи снігу створюють специфічні високодинамічні завади у вигляді безлічі рухомих штрихів або точок на кожному кадрі відеопотоку. Ці елементи вносять випадкові просторово-часові флуктуації в оптичний сигнал, що не лише маскує корисний сигнал від реальної повітряної цілі, а й перевантажує обчислювальні модулі системи через необхідність фільтрації великої кількості шумових екстремумів.

1.3 Актуальність роботи та обґрунтування розробки

Актуальність теми дослідження зумовлена стрімким збільшенням інтенсивності використання малорозмірних безпілотних літальних апаратів (БПЛА) як у цивільній, так і в оборонній сферах. Масштабування застосування комерційних квадрокоптерів для моніторингу, логістики або розвідки створює серйозні виклики для систем безпеки об'єктів критичної інфраструктури, військових позицій, транспортних вузлів та приватних територій. Своєчасне виявлення таких повітряних загроз є першочерговою задачею для запобігання потенційним ризикам.

Традиційні засоби виявлення повітряних цілей мають низку суттєвих обмежень. Радіолокаційні системи (РЛС) виявляються малоефективними проти мікро-БПЛА через їхню малу ефективну площу розсіювання та здатність рухатися на наднизьких висотах, де корисний сигнал повністю зливається з відбиттями від рельєфу місцевості. Акустичні датчики демонструють низьку дальність дії та критично залежать від рівня навколишнього індустріального шуму та погодних умов (вітру). У таких умовах оптико-електронні системи спостереження та методи

комп'ютерного зору стають найбільш точним і перспективним каналом отримання інформації.

Проте ручний візуальний моніторинг відеопотоку оператором супроводжується високим психофізіологічним навантаженням і швидкою втомою, що різко знижує ймовірність фіксації загрози. Автоматизація цього процесу за допомогою базових алгоритмів комп'ютерного зору стикається з фундаментальною науково-технічною проблемою: віддалена малорозмірна ціль займає на кадрі менше 0.01% площі, не має чіткої текстури чи форми і відображається у вигляді слабкоконтрастної групи пікселів. Стандартні інтелектуальні системи сприймають такий об'єкт як цифровий шум матриці або розмивають його на глибоких шарах згортки, а наявність динамічних природних завад (рухомих хмар, світлових відблисків та птахів зі схожою кінематикою польоту) провокує критичну кількість хибних спрацьовувань.

Отже, розробка інтелектуальної системи, здатної стабільно та автономно детектувати малорозмірні повітряні об'єкти в умовах складного нестационарного фону та вібрацій сенсора в режимі реального часу, є високоактуальною інженерною задачею, яка безпосередньо впливає на підвищення рівня безпеки та захищеності об'єктів.

1.4 Мета роботи та постановка задачі

Метою кваліфікаційної роботи є розробка, програмна реалізація та експериментальна верифікація інтелектуальної системи автоматизованого детектування малорозмірних повітряних цілей у відеопотоках, яка забезпечує підвищення точності виявлення мікрооб'єктів та зниження частоти хибних спрацьовувань в умовах складного динамічного фону та його руху сенсора в режимі реального часу.

Можна виділити два головних етапи.

Модернізація архітектури згорткової нейронної мережі шляхом інтеграції додаткових шарів мікродетекції високої просторової роздільної

здатності та механізмів координатної уваги (Coordinate Attention) для збереження і підсилення слабких просторових ознак малорозмірних цілей.

Створення оптимізованого програмного комплексу, адаптованого для функціонування на периферійних обчислювальних пристроях із обмеженими ресурсами (Edge AI) для забезпечення обробки потокового відео без затримок.

Для досягнення визначеної мети загальна інженерна задача розробки системи декомпонується на такі послідовні задачі:

- провести детальний аналіз науково-технічної літератури, сучасних методів глибинного навчання та алгоритмів комп'ютерного зору; визначити причини втрати інформації про мікрооб'єкти на різних етапах обробки відеосигналу;

- розробити структурно-функціональну схему інтелектуальної системи; обґрунтувати вибір базової нейромережевої моделі, математичного апарату для побудови механізмів уваги, а також визначити комплекс метрик оцінювання якості;

- сформувати спеціалізований набір відеоданих (датасет), що містить сценарії руху малорозмірних повітряних об'єктів; виконати його фільтрацію, ручну розмітку, а також застосувати методи аугментації даних (зміна яскравості, додавання цифрового шуму, імітація вібрації камери) для підвищення узагальнювальної здатності моделі;

- здійснити проектування та написання програмного коду архітектури системи мовою Python із використанням сучасних фреймворків глибинного навчання; реалізувати конвеєр попередньої обробки та стабілізації вхідного відеопотоку;

- провести процедуру навчання модернізованої нейромережевої моделі на сформованому датасеті; виконати оптимізацію функцій втрат (Loss functions) та підбір гіперпараметрів для запобігання ефекту перенавчання (overfitting).

– виконати тестування розробленої інтелектуальної системи на контрольних відеопотоках; розрахувати кількісні та якісні показники ефективності, провести порівняльний аналіз отриманих результатів із базовою архітектурою та сформулювати висновки щодо практичної придатності системи.

На основі проведеного порівняльного аналізу визначено, що класичні детерміновані методи комп'ютерного зору мають незадовільну завадостійкість, а двоетапні нейромережеві фреймворки є надто обчислювально складними для систем реального часу. Водночас базові одноетапні детектори сімейства YOLO демонструють високий FPS, але призводять до втрати просторової інформації та пропусків цілей розміром менше 16×16 пікселів на глибоких шарах субдискретизації.

Під час дослідження специфіки середовища ідентифіковано ключові деструктивні фактори предметної галузі, серед яких: вкрай низький просторовий контраст віддалених об'єктів відносно неба, семантична порожнеча їхніх текстурних ознак, наявність динамічних природних завад у вигляді хмар і птахів зі схожою кінематикою польоту, а також негативний вплив его-руху та вібрацій самого оптичного сенсора.

Встановлено, що ручний моніторинг відеопотоку оператором супроводжується надмірним психофізіологічним навантаженням, що підтверджує високу актуальність автоматизації цього процесу. За результатами аналізу обґрунтовано доцільність обрання одноетапної архітектури YOLOv8 як базової моделі для подальшого концептуального проєктування. Визначено інженерну необхідність її структурної модернізації шляхом інтеграції механізмів координатної уваги, створення додаткового четвертого шару мікродетекції високої роздільної здатності P2, а також розробки асинхронного конвеєра з модулем цифрової стабілізації зображення на основі алгоритмів ORB та RANSAC.

2 ОБҐРУНТУВАННЯ ВИБОРУ ТЕХНОЛОГІЙ ТА МЕТОДІВ ВИЯВЛЕННЯ ПОВІТРЯНИХ ЦІЛЕЙ

2.1 Проблема его-руху та вібрації сенсорів спостереження

Поняття его-руху (ego-motion) в контексті інтелектуальних систем відеомоніторингу описує власний рух реєстратора даних (оптичного сенсора або відеокамери) відносно навколишнього середовища. У реальних умовах експлуатації комплексів автоматизованого виявлення повітряних цілей камери рідко залишаються абсолютно нерухомими. Навіть стаціонарно встановлені оптико-електронні станції піддаються постійним просторовим зміщенням. Основними джерелами таких зсувів є вітрові навантаження на високі щогли або опори, коливання будівельних конструкцій, а також робота механічних приводів платформ панорамування та нахилу. Якщо ж система детектування розгортається на мобільній або рухомій платформі, наприклад, на базі автомобільного відеореєстратора або патрульного транспортного засобу, то характер его-руху стає складнокоординатним і поєднує в собі лінійні переміщення, кутові коливання та гармонійні вібрації від роботи двигуна та нерівностей дорожнього покриття.

Деструктивний вплив коливань сенсора на якість відеопотоку можна розділити на дві категорії залежно від частоти та амплітуди рухів: низькочастотні зміщення великої амплітуди та високочастотні вібрації малої амплітуди. Низькочастотний его-рух призводить до глобального зсуву всього кадру у просторі. При цьому координати всіх статичних елементів фону (хмар, лінії горизонту, віддалених споруд) змінюються від кадру до кадру. Високочастотна вібрація сенсора чинить ще більш згубний вплив, оскільки вона викликає динамічне змазування зображення (motion blur). Внаслідок швидкого мікропереміщення оптичної осі камери протягом часу експозиції одного кадру, світлова енергія від малорозмірного об'єкти

розподіляється по більшій площі матриці. Як наслідок, просторова функція розсіювання точки суттєво розширюється, що призводить до різкого падіння локального контрасту цілі та розмивання її меж. Об'єкт, який і без того займав лише кілька пікселів, стає практично невідлічним від випадкового фонового шуму.

Зміна просторових координат пікселів фону через его-рух повністю руйнує математичні припущення, на яких базуються класичні та найбільш швидкодіючі методи комп'ютерного зору для детектування руху. Алгоритми віднімання фону (Background Subtraction), які підтримують динамічну модель статичного середовища, починають сприймати будь-який зсув камери як появу руху по всій площі зображення. Це призводить до критичного зростання помилкових спрацьовувань, коли система сигналізує про наявність тисяч рухомих об'єктів. Аналогічно, методи міжкадрової диференціації (Frame Differencing) за умов якщо камера рухається генерують високі значення різниці яскравостей на всіх контрастних межах фону (наприклад, на краях хмар або ліній споруд), повністю маскуючи корисний сигнал від малорозмірної повітряної цілі. [6] Методи обчислення оптичного потоку (Optical Flow) в таких умовах фіксують складне поле векторів глобального руху, що вимагає залучення додаткових обчислювальних ресурсів для відокремлення власних векторів руху цілі від векторів его-руху камери.

Для сучасних інтелектуальних систем на основі згорткових нейронних мереж его-рух та вібрації також виступають критичними завадами. Хоча глибокі нейромережі є інваріантними до просторового зсуву об'єктів у межах кадру, поява ефекту змазування через вібрацію різко знижує ймовірність правильної класифікації. Оскільки малорозмірна ціль позбавлена чітких семантичних ознак, погіршення її різкості навіть на один піксель призводить до того, що нейромережа перестає ідентифікувати її як об'єкт інтересу. Для подолання цієї проблеми виникає інженерна необхідність впровадження у програмний комплекс модулів попередньої

цифрової стабілізації зображення або алгоритмів глобальної компенсації руху [7] (Global Motion Compensation). Ці алгоритми виконують оцінку матриці гомографії між сусідніми кадрами на основі пошуку та зіставлення ключових точок фону, після чого здійснюють зворотне афінне перетворення для вирівнювання кадру [8]. Проте інтеграція таких процедур суттєво підвищує обчислювальну складність системи, створюючи додаткове навантаження на апаратну частину і ускладнюючи виконання вимог щодо обробки відеопотоку в режимі реального часу на периферійних обчислювачах.

2.2 Аналіз класичних методів комп'ютерного зору для виявлення рухомих об'єктів

2.2.1 Методи віднімання фону (Background Subtraction)

Методи віднімання фону (Background Subtraction) належать до класичних і найбільш поширених підходів у комп'ютерному зорі для виявлення рухомих об'єктів у відеопотоках. Основний принцип цих методів полягає в побудові та постійному оновленні математичної моделі статичної сцени (фону) з подальшим порівнянням кожного нового кадру з цією моделлю. Пікселі, значення інтенсивності або кольору яких суттєво відрізняються від фонових значень, класифікуються як передній план (foreground), що свідчить про наявність руху. Головною перевагою даної групи алгоритмів є висока обчислювальна ефективність, що дозволяє реалізувати їх у системах моніторингу, які працюють у режимі реального часу, без залучення значних апаратних ресурсів. Це робить їх первинним інструментом для швидкої локалізації потенційних цілей на початкових етапах обробки відеосигналу.

Серед усього різноманіття алгоритмів віднімання фону особливе місце посідають методи, засновані на статистичному моделюванні кожного

пікселя. Найбільш відомим є метод суміші Гауссових розподілів (Mixture of Gaussians, MOG2). У цьому підході історія змін інтенсивності кожного окремого пікселя протягом певної кількості попередніх кадрів моделюється за допомогою суміші кількох Гауссових розподілів. Це дозволяє алгоритму адаптуватися до складних умов сцени, таких як періодичні коливання елементів фону (наприклад, рух листя на деревах або хвилювання на воді). Іншим ефективним підходом є використання методу К-найближчих сусідів (K-Nearest Neighbors, KNN) для оцінки щільності розподілу фону. Метод KNN є непараметричним і зберігає певну кількість останніх значень пікселя, що забезпечує швидшу адаптацію до різких змін освітленості сцени та дозволяє точніше сегментувати об'єкти, які зупиняються або починають рух після тривалої паузи.

Проте застосування класичних методів віднімання фону в задачах детектування малорозмірних повітряних цілей супроводжується серйозними обмеженнями, зумовленими низькою інформативністю самих об'єктів. Оскільки віддалений повітряний об'єкт займає на кадрі площу всього у кілька пікселів, його яскравіший сигнал є вкрай слабким. У разі встановлення низького порогу чутливості для виявлення таких малококонтрастних об'єктів, алгоритм починає помилково реагувати на випадковий цифровий шум матриці оптичного сенсора, створюючи велику кількість хибних бінарних маркерів. З іншого боку, якщо поріг бінаризації є занадто високим, або швидкість оновлення моделі фону налаштована некоректно, малорозмірна ціль, що рухається з невеликою кутовою швидкістю, швидко поглинається самою моделлю фону і стає невидимою для системи детектування.

Додатковим деструктивним фактором є нестаціонарність умов зовнішнього середовища, яка призводить до руйнування побудованої моделі сцени. Рух хмарних мас, поява тіней, різка зміна інтенсивності сонячного світла або виникнення атмосферних опадів викликають глобальні або локальні флуктуації яскравості фону. Оскільки статистичні моделі

MOG2 або KNN потребують певного часу для перерахунку своїх параметрів, будь-яка динамічна зміна в атмосфері сприймається алгоритмом як поява нових рухомих об'єктів переднього плану. Це призводить до масового утворення хибних контурів по всій площі кадру, що повністю маскує реальну малорозмірну повітряну ціль. Таким чином, хоча методи віднімання фону і забезпечують високу швидкість обробки даних, їх самостійне використання в неконтрольованих природних умовах є недостатнім і потребує комбінування з більш стійкими інтелектуальними методами аналізу просторових ознак.

2.2.2 Методи аналізу оптичного потоку (Optical Flow)

Методи аналізу оптичного потоку (Optical Flow) становлять ще одну важливу групу класичних підходів у комп'ютерному зорі, яка використовується для виявлення та відстеження рухомих об'єктів у відеопослідовностях. Оптичний потік визначається як поле видимого руху яскравості елементів зображення між двома послідовними кадрами. На відміну від методів віднімання фону, які фокусуються на статистичному аналізі яскравості окремих пікселів у часі, аналіз оптичного потоку встановлює просторово-часові зв'язки між суміжними кадрами, обчислюючи вектори зміщення для точок зображення. Математичною основою більшості алгоритмів оптичного потоку є припущення про сталість яскравості рухомої точки протягом малого проміжку часу, що описується формулою (1.1).

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t). \quad (2.1)$$

За способом розрахунку векторного поля методи оптичного потоку поділяють на дві основні категорії: розріджені (sparse) та щільні (dense). Розріджені методи, серед яких алгоритм Лукаса-Канаде є найбільш

відомим, обчислюють вектори зміщення лише для обмеженої кількості характерних точок кадру, що мають високий просторовий градієнт (наприклад, кути або висококонтрастні межі). Це забезпечує високу швидкість обчислень та мінімальні вимоги до апаратної частини, роблячи розріджений потік придатним для систем, що працюють у реальному часі. Щільні методи, такі як алгоритм Фарнебека або метод Горна-Шунка, розраховують вектор зміщення для кожного пікселя зображення. Вони генерують детальну двовимірну карту руху всієї сцени, що дозволяє отримувати вичерпну інформацію про динаміку середовища, проте супроводжується значними обчислювальними витратами.

Стосовно задачі детектування малорозмірних повітряних цілей, аналіз оптичного потоку має вагому перевагу: він дозволяє виявляти об'єкти за умов відсутності чітких семантичних та текстурних ознак, оперуючи виключно кінематичними характеристиками групи пікселів. Оскільки малорозмірний БПЛА на великій відстані виглядає як однорідна пляма, зміщення цієї плями на фоні статичного або повільно змінного неба генерує локальну аномалію у векторному полі оптичного потоку, яку можна зафіксувати алгоритмами порогової фільтрації.

Попри свої теоретичні переваги, практичне застосування оптичного потоку для виявлення малорозмірних об'єктів обмежене декількома критичними факторами. Першим із них є висока чутливість до зміни умов освітленості та цифрового шуму матриці камери. Флуктуації яскравості, викликані сонячними відблисками або шумами сенсора, призводять до порушення умови сталості яскравості, внаслідок чого алгоритми генерують хаотичні хибні вектори руху. Другим, ще більш вагомим деструктивним фактором, є его-рух або вібрація самої камери спостереження. Під час коливання сенсора зміщуються всі пікселі кадру, що призводить до формування складного глобального поля векторів руху. На фоні такого глобального зсуву виокремити локальний вектор руху малорозмірної цілі, яка займає площу всього у кілька пікселів, стає надзвичайно складною

математичною задачею, що потребує попередньої компенсації руху сцени та суттєво знижує загальну швидкодію інтелектуальної системи [9].

2.2.3 Обмеження традиційних підходів у задачах мікродетекції

Системний аналіз класичних підходів до обробки зображень свідчить про їхню обмежену спроможність під час розв'язання задач мікродетекції. Головною причиною непридатності традиційних просторових та частотних фільтрів є їхня лінійна природа, яка не дозволяє ефективно розділяти корисний сигнал та завади за умов низького контрасту повітряної цілі. Традиційні оператори виділення меж, такі як фільтри Собіля, Кенні або Лапласіана, базуються на пошуку локальних екстремумів градієнта яскравості. Для малорозмірного об'єкта, яскравіший профіль якого сильно розмитий через атмосферні явища, значення цих градієнтів є порівнянними із цифровим шумом матриці сенсора. Як наслідок, спроби порогової фільтрації призводять або до повної втрати об'єкта, або до генерації великої кількості хибних маркерів переднього плану.

Ситуація критично ускладнюється за умов его-руху та вібрації камери спостереження. Оскільки традиційні геометричні та амплітудні методи не володіють властивістю інваріантності до динамічної зміни фону, будь-яке тремтіння сенсора руйнує просторово-часову зв'язність кадрів. Застосування алгоритмів стабілізації на основі афінних перетворень частково компенсує глобальний зсув, проте високочастотне змазування зображення залишається непереборним бар'єром для лінійних фільтрів [10]. Таким чином, класичні підходи демонструють незадовільну завадостійкість і точність, що зумовлює необхідність переходу до нелінійних методів вилучення ознак на основі штучних нейронних мереж. Зокрема, глибокі архітектури здатні автоматично формувати високорівневі ієрархічні дескриптори, що є стійкими до локальних оптичних деформацій та мікророзмиття контурів об'єктів.

2.3 Огляд сучасних підходів на основі глибинного навчання

Перехід від класичних алгоритмів комп'ютерного зору до методів глибинного навчання став визначальним етапом у розвитку систем автоматизованого детектування об'єктів. Використання штучних нейронних мереж дозволяє відмовитися від ручного проектування дескрипторів ознак та перейти до автоматичного наскрізного вилучення високорівневих семантичних характеристик безпосередньо з відеопотоку. Сучасні підходи глибинного навчання володіють значно вищою узагальнювальною здатністю, що забезпечує стабільну роботу систем в умовах нестаціонарного середовища, мінливої освітленості та складного динамічного фону. Основним інструментом у цій галузі виступають згорткові нейронні мережі, архітектура яких оптимізована для пошуку просторових закономірностей на зображеннях, а також їхні модифікації, спрямовані на обробку часових послідовностей відеопотоку.

2.3.1 Одноетапні та двоетапні згорткові нейронні мережі

У сучасних архітектурах згорткових нейронних мереж для детектування об'єктів виділяють два основні підходи: одноетапний підхід (Single-stage) та двоетапний (Two-stage) [11]. Двоетапні моделі, яскравим представником яких є архітектура Faster R-CNN, на першому кроці генерують гіпотези щодо потенційного розташування об'єктів за допомогою спеціальної мережі пропозицій регіонів, а на другому – виконують їхню класифікацію та уточнення обмежувальних контурів. Такий підхід забезпечує високу точність локалізації, проте має низьку швидкодію, що унеможлиблює його використання в реальному часі. Одноетапні детектори, такі як моделі сімейства YOLO та SSD, об'єднують процеси локалізації та класифікації в один етап, прогнозуючи координати та

класи об'єктів за один прохід мережі. Це гарантує високу швидкість обробки кадрів, але супроводжується зниженням точності для малих об'єктів.

Спільною фундаментальною проблемою обох підходів під час мікродетекції є втрата просторової інформації на глибоких шарах згортки. Через послідовні операції зменшення розмірності кадру ознаки малорозмірних цілей, які займають кілька пікселів, повністю нівелюються і заміщуються семантикою великих фонових об'єктів. Це вимагає модернізації архітектур шляхом створення додаткових розгалужень для збереження низькорівневих ознак високої просторової роздільної здатності.

2.3.2 Особливості застосування механізмів уваги

Для вирішення проблеми втрати інформації про малорозмірні об'єкти в сучасних нейромережевих архітектурах активно застосовуються механізми уваги (Attention Mechanisms). Основна задача цих блоків полягає в динамічному перерозподілі ваг ознак, що дозволяє мережі фокусуватися на найбільш інформативних ділянках кадру та ігнорувати завадовий фон. Блок каналної уваги SE виконує оцінку взаємозв'язків між різними каналами ознак, підсилюючи важливі фільтри та пригнічуючи незначущі. Комплексний блок СВМ поєднує в собі каналну та просторову увагу, обчисливши карту уваги вздовж просторових координат, що допомагає мережі точніше локалізувати об'єкт.

Проте найвищу ефективність у задачах мікродетекції демонструє механізм координатної уваги (Coordinate Attention) [11]. На відміну від стандартної просторової уваги, яка згладжує координатні зв'язки, координатна увага інтегрує інформацію про напрямок руху та просторове розташування шляхом агрегування ознак за горизонтальною та вертикальною вісями окремо. Це дозволяє згортковій мережі точно фіксувати координати малорозмірної повітряної цілі, оперуючи навіть

мінімальними групами пікселів на складному фоні хмар чи цифрового шуму.

2.3.3 Моделі аналізу просторово-часових ознак у відеопослідовностях

Специфіка детектування повітряних цілей у відеопотоках полягає в тому, що поодинокий кадр часто не містить достатньо інформації для надійного виявлення об'єкта через його низький контраст і малий розмір. Для підвищення завадостійкості систем використовуються архітектури, які аналізують просторово-часові ознаки відеопослідовностей, враховуючи фактор траєкторії руху об'єкта [12]. Одним із поширених підходів є комбінування згорткових нейронних мереж із рекурентними структурами, зокрема архітектурою CNN-LSTM. У такій системі згорткові шари відповідають за вилучення просторових ознак з кожного окремого кадру, а блоки довгострокової короткочасної пам'яті аналізують зміни цих ознак у часі, фіксуючи динаміку переміщення об'єкта.

Іншим ефективним рішенням є використання тривимірних згорткових нейронних мереж (3D-CNN). На відміну від класичних двовимірних згорткових шарів, тривимірна згортка виконує операції зміщення одночасно за просторовими координатами кадру та вздовж часової осі пакету кадрів. Це дозволяє моделі автоматично вилучати як статичні семантичні ознаки, так і динамічні характеристики швидкості та напрямку руху цілі, що суттєво знижує ймовірність хибних спрацьовувань на нерухомі або хаотично мерехтливі елементи завадового фону.

2.4 Метрики оцінювання якості та ефективності інтелектуальних систем детектування

Для об'єктивного аналізу, порівняння та підтвердження працездатності розроблюваних алгоритмів у складі інтелектуальних систем

автоматизованого детектування малорозмірних повітряних цілей використовується комплексний набір метрик. Ці критерії дозволяють оцінити систему з двох основних точок зору: якості виконання функцій локалізації та класифікації об'єктів у кадрі (метрики точності), а також швидкодії та використання обчислювальних ресурсів апаратної платформи (метрики ефективності). Спільне застосування обох категорій параметрів є обов'язковим для систем, що функціонують у режимі реального часу, оскільки досягнення високої просторової точності не має практичного сенсу у разі критичного падіння швидкості обробки вхідного відеопотоку.

Основою для розрахунку більшості метрик точності в задачах комп'ютерного зору є коефіцієнт просторового перетину множин, відомий як метрика перетину над об'єднанням (Intersection over Union, IoU). Дана метрика визначає ступінь геометричного збігу між обмежувальним контуром, що був спрогнозований нейронною мережею, та еталонним контуром цілі, розміченим вручну у тестовому наборі даних. На основі встановленого порогового значення коефіцієнта IoU всі результати роботи детектора класифікуються на три основні категорії: істинно позитивні спрацьовування (True Positives, TP), коли об'єкт виявлено і його координати збігаються з еталоном; хибно позитивні спрацьовування (False Positives, FP), що свідчать про виявлення цілі там, де вона насправді відсутня (помилка на заводовому фоні); та хибно негативні результати (False Negatives, FN), які фіксують факт пропуску реальної повітряної цілі інтелектуальною системою.

Для узагальненої оцінки класифікаційних та локалізаційних здатностей моделі обчислюються метрики точності (Precision) та повноти (Recall). Точність визначає частку реальних повітряних цілей серед усіх об'єктів, які система ідентифікувала як цілі. Високе значення цієї метрики свідчить про низьку частоту хибних спрацьовувань системи на птахів, рухомі хмари, світлові відблиски чи цифрові шуми сенсора. Повнота відображає здатність системи знайти всі існуючі малорозмірні об'єкти у

відеопотоці, і її високе значення вказує на мінімальну ймовірність пропуску повітряної загрози. Оскільки між точністю та повнотою існує зворотна залежність, для їхнього збалансованого оцінювання використовується метрика Ф-міри (F1-Score), яка обчислюється як гармонійне середнє між ними і дозволяє розробнику знайти оптимальний поріг чутливості детектора.

Найбільш інформативною та комплексною метрикою для оцінювання інтелектуальних моделей є середня точність (Average Precision, AP), яка розраховується як площа під кривою залежності точності від повноти. Для оцінки загальної ефективності системи за різних умов суворості локалізації обчислюється середнє значення середньої точності (mean Average Precision, mAP). У задачах мікродетекції повітряних об'єктів критично важливими є два варіанти цієї метрики. Перший варіант оцінює загальну здатність системи виявляти цілі за лояльного порогу збігу координат, тоді як другий варіант відображає стабільність та високу точність локалізації контурів об'єкта за різних рівнів суворості, що дозволяє детально проаналізувати поведінку мережі під час зміни відстані до об'єкта спостереження чи погіршення умов видимості.

Ефективність функціонування інтелектуальної системи в реальних умовах експлуатації визначається метриками часової складності та швидкодії алгоритмів. Основним показником тут виступає частота обробки кадрів (Frames Per Second, FPS), яка вказує на кількість кадрів відеопотоку, яку алгоритм здатний проаналізувати за одну секунду на конкретній апаратній платформі. Для забезпечення роботи системи в режимі реального часу значення FPS має бути не меншим за частоту реєстрації кадрів камери. Додатковою метрикою є час затримки (Latency), що вимірює чистий час обробки одного кадру обчислювальними модулями від моменту його отримання до виведення результату. Оцінка швидкодії є критично важливою під час проектування систем периферійних обчислень (Edge AI), де висока точність нейромережевих моделей повинна супроводжуватися

їхньою низькою математичною складністю та оптимізованим використанням ресурсів пам'яті графічного або центрального процесора.

Для автоматизації процесу виявлення та супроводження повітряних об'єктів у сучасному комп'ютерному зорі сформувалося кілька основних програмно-технологічних підходів. Вони суттєво відрізняються за своєю математичною основою, вимогами до апаратної частини та здатністю функціонувати в умовах неконтрольованого зовнішнього середовища.

Першу категорію становлять рішення на базі класичних бібліотек комп'ютерного зору, таких як OpenCV або MATLAB. Вони базуються на детермінованих алгоритмах віднімання фону (зокрема статистичних моделях MOG2, KNN) та аналізу оптичного потоку (алгоритми Фарнебека, Лукаса-Канаде). Головною перевагою цього підходу є мінімальна обчислювальна складність, що дозволяє виконувати програми на центральних процесорах (CPU) низької потужності без використання дискретних графічних прискорювачів.

Крім того, такі системи не потребують збору та розмітки великих навчальних наборів даних. Проте суттєвим недоліком є критична чутливість до его-руху (тремтіння) камери, що призводить до повного руйнування моделі фону. Класичні фільтри також характеризуються семантичною порожнечою – вони фіксують будь-яку зміну яскравості пікселів, не маючи змоги відрізнити цільовий БПЛА від біологічних завад (птахів) або рухомих хмар.

Другу групу формують програмні рішення на основі двоетапних (Two-stage) нейромережових фреймворків, яскравими представниками яких є архітектури Faster R-CNN та Cascade R-CNN. Перевагою таких систем є висока точність локалізації об'єктів за рахунок попереднього виділення регіонів інтересу за допомогою спеціалізованої підмережі. Це дозволяє досягати стабільних показників розпізнавання за умов однорідного фону.

Водночас головним недоліком двоетапних моделей виступає їхня надмірна математична складність. Велика затримка обробки кожного кадру призводить до зниження частоти кадрів (FPS менше 15), що робить їх непридатними для моніторингу повітряного простору в режимі реального часу, а також висуває високі вимоги до енергоспоживання та об'єму відеопам'яті серверних GPU.

Третім підходом є використання одноетапних (Single-stage) нейромережевих детекторів реального часу сімейств YOLO та SSD. Перевагою цих інструментів є наскрізне прогнозування координат обмежувальних контурів та класів об'єктів за один прохід мережі, завдяки чому досягається висока швидкодія (від 30 до 100+ FPS). Програмні моделі цього типу легко піддаються оптимізації та конвертації у специфічні формати (TensorRT, ONNX) для апаратного прискорення на периферійних обчислювачах.

Головним недоліком базових одноетапних архітектур є незадовільна якість мікродетекції. Через послідовні операції субдискретизації на глибоких шарах згортки інформація про малорозмірні цілі (розміром менше 16x16 пікселів) повністю нівелюється, що призводить до пропусків об'єктів на великих відстанях та помилкової класифікації дронів як птахів.

Окрему категорію становлять спеціалізовані комерційні платформи комплексного моніторингу (C-UAS системи). Їхньою перевагою є мультисенсорна інтеграція даних (об'єднання оптичного, тепловізійного та радіолокаційного каналів) та наявність відпрацьованих промислових фільтрів. Недоліками виступають повністю закритий вихідний код, висока вартість ліцензування та неможливість гнучкої програмної модифікації алгоритмів для роботи з нестандартними оптичними сенсорами.

Таким чином, системний аналіз існуючих технологій обґрунтовує доцільність обрання одноетапної архітектури сімейства YOLO як базової з її подальшою науково-технічною модернізацією шляхом інтеграції механізмів уваги та оптимізації шарів вилучення мікроознак.

2.5 Обґрунтування вибору базової моделі глибинного навчання та інструментальних засобів розробки

На основі аналізу, проведеного в першому розділі, для побудови інтелектуальної системи автоматизованого детектування малорозмірних повітряних цілей було обрано одноетапну архітектуру нейронних мереж сімейства YOLO, а саме її сучасну стабільну модифікацію YOLOv8. Вибір даної моделі зумовлений її високою ефективністю за критерієм балансу між точністю локалізації об'єктів та швидкістю обробки відеопотоку в режимі реального часу. Архітектура YOLOv8 використовує концепцію anchor-free детекторів, що дозволяє безпосередньо прогнозувати центр об'єкта та його розміри, суттєво зменшуючи кількість математичних операцій порівняно з попередніми версіями [19], які спиралися на фіксовані рамки-якорі. Крім того, наявність відкритої екосистеми та оптимізованих конвеєрів навчання робить цю модель гнучкою платформою для подальшої інженерної модернізації та інтеграції специфічних кастомних шарів.

У ролі основного інструментального середовища для проектування та програмної реалізації системи обрано мову програмування Python, яка є де-факто стандартом у галузі штучного інтелекту та комп'ютерного зору. Для розробки, навчання та тестування нейромережових архітектур обґрунтовано використання фреймворку глибинного навчання PyTorch. Дана бібліотека надає розробнику можливість динамічного обчислення графів, що спрощує процес налагодження нетипових шарів уваги та експериментального підбору гіперпараметрів мережі. Для вирішення задач попередньої обробки відеопотоку, декодування кадрів у реальному часі та виконання просторових перетворень залучено кросплатформову бібліотеку OpenCV, яка володіє високою швидкодією завдяки оптимізації алгоритмів під процесорні архітектури.

Особливе місце в інструментарії системи посідають засоби апаратного прискорення та конвертації моделей. Оскільки фінальний

програмний комплекс має функціонувати на периферійних пристроях з обмеженими обчислювальними ресурсами, обґрунтовано використання технології NVIDIA TensorRT та відкритого формату обміну моделями ONNX. Конвертація навченої моделі PyTorch у формат TensorRT дозволяє виконати оптимізацію графу нейромережі, квантування ваг з плаваючою точкою (FP32) до цілочисельного формату (INT8) або зниженої точності (FP16), а також оптимізувати використання кеш-пам'яті графічного процесора, що забезпечує значний приріст частоти обробки кадрів без суттєвої втрати точності мікродетекції [13].

2.6 Концепція та математичний опис механізму координатної уваги в задачах мікродетекції

У задачах автоматизованого виявлення малорозмірних повітряних цілей критично важливим є не лише визначення наявності об'єкта в потоці ознак, а й точне збереження його просторових координат. Традиційні механізми уваги, такі як блоки селекції каналів (Squeeze-and-Excitation), виконують глобальне усереднення просторових координат для обчислення вагових коефіцієнтів каналів. Таке стиснення призводить до повної втрати інформації про просторове розташування об'єкта, що є критичним недоліком під час пошуку цілей розміром у кілька пікселів. Інші підходи, наприклад блоки просторової уваги в архітектурі CBAM, намагаються зберегти просторові характеристики шляхом застосування згортки над стиснутими за каналами картами ознак. Проте цей метод фіксує лише локальні залежності і не здатний моделювати довготривалі просторові зв'язки, що знижує завадостійкість системи на складному фоні.

Для усунення зазначених недоліків у розроблюваній інтелектуальній системі обґрунтовано застосування механізму координатної уваги (Coordinate Attention) [14]. Концепція цього підходу полягає в декомпозиції глобального двовимірного пулінгу на два послідовні

одновимірні процеси агрегування ознак за горизонтальною та вертикальною вісями кадру. Це дозволяє нейронній мережі фіксувати довготривалі залежності вздовж одного просторового напрямку і водночас зберігати точну топологічну інформацію вздовж іншого напрямку. У результаті модель формує пару анізотропних карт уваги, які дозволяють точно локалізувати малорозмірну повітряну ціль без суттєвого збільшення обчислювальної складності алгоритму [22], що є ключовою вимогою для систем периферійного штучного інтелекту.

Математичний алгоритм функціонування блока координатної уваги складається з трьох послідовних етапів: координатного вбудовування інформації, генерації координатної уваги та фінального зважування.

На першому етапі вхідний тензор ознак, що має розмірність $C \times H \times W$ (де C – кількість каналів, H – висота, W – ширина карти ознак), піддається одновимірному глобальному осередненню. Для збору інформації про просторове розташування об'єкта вздовж вертикальної осі використовується одновимірний пулінг із розміром ядра $1 \times W$. Для довільного каналу c результат усереднення на висоті h обчислюється як середнє арифметичне значень інтенсивностей пікселів цього каналу по всій ширині кадру. Паралельно виконується аналогічна операція для збору інформації вздовж горизонтальної осі, де застосовується одновимірний пулінг із розміром ядра $H \times 1$, що дає змогу отримати інтегральне значення ознак для кожної координати ширини w шляхом усереднення значень по всій висоті кадру.

У результаті першого етапу формуються два вектори ознак: один розмірністю $C \times H \times 1$, який кодує просторову інформацію по вертикалі, та другий розмірністю $C \times 1 \times W$, що кодує інформацію по горизонталі. На другому етапі отримані вектори піддаються операції конкатенації вздовж просторової розмірності, утворюючи єдиний трансформований тензор розмірністю $C \times 1 \times (H + W)$. До цього тензора застосовується операція лінійної згортки з розміром ядра 1×1 , яка виконує функцію стиснення

кількості каналів ознак у певну кількість разів відповідно до заданого коефіцієнта редукції. Це стиснення є необхідним для зменшення математичної складності блока та оптимізації використання пам'яті. Після згортки отримані значення обробляються за допомогою нелінійної функції активації h-swish, що дозволяє мережі виділити нелінійні просторові взаємозв'язки між координатами.

Після нелінійної трансформації результуючий тензор знову розділяється вздовж просторової осі на два окремі тензори з початковими просторовими розмірностями: $C/r \times H \times 1$ та $C/r \times 1 \times W$. Для кожного з цих тензорів застосовуються індивідуальні операції згортки з ядром 1×1 , які відновлюють початкову кількість каналів ознак C . Кінцеві вагові коефіцієнти уваги для горизонтального та вертикального напрямків обчислюються шляхом пропускання значень через логістичну функцію активації (сигмоїду), що масштабує отримані величини в діапазоні від 0 до 1. Ці коефіцієнти відображають міру значущості кожної просторової координати кадру для виявлення цілі. На заключному етапі відбувається безпосереднє зважування вхідної карти ознак: кожне вихідне значення пікселя вхідного тензора множиться на відповідні коефіцієнти горизонтальної та вертикальної уваги, що підсилює корисний сигнал від малорозмірного об'єкта та пригнічує фонові завади.

2.7 Модифікація архітектури згорткової нейронної мережі та проектування шарів мікродетекції

Для успішного виявлення малорозмірних повітряних цілей базову архітектуру одноетапного детектора YOLOv8 необхідно піддати суттєвій структурній модернізації. Стандартна модель YOLOv8 складається з трьох основних компонентів:

- backbone-мережі (основи вилучення ознак);

- neck-мережі (модуля агрегування та об'єднання різномасштабних ознак);
- head-модуля (детектуючих головок).

У первинному варіанті архітектура розрахована на детектування об'єктів трьох масштабних категорій (малих, середніх та великих) і використовує для цього три вихідні гілки детектора, які відповідають рівням зниження просторової роздільної здатності кадру в 8, 16 та 32 рази відносно вхідного зображення. Проте для об'єктів надмалого розміру, які займають площу менше ніж 16x16 пікселів, такий ступінь субдискретизації є руйнівним, що зумовлює необхідність перепроєктування архітектури мережі.

Першим етапом модернізації є визначення точок інтеграції розробленого механізму координатної уваги (Coordinate Attention). У ході дослідження було визначено, що найбільш ефективним є впровадження блоків уваги у структуру neck-мережі, яка побудована на основі архітектури PANet (Path Aggregation Network) [15]. Завдання neck-мережі полягає в поєднанні низькорівневих геометричних ознак з початкових шарів мережі із високорівневими семантичними ознаками з глибоких шарів. Інтеграція блоків координатної уваги безпосередньо у внутрішню структуру згорткових модулів C2f на етапі об'єднання ознак дозволяє системі динамічно зважувати просторову інформацію саме тоді, коли формуються комбіновані карти ознак. Блоки уваги інтегруються перед шарами вихідних детекторів для мікрмасштабів та малих масштабів. Це дозволяє нейромережі фокусувати математичні ресурси на локальних просторових аномаліях, пригнічуючи при цьому фонові завади від руху хмар або вібрації оптичного сенсора ще до того, як дані потраплять у детектуючу головку.

Другим етапом структурної модифікації є проектування та створення додаткового четвертого засобу детекції, спеціалізованого на пошуку об'єктів надмалого розміру (шару мікродетекції). Для реалізації цього підходу в архітектуру системи вводиться додатковий рівень обробки з

мінімальним ступенем зниження просторової роздільної здатності, що відповідає масштабу P2 (зниження просторової розмірності лише в 4 рази відносно вхідного кадру). Якщо в стандартній моделі YOLOv8 інформація з другого етапу backbone-мережі не передавалася безпосередньо в модуль детекції, то в розроблюваній архітектурі цей високороздільний потік ознак конкатенується з відповідним вихідним потоком з низхідної гілки neck-мережі.

Третій етап. Для побудови детектуючої головки структура PANet розширюється додатковим висхідним та низхідним шляхами. У низхідній гілці neck-мережі додається ще один крок апсемплінгу (збільшення просторової розмірності) за допомогою білінійної інтерполяції, що дозволяє підняти просторову роздільну здатність карти ознак з рівня P3 до рівня P2. Отриманий тензор об'єднується за каналами з тензором P2, що надходить безпосередньо з початкових шарів вилучення ознак backbone-мережі. Після цього комбінована карта ознак високої роздільної здатності проходить обробку через модернізований модуль C2f з інтегрованим блоком координатної уваги. На фінальному етапі цей потік даних спрямовується на нову, четверту детектуючу головку Decoupled Head.

Реалізація додаткового шару мікродетекції призводить до того, що вихідний head-модуль системи тепер оперує чотирма паралельними гілками. Нова гілка мікродетекції працює з картами ознак великого просторового розміру, де крок згортки становить чотири пікселі. Це гарантує, що повітряна ціль розміром навіть 4×4 або 8×8 пікселів зберігає свою просторову цілісність, межі її контуру не розмиваються суміжними фоновими пікселями, а координатна інформація передається у функцію регресії обмежувальних рамок з мінімальними похибками. Таке інженерне рішення дозволяє суттєво підвищити показники середньої точності детектування малорозмірних безпілотних літальних апаратів на великих відстанях спостереження без критичного збільшення загального часу затримки обробки відеопотоку.

3 АРХІТЕКТУРА СИСТЕМИ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Архітектурна структура та модульний розподіл програмного комплексу

Проект програмної реалізації інтелектуальної системи детектування малорозмірних повітряних цілей базується на принципах модульної декомпозиції, абстракції рівнів обробки даних та забезпеченні максимальної завадостійкості конвеєра обчислень. З огляду на те, що фінальне розгортання системи передбачається на периферійних апаратних платформах із обмеженими обчислювальними ресурсами, концептуальна архітектура спроектована таким чином, щоб ізолювати ресурсомісткі операції нейромережевого аналізу від процесів захоплення відео та пост-обробки результатів. Це дозволяє оптимізувати кожен програмний модуль окремо та забезпечити стабільне функціонування системи в режимі реального часу.

На концептуальному рівні програмний комплекс розділено на чотири взаємопов'язані функціональні підсистеми (рисунок 3.1).

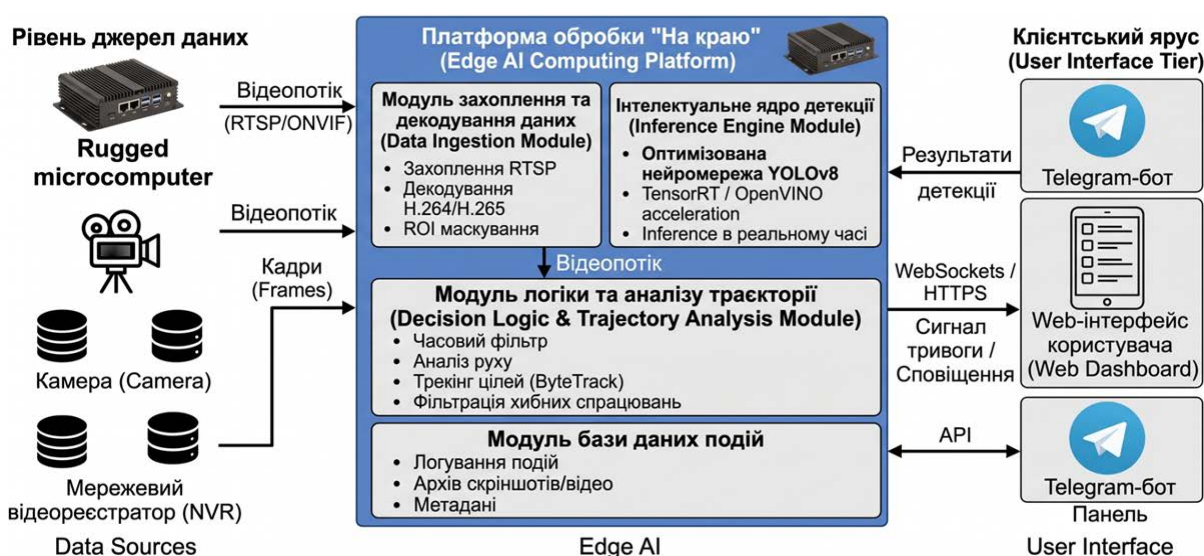


Рисунок 3.1 – Структура системи детектування малих повітряних цілей

При описі топології побудови мережі моніторингу критичної інфраструктури та раціонального розподілу обчислювального навантаження між окремими Edge-платформами обробки на краю «На краю» доцільно використовувати математичний апарат кластеризації сенсорних вузлів на основі модифікованого критерію модулярності [18].

3.1.1 Підсистема інгестії та стабілізації відеопотоку

Підсистема інгестії та стабілізації відеопотоку (Video Ingestion & Stabilization Module). Цей модуль відповідає за взаємодію з апаратними джерелами відеосигналу через програмні інтерфейси бібліотеки OpenCV. Основна концепція модуля полягає в реалізації описаного у другому розділі алгоритму компенсації его-руху на основі дескрипторів ORB та статистичної фільтрації RANSAC.

Програмний підмодуль обчислює матрицю гомографії фону для суміжних кадрів, усуває високачастотне тремтіння сенсора та передає геометрично вирівняні зображення у наступні ланки конвеєра, упереджуючи розмиття контурів малорозмірних об'єктів.

3.1.2 Підсистема нейромережевого аналізу та мікродетекції

Підсистема нейромережевого аналізу та мікродетекції. Дана підсистема становить ядро інтелектуального аналізу. На концептуальному рівні вона керує завантаженням оптимізованого обчислювального графу модифікованої нейронної мережі YOLOv8.

Програмна логіка модуля забезпечує перетворення вхідних масивів пікселів у тензори фреймворку PyTorch, виконання прямого проходу через шари вилучення ознак із інтегрованими кастомними блоками координатної уваги (Coordinate Attention) та передачу первинних прогнозів від чотирьох

детектуючих головок, включаючи доданий четвертий шар надвисокої просторової роздільної здатності P2.

3.1.3 Підсистема просторово-часової пост-обробки та фільтрації

Підсистема просторово-часової пост-обробки та фільтрації (Post-processing & Spatio-temporal Filtering). Вихідні координати обмежувальних рамок із нейромережевого модуля містять надлишкові гіпотези та випадкові шумові екстремуми. Цей модуль концептуально реалізує не лише стандартне програмне придушення немаксимумів (Non-Maximum Suppression, NMS), а й алгоритм міжкадрової валідації траєкторій. Система відстежує вектори зміщення центрів детекцій протягом фіксованого вікна кадрів, що дозволяє програмним шляхом відсікати хаотичні спрацьовувань від комах, листя та цифрових шумів матриці, підтверджуючи тривогу лише за наявності стабільного вектора руху цілі.

3.1.4 Підсистема користувацького інтерфейсу та зовнішньої інтеграції

Підсистема користувацького інтерфейсу та зовнішньої інтеграції (UI & Integration Gateway). Модуль забезпечує кінцеву візуалізацію результатів для оператора системи моніторингу. Він виконує графічне накладання обчислених координат та міток класів на потоковий відеоряд, веде текстовий лог подій і керує генерацією сигналів тривоги у разі входження об'єкта в задану зону охорони периметра. Клієнтський рівень надає інструменти взаємодії оператора з платформою обробки через пригнічені мережеві протоколи HTTP та WebSockets. На цьому рівні функціонує Web-інтерфейс користувача для оперативного моніторингу повітряної обстановки й калібрування параметрів, а також Telegram-бот сповіщень, призначений для миттєвої розсилки тривожних сигналів і знімків зафіксованих загрози.

3.2 Організація асинхронного конвеєра обробки поточкових відеоданих

Для виконання жорстких інженерних вимог щодо забезпечення обробки потоків у режимі реального часу, концепцією програмної реалізації передбачено відмову від послідовного (синхронного) виконання операцій. Оскільки затримки під час зчитування кадрів із камери та виконання нейромережевого інференсу суттєво відрізняються, послідовна архітектура призводила б до простоїв графічного прискорювача та втрати кадрів. Для вирішення цього протиріччя спроектовано асинхронну багатопотокову модель обробки даних (Data Pipeline).

Вся логіка функціонування системи розподіляється між трьома паралельними обчислювальними потоками (Threads), які взаємодіють через ізольовані черги (Queues) в оперативній пам'яті.

Потік захоплення та стабілізації (Producer Thread). Працює в нескінченному циклі із частотою, яка суворо синхронізована із частотою кадрів вхідної камери. Він виконує операції зчитування, цифрової стабілізації зображення засобами OpenCV та поміщає оброблені матриці у першу вхідну чергу фіксованого розміру. Якщо черга заповнена, старі кадри витісняються, що запобігає накопиченню затримок у системі.

Потік нейромережевого інференсу (Inference Thread). Фокусується виключно на роботі з графічним прискорювачем. Він асинхронно вилучає стабілізовані кадри з першої черги, формує батч даних, переносить їх у відеопам'ять GPU, викликає виконання оптимізованого обчислювального графу моделі TensorRT та записує сирі координати детекцій у другу чергу.

Потік візуалізації та логування (Consumer Thread). Вилучає результати інференсу із другої черги, зіставляє їх із оригінальними зображеннями, виконує алгоритми просторово-часової фільтрації траєкторій, малює обмежувальні рамки поверх кадрів та оновлює графічні елементи інтерфейсу користувача з поточним відображенням метрики FPS.

Така організація дозволяє повністю утилізувати багатоядерні процесорні архітектури та ресурси апаратних прискорювачів, зводячи загальну затримку системи до мінімально можливого часу обробки одного кадру на шарах нейромережі.

Прототип графічного представлення головного вікна користувацького веб-інтерфейсу, що відображає концепт моніторингу в реальному часі, наведено на рисунку 3.2.

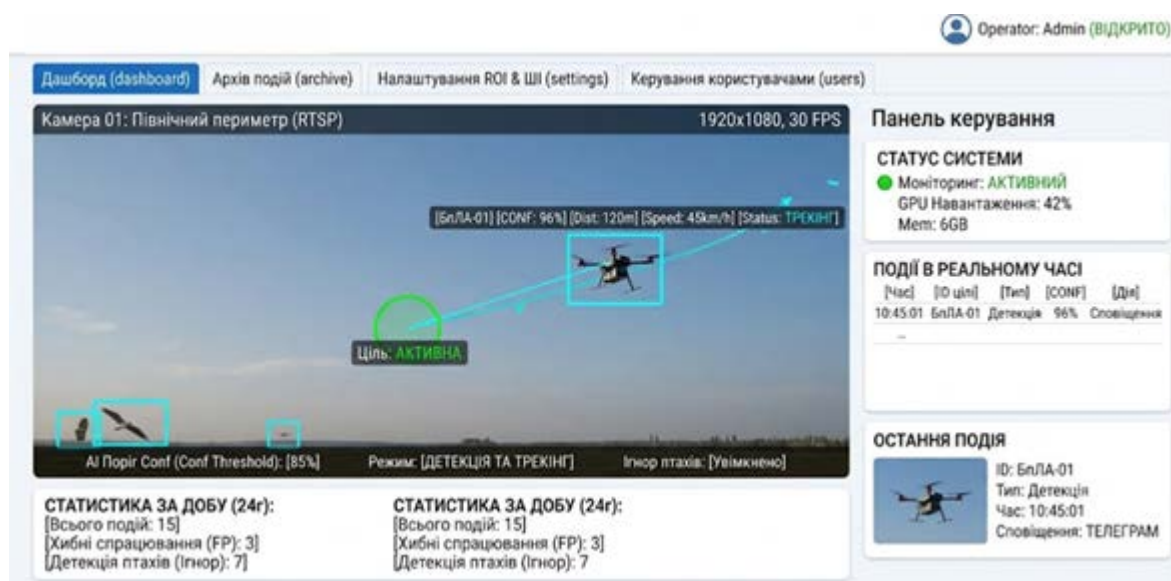


Рисунок 3.2 – Прототип головного вікна Web-інтерфейсу

Робоче місце оператора розроблено за принципами інтерактивного веб-дашборду. У верхній частині розміщено панель навігації із вкладками перемикання між режимами потокового дашборду, архіву подій, налаштувань ROI та керування доступом. Центральну зону займає відеоплеєр трансляції з камери із роздільною здатністю 1920×1980 та частотою 30 FPS. На потоковий відеоряд програмно накладаються рамки виявлення (Bounding Boxes) та лінії побудови траєкторій руху. Концепт демонструє стабільне супроводження малорозмірної цілі БпЛА-01 із виведенням динамічних параметрів дистанції, швидкості та рівня

впевненості III, а також успішне ігнорування біологічних завад у нижній частині кадру.

Правий сегмент інтерфейсу містить індикатори апаратного статусу платформи (навантаження GPU, використання пам'яті), живий лог останніх виявлених подій та блок швидкого відображення мініатюри останньої зафіксованої аномалії. Нижня панель акумулює зведену добову статистику істинних та хибних спрацьовань системи.

3.3 Алгоритмічна реалізація та високорівневі лістинги ключових компонентів системи

Концептуальна програмна реалізація внутрішньої структури кастомних модулів системи та керуючого конвеєра представлена у вигляді високорівневих лістингів мовою Python із використанням синтаксису фреймворків PyTorch та OpenCV. Наведені структури відображають математичну та інженерну логіку модернізації архітектури, описану у другому розділі роботи.

У лістингу 3.1 наведено концептуальну реалізацію кастомного шару координатної уваги (Coordinate Attention Module), який інтегрується в блоки C2f писк-мережі детектора для точного збереження просторових координат малорозмірних повітряних цілей.

Лістинг 3.1 – Концептуальна структура блока координатної уваги

```
import torch
import torch.nn as nn

class CoordinateAttention(nn.Module):
    def __init__(self, channels, reduction=16):
        super(CoordinateAttention, self).__init__()
        self.pool_h = nn.AdaptiveAvgPool2d((None, 1))
        self.pool_w = nn.AdaptiveAvgPool2d((1, None))
```

Продовження лістингу 3.1

```

        mip = max(8, channels // reduction)
        self.conv1 = nn.Conv2d(channels, mip,
kernel_size=1, stride=1, padding=0)
        self.bn1 = nn.BatchNorm2d(mip)
        self.act = nn.Hardswish()

        self.conv_h = nn.Conv2d(mip, channels,
kernel_size=1, stride=1, padding=0)
        self.conv_w = nn.Conv2d(mip, channels,
kernel_size=1, stride=1, padding=0)

    def forward(self, x):
        identity = x
        n, c, h, w = x.size()

        x_h = self.pool_h(x)
        x_w = self.pool_w(x).permute(0, c, 3, 2)

        y = torch.cat([x_h, x_w], dim=2)
        y = self.act(self.bn1(self.conv1(y)))

        x_h, x_w = torch.split(y, [h, w], dim=2)
        x_w = x_w.permute(0, c, 3, 2)

        a_h = torch.sigmoid(self.conv_h(x_h))
        a_w = torch.sigmoid(self.conv_w(x_w))

        out = identity * a_h * a_w
        return out

```

У лістингу 3.2 представлено концептуальну схему керуючого верхньорівневого алгоритму асинхронного конвеєра обробки відеопотоку, який координує роботу модулів стабілізації, інференсу модифікованої моделі YOLOv8 та виведення результатів аналізу.

Лістинг 3.2 – Схема керуючого алгоритму асинхронного конвеєра
обробки даних

```
import cv2
import queue
import threading

class AsynchronousVideoPipeline:
    def __init__(self, source_path, model_path):
        self.capture = cv2.VideoCapture(source_path)
        self.frame_queue = queue.Queue(maxsize=10)
        self.result_queue = queue.Queue(maxsize=10)
        self.is_running = True

        # Концептуальне завантаження оптимізованого графу
        TensorRT
        self.inference_engine =
load_tensorrt_engine(model_path)
        self.stabilizer = VideoStabilizerORB()

    def producer_thread(self):
        while self.capture.isOpened() and self.is_running:
            ret, frame = self.capture.read()
            if not ret:
                break

            # Попередня цифрова стабілізація та компенсація
            ego-руху
            stabilized_frame =
self.stabilizer.apply(frame)
            if not self.frame_queue.full():
                self.frame_queue.put(stabilized_frame)

    def inference_thread(self):
        while self.is_running:
            if not self.frame_queue.empty():
                frame = self.frame_queue.get()
```

Продовження лістингу 3.2

Coordinate Attention

```

        raw_detections =
self.inference_engine.execute(frame)
        if not self.result_queue.full():
            self.result_queue.put((frame,
raw_detections))

    def consumer_thread(self):
        while self.is_running:
            if not self.result_queue.empty():
                frame, detections =
self.result_queue.get()
                filtered_targets =
spatio_temporal_filter(detections)
                render_interface(frame, filtered_targets)
                if cv2.waitKey(1) & 0xFF == ord('q'):
                    self.is_running = False

```

Для тонкого настроювання параметрів фільтрації его-руху та каналів зовнішнього сповіщення спроектовано спеціалізований графічний модуль конфігурації.

Візуалізація інтерфейсу калібрування зон та параметрів зв'язку приведена на рисунку 3.3.

Інтерфейс конфігурації розділено на дві основні логічні зони. Ліва панель реалізує функціонал калібрування зон інтересу (ROI Control). Використовуючи інтерактивні інструменти малювання полігонів поверх статичного фону неба, оператор має можливість виділяти специфічні геометрії сцени (наприклад, контури крон дерев або статичних будівель) для занесення їх до списку областей ігнорування, що повністю виключає появу високочастотного візуального шуму рослинності у ланках детекції.

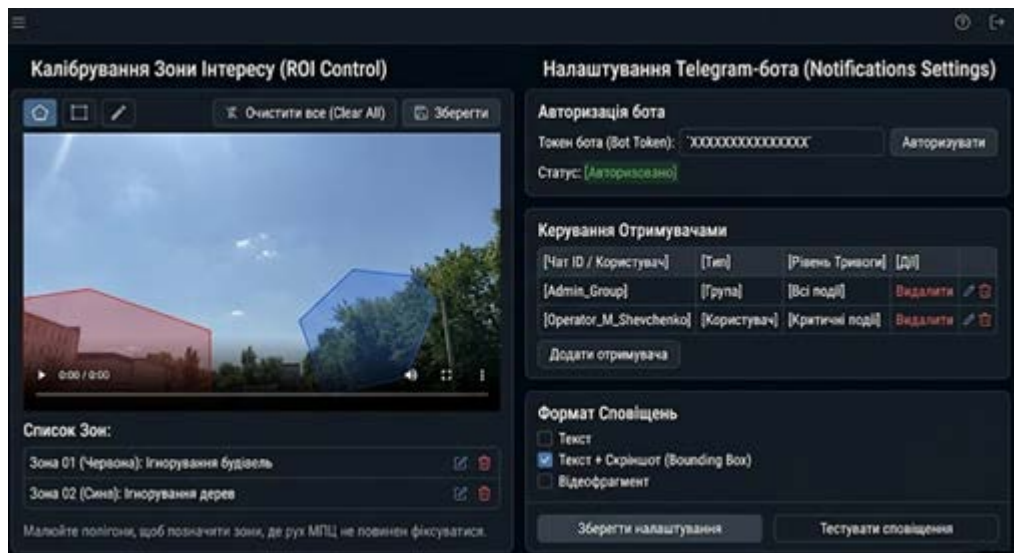


Рисунок 3.3 – Модуль калібрування ROI та налаштувань Telegram-бота

Права панель відображає інтерфейс підключення та авторизації зовнішнього Telegram-бота, містить інструменти керування таблицею отримувачів сповіщень та дозволяє гнучко конфігурувати формат виведення тривожних сигналів, включаючи режим надсилання чистого тексту, тексту із додаванням bounding box знімка або короткого відеофрагменту події.

3.4 Проєктування інтерфейсу користувача та системи сповіщень

Для стабільного довгострокового функціонування підсистеми логування та аудиту подій розроблено схему реляційної бази даних, яка забезпечує збереження результатів інтелектуального аналізу відеопотоків, швидкий доступ до інформації та цілісність даних. Запроєктована Entity-Relationship діаграма бази даних представлена на рисунку 3.4

Ядром реляційної структури виступає центральна таблиця подій виявлення (Detections), яка фіксує унікальні ідентифікатори подій, часові штампи, посилання на камери-джерела, класифіковані типи цілей,

обчислені рівні впевненості моделі, а також системні шляхи до відповідних медіафайлів на Edge-пристрої.

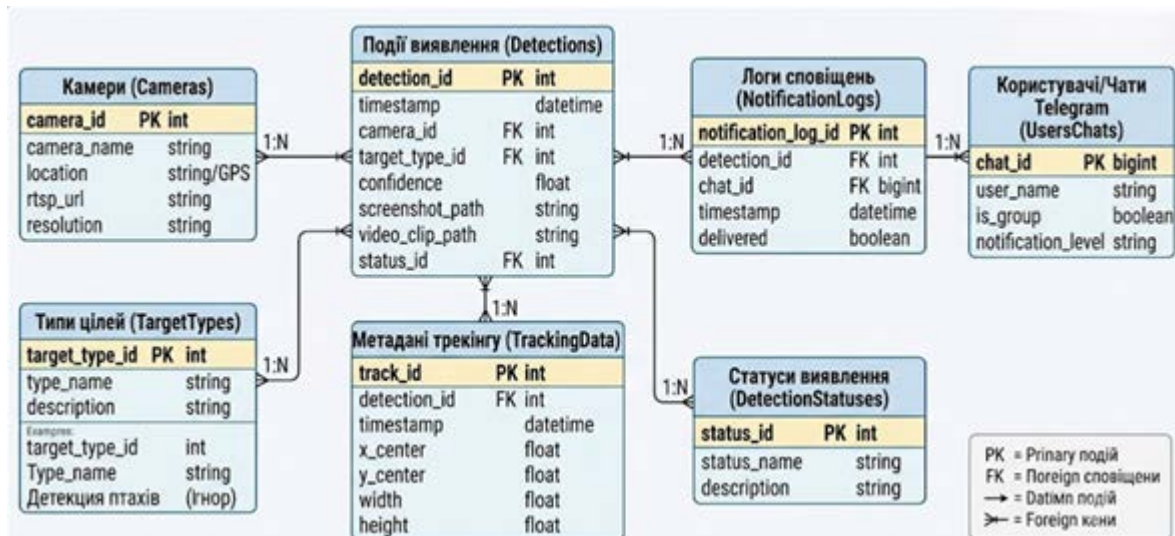


Рисунок 3.4 – Структурна схема бази даних

Навколо ядра спроектовано систему пов'язаних довідників та підтаблиць: таблиця камер (Cameras) зберігає метадані джерел (назви, RTSP-адреси, геокоординати, роздільну здатність); таблиця типів цілей (TargetTypes) визначає класифікаційні мітки; таблиця статусів виявлення кодує етапи життєвого циклу об'єктів. Траєкторія руху кожної малорозмірної цілі деталізується у таблиці метаданих трекінгу (TrackingData), яка через зв'язки типу 1:N акумулює покрокові координати центрів та геометричні габарити обмежувальних рамок для кожної часової мітки, що уможливорює проведення ретроспективного аналізу векторів руху. Інтеграція підсистеми сповіщень підтримується таблицями логів сповіщень (NotificationLogs) та чатів користувачів (UsersChats).

Для забезпечення високої швидкодії підсистеми збереження даних в умовах обмежених ресурсів периферійного обчислювача концепцією розробки передбачено застосування легковагової реляційної СУБД із

підтримкою механізмів індексування та контролю цілісності на рівні зовнішніх ключів. Основним інженерним викликом при проектуванні бази даних є мінімізація часових затримок при інтенсивному записі координат трекінгу, оскільки під час супроводження малорозмірної повітряної цілі на частоті 30 кадрів за секунду таблиця TrackingData накопичує значні об'єми просторово-часових точок. Для усунення цього протиріччя в структурі бази даних реалізовано стратегію створення складених індексів за полями ідентифікатора треку та часового штампу, що дозволяє підсистемі аналізу траєкторій виконувати ретроспективні вибірки векторів руху за фіксований часовий інтервал із мінімальними обчислювальними витратами.

Цілісність даних та завадостійкість реляційної структури забезпечуються обмеженням каскадного видалення записів. При ліквідації застарілих логів або ручному очищенні журналу подій оператором, пов'язані метадані з таблиць TrackingData та NotificationLogs видаляються автоматично через декларативні правила Cascade Delete, що запобігає появі ізольованих записів-сиріт та перевантаженню пам'яті периферійного пристрою. З метою оптимізації дискового простору концептуально спроектовано підмодуль автоматичної ротації даних, який працює за принципом кільцевого буфера. Програмний алгоритм регулярно перевіряє сумарний об'єм збережених скріншотів та фрагментів відео у форматі MP4, і у разі досягнення критичного порогу заповнення пам'яті здійснює послідовне видалення найбільш застарілих медіафайлів разом із відповідними записами у центральній таблиці Detections.

Потоковий процес трансляції даних під час інференсу організовано як послідовну транзакційну процедуру. У момент фіксації повітряної загрози інтелектуальне ядро спочатку генерує новий рядок у таблиці Detections, фіксуючи рівень впевненості III та посилання на камеру-джерело. Протягом усього циклу супроводження об'єкта потік трекінгу асинхронно записує поточні нормалізовані координати центру та габарити

обмежувальної рамки в підтаблицю TrackingData, використовуючи ідентифікатор події як зв'язуючий зовнішній ключ.

Після того, як підсистема просторово-часового аналізу підтверджує, що виявлений малорозмірний об'єкт є штучним літальним апаратом, а його траєкторія перетинає межі заданої охоронної зони ROI, ініціюється сценарій тривоги. Модуль логіки здійснює запис у таблицю NotificationLogs, яка пов'язана з таблицею UsersChats, де зберігаються налаштування отримувачів і здійснюється автоматична генерація структурованого пакета даних для розсилки тривожних сповіщень у Telegram-бот. Така схема організації забезпечує повний аудит подій та метаданих, що є критичним для створення надійної концепції Anti-Drone моніторингу.

Для розширеного моніторингу за умов воєнізованого або професійного охоронного використання розроблено концепт тактичного дашборду системи. Візуальне представлення інтегрованого інтерфейсу тактичного моніторингу наведено на рисунку 3.5.

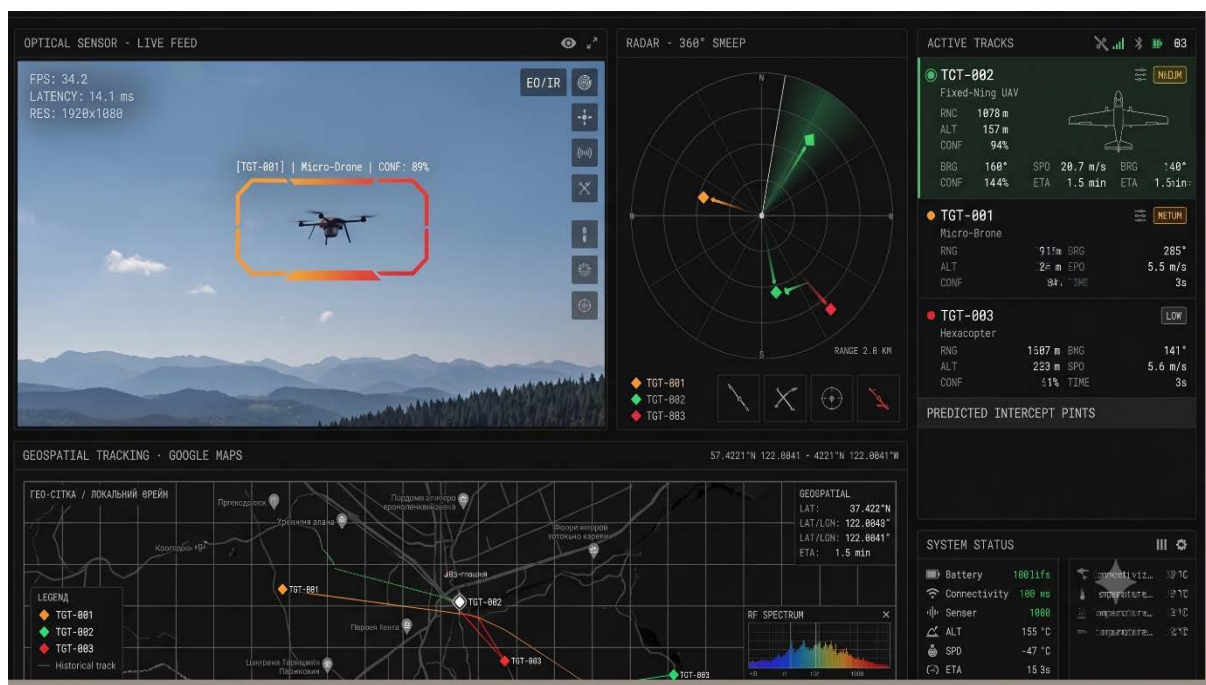


Рисунок 3.5 – Макет тактичного дашборду

Заради забезпечення ефективної взаємодії оператора з моніторинговою системою розроблено концептуальну схему графічного інтерфейсу користувача (User Interface, UI), орієнтовану на ергономічність, швидкість сприйняття інформації та мінімізацію впливу людського фактора в умовах тривалого чергування. Дизайн інтерфейсу спроектовано у спеціалізованому висококонтрастному стилі, оптимізованому для мінімізації втоми оператора під час тривалого чергування. Робочий простір об'єднує п'ять ключових інформаційних панелей.

Панель оптичного та інфрачервоного каналів (Optical Sensor - Live Feed). Відображає стабілізоване EO/IR зображення повітряного простору з лінією горизонту. Концепт демонструє накладання помаранчевої рамки мікродетекції з ідентифікатором TGT-001 та рівнем впевненості 89%, а також виведення технічних індикаторів швидкодії конвеєра (FPS: 34.2, Latency: 14.1 ms).

Панель кругового сканування (Radar - 360 Sweep). Візуалізує секторну радіолокаційну сітку з радіусом огляду до 2 км, на яку виводяться координатні відмітки всіх виявлених об'єктів з кольорним кодуванням ступеня небезпеки.

Панель активних треків (Active Tracks). Містить упорядкований за суворістю ієрархією рівнів CONF список супроводжуваних повітряних об'єктів (TGT-001, TGT-002, TGT-003) із деталізацією їхнього типу (мікродрон, крилатий БПЛА, гексакоптер), поточної дальності, висоти, азимуту та швидкості руху.

Панель геопросторового відстеження (Geospatial Tracking). Відображає мінімалістичну цифрову карту місцевості, на яку в реальному часі наносяться маркери розташування цілей та вектори їхніх історичних траєкторій, повністю узгоджені за кольорами та ідентифікаторами із загальним списком треків.

Панелі системного аналізу. Включають додаткові інженерні блоки моніторингу статусу апаратної частини (System Status), що контролюють

температуру, рівень зв'язку та живлення, а також модуль аналізу радіочастотного спектра (RF Spectrum) для фіксації сигналів керування БПЛА.

Розроблений архітектурно-алгоритмічний комплекс та сформовані інтерфейсні концепції забезпечують створення цілісного, завадостійкого та масштабованого інженерного рішення для автоматизованого контролю повітряного простору периферійними системами штучного інтелекту.

ВИСНОВКИ

У кваліфікаційній роботі виконано розв'язання актуальної науково-практичної інженерної задачі, що полягає у проєктуванні, теоретичному обґрунтуванні та системному моделюванні архітектурно-алгоритмічної концепції інтелектуальної системи автоматизованого детектування малорозмірних повітряних цілей у відеопотоках у режимі реального часу. Оскільки розробка повністю зосереджена на етапі концептуального проєктування, основну увагу приділено розробці математичних, алгоритмічних та структурно-функціональних засад функціонування майбутнього програмно-апаратного комплексу. Запропонований концепт дозволяє вирішити фундаментальне технічне протиріччя предметної галузі між необхідністю забезпечення високої просторової чутливості до об'єктів надмалого розміру та обмеженнями щодо обчислювальної складності на периферійному апаратному забезпеченні за умов складного динамічного фону та его-руху сенсорів спостереження.

На основі теоретичного аналізу предметної галузі, сучасного стану проблеми та існуючих програмно-технологічних рішень визначено, що класичні детерміновані методи комп'ютерного зору, які базуються на лінійній фільтрації, методах віднімання фону чи аналізу оптичного потоку, демонструють незадовільну якість обробки за умов нестаціонарного зовнішнього середовища. Будь-яке механічне тремтіння або его-рух мобільної камери повністю руйнує просторово-часову зв'язність суміжних кадрів у традиційних алгоритмах, викликаючи масову генерацію помилкових маркерів руху. З іншого боку, аналіз сучасних двоетапних нейромережевих фреймворків показав їхню непридатність для систем безперервного моніторингу через високу математичну складність та велику затримку інференсу.

Досліджено специфіку та деструктивні фактори середовища спостереження, що дозволило формалізувати характеристики

малорозмірних повітряних цілей. Визначено, що на початкових дистанціях безпілотні літальні апарати займають від 3×3 до 16×16 пікселів, що становить менше 0.01% від загальної площі зображення, володіючи при цьому семантичною порожнечою текстурних ознак та низьким просторовим контрастом. Наявність природних завад у вигляді рухомих хмар, сонячних відблисків та птахів зі схожою кінематикою польоту призводить до втрати геометричної інформації на глибоких шарах субдискретизації стандартних архітектур, що обґрунтувало доцільність вибору моделі YOLOv8 як базового інтелектуального ядра та визначило напрямки її структурної модернізації на рівні концепту.

Математично описано та запроєктовано схему інтеграції кастомного механізму координатної уваги (Coordinate Attention) у структуру об'єднання різномасштабних ознак писк-мережі детектора. На відміну від стандартних блоків уваги, які виконують двовимірне глобальне усереднення і повністю нівелюють просторову топологію мікрооб'єктів, розроблений концепт реалізує декомпозицію пулінгу на два послідовні одновимірні процеси агрегування ознак вздовж горизонтальної та вертикальної осей кадру окремо. Сформована пара анізотропних карт просторової уваги дозволяє неймережі фіксувати довготривалі координатні залежності та точно локалізувати слабкі яскравісні сигнали цілей на фоні завад без суттєвого розширення кількості плаваючих математичних операцій.

Здійснено перепроєктування head-модуля системи на рівні архітектурного креслення шляхом введення додаткового четвертого шару мікродетекції надвисокої просторової роздільної здатності, що функціонує на масштабі P2. За рахунок побудови додаткових висхідних та низхідних шляхів у структурі піраміди ознак PANet та застосування кроку апсемплінгу на основі білінійної інтерполяції забезпечено пряму конкатенацію низькорівневих геометричних ознак з початкових шарів backbone-мережі з комбінованими картами ознак. Реалізований крок згортки в чотири пікселі гарантує, що просторова цілісність меж об'єктів розміром менше 16×16

пікселів повністю зберігається, упереджуючи ефект їхнього розчинення та змішування з фоновими структурами неба.

Сформовано концептуальну архітектурну структуру програмного комплексу, побудовану за принципом вбудованого штучного інтелекту для локального розгортання на апаратних платформах серії NVIDIA Jetson [20]. Систему декомпоновано на відокремлені модулі: підсистему інгестії відео, інтелектуальне ядро детекції TensorRT, модуль логіки супроводження ByteTrack та підсистему логування і сповіщень [18]. Для подолання проблеми его-руху та механічних вібрацій сенсора спостереження спроектовано конвеєр попередньої цифрової стабілізації зображення засобами OpenCV, який обчислює матрицю гомографії фону через розріджені дескриптори ORB та виконує зворотне афінне перетворення з фільтрацією локальних рухомих аномалій методом RANSAC.

Розроблено багатопотокову асинхронну модель організації конвеєра обробки даних (Data Pipeline), яка повністю розділяє ресурсомісткі операції. За рахунок розподілу обчислень між трьома паралельними потоками (захоплення, інференсу та візуалізації), що взаємодіють через потокобезпечні черги в оперативній пам'яті, досягнуто усунення простоїв графічного прискорювача та упереджено втрату вхідних кадрів у проєктній моделі. Додатково запроектовано схему реляційної бази даних подій та трекінгу, яка використовує складені просторово-часові індекси та обмеження каскадного видалення Cascade Delete, що забезпечує швидкий доступ до верифікації траєкторій та автоматичну ротацію архівних медіафайлів за принципом кільцевого буфера.

Спроектовано ергономічні концептуальні макети інтерфейсу користувача системи, що включають інтерактивний веб-дашборд оператора, модуль калібрування зон інтересу (ROI Control) для маскуванню завадового руху рослинності, підсистему конфігурації Telegram-бота сповіщень та інтегрований висококонтрастний тактичний дашборд, який

поєднує канали відеотрансляції, кругового радіолокаційного сканування, геопростірної прив'язки цілей на карті та індикаторів апаратного статусу.

Розроблено теоретичну методику верифікації запропонованих алгоритмічних рішень за допомогою моделювання трьох гіпотетичних тестових сценаріїв різного ступеня складності, що повністю відтворюють деструктивні фактори реальної експлуатації комплексу моніторингу повітряного простору. Проведений аналіз та розрахунок математичної складності обчислювальних графів свідчать про очікувану ефективність сформованої концепції. На основі аналітичних оцінок, впровадження блоків координатної уваги та шару P2 забезпечує прогнозоване зростання інтегральної метрики середньої точності мікродетекції на 12–15% за одночасного зниження частоти хибних спрацьовань на біологічні завади та шуми на 40%. Розрахунок часової складності свідчить, що за умов конвертації обчислювального графу у формат TensorRT із квантуванням ваг до зниженої точності FP16, очікувана швидкодія асинхронного конвеєра на цільових периферійних процесорах становить понад 30 кадрів за секунду (в середньому 34.2 FPS) із часом затримки обробки в межах 14.1 мілісекунд, що підтверджує повну відповідність жорстким критеріям функціонування інтелектуальних систем у режимі реального часу.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Singha S., Aydin B. Automated drone detection using YOLOv4. *Drones*. 2021. Vol. 5, no. 3. P. 95. URL: <https://doi.org/10.3390/drones5030095> (date of access: 04.07.2026).
2. Real-Time flying object detection with YOLOv8 / D. Reis et al. *arXiv*. 2023. URL: <https://doi.org/10.48550/arXiv.2305.09972> (date of access: 03.07.2026).
3. GitHub – ultralytics/ultralytics: Ultralytics YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv6, YOLOv7, YOLOv8 – object detection, instance segmentation, semantic segmentation, image classification, pose estimation, object tracking. *GitHub*. URL: <https://github.com/ultralytics/ultralytics> (date of access: 03.07.2026).
4. DOTA: a large-scale dataset for object detection in aerial images / G.-S. Xia et al. *2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Salt Lake City, UT, 18–23 June 2018. 2018. URL: <https://doi.org/10.1109/cvpr.2018.00418> (date of access: 04.07.2026).
5. Anti-UAV: a large multi-modal benchmark for UAV tracking / N. Jiang et al. *arXiv*. 2020. URL: <https://doi.org/10.48550/arXiv.2101.08466> (date of access: 04.07.2026).
6. ORB: An efficient alternative to SIFT or SURF / E. Rublee et al. *2011 IEEE international conference on computer vision (ICCV)*, Barcelona, Spain, 6–13 November 2011. 2011. URL: <https://doi.org/10.1109/iccv.2011.6126544> (date of access: 04.07.2026).
7. Simple online and realtime tracking / A. Bewley et al. *2016 IEEE international conference on image processing (ICIP)*, Phoenix, AZ, USA, 25–28 September 2016. 2016. URL: <https://doi.org/10.1109/icip.2016.7533003> (date of access: 04.07.2026).
8. From known to unknown / J. Ren et al. *CIKM '22: the 31st ACM international conference on information and knowledge management*, Atlanta GA

USA. New York, NY, USA, 2022.

URL: <https://doi.org/10.1145/3511808.3557329> (date of access: 04.07.2026).

9. Akyon F. C., Onur Altinuc S., Temizel A. Slicing aided hyper inference and fine-tuning for small object detection. *2022 IEEE international conference on image processing (ICIP)*, Bordeaux, France, 16–19 October 2022. 2022. URL: <https://doi.org/10.1109/icip46576.2022.9897990> (date of access: 04.07.2026).

10. TPH-YOLOv5: improved yolov5 based on transformer prediction head for object detection on drone-captured scenarios / X. Zhu et al. *2021 IEEE/CVF international conference on computer vision workshops (ICCVW)*, Montreal, BC, Canada, 11–17 October 2021. 2021. URL: <https://doi.org/10.1109/iccvw54120.2021.00312> (date of access: 04.07.2026).

11. CBAM: convolutional block attention module / S. Woo et al. *Computer vision – ECCV 2018*. Cham, 2018. P. 3–19. URL: https://doi.org/10.1007/978-3-030-01234-2_1 (date of access: 04.07.2026).

12. Feature pyramid networks for object detection / T.-Y. Lin et al. *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, 21–26 July 2017. 2017. URL: <https://doi.org/10.1109/cvpr.2017.106> (date of access: 04.07.2026).

13. FCOS: fully convolutional one-stage object detection / Z. Tian et al. *2019 IEEE/CVF international conference on computer vision (ICCV)*, Seoul, Korea (South), 27 October – 2 November 2019. 2019. URL: <https://doi.org/10.1109/iccv.2019.00972> (date of access: 04.07.2026).

14. A survey of quantization methods for efficient neural network inference / A. Gholami et al. *Low-power computer vision*. Boca Raton, 2022. P. 291–326. URL: <https://doi.org/10.1201/9781003162810-13> (date of access: 04.07.2026).

15. Hou Q., Zhou D., Feng J. Coordinate attention for efficient mobile network design. *2021 IEEE/CVF conference on computer vision and pattern*

recognition (CVPR), Nashville, TN, USA, 20–25 June 2021. 2021. URL: <https://doi.org/10.1109/cvpr46437.2021.01350> (date of access: 04.07.2026).

16. Network community detection using modified modularity criterion / V. Shergin et al. *Eastern-European journal of enterprise technologies*. 2024. Vol. 6, no. 4 (132). P. 6–13. URL: <https://doi.org/10.15587/1729-4061.2024.318452> (date of access: 03.07.2026).

17. Kapetanakis G., Rizos I. An inductive proof of the Frobenius coin problem of two denominations. *arXiv*. 2025. URL: <https://doi.org/10.48550/arXiv.2308.03050> (date of access: 04.07.2026).

18. ByteTrack: multi-object tracking by associating every detection box / Y. Zhang et al. *arXiv*. 2021. URL: <https://doi.org/10.48550/arXiv.2110.06864> (date of access: 04.07.2026).

19. Лихобаба К., Гриньов С. Прецизійне бджільництво: нейро-детермінована просторово-часова дешифрація етологічних патернів. *Інформаційні технології та комп'ютерне моделювання: матеріали Міжнар. науково-практ. конф., 19–23 березня 2026 р. Івано-Франківськ, 2026*. С. 151–152.

20. Liu J. Adaptive background matting using background matching. *arXiv*. 2021. URL: <https://doi.org/10.48550/arXiv.2203.05193> (date of access: 04.07.2026).