

УДК 004.8

DOI <https://doi.org/10.36994/2788-5518-2025-02-10-08>

ЗАСТОСУВАННЯ АЛГОРИТМІВ ДЕТЕКЦІЇ ОБ'ЄКТІВ У БОРТОВИХ СИСТЕМАХ ВІДЕОАНАЛІЗУ БПЛА З ВИКОРИСТАННЯМ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ

Дьяконов О. С., к.т.н., доцент Національного університету кораблебудування ім. адм. Макарова, Миколаїв, Україна. <https://orcid.org/0000-0001-7438-7066>, alex.s.dyakonov@gmail.com

Ушкарєнко О. О., д.т.н., професор Національного університету кораблебудування ім. адм. Макарова, Миколаїв, Україна. <https://orcid.org/0000-0002-3159-330X>, oleksandr.ushkarenko@nuos.edu.ua

Обрубів А. В., д.т.н., доцент Національного університету кораблебудування ім. адм. Макарова, Миколаїв, Україна. <https://orcid.org/0000-0001-9667-1703>, andrii.obrubov@nuos.edu.ua

Нянько С. В., к.ю.н., ст.викл. Університету Короля Данила, Івано-Франківськ, Україна. <https://orcid.org/0009-0002-0218-8409>, s.nianko@icloud.com

Сірієчук А. С., к.т.н., доцент Національного університету кораблебудування ім. адм. Макарова, Миколаїв, Україна. <https://orcid.org/0000-0003-2927-2600>, sirivchuka@gmail.com

Анотація. Ефективна детекція об'єктів у реальному часі з використанням бортових систем комп'ютерного зору на безпілотних літальних апаратах (БПЛА) є фундаментальною задачею для багатьох прикладних галузей. Традиційні підходи часто не справляються з унікальними викликами, такими як обмежені обчислювальні ресурси (SWaP), висока швидкість руху, та складні умови зйомки з повітря. Використання алгоритмів глибокого навчання (Deep Learning) відкриває нові можливості для створення автономних систем, здатних аналізувати відеопотік безпосередньо на борту. У даній роботі проведено комплексний аналіз придатності моделі YOLOv8, як однієї з провідних архітектур, для задач детекції об'єктів з БПЛА, порівняно з альтернативними архітектурами глибоких нейронних мереж. Досліджено теоретичні можливості застосування різних варіантів моделей сімейства YOLO, а також архітектур Faster R-CNN, RetinaNet та EfficientDet в контексті специфічних умов аерозйомки. Проаналізовано технічні характеристики кожної архітектури, включаючи швидкість обробки, точність детекції на стандартних датасетах, обчислювальні вимоги та придатність для роботи в режимі реального часу. Виконано оцінку відповідності можливостей різних моделей специфічним вимогам складних задач детекції, зокрема виявленню малорозмірних та замаскованих об'єктів. Як приклад однієї з найбільш критичних таких задач розглядається виявлення вибухонебезпечних предметів з повітря. Результати аналізу демонструють, що YOLOv8 теоретично забезпечує оптимальний баланс між точністю виявлення та швидкістю обробки відеопотоку, необхідними для практичного застосування на борту БПЛА. Визначено ключові переваги та обмеження кожної архітектури, а також окреслено перспективи їх практичного впровадження для вирішення широкого кола задач, що вимагають автономного аналізу з борту БПЛА.

Ключові слова: комп'ютерний зір, аерозйомка, гуманітарне розмінування, безпілотні літальні апарати (БПЛА), штучна нейронна мережа, детекція об'єктів, інформаційно-комп'ютерна система, електронні елементи, глибоке навчання

APPLICATION OF OBJECT DETECTION ALGORITHMS IN UAV ONBOARD VIDEO ANALYSIS SYSTEMS USING DEEP LEARNING MODELS

Oleksii Dyakonov, Ph.D., Ass. Prof. Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine. <https://orcid.org/0000-0001-7438-7066>, alex.s.dyakonov@gmail.com

Oleksandr Ushkarenko, Dr. habil., Prof. Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine. <https://orcid.org/0000-0002-3159-3300>, oleksandr.ushkarenko@nuos.edu.ua

Andrii Obrubov, Dr. habil., Ass. Prof. Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine. <https://orcid.org/0000-0001-9667-1703>, andrii.obrubov@nuos.edu.ua

Stanislav Nianko, Ph.D., Senior Lecturer King Danylo University, Ivano-Frankivsk, Ukraine. <https://orcid.org/0009-0002-0218-8409>, s.nianko@icloud.com

Andrii Sirivchuk, Ph.D., Ass. Prof. Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine. <https://orcid.org/0000-0003-2927-2600>, sirivchuka@gmail.com

Abstract: Efficient real-time object detection using onboard computer vision systems on unmanned aerial vehicles (UAVs) is a fundamental task for many applied domains. Traditional approaches often fail to cope with unique challenges such as limited computational resources (SWaP), high movement speeds,

and complex aerial imaging conditions. This paper presents a comprehensive analysis of the suitability of the YOLOv8 model for UAV-based object detection tasks compared to Faster R-CNN, RetinaNet, and EfficientDet architectures. The technical characteristics of the models, including detection accuracy, processing speed, and computational requirements, are examined in the context of detecting small and camouflaged objects, such as explosive devices. The analysis results demonstrate that YOLOv8 provides an accuracy rate of 50.2 % (mAP50-95) with a processing speed of 60-65 frames per second and a computational complexity of 78.9 GFLOPS. This enables the system to operate on compact platforms with power consumption of 10–20 W. It is established that alternative architectures (Faster R-CNN, EfficientDet), despite comparable accuracy values (42–52 %), demonstrate significantly lower performance (12–30 frames per second) and require substantially more resources (150–240 GFLOPS). The obtained data confirm that YOLOv8 provides the optimal balance for practical implementation in autonomous UAV onboard monitoring systems.

Keywords: computer vision, aerial photography, humanitarian demining, unmanned aerial vehicles (UAVs), artificial neural network, object detection, information and computer system, electronic elements, deep learning

Вступ

Швидкий розвиток безпілотних літальних апаратів та методів комп'ютерного зору суттєво розширив можливості дистанційного моніторингу територій, об'єктів інфраструктури, транспортних потоків, сільськогосподарських угідь та зон надзвичайних ситуацій [1–3]. Сучасні БПЛА здатні виконувати тривалі польоти, отримувати високодеталізовані зображення та відеопотоки з різних висот і ракурсів, що дає змогу оперативно оцінювати стан місцевості без безпосередньої присутності людини в небезпечних або важкодоступних районах. Ефективність таких систем дедалі більше залежить не лише від льотно-технічних характеристик, а й від якості автоматизованого аналізу даних на борту, зокрема від здатності виявляти, класифікувати й відстежувати об'єкти у реальному часі в умовах обмежених обчислювальних ресурсів.

Аналіз останніх досліджень

Зі зростанням обсягів даних, які генерують бортові сенсори, ручна обробка відеоінформації стає малопридатною для оперативних задач. Аналіз тривалих місій вимагає значних людських ресурсів, виснажує операторів і підвищує ймовірність помилок, пов'язаних з людським фактором [4]. Інтеграція алгоритмів штучного інтелекту та комп'ютерного зору в бортові системи БПЛА дає змогу автоматизувати обробку аерофотознімків і відеопотоків, підвищити швидкість прийняття рішень, забезпечити масштабованість моніторингу великих територій та зменшити залежність від суб'єктивного сприйняття аналітиків [5; 6]. Глибокі нейронні мережі, зокрема сучасні архітектури детекції об'єктів, продемонстрували високі показники точності та стійкості в умовах реального середовища. Особливу увагу привертають моделі, здатні працювати у режимі реального часу на вбудованих обчислювальних платформах, що є ключовою вимогою для бортових систем БПЛА. До таких рішень належать сімейства моделей YOLO, а також інші одностадійні та двостадійні детектори, які забезпечують компроміс між точністю, швидкодією та ресурсомісткістю [7]. Останні версії моделей, зокрема YOLOv8, характеризуються вдосконаленою архітектурою, покращеним представленням ознак, більш ефективними механізмами фокусування уваги та оптимізованими підходами до постобробки детекцій [8], що робить їх перспективними для використання на борту безпілотних платформ.

Одним із важливих напрямів застосування таких систем є підвищення ефективності виявлення потенційно небезпечних об'єктів на поверхні, включно з вибухонебезпечними, без прямого залучення людини до зони ризику. У цьому контексті БПЛА з інтегрованими моделями глибокого навчання можуть доповнювати традиційні методи обстеження, слугувати інструментом попереднього скринінгу та пріоритизації ділянок для детальної дослідження. Водночас такі сценарії є лише окремим прикладом ширшого класу задач, пов'язаних з автоматизованим аналізом сцен, виявленням малих і слабо контрастних об'єктів, відстеженням динамічних цілей та контролем змін у просторі й часі.

Метою даної роботи є дослідження можливостей застосування сучасних алгоритмів детекції об'єктів на основі глибоких нейронних мереж у бортових системах відеоаналізу БПЛА. Основна увага приділяється порівнянню архітектур з точки зору точності, швидкодії, стійкості до завад і придатності для реалізації на обмежених обчислювальних платформах, а також оцінці їхнього потенціалу для вирішення прикладних задач дистанційного моніторингу, включно з підвищенням безпечності операцій у зонах підвищеного ризику.

Виклад основного матеріалу

Проблематика бортових систем комп'ютерного зору БПЛА та еволюція архітектур детекції об'єктів. Застосування комп'ютерного зору на борту безпілотних літальних апаратів пов'язане з низкою специфічних викликів, що відрізняють такі системи від класичних сценаріїв обробки зображень. Бортові алгоритми мають працювати з аерофотоданими, у яких об'єкти інтересу часто є малорозмірними, слабо контрастними та частково замаскованими елементами довкілля. Йдеть-

ся про транспортні засоби, інженерні споруди, технічні елементи інфраструктури, засоби зв'язку, фрагменти техніки, а також потенційно небезпечні об'єкти, включно з вибухонебезпечними [9]. Такі об'єкти можуть мати суттєву варіативність форми, розмірів і текстурних характеристик, змінювати зовнішній вигляд під впливом погодних умов, забруднення, корозії або тривалого перебування у ґрунті чи на відкритій місцевості, що ускладнює їх надійну автоматизовану ідентифікацію [9; 10]. Додатковим ускладнювальним фактором є те, що об'єкти спостереження нерідко частково перебиваються рослинністю, уламками, ґрунтом або штучними конструкціями. Це призводить до появи лише фрагментарних ознак на зображенні та зменшує інформативність локальних контекстів [10]. Дослідження, виконані для задач дистанційного моніторингу та виявлення небезпечних об'єктів, демонструють, що значна частина цілей може бути частково прихованою, що потребує від детекторів підвищеної чутливості до слабо виражених і неповних візуальних патернів [9–11]. Умови зйомки з борту БПЛА суттєво варіюються за освітленням, кутом огляду, висотою польоту, станом атмосфери та динамікою руху платформи [11]. Зміни освітленості протягом доби, відблиски, тіні, туман, дощ, пил або дим призводять до деградації якості зображень та нестабільності ознак. Вібрації і мікрорухи камери спричиняють розмиття, особливо на малих витримках, тоді як змінні кути огляду між приблизно 45 і 90 градусами впливають на перспективні спотворення та вигляд об'єктів. Роздільна здатність та масштаб цілі напряму залежать від висоти польоту. Наприклад, об'єкт розміром близько 10 см при зйомці з 30 м займає порядку двох десятків пікселів у межах кадру камери 12 Мп, а на висоті 75 м його проекція скорочується до кількох пікселів, що наближається до порогу надійної детекції [12]. Співвідношення площі цілі до площі зображення часто є меншим за десяти частки відсотка, що призводить до суттєвого дисбалансу класів та потребує спеціалізованих методів формування вибірок, стратифікованого семплінгу, важчених функцій втрат та технік *hard example mining* [12].

Для практичного використання у складі бортових систем відеоаналізу БПЛА алгоритми детекції об'єктів мають відповідати сукупності вимог до швидкодії, точності, робастності та ресурсоемності. Обробка відеопотоку в режимі, наближеному до реального часу, вимагає підтримки частоти щонайменше 25–30 кадрів за секунду, що забезпечує безперервність спостереження та своєчасність реакції оператора або автономної системи керування. У задачах моніторингу безпеки, включно з виявленням потенційно вибухонебезпечних об'єктів, особливо критичною є висока чутливість системи, отже метрика *recall* повинна досягати значень понад 90 % для зменшення ймовірності пропуску небезпечної цілі [9–12]. Водночас надмірна кількість хибнопозитивних спрацювань знижує ефективність використання ресурсів, ускладнює подальший аналіз та може перевантажувати операторів, тому необхідний контроль рівня *false positive* у межах, прийнятних для конкретного сценарію застосування. Апаратні обмеження БПЛА диктують вимоги до компактності та енергоефективності бортових обчислювальних засобів. Для дронів малого та середнього класу корисне навантаження зазвичай обмежене сотнями грамів до кількох кілограмів, а споживана потужність системи комп'ютерного зору має залишатися на рівні, що не призводить до істотного скорочення тривалості польоту, орієнтовно до 15–20 Вт. Це стимулює використання оптимізованих моделей глибокого навчання, прискорених бібліотек та технік квантування. Також важливо забезпечити стабільну роботу алгоритмів у широкому діапазоні умов освітлення, типів підстилаючої поверхні та рельєфу, а для прикладних задач геопросторового аналізу необхідна коректна просторово-часова прив'язка детекцій до координат місцевості з точністю порядку 1–2 метрів, що досягається інтеграцією з навігаційними системами та сенсорами положення.

Розвиток глибокого навчання для детекції об'єктів за останнє десятиліття визначив перехід від двостадійних до одностадійних архітектур, що є принципово важливим для бортового аналізу даних. Перші підходи, представлені сімейством R-CNN, використовували попередню генерацію регіонів-кандидатів із подальшою екстракцією ознак та класифікацією згортковими нейронними мережами. Модифікація Fast R-CNN об'єднала ці етапи та підвищила ефективність обчислень [13], а Faster R-CNN ввела мережу Region Proposal Network, що забезпечило наскрізне навчання та зменшило час обробки [14]. Подальші рішення, такі як Cascade R-CNN, застосували каскадне уточнення меж об'єктів з послідовним підвищенням порогів IoU, а Mask R-CNN додала можливість сегментації на рівні пікселів, що є релевантним для задач точної локалізації об'єктів складної форми. Ці підходи демонструють високі значення mAP на стандартних наборах даних на кшталт COCO, але потребують значних обчислювальних ресурсів і зазвичай не забезпечують необхідної швидкодії на вбудованих платформах, обмежуючи їх придатність до використання безпосередньо на борту БПЛА. Одностадійні детектори стали відповіддю на потребу в поєднанні точності та високої швидкості. Архітектура YOLO, запропонована у 2016 році [15], інтерпретувала детекцію об'єктів як єдину регресійну задачу для координат обмежувальних рамок та ймовірностей класів, що дозволило досягти обробки десятків і більше кадрів за секунду на сучасних GPU і вбудованих прискорювачах.

Порівняльний аналіз архітектур нейронних мереж для бортових систем детекції об'єктів БПЛА. Оцінювання придатності сучасних архітектур глибоких нейронних мереж для використання

у бортових системах відеоаналізу безпілотних літальних апаратів потребує врахування сукупності технічних та експлуатаційних параметрів. До ключових критеріїв належать базова точність на репрезентативних датасетах детекції об'єктів, швидкість обробки на різних апаратних платформах, обчислювальна складність, обсяг оперативної пам'яті, здатність до виявлення дрібних об'єктів з розміром проєкції менш ніж 32×32 пікселі, стійкість до варіативних умов аерофотозйомки та потенціал оптимізації для розгортання на обладнанні з обмеженими ресурсами. Для задач дистанційного моніторингу, інфраструктурного нагляду, виявлення транспортних засобів, техніки, інженерних елементів та інших критичних об'єктів ці характеристики визначають можливість інтеграції алгоритмів безпосередньо на борту БПЛА. Окремим важливим напрямом є застосування таких моделей для підвищення ефективності виявлення потенційно вибухонебезпечних об'єктів, де поєднання високого показника recall та робастності до складних умов зйомки має критичне значення.

Модель YOLOv8 у конфігурації medium демонструє на датасеті COCO значення mAP50 близько 63,9 і mAP50-95 на рівні 50,2 % при швидкості обробки орієнтовно 60–65 кадрів за секунду на GPU NVIDIA RTX 3090 та 18–22 кадри за секунду на CPU Intel Core i9-12900K [16]. Кількість параметрів становить приблизно 25,9 млн при обсязі моделі близько 52 МБ та обчислювальній складності 78,9 GFLOPS для вхідного зображення 640×640 пікселів [16]. Такі показники свідчать про задовільний компроміс між точністю та ресурсоемністю, що є важливим для бортових систем із обмеженими енергетичними та масогабаритними характеристиками. Багаторівнева структура детекції на основі Feature Pyramid Network з трьома масштабами (P3, P4, P5) забезпечує покращену чутливість до об'єктів різного розміру, включно з дрібними наземними цілями, характерними для аерозйомки. Архітектура YOLOv8 використовує backbone із Cross Stage Partial connections, що зменшує дублювання обчислень та підвищує ефективність використання параметрів. Модуль Spatial Pyramid Pooling Fast агрегує ознаки на різних просторових масштабах, що підсилює здатність моделі коректно описувати як великі об'єкти інфраструктури, так і компактні цілі. Застосування anchor-free підходу, за якого мережа прогнозує параметри обмежувальних рамок без фіксованого набору anchor boxes, підвищує гнучкість щодо форм, масштабів та співвідношень сторін виявлюваних об'єктів [17]. Це є принципово корисним для задач аналізу аерофотозображень, де спостерігається значна варіативність геометрії та проєкції цілей, а також для сценаріїв пошуку потенційно небезпечних об'єктів зі складними контурами та частковим маскуванням. YOLOv8 доповнена вдосконаленою функцією втрат, що поєднує Distribution Focal Loss для класифікації та Complete IoU Loss для регресії меж рамок, завдяки чому досягається кращий баланс між якістю локалізації та розмежуванням класів. Використання task-aligned assigner сприяє більш коректному формуванню позитивних і негативних зразків під час навчання на даних з домінуванням фону, що є типовим для аерофотозйомки над великими територіями [17]. Додаткове застосування механізмів просторово-канального виділення важливих ознак дозволяє моделі зосереджуватись на інформативних ділянках сцени, підвищуючи стійкість до шумів, неоднорідного фону та часткового перекриття об'єктів. Сукупність цих рішень робить YOLOv8 перспективною для бортового розгортання на БПЛА у задачах широкого спектра, включно з раннім виявленням небезпечних цілей.

Попереднє покоління архітектур, представлене YOLOv7, забезпечує mAP50 близько 56,8 % та mAP50-95 на рівні 43,5 % на COCO при швидкості 55–60 кадрів за секунду на GPU RTX 3090 [18]. При цьому модель містить орієнтовно 36,9 млн параметрів і вимагає близько 104,7 GFLOPS, що робить її більш ресурсомісткою порівняно з YOLOv8 [18]. Ключовим конструктивним елементом виступає Extended Efficient Layer Aggregation Network, який забезпечує ефективну агрегацію ознак з різних рівнів. Підхід до масштабування архітектури дозволяє добирати компроміс між швидкістю та точністю. Водночас використання anchor-based механізму детекції дещо знижує гнучкість при роботі з об'єктами сильно змінних форм та масштабів, що особливо відчутно в умовах аерозйомки зі змінною висотою і ракурсами. Таким чином, YOLOv7 залишається конкурентоспроможною для високопродуктивних наземних систем обробки даних, але потребує додаткової оптимізації для повноцінного застосування на борту БПЛА.

Модель YOLOv5 у варіанті medium демонструє mAP50 близько 56,0 % і mAP50-95 на рівні 45,4 % при дуже високій швидкості, яка може досягати 70–75 кадрів за секунду завдяки меншій кількості параметрів (приблизно 21,2 млн) та обчислювальній складності близько 48,2 GFLOPS [19]. Обсяг моделі на рівні 42 МБ робить її привабливою для систем із жорсткими обмеженнями за пам'яттю та енергоспоживанням. Архітектура поєднує CSPDarknet як backbone і PANet для багаторівневої агрегації ознак, що забезпечує стабільну роботу в типових задачах детекції об'єктів у відеопотоці. Водночас попередні версії YOLO поступаються новішим модифікаціям за ефективністю виявлення дуже дрібних об'єктів у складних сценах, зокрема на великих висотах польоту, де проєкція цілі на зображенні має невелику кількість пікселів, що обмежує їх застосування у більш вимогливих сценаріях аероспостереження.

Faster R-CNN з backbone ResNet-101 демонструє на датасеті COCO одні з найвищих показників точності серед класичних архітектур, досягаючи mAP50 близько 59 % та mAP50-95 42 % [20]. Двостадійна природа архітектури забезпечує високу якість локалізації об'єктів з precision на рівні 85–90 %

для високих порогів IoU понад 0.75 та меншу кількість хибнопозитивних спрацювань порівняно з одностадійними детекторами. Однак швидкість обробки на рівні від 12 до 15 кадрів за секунду навіть на потужних GPU NVIDIA RTX 3090 та обчислювальна складність близько 240 GFLOPS створюють значні обмеження для практичного застосування на борту БПЛА [21].

RetinaNet з backbone ResNet-50 представляє компромісний підхід, поєднуючи елементи одностадійної та двостадійної архітектур. Модель досягає точності mAP50 на рівні 55 % та mAP50-95 39,1 % при швидкості близько 30–35 кадрів за секунду на GPU RTX 3090 [22]. Ключовою особливістю є використання focal loss функції втрат, що модифікує стандартну cross-entropy втрату шляхом додавання модулюючого фактора $(1-pt)^{\gamma}$, де pt є ймовірністю правильного класу, а γ зазвичай дорівнює 2 [23]. Це дозволяє ефективніше навчати модель на несбалансованих датасетах з великою кількістю фонових регіонів, зменшуючи вплив легко класифікованих негативних прикладів.

Focal loss теоретично може бути значною перевагою для задач виявлення мін, де співвідношення позитивних до негативних прикладів може досягати 1:1000 або більше. Feature Pyramid Network з п'яти рівнями від P3 до P7 забезпечує хорошу детекцію об'єктів з великим діапазоном розмірів від 32x32 до 512x512 пікселів [23], що важливо для одночасного виявлення протитанкових мін на низьких висотах та протипіхотних мін на більших висотах польоту. Архітектура використовує дев'ять anchor boxes на кожній просторовій позиції з трьома співвідношеннями сторін 1:2, 1:1, 2:1 та трьома масштабами, що забезпечує гнучкість у виявленні об'єктів різних форм.

EfficientDet представляє сімейство моделей різної складності від D0 до D7, розроблених з використанням compound scaling для збалансованого масштабування всіх компонентів мережі включаючи роздільність вхідного зображення, глибину backbone, ширину каналів та глибину BiFPN [24]. Варіант EfficientDet-D4 демонструє точність на рівні 52,2 % mAP50-95 при швидкості близько 27–30 кадрів за секунду на GPU RTX 3090 та містить 20,7 мільйона параметрів при обчислювальній складності 55.2 GFLOPS [25]. Архітектура характеризується високою ефективністю параметрів, досягаючи конкурентної точності при меншій кількості параметрів порівняно з іншими моделями аналогічної продуктивності.

Трансформерні архітектури на кшталт DETR (Detection Transformer) та його варіантів Deformable DETR, Conditional DETR представляють принципово інший підхід до детекції об'єктів, використовуючи механізми self-attention для глобального аналізу зображення замість локальних згорткових операцій [26]. DETR використовує CNN backbone для екстракції ознак з подальшою обробкою через encoder-decoder трансформер, що генерує фіксовану кількість передбачень (зазвичай 100 об'єктів). Модель досягає точності близько 42 % mAP50-95 на COCO при швидкості обробки лише 8–10 кадрів за секунду на потужному GPU A100 [27]. Обчислювальна складність становить понад 180 GFLOPS при 41,3 мільйонах параметрів.

DETR використовує bipartite matching для призначення передбачень до ground truth об'єктів під час навчання, що теоретично спрощує pipeline детекції, усуваючи необхідність у non-maximum suppression. Однак модель вимагає значних обчислювальних ресурсів для навчання порядку 300–500 епох порівняно з 100–200 для YOLO та потребує великих обсягів даних для досягнення конкурентної точності. Deformable DETR покращує швидкість збіжності та точність детекції дрібних об'єктів шляхом використання deformable attention модулів, що фокусуються на невеликій кількості ключових точок замість повного простору ознак, підвищуючи точність до 46,9 % mAP50-95 при швидкості 19 FPS. Незважаючи на теоретичні переваги глобального аналізу контексту, практичне застосування трансформерних архітектур на борту БПЛА наразі обмежене їх високими обчислювальними вимогами та повільною швидкістю інференсу.

Аналіз потенційного впливу висоти польоту на можливість детекції різними моделями показує критичну залежність від роздільної здатності об'єктів на зображенні та здатності архітектури виявляти дрібні об'єкти. При польоті на висоті 30 м з камерою роздільністю 12 Мп протипіхотна міна діаметром 10 см займатиме близько 15–50 пікселів на зображенні, що відповідає категорії small objects згідно з класифікацією COCO area менше 32x32 пікселів. На цій висоті YOLOv8 теоретично може досягати recall понад 85 % для таких об'єктів завдяки ефективній Feature Pyramid Network з детекцією на рівні P3 з роздільністю 80x80 комірок. Збільшення висоти до 50 м зменшить розмір об'єкта до 5–15 пікселів, що знижує теоретичний recall до 70–75 % для моделей без спеціалізованої оптимізації.

На висоті 75 м протипіхотна міна займатиме лише 8–10 пікселів, що наближається до нижньої межі можливостей детекції для більшості архітектур. EfficientDet з BiFPN теоретично може демонструвати кращу продуктивність на великих висотах завдяки ефективній агрегації ознак з кількох масштабів. Faster R-CNN з RPN також може мати переваги для виявлення граничних дрібних об'єктів завдяки генерації великої кількості регіонів-кандидатів різних масштабів. Протитанкові міни діаметром 30 см при тих же умовах займатимуть 45–150 пікселів на висоті 30 м та 15–50 пікселів на 50 м, що значно полегшує їх детекцію для всіх розглянутих архітектур.

Приклад апаратної платформи для розгортання системи детекції на БПЛА показано на рис. 1. Nvidia Jetson Nano з модулем камери представляє компактне рішення для бортових обчислень, а схема підключення платформи Jetson до БПЛА Skypad TBS Discovery наведена на рис. 2.

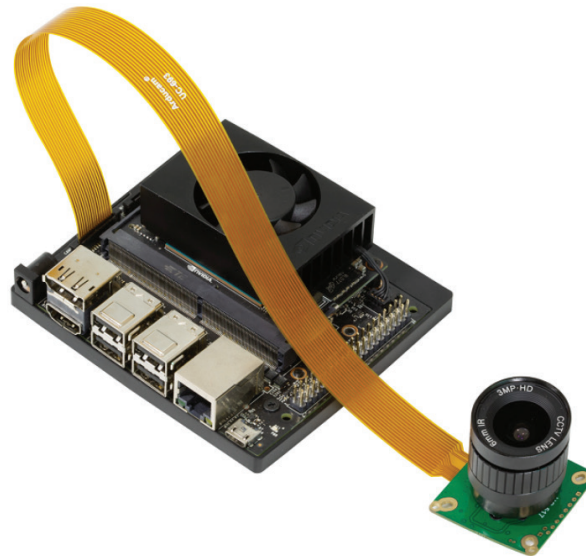


Рис. 1. Nvidia Jetson Nano з модулем камери

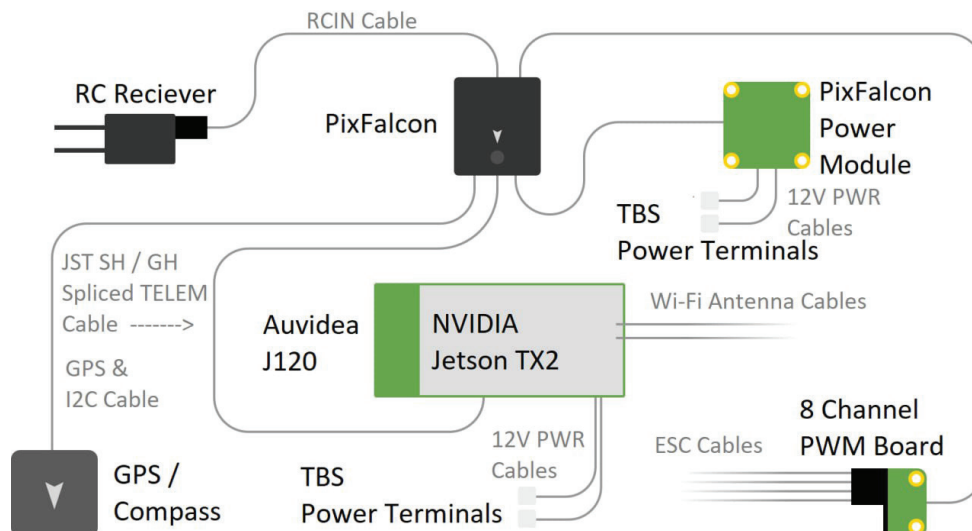


Рис. 2. Функціональна схема підключення платформи Jetson до БПЛА Skypad TBS Discovery

Питання енергоефективності та можливості розгортання на обладнанні з обмеженими ресурсами є критичним для практичного застосування систем детекції на борту БПЛА. YOLOv8 у варіанті nano містить лише 3,2 мільйона параметрів при розмірі 6 МБ та обчислювальній складності 8.7 GFLOPS, що дозволяє розгорнути її на компактних обчислювачах типу NVIDIA Jetson Nano з 4 ГБ пам'яті та споживанням потужності 5–10 Вт. На Jetson Nano YOLOv8-nano теоретично може досягати швидкості 15–20 FPS з точністю близько 37,3 % mAP50-95 на стандартних датасетах. Більш потужний Jetson Xavier NX з 8 ГБ пам'яті та споживанням 10–15 Вт може забезпечити обробку YOLOv8-small зі швидкістю 25–30 FPS при точності 44,9 % mAP50-95.

Графік залежності точності детекції від швидкості обробки для різних моделей представлено на рис. 3. Цей графік демонструє компроміс між точністю та швидкістю для різних архітектур нейронних мереж. Моделі, розташовані ближче до правого верхнього кута, є найбільш привабливими для практичного застосування. Дані по точності та швидкості для цих моделей наведені в табл. 1.

Графік на рис. 4 ілюструє критичну залежність якості детекції від висоти польоту БПЛА для протипілотних мін діаметром 10 см при використанні камери 12 Мп. Чисельні залежності теоретичного recall YOLOv8 та EfficientDet від висоти польоту БПЛА наведені у табл. 2.

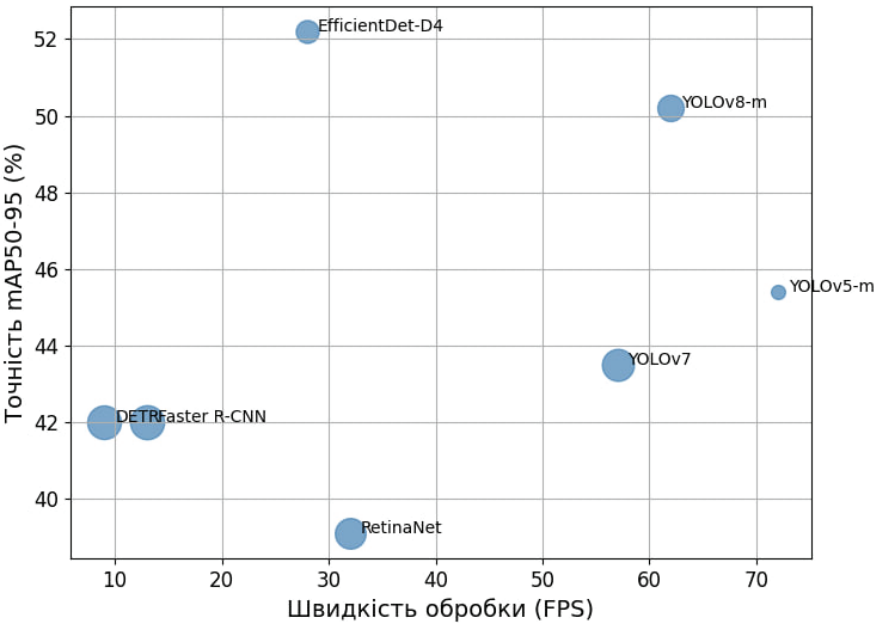


Рис. 3. Графік залежності точності детекції (mAP50-95) від швидкості обробки (FPS) для різних моделей

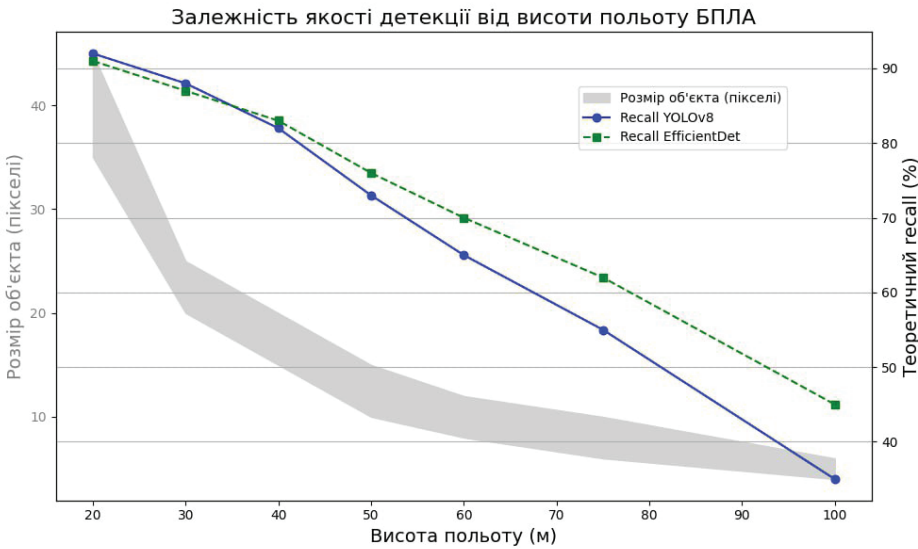


Рис. 4. Вплив висоти польоту БПЛА на розмір об'єкта в пікселях та теоретичний recall

Таблиця 1
Дані точності та швидкості

Модель	mAP50-95, %	Швидкість GPU, FPS
DETR	42.0	9
Faster R-CNN	42.0	13
YOLOv7	43.5	57
YOLOv5-m	45.4	72
YOLOv8-m	50.2	62
EfficientDet-D4	52.2	28
RetinaNet	39.1	32

Таблиця 2

Чисельні залежності теоретичного recall YOLOv8 та EfficientDet від висоти польоту БПЛА

Висота польоту, м	Розмір об'єкта, пікселі	Теоретичний recall YOLOv8, %	Теоретичний recall EfficientDet, %
20	35–45	92	91
30	20–25	88	87
40	15–20	82	83
50	10–15	73	76
60	8–12	65	70
75	6–10	55	62
100	4–6	35	45

Примітки до даних:

1. Дані про продуктивність моделей базуються на тестуванні на датасеті COCO 2017.
2. Швидкість вимірювалася на GPU NVIDIA RTX 3090.
3. Теоретичні значення recall для різних висот є екстраполяцією на основі продуктивності моделей на дрібних об'єктах.

4. Розмір об'єкта розрахований для камери 12 Мп з фокусною відстанню 24 мм (еквівалент).

Квантизація моделей до 8-бітного цілочисельного представлення INT8 потенційно дозволяє зменшити розмір у 4 рази та прискорити інференс на 2–3 рази при типовій втраті точності лише 1–3 % mAP. Для YOLOv8-medium квантизація може зменшити розмір з 52 МБ до 13 МБ при збереженні mAP50-95 на рівні 48–49 %. Спеціалізовані прискорювачі на базі Intel Movidius Myriad X зі споживанням 1–2 Вт або Google Coral Edge TPU 2 Вт теоретично можуть забезпечити обробку YOLOv8-nano зі швидкістю 20–40 FPS при ультранизькому енергоспоживанні, що робить можливим використання на мініатюрних БПЛА вагою менше 1 кг. Для великих БПЛА корисним навантаженням понад 3–5 кг можливе використання компактних GPU на кшталт NVIDIA RTX 4060 Mobile зі споживанням 40–60 Вт, що дозволить обробляти YOLOv8-medium або large зі швидкістю 40–50 FPS при потенційно вищій точності. Важливим аспектом є також енергоефективність зв'язку для передачі відеопотоку: при використанні бортової обробки необхідна передача лише координат виявлених об'єктів порядку 100–500 байт на кадр замість повного відеопотоку 5–20 Мбіт/с, що значно знижує енергоспоживання та збільшує радіус дії БПЛА. Робастність до різних умов освітлення є важливим фактором для цілодобової роботи систем виявлення мін. Архітектури з механізмами attention, такі як YOLOv8 та трансформерні моделі теоретично можуть краще адаптуватися до змінних умов освітлення, фокусуючись на найбільш інформативних ознаках незалежно від загальної яскравості зображення. Дослідження показують, що моделі, навчені з агресивною аугментацією освітлення, включаючи зміни яскравості $\pm 50\%$, контрасту $\pm 40\%$ та насиченості $\pm 30\%$, демонструють деградацію точності не більше 10–15 % при переході від оптимальних до субоптимальних умов освітлення.

Порівняльні характеристики моделей детекції об'єктів наведені в табл. 3.

Таблиця 3

Порівняльні характеристики моделей детекції об'єктів

Модель	mAP50-95, %	mAP50, %	FPS GPU	FPS CPU	Параметри, млн	GFLOPS	Розмір, МБ	Придатність
YOLOv8-m	50,2	63,9	60–65	18–22	25,9	78,9	52	Висока
YOLOv7	56,8	55–60	14–18	36,9	104,7	75	43,5	Середня
Faster R-CNN	42,0	59,0	12–15	3–5	41,8	240	167	Низька
RetinaNet	39,1	55,0	30–35	8–10	36,4	145	145	Середня
EfficientDet-D4	52,2	68,4	27–30	6–8	20,7	55,2	83	Середня
DETR	42,0	62,4	8–10	2–3	41,3	180	165	Низька

Використання інфрачервоних (IR) або теплових камер у нічний час вимагає адаптації моделей до монохромних зображень з іншим розподілом контрасту порівняно з RGB. Міни можуть мати відмінну від навколишнього середовища температуру завдяки різниці у теплоємності та теплопровідності матеріалів, що може надавати додаткову інформацію для детекції. Однак ефективність теплової зйомки залежить від часу доби, погодних умов та типу ґрунту. Мультиспектральні системи, що поєднують RGB, IR та теплові канали, теоретично можуть значно підвищити робастність та точність детекції за рахунок комплементарної інформації з різних частин спектру.

Практичне значення одержаних результатів полягає у розробці алгоритмічного забезпечення та рекомендацій щодо налаштування гіперпараметрів детекторів YOLOv8, що дозволяє імплементувати систему автоматизованого пошуку вибухонебезпечних предметів на серійних БПЛА з продуктивністю обробки не менше 25 кадрів за секунду. Результати дослідження можуть бути використані підрозділами ДСНС та операторами протимінної діяльності для прискорення обстеження територій.

Висновки

Розглянуті технічні характеристики та функціональні можливості сучасних архітектур глибоких нейронних мереж свідчить, що сімейство моделей YOLOv8 є одним із найбільш перспективних варіантів для використання в бортових системах відеоаналізу безпілотних літальних апаратів. Поеднання показника точності на рівні близько 50,2 % mAP50-95 на стандартних наборах даних, високої швидкості обробки порядку 60–65 кадрів за секунду та обчислювальної складності близько 78,9 GFLOPS забезпечує можливість функціонування в режимі, наближеному до реального часу, на компактних енергоефективних платформах з потужністю на рівні 10–20 Вт. Наявність моделей різної місткості від nano до extra-large полегшує адаптацію до конкретних обмежень корисного навантаження БПЛА, вимог до продуктивності та цільових сценаріїв застосування. Anchor-free підхід, удосконалені механізми агрегації багатомасштабних ознак і сучасні стратегії формування функції втрат підвищують гнучкість моделі під час роботи з об'єктами різних розмірів, форм і співвідношень сторін, що є принциповим для аерофотозйомки складних сцен.

Порівняння з альтернативними архітектурами, такими як Faster R-CNN, EfficientDet та інші двостадійні або гібридні детектори, показує, що попри потенційно вищі значення mAP50-95 у діапазоні приблизно 42–52 %, їх практичне використання в якості бортових модулів для аналізу відеопотоку обмежується нижчою швидкістю (орієнтовно 12–30 кадрів за секунду) та зростанням обчислювальних витрат до рівня 150–240 GFLOPS. Такі вимоги ускладнюють впровадження на енергообмежених платформах та знижують доцільність використання у режимі реального часу при тривалих місіях моніторингу. На цьому тлі конфігурації на основі YOLOv8 забезпечують більш збалансоване співвідношення точності, швидкодії та ресурсоефективності, що робить їх придатними для широкого спектра задач дистанційного спостереження, включно з моніторингом інфраструктури, транспортних потоків, пошуково-рятувальними операціями та попереднім виявленням потенційно небезпечних об'єктів без безпосередньої присутності людини в зоні ризику.

Література:

1. Landmine Monitor, 2023, Current conflicts, long-lasting contamination cause high number of mine casualties. *Humanity & Inclusion US*. URL: <https://www.hi-us.org/en/landmine-monitor-2023>
2. Facts About Landmines. *Minesweepers*. URL: <https://landminefree.org/facts-about-landmines/>
3. Minesweepers and SDGs. *Minesweepers*. URL: <https://landminefree.org/blog/minesweepers-and-sdgs/>
4. Heuschmid D., Wacker O., Zimmermann Y., Penava P., Buettner R. Advancements in Landmine Detection: Deep Learning-Based Analysis With Thermal Drones. *IEEE Access*. 2025. Vol. 13. P. 91777–91794. DOI: 10.1109/ACCESS.2025.3572196.
5. Ball M. Landmine Detection with Drones. 2021. URL: <https://www.unmannedsystemstechnology.com/2021/08/landmine-detection-with-drones/>
6. Baur J., Steinberg G., Nikulin A., de Smet T. Applying Deep Learning to Automate UAV-Based Detection of Scatterable Landmines. *Remote Sensing*. 2020. Vol. 12, No. 5. P. 859. DOI: 10.3390/rs12050859.
7. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 779–788. DOI: 10.48550/arXiv.1506.02640.
8. Jocher G., Chaurasia A., Qiu J. Ultralytics YOLOv8. 2023. URL: <https://docs.ultralytics.com/models/yolov8/>
9. Joint Fusion and Detection via Deep Learning in UAV-Borne Multispectral Sensing of Scatterable Landmine. *Sensors*. 2023. DOI: 10.3390/s23125693.
10. How to Implement Drones and Machine Learning to Reduce Time, Costs, and Dangers Associated with Landmine Detection. *The Journal of Conventional Weapons Destruction*. 2021. URL: <https://commons.lib.jmu.edu/cgi/viewcontent.cgi?article=2951&context=cisr-journal>
11. Air Drones for Explosive Landmines Detection. *Springer*. 2014. DOI: 10.1007/978-3-319-03653-3_9.
12. Advancements in Landmine Detection: Deep Learning-Based Analysis With Thermal Drones. *IEEE Xplore*. 2024. DOI: 10.1109/ACCESS.2025.3572196
13. Girshick R. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision. 2015. P. 1440–1448. DOI: 10.48550/arXiv.1504.08083
14. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*. 2015. Vol. 28. P. 91–99. DOI: 10.48550/arXiv.1506.01497.
15. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection, CVPR. 2016. DOI: 10.1109/CVPR.2016.91.
16. YOLOv8: State-of-the-Art Computer Vision Model, Roboflow. URL: <https://yolov8.com/>
17. A Review on YOLOv8 and Its Advancements, ResearchGate. 2024. DOI: 10.1007/978-981-99-7962-2_39.
18. Wang C. Y., Bochkovskiy A., Liao H. Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. DOI: 10.48550/arXiv.2207.02696.
19. Jocher G. YOLOv5 [Електронний ресурс]. 2020. URL: <https://github.com/ultralytics/yolov5>

20. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017. Vol. 39. No. 6. P. 1137-1149. DOI: 10.48550/arXiv.1506.01497
21. Faster R-CNN Explained for Object Detection Tasks. *DigitalOcean*. 2025. URL: <https://www.digitalocean.com/community/tutorials/faster-r-cnn-explained-object-detection>
22. RetinaNet: The beauty of Focal Loss. *Towards Data Science*. 2025. URL: <https://towardsdatascience.com/retinanet-the-beauty-of-focal-loss-e9ab132f2981/>
23. Lin T. Y., Goyal P., Girshick R., He K., Dollár P. Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. 2017. P. 2980–2988. DOI: 10.48550/arXiv.1708.02002.
24. Tan M., Pang R., Le Q. V. EfficientDet: Scalable and Efficient Object Detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020. P. 10781–10790. DOI: 10.48550/arXiv.1911.09070.
25. EfficientDet: Towards Scalable and Efficient Object Detection. *Google Research Blog*. 2020. URL: <https://research.google/blog/efficientdet-towards-scalable-and-efficient-object-detection/>
26. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. End-to-End Object Detection with Transformers. *European Conference on Computer Vision (ECCV)*. 2020. P. 213–229. DOI: 10.48550/arXiv.2005.12872.
27. Shehzadi T., Hashmi K. A., Stricker D., Afzal M. Z. Object Detection with Transformers: A Review. *arXiv*. 2023. URL: <https://arxiv.org/abs/2306.04670>

Дата першого надходження статті до видання: 21.10.2025
Дата прийняття статті до друку після рецензування: 24.11.2025
Дата публікації (оприлюднення) статті: 26.12.2025