

МЕТОДИ І СИСТЕМИ ОПТИЧНО-ЕЛЕКТРОННОЇ ТА ЦИФРОВОЇ ОБРОБКИ СИГНАЛІВ

DOI: 10.20535/1970.68(2).2024.318080

УДК 004.93

АНАЛІЗ ІСНУЮЧИХ МОДЕЛЕЙ ВИЯВЛЕННЯ БПЛА НА ТЕПЛОВІЗІЙНИХ ЗОБРАЖЕННЯХ

*Танчук В. С., Колобродов В. Г.**Національний технічний університет України**«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна**E-mail: tanchuk.vladislav@gmail.com, thermo@ukr.net*

У наші часи широкого використання набувають безпілотні літальні апарати (БПЛА). Вони підтвердили свою ефективність, надійність та доцільність. Однак неналежне використання цієї технології може призвести до суттєвих порушень прав людини та створити значні загрози громадській безпеці. У зв'язку з чим повинні створюватися відповідні методи протидії БПЛА. Одним із напрямків, який активно розвивається в наш час, є виявлення БПЛА на основі оптико-електронного випромінювання. Оскільки більшість таких систем розраховані на видимий діапазон, інфрачервоний (ІЧ) діапазон не так широко був досліджений. Виявлення БПЛА на основі тепловізійних зображень може бути виконано, використовуючи штучний інтелект (ШІ). Наразі вже існують готові підходи та моделі, які виконують задачі виявлення об'єктів, проте вони мають більш загальне призначення.

Для огляду проблеми виявлення БПЛА на основі тепловізійних зображень та результатів ШІ, необхідно провести аналіз найбільш ефективних моделей виявлення об'єктів для вирішення задачі визначення БПЛА у ІЧ діапазоні. У цій роботі було проведено аналіз та порівняння таких моделей виявлення об'єктів, як YOLOv5 та YOLOv8, Faster RCNN, DETR.

Для навчання моделей використовувались датасети, як в ІЧ діапазоні, так і у видимому. Було проведено попередню цифрову обробку датасету зображень БПЛА у видимому діапазоні, для перетворення їх у псевдо теплові. Це виконувалось з метою збільшення кількості тренувальних даних для моделей і тим самим покращення їх точності.

Результати показали, що такі моделі, як YOLOv8 та DETR, є найефективнішими для задач виявлення БПЛА на тепловізійних зображеннях, проте їх точність і досі є недостатньою для ефективного використання в системах захисту від БПЛА у реальному часі.

Ключові слова: безпілотний літальний апарат, штучний інтелект, модель виявлення об'єктів, YOLOv8, YOLOv5, Faster RCNN, DETR, датасет, попередня обробка, середня точність виявлення, середнє значення повноти.

Вступ

Безпілотні літальні апарати (БПЛА) набули значного використання в реаліях нашого часу завдяки своїй здатності ефективно та надійно виконувати різноманітні завдання. Використання даної технології є економічно виправданим, оскільки вартість виготовлення БПЛА є відносно низькою порівняно зі складністю завдань, які вони можуть вирішувати. Проте неналежне їх використання може призвести до суттєвих порушень прав людини та створити значні загрози громадській безпеці. Тому на протидію БПЛА повинні існувати відповідні ефективні методи захисту від них.

Однією з ключових складових задач протидії БПЛА є автоматизація процесу виявлення та націлювання на БПЛА, яка повинна забезпечувати ефективну роботу за будь-яких умов: і вдень, і вночі, і в туман, і в дощ тощо. В такому випадку,

можна скористатися тепловізійними системами, через те, що вони мають змогу цілодобово виявляти об'єкт спостереження у складних умовах.

Для виконання задачі виявлення БПЛА на основі тепловізійного зображення можна скористатися системами штучного інтелекту (ШІ).

Наразі, активно досліджується виявлення БПЛА на основі тепловізійних зображень та ШІ, проте дані дослідження більше сконцентровані на видимому діапазоні. В свою чергу, інфрачервоний (ІЧ) діапазон потребує більш глибоких досліджень. Вже існують певні системи та моделі, які більш менш здатні виявляти БПЛА.

Тому, є необхідність в дослідженні наявних моделей виявлення БПЛА на основі ШІ, використовуючи тепловізійні зображення, для їх подальшого аналізу та огляду стану проблеми.

Огляд моделей виявлення об'єктів

YOLOv8 та YOLOv5. YOLO (You Only Look Once) – це серія моделей глибокого навчання, які використовуються для виявлення об'єктів у реальному часі. Серед різних моделей виявлення об'єктів фреймворк YOLO виділяється гарним балансом швидкості та точності, що дозволяє швидко та надійно ідентифікувати об'єкти на зображеннях. YOLOv8 та YOLOv5 є одними з найточніших моделей сімейства YOLO. YOLOv8 остання модель створена компанією Ultralytics, характеризується високою точністю та швидкістю. YOLOv5 старша, проте має більшу надійність та більш оптимізована під різні задачі. Загальна схема алгоритму цих моделей зображена на рис. 1.

Принцип роботи YOLOv5 починається з вхідного зображення, яке спочатку змінюється до фіксованого розміру (зазвичай 640x640 пікселів). Після цього зображення проходить через основну мережу

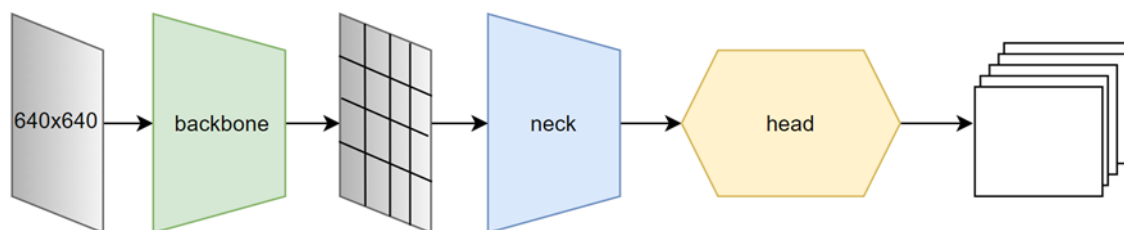


Рис. 1. Принцип роботи YOLOv5 та YOLOv8 моделей

YOLOv8, як і його попередник, використовує схожий принцип роботи, але з деякими покращеннями. Початок процесу також полягає у зміні розміру вхідного зображення до фіксованого розміру та проходження через основну мережу, яка є вдосконаленою версією CSPDarknet53. Відмінність YOLOv8 полягає в більш вдосконаленій архітектурі проміжної мережі (neck), яка може включати додаткові шари та оптимізації для покращення обробки характеристик. На виході модель також генерує передбачення координат рамок, ймовірностей об'єктів та їх класифікації. Однак, YOLOv8 може використовувати anchor-free підхід, що усуває необхідність у попередньо визначених анкерних рамках. Замість цього, модель безпосередньо передбачає центри, ширину та висоту рамок об'єктів. Це значно спрощує процес налаштування моделі [3].

Faster R-CNN. Faster R-CNN (Region-based Convolutional Neural Network) є однією з провідних моделей для виявлення об'єктів, яка поєднує в собі високу точність та ефективність. Faster R-CNN побудована на основі ідей попередніх моделей, таких як R-CNN [4] і Fast R-CNN [5]. Головна відмінність від попередніх моделей – це використання регіональної мережі пропозицій (Region Proposal Network, RPN), яка замінює зовнішні методи генерації регіонів-кандидатів, роблячи модель швидшою та більш інтегрованою.

(backbone), назва якої CSPDarknet53. Ця мережа відповідає за визначення характеристик зображення, перетворюючи його на множину ознак (features). Проміжна мережа (neck), яка використовує PANet (Path Aggregation Network), комбінює характеристики з різних рівнів мережі, що допомагає краще виявляти об'єкти різних розмірів. На виході (head) модель генерує передбачення для кожної комірки сітки зображення. Це передбачення включає координати рамок об'єктів, ймовірність наявності об'єктів та класифікацію об'єктів. Для обробки рамок об'єктів використовуються анкерні рамки (anchors), які представляють собою попередньо визначені розміри та співвідношення сторін рамок. Модель прогнозує зміщення від цих анкерів, щоб краще відповідати об'єктам на зображенні. Фінальний крок, застосування алгоритму Non-Maximum Suppression (NMS) [1] для усунення дублюючих рамок та вибору найбільш вірогідних передбачень [2].

Принцип роботи Faster R-CNN можна розподілити на кілька етапів (рис. 2) [6].

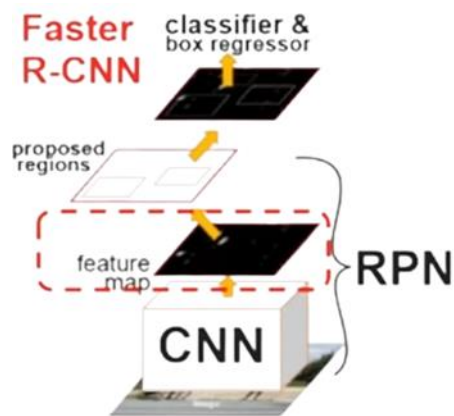


Рис. 2. Принцип роботи Faster R-CNN моделі

Вхідне зображення подається на вхід моделі, де спочатку проходить через базову згорткову мережу (CNN). Ця мережа витягує високорівневі характеристики зображення, перетворюючи його на багатовимірний тензор характеристик. RPN проходить по вихідному тензору характеристик та генерує регіони-кандидати (proposed regions), які можуть містити об'єкти. Для кожної просторової позиції в тензорі характеристик RPN передбачає набір анкерних рамок (anchors) різних розмірів і співвідношень сторін. RPN визначає, які з цих ан-

керів є ймовірними рамками об'єктів, передбачаючи зсуви та ймовірності наявності об'єктів в кожному анкері. Обрізані та нормалізовані регіони проходять через додаткові шари, які передбачають клас об'єкта та точні координати рамки для кожного регіону. Це включає в себе повнозв'язні шари, які виконують класифікацію та регресію рамок. Алгоритм NMS застосовується для видалення дублюючих рамок об'єктів, залишаючи лише ті рамки, які мають найвищі ймовірності для кожного об'єкта [6].

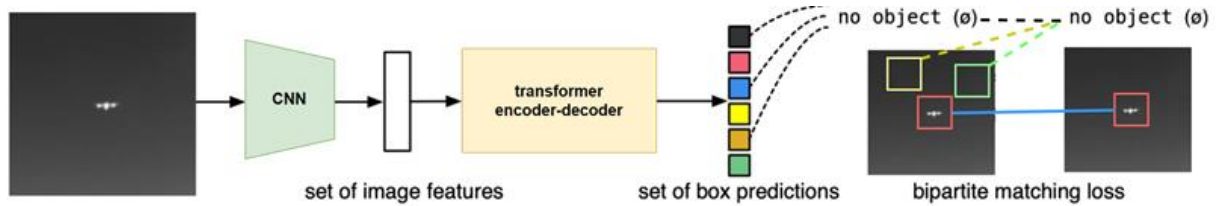


Рис. 3. Принцип роботи DETR моделі [7]

Вхідне зображення подається на вхід моделі та змінюється до фіксованого розміру. Це зображення проходить через базову згорткову нейронну мережу (CNN), з метою витягти ознаки зображення та перетворити його на багатовимірний тензор характеристик. Витягнуті характеристики зображення подаються до енкодера та трансформера. Енкодер складається з кількох шарів самоуваги та повнозв'язних шарів, які обробляють характеристики та виявляють зв'язки між різними частинами зображення. Самоувага дозволяє моделі фокусуватися на важливих частинах зображення та враховувати глобальний контекст. Декодер трансформера отримує вхідні запити (queries), які є набором векторів, що відповідають потенційним об'єктам. Декодер використовує механізм уваги для зіставлення цих запитів з ознаками, витягнутими енкодером. Кожен запит генерує прогноз щодо наявності об'єкта, його координат та класу. Виходи декодера є наборами прогнозів для кожного запиту. Це включає ймовірність наявності об'єкта, його коор-

динати та клас. Оскільки трансформер обробляє глобальний контекст зображення, ці прогнози можуть бути дуже точними навіть для складних сцен. Прогнози об'єктів обробляються для видалення дублюючих рамок та вибору найбільш вірогідних передбачень, використовуючи NMS алгоритм [7].

Створення датасету

Для тестування та порівняння наявних моделей виявлення БПЛА, необхідно отримати відповідний датасет, в якому будуть зібрані тепловізійні зображення та відео БПЛА. В цьому дослідженні використовувався датасет CVPR 2023 Anti-UAV Challenge, в якому зібрано близько 200 різних тепловізійних відео БПЛА з відповідними для них анотаціями. Декілька зображень з цього датасету зображені на рис. 4. Проте для навчання моделей, необхідно набагато більше даних. Це є однією з проблем навчання моделей для виявлення БПЛА в ІЧ спектрі – складність в розробці датасету і постійна нестача даних.



Рис. 4. Зображення дронів з CVPR 2023 Anti-UAV Challenge датасету

У зв'язку з вищеописаною проблемою, у роботі [8] запропонований та досліджений метод перетворення зображень у видимому діапазоні в зображення в ІЧ. Для цього необхідно провести наступні дії:

1. Переведення зображення у відтінки сірого;
2. Інверсія зображення.

В дослідженні [8], ще було запропоновано переведення зображення у HSV, проте автори не побачили в цьому якогось сенсу чи приросту для

навчання моделей, тому для даного дослідження достатньо перетворити зображення у відтінок сірого та виконати їх інверсію.

Для перетворення зображення у відтінок сірого, необхідно перемножити кожен складову кольорового зображення на коефіцієнти відтінку сірого та знайти їх суму:

$$X_g = 0,299 \cdot X_r + 0,587 \cdot X_g + 0,144 \cdot X_b . \quad (1)$$

Для інверсії зображення використовуємо наступну формулу:

$$X_i = 255 - X_g . \quad (2)$$

За основу зображень БПЛА у видимому діапазоні, було обрано датасет Drone vs Bird Challenge [9]. В цьому датасеті зібрано велику кількість зображень дронів в різній місцевості та з різним природним фоном.

Результат попередньої обробки зображень видимого діапазону для перетворення їх у псевдо-теплові можна побачити на рис. 5.



Рис. 5. Геометричні функції концентрації енергії в кругових зонах синтезованих об'єктивів, де: *a* – з кутовим полем зору 10° (поля зору №1-4 відповідають предметним точкам з кутовими координатами 0°, 2,5°, 3,5° і 5°); *б* – з кутовим полем зору 37° (поля зору №1-5 відповідають предметним точкам з кутовими координатами 0°, 5°, 10°, 15° і 18,5°); *в* – з кутовим полем зору 60° (поля зору №1-5 відповідають предметним точкам з кутовими координатами 0°, 7,5°, 15°, 22,5° і 30°)

Перетворені зображення використовувались виключно для навчання моделей. Для валідації та тестування використовувались реальні тепловізійні зображення БПЛА із Anti-UAV Challenge датасету.

Порівняння моделей

Для порівняння якості та ефективності моделей для задач виявлення БПЛА використовуючи тепловізійні зображення застосовуємо такі метрики, як середня точність виявлення (Average Precision або AP) при різних співвідношеннях площі перетину прогнозованої рамки та фактичної рамки до площі їх об'єднання (Intersection over Union або IoU), середнє значення повноти (Average Recall) та кількість кадрів, які модель може обробити за секунду (FPS).

IoU розраховується як співвідношення площі перетину до площі їх об'єднання. Для більшого розуміння, дана формула зображена на рис. 6.

$$IoU = \frac{\text{площа перетину}}{\text{площа об'єднання}} = \frac{\text{img1} \cap \text{img2}}{\text{img1} \cup \text{img2}}$$

Рис. 6. Intersection over Union

Точність (Precision) та повноту (Recall) можна розрахувати, використовуючи формули (3) та (4) відповідно [10]:

$$Precision = \frac{TP}{TP + FP} , \quad (3)$$

$$Recall = \frac{TP}{TP + FN} , \dots\dots\dots(4)$$

де TP – кількість справжніх позитивних прогнозів (true positive); FP – кількість помилкових позитивних (false positive); FN – кількість помилкових негативних прогнозів (false negative).

Середню точність виявлення можна розрахувати виходячи з розуміння, що AP – це площа під кривою PR (Precision-Recall), тобто середня точність виявлення – це інтеграл функції точності для різних значень повноти [10]:

$$AP = \int_0^1 p(r) dr . \quad (5)$$

Для більш практичного розрахунку AP формулу (5) можна представити в наступному вигляді, виконавши апроксимацію [10]:

$$AP = \sum_{k=0}^{k=n-1} (r(k) - r(k+1)) \cdot p(k) , \quad (6)$$

де $p(k)$ та $r(k)$ – точність та повнота відповідно при пороговому значенні k ; n – загальна кількість порогів функції.

Відповідно до табл. 1, найкращі результати виявлення мають моделі DETR та YOLOv8. DETR має кращі показники точності, проте у YOLOv8 більше кількість кадрів, що є більш важливим для систем

виявлення у реальному часі. DETR має найвищу середню точність AP, що вказує на його здатність точно розпізнавати та локалізувати об'єкти.

Таблиця 1. Результати виявлення БПЛА у ІЧ діапазоні

Модель	AP	AP _{0.75}	AP _{0.5}	AR	FPS
YOLOv8	0.605	0.688	0.921	0.614	40
YOLOv5	0.593	0.672	0.905	0.619	35
Faster RCNN	0.533	0.631	0.893	0.612	17
DETR	0.618	0.692	0.925	0.622	22

У YOLOv8 показник середньої точності ненабагато гірший від DETR. При високому значенні IoU (0.75) найбільшу середню точність показують DETR та YOLOv8, що свідчить про їхню здатність до точного розташування рамок об'єктів. DETR має найвищу середню повноту AR, що вказує на його здатність виявляти більшість об'єктів у зображенні. YOLOv8 має найвищий показник FPS, що робить його найшвидшою моделлю серед розглянутих, ідеальною для застосувань у реальному часі.

Faster RCNN показує найгірший результат, проте у роботі [11] розглядалось використання Faster RCNN моделі для крос-модального виявлення БПЛА, тобто з використанням тепловізійного зображення та видимого. Але виконуючи подібну задачу, Faster RCNN модель втрачає ще більше FPS, проте точність такого підходу є доволі високою.

Висновки

Дослідження показало, що кожна із аналізованих моделей має як свої переваги, так і недоліки. В різних умовах та при різних задачах кожна модель проявляє себе по-різному. Виходячи з порівняння, найшвидшою моделлю є YOLOv8 проте вона є менш точною порівняно з DETR.

В свою чергу, DETR має кращу точність, але у випадку трекінгу цілі в реальному часі такої кількості кадрів може бути не достатньо. Faster RCNN не дуже добре себе показало у даній задачі, але якщо розглядати систему, яка використовує як видимі, так і тепловізійні зображення, дана модель може значно покращити точність, але втратити у кількості кадрів. Виникає ідея у використанні даних моделей у комплексі, маючи змогу переключатися в різні режими виявлення, проте подібний підхід повинен бути досліджений та перевірений його резонансність та ефективність.

В загальному, отримані результати є задовільними, проте даної точності може бути не достатньо для виконання реальних задач з виявлення БПЛА на полі бою чи при захисті стратегічно важливих об'єктів. Тому є необхідність у подальшому дослідженні цього напрямку та в знаходженні шляхів покращення виявлення БПЛА на основі тепловізійних зображень.

Література

[1] A. Neubeck & L. Van Gool, "Efficient non-maximum suppression," in *18th international con-*

ference on pattern recognition (ICPR'06), pp. 850-855, August 2006. DOI: 10.1109/ICPR.2006.479.

- [2] J. Terven, D.-M. Córdova-Esparza & J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction* 5.4, pp. 1680-1716, 2023. DOI:10.3390/make5040083
- [3] G. Jocher, A. Chaurasia & J. Qiu, "YOLO by Ultralytics," 2023.
- [4] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 580-587, 2014. DOI: 10.1109/CVPR.2014.81.
- [5] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440-1448, 2015. DOI: 10.1109/ICCV.2015.169.
- [6] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, DOI: 10.1109/TPAMI.2016.2577031.
- [7] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov & S. Zagoruyko, "End-to-end object detection with transformers," in *Computer Vision – ECCV 2020, 16th European Conference*, Glasgow, UK, August 23–28, 2020, Proceedings, Part I, pp.213-229, 2020. DOI: 10.1007/978-3-030-58452-8_13.
- [8] L. Sommer and A. Schumann "Deep learning-based drone detection in infrared imagery with limited training data", in *Proc. SPIE 11542, Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies IV*, 1154204 (20 September 2020). DOI: 10.1117/12.2574171
- [9] A. Coluccia, A. Fascista, L. Sommer, A. Schumann, A. Dimou & D. Zarpalas, "The Drone-vs-Bird Detection Grand Challenge at ICASSP 2023: A Review of Methods and Results," *IEEE Open Journal of Signal Processing*, pp. 1-15, 2024. DOI: 10.1109/OJSP.2024.3379073.
- [10] R. Padilla, S. L. Netto and E. A. B. da Silva, "A Survey on Performance Metrics for Object-Detection Algorithms," in *2020 International Con-*

ference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 2020, pp. 237-242.

DOI: 10.1109/IWSSIP48289.2020.9145130.

[11] Q. Ma, B. Zhu, Z. Cheng & Y. Zhang, "Fast detection and recognition method of UAV in sky

background," in *Eleventh International Conference on Information Optics and Photonics (CIOP 2019)*, *Proc. of the SPIE*, vol. 1120907, pp. 36-41, 2019. DOI: <https://doi.org/10.1117/12.2542273>

UDC 004.93

Tanchuk V. S., Kolobrodov V. G.

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

ANALYSIS OF EXISTING OBJECT DETECTION MODELS FOR UAV DETECTION ON THERMAL IMAGERY

Nowadays, unmanned aerial vehicles (UAVs) are becoming widely used. They have proven their effectiveness, reliability, and feasibility. However, improper use or abuse of this technology can lead to significant human rights violations and threaten public safety. In connection with this, appropriate methods of countering UAVs should be made. One of the areas that is actively developing nowadays is the detection of UAVs based on optoelectronic radiation. Since most of such systems are designed for the visible range, the IR range has not been so widely investigated. UAV detection based on thermal imaging can be performed using artificial intelligence. Currently, there are ready-made approaches and models that perform object detection tasks, but they are more general purpose.

To review the problem of UAV detection based on thermal imaging and AI, it is necessary to analyze the most effective object detection models for the problem of UAV detection in the IR range. In this work, an analysis and comparison of such object detection models as YOLOv5 and YOLOv8, Faster RCNN, DETR was carried out.

Datasets were used for model training, both in the IR range and in the visible range. Preliminary digital processing of the dataset of UAV images in the visible range was carried out to convert them into pseudo-thermal images. This was done in order to increase the amount of training data for the models and thereby improve their accuracy.

The results showed that models such as YOLOv8 and DETR are the most effective for UAV detection tasks in thermal imaging, but their accuracy is still insufficient to be effectively used for real-time UAV defense systems.

Key words: unmanned aerial vehicle, artificial intelligence, object detection model, YOLOv8, YOLOv5, Faster RCNN, DETR, dataset, pre-processing, average precision.

*Надійшла до редакції
20 червня 2024 року*

*Рецензовано
25 серпня 2024 року*



© 2024 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).