

ВИЯВЛЕННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННІ В ПОТОКОВОМУ РЕЖИМІ ПРИ ВИКОРИСТАННІ YOLOv5 і FASTER R-CNN

Анотація. Для виконання щоденних завдань (виявлення порушень правил дорожнього руху, щоденне виявлення натовпу та тощо) можуть використовуватися безпілотних літальних апаратів. Порівняння підходів YOLOv5 і Faster R CNN надає можливість вирішення проблеми низької точності безпілотних літальних апаратів у виявленні цілей.

У цій роботі розглядається можливість збору та анотування відповідного набору даних для навчання, наведено опис найбільш поширених моделей YOLOv5 і Faster R CNN виявлення об'єктів, а також наведені результати навчання і валідації на зібраному наборі даних. Для порівняння результатів двох різних моделей глибокого навчання для виявлення цілей застосовано низку стандартних метрик. На основі отриманих під час експерименту результатів пропонується вибір найбільш відповідної моделі для вирішення визначеної проблеми.

Ключові слова: машинне навчання, виявлення об'єктів, Faster R CNN, YOLOv5, виявлення в потоковому режимі, комп'ютерний зір, БПЛА.

Постановка проблеми

Починаючи з 20-го століття, виявлення об'єктів стало важливим результатом досліджень в комп'ютерному зорі. Ця задача широко розвивається і застосовується для виявлення транспорту, виявлення пішоходів і безпілотного водіння. З підвищенням точності зображення алгоритми виявлення об'єктів на основі згорткових нейронних мереж поступово стають актуальним напрямком досліджень.

Технологія безпілотних літальних апаратів (БПЛА) стала важливою сферою застосування дронів з поглибленим вивченням відповідної теорії та практики. Зі створенням високоточних датчиків і розвитком суміжних технологій продуктивність БПЛА постійно покращується.

Використання інформаційних технологій та підходів машинного навчання може вирішити завдання автономного польоту БПЛА за умов зникнення сигналу GPS, орієнтуючись лише на ціль.

Аналіз останніх досліджень і публікацій

Машинне навчання широко застосовується в комп'ютерному зорі та обробці зображень. Рівень розвитку машинного навчання та наявність різноманітних комплексних моделей можуть дозволити з достатньою точністю виявляти на зображенні потрібну ціль.

Точність моделі може бути одним з основних показників, на основі якого можна зробити висновок про придатність моделі для її практичної експлуатації. Однак з урахуванням специфіки виявленого завдання, також варто звернути увагу на швидкість роботи моделі, оскільки є необхідність оброблення даних у потоковому режимі. При занадто довгому обробленні даних моделлю можна отримати неактуальну інформацію.

Для дослідження можливостей використання машинного навчання в прикладній задачі розглянуто дві групи моделей розпізнавання об'єктів: YOLOv5 і Faster R-CNN. Розглянуто такі характеристики цих підходів, як швидкість і точність виявлення для вибору найкращої моделі для вирішення прикладного завдання або для подальшої комбінації їх.

Обидві архітектури належать до групи згорткових нейронних мереж, які є найбільш поширеними алгоритмами глибокого навчання, що застосовують кілька надточних шарів і обчислень. Алгоритми надають ефективні способи вилучення ознак, а також є найкращим вибором для вирішення проблем виявлення об'єктів.

Поточні підходи з використанням методів глибокого навчання для задач класифікації та регресії об'єктів можна розділити на дві категорії.

Двоетапні методи, які представлені такими архітектурами, як R-CNN, Fast R-CNN і Faster R-CNN. Спочатку використовується селективний пошук або мережу регіональних припущень (RPN) для виділення регіонів інтересу - областей, які з високою ймовірністю містять усередині себе об'єкти. Надалі за допомогою класифікатора визначається клас об'єкта та за допомогою регресора визначаються обмежувальні рамки. Цей метод має високу точність [3], але при цьому обмежений у швидкості виявлення.

До іншої категорії належать одноетапні алгоритми, які представлені різними версіями YOLO та ін. Ці алгоритми не використовують окрему мережу для генерації регіонів і ґрунтуються на методах регресії, переглядаючи зображення цілком. Оскільки ці алгоритми не використовують RPN, швидкість виявлення вища, але точність виявлення, особливо малих об'єктів, не така висока, як у двоетапних методів.

Мета дослідження

Метою дослідження є аналіз архітектурних рішень найбільш поширених моделей виявлення об'єктів YOLOv5 і Faster R-CNN для побудови моделі щодо покращення швидкості і точності виявлення об'єктів в прикладній задачі або для подальшої комбінації їх. Метою дослідження є отримання результатів навчання і валідації на зібраному наборі даних.

Faster R-CNN є алгоритмом виявлення об'єктів та складається із чотирьох частин: мережа вилучення ознак (feature extraction network), мережа регіональних припущень (region proposal network), об'єднання регіонів інтересу (RoI Pooling) і повнозв'язний шар. Faster R-CNN це покращена версія алгоритмів R-CNN і Fast R-CNN. Різниця між ними в тому, що алгоритм Faster R-CNN уникає використання обчислювально дорогого алгоритму селективного пошуку [2]. Замість цього використовує RPN для створення областей кандидатів. RPN дає змогу визначати припущення на основі ознак всього зображення за один прохід, що дає змогу уникнути повторних обчислень і збільшує швидкість виявлення.

YOLO пропонує нову ідею розв'язання задачі виявлення об'єктів завдяки перетворенню задачі на регресійну. YOLO - одноетапний алгоритм глибокого навчання, який використовує надточні нейронні мережі для виявлення об'єктів [1]. Існують різні версії цього алгоритму. YOLOv1 ділить зображення на комірки розміром n на n однакового розміру. Кожна комірка сітки відповідає за виявлення центру об'єкта всередині комірки. Кожна комірка може передбачити фіксовану кількість обмежувальних рамок зі значенням достовірності (confidence score). Кожна обмежувальна рамка містить у собі 5 значень: координата центру за віссю абсцис, координата центру за віссю ординат, висота рамки, ширина рамки та значення достовірності. Після визначення обмежувальних рамок YOLO використовує перетин над об'єднанням (IoU), щоб вибрати найбільш підходящі. Для видалення зайвих рамок використовується немаксимальне придушення (NMS). У YOLOv2 було додано пакетну нормалізацію (batch normalization) разом зі згортковими шарами для збільшення точності та зменшення можливості перенавчання. У YOLOv3 магістральна мережа вилучення ознак (feature extraction backbone) Darknet19, яка погано виявляла маленькі об'єкти, була замінена на Darknet53 для вирішення цієї проблеми. У YOLOv4 знову було замінено магістральну мережу вилучення ознак на CSPDarknet53, що значно поліпшило швидкість і точність алгоритму. YOLOv5 - найбільш полегшена версія з попередніх, яка використовує фреймворк PyTorch замість

Darknet. Також у неї було додано новий шар фокусу (focus layer), який замінив перші три шари магістральної мережі YOLOv3, що дозволило збільшити швидкість при мінімальних втратах у точності.

Викладення основного матеріалу дослідження

Підготовка навчального набору даних.

Для навчання моделей глибокого навчання потрібна велика вибірка зображень для отримання досить ефективної моделі. Існує безліч різних наборів даних, однак було обрано набір даних Kitti.

Загалом було зібрано 550 зображень для навчання і 105 зображень для валідації (рис. 1).



Рисунок 1 – Приклади зображень навчального набору даних

Також було зібрано набір даних із 573 зображень з нового місця розташування для остаточної валідації моделей.

Анотування зображень.

Оскільки використовуваний метод розпізнавання об'єктів є різновидом методу навчання з вчителем, виникає необхідність передавати моделі інформацію про місце розташування обмежувальних рамок цілей на зображенні. Тому, для анотування зображень передбачено використання Roboflow, який дає можливість не тільки розмічати зображення, а й експортувати анотовані набори даних у різних форматах. Цей підхід прискорює процес, оскільки YOLOv5 і Faster R-CNN мають різні формати анотацій.

Початкова конфігурація експериментального стенда.

Навчання і валідацію моделей проведено на платформі Google Colab, яка має конфігурацією: ОС Ubuntu 18.04, графічний процесор NVIDIA Tesla T4, процесор Intel(R) Xeon(R) CPU @ 2.20ГГц. На платформі використовується мова програмування Python і фреймворк PyTorch. Для навчання та валідації моделей

використано бібліотеки yolov5 і detecron2 для YOLOv5 і Faster R-CNN відповідно.

Для покращення результатів навчання обрано вже попередньо навчені моделі. Бібліотекою YOLOv5 надаються 5 різних переднавчених на наборі даних Kitti моделей, які відрізняються розміром, швидкістю виконання і точністю.

Обрано YOLOv5m, тому що вона порівняно з YOLOv5s має набагато більшу точність і не втрачає швидкості, а порівняно з YOLOv5l має трохи меншу точність, втрачаючи при цьому в швидкості. Бібліотекою detecron2 надається велика кількість переднавчених моделей на різних наборах даних (кількість епох-20, розмір пакету-16, швидкість навчання-0.001, моментум-0.9, скорочення вагів-0.0005).

Критерії оцінювання.

Для порівняння результатів двох різних моделей глибокого навчання для виявлення цілей застосовано низку стандартних метрик, які використовують для оцінювання моделей машинного навчання. Є всього 4 можливі результати під час виділення обмежувальних рамок і виявлення об'єкта: істинно позитивні (*TP*) - результат означає, що модель коректно визначила позитивний клас, істинно негативні (*TN*) - результат означає, що модель коректно визначила негативний клас, хибно позитивні (*FP*) - результат означає, що модель некоректно визначила позитивний клас і хибно негативні (*FN*) - результат означає, що модель некоректно визначила негативний клас.

У цій задачі однокласового виявлення позитивним класом є об'єкт, а негативним задній фон. Таким чином, *TP* - вірне виявлення об'єкта, *TN* - вірне виявлення фону (цей показник не використовується, оскільки область заднього фону не виділяється під час анування), *FP* - некоректне виявлення об'єкта, *FN* - відсутність виявлення.

Для визначення належності результату до однієї з чотирьох груп використовується метрика *IOU*, яка є відношенням області перетину до області об'єднання коректної та передбаченої обмежувальних рамок.

На основі отриманих результатів *TP*, *FP* і *FN* обчислені стандартні показники: точність (*Precision*), повнота (*Recall*) і середня точність (*AP*). Також розраховано окремо кількість кадрів, яку здатна обробити модель за одну секунду (*FPS*).

Точність визначає відношення числа коректно виявлених об'єктів до всієї кількості виявлених об'єктів і обчислюється за формулою $Precision = TP / (TP + FP)$.

Повнота визначає відношення числа коректно виявлених об'єктів до всіх об'єктів класу і обчислюється за формулою $Recall = TP / (TP + FN)$. Середня точність AP визначається площею під кривою точності-повноти і визначається як інтеграл від функції $p(r)$ на відрізку $[0;1]$, де p – точність, r – повнота.

Крива точності-повноти показує залежність зміни точності при зміні повноти для різних порогових значень. У цьому дослідженні графіки кривих точності-повноти побудовано на основі широко використовуваного порогового значення IOU , що дорівнює 0.6.

Середня точність є однією з найбільш використовуваних метрик для визначення якості моделі виявлення об'єктів. При цьому середня точність може містити різні метрики залежно від IOU . Найбільш використовуваними є значення AP при IOU не менше 0.5 і середнє значення AP при порогових значеннях IOU в діапазоні від 0.5 до 0.95 з кроком 0.5, які позначаються $AP@.5$ і $AP@.5:.95$ відповідно.

Кількість кадрів на секунду - це основний показник швидкості моделі. Здебільшого вважається, що показник FPS вище 30 достатній для виконання виявлення в потоковому режимі.

Порівняння результатів роботи моделей. Розмір навчених моделей YOLOv5 і Faster R-CNN склав 40.2 МБ і 230.8 МБ відповідно. Після навчання моделі протестовані на першому валідаційному наборі (табл. 1, рис. 2).

Таблиця 1

Результати валідації моделей на першому наборі даних

Архітектура	Точність, %	Повнота, %	$AP@.5$, %	$AP@.5:.95$, %	FPS, к/с
YOLOv5	93,6	93,7	96	50,2	56,5
Faster R-CNN	94,2	92,8	95,1	51	6,56

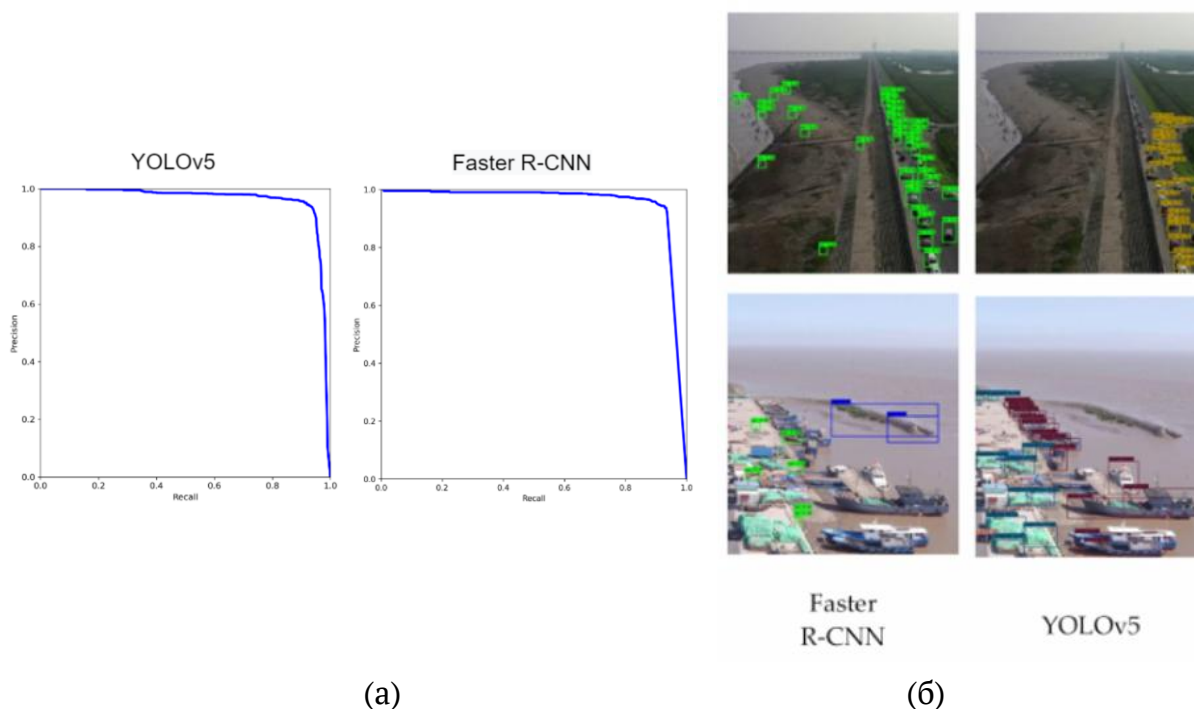


Рисунок 2 – (а) Криві точності-повноти моделей після валідації на першому наборі даних; (б) Візуалізація після валідації на першому наборі даних

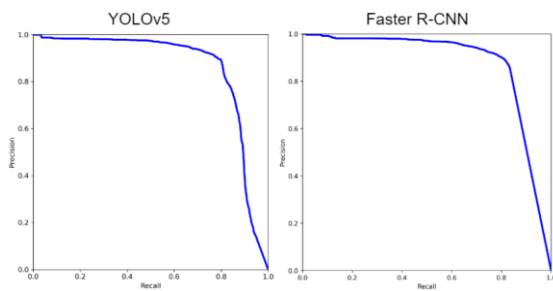
YOLOv5 має показники $AP@.5$ і FPS вищі на 0.9% і 49.94 к/с відповідно, ніж у Faster R-CNN. При цьому Faster R-CNN має показник $AP@.5:.95$ вищий на 0.8%, ніж у YOLOv5. Можна зробити висновок, що YOLOv5 швидше справляється з виявленням і має більшу точність за невеликого порогового значення IoU , а Faster RCNN зберігає більшу точність під час передбачення рамок, які більше збігаються.

Далі моделі були протестовані на другому валідаційному наборі (табл. 2, рис. 3).

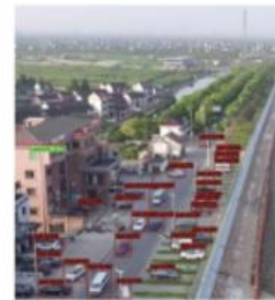
Таблиця 2

Результати валідації моделей на другому наборі даних

Архітектура	Точність, %	Повнота, %	$AP@.5$, %	$AP@.5:.95$, %	FPS, к/с
YOLOv5	89,3	79,9	86,1	34,1	100
Faster R-CNN	89,2	81,4	87,7	40,9	6,1



Faster
R-CNN



YOLOv5

(a)

(б)

Рисунок 3 – (а) Криві точності-повноти моделей після валідації на другому наборі даних; (б) Візуалізація після валідації на другому наборі даних

При отриманні зображення з нового місця розташування точність двох моделей зменшилася, проте Faster R-CNN має показник середньої точності вищий, ніж YOLOv5 на 1.6% та 6.8% для $AP@.5$ та $AP@.5:.95$ відповідно. З цього можна зробити висновок, що Faster R-CNN має більш глибоке навчання, що дозволяє не так сильно втрачати точність при отриманні зображення з новим середовищем.

Faster R-CNN не задовольняє вимогам виявлення в потоковому режимі, проте дає змогу досить точно визначати об'єкти в новому середовищі без донавчання. Цей алгоритм підходить у тих випадках, коли потрібно виявлення точних обмежувальних рамок і виявлення в різних середовищах без попереднього донавчання.

YOLOv5 задовольняє вимогам виявлення в потоковому режимі, при цьому зберігає досить високу точність. Однак при отриманні зображень з іншим середовищем може втратити точність виявлення як при порогового значення IoU , що дорівнює 0.5, так і за великих.

Висновки

У результаті дослідження зібрано набір даних із камер відеоспостереження і проанотоване за допомогою RoboFlow. Головні представники двох груп алгоритмів виявлення об'єктів YOLOv5 і Faster R-CNN навчені з використанням підготовленого набору даних. Експериментальним чином отримано метрики моделей. Результати показали, що обидві моделі мають свої переваги та недоліки, обидві моделі застосовні для різних завдань. Faster R-CNN зберігає високу точність за різних вхідних даних, але при цьому має повільну швидкість оброб-

ки, що не дає змоги її використовувати для виявлення в потоковому режимі. YOLOv5 трохи втрачає точність на нових наборах даних, маючи AP@.5 і AP@.5:.95 на 1.6% і 6.8% відповідно, ніж Faster R-CNN, проте володіє швидкістю оброблення в кілька разів вищою, ніж Faster R-CNN, що дає змогу використовувати її для обробки в потоковому режимі. Незважаючи на те, що YOLOv5 втрачає точність на новому наборі даних, показника AP має бути достатньо для вирішення завдання обчислення кількості об'єктів, похибка обчислення з такою точністю вважається допустимою. Якщо взяти до уваги, що досліджуване середовище з плином часу не змінюється, то можна донавчити модель для отримання більш точних результатів.

ЛІТЕРАТУРА

1. B. Li, M. M. Fu, and Q. Li Runway crack detection based on YOLOV5 // in Proc. IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Changsha, China, 2021, pp. 1252–1255.
2. B. Liu, W.C. Zhao, and Q. Q. Sun Study of object detection based on Faster R-CNN // in Proc. Chinese Automation Congress (CAC), Ji'nan, China, 2017, pp.6233–6236
3. Голенко М.Ю., Воротніков В.В., Єфіменко А.А. Методи покращення розпізнавання малих об'єктів алгоритму faster r-cnn для застосування на безпілотних літальних апаратах // Тези XIII Міжнародної науково-технічної конференції «Інформаційно-комп'ютерні технології», м. Житомир, 30–31 березня 2023 р. – Житомир: Житомирська політехніка, 2023. – с. 5-6. - URL: <https://conf.ztu.edu.ua/wp-content/uploads/2023/06/povnyy-tekst.pdf>

REFERENCES

1. B. Li, M. M. Fu, and Q. Li Runway crack detection based on YOLOV5 // in Proc. IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Changsha, China, 2021, pp. 1252–1255.
2. B. Liu, W.C. Zhao, and Q. Q. Sun Study of object detection based on Faster R-CNN // in Proc. Chinese Automation Congress (CAC), Ji'nan, China, 2017, pp.6233–6236
3. Golenko M.Yu., Vorotnikov V.V., Yefimenko A.A. Methods of improving the recognition of small objects of the faster r-cnn algorithm for use on unmanned aerial vehicles // Abstracts of the XIII International Scientific and Technical Conference "Information and Computer Technologies", Zhytomyr, March 30-31, 2023 - Zhytomyr: Zhytomyr Polytechnic, 2023. - p. 5-6. - URL: <https://conf.ztu.edu.ua/wp-content/uploads/2023/06/povnyy-tekst.pdf>.

Received 08.01.2024.
Accepted 10.01.2024.

Detection of objects in the image in streaming mode using YOLOv5 i Faster R-CNN

The accuracy of the model can be one of the main indicators, on a basis of which it is possible to conclude about the suitability of the model for its practical operation. However, taking into account the specifics of the identified task, it is also worth paying attention to the speed of the model, since there is a need to process data in streaming mode.

To investigate the possibilities of using machine learning in an applied problem, two groups of object recognition models considered: YOLOv5 and Faster R-CNN.

The purpose of the study is to analyze the architectural solutions of the most common object detection models YOLOv5 and Faster R-CNN to build a model to improve the speed and accuracy of object detection in an applied task or further combine them.

A total of 550 training images and 105 validation images collected. A dataset of 573 images from the new location also collected for final validation of the models. The use of Roboflow provided for image annotation, which allows not only to mark images, but also to export annotated data sets in various formats. Training and validation of the models carried out on the Google Colab platform. The platform uses the Python programming language and the PyTorch framework. The yolov5 and detecron2 libraries for YOLOv5 and Faster R-CNN, respectively, used for model training and validation.

To determine whether the result belongs to one of the four groups, the IOU metric is used, which is the ratio of the intersection area to the area of the union of the correct and predicted bounding frames.

The size of the trained YOLOv5 and Faster R-CNN models was 40.2 MB and 230.8 MB, respectively. The models tested on the second validation set.

As result of the study, a set of data from video surveillance cameras collected and annotated using RoboFlow. The main representatives of two groups of object detection algorithms YOLOv5 and Faster R-CNN trained using the prepared data set. The results showed that both models have their advantages and disadvantages, both models are applicable for different tasks.

Божуха Лілія Миколаївна - к.ф.-м., доцент Дніпровський національний університет імені Олеся Гончара.

Сизоненко Олександра Дмитрівна – аспірант, Дніпровський національний університет імені Олеся Гончара.

Bozhukha Liliia - Dnipro National University named after Oles Gonchar.

Syzonenko Oleksandra - Dnipro National University named after Oles Gonchar.