

УДК 004.932:004.8:629.735

DOI: 10.18372/2073-4751.86.21283

**Троцький Я. В.,**  
orcid.org/0009-0000-4725-9544,  
e-mail: 4640007@stud.kai.edu.ua,

**Зіатдінов Ю. К.,** д.т.н.,  
orcid.org/0000-0003-2035-7376,  
e-mail: yuriy.ziatdinov@npp.kai.edu.ua,

## АНАЛІЗ СУЧАСНИХ МЕТОДІВ КОМП'ЮТЕРНОГО ЗОРУ ДЛЯ МОНІТОРИНГУ ТА ІДЕНТИФІКАЦІЇ ЛІТАЮЧИХ ОБ'ЄКТІВ: ВИКЛИКИ ТА ПЕРСПЕКТИВИ

Державний університет “Київський авіаційний інститут”

### **Вступ**

Стрімкий розвиток комп'ютерних технологій та БПЛА зумовив потребу в ефективному контролі повітряного простору. Традиційні радіолокаційні системи часто є дорогими або малоефективними для виявлення малорозмірних цілей, таких як безпілотні коптери чи птахи, що актуалізує пошук альтернативних підходів до моніторингу.

Одним із перспективних рішень є методи комп'ютерного зору, які поєднують алгоритми обробки зображень, глибинного навчання та інтелектуального аналізу даних. На відміну від традиційних сенсорних систем, вони дозволяють не лише виявляти об'єкти, а й визначати їх тип, параметри руху та інші характеристики.

Попри високу ефективність, такі системи мають низку проблем, зокрема вплив шумів і перекриття поля зору [1], складність керування камерами та обмеженість якісних датасетів [2]. Це зумовлює необхідність систематизації сучасних підходів і оцінки їх ефективності для реальних умов застосування.

У статті проведено порівняльний аналіз сучасних алгоритмів комп'ютерного зору для моніторингу повітряного простору, запропоновано класифікацію задач таких систем та

узагальнено основні технічні виклики й тенденції розвитку глибоких архітектур.

### **Мета**

Метою статті є систематизація та порівняльний аналіз сучасних методів комп'ютерного зору для розв'язання науково-технічної проблеми моніторингу повітряного простору в умовах високої невизначеності. Дослідження спрямоване на визначення ефективності алгоритмів глибокого навчання при виявленні, супроводі та ідентифікації малорозмірних об'єктів на складному фоні.

Класифікація задач комп'ютерного зору в аерокосмічному моніторингу

Процес моніторингу повітряного простору з застосуванням методів комп'ютерного зору не є монолітною алгоритмічною задачею. Останні публікації можуть показати, що це завдання можна розбити на чотири пов'язані задачі зі своїми специфіками [3]. Хоча в сучасних системах ці задачі часто виконуються паралельно, їхній поділ все ж залишається актуальним для розуміння архітектури систем безпеки.

### **Виявлення об'єктів**

Виявлення об'єктів є базовою задачею систем моніторингу повітряного простору. Її мета полягає у локалізації потенційних цілей на фоні неба або складного рельєфу. Положення об'єкта зазвичай описується

прямокутником Axis-Aligned Bounding Box (AABB). У такому вигляді задачу можна привести до вигляду:  $B = \{(X_{min}, Y_{min}, X_{max}, Y_{max}, C)\}$ , де  $C$  - клас об'єкта, а координати - описують його межі.

Для розв'язання цієї задачі використовують дві основні групи методів: регіонно-орієнтовані підходи родини R-CNN та одностадійні детектори. Перші забезпечують високу точність, але мають значну обчислювальну складність [4]. Одностадійні архітектури, такі як YOLO, виконують локалізацію та класифікацію за один прохід мережі, що дозволяє працювати в режимі реального часу [5].

#### Класифікація та ідентифікація

Після виявлення об'єкта система повинна визначити його тип і рівень загрози. Це ускладнюється високою міжкласовою подібністю, оскільки на великих відстанях птахи та дрони можуть виглядати однаково. Для вирішення цієї задачі зазвичай використовують багаторівневу ієрархічну таксономію.

Першим етапом є бінарна класифікація — відокремлення об'єктів від фонових аномалій або поділ на великі групи. Це особливо важливо для систем орнітологічної безпеки аеропортів, де необхідно фільтрувати хибні цілі та зберігати ресурси [6].

Після цього виконується класифікація типу апарату: БПЛА, літак, гелікоптер тощо. Однак зі збільшенням кількості класів ускладнюється визначення меж між ними, що знижує швидкість та точність системи [7].

Найскладнішим етапом є тонка ідентифікація, спрямована на визначення конкретної моделі об'єкта. Вона використовується для виявлення потенційно небезпечної або нелегальної діяльності. Основні труднощі виникають під час ідентифікації малих БПЛА через їхню схожість з птахами, складний фон та малі розміри цілей [8].

#### Локалізація та супровід

Після того як об'єкт було виявлено та класифіковано, виникає необхідність відслідковування траєкторії його руху. Цей процес полягає в пов'язуванні просторово розрізнених визначень між послідовними кадрами відеопотоку для побудови безперервної траєкторії руху та, інколи, присвоєння об'єкту унікального ідентифікатора.

В залежності від типу системи та кількості цілей цю задачу можна поділити на два напрями:

- Супровід однієї цілі (SOT): система захоплює один об'єкт на першому кадрі і далі самостійно утримує його в центрі кадру, передаючи координати на поворотні механізми.

- Багатооб'єктний супровід (MOT): набагато складніше завдання, яке виникає при виявленні декількох об'єктів одночасно. Тут застосовується парадигма Tracking-by-Detection або Joint Detection and Tracking. Алгоритми MOT повинні вирішувати задачу асоціації даних: математично визначати, яка рамка на поточному кадрі відповідає якому об'єкту з попередніх кадрів [9].

Оцінка просторового положення та намірів

Оцінка просторового положення та поведінки об'єкта є найбільш складною задачею моніторингу. Вона передбачає аналіз траєкторії руху, орієнтації та прогнозування майбутньої поведінки цілі. Для цього використовуються фільтри Калмана, марковські моделі та рекурентні нейронні мережі [10].

Виклики та обмеження систем комп'ютерного зору

Незважаючи на значний прогрес у галузі штучного інтелекту, створення повністю автономних систем моніторингу повітряного простору залишається складним завданням. Основними проблемами є малі розміри цілей, висока швидкість руху, складні погодні умови, нестача якісних датасетів та обмеження апаратного забезпечення.

Особливо складним є виявлення малорозмірних об'єктів, які можуть займати менше  $10 \times 10$  пікселів на зображенні [11]. Додаткові труднощі створюють динамічний фон, сонячні відблиски та оклюзії.

Використання синтетичних даних та генеративних моделей частково вирішує проблему нестачі навчальних вибірок, однак моделі, натреновані у віртуальному середовищі, часто демонструють падіння точності в реальних умовах [12]. Окрему загрозу становлять адверсаріальні атаки, здатні навмисно вводити нейромережі в оману шляхом нанесення спеціальних патернів на корпус БПЛА [13].

Аналіз методів детекції літаючих об'єктів

Історично еволюція методів виявлення пройшла шлях від простої піксельної математики до складних архітектур глибокого навчання.

Традиційні методи комп'ютерного зору

До широкого поширення глибокого навчання основними методами детекції були алгоритми віднімання фону та аналізу оптичного потоку.

Методи віднімання фону моделюють статистичний розподіл пікселів і дозволяють виділяти рухомі об'єкти, однак мають низьку ефективність у випадку динамічного фону [14]. Алгоритми оптичного потоку аналізують зміщення пікселів між кадрами та забезпечують кращу сегментацію рухомих цілей, проте характеризуються високою обчислювальною складністю.

Згорткові нейронні мережі (CNN)

Де-факто стандарт галузі у задачах візуального виявлення об'єктів, здатні автоматично вилучати ієрархічні ознаки. Подібні системи можна поділити на дві архітектурні гілки: Одно- та двостадійні детектори.

Одностадійні детектори створювалися з однією метою: швидкість. Не маючи етапу генерації регіонів, вони одразу регресують координати рамок та ймовірності класів за один прохід нейромережі. Сімейство алгоритмів YOLO (You Only Look Once) (Рис. 1) стало промисловим стандартом завдяки простоті та потужності використаних екстракторів ознак, наприклад CSPDarknet.

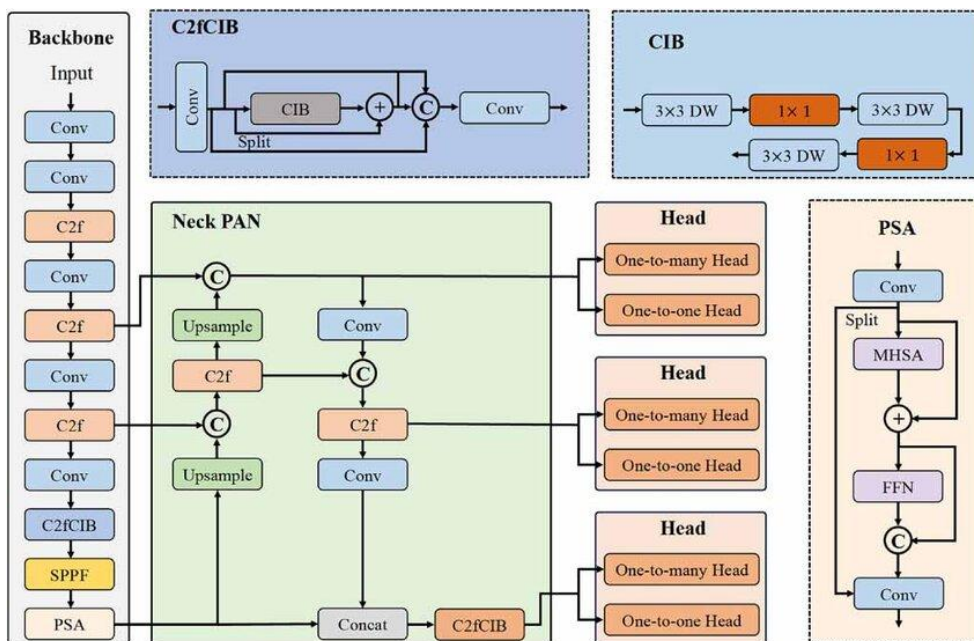


Рис. 1 - Архітектура YOLOv10

Проте в останніх поколіннях (YOLOv8 та YOLOv10) почали використовувати так звану без'якірну архітектуру. Замість того, щоб підганяти рамку дрона під заздалегідь визначені пропорції, мережа прогнозує безпосередньо центр об'єкта та відстань до його країв. А для боротьби з дисбалансом класів, коли порожнє небо займає більше 90% кадру, застосовується спеціальна функція втрат, яка зменшує вагу пікселів фону, фактично змушуючи мережу фокусуватися на пікселях об'єкту [15, 16]. Математично цей процес описується за допомогою фокусувальної функції втрат (Focal Loss), яка модифікує стандартну крос-ентропію. Для бінарної класифікації вона має вигляд:

$$FL(p_t) = -a_t(1 - p_t)^\gamma \log(p_t) \quad (2)$$

де  $p_t$  - ймовірність, передбачена моделлю для правильного класу;  $a_t$  - ваговий коефіцієнт, що балансує

співвідношення між класами (наприклад, об'єкт/фон);  $\gamma \geq 0$  - фокусувальний параметр, який плавно зменшує штраф для "легких" прикладів (наприклад, пікселів чистого неба), концентруючи увагу градієнтного спуску на "складних" прикладах (малорозмірних цілях).

Двостадійні детектори ж діють послідовно. Такі архітектури як Faster R-CNN та Cascade R-CNN (Рис. 2) використовують підмережу для сканування карти ознак та генерації тисяч гіпотез - місць, де є ймовірність перебування об'єкта. Після чого мережа стандартизує ці ділянки і повнозв'язні шари мережі виконують точну класифікацію. Такий підхід має вищу точність і кращі показники при роботі зі скупченнями об'єктів. Проте саме через послідовну структуру і складність алгоритму вони працюють повільніше, що робить їх менш привабливими для роботи в обмежених умовах [17].

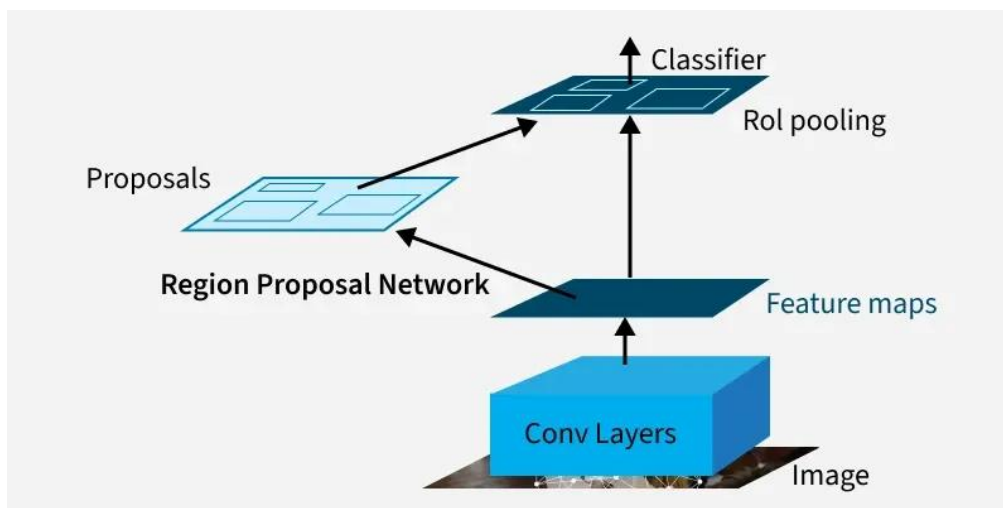


Рис. 2 - Приклад архітектури Faster R-CNN

Візуальні трансформери (ViT)  
Впровадження архітектур типу ViT здійснило концептуальну революцію в області комп'ютерного зору. Рішення на зразок DETR (DEtection TRansformer), Deformable DETR або RT-DETR

розглядають детекцію не як задачу попиксельної регресії, а як задачу прогнозування фіксованої множини об'єктів за допомогою алгоритму двочасткового зіставлення (Рис. 3) [18].

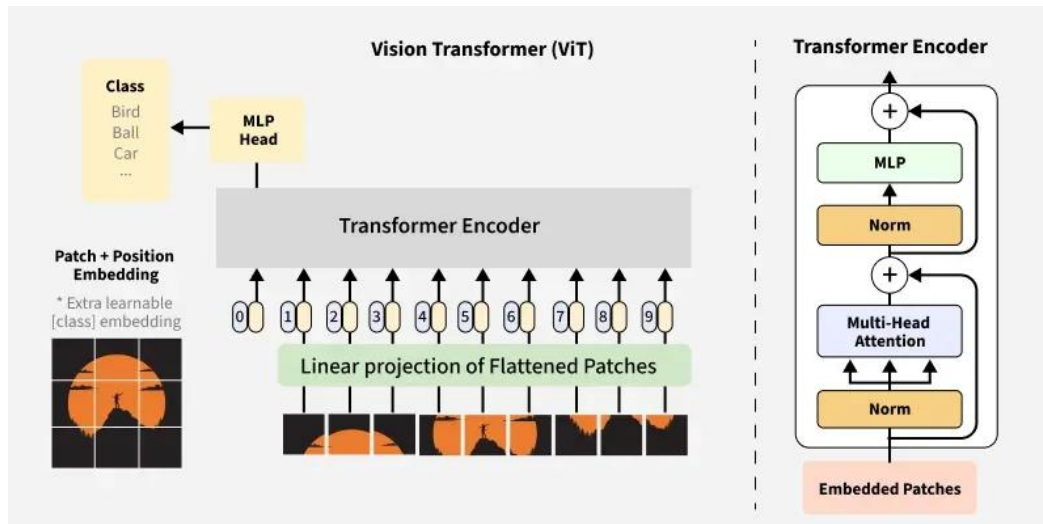


Рис. 3 - Приклад класичної архітектури ViT

Основа будь-якого візуального трансформеру - механізм самоуваги. Трансформер розбиває зображення на частини, лінійно їх проектує у вектори і для кожного вектора створює матриці Запиту, Ключа та Значення. Перемноження запиту та ключа дозволяє мережі математично визначити рівень важливості (уваги) кожної частини зображення по відношенню до всіх інших частин одночасно.

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V \quad (3)$$

$d_K$  - розмірність векторів ключів. Ділення на  $\sqrt{d_K}$  (масштабування) є критично важливим кроком, оскільки при великих розмірностях скалярні добутки можуть ставати занадто великими, що призводить до проблеми зникаючого градієнта у функції *softmax*. *softmax* - функція, яка перетворює отримані "сирі" оцінки зв'язку у ймовірності (від 0 до 1), сума яких дорівнює одиниці, що і створює "карту уваги".

На відміну від CNN із локальним рецептивним полем, трансформери аналізують глобальний контекст, що підвищує ефективність в умовах обмеженої видимості [19]. Проте їхнім головним недоліком є квадратичне зростання складності механізму самоуваги відносно розміру зображення,

через що обробка відео високої роздільної здатності стає надто ресурсоємною. Для часткового вирішення цієї проблеми застосовують архітектури Swin Transformer та Deformable DETR, однак вони або знижують ефективність, або створюють нові обмеження [20].

Шляхи вирішення проблеми виявлення малих об'єктів

Навіть сучасні версії YOLO втрачають ефективність під час виявлення малих цілей розміром менше  $10 \times 10$  пікселів. Для вирішення цієї проблеми застосовують спеціалізовані підходи [21]:

- Feature Pyramid Networks (FPN, BiFPN, PANet) — архітектури, що поєднують високорівневі семантичні ознаки з деталізованими картами ознак ранніх шарів для кращого виявлення дрібних об'єктів [22].

- Надшвидкісна генерація ознак — використання GAN або SRCNN для штучного збільшення фрагментів зображення з підозрілими об'єктами перед класифікацією [23].

- SAHI (Slicing Aided Hyper Inference) — метод розбиття зображень високої роздільної здатності на менші перекривні фрагменти (Рис. 4), що дозволяє ефективніше виявляти субпіксельні об'єкти та об'єднувати результати в єдину сцену [24].

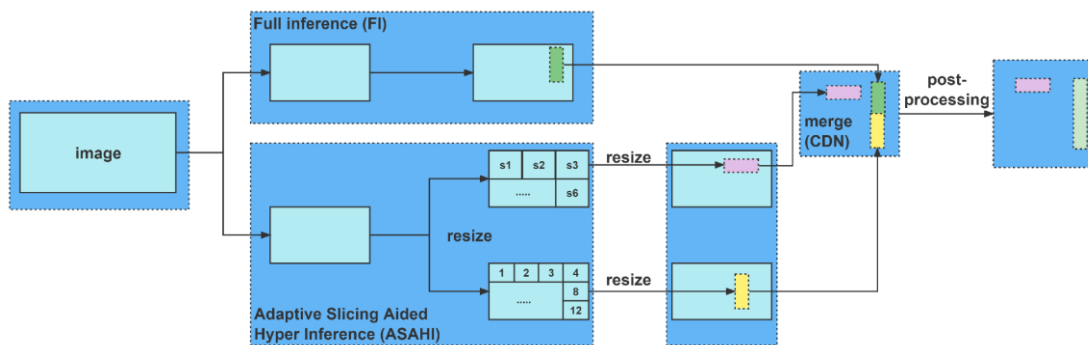


Рис. 4 - Приклад архітектури SAHI

Аналіз методів ідентифікації та класифікації

Хибна ідентифікація об'єкту може дорого обійтись: від втрати ресурсів, чого можна було уникнути до більш катастрофічних наслідків. Саме тому правильна ідентифікація та класифікація є дуже важливою складовою завдання моніторингу.

Від дескрипторів до глибоких ознак

У еру “класичного” комп'ютерного зору класифікація покладалася на ручний вибір ознак. Дескриптори гістограми напрямлених градієнтів фіксували контури, а масштабоінваріантне ознакове перетворення - локальні ключові точки. Отримані вектори оброблялися класифікаторами опорних векторів або “випадкового лісу”. Ці підходи назавжди відійшли у минуле через нездатність працювати при зміні ракурсу та низькому контрасті [25].

Сьогодні монополію утримують архітектури класифікації зображень:

ResNet, DenseNet, EfficientNet. Вони автоматично вивчають ієрархію ознак: від базових ліній та кутів на перших парах до складних семантичних патернів, такі як пропелери, форма крила, шасі, на глибоких рівнях [26].

Edge AI та метричне навчання

Розгортання систем на борту мобільних (або просто малопотужних) систем накладає свої обмеження: за розмірами, доступними обчислювальними потужностями, енергоефективністю і так далі. Для обходу таких обмежень, зазвичай використовуються сімейства оптимізованих мереж: MobileNet (v2, v3) та ShuffleNet (Рис. 5). Їхній секрет - використання глибоких відокремлених згорток, що розбивають стандартну 3D-згортку на дві простіші операції, зменшуючи кількість математичних множень та розмір моделі в десятки разів при падінні точності лише, приблизно, на 2-4% [27].

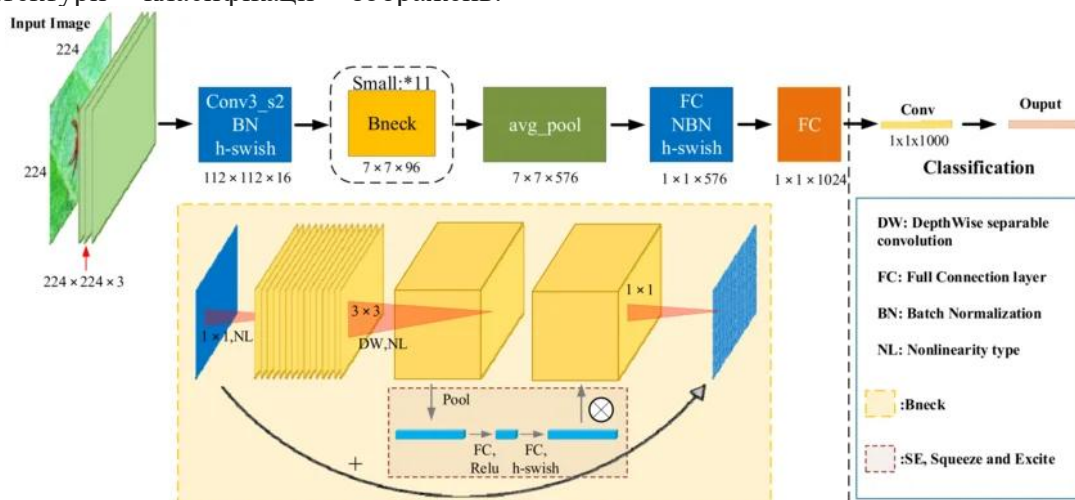


Рис. 5 - Архітектура MobileNet v3

Виникає інша проблема: подібні системи з функцією Softmax не можуть ідентифікувати абсолютно нові, раніше не бачені типи об'єктів, тобто об'єкти про які мало інформації, або її взагалі немає. Цю проблему можна вирішити, використовуючи Сіамські мережі з функціями втрат Contrastive Loss або Triplet Loss. Такі мережі навчаються не передбачати клас, а генерувати числовий зліпок.

Для функції Triplet Loss оптимізація спрямована на мінімізацію відстані між опорним зображенням (Anchor,  $a$ ) та позитивним прикладом (Positive,  $p$  - той самий клас) і максимізацію відстані до негативного прикладу (Negative,  $n$  - інший клас). Функція втрат формалізується як:

$$L = \max(d(f(a), f(p)) - d(f(a), f(n)) + m, 0) \quad (4)$$

де  $f(\cdot)$  - функція нейромережі, що генерує вектор ознак,  $d(\cdot, \cdot)$  - обрана метрика відстані (наприклад, евклідова), а  $m$  (margin) - заданий проміжок, який гарантує надійне розділення класів у просторі ознак. Ідентифікація відбувається шляхом вимірювання косинусної відстані між зліпком невідомого об'єкта в небі та еталонними зразками в базах даних [28].

Кінематика: Проблема "Bird vs. Drone"

Проблема "Птах чи Дрон" є класичною у моніторингу повітряного простору. На великих відстані (понад 3 км) об'єкт є безструктурним кластером пікселів, а просторовий аналіз стає нерелевантним [29].

Вирішення лежить у площині просторово-часового аналізу за допомогою рекурентних мереж або оптичного потоку. Біологічні об'єкти, такі як птахи, мають періодичні зміни форми кластера та рухаються за плавними кривими. Натомість дрони або літаки мають жорсткий корпус і здатні до неприродних для птахів маневрів: лінійні рухи, миттєві зміни вектора, швидкий набір висоти або ідеально стабільне зависання в одній точці

простору. Аналіз мікро-змін векторів швидкості дозволяє досягти точності понад 98% у визначенні цих класів [30].

Аналіз методів локалізації та супроводу

Супровід - найбільш вимоглива частина процесу моніторингу, якщо говорити про час. Часу потрібно багато, адже поки БПЛА або птах в об'єктиві, його потрібно відслідковувати. До того ж, сучасні алгоритми повинні боротися з екстремальною зміною масштабу об'єкта, оклюзіями та деформаціями візуального вигляду об'єктів під час руху [31].

Математичні фільтри та кореляційний аналіз

Традиційною основою методів супроводу є прогнозуючі моделі. Вони використовують фільтр Калмана для оновлення стану об'єкта у два етапи: екстраполяція, або передбачення, та корекція на основі нових вимірювань.

Формально стан об'єкта в момент часу  $k$  описується вектором  $x_k$  (наприклад, координати та швидкість). Етап екстраполяції (передбачення) задається рівняннями:

$$\hat{x}_{kvk-1} = F_k \hat{x}_{k-1vk-1} \quad (5)$$

$$P_{kvk-1} = F_k P_{k-1vk-1} F_k^T + Q_k \quad (6)$$

де  $F_k$  - матриця переходу станів,  $P$  - матриця коваріації помилки,  $Q_k$  - коваріація шуму процесу. Етап корекції при отриманні нового вимірювання  $z_k$  від детектора виглядає так:

$$K_k = P_{kvk-1} H_k^T (H_k P_{kvk-1} H_k^T + R_k)^{-1} \quad (7)$$

$$\hat{x}_{kvk} = \hat{x}_{kvk-1} + K_k (z_k - H_k \hat{x}_{kvk-1}) \quad (8)$$

де  $K_k$  - коефіцієнт підсилення Калмана,  $H_k$  - матриця спостережень, а  $R_k$  - коваріація шуму вимірювань. Цей метод ідеальний для лінійних траєкторій, але дає катастрофічні збої при різких рухах маневрених об'єктів через надмірне зростання шуму процесу. Для нелінійних рухів можна використовувати Фільтр часток, що генерує велику кількість ймовірнісних гіпотез, проте вони споживають багато обчислювальних ресурсів [32].

В цьому плані високу ефективність продемонстрували кореляційні фільтри (KCF, CSR-DCF, ECO). Алгоритм навчає дискримінантний класифікатор на першому кадрі. Завдяки перетворенню Фур'є та використанню матриць циркулянтів, складні згортки замінюються на поелементне множення у частотній області. Це забезпечує рекордну частоту роботи без використання процесорів відеокарт. Проте кореляційні треки страждають від "дрейфу": при частковому перекритті об'єкта, наприклад хмарою, фільтр починає вивчати хмару як частину об'єкта і безповоротно втрачає ціль [33].

Багатооб'єктний супровід (MOT) та алгоритм ByteTrack

Але що якщо потрібно відслідкувати не один, а декілька десятків об'єктів? Саме це питання вивело на перший план технології MOT. Парадигма "Tracking-by-Detection" розбиває процес на кілька етапів: спочатку працює YOLO, потім алгоритм асоціації даних зв'язує детекції у траєкторії. Класичний алгоритм DeepSORT використовує Угорський алгоритм оптимізації, комбінуючи відстань Махаланобіса (оцінку Калмана) та косинусну відстань ембедінгів зовнішнього вигляду (Re-Identification).

Математично задача MOT на цьому етапі зводиться до задачі двочасткового зіставлення, де метою є мінімізація загальної вартості призначень між наявними треками (існуючими об'єктами) та новими виявленнями

(детекціями). Матриця вартості  $C = [c_{i,j}]$  розраховується як зважена сума кінематичної та візуальної метрик:

$$c_{i,j} = \lambda d^{(1)}(i,j) + (1 - \lambda) d^{(2)}(i,j) \quad (9)$$

де  $\lambda$  - гіперпараметр балансування. Кінематична відстань  $d^{(1)}(i,j)$  обчислюється як відстань Махаланобіса між прогнозованим станом треку  $y_i$  та виявленням  $d_j$  з урахуванням коваріації  $S_i$ :

$$d^{(1)}(i,j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i) \quad (10)$$

А візуальна відстань  $d^{(2)}(i,j)$  є косинусною відстанню між векторами ознак, згенерованими нейромережею.

Однак при перетині траєкторій об'єктів часто виникає плутанина ідентифікаторів. [34].

Фундаментальний прорив здійснив алгоритм ByteTrack (Рис. 6). Раніше всі детекції з низькою впевненістю просто відкидали результати для уникнення хибних спрацювань. Але малий об'єкт на великій дистанції завжди має низький score, і його відкидання призводило до "розірваних" треків. ByteTrack застосовує дворівневу асоціацію: спочатку він намагається зв'язати надійні треки з високоточними виявленнями. Потім, ті треки, які "загубилися", він асоціює з виявленнями низької впевненості. Завдяки цьому система здатна безперервно вести ледь помітну ціль розміром з піксель в умовах інтенсивних візуальних перешкод [35].

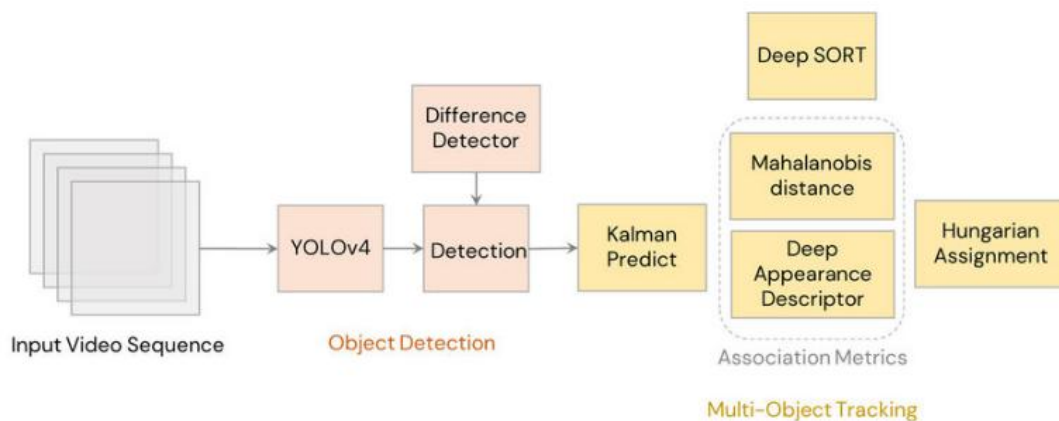


Рис. 6 - Архітектура ByteTrack

Порівняльний аналіз ефективності методів комп'ютерного зору

Вибір оптимального алгоритму для задач моніторингу повітряного простору вимагає пошуку компромісу між точністю виявлення, обчислювальною швидкістю, здатністю працювати в режимі реального часу та ефективністю локалізації малорозмірних об'єктів. Проведений аналіз дозволяє виділити кілька ключових тенденцій щодо ефективності розглянутих підходів.

Традиційні методи комп'ютерного зору, такі як віднімання фону та математична морфологія, вирізняються найменшими вимогами до апаратного забезпечення та високою швидкістю. Однак їхня точність залишається низькою через повну нездатність адаптуватися до складного або динамічного фону (рух хмар, хитання дерев). Використання щільного оптичного потоку частково вирішує проблему складного фону завдяки сегментації за кінематикою, проте цей підхід має надто високу обчислювальну складність, що унеможлиблює його застосування для відео високої роздільної здатності в реальному часі.

Одностадійні детектори на базі згорткових нейромереж (такі як родина YOLO, SSD) на сьогодні є індустріальним стандартом для систем спостереження. Вони забезпечують найкращий баланс між високою точністю та надзвичайною швидкістю, що робить їх ідеальними для роботи в реальному часі. Їхнім головним недоліком є посередня здатність до виявлення мікроцілей, що вимагає обов'язкового застосування додаткових архітектурних надбудов (наприклад, FPN або методів розрізання зображень SAH).

На противагу їм, двостадійні детектори (такі як Faster R-CNN, Cascade R-CNN) демонструють дуже високу точність та чудові результати при роботі з малими об'єктами або в умовах їх скупчення. Проте послідовна архітектура генерації гіпотез суттєво знижує частоту їх роботи. Їх застосування в реальному часі є суттєво обмеженим і можливе лише на потужних стаціонарних

обчислювальних станціях, що ускладнює їх розгортання в польових умовах.

Візуальні трансформери (ViT, DETR) демонструють найвищу точність серед усіх розглянутих методів завдяки феноменальному розумінню глобального контексту зображення та високій стійкості до перекриттів цілі. Водночас їхня головна проблема полягає у квадратичному зростанні обчислювальної складності залежно від розміру кадру. Це робить їх найповільнішими серед сучасних алгоритмів і на даному етапі розвитку практично непридатними для потокової обробки відео високої роздільної здатності в реальному часі.

Для мобільних платформ та систем безпосередньо на борту БПЛА найдоцільнішим є використання спеціалізованих архітектур (MobileNet, ShuffleNet). Вони забезпечують високу швидкість обробки при жорстких обмеженнях енергоспоживання та обсягу пам'яті. Компромісом у цьому випадку є зниження загальної точності на 2-4% порівняно з базовими моделями та обмежена здатність ідентифікувати малорозмірні об'єкти на великих дистанціях.

### **Висновок**

Стрімкий розвиток БПЛА зумовив гостру потребу в нових методах візуального контролю, оскільки традиційні радары малоефективні для малорозмірних цілей. У цій частині дослідження доведено, що перехід до згорткових нейромереж (YOLO) та візуальних трансформерів забезпечив значний стрибок у швидкості детекції та розумінні глобального контексту.

Подолання проблеми мікророзмірних цілей вимагає застосування архітектурних надбудов (SAH, FPN). Розрізнення птахів і дронів на великих відстанях найбільш ефективно вирішується алгоритмічно, шляхом просторово-часового аналізу кінематики. Встановлено, що еволюція алгоритмів локалізації завершилася багатоб'єктними трекерами (ByteTrack), здатними утримувати ціль навіть за умов

екстремально низької впевненості виявлення.

Однак будь-які програмні алгоритми залишаються залежними від фізичних обмежень самих оптичних сенсорів. Друга частина цього дослідження буде присвячена апаратно-системним рішенням: мультимодальному злиттю даних, нейроморфним камерам подій та розподіленим Edge-Cloud обчисленням для створення всепогодних систем моніторингу повітряного простору.

### Література

1. Taha, B., & Shoufan, A. Machine learning-based drone detection and classification: State-of-the-art in research. IEEE Access, 2019, Vol. 7., p. 138669 - 138682. DOI: 10.1109/ACCESS.2019.2942944
2. Coluccia, A., Fascista, A., Schumann, A., et al. (2021). Drone vs. bird detection: Deep learning algorithms and results from a grand challenge. Sensors, 21(8).
3. Smyers E., Katz S., Corso A., Kochenderfer M. AVOIDDS: Aircraft Vision-based Intruder Detection Dataset and Simulator. URL: <https://arxiv.org/abs/2306.11203>
4. Girshick R., Donahue J., Darrell T., Malik R. feature hierarchies for accurate object detection and semantic segmentation. URL: <https://arxiv.org/abs/1311.2524>
5. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, DOI: 10.1109/CVPR.2016.91
6. A. Coluccia et al., "Drone-vs-Bird detection challenge at IEEE AVSS2017," 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, 2017, pp. 1-6, doi: 10.1109/AVSS.2017.8078464
7. Thrampoulidis C., Oymak S., Soltanolkotabi M. Theoretical Insights Into Multiclass Classification: A High-dimensional Asymptotic View. URL: <https://arxiv.org/abs/2011.07729>
8. M. Nalamati, A. Kapoor, M. Saqib, N. Sharma and M. Blumenstein, "Drone Detection in Long-Range Surveillance Videos," 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Taipei, Taiwan, 2019, pp. 1-6, doi: 10.1109/AVSS.2019.8909830
9. Zhu P., Wen L., Du D., Bian X., Fan H., Hu Q., Ling H. Detection and Tracking Meet Drones Challenge. URL: <https://arxiv.org/abs/2001.06303>
10. Khanpour A., Wang T., Vahidi-Shams A., Ectors W., Nakhaie F., Taheri A., Claudel C. UAV-Based Intelligent Traffic Surveillance System: Real-Time Vehicle Detection, Classification, Tracking, and Behavioral Analysis. URL: <https://arxiv.org/abs/2509.04624>
11. Kaur D., Battish N., Bhavsar A., Poddar S. YOLOBirDrone: Dataset for Bird vs Drone Detection and Classification and a YOLO based enhanced learning architecture. URL: <https://arxiv.org/abs/2601.08319>
12. C. Aker and S. Kalkan, "Using deep networks for drone detection," 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, 2017, pp. 1-6, doi: 10.1109/AVSS.2017.8078539.
13. Xi H, Ru L, Tian J, Wang W, Zhu R, Li S, Zhang Z, Liu L, Luan X. Towards Robust Physical Adversarial Attacks on UAV Object Detection: A Multi-Dimensional Feature Optimization Approach. Machines. 2025; 13(11):1060. <https://doi.org/10.3390/machines13111060>
14. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.
15. T. -Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 318-327, 1 Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.
16. Varghese, R., & M., S. (2024). YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and

Robustness. In 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS) (pp. 1-6).

17. 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, Vol. 28*. Curran Associates, Inc. 2015.

18. Yang M. Visual Transformer for Object Detection. URL: <https://arxiv.org/abs/2206.06323>

19. Shah, S., Tembhurne, J. Object detection using convolutional neural networks and transformer-based models: a review. *Journal of Electrical Systems and Inf Technol* 10, 54 (2023). <https://doi.org/10.1186/s43067-023-00123-z>

20. M. Ali A., Benjdira B., Koubaa A., El-Shafai W., Khan Z., Boulila W. Vision Transformers in Image Restoration: A Survey. [https://www.researchgate.net/publication/368691069\\_Vision\\_Transformers\\_in\\_Image\\_Restoration\\_A\\_Survey](https://www.researchgate.net/publication/368691069_Vision_Transformers_in_Image_Restoration_A_Survey)

21. Brais Bosquet, Manuel Mucientes, & Victor M. Brea (2021). STDnet-ST: Spatio-temporal ConvNet for small object detection. *Pattern Recognition*, 116, 107929.

22. T. -Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 936-944, doi: 10.1109/CVPR.2017.106.

23. Xiao, Y., Li, Z., & Zhang, Z. (2024). Enhancing Image Perception Quality: Exploring Loss Function Variants in SRCNN. In Proceedings of the 2024 7th International Conference on Computer Information Science and Artificial Intelligence (pp. 352–356). Association for Computing Machinery.

24. Zhang, H., Hao, C., Song, W., Jiang, B., & Li, B. (2023). Adaptive Slicing-Aided Hyper Inference for Small Object Detection in High-Resolution Remote Sensing Images. *Remote Sensing*, 15(5).

25. Nassi, B., Mirsky, Y., Nassi, D., Ben-Netanel, R., Drokin, O., & Elovici, Y.

(2020). Phantom of the ADAS: Securing Advanced Driver-Assistance Systems from Split-Second Phantom Attacks. In Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (pp. 293–308). Association for Computing Machinery.

26. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning (pp. 6105–6114). PMLR.

27. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.

28. Koch G., Zemel R. Siamese Neural Networks for One-shot Image Recognition. URL: <https://www.cs.cmu.edu/~rsalakhu/papers/one-shot1.pdf>

29. Nesteruk, A., Nikitin, V., Albrecht, Y., Ścisło, Ł., Grela, D., & Król, P. (2026). Distinguishing a Drone from Birds Based on Trajectory Movement and Deep Learning. *Sensors*, 26(3).

30. Khan, S., Naseer, M., Hayat, M., Zamir, S., Khan, F., & Shah, M. (2022). Transformers in Vision: A Survey. *ACM Comput. Surv.*, 54(10s).

31. Y. Wu, J. Lim and M. -H. Yang, "Online Object Tracking: A Benchmark," 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 2013, pp. 2411-2418, doi: 10.1109/CVPR.2013.312.

32. M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," in *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174-188, Feb. 2002, doi: 10.1109/78.978374.

33. Henriques, J., Caseiro, R., Martins, P., & Batista, J. (2015). High-Speed Tracking with Kernelized Correlation Filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(3), 583–596.

34. N. Wojke, A. Bewley and D. Paulus,

"Simple online and realtime tracking with a deep association metric," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 2017, pp. 3645-3649, doi: 10.1109/ICIP.2017.8296962.

35. Zhang, Y., Sun, P., Jiang, Y., Yu, D.,

Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object Tracking by Associating Every Detection Box. In Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXII (pp. 1–21). Springer-Verlag.

**Троцький Я. В., Зіатдінов Ю. К.**

### **АНАЛІЗ СУЧАСНИХ МЕТОДІВ КОМП'ЮТЕРНОГО ЗОРУ ДЛЯ МОНІТОРИНГУ ТА ІДЕНТИФІКАЦІЇ ЛІТАЮЧИХ ОБ'ЄКТІВ: ВИКЛИКИ ТА ПЕРСПЕКТИВИ**

*У оглядовій статті проведено комплексний аналіз сучасних програмно-алгоритмічних методів комп'ютерного зору, що застосовуються для моніторингу та ідентифікації літаючих об'єктів у повітряному просторі. Актуальність дослідження зумовлена стрімким розвитком безпілотних літальних апаратів та зростанням потреби в ефективних, доступних і високоточних системах спостереження, здатних працювати в умовах невизначеності, обмежених ресурсів і складного фону. Розглянуто класифікацію основних задач комп'ютерного зору, зокрема виявлення об'єктів, класифікацію та ідентифікацію, локалізацію і супровід, а також оцінку просторового положення та поведінкових характеристик цілей. Проаналізовано еволюцію методів детекції від традиційних підходів до сучасних архітектур глибокого навчання, включаючи згорткові нейронні мережі, одностадійні та двостадійні детектори, візуальні трансформери та спеціалізовані підходи для виявлення малорозмірних об'єктів. Особливу увагу приділено проблемі розрізнення птахів і дронів, методам багатоб'єктного супроводу, алгоритмам асоціації даних, а також стійкості систем до адверсаріальних атак. Здійснено порівняльний аналіз ефективності методів комп'ютерного зору.*

**Ключові слова:** Комп'ютерний зір, моніторинг повітряного простору, нейронні мережі, глибинне навчання, ідентифікація літаючих об'єктів.

**Trotskyi Y., Ziatdinov Y.**

### **ANALYSIS OF MODERN COMPUTER VISION METHODS FOR MONITORING AND IDENTIFYING FLYING OBJECTS: CHALLENGES AND PROSPECTS**

*This review article presents a comprehensive analysis of modern software and algorithmic computer vision methods used for monitoring and identifying flying objects in airspace. The relevance of the study is driven by the rapid development of unmanned aerial vehicles and the growing need for efficient, affordable, and highly accurate surveillance systems capable of operating under uncertainty, limited resources, and complex backgrounds. The classification of the main computer vision tasks is considered, including object detection, classification and identification, localization and tracking, as well as estimation of the spatial position and behavioral characteristics of targets. The evolution of detection methods from traditional approaches to modern deep learning architectures is analyzed, including convolutional neural networks, one-stage and two-stage detectors, visual transformers, and specialized approaches for detecting small-sized objects. Particular attention is paid to the problem of distinguishing birds from drones, multi-object tracking methods, data association algorithms, and the robustness of systems against adversarial attacks. A comparative analysis of the effectiveness of computer vision methods is also carried out.*

**Keywords:** Computer vision, airspace monitoring, neural networks, deep learning, flying object identification.

Стаття подана до редакції: 23/03/2026

Стаття прийнята до опублікування: 07/04/2026

Стаття опублікована: 30/05/2026

Стаття поширюється на умовах ліцензії CC BY 4.0