

БАГАТОРІВНЕВИЙ ПІДХІД ВИЯВЛЕННЯ ТА РОЗПІЗНАВАННЯ НЕБЕЗПЕЧНИХ ПРЕДМЕТІВ ЗА ДОПОМОГОЮ БПЛА ІЗ ЗАСТОСУВАННЯМ КОМП'ЮТЕРНОГО ЗОРУ

Роботько С. П.¹, Луцак Д. Л.²

¹Національний університет кораблебудування імені адмірала Макарова

²Університет короля Данила

Анотація. Запропоновано багаторівневу систему виявлення мін за допомогою безпілотного літального апарата (БПЛА) зі швидкою детекцією YOLOv8n на наземній станції та хмарній семантичній верифікації ChatGPT-4o Vision із проміжною впевненістю. Відео 1080p з Raspberry Pi синхронізується з телеметрією Pixhawk 6C й геоприв'язується, після чого контейнеризація та повне логування гарантують відтворювання. Запропонована схема прийняття рішень дозволяє зберігати високу повноту на етапі YOLOv8 і скорочувати хибні спрацьовування завдяки семантичній перевірці проблемних ділянок зображення в хмарі без істотного впливу на середню затримку та смугу каналу.

Ключові слова: гуманітарне розмінування, БПЛА, візуально-мовні моделі, бездротова комп'ютерна мережа, комп'ютерний зір, YOLOv8, геоприв'язка, багаторівневий аналіз зображення, ChatGPT.

Вступ

Мінна небезпека на великих площах потребує інструментів швидкого та безпечного обстеження, здатних працювати в різноманітних умовах фону, освітлення та погоди. Класичні металодетектори надійні для масивної мети, однак малі протипіхотні міни (зокрема ПФМ-1) часто залишаються поза чутливістю або подають неоднозначні сигнали за нестабільних траєкторій БПЛА й часткових оклюзій.

Комп'ютерний зір з безпілотних платформ довів ефективність у візуальному виявленні, але вимагає архітектурних рішень, що збалансують точність, затримку, смугу каналу й енергоспоживання та зможуть забезпечити відтворюваність і прозору трасованість інформації.

У попередніх дослідженнях аналізували успішні кейси використання глибокого навчання для оптичної детекції мін і багатосенсорні підходи з інтеграцією магнітометрії й оптики.

Водночас у доступній літературі майже не відтворювали наскрізні рішення, які поєднують узгоджене за часом захоплення відео з БПЛА та телеметрії, швидку детекцію на наземному рівні з просторово-часовою консистентністю, ескалаційну хмарну верифікацію «важкими» моделями для суперечливих випадків формалізоване злиття аналізу в єдину інтегральну ймовірність рішення з можливістю аудиту.

Ця прогалина особливо відчутна для випадків малорозмірної мети на складному фоні, де одинична модальність часто має компроміси між повнотою та точністю (recall та precision).

У цій роботі запропоновано багаторівневу архітектуру edge→ground→cloud, що чітко розподіляє функції між рівнями й експлуатує їхні переваги. На edge реалізовано безперервне захоплення і гарантовану доставку сегментів відео з уніфікованими часовими мітками; на ground – основне інтелектуальне оброблення з використанням YOLOv8 і геоприв'язкою, яка забезпечує просторово-часову консистентність і близький до реального часу аналіз; на cloud – селективну семантичну верифікацію візуально-мовними моделями VLM (ChatGPT-4o Vision та Gemini 2.5) з проміжною впевненістю, що знижує хибні спрацьовування без істотного впливу на середню затримку. Архітектура підтримує повну відтворюваність через контейнеризацію модулів і журналювання версій моделей, порогів, метаданих і контрольних сум.

Основними науковими внесками роботи є формалізована ієрархія оброблення із прозорою часовою узгодженістю, узагальнена модель P_{final} для злиття аналізу детектора та VLM і відтворювана експериментальна методологія з роботоздатністю в польових умовах.

Отримані результати демонструють, що запропонована система досягає високої якос-

ті в однокласовому визначені ПФМ-1 і створює основу для подальшого розширення – до багатокласових сценаріїв і інтеграції фізичного каналу металодетектора у формальну модель злиття.

Аналіз публікацій

Нещодавно опубліковані роботи демонструють ефективність глибокого навчання для оптичного виявлення наземних мін. Зокрема Emanuele Vivoli та співавтори проаналізували виявлення поверхневих мін (PFM-1, PMA-2) із використанням YOLOv8, вбудованого в роботизовану платформу, зазначивши високу повноту та придатність до практичного використання [1].

Окремий напрям – моделювання чинників середовища. Зазначено, що рослинний покрив і часткові перекриття істотно впливають на чутливість YOLO-детекторів з БПЛА. Автори пропонують визначати точність як функцію оклюзії, отриманої з фотограмметричних моделей місцевості. Це доводить необхідність урахування умов зйомки в польових експериментах [2].

У суміжній гільці активно досліджують інтеграцію оптичних і магнітометричних даних, застосовуючи байєсівські підходи, теорію Демпстера-Шефера, фільтри Калмана тощо.

Це доводить перспективність багатосенсорних систем та дозволяє здійснити процес подальшого об'єднання з металодетекторами.

У публікації “Optical and Magnetometric Data Integration for Landmine Detection with UAV” [3] автори описують побудову «магнітних карт» та використання CNN/глибокого навчання для розпізнавання мін за магнітометричними зображеннями, зазначаючи, що поєднання сигналів різного фізичного походження збільшує ефективність виявлення.

З огляду на це наша робота відрізняється мультирівневим аналізом відеоданих (попередній відбір/детекція YOLOv8 [4] та подальша верифікація на рівні VLM (ChatGPT/Gemma) у хмарі), що дозволяє зменшити хибні напрацювання та збільшити рівень довіри до рішень завдяки ескалації між рівнями.

Література доводить ефективність оптичної детекції YOLOv8 [5] та багатосенсорної інтеграції (оптика+магнітометрія). Водночас невелика кількість робіт формалізує ієрархічну (edge→ground→cloud) архітектуру з

узгодженими часовими мітками та прозорою трасованістю.

Запропонований нами підхід вирішує це завдання: він поєднує швидку детекцію на відео (YOLOv8) з подальшою семантичною верифікацією (ChatGPT/Gemma) у хмарі та готує підґрунтя для подальшого злиття з металодетектором.

Мета та постановка завдання

Мета дослідження – розробити та експериментально верифікувати модуль комп'ютерного зору для виявлення протипіхотних мін з борту БПЛА. Для її досягнення потрібно спроектувати відтворюваний конвеєр збору, геоприв'язки та оброблення відео, сформувати й розмітити репрезентативний датасет для навчання та налаштування моделі YOLOv8n. Передбачається аналіз точності та продуктивності системи в польових умовах, а також окреслення інтеграції з металодетектором.

Архітектура системи. Запропонована система реалізує ієрархію оброблення edge→ground→cloud, оптимізуючи вектор показників «точність – затримка – смуга – енергія – відтворюваність».

На (рис. 1) edge (борт БПЛА) здійснює безперервне захоплення та кільцеву буферизацію відео з часовими мітками, синхронізованими з навігаційним трактом (GPS/INS). Цей рівень працює за принципом проміжного зберігання та перенаправлення із гарантованою доставкою.



Рис. 1. Зображення комплексу виявлення мін у польоті

Сегменти відео передаються на наземну станцію тоді, коли це дозволяє канал, а в разі перебоїв зв'язку накопичуються локально й відновлюються без втрати інформації. Такий процес зменшує вимоги до миттєвої пропускної здатності та енергоспоживання борта, але зберігає повну трасованість інформації: кожен кадр має уніфіковану часову мітку, що в подальшому забезпечує коректну геоприв'язку й аудит.

Основне інтелектуальне оброблення здійснюється на *ground*-рівні. Тут формалізовано центральну новизну, а саме поділ функцій між швидкою візуальною детекцією та відкладеною семантичною верифікацією.

Відеокадри, узгоджені з телеметрією, подаються на детектор YOLOv8 [6], який оцінює ймовірність належності об'єкта до класу мін. Вихід YOLOv8 інтерпретується як рівень s_Y і є першим рівнем прийняття рішень з низькою латентністю.

Для підвищення стійкості до оклюзій, змін масштабу та умов фону передбачено часове об'єднання детекцій (агрегація s_Y у вікнах кадрів уздовж траєкторії) та Post-hoc геоприв'язку: кожне спрацювання отримує координати WGS-84 і висоту, що дає можливість здійснювати просторовий аналіз та повторювані вимірювання в тих самих точках. Таким чином, *ground*-рівень поєднує «субсекундну» інференс-логіку з просторово-часовою консистентністю.

Рівень *cloud* реалізує ескалаційний шар для випадків невизначеності. На ньому застосовуються «важкі» моделі семантичної верифікації (VLM): ChatGPT-4o Vision або Gemma-3 Vision, які зображенням підозрілих ділянок визначають незалежний рівень s_C (semantic confidence).

Архітектурно хмара інтегрована як сервіс валідації з пріоритетом «on demand»: до неї спрямовуються лише ті події, для яких *ground*-рівень формує проміжну впевненість (зона невизначеності).

Така ескалація зберігає пропускну здатність і контролює затримку, де більшість рішень приймається локально, а хмара під'єднується точково, коли очікувана користь від семантичного сигналу перевищує вартість виклику.

Функціональна взаємодія рівнів описується узагальненим часовим конвеєром із явними бюджетами, а загальна затримка визначається як

$$L_{\text{tot}} = L_{\text{cap}} + L_{\text{tx}} + L_{\text{YOLO}} + L_{\text{fuse}} + L_{\text{cloud}}. \quad (1)$$

Внесок L_{cloud} активується лише для ескалаційних випадків у режимі роботи “AND”, завдяки цьому середня L_{tot} нижча за монолітні хмарні схеми, а вимоги до каналу описуються як $B_{\text{avg}} \ll B_{\text{raw}}$ завдяки сегментації і вибіркоким відправкам ROI (Region of Interest).

Якщо використовують сценарій “OR”, до хмари надсилаються всі можливі кадри, що збільшує вірогідність виявлення. Архітектура передбачає відмовостійкість – у разі втрати мережі рішення приймаються на *ground*-рівні, а хмарна верифікація здійснюється пакетно після відновлення зв'язку без деградації трасованості.

На рівні даних система забезпечує відтворюваність і аудит. Усі модулі використовуються як контейнеризовані компоненти з фіксованими версіями моделей і залежностей. Метадані експериментів (хеші відеосегментів, конфігурації порогів, ревізії ваг детекторів) логуються в машинному журналі. Єдина вісь часу будується з використанням GPS-часу автопілота та NTP-синхронізації борту.

Під час злиття інформації застосовується інтерполяційне зіставлення «кадр $\leftrightarrow$ телеметрія» з контролем дрейфу. Це гарантує, що геоприв'язані детекції можуть бути відтворені, перевірені та повторно проаналізовані під іншими порогами або моделями без повторного польоту.

Ключовим науковим елементом є ескалаційне правило рішень, яке пов'язує архітектуру з формальною моделлю. На *ground*-рівні формується інтегральна впевненість $P_{\text{local}} = g(s_Y)$, після чого події з $P_{\text{local}} \in [\theta_l, \theta_u]$ передаються до хмари для семантичної верифікації; основою для фінального рішення є $P_{\text{final}} = f(s_Y, s_C, \dots)$ із порогом Θ , де $f(\cdot)$ – калібрована агрегуюча функція (наприклад, з урахуванням Platt/temperature-scaling для зведення значень до єдиної ймовірнісної шкали).

Така конструкція узагальнює OR/AND і k-із-n-правила як окремі випадки порогування й демонструє, що багаторівнева організація дозволяє одночасно знизити хибні спрацювання та зберегти високу.

Практична схема цього підходу подана на рис. 2: відео, інформація з металодетектора та GPS з борта надходять на наземну станцію, де відбуваються злиття та детекція YOLOv8.

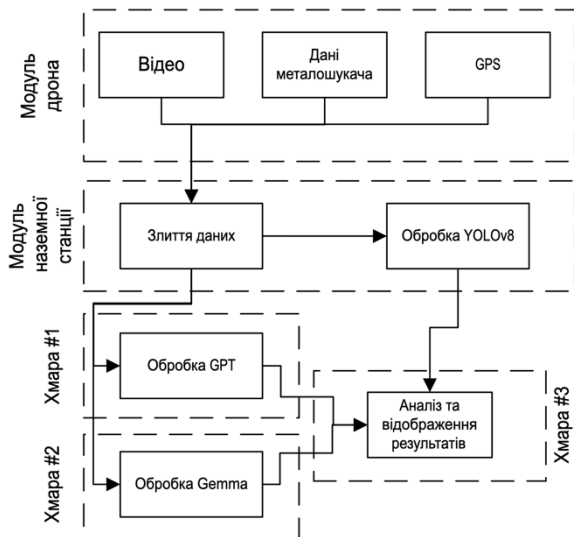


Рис. 2. Схема багаторівневого підходу оброблення відеоданих

Невизначені випадки ескалюються у хмари з обробленням ChatGPT/Gemma, а результати консоліднуються в модулі аналізу та візуалізації. Саме таке розмежування функцій між рівнями забезпечує керований компроміс між точністю, затримкою та смугою, а також гарантує відтворюваність і аудит усіх рішень завдяки єдиній часовій шкалі, контейнеризації і повному журналюванню експериментів.

Методологія оброблення. Вхідними даними є відеопотік з борта БПЛА у форматі MPEG-4 (1920×1080) із уніфікованими часовими мітками, синхронізованими з телеметрією автопілота (GPS/курс/висота). На наземній станції відео декодується в реальному часі або пакетно.

Застосовується легка стабілізація від мікрівібрацій через аналіз міжкадрового зсуву. Детектор комп'ютерного зору реалізовано на основі Ultralytics YOLOv8n, створеної за попередньо розміченими [7] прикладами мін малого розміру (деталі складу датасету, протокол анотування та розподіл train/val/test наведено у [8]).

Під час інференсу використовується пакування кадрів для зменшення накладних витрат препроцесингу, а постпроцесинг здійснюється стандартною NMS (Non-Maximum Suppression), що прибирає дублікати виявлень із каліброваними порогами score/IoU, що було вибрані за валідаційними PR-кривими.

Вихід YOLOv8 інтерпретується як безперервний рівень $s_Y \in [0,1]$ для кожної гіпотези з її просторовими координатами (x_1, y_1, x_2, y_2) .

Щоб зменшити локальні флуктуації впевненості та чутливість до оклюзій, результати YOLOv8 (рис. 3) агрегуються в часі. Гіпотези з подібними положеннями на сусідніх кадрах об'єднуються в треки з експоненційним згладжуванням s_Y , а до кожної події прив'язується телеметрія через лінійну інтерполяцію за часовими мітками.

На цьому етапі формується список ROI-фрагментів з обрізаними зображеннями об'єктів і метаданими (час, Lat/Lon/Alt, курс, s_Y). Для рішень «очевидно позитивних» та «очевидно негативних» застосовуються два пороги $\theta_{low} < \theta_{high}$: якщо $s_Y \geq \theta_{high}$, спрацювання фіксується локально без ескалації; якщо $s_Y \leq \theta_{low}$, відкидається. Події «сірої зони» $s_Y \in (\theta_{low}, \theta_{high})$ передаються до хмари для семантичної верифікації, що мінімізує як середню затримку, так і обсяг переданої інформації.



Рис. 3. Зображення частини обробленого відео з анотацією

Семантична верифікація здійснюється в хмарі за допомогою VLM із зоровими можливостями (ChatGPT-4o Vision API) [9]. До запиту передаються ROI вирізка та службові метадані (час/координати). Для стимулювання бінарної відповіді використовується короткий інструктивний промпт: *“You are a land mine detector. Say ‘YES’ if you can see any suspicion that the photo shows an explosive object. Shortly describe its type if you can”* (укр.: «Ви детектор мін. Скажіть “YES”, якщо бачите будь-які підозри, що на фото зображено вибухонебезпечний предмет. Коротко опишіть його тип, якщо зможете»). Модель повертає перший токен YES/NO та

короткий опис. Для перетворення текстової відповіді в числову семантичну впевненість s_C використано два режими. У швидкому режимі задається детермінований мапінг: $YES \rightarrow s_C = 1$, $NO \rightarrow s_C = 0$, що адекватно здійснює функцію жорсткого фільтра на сумнівних кейсах. У розширеному режимі оцінюється «м'яка» впевненість через ансамблювання.

Модель аналізується Т разів з нейтральними парафразами промпта та фіксованими обмеженнями формату. s_C визначається як частка відповідей YES і надалі калібрується логістичною регресією (Platt/temperature-scaling) за валідаційною вибіркою, щоб звести s_C до імовірнісної шкали. Обидва режими допускають автоматичний парсинг та журналювання відповіді разом із зображенням ROI і геоприв'язкою.

Фінальний розв'язок для кожної події приймається як пороговування інтегральної імовірності P_{final} , яка для дворівневої схеми задається агрегуючою функцією $P_{final} = f(s_Y, s_C)$. На практиці використовувалося каліброване зважування

$$P_{final} = \lambda_Y s_Y + \lambda_C s_C, \quad (2)$$

якщо $\lambda_Y + \lambda_C = 1$.

В експериментах вагові коефіцієнти було зафіксовано як $\lambda_Y = 0.7$ та $\lambda_C = 0.3$. Більша вага λ_Y на ознаці наявності об'єкта підсилює чутливість до виявлення мін, тоді як менша вага λ_C для впевненості класифікації є достатньою в умовах однокласового визначення, забезпечуючи збалансований компроміс між чутливістю та точністю. Така конструкція узагальнює класичні правила, де OR-логіка дорівнює низькому порогу Θ або високій вазі λ_C для YES-вердикту чат-моделі; AND-логіка – високому Θ або вимозі одночасно $s_Y \geq \theta_Y$ і $s_C = 1$. Після прийняття рішення подія маркується як «детекція міні» або «відхилено», а всі проміжні значення (s_Y , s_C , P_{final}) та версії моделей зберігаються в журналі для аудиту й повторюваності.

Сумарний часовий бюджет конвеєра можна записати так:

$$L_{tot} = L_{decode} + L_{stabilize} + L_{YOLO} + L_{agg} + L_{esc} \cdot (L_{upl} + L_{VLM}), \quad (3)$$

де L_{esc} – індикатор ескалації у хмару. Оскільки лише події «сірої зони» спрямовуються на верифікацію, середня L_{tot} залишається наближеною до латентності YOLOv8, а пропускна здатність каналу визначається кількі-

стю ROI та їхнім розміром, а не сирих відеопотоком. Такий режим забезпечує керований компроміс між чутливістю та хибними спрацьовуваннями: швидка візуальна детекція формує високий recall, а семантична перевірка відсікає типові помилки на «схожих» об'єктах (листя, каміння, сміття), підвищуючи precision без істотної втрати пропускної здатності або збільшення часу затримки. Усі кроки реалізовані в контейнерах з фіксованими версіями залежностей; часові мітки, конфігурації порогів і ваг-агрегування логуються разом із контрольними сумами відеосегментів, що гарантує повну відтворюваність і можливість повторного аналізу під різними θ_{low} , θ_{high} , Θ , λ без повторних польотів.

Експерименти та метрики. Польові випробування здійснювалися на ділянках із розкладеними муляжами ПФМ-1. БПЛА здійснював низьковисотні проходи ($\approx 2\text{--}3$ м) над hetero-фоном (трава/грунт/тінь). Відео 1920×1080 (MPEG-4) записувалося на Raspberry Pi 4B [10] та передавалося на польовий ПК за допомогою Wi-Fi/SSH-сегментів з гарантованим відновленням; телеметрія Pixhawk 6C [11] логувалася паралельно. Оброблення здійснювалося покадрово конвеєром «стабілізація/препроцесинг $\rightarrow$ інференс YOLOv8n $\rightarrow$ NMS $\rightarrow$ геоприв'язка $\rightarrow$ журналювання». Кадри вирівнювалися від мікрівібрацій, масштабувалися з letterbox-падингом і нормалізувалися. Детекції отримували часову мітку та координати (інтерполяція за ULog/MAVLink) [12]. Для визначення якості застосовано стандартні метрики Precision, Recall, mAP@0.5 і mAP@0.5:0.95, для продуктивності – латентність за стадіями (preprocess/inference/postprocess на кадр) і ефективний FPS. Додатково фіксувалися показники операційного рівня для двоступеневої схеми. Частка ROI, що ескалюються до хмари (rate of escalation), середній час запиту до VLM та додаткова передана смуга (на ROI-виризках, а не на опрацьованому відео). Усі події логувалися в машинночитаному форматі (frame_id, timestamp, bbox, conf, геоприв'язка, версії моделей, конфігурації порогів), що забезпечує повну відтворюваність експериментів.

Налаштування середовища дорівнювали таким версіям: Ultralytics 8.3.178, Python 3.9.6, torch-2.8.0 із апаратним прискоренням MPS (Apple M3 Pro); підсумкова модель – YOLOv8n (fused), 72 шари, 3 005 843 пара-

метри, 8.1 GFLOPs. На валідаційному піднаборі 53 зображення / 23 інстанси отримано: Precision = 0.957, Recall = 1.000, mAP@0.5 = 0.981, mAP@0.5:0.95 = 0.688. Час на кадр становив ≈ 27.7 мс: 0.5 мс (preprocess) + 21.2 мс (inference) + 6.0 мс (postprocess), що дорівнює ≈ 36 FPS; внесок стадій у затримку – $\sim 1.8\%$ / 76.5% / 21.7% . Такі значення свідчать про високу якість в однокласовій постановці та вказують на очікувану чутливість до масштабу/ракурсу й часткових оклюзій (нижчий mAP@0.5:0.95).

Дослідницьку верифікацію в хмарі здійснювали для «сірої зони» впевненості детектора. ROI з проміжними значеннями скору передавалися до ChatGPT-4o Vision із коротким інструктивним промптом: “You are a land mine detector. Say ‘YES’ if you can see any suspicion that the photo shows an explosive object. Shortly describe its type if you can”.

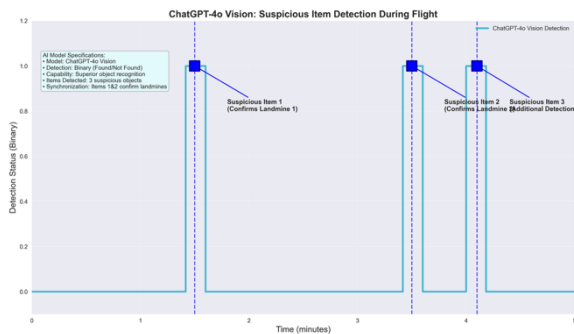


Рис. 4. Графік спрацювань ChatGPT-4o Vision під час аналізу зображень

Відповідь аналізувалась до семантичної впевненості s_C (режим «жорсткий» YES/NO або «м'який» через ансамблювання запитів з подальшим Platt/temperature-scaling). На рисунку 4 подано спрацювання моделі, де 1 – це YES, а 0 – NO. Фінальна імовірність P_{final} формувалася як калібрована агрегація $P_{\text{final}} = \lambda_Y s_Y + \lambda_C s_C$ з підбором λ і межі Θ за валідаційними кривими PR під обмеженням FAR. Для відтворювання зберігалися всі первинні відповіді VLM і контрольні суми переданих ROI. Якісний аналіз ефекту верифікації наводили як порівняння PR/ROC-кривих «YOLO-baseline» проти «YOLO+ChatGPT» та зміни F1 за фіксованого FAR, а числова бенчмаркінг-програма розширюється в межах подальших робіт, однак уже зараз журналювання демонструє керовану ціну ескалації, де додаткова латентність і смуга залежать від частки спірних ROI, тоді як більшість рішень приймається локально на наземній станції.

Сукупно отримані результати доводять спроможність системи до близького до реального часу післяпольотного аналізу на польовому ноутбучі та демонструють, що двоступенева схема дозволяє зберегти високий recall детектора і водночас підвищити рівень довіри до рішень завдяки семантичній перевірці складних випадків. Відтворюваність забезпечено контейнеризацією модулів, єдиною часовою шкалою інформації і повним журналюванням, що дозволяє повторно здійснити експерименти та перевірити результати під іншими межами без повторних польотів.

Висновки

У роботі подано багаторівневу систему виявлення протипіхотних мін з БПЛА, побудовану за архітектурою edge→ground→cloud із чітким розподілом функцій та часовою узгодженістю інформації. На наземному рівні реалізовано швидку візуальну детекцію (YOLOv8), а на хмарному – селективну семантичну верифікацію (ChatGPT-4o Vision) для випадків невизначеності. Така ескалаційна логіка дозволяє досягти кращого балансу «точність – затримка – смуга», як порівняти з однорівневими схемами, і водночас зберегти відтворюваність завдяки контейнеризації, єдиній часовій шкалі та повному журналюванню.

Польові випробування на муляжах ПФМ-1 довели практичну придатність підходу. Базовий детектор досяг показників Precision = 0.957, Recall = 1.000, mAP@0.5 = 0.981, mAP@0.5:0.95 = 0.688, які довели ефективну продуктивність ≈ 36 FPS на польовому ноутбучі. Запропонована двоступенева схема прийняття рішень дозволяє зберігати високий recall на етапі YOLOv8 [13] і скорочувати хибні спрацювання завдяки семантичній перевірці спірних ROI в хмарі без істотного впливу на середню затримку та смугу каналу.

Наукова новизна полягає у формалізованій ієрархії оброблення з інтегрованою моделлю злиття, де локальна впевненість детектора s_Y поєднується із семантичною впевненістю s_C для формування P_{final} та порогового рішення. Така конструкція узагальнює класичні OR/AND і k-із-n-правила та демонструє керований компроміс між чутливістю та специфічністю завдяки вибіркового використанню «важкого» семантичного сигналу.

Обмеження роботи пов'язані з однокласовою постановкою та чутливістю до

зміни масштабу/ракурсу та часткових оклюзій. Подальші дослідження передбачають розширення до багатокласового детектора, приєднання фізичного каналу металодетектора до формальної моделі P_{final} , систематичний бенчмаркінг PR/ROC для «YOLO vs. YOLO+VLM» на розширеному наборі сцен та оптимізацію латентності (квантовка/експорт у прискорювачі). У сукупності запропонована архітектура та методологія формують відтворювану та масштабовану основу для оперативного гуманітарного розмінування.

Література

1. Emanuele Vivoli, Marco Bertini, Lorenzo Capineri Deep Learning-Based Real-Time Detection of Surface Landmines Using Optical Imaging. *Remote Sens.* 2024. URL: <https://www.mdpi.com/2072-4292/16/4/677>
2. Jasper Baur, Kyle Dewey, Gabriel Steinberg, Frank O. Nitsche. Modeling the Effect of Vegetation Coverage on Unmanned Aerial Vehicles-Based Object Detection: A Study in the Minefield Environment. *Remote Sens.* 2024. URL: <https://www.mdpi.com/2072-4292/16/12/2046>
3. Sergey A. Stankevich, Ievgen Y. Saprykin. Optical and Magnetometric Data Integration for Landmine Detection with UAV. *WSEAS Transactions on Environment and Development*, 2024. DOI: 10.37394/232015.2024.20.96.
4. Dillon Reis, Jordan Kupec, Jacqueline Hong, Ahmad Daoudi. Real-Time Flying Object Detection with YOLOv8, arXiv, 2023. URL: <https://arxiv.org/abs/2305.09972>
5. Ultralytics. “YOLOv8 Documentation,” Ultralytics Docs, 2025. URL: <https://docs.ultralytics.com/>
6. YOLOv5, YOLOv8 and YOLOv10 / Muhammad Hussain, et al. Muhammad Hussain, et al. The Go-To Detectors for Real-Time Object Detection, arXiv, 2024. URL: <https://arxiv.org/pdf/2407.02988>
7. Label Studio Team. Label Studio: Open Source Data Labeling Platform. Documentation, 2025. URL: <https://labelstud.io/>
8. Yolo model training dataset. URL: https://github.com/robotkos/yolo_dataset
9. OpenAI. GPT-4o System Card, OpenAI, 2024. URL: <https://cdn.openai.com/gpt-4o-system-card.pdf>
10. Simon Monk “Raspberry Pi Cookbook: Software and Hardware Problems and Solutions 4th Edition”. 2023
11. PX4 Dev Team. Holybro Pixhawk 6C | PX4 User Guide, PX4 Docs, 2024. URL: https://docs.px4.io/main/en/flight_controller/pixhawk6c
12. MAVLink Community. MAVLink Developer Guide, 2025. URL: <https://mavlink.io/en/>
13. Jasper Baur, Kyle Dewey, Gabriel Steinberg, Frank O. Nitsche. A False-Positive-Centric Framework for Object Detection in Aerial Imagery. *Remote Sensing*, 2025. URL: <https://www.mdpi.com/2072-4292/17/14/2429>

References

1. Emanuele Vivoli, Marco Bertini, Lorenzo Capineri Deep Learning-Based Real-Time Detection of Surface Landmines Using Optical Imaging. *Remote Sens.* 2024. URL: <https://www.mdpi.com/2072-4292/16/4/677>
2. Jasper Baur, Kyle Dewey, Gabriel Steinberg, Frank O. Nitsche. Modeling the Effect of Vegetation Coverage on Unmanned Aerial Vehicles-Based Object Detection: A Study in the Minefield Environment. *Remote Sens.*, 2024. URL: <https://www.mdpi.com/2072-4292/16/12/2046>
3. Sergey A. Stankevich, Ievgen Y. Saprykin. Optical and Magnetometric Data Integration for Landmine Detection with UAV. *WSEAS Transactions on Environment and Development*, 2024. DOI: 10.37394/232015.2024.20.96
4. Dillon Reis, Jordan Kupec, Jacqueline Hong, Ahmad Daoudi. Real-Time Flying Object Detection with YOLOv8, arXiv, 2023. URL: <https://arxiv.org/abs/2305.09972>
5. Ultralytics. “YOLOv8 Documentation,” Ultralytics Docs, 2025. URL: <https://docs.ultralytics.com/>
6. YOLOv5, YOLOv8 and YOLOv10 / Muhammad Hussain, et al. Muhammad Hussain, et al. The Go-To Detectors for Real-Time Object Detection, arXiv, 2024. URL: <https://arxiv.org/pdf/2407.02988>
7. Label Studio Team. Label Studio: Open Source Data Labeling Platform. Documentation, 2025. URL: <https://labelstud.io/>
8. Yolo model training dataset Yolo model training dataset. URL: https://github.com/robotkos/yolo_dataset
9. OpenAI. GPT-4o System Card, OpenAI, 2024. URL: <https://cdn.openai.com/gpt-4o-system-card.pdf>
10. Simon Monk “Raspberry Pi Cookbook: Software and Hardware Problems and Solutions 4th Edition”, 2023
11. PX4 Dev Team. Holybro Pixhawk 6C | PX4 User Guide, PX4 Docs, 2024. URL: https://docs.px4.io/main/en/flight_controller/pixhawk6c
12. MAVLink Community. “MAVLink Developer Guide,” 2025. URL: <https://mavlink.io/en/>
13. Jasper Baur, Kyle Dewey, Gabriel Steinberg, Frank O. Nitsche. A False-Positive-Centric Framework for Object Detection in Aerial Imagery, *Remote Sensing*, 2025. URL: <https://www.mdpi.com/2072-4292/17/14/2429>

Роботько Сергій¹, аспірант кафедри комп’ютеризованих систем управління, robotkos@gmail.com

Луцак Дмитро², старший викладач кафедри інформаційних технологій D.L.LUTSAK@GMAIL.COM

¹Національний університет кораблебудування імені адмірала Макарова,

²Університет короля Данила

Multilevel approach to uav-based detection of anti-personnel mines using computer vision

Abstract. Goal. Develop and validate a multilevel edge→ground→cloud decision pipeline for UAV-based detection of small anti-personnel mines (PFM-1), combining a fast visual detector with on-demand semantic verification. The objective is to maximize recall at a controlled false-alarm rate while preserving low latency, narrow bandwidth, and full temporal/geospatial traceability for audit and reproducibility. **Methodology.** The edge (Raspberry Pi) records 1080 p video time-aligned with Pixhawk 6C telemetry and forwards segments reliably (store-and-forward). The ground station runs Ultralytics YOLOv8n with frame-level inference, temporal aggregation of confidences (sY), and WGS-84 georeferencing. Events with intermediate confidence are escalated to the cloud, where a vision-language model (ChatGPT-4o Vision) returns a semantic score sC under a constrained YES/NO prompt. Final decisions use calibrated fusion $P_{final} = \lambda_Y s_Y + \lambda_C s_C$ with thresholding (generalizing AND/OR and k-of-n rules). All modules are containerized; timestamps, model versions, thresholds, and checksums are logged, with a unified GPS/NTP time axis. Field trials at $\approx 2\text{--}3$ m altitude over heterogeneous backgrounds with PFM-1 mock-ups yielded Precision 0.957, Recall 1.000, mAP@0.5 0.981, mAP@0.5:0.95 0.688 at ≈ 36 FPS on a field laptop. Selective escalation of ROI crops reduced false positives with negligible added latency and bandwidth. **Originality.** We provide the first reproducible end-to-end architecture that links detector confidence to a semantic VLM signal via an explicit

escalation rule and probabilistic calibration (Platt/temperature scaling), under an auditable data plane. The design formalizes timing and bandwidth budgets, remains resilient to link loss, and unifies temporal alignment, georeferencing, and provenance via containerized modules and machine-readable logs. **Practical value.** The system enables near-real-time post-flight analysis on commodity hardware, preserves high recall while suppressing clutter-driven false alarms, and reduces operator workload. By transmitting only ROI crops, it lowers communication energy demands and eases deployment in bandwidth-constrained field settings. The methodology supports governance and evidence through verifiable logs and is readily extensible to multi-class detection and to fusion with a metal detector within the same probabilistic framework, facilitating technology transfer to humanitarian demining workflows. **Key words:** humanitarian demining; UAV; vision-language models; wireless computer network; computer vision; YOLOv8; georeferencing; multilevel image analysis; ChatGPT.

Robotko Serhii, PhD student of the Department of Computerized Control Systems,
robotkos@gmail.com.

Lutsak Dmytro, Senior Lecturer of the Department of Information Technologies,
D.L.LUTSAK@GMAIL.COM

¹Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine

²King Danylo University, Ivano-Frankivsk, Ukraine.