

РОЗУМОВСЬКИЙ МИКОЛА

Національний лісотехнічний університет України

<https://orcid.org/0009-0003-7291-811X>e-mail: mykolarozumovskyy@nltu.edu.ua

КРОШНИЙ ІГОР

Національний лісотехнічний університет України

<https://orcid.org/0000-0003-0018-7998>e-mail: kroshny.igor@nltu.edu.ua

ПОШУК ОБ'ЄКТІВ НА СУПУТНИКОВИХ ЗНІМКАХ ЗА ДОПОМОГОЮ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

З прискореним розвитком штучного інтелекту технології аналізу зображень дистанційного зондування привернули значну увагу у розумних містах. Останніми роками дослідження у сфері виявлення об'єктів на супутникових знімках зосередилися на виявленні та підрахунку невеликих щільно розташованих об'єктів на великих просторах. Виявлення дрібних об'єктів, як окрема гілка обробки зображень, залишається значною проблемою через обмежену роздільну здатність, розмір, кількість та орієнтацію об'єктів, а також інші фактори. У цій статті розглядаються методи виявлення об'єктів на основі глибокого навчання та їх застосування для виявлення дрібних об'єктів на супутникових знімках. Мета статті — забезпечити читачів глибоким розумінням дослідницьких цілей. Зокрема, об'єднати основні набори даних і методи оцінки, які широко використовуються в останніх техніках виявлення об'єктів. Також проаналізувати проблеми нерегулярності виявлення об'єктів на супутникових знімках і надати огляд методів виявлення дрібних об'єктів. Додатково, відібрати методи з найкращими показниками за останні роки для експериментів та аналізу. Нарешті, окреслюються виклики та перспективи для майбутніх досліджень, пов'язаних із виявленням дрібних об'єктів на супутникових зображеннях.

Ключові слова: згорткові нейронні мережі, глибоке навчання, виявлення об'єктів, семантична сегментація, дистанційне зондування.

RAZUMOVSKYY MYKOLA, KROSHNYI IGOR

National Forestry University of Ukraine

SEARCH FOR OBJECTS ON SATELLITE IMAGES USING ARTIFICIAL NEURAL NETWORKS

With the accelerated development of artificial intelligence, remote sensing image analysis technologies have gained considerable attention in smart cities. In recent years, research on object detection in satellite imagery has focused on detecting and counting small, densely spaced objects over large areas. Small object detection, as a separate branch of image processing, remains a significant challenge due to the limited resolution, size, number and orientation of objects, as well as other factors.

To solve this problem, this article presents a series of methods for small object detection that will help the reader understand the current state of the art and choose the appropriate solution for specific tasks. To date, traditional object detection methods include the Histogram of Oriented Gradients (HOG) algorithm and the Deformed Particle Model (DPM) algorithm. However, since these algorithms are mostly focused on pedestrian detection, their effectiveness in processing remote sensing images is limited.

In recent years, convolutional neural networks (CNNs), a variant of feed-forward networks, have become widely used. This architecture effectively reduces the amount of memory required for deep neural networks. A work in deep learning-based object detection algorithms, Regional Convolutional Neural Networks (RCNNs) successfully combine convolutional neural networks with object detection. However, due to the fact that RCNN consists of four components, the efficiency of this algorithm is quite low. With the advent of detectors such as the YOLO and SSD series, the performance of object detection algorithms has improved significantly, and the technology has been constantly evolving.

The purpose of this article is to provide readers with a deep understanding of the research goals. In particular, to combine the main datasets and evaluation methods that are widely used in the latest object detection techniques. Also, to analyze the problems of irregularity of object detection on satellite images and provide an overview of small object detection methods. Additionally, to select the methods with the best performance in recent years for experiments and analysis. Finally, the challenges and prospects for future research related to small object detection in satellite images are outlined.

Keywords: convolutional neural networks, deep learning, object detection, semantic segmentation, remote sensing.

Постановка проблеми

Швидкий розвиток науки і технологій зробив технології дистанційного зондування незамінними для різноманітних застосувань. Серед таких застосувань можна виділити моніторинг захворювань, планування транспорту, екологічний моніторинг, аналіз урожаю, геологічні дослідження та ідентифікацію об'єктів у військових операціях. Основною метою виявлення та розпізнавання об'єктів, що є однією з ключових задач дистанційного зондування, є визначення елементів на цифрових зображеннях. Однак зображення, отримані методом дистанційного зондування, охоплюють значні території, що призводить до різноманітності розмірів об'єктів. Це створює значні труднощі при застосуванні алгоритмів для виявлення об'єктів на таких зображеннях.

Щоб вирішити цю проблему, у статті представлено серію методів для виявлення дрібних об'єктів, що допоможуть читачеві зрозуміти поточний стан розвитку цієї технології і вибрати відповідне рішення для конкретних задач. На сьогоднішній день традиційні методи виявлення об'єктів включають алгоритм Гістограми Орієнтованих Градієнтів (НОГ) та алгоритм Моделі Деформованих Частинок (DPM). Алгоритм НОГ працює шляхом поділу вхідного зображення на сітку, створення гістограми для кожної комірки сітки, і виявлення області інтересу для подальшої класифікації за допомогою методу опорних векторів (SVM).

Алгоритм DPM є вдосконаленням НОГ і пропонує ефективніші рішення для проблеми множинних

перспектив об'єкта. Проте, через те, що ці алгоритми здебільшого націлені на виявлення пішоходів, їх ефективність при обробці зображень дистанційного зондування обмежена.

Останніми роками згорткові нейронні мережі (CNN), що є варіантом мереж з прямим розповсюдженням, набули широкого використання. Ця архітектура ефективно зменшує обсяг пам'яті, необхідний для роботи глибоких нейронних мереж. Зменшення кількості параметрів у мережі та уникнення перенавчання досягаються завдяки трьом ключовим операціям: локальним перцептивним полям, спільному використанню ваг і шару об'єднання. Загалом, CNN складається з кількох чергувальних шарів згортки та об'єднання. Кожен нейрон вихідної карти ознак у шарі згортки локально підключений до свого входу. Відповідні ваги множаться на локальні входи, і до результату додаються значення зсуву для отримання вхідних значень нейрону. З розвитком глибокого навчання було запропоновано велику кількість алгоритмів виявлення об'єктів на основі CNN, які досягли вражаючих результатів при роботі з наборами даних супутникових знімків.

Робота з алгоритмів виявлення об'єктів на основі глибокого навчання, Регіональні згорткові нейронні мережі (RCNN) успішно поєднують згорткові нейронні мережі з виявленням об'єктів. Однак, через те, що RCNN складається з чотирьох компонентів — генерація кандидатних вікон, витяг ознак, класифікація за допомогою SVM та регресія вікон, — ефективність цього алгоритму досить низька. Вирішенням цієї проблеми стали наступні модифікації, такі як Spatial Pyramid Pooling (SPPNet), Fast RCNN, Faster RCNN, Feature Pyramid Network (FPN), Mask RCNN, які вдосконалили недоліки попередніх підходів та покращили продуктивність детекторів.

З появою таких детекторів, як серії You Only Look Once (YOLO) та Single Shot MultiBox Detector (SSD), продуктивність алгоритмів виявлення об'єктів значно покращилася, і технологія постійно розвивалася.

Вступ

У цій статті детально проаналізовано супутникові знімки та метрики оцінки, які зазвичай використовуються для виявлення об'єктів, що відрізняється від існуючої літератури. Стаття зосереджується на різних категоріях технік виявлення об'єктів, обмеженнях, пов'язаних із зображеннями дистанційного зондування, та викликах, спричинених нерегулярностями об'єктів, а також стратегіях для їх вирішення. Крім того, досліджуються методи виявлення дрібних об'єктів на супутникових зображеннях.

У статті підсумовані застосування методів виявлення дрібних об'єктів на супутникових знімках. Подано класифікацію існуючих процесів на шість категорій на основі різних технічних основ, включаючи найсучасніші техніки останніх двох років. Окрім цього, повторно оцінено середню точність виявлення (mAP), кількість операцій з плаваючою точкою (FLOPs), кількість параметрів (Params) та кількість кадрів за секунду (FPS) для шести найкращих методів. Ці алгоритми були оцінені на основі цих показників.

Крім того, приділено велику увагу розробці алгоритмів та застосуванню для виявлення дрібних об'єктів.

Основні внески цієї статті є наступними:

- Представлено детальний огляд процесу виявлення об'єктів за допомогою глибокого навчання, охоплюючи такі теми, як визначення проблеми, історія розвитку, сучасний стан досліджень, набори даних та методології оцінювання.
- Підхід до організації, класифікації та порівняння різних методів виявлення об'єктів, спираючись на різні принципи диференціації. Проблему нерегулярності у виявленні об'єктів для дистанційного зондування вирішують різноманітні підходи та методології. У статті розглядаються найсучасніші методи, а також методи обробки дистанційних знімків і виявлення малих об'єктів.
- Для методів виявлення малих об'єктів у дистанційному зондуванні проведено детальну класифікацію та аналіз літератури. Класифіковано алгоритми виявлення малих об'єктів на шість категорій, включаючи багатоскальне прогнозування, підвищення роздільної здатності ознак, контекстну інформацію, покращення даних, нову архітектуру мережі та стратегію навчання, а також проблему розриву на кордонах.
- Надано глибокий аналіз проблем і труднощів, пов'язаних з виявленням малих об'єктів у зображеннях дистанційного зондування з різних точок зору, і визначено тенденції та напрямки майбутнього розвитку.

Набори даних та методи оцінки

Набори даних

1. DIOR [1]: Публічний набір даних для масштабної оцінки виявлення об'єктів на оптичних зображеннях дистанційного зондування. Він містить 20 типів об'єктів, таких як літаки, стадіони, мости, дамби, порти тощо. Загалом у ньому 23,463 зображення.
2. RSOD [2]: Цей набір даних значно менший за кількістю типів та зображень порівняно з DIOR. Він включає лише 4 типи об'єктів: літаки, нафтові барабани, естакади та спортивні майданчики — загалом 976 зображень.
3. NWPU VHR-10 [3]: Набір даних NWPU-RESISC45 був запропонований дослідниками з Північно-Західного політехнічного університету і включає 45 категорій та 31,500 зображень. Результати експериментів були не зовсім задовільними через низьку роздільну здатність зображень. Як рішення було створено набір даних NWPU VHR-10 з 10 категоріями високої роздільної здатності (VHR), які широко використовуються в задачах виявлення об'єктів. Недоліком цього набору є відсутність позначених малих об'єктів, що знижує ефективність алгоритмів для їх виявлення.

4. DOTA [4]: Цей набір містить 2806 зображень і 15 типів об'єктів різних масштабів, орієнтацій і форм, що робить алгоритми виявлення більш вимогливими.
5. VEDAI [5]: Набір містить велику кількість транспортних засобів і призначений для виявлення автомобілів на зображеннях дистанційного зондування. Оскільки автомобілі є малими об'єктами, цей набір добре підходить для задач виявлення малих об'єктів.
6. ITCVD [6]: Набір для виявлення транспортних засобів на зображеннях дистанційного зондування, що містить 23,678 зображень. Із них 23,543 зображення використовуються для тестування, і кожен транспортний засіб був позначений вручну.
7. COCO [7]: Один із найпоширеніших наборів для виявлення об'єктів, особливо малих. Він містить 91 тип об'єктів і 328,000 зображень.
8. UCAS-AOD [8]: Набір містить 2819 зображень автомобілів та 3210 зображень літаків.
9. RSC11 [9]: Набір містить 11 схожих класів сцен, що ускладнює класифікацію.

Горизонтальна обмежувальна рамка (HBB) зазвичай використовується для представлення об'єктів, орієнтованих горизонтально, у маркуванні наборів даних. Об'єкти, які не підлягають обертанню, зазвичай зображуються за допомогою методу орієнтовано обмежувальної рамки (OBV).

HBB (ієрархічно обмежувальна рамка) вимагає, щоб рамка була орієнтовано перпендикулярна до координатних осей. Ця орієнтація обмежує рамку у здатності повністю охоплювати частково спотворені великі об'єкти. Орієнтація та масштаб рамки визначаються OBV, який враховує форму об'єкта. Рамка не обов'язково має бути перпендикулярним до координатних осей, а утворена активуюча рамка є порівняно компактніша, ніж орієнтовано обмежувальна рамка.

Щодо методів створення OBV, домінуючим методом є метод аналізу головних компонент (PCA). Спочатку OBV використовує PCA для отримання трьох головних напрямків хмари точок, щоб визначити центр мас і обчислити коваріацію. Потім отримується матриця коваріації, з якої знаходяться власні значення та власні вектори. Серед них власні вектори є головними напрямками. На другому етапі OBV перетворює вхідну хмару точок в початок координат за допомогою головних напрямків і центру мас.

Головний напрямок збігається з напрямком системи координат для побудови обмежувальної рамки хмари точок, перетвореної в початок координат. Нарешті, OBV задає головний напрямок і обмежувальну рамку для вхідної хмари точок, досягаючи остаточного ефекту через обернене перетворення вхідної хмари точок до початкової хмари точок.

Як показано в таблиці 1, склад базового набору даних, включаючи кількість об'єктів, класів, екземплярів і анотації, суттєво впливає на навчання та тестування моделі. Ефективне навчання моделі можна полегшити, використовуючи багаті екземпляри, різноманітні класи та відповідний стиль анотації. У таблиці 1 класи DIOR і DOTA становлять 20 і 15 відповідно, і їхня кількість екземплярів значно перевищує інші набори даних. Крім того, стиль анотації OBV допомагає покращити виявлення обернених об'єктів. DIOR використовує комбінацію обох стилів анотації — HBB і OBV, тоді як RSOD і NWPU VHR-10 використовують лише стиль анотації HBB. Крім того, DOTA включає всі OBV. Відмінні характеристики DIOR і DOTA відрізняють їх від інших наборів даних дистанційного зондування.

Таблиця 1

Порівняння класичних наборів даних

Набір даних	Кількість	Класи	Екземпляри	Анотація
DIOR [1]	23,463	20	192,472	HBB + OBV
RSOD [2]	976	4	6950	HBB
NWPU VHR-10 [3]	800	10	3775	HBB
DOTA [4]	2806	15	188,282	OBV
VEDA [5]	1210	9	3640	OBV
ITCVD [6]	173	1	29,088	OBV
UCAS-AOD [8]	910	2	6029	HBB + OBV
RSC11 [9]	1213	11	-	Scene Class

Методи оцінювання

У цьому розділі зосереджено на п'яти метриках оцінювання, які зазвичай використовуються для оцінки ефективності виявлення об'єктів: перетин над об'єднанням (IoU), точність, відзив, середня точність (AP) і усереднена середня точність (mAP).

1. IoU: При виявленні об'єкта створюється детекційний кадр. IoU — це відношення перетину та об'єднання площі передбачуваного та реального кадру. Зазвичай поріг встановлюється на рівні 0.5, що також є порогом для коефіцієнта перетину. Якщо значення перевищує 0.5, виявлений об'єкт вважається знайденим. Коефіцієнт перетину визначається наступним чином:

$$IoU = \frac{A \cap B}{A \cup B}, \quad (1)$$

2. Precision: Точність представляє собою відношення правильно виявлених зразків до загальної кількості

зразків у результатах прогнозування. Коли коефіцієнт перетину та об'єднання перевищує поріг, результат класифікується як позитивний (True Positive, TP), а в іншому випадку — як хибнопозитивний (False Positive, FP). Якщо детектор не виявляє об'єкт у кадрі виявлення, позначеному зразком, цей об'єкт класифікується як хибнонегативний (False Negative, FN). Точність визначається наступним чином:

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} = \frac{\text{TruePositive}}{\text{AllObservations}}, \quad (2)$$

3. Recall: Рівень вказує на кількість позитивних зразків, які модель виявила в загальному обсязі позитивних зразків. Це важливий показник для вимірювання того, чи модель "знайшла всі" об'єкти. Recall визначається наступним чином:

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} = \frac{\text{TruePositive}}{\text{AllGroundTruth}}, \quad (3)$$

4. Середня точність (AP): Середня точність є середнім значенням точності в межах інтервалу [0,1] відзиву. Чим вищий значення AP, тим краща продуктивність детектора у виявленні певного типу об'єкта в наборі даних. Середня точність визначається наступним чином:

$$\text{AP}_u = \frac{1}{\Omega_u} \sum_{i \in \Omega_u} \frac{\sum_{j \in \Omega_u} h(p_{uj} < p_{ui}) + 1}{p_{uj}}, \quad (4)$$

де Ω_u позначає істинний результат (Ground Truth), p_{uj} — це місцезнаходження об'єкта j , а $p_{uj} < p_{ui}$ означає, що об'єкт j ранжується перед елементом i у списку рекомендацій.

5. Усереднена середня точність (mAP): є середнім значенням середньої точності для кожного класу об'єктів, виявлених детектором. Вищі значення mAP вказують на кращу продуктивність детектора для всього набору даних. Усереднена середня точність визначається наступним чином:

$$\text{mAP} = \frac{\sum_{u \in U} \text{AP}_u}{|U|}, \quad (5)$$

6. FPS (Frames Per Second): FPS використовується для оцінки швидкості виявлення об'єктів, тобто кількість зображень, які можуть бути оброблені за секунду. Чим вищий FPS, тим швидша швидкість виявлення моделі.
7. FLOPs (Floating Point Operations): FLOPs означає кількість операцій з плаваючою комою, що також можна інтерпретувати як кількість обчислень. Чим менше FLOPs, тим менша складність моделі.
8. Params (Параметри): Params представляє кількість параметрів, необхідних для моделі. Чим менше Params, тим менше параметрів потребує модель і тим легшою вона є.

Виявлення об'єктів

Виявлення об'єктів досліджується та вдосконалюється протягом понад двох десятиліть. Ідентифікація об'єктів, яка вважається одним із ключових напрямів і основних викликів комп'ютерного зору, на даний момент розвивається разом із вимогами багатьох застосувань. Протягом історії свого розвитку, виявлення об'єктів можна розділити на два основні періоди: алгоритми традиційного виявлення об'єктів та алгоритми на основі глибокого навчання.

Традиційні методи виявлення об'єктів

У цьому розділі ми обговорюємо три основні традиційні детектори: детектор Віола–Джонса, HOG (Histogram of Oriented Gradients) та DPM (Deformable Part-based Models).

Детектор Віола–Джонса був запропонований у 2001 році, і здебільшого використовувався для виявлення облич. Він поєднує такі методи, як цілісні зображення, каскадні класифікатори та інші технології, досягаючи високої продуктивності.

Детектор HOG, представлений у 2005 році, має ширші функціональні можливості, виходячи за межі розпізнавання облич. Він зосереджений на вилученні ознак з об'єктів, що дозволяє ефективніше їх виявляти.

Детектор DPM, розроблений у 2009 році, базується на принципах детектора HOG. Він здатний виявляти окремі компоненти об'єктів, що дозволяє досягти більшої точності у виявленні.

Хоча ці традиційні детектори мали успіх на ранніх етапах, їх продуктивність виявилася недостатньою для великих наборів зображень або коли потрібно виявляти багато об'єктів. Поява методів глибокого навчання призвела до створення нових детекторів, які значно перевершують традиційні методи.

Методи виявлення об'єктів на основі прив'язки

Методологія виявлення об'єктів на основі прив'язки (anchor-based detection) передбачає створення набору дискретних кадрів-кандидатів для кожного пікселя зображення, після чого до цих кадрів застосовуються методи фільтрації, класифікації та регресії. Такий підхід частково вирішує проблеми, пов'язані з різними розмірами об'єктів і їхньою оклюзією (перекриттям).

Ця технологія має потенціал для підвищення рівня запам'ятовування, особливо при виявленні малих об'єктів. Проте вона також має свої недоліки: вона сильно залежить від великої кількості розроблених вручну компонентів. Крім того, процес навчання та тестування вимагає значної кількості часу, що знижує загальну ефективність.

Методологія на основі прив'язки включає два типи детекторів: одноступеневий та двоступеневий. У двоступеневих детекторах процес складається з двох етапів:

- Генерація кадрів-кандидатів зображення.
- Коригування обраних областей для отримання остаточних результатів виявлення.

Ця структура забезпечує більшу точність за рахунок складнішого процесу навчання, тоді як одноступеневі детектори зазвичай швидші, але менш точні. Блок-схема процесу для двоступеневих детекторів зображена на рисунку 1.

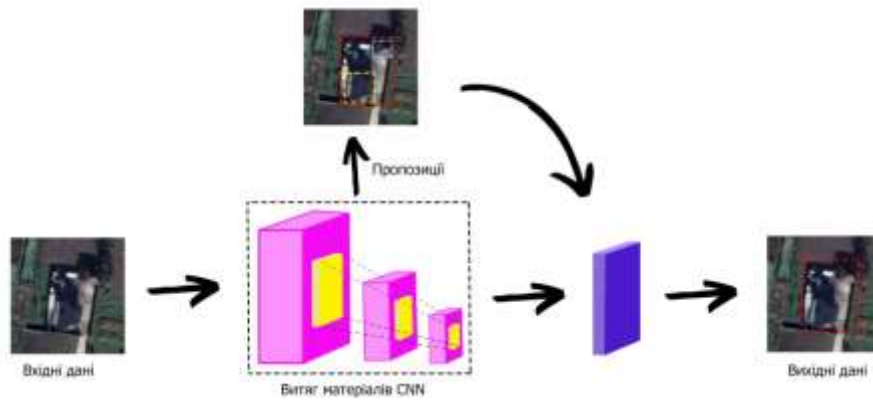


Рис 1. Фреймворк для двоетапного виявлення об'єктів

Детектори двох етапів включають такі методи, як R-CNN та SPPNet. R-CNN став піонером у використанні нейронних мереж для вирішення проблеми виявлення об'єктів. Проте цей підхід має кілька обмежень, зокрема: розподілене навчання для кожного компонента, незалежне зберігання даних та велика кількість надлишкових регіонів-кандидатів. Це призводить до значного навантаження на час і простір обчислень.

Наприклад, SPPNet вирішує проблему втрати об'єктів через масштабування або зміну їх розмірів. Воно пропонує використання шару просторового пірамідального згортання (SPP), який дозволяє конволюювати зображення лише один раз, уникаючи надмірних обчислень і економлячи час. Однак, попри поліпшення продуктивності, SPPNet має ті самі обмеження, що й R-CNN, зокрема, відсутність можливості здійснення детекції "від кінця до кінця" (end-to-end). Схематичне зображення SPPNet представлено на рисунку 2.

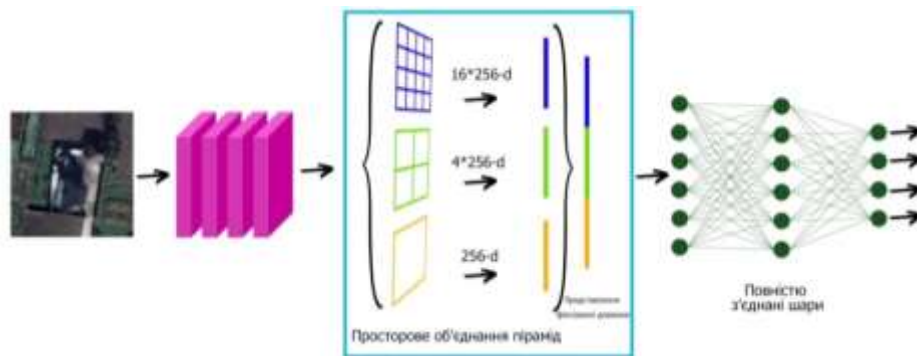


Рис 2. Структура мережі SPPNet

Підхід Fast RCNN використовує softmax-класифікатор для вирішення проблеми синхронізації класифікації. Крім того, застосовуються RoI (Region of Interest) шари для полегшення відображення багатомасштабних ознак, що допомагає вирішити проблему зміни масштабу. Багатофункційна функція втрат Fast RCNN дозволяє виконувати навчання "кінець до кінця" для багатозадачних цілей. Проте, через складний алгоритм вибору регіонів-кандидатів, швидкість виявлення виявляється повільною.

Faster RCNN успадковує переваги Fast RCNN, запропонувавши новаторське використання мережі для вибору регіонів (Region Proposal Network), що значно підвищує швидкість обчислень. Однак, у нього є недоліки: неточне визначення меж об'єктів та низька ефективність виявлення малих об'єктів.

FPN будує багатомасштабні піраміди ознак для різних масштабів зображень, що суттєво покращує точність мережі. Проте, оскільки мережа може бути навчена лише для конкретної роздільної здатності, це може суперечити багатомасштабному висновку.

Cascade RCNN використовує каскадний детектор для вибору порогів на основі критеріїв якості. Це рішення ефективно вирішує проблему перенавчання, яка може виникнути через високі порогові, але не підтримує роботу в реальному часі.

Region-based-FCN додає карту позиційно-чутливих оцінок, щоб підвищити чутливість згорткової мережі до місцезнаходження об'єкта. Хоча це вирішує проблему нечутливості до положення об'єктів, швидкість обчислень залишається незмінною.

Mask RCNN вирішує проблему одночасної локалізації, класифікації та сегментації об'єктів, вводячи гілку для сегментації на рівні екземплярів, що дозволяє досягти виявлення на рівні пікселів. Проте через високу вартість сегментації його продуктивність залишається нижчою за реальні вимоги до роботи в реальному часі.

У виявленні багатомасштабних об'єктів TridentNet відрізняється від багатомасштабної піраміди ознак FPN. Він використовує багатомасштабну структуру з різними полями сприйняття і спільними вагами для гілок, що покращує точність виявлення. Проте, через повільну швидкість виявлення, моніторинг в реальному часі неможливий.

Одноступеневі детектори можуть отримати фінальний результат виявлення одразу після одного етапу, що робить їх швидшими за двоступеневі детектори. На рисунку 3 представлена блок-схема одноступеневих детекторів. Серія YOLO, яка є одним із одноступеневих детекторів, постійно еволюціонує.

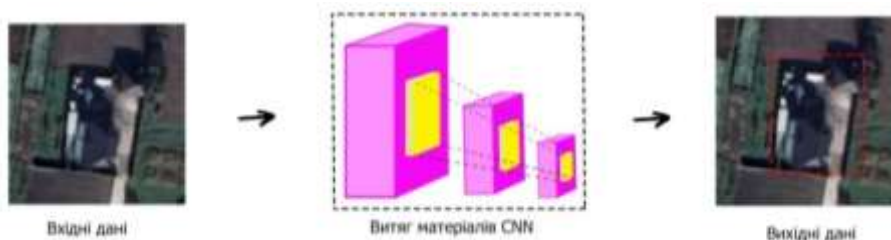


Рис 3. Фреймворк для одноетапного виявлення об'єктів

YOLOv1 був першим, хто перетворив задачу виявлення об'єктів на проблему регресії. Він має простішу структуру мережі та високу швидкість виявлення. Однак точність локалізації об'єктів значно нижча, ніж могла б бути. Коли об'єкт невеликий або є багато об'єктів, ефективність виявлення YOLOv1 знижується.

YOLOv2 підвищує як точність, так і швидкість виявлення, але не вирішує обмежень YOLOv1. Для вирішення проблеми недостатнього виявлення малих об'єктів, YOLOv3 використовує метод екстракції багатомасштабних ознак і покращену класифікаційну мережу. Проте вона неефективна для виявлення середніх і великих об'єктів.

YOLOv4 використовує стратегії підвищення даних Mosaic і самосуперницьке навчання. Він інтегрує FPN і Path Aggregation Network (PAN), що ще більше покращує продуктивність моделі. YOLOv5 має трохи гіршу продуктивність у порівнянні з YOLOv4, але є гнучкішою, швидшою та краще підходить для швидкого розгортання моделей. YOLOv6 додатково підвищує точність і швидкість, досягаючи найвищої точності в режимі реального часу.

SSD використовує багатомасштабну екстракцію ознак та згорткове виявлення ознак, забезпечуючи швидкість і високу точність. Проте SSD більше покладається на ручну настройку параметрів для попередньо вибраних рамок, що знижує його точність виявлення малих і кількох об'єктів.

RetinaNet використовує функцію втрат Focal Loss, що вирішує проблему дисбалансу категорій, проте він не підтримує виявлення в реальному часі та має погані результати для малих і кількох об'єктів.

EfficientDet пропонує зважену двонаправлену мережу піраміди ознак, яка спрощує та прискорює багатомасштабну інтеграцію ознак. EfficientDet також використовує складене масштабування, що одночасно масштабує роздільну здатність, глибину та ширину базової мережі. Однак він має повільнішу швидкість виявлення.

Методи виявлення об'єктів без якоря

Методи виявлення об'єктів без якорів (anchor-free) не вимагають попередньо визначених якорів. Вони визначають об'єкт за допомогою кількох ключових точок або центроїдів і здійснюють його виявлення безпосередньо. Ці методи можна поділити на алгоритми на основі центроїда і алгоритми на основі ключових точок.

Алгоритми на основі центроїда поєднують класифікацію та регресію у дві підсітки та безпосередньо виявляють центральну область і межі об'єкта. Представниками таких алгоритмів є CenterNet, FCOS та TTFNet.

Наприклад, CenterNet використовує алгоритм виявлення ключових точок, де об'єкти виявляються як точки з використанням центрованого пулінгу та каскадного пулінгу кутових точок. Однак цей підхід не підходить для виявлення малих об'єктів та об'єктів у великій кількості через високу обчислювальну складність моделі.

FCOS застосовує повністю згорткову мережу для виконання регресії на основі відстані від кожного пікселя на карті ознак до меж об'єкта. Подібно до принципу FCN, FCOS розглядає кожну позицію як навчальний приклад. Порівняно з алгоритмами, що використовують якорі, FCOS заощаджує значну кількість пам'яті під час тренування, що робить його придатним для сегментації на рівні екземплярів.

TTFNet є покращеною версією CenterNet. Він використовує еліптичне ядро Гауса для генерування сигналів навчання на основі негативних зразків та регіонів навколо центроїда. TTFNet підвищує ефективність навчання та якість сигналів нагляду, зменшуючи кількість попередньої обробки даних.

Алгоритми на основі ключових точок (також відомі як corner point-based algorithms) використовують позиції верхньої лівої та нижньої правої точки об'єкта для створення рамки виявлення. Представники цих алгоритмів – CornerNet та ExtremeNet. Їм притаманні помилкові спрацьовування через нестачу інформації всередині об'єкта.

На відміну від CornerNet, ExtremeNet використовує п'ять ключових точок: верхню, нижню, ліву, праву та центральну. ExtremeNet витягує локальну інформацію з меншою кількістю шуму, що дозволяє покращити продуктивність виявлення.

Методи виявлення об'єктів на основі трансформаторів

Методи виявлення об'єктів на основі Transformer значно розширилися останніми роками. Початково

модель Transformer використовувалася переважно для завдань обробки природної мови (NLP), але згодом знайшла застосування у сфері комп'ютерного зору під назвою Vision Transformer (ViT).

Серед основних алгоритмів виявлення об'єктів на основі Transformer виділяються DETR та YOLOs:

DETR (Detection Transformer) подібний до традиційних методів виявлення об'єктів, але відрізняється підходом до їх представлення. Традиційні методи на основі якорів класифікують заздалегідь визначені якірні точки і регресують коефіцієнти країв. Натомість DETR перетворює задачу виявлення об'єктів на завдання прогнозування ансамблів. Зображення перетворюється в послідовність ансамблів, і ця послідовність обробляється для визначення об'єктів.

YOLOS об'єднує енкодер-декодер з DETR із лише енкодерною частиною ViT. YOLOS демонструє потужні можливості трансферу та узагальнення ViT із мінімальними змінами, що робить його гнучким для різних завдань виявлення об'єктів.

Один із популярних варіантів — Swin Transformer, який вирішує проблему надмірної обчислювальної складності, властиву багатьом Transformer-моделям. На відміну від ViT, що застосовує зниження роздільної здатності в 16 разів, Swin Transformer пропонує три різні розміри функціональних ознак: зменшені в 4, 8 і 16 разів. У контексті обчислювальних витрат Swin Transformer вводить концепцію W-MSA (Window-based Multi-head Self-Attention). Для зменшених карт ознак у $4\times$ та $8\times$ цей метод ділить карту ознак на декілька областей, що не перетинаються, і виконує операції MSA всередині кожного вікна. Це дозволяє ефективно передавати дані між вікнами, знижуючи обчислювальну складність.

Обробка зображень дистанційного зондування

Обробка зображень дистанційного зондування

Обмеження зображень дистанційного зондування

Зображення дистанційного зондування мають кілька особливостей, що ускладнюють виявлення об'єктів. Основні проблеми включають: різну роздільну здатність зображень, через яку об'єкти виглядають по-різному на різних масштабах; малий відносний розмір об'єктів, що призводить до втрати деталей під час обробки; і обмежену кількість позначених даних, що ускладнює навчання моделей. Ці фактори роблять завдання виявлення маломасштабних об'єктів у зображеннях дистанційного зондування особливо складним.

Методи покращення зображень

Об'єкти з нерегулярними масштабами та формами знижують ефективність детекторів. Тому використання геометричних змін для обробки навчальних зображень стало нагальною проблемою. В даний час виник новий режим когнітивної обробки. Стиснення навчальних зображень може зменшити обсяг пам'яті для високоякісних зображень, а також скоротити час кодування. Використання цієї техніки підвищує ефективність виявлення об'єктів.

Методи покращення зображень є зменшення шуму [10], фільтрація зображень [11], посилення країв [12], обертання зображень [13], масштабування зображень [14] та стиснення зображень [15].

Нерегулярне виявлення об'єктів на зображеннях дистанційного зондування

На зображеннях дистанційного зондування багато об'єктів розташовані близько один до одного та неправильно орієнтовані, що впливає на ефективність виявлення об'єктів у різній мірі. У цьому розділі обговорюються можливі рішення проблеми нерегулярності об'єктів на зображеннях дистанційного зондування.

Виявлення об'єктів з урахуванням напрямку

Останні методи виявлення об'єктів в основному вводять задачі регресії напрямків у класичних детекторах об'єктів. Серед них SCRDet, CADNet, DRN, R3Det, ReDet та Oriented RCNN підвищують ефективність, прогножуючи кут обертання рамки. GlidingVertex та RSDet поліпшили ефективність, повертаючись до чотирикутників. Наприклад, Yang та ін. [16] використали задачу класифікації кутів для вирішення проблеми розриву кордонів при оцінці кутового напрямку. Усі вищезазначені методи покращують оцінку орієнтації на основі представлення кута обертання, що, в свою чергу, підвищує ефективність виявлення об'єктів.

Навчання нестандартних ознак

Методи виявлення об'єктів, такі як YOLOv1, Faster RCNN, FCOS, RepPoints, APD, FAN та CenterNet, орієнтовані на вертикально розташовані або аксіально вирівняні об'єкти. Вони погано справляються з виявленням об'єктів, які розподілені щільно і не є аксіально вирівняними. Щоб вирішити цю проблему, Хан та ін. [17] розробили модуль вирівнювання ознак, щоб зменшити невідповідність між аксіально симетричними та неаксіально симетричними об'єктами. Дінг та ін. [18] застосували просторову трансформацію аксіально симетричних Rols, що дозволяє моделі навчатися представленням неаксіально симетричних об'єктів під контролем в умовах обертання рамки. SCRDet++ отримав вищі відповіді об'єктів у навчальній мережі, покращивши неаксіально симетричні ознаки. Гуо та ін. [19] впровадили представлення опуклої оболонки для навчання сприйняття форми та розподілу нерегулярних об'єктів. Також використовується адаптація ознак, що навчається, щоб уникнути змішування ознак. DRN використовує модуль вибору ознак для агрегування інформації про неаксіально симетричні об'єкти різних форм, напрямків та розмірів ядра. Він також використовує динамічний генератор фільтрів для регресії цієї інформації. Вищезазначені методи спрямовані на поліпшення ефективності виявлення неаксіально симетричних ознак.

Розподіл зразків для виявлення об'єктів

При встановленні порогу IoU більшість досліджень намагаються визначити поріг, який дозволяє вибрати позитивні зразки. Надійність навчальних зразків не гарантована через різні фактори, такі як шум.

Напрями об'єктів у зображеннях дистанційного зондування різноманітні і щільно розподілені. Тому вибір високоякісних зразків є важливим для навчання детекторів з урахуванням напрямку. Нещодавно для вирішення цієї проблеми було запропоновано кілька ілюстративних стратегій розподілу зразків.

ATSS, FreeAnchor, PAA, OTA та інші методи використовують стратегії оптимізації відповідностей для вибору найкращих зразків. Наприклад, Мінг та ін. [20] запропонували метод оцінки відповідності, що використовує втрати чутливості до відповідності для оцінки просторового вирівнювання об'єктів. Цей метод вимірювання може ефективно підвищити кореляцію між виявленням об'єктів з урахуванням напрямку та класифікацією. Усі вищезазначені техніки можуть оптимізувати розподіл зразків для виявлення об'єктів, що допомагає вирішити проблему нерегулярності об'єктів.

Виявлення малих об'єктів на зображеннях дистанційного зондування

Термін "малі об'єкти" не має загальноприйнятого визначення, однак існуючі академічні визначення можна умовно поділити на дві основні категорії. Перша категорія стосується порівняльного розміру. Відповідно до визначення, наданого Ченем та ін. [21], малі об'єкти характеризуються відносною площею, що знаходиться в межах від 0.08% до 0.58%. Ця відносна площа визначається шляхом обчислення медіани співвідношення між площею обмежувальної рамки та площею зображення для всіх об'єктів, що належать до однієї категорії. Друга категорія визначається абсолютним масштабом, що визначає розмір малих об'єктів на основі абсолютного розміру в пікселях. Загальний набір даних MS COCO для виявлення об'єктів визначає малий об'єкт як об'єкт з роздільною здатністю менше 32 пікселів \times 32 пікселів. Експериментальний набір даних DOTA в цій статті визначає об'єкти з піксельними значеннями в діапазоні [10, 50] як малі об'єкти. Проте через невелику кількість пікселів, малий масштаб та легкі для затінення тіні виявлення малих об'єктів все ще стикається зі значними викликами.

Багатошкільне прогнозування

Багатошкільне прогнозування передбачає визначення координат і класів об'єктів на різних масштабах. Раніше використовували піраміди зображень, але з розвитком глибокого навчання їх замінили CNN, як у SSD. Проте SSD неефективний для малих об'єктів, тому Deconvolutional Single Shot Detector (DSSD) і FPN інтегрують ознаки різних рівнів для покращення виявлення. PANet вдосконалює FPN, зберігаючи більше інформації на нижчих шарах. Adaptively Spatial Feature Fusion адаптує злиття ознак для покращення точності, а AugFPN зменшує семантичні відмінності між шарами. TridentNet вирішує проблему масштабування, а RTMDet вдосконалює одноетапні детектори, використовуючи CSPDarkNet і нові стратегії навчання.

Підвищена роздільна здатність ознак

Цей метод покращує точність виявлення, збільшуючи роздільну здатність карт ознак або перетворюючи ознаки малих об'єктів у представлення більших. STDN реалізує це через модуль трансферу масштабу. Генеративні змагальні мережі (GAN) на основі PGAN і SOD-MTGAN використовують генератор і дискримінатор, навчаючись на прикладах як малих, так і великих об'єктів для підвищення точності. ViTAE-V+RVSA_ORCN застосовує метод MAE для виділення ознак зображення, адаптуючи ViTAE для виявлення об'єктів у дистанційному зондуванні. Enhanced Super Resolution GAN (EESRGAN) покращує виявлення, зосереджуючись на краях зображення для підвищення ефективності.

Контекстна інформація

Метод використовує контекстуальні зв'язки між об'єктами на зображенні для підвищення точності виявлення малих об'єктів. Тан та ін. [22] застосували контекстуально асистовані методи з апіорними рамками для виявлення обличчів різних масштабів, де апіорні рамки служили додатковою інформацією для навчання. Ху та ін. [23] запропонували модуль, що використовує кореляції між об'єктами, об'єднуючи їх форму та геометричні ознаки. Чен та ін. [24] включили контекстну інформацію для опису взаємовідносин між об'єктами, що знижує ймовірність помилкової ідентифікації.

Поліпшення даних

Поліпшення даних збільшує кількість навчальних зразків шляхом панорамування, обертання та повторної вибірки. Кісантал і колеги [25] збільшили навчальні зразки, повторно вибираючи зображення, що містять малі об'єкти, і копіюючи та вставляючи ці малі об'єкти.

Зоф і колеги [26] застосували пошук нейронних архітектур (NAS) для збільшення даних, щоб знайти оптимальні стратегії покращення ефективності RetinaNet.

Нова основна мережа та стратегії навчання

Відмінності між наборами даних для класифікації та виявлення можуть ускладнювати виявлення малих об'єктів. Для вирішення цієї проблеми запропоновано нові основні мережі та стратегії навчання. DEA-net використовує динамічно поліпшену мережу анкерів для запобігання втраті або неправильній мітці малих об'єктів. GGHF підходить для виявлення об'єктів у будь-якому напрямку, застосовуючи адаптивну стратегію міток (OLA) на основі гауссової теплової карти. APE (Adaptive Period Embedding) представляє орієнтовані об'єкти за допомогою кутових векторів. CFC-Net захоплює ключові ознаки, поляризує модуль уваги і використовує динамічну стратегію для вибору анкерів.

Проблема розриву границі

Проблема розриву границі в певній мірі впливає на ефективність виявлення об'єктів. PP-YOLOE-R суттєво покращила розпізнавання обертання об'єктів. Вона вводить втрату ProbloU, щоб уникнути проблеми розриву границі. PP-YOLOE-R також використовує навчання з вирівнюванням завдань обертання для виявлення обертових об'єктів. Вона отримує більш точні прогнози кутів завдяки голові прогнозування кута та

втраті DFL.

Для вирішення проблеми розриву границі DCL розроблений, щоб замінити існуючі рідкокодовані мітки на щільно закодовані мітки (DCL). Це досягає значного покращення швидкості навчання та точності виявлення. Метод чутливого зважування кутової відстані та співвідношення сторін у DCL робить детектор більш чутливим до цих двох аспектів об'єкта. Це покращує продуктивність виявлення і робить DCL особливо придатним для виявлення квадратних об'єктів.

Порівняння та аналіз сучасних моделей

Експериментування та аналіз типових алгоритмів

Серія YOLO є прикладом регресійного алгоритму на основі глибокого навчання. Ці алгоритми стали важливими представниками методів виявлення об'єктів. Проте серія YOLO рідко використовується в сучасних дослідженнях, що стосуються тестування зображень дистанційного зондування. Тому ми вирішили обрати чотири моделі з серії YOLO для навчання на наборі даних DOTA, щоб оцінити їхню ефективність.

Результати наведено на рисунку 4. Проаналізувавши отримані дані, ми дійшли висновку, що YOLOv3 має найвищий середній показник точності (mAP), який становить 0,495, при 100 ітераціях навчання. Однак ми зовсім не очікували такого результату. Отже, традиційні моделі YOLO виявляють низьку ефективність при застосуванні до зображень дистанційного зондування.

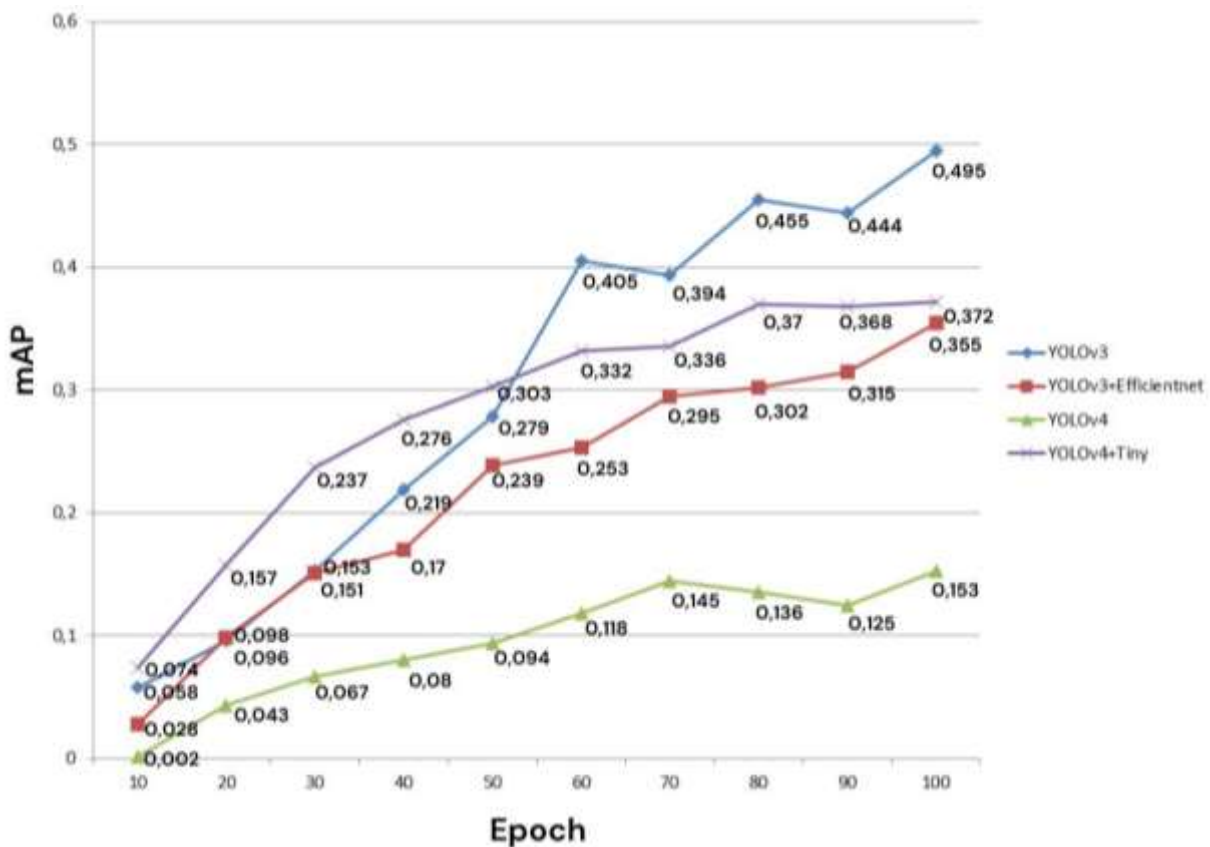


Рис. 4. Порівняння серії YOLO

Порівняння сучасних методів виявлення об'єктів на зображеннях дистанційного зондування

Щоб оцінити ефективність алгоритмів виявлення об'єктів на зображеннях дистанційного зондування, ми провели порівняння нещодавно розроблених високоєфективних алгоритмів на наборі даних DOTA. Результати представлені в таблиці 2.

Таблиця 2

Порівняння сучасних методів на наборі даних DOTA.

Methods	Backbone	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
One-stage Methods																	
R3Det-DCL	R-152-FPN	89,78	83,95	52,63	69,7	76,84	81,26	87,3	90,81	84,67	85,27	63,5	64,16	68,96	68,79	65,45	75,54
R3Det	R-152-FPN	89,49	81,17	50,53	66,1	70,92	78,66	78,21	90,81	85,26	84,23	61,81	63,77	68,16	69,83	67,17	73,74
S2A-Net	R-50-FPN	89,11	82,84	48,37	71,11	78,11	78,39	87,25	90,83	84,9	85,64	60,36	62,6	65,2	69,31	57,94	74,12
RetinaNet-O	R-50-FPN	88,67	77,62	41,81	58,71	74,58	71,64	79,11	90,29	82,18	74,32	54,75	60,6	62,57	69,67	60,64	68,43
RSDet	R-152-FPN	90,1	82	53,8	68,5	70,2	78,7	73,6	91,2	87,1	84,7	64,3	68,2	66,1	69,3	63,7	74,1
DAL	R-101-FPN	88,61	79,69	46,27	70,31	65,89	76,1	78,53	90,84	79,98	78,41	58,71	62,02	69,23	71,32	60,65	71,78
CFA	R-152	89,08	83,2	54,37	66,87	81,23	80,96	87,17	90,21	84,32	86,09	52,34	69,94	75,52	80,76	67,96	76,67
DAFNet	R-101	89,4	86,27	53,7	60,51	82,04	81,17	88,66	90,37	83,81	87,27	53,93	69,38	75,61	81,26	70,86	76,95
SASM	RX-101	89,54	85,94	57,73	78,41	79,78	84,19	89,25	90,87	88,12	87,27	63,82	67,81	78,67	79,35	69,37	79,17
AO2-DETR	R-50	89,95	84,52	56,9	74,83	80,86	83,47	88,47	90,87	86,12	88,55	63,24	65,09	79,09	82,88	73,46	79,22
Two-stage Methods																	
Oriented R-CNN	R-101-FPN	88,86	83,48	55,27	76,92	74,27	82,1	87,52	90,9	85,56	85,33	65,51	66,82	74,36	70,15	57,28	76,28
ReDet	ReR-50-ReFPN	88,79	82,64	53,97	74	78,1	84,06	88,04	90,89	87,78	85,75	61,76	60,39	75,96	68,07	63,59	76,25
CenterMap	R-50-FPN	88,88	81,24	53,15	60,65	78,62	66,55	78,1	88,83	77,8	83,61	49,36	66,19	72,1	72,36	58,7	71,74
MaskOBB	R-50-FPN	89,61	85,09	51,85	72,9	75,28	73,23	85,57	90,37	82,08	85,05	55,73	68,39	71,61	69,87	66,33	74,86
Gliding Vertex	R-101-FPN	89,64	85	52,26	77,34	73,01	73,14	86,82	90,74	79,02	86,81	59,55	70,91	72,64	70,86	57,32	75,02
RoI-Trans	R-101-FPN	88,65	82,6	52,53	70,87	77,93	76,67	86,87	90,71	83,83	82,51	53,95	67,61	74,67	68,75	61,03	74,61
FAOD	R-101-FPN	90,21	79,58	45,49	76,41	73,18	68,27	79,56	90,83	83,4	84,68	53,4	65,42	74,17	69,69	64,86	73,28
SCRDet	R-101-FPN	89,98	80,65	52,09	68,36	68,36	60,32	72,41	90,85	87,94	86,86	65,02	66,68	66,25	68,24	65,21	72,61
CAD-Net	R-101-FPN	87,8	82,4	49,4	73,5	71,1	63,5	76,6	90,9	79,2	73,3	48,4	60,9	62	67	62,2	69,9
Faster RCNN-O	R-50-FPN	88,44	73,06	44,86	59,09	73,25	71,49	77,11	90,84	78,94	83,9	48,59	62,95	62,18	64,91	56,18	69,5
CSL	R-152	90,25	85,53	54,64	75,31	70,44	73,51	77,62	90,84	86,15	86,69	69,6	68,04	73,83	71,1	68,93	76,17
DODet	R-50-FPN	89,96	85,52	58,01	81,22	78,71	85,46	88,59	90,89	87,12	87,8	70,5	71,54	82,06	77,43	74,47	80,62
AOPG	R-50-FPN	89,88	85,57	60,9	81,51	78,7	85,29	88,85	90,89	87,6	87,65	71,66	68,69	82,31	77,32	73,1	80,66
LSKNet-S *	LSKNet	89,69	85,7	61,47	83,23	81,37	86,05	88,64	90,88	88,49	87,4	71,67	71,35	79,19	81,77	80,86	81,85
LSKNet-S	LSKNet	89,57	86,34	63,13	83,67	82,2	86,1	88,66	90,89	88,41	87,42	71,72	69,58	78,88	81,77	76,52	81,64
Anchor-free Methods																	
Oriented RepPoints	R-50-FPN	87,02	83,17	54,13	71,16	80,81	78,4	87,28	90,9	85,97	86,25	59,9	70,49	73,53	72,27	58,97	75,97
Oriented RepPoints	R-101-FPN	89,53	84,07	59,86	71,76	79,95	80,03	87,33	90,84	87,54	85,23	59,15	66,37	75,23	73,75	57,23	76,52
Oriented RepPoints	Swin-T-FPN	89,11	82,32	56,71	74,95	80,7	83,73	87,67	90,81	87,11	85,85	63,6	68,6	75,95	73,54	63,76	77,63
DRN [73]	H-104	89,71	82,34	47,22	64,1	76,22	74,43	85,84	90,57	86,18	84,89	57,65	61,93	69,3	69,63	58,48	73,23
PloU	DLA-34	80,9	69,7	24,1	60,2	38,3	64,4	64,8	90,9	77,2	70,4	46,5	37,1	57,1	61,9	64	60,5
CenterNet-O	DLA-34	81	64	22,6	56,6	38,6	64	64,9	90,8	78	72,5	44	41,1	55,5	55	57,4	59,1

У цьому документі ми додатково класифікували набір даних DOTA відповідно до категорій об'єктів. Великі об'єкти включають бейсбольні майданчики, тенісні корти, баскетбольні майданчики, легкоатлетичні стадіони, кругові роз'язки та футбольні поля. Середні об'єкти включають літаки, резервуари для зберігання, порти, мости та басейни. Малі об'єкти складаються з великих транспортних засобів, малих транспортних засобів, вертольотів і кораблів. Значення середнього показника точності (mAP) для кожного методу на наборі даних DOTA для об'єктів різних розмірів представлені в таблиці 3.

Таблиця 3

Середня точність для різних типів об'єктів на наборі даних DOTA.

Methods	Backbone	mAP	APLAPL	APMAPM	APSAPS
One-stage Methods					
R3Det-DCL	R-152-FPN	75,54	76,13	73,09	73,09
R3Det	R-152-FPN	73,74	74,82	72,45	71,65
S2A-Net	R-50-FPN	74,12	75,44	71,53	70,94
RetinaNet-O	R-50-FPN	68,43	70,69	67,41	69,13
RSDet	R-152-FPN	74,1	76,88	72,8	70,48
DAL	R-101-FPN	71,78	73,59	70,77	68,49
CFA	R-152	76,67	74,48	77,16	77,73
DAFNet	R-101	76,95	74,05	77,45	78,83
SASM	RX-101	79,17	74,28	78,51	78,17
AO2-DETR	R-50	79,22	77,45	79,47	80,16
Two-stage Methods					
Oriented R-CNN	R-101-FPN	76,28	78,2	74,79	70,95
ReDet	ReR-50-ReFPN	76,25	76,24	74,51	73,46
CenterMap	R-50-FPN	71,74	70,68	74,02	69,06
MaskOBB	R-50-FPN	74,86	75,76	73,6	71,18
Gliding Vertex	R-101-FPN	75,02	77,09	74,44	68,58
RoI-Trans	R-101-FPN	74,61	74,93	73,42	74,1
FAOD	R-101-FPN	73,28	74,84	72,85	69
SCRDet	R-101-FPN	72,61	76,58	72,68	65,53
CAD-Net	R-101-FPN	69,9	72,55	85,46	65,95
Faster RCNN-O	R-50-FPN	69,5	68,91	68,86	66,46

CSL	R-152	76,17	79,21	75,3	72,22
DODet	R-50-FPN	80,62	81,13	79,05	79,02
AOPG	R-50-FPN	80,66	80,99	79,61	78,6
LSKNet-S *	LSKNet	81,85	81,89	79,9	82,5
LSKNet-S	LSKNet	81,64	81,77	80,15	81,65
Anchor-free Methods					
Oriented RepPoints	R-50-FPN	75,97	76,93	74,64	72,61
Oriented RepPoints	R-101-FPN	76,52	76,62	76,72	72,74
Oriented RepPoints	Swin-T-FPN	77,63	77,9	76,23	75,43
DRN	H-104	73,23	73,8	72,15	69,69
PloU	DLA-34	60,5	63,6	58,88	57,15
CenterNet-O	DLA-34	59,1	62,42	57,32	53,75

Як показано в таблиці 3, серед моделей першого етапу, mAP, APL, APM та APS AO2-DETR є найвищими. Серед детекторів другого етапу, LSKNet-S* має найвищі APL та APS, а його продуктивність виявляється найкращою для великих та малих об'єктів. APM CAD-Net також є найвищим, що свідчить про його найкращу продуктивність для середніх об'єктів. Загалом, LSKNet-S* демонструє найвищий mAP серед детекторів другого етапу, досягаючи 81,85%. У категорії безякорних методів, Oriented RepPoints із мережею Swin-T-FPN має найкращу продуктивність для великих та малих об'єктів, з високими APL та APS на рівні 77,90% та 75,43% відповідно. Щодо середніх об'єктів, Oriented RepPoints із R-101-FPN продемонстрували найкращі результати з APM, що досягає 76,72%. Таким чином, Oriented RepPoints із мережею Swin-T-FPN мають найвищий mAP та найкращі результати. Порівнюючи всі три типи методів, можна помітити, що LSKNet-S* має вищі значення mAP, APL, APM та APS у порівнянні з AO2-DETR та Oriented RepPoints на базі R-101-FPN. Таким чином, LSKNet-S* демонструє найкращу продуктивність в об'єктному детекції.

Ось деякі висновки, які можна зробити на основі порівняння:

Наразі Resnet-FPN слугує основною мережею для більшості методів об'єктного детекції, що застосовуються до дистанційно-знятих зображень. Аналізуючи їх mAP, можна відзначити, що продуктивність цих методів є стабільнішою, тобто на середньому рівні. Загальний рівень продуктивності першого методу кращий, ніж у другого, при порівнянні методів на базі якорів і безякорних методів. Нещодавно запропонована основна мережа LSKNet демонструє значні переваги на наборі даних DOTA, включаючи різні категорії точності (AP) та mAP.

Аналізуючи вищезгадану інформацію, можна зробити висновок, що, з одного боку, методи об'єктного детекції на дистанційно-знятих зображеннях постійно еволюціонують та покращують свою продуктивність. З іншого боку, пропозиція нових основних мереж суттєво допомагає покращити продуктивність об'єктної детекції. Отже, дизайн основних мереж може стати основним напрямком майбутніх досліджень.

Актуальні Виклики та Майбутні Напрямки

Незважаючи на значний прогрес, область розпізнавання об'єктів на зображеннях все ще стикається з численними проблемами та викликами. Дистанційно-зняті зображення створюють унікальні труднощі для застосувань, особливо в порівнянні з природними зображеннями. Серед цих викликів—різноманітні сценарії застосування, велика кількість об'єктів та різні напрямки. Крім того, зовнішні фактори, такі як погодні умови та освітлення, також можуть істотно вплинути на дистанційно-зняті зображення. Цей розділ окреслює кілька перспективних підходів для покращення застосувань ідентифікації об'єктів, зокрема в контексті ідентифікації малих об'єктів на дистанційно-знятих зображеннях.

Обробка Зображень

Через великі витрати на отримання високоякісних дистанційно-знятих зображень великої площі та значні капіталовкладення, що споживаються в практичних застосуваннях, виникає потреба в нових схемах для виявлення малих об'єктів на зображеннях з низькою роздільною здатністю. Для вирішення цієї проблеми технологія супервисокої роздільної здатності мають на меті реконструювати зображення з низькою роздільною здатністю у високоякісні зображення. Це дозволяє зменшити вплив зовнішніх факторів, таких як навколишнє середовище, на дистанційно-зняті зображення. У порівнянні з зображеннями з низькою роздільною здатністю, високоякісні зображення містять більше характеристик об'єктів, що сприяє покращенню продуктивності моделей об'єктного детекції.

Стратегія Навчання

Необхідно дослідити найбільш підходящу функцію втрат для виявлення малих об'єктів на дистанційно-знятих зображеннях. Нормалізація пакетів широко використовується в галузі об'єктного детекції для прискорення навчання моделі. Проте вплив згаданої технології на виявлення малих об'єктів на дистанційно-знятих зображеннях потребує покращення. Потрібно провести подальше дослідження, щоб визначити відповідні техніки нормалізації для завдань виявлення малих об'єктів у дистанційних зображеннях.

Проектування Мережі

Відмінні мережеві структури не лише забезпечують вищі оціночні показники, але й потребують менше часу та обчислювальних ресурсів для ефективного навчання. Тому більш ефективні та легкі мережі

залишаються важливим напрямом досліджень. Однак багато сучасних моделей глибоких нейронних мереж демонструють відмінні результати на еталонних наборах даних. Проте мережеві структури для виявлення малих об'єктів досі є дуже обмеженими. Тому існує потреба в розробці нових об'єктних, легких та ефективних архітектур мереж для покращення продуктивності виявлення.

Побудова Набору Даних

Дистанційно-зняті зображення підлягають вродженим обмеженням і вимагають значних витрат часу та ресурсів для створення в межах набору даних. У порівнянні з природними зображеннями, обсяг даних цих зображень обмежений. Дистанційно-зняті зображення різних регіонів і сцен демонструють помітні відмінності через вплив рельєфу та рослинності. Це призводить до обмеженої репрезентативності та узагальненості набору даних. Генерація дистанційно-знятих зображень залишається значною перешкодою.

Висновки

У цьому дослідженні надано детальний огляд методів виявлення об'єктів, зокрема малих об'єктів. Обговорюються загальні набори даних, методи оцінки, критерії класифікації, а також проблеми дистанційного зондування, включаючи труднощі з виявленням нерегулярних об'єктів. Також розглянуто різні приклади застосування методів виявлення об'єктів у зображеннях дистанційного зондування.

Незважаючи на значний прогрес у цій сфері, залишаються виклики, такі як низька продуктивність моделей і незадовільні результати виявлення. Рекомендовано перспективні напрями досліджень, зокрема вдосконалення обробки зображень, створення більш ефективних базових мереж та оптимізація стратегій навчання.

Ця стаття сприятиме глибшому розумінню методів виявлення об'єктів, особливо застосування методів виявлення малих об'єктів у дистанційно-знятих зображеннях. Очікується, що це сприятиме розвитку та прогресу технологій дистанційного зображення.

References

1. Li, K.. Object detection in optical remote sensing images: A survey and a new benchmark / Wan, G., Cheng, G., Meng, L., Han, J // *ISPRS J. Photogramm. Remote Sens.* – 2020 – № 159 – P. 296–307.
2. Long, Y. Accurate object localization in remote sensing images based on convolutional neural networks. / Gong, Y., Xiao, Z., Liu, Q // *IEEE Trans. Geosci. Remote Sens.* – 2017 – № 55 – P. 2486–2498.
3. Cheng, G. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. / Zhou, P., Han, J // *IEEE Trans. Geosci. Remote Sens.* – 2016 – №54 – P. 7405–7415.
4. Xia, G.S. DOTA: A large-scale dataset for object detection in aerial images. / Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M.; Zhang, L. // *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018* – P. 3974–3983.
5. Razakarivony, S. Vehicle detection in aerial imagery: A small target detection benchmark. / Jurie, F. // *J. Vis. Commun. Image Represent.* – 2016 – №34 – P. 187–203.
6. Yang, M.Y. Deep learning for vehicle detection in aerial images. / Liao, W., Li, X., Rosenhahn, B. // *In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*. – 2018 – P. 3079–3083.
7. Lin, T.Y. Microsoft coco: Common objects in context. / Maire, M., Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. // *In Proceedings of the Computer Vision—ECCV – 2014 – № 13* – P. 740–755.
8. Zhu, H. Orientation robust object detection in aerial images using deep convolutional neural network. / Chen, X., Dai, W., Fu, K., Ye, Q., Jiao, J. // *In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP)*. – 2015 – P. 3735–3739.
9. Zhao, L. Feature significance-based multibag-of-visual-words model for remote sensing image scene classification. / Tang, P., Huo, L. // *J. Appl. Remote Sens.* – 2016 – №10 – 035004.
10. Wu, X. Unpaired learning of deep image denoising. / Liu, M., Cao, Y., Ren, D., Zuo, W. // *In Proceedings of the Computer Vision—ECCV – 2020 – № 4* – P. 352–368.
11. He, K. Guided image filtering. / Sun, J., Tang, X. // *IEEE Trans. Pattern Anal. Mach. Intell.* – 2012 – № 35 – P. 1397–1409.
12. Lev, B. Sharpening the intangibles edge. // *Harv. Bus. Rev.* – 2004 – № 6 – P. 109–116.
13. Lin, C.Y. Rotation, scale, and translation resilient watermarking for images. / Wu, M., Bloom, J.A., Cox, I.J., Miller, M.L., Lui, Y.M. // *IEEE Trans. Image Process.* – 2001 – №10 – P. 767–782.
14. Lin, X. A survey for image resizing. / Ma, Y.I., Ma, L.z., Zhang, R.I. // *J. Zhejiang Univ. Sci.* – 2014 – №15 – P. 697–716.
15. Dhawan, S. A review of image compression and comparison of its algorithms. *Int. J. Electron. Commun. Technol.* – 2011 – №2 – P. 22–26.
16. Yang, X. Arbitrary-Oriented Object Detection with Circular Smooth Label. / Yan, J. // *In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August – 2020*.
17. Han, J. Align Deep Features for Oriented Object Detection. / Ding, J., Li, J., Xia, G. // *IEEE Trans. Geosci. Remote Sens.* – 2020 – №60 – 5602511.
18. Ding, J. Learning RoI Transformer for Oriented Object Detection in Aerial Images. / Xue, N.; Long, Y.; Xia, G.; Lu, Q. // *In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) – 2019* – P. 2844–2853.

-
19. Guo, Z.. Beyond Bounding-Box: Convex-hull Feature Adaptation for Oriented and Densely Packed Object Detection. / Liu, C.; Zhang, X., Jiao, J., Ji, X., Ye, Q // In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) – 2021 – P. 8788–8797.
 20. Ming, Q. Dynamic Anchor Learning for Arbitrary-Oriented Object Detection. / Zhou, Z.; Miao, L.; Zhang, H.; Li, L. // In: Proceedings of the AAAI conference on artificial intelligence. – 2021 – P. 2355-2363.
 21. Chen, C. R-CNN for small object detection. / Liu, M.Y., Tuzel, O., Xiao, J. // In Proceedings of the Computer Vision–ACCV 2016: Revised Selected Papers. – 2017 – №13 – P. 214–230.
 22. Tang, X.. PyramidBox: A Context-assisted Single Shot Face Detector. / Du, D.K., He, Z., Liu, J // In. Proceedings of the European conference on computer vision (ECCV). – 2018 – P. 797-813.
 23. Hu, H. Relation Networks for Object Detection. / Gu, J., Zhang, Z., Dai, J., Wei, Y. // In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2018 – P. 3588–3597.
 24. Chen, X. Spatial Memory for Context Reasoning in Object Detection. / Gupta, A.K. // In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). – 2017 – P. 4106–4116.
 25. Kisantal, M. Augmentation for small object detection. / Wojna, Z., Murawski, J., Naruniec, J., Cho, K. // arXiv – 2019 – arXiv:1902.07296.
 26. Zoph, B. Learning Data Augmentation Strategies for Object Detection. / Cubuk, E.D., Ghiasi, G., Lin, T.Y., Shlens, J., Le, Q.V. // In Proceedings of the European Conference on Computer Vision. – 2020 – №16 – P. 566-583