

МАРЧУК ДМИТРО

Державний університет «Житомирська політехніка»

ORCID ID: [0000-0001-8675-8047](https://orcid.org/0000-0001-8675-8047)e-mail: kipz_mdk@ztu.edu.ua

АНАЛІЗ СУЧАСНИХ АЛГОРИТМІВ ВИЯВЛЕННЯ І РОЗПІЗНАВАННЯ ОБ'ЄКТІВ З ВІДЕОПОТОКУ ДЛЯ СИСТЕМ УПРАВЛІННЯ ПАРКУВАННЯМ В РЕАЛЬНОМУ ЧАСІ

Одною зі сфер штучного інтелекту є комп'ютерний зір, який використовує глибоке навчання для виявлення, розпізнавання, класифікації об'єктів на зображеннях та відео. Для більш ефективного застосування таких систем часто використовуються методи, засновані на нейронних мережах. Розвиток технологій глибокого навчання дозволив створювати більш точні та складні моделі комп'ютерного зору. Методи глибокого навчання, які використовуються для розпізнавання об'єктів на відео, можуть включати пропозицію регіону як частину системи або використовувати нерегіональні методи, які базуються на пропозиціях детектору. У статті розглядаються сучасні, найбільш відомі алгоритми розпізнавання об'єктів на відео. Наводиться опис особливостей різних архітектурних рішень нейронних мереж. Дослідження публікацій, присвячених проблемам аналізу відеоданих, вказав на пріоритет застосування алгоритмів на основі архітектури згорткової нейронної мережі. В роботі більше уваги приділяється таким архітектурним рішенням як YOLO і Mask R-CNN. Порівнюються продуктивність, швидкість обробки і точність. Результати проведеного дослідження показують, що YOLO одна з найсучасніших систем виявлення об'єктів у реальному часі, яка обробляє зображення зі швидкістю від 45 до 150 кадрів в секунду і має mAP 63,4% на тестовому наборі MS-COCO, а наприклад, Mini-YOLOv3 досягає mAP 52,1% зі швидкістю 67 кадрів в секунду. Але, якщо порівнювати різні версії і модифікації YOLO з іншими системами, то можна з впевненістю сказати, що YOLO допускає більше помилок локалізації. Mask R-CNN є розширенням Faster R-CNN, де паралельно проходить прогнозування маски об'єкта і розпізнавання обмежувальної рамки. Таким чином було виявлено, що для системи управління паркуванням, яка може відстежувати вільні місця на парковці з відеопотоку камери, найкраще підходить Mask R-CNN. Дана нейронна мережа має ряд переваг в порівнянні з такими, як R-CNN, Fast R-CNN та YOLO. Основною з її переваг є продуктивність та точність.

Ключові слова: нейронна мережа, архітектура, YOLO, CNN, R-CNN, Mask R-CNN.

MARCHUK DMYTRO

Zhytomyr Polytechnic State University

ANALYSIS OF MODERN ALGORITHMS FOR DETECTING AND RECOGNIZING OBJECTS FROM A VIDEO STREAM FOR REAL-TIME PARKING MANAGEMENT SYSTEMS

One of the areas of artificial intelligence is computer vision, which uses deep learning to detect, recognize, and classify objects in images and videos. To make such systems more efficient, methods based on neural networks are often used. The development of deep learning technologies has made it possible to create more accurate and complex computer vision models. Deep learning methods used to recognize objects in video can include a region proposal as part of the system or use non-regional methods based on detector proposals. The article discusses modern, best-known algorithms for object recognition in video. The features of different architectural solutions of neural networks are described. A study of publications on the problems of video data analysis has indicated the priority of using algorithms based on the convolutional neural network architecture. The paper pays more attention to such architectural solutions as YOLO and Mask R-CNN. Performance, processing speed, and accuracy are compared. The results of the study show that YOLO is one of the most advanced real-time object detection systems that processes images at a speed of 45 to 150 frames per second and has an mAP of 63.4% on the MS-COCO test set, and, for example, Mini-YOLOv3 reaches an mAP of 52.1% at 67 frames per second. However, if we compare different versions and modifications of YOLO with other systems, we can say with certainty that YOLO makes more localization errors. Mask R-CNN is an extension of Faster R-CNN, where the object mask prediction and bounding box recognition are performed in parallel. Thus, it was found that Mask R-CNN is best suited for a parking management system that can track free parking spaces from a camera video stream. This neural network has a number of advantages compared to R-CNN, Fast R-CNN, and YOLO. The main advantages of Mask R-CNN are performance and accuracy.

Keywords: neural network, architecture, YOLO, CNN, R-CNN, Mask R-CNN.

Постановка проблеми

На сьогоднішній день широко і успішно використовуються різні системи працюючі на основі комп'ютерного зору, які дозволяють отримувати інформацію з цифрових зображень, відео та інших візуальних вхідних даних і вживати дії або давати рекомендації на основі отриманої інформації. Такі системи використовуються в різних галузях промисловості, від енергетики та комунальних послуг до виробництва та автомобілебудування, торгівлі, охороні здоров'я, для побудови траєкторії між двома заданими точками призначення з можливістю обходу перешкод [1] і ринок застосування продовжує постійно зростати.

За останні роки досягнення в алгоритмах комп'ютерного зору на базі глибокого навчання показали багатообіцяючі результати в різноманітних завданнях. Подібні методи можна використовувати для вирішення проблеми виявлення вільного місця на автостоянках або зонах для паркування. Проблема знаходження місць для паркування є досить поширеною в сучасному світі, особливо в великих містах. Це може мати різні причини, серед яких:

- Недостатня кількість місць для паркування на автостоянці. Існує дисбаланс між попитом на паркування та пропозицією. Це можуть бути старі норми планування міста, кварталу, вулиці, де було непередбачено великої кількості машин у містян.
- Невідповідність розміру автомобілів та розміру місць для паркування. Деякі місця можуть бути занадто маленькими для сучасних автомобілів, що ускладнює паркування і може призводити до пошкоджень автомобілів.
- Недостатня організація місць для паркування. Місця можуть бути погано позначені, розташовані у незручних місцях або не підтримуються в належному стані, що ускладнює пошук вільного місця для паркування.
- Висока вартість паркування. В такому випадку водієві приходится шукати місце на безкоштовній стоянці або у сусідньому дворі, що приводить до дисбалансу місць для паркування.
- Недостатній контроль за паркуванням. Автомобіль може бути припарковано неправильно, що може ускладнювати рух іншого автотранспорту.

На сьогодні система управління паркуванням, яка може відстежувати вільні місця, стала необхідністю для майже всіх великих міст. Система надасть можливість користувачу легко та швидко визначити, чи має конкретна зона для паркування вільне місце та зорієнтуватися в якій частині це місце розташоване.

Аналіз останніх досліджень і публікацій

Останнім часом технології пов'язані з комп'ютерним баченням привернули значну увагу дослідників. Традиційні методи виявлення об'єктів, які базувалися на створених вручну функціях та неглибоких архітектурах, вже застаріли і не задовольняють поточним вимогам. Дослідження [2] містить детальний огляд фреймворків виявлення об'єктів на основі глибокого навчання, які вирішують різні проблеми, такі як оклюзія, перешкоди та низька роздільна здатність, з різними ступенями модифікацій на R-CNN (Region Based Convolutional Neural Networks).

Шляхом внесення численних змін алгоритми виявлення об'єктів були покращені з точки зору швидкості і точності. Завдяки невпинним зусиллям великої кількості дослідників алгоритми глибокого навчання швидко розвиваються, пропонуючи вдосконалену продуктивність у сфері виявлення об'єктів. Програми, такі як розпізнавання пішоходів, медична візуалізація, робототехніка, безпілотні автомобілі, розпізнавання обличчя та інші, допомагають економити людські ресурси у багатьох галузях. Стаття [3] представляє фундаментальний огляд методів виявлення об'єктів, включаючи два класи детекторів об'єктів. У двоступеневому детекторі розглядаються алгоритми R-CNN, Fast R-CNN і Faster R-CNN, тоді як в одноступінчастому детекторі - YOLO v1, v2, v3 і SSD. Двоступеневі детектори більше зосереджені на точності, тоді як головною перевагою одноступеневих детекторів є швидкість. Автори статті представляють вдосконалену версію YOLO під назвою YOLO v3-Tiny і порівнюють її з попередніми версіями.

Процес виявлення об'єктів в реальному часі на відеозаписі, запущеного на вбудованому пристрої, дуже складний через обмежену пам'ять і обчислювальну потужність вбудованих пристроїв. Щоб впоратися з цими проблемами, автори статті [4] вдосконалили спрощену мережу без значного зниження точності виявлення об'єктів. Ці-Чао Мао та інші пропонують модель виявлення об'єктів під назвою Mini-YOLOv3, яка має менший розмір моделі та менше параметрів, які використовуються при навчанні порівняно з YOLOv3. Модель навчається на базі даних MS-COCO. Розмір параметра Mini-YOLOv3 становить лише 23% від YOLOv3 і забезпечує майже таку точність виявлення як і YOLOv3. Зокрема, Mini-YOLOv3 досягає mAP-50 52,1 зі швидкістю 67 кадрів в секунду.

Автори статті [5] представляють вдосконалену версію YOLOv3 під назвою F1% Light-YOLOv3. Коли робот збирає зелений плід при природному освітленні, колір фрукта схожий на колір фону; нерівномірне освітлення та оклюзія плодів і листя часто впливають на продуктивність методу виявлення. Дослідниками пропонується допоміжний алгоритм покращення зображення зеленого манго (CLANE-Mango), який підходить для середовищ виявлення низької яскравості та покращує точність методу виявлення зеленого манго. Результати експерименту показують, що F1% Light-YOLOv3 у тестовому наборі становить 97,7%. Порівняно з YOLOv3, F1% Light-YOLOv3 збільшено на 4,5%, а швидкість обробки збільшена в 5 разів, що може відповідати вимогам до роботи в режимі реального часу на вбудованій платформі.

У статті [6] розроблено та впроваджено техніку відкидання для алгоритмів на основі YOLOv3 для підвищення швидкості та підтримки точності. Після застосування техніки відкидання YOLOv3 може досягти 22% покращення швидкості. Більше того, результати цієї нової техніки відкидання було протестовано на Tiny-YOLOv3 з трьома вихідними шарами на автономному транспортному засобі для виявлення пішоходів, і вона досягла покращення швидкості на 48,7%.

Джозеф Редмон та інші [7] представили новий підхід до виявлення об'єктів на основі мережі YOLO. Уніфікована архітектура, що була розроблена науковцями, надзвичайно швидка. Fast YOLO, обробляє приголомшливі 155 кадрів на секунду, при цьому досягаючи вдвічі більшої mAP, ніж інші детектори реального часу. Порівняно з іншими системами виявлення, YOLO допускає більше помилок локалізації, але з набагато меншою ймовірністю передбачає помилкові виявлення там, де нічого не існує.

Розпізнавання зображень за кількома мітками набуло значення в останні роки через збільшення використання в багатьох областях. У проекті [8] автори пропонують гібридний метод, який реалізує підхід,

– Блок згорткового шару + пакетна нормалізація + функція активації (Convolutional Layer + Batch Normalization + Leaky ReLU): призначений для додавання нелінійності та підсилення розпізнавання об'єктів.

– Повнозв'язний шар (Fully Connected Layer): призначений для зведення вихідних значень з попередніх шарів до фінального розміру та отримання вектору ознак.

– Блок векторної розмітки (Vectorization Layer): призначений для перетворення вектора ознак в вектор розмітки, який містить інформацію про розташування та клас об'єкта.

– Блок об'єднання результируючих векторів (Anchors Layer): призначений для об'єднання результируючих векторів розмітки в один вектор, який містить інформацію про всі об'єкти на зображенні.

– Далі йде п'ять повторювальних блоків, кожен з яких складається зі свого власного детектора об'єктів, який працює на основі вихідних даних з передньої частини мережі. Кожен детектор об'єктів використовує фільтри-якорі для прогнозування координат обмежувачої рамки і ймовірності класів для кожної рамки. Якщо ймовірність перевищує поріг, то рамка пов'язується з даними класу.

YOLO використовує функцію втрати суми квадратів помилок, яку легко оптимізувати. Однак ця функція надає однакову вагу завданням класифікації та локалізації.

$$\begin{aligned} & \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ & + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\ & + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (C_i - \hat{C}_i)^2 \\ & + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{S^2} I_i^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2 \end{aligned} \tag{1}$$

де I_{ij}^{obj} - позначає, чи присутній об'єкт у клітинці i ,

I_{ij}^{noobj} - позначає j_{th} обмежувальну рамку, відповідальну за передбачення об'єкта в клітинці i ,

λ_{coord} і λ_{noobj} - це параметри регуляризації, які необхідні для балансування функції втрат.

Основна ідея YOLO полягає в тому, щоб розділити зображення на сітку, складаючи її з комірок, які

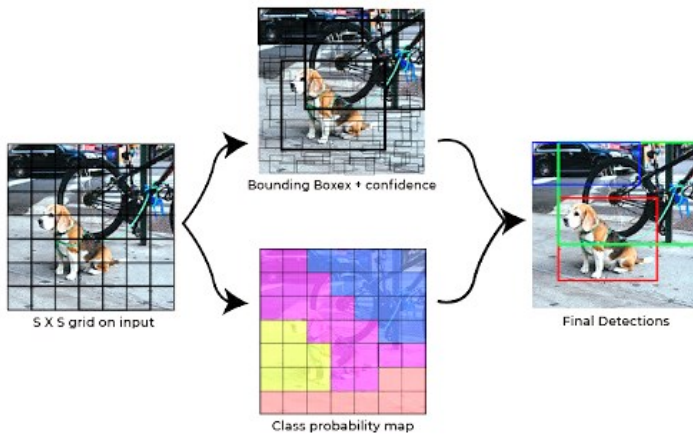


Рис. 2. Приклад накладання сітки на зображення

накладаються на зображення. Кожна комірка відповідає деякому регіону на зображенні, але вона також може бути причетною до декількох об'єктів. Якщо центр обмежувальної рамки об'єкта знаходиться в цій сітці, то ця сітка відповідає за виявлення цього об'єкта. Кожен об'єкт позначається як bounding box з деякими характеристиками, такими як координати кутів, розміри, клас об'єкта тощо(рис. 2).

Також кожна комірка сітки передбачає умовну ймовірність класу об'єкту $P_r(Class_i | Object)$. Ймовірність залежать від комірки сітки, яка містить

передбачає умовну ймовірність класу об'єкту $P_r(Class_i | Object)$. Ймовірність залежать від комірки сітки, яка містить

оцінку надійності для кожного класу можна визначити наступним чином [7]:

$$P_r(Class_i | Object) * P_r(Object) * IOU_{pred}^{truth} = P_r(Object) * IOU_{pred}^{truth} \tag{2}$$

Для визначення bounding box та класів об'єктів, YOLO використовує останній блок CNN, який з'єднує розміри зображення з вимірами сітки та використовує цю інформацію для прогнозування характеристик bounding box та класів об'єктів. Завдяки цій архітектурі YOLO здатна визначити об'єкти на зображеннях та відео в реальному часі, з високою точністю та відносно невисокими обчислювальними

витратами. Однак, через те, що YOLO розділяє зображення на сітку комірок, вона може мати деяку схильність до помилок в обробці деяких об'єктів, а також відносно невеликої кількості об'єктів, розташованих дуже близько один до одного (рис. 2).

Для навчання YOLO використовують функцію втрат, яка обчислює різницю між прогнозованими та дійсними значеннями. Ця функція включає в себе складові, які стимулюють мережу прогнозувати об'єкти різних розмірів та забезпечувати більш точне позиціонування обмежуючих рамок.

Переваги:

- Обробляє кадри зі швидкістю від 45 кадрів/с до 150 кадрів/с.

- Краще знаходить рамки зображення.

Недоліки:

- Досить низький рівень запам'ятовування відокремлених об'єктів.

- Багато помилок локалізації порівняно з Faster R-CNN.

- Важко виявити близькі об'єкти, оскільки кожна сітка може запропонувати лише дві обмежувальні рамки.

- Погано виявляє дрібні предмети.

YOLO є ефективною для задачі об'єктного детектування, оскільки вона дозволяє швидко та точно прогнозувати обмежуючі рамки та класи об'єктів на зображеннях і відео. Вона також має високу швидкість роботи та низькі вимоги до обчислювальних ресурсів порівняно з деякими видами моделей.

Один із підходів до виявлення і розпізнавання об'єктів є об'єднання регіонів зі згортковими нейронними мережами (Регіональна згорткова нейронна мережа, Region-based Convolutional Neural Networks (R-CNN)). R-CNN – це двоетапний алгоритм виявлення об'єктів. Перший етап визначає підмножину областей зображення, які можуть містити об'єкт розпізнавання. Наприклад, взяти вхідне зображення та створити набір обмежувальних рамок як вихід, де кожна обмежувальна рамка містить об'єкт, а також категорію об'єкта (наприклад, автомобіль чи людина). Другий етап класифікує об'єкт у кожному регіоні.

Модель R-CNN була створена щоб обійти проблему вибору величезної кількості регіонів. Росс Гіршик та ін. [11] запропонували метод, за допомогою якого можна використати вибіркового пошуку (selective search), щоб виділити лише 2000 регіонів із зображення. Тому тепер замість того, щоб намагатися класифікувати величезну кількість регіонів, можна просто працювати з 2000 регіонами (рис. 4).

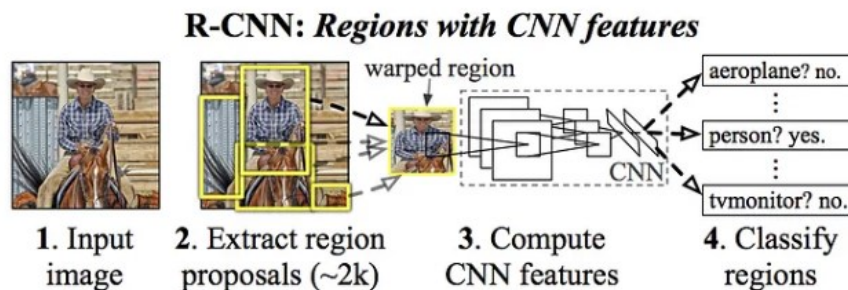


Рис. 4. Модель R-CNN

Пропозиції регіонів створюються за допомогою алгоритму вибіркового пошуку наступним чином:

1. Створюється початкова сегментація. Генерується багато регіонів-кандидатів на основі кольору, текстури, розміру і форми.

2. Використовується жадібний алгоритм для рекурсивного об'єднання подібних регіонів.

3. Регіони, які були згенеровано, використовуються для створення остаточних пропозицій регіонів-кандидатів.

Вилучення ознак кожного отриманого регіону здійснюється за допомогою згорткової нейронної мережі, а класифікація опрацьованих ознак за допомогою методу опорних векторів (Вилучення ознак кожного отриманого регіону за допомогою згорткових нейронних мереж. Класифікація опрацьованих ознак за допомогою методу опорних векторів (SVM, Support Vector Machine) та уточнення меж регіонів за допомогою лінійної регресії.), а уточнення рамок регіонів за допомогою лінійної регресії.

Але ця модель має більше недоліків чим переваг. Навчання мережі займає величезну кількість часу, оскільки доводиться класифікувати 2000 регіонів на зображенні. Неможливість реалізації у реальному часі, оскільки для кожного тестового зображення потрібно близько 47 секунд. Алгоритм вибіркового пошуку є фіксованим алгоритмом. Тому навчання на деякому етапі припиняється. Це може призвести до створення поганих пропозицій регіонів-кандидатів, які в подальшому повинні бути класифіковані. Тому R-CNN було розширено для виконання завдань комп'ютерного зору.

Архітектура Faster R-CNN складається з RPN як алгоритму пропозиції регіону та Fast R-CNN як мережі детектора. Faster R-CNN, модель розроблена групою дослідників Microsoft[12], і представляє собою

глибоку згорткову нейронну мережу, яка використовується для виявлення об'єктів. Мережа може точно і швидко передбачити розташування різних об'єктів. Швидкість досягається завдяки регіональній мережі пропозицій (RPN ,Region Proposal Network). RPN і Faster R-CNN можна об'єднати в єдину мережу, скориставшись їхніми функціями згортки, RPN буде повідомляти об'єднаній мережі, де шукати об'єкт.

Як вхідні дані RPN приймає карту об'єктів згортки і виводить прив'язки, згенеровані згорткою ковзного вікна, застосованою до карти вхідних об'єктів. Для кожного ковзного вікна мережа генерує максимальну кількість k-прив'язок. За замовчуванням значення k=9 (3 масштаби (128*128, 256*256 і 512*512) і 3 співвідношення сторін (1:1, 1:2 і 2:1)) для кожного з різних положень об'єкта на зображенні. Отже, для карти згортки W * H можна отримати N = W * H * k опорних блоків. Ці пропозиції областей потім передаються на проміжний рівень згортки 3*3 і 1 заповнення та 256 вихідних каналів. Вихідні дані, отримані з цього шару, передаються на два шари згортки 1*1, рівень класифікації та рівень регресії. Рівень регресії має 4*N (W * H * (4*k)) вихідних параметрів (що позначають координати обмежувальних рамок), а рівень класифікації має 2*N (W * H * (2*k)) вихідних параметрів (позначаючи ймовірність об'єкта чи не об'єкта) (рис. 5).

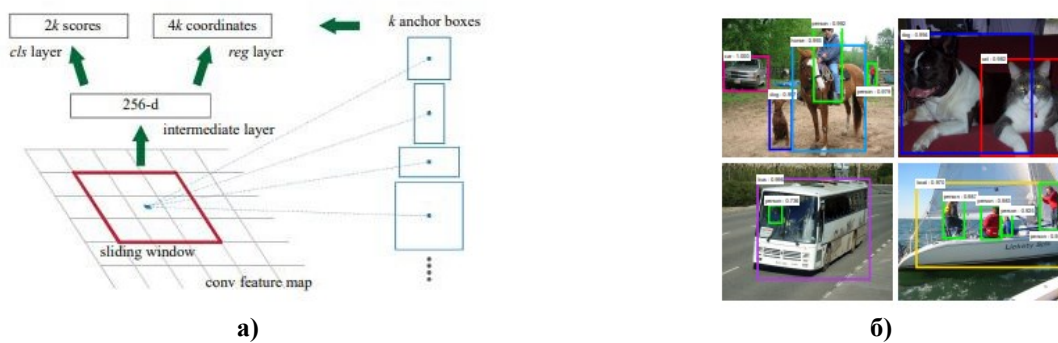


Рис. 5. а) Мережа регіональних пропозицій (RPN) б) Приклад виявлення за допомогою RPN

Подібно до Fast R-CNN, зображення надається вхідні дані для згорткової мережі, яка забезпечує карту згорткових функцій. Замість використання алгоритму вибіркового пошуку на карті функцій для визначення пропозицій регіону використовується окрема мережа. Прогнозовані області потім змінюються і використовується для класифікації зображення в межах запропонованої області та прогнозування значень зсуву для обмежувальних рамок.

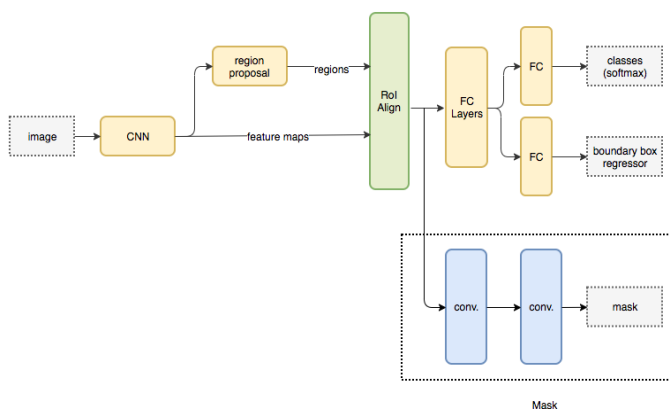


Рис. 6. Архітектура Mask R-CNN

Faster R-CNN набагато швидший за своїх попередників (R-CNN, Fast R-CNN). Тому алгоритм можна використовувати для виявлення об'єктів у реальному часі.

Однією з покращених версій R-CNN є Mask R-CNN. У той час як деякі версії R-CNN були зосереджені на виявленні об'єктів, Mask R-CNN додає сегментацію екземплярів. Автори роботи [13] представили концептуально просту, гнучку структуру для сегментації об'єктів. Метод під назвою Mask R-CNN розширює Faster R-CNN шляхом додавання гілки для прогнозування маски об'єкту паралельно з існуючою гілкою для розпізнавання обмежувальної рамки. Підхід, який представили розробники, дозволяє ефективно виявляти об'єкти на зображенні,

одночасно генеруючи високоякісну маску сегментації для кожного екземпляра.

Mask R-CNN являє собою двоетапну структуру: на першому етапі сканується зображення і генеруються області, які можуть містити об'єкт. На другому етапі пропозиції класифікуються та створюються обмежувальні рамки та маски. На рисунку 6 представлена архітектура Mask R-CNN.

Опис роботи мережі:

- Зображення проходить через згорткову мережу.
- Вихідні дані передаються до мережі регіональних пропозицій (RPN), яка створює різні блоки прив'язки (регіони або карти функцій, які містять певний об'єкт), що підлягають виявленню.
- Блоки прив'язки відправляються на етап вирівнювання області інтересу, який перетворює ці області до розміру, який потрібно для подальшої обробки.
- Вихідні дані після формування області інтересу переходять до формування (вирівнювання)

рамки, що обмежує об'єкт.

– Вихідні дані етапу вирівнювання області інтересу паралельно відправляються до Convolutional Networks для створення маски пікселів об'єкту.

Mask R-CNN генерує маску сегментації. Для цього спочатку обчислюється область інтересу, щоб можна було скоротити час обчислення. Після чого для всіх передбачуваних регіонів обчислюється перетин через об'єднання (IoU) :

$$IoU = \frac{\text{Площа перетину}}{\text{Площа об'єднання}} \quad (3)$$

Якщо $IoU \geq 0.5$, будемо вважати регіон регіоном інтересу. В іншому випадку нехтуємо регіоном. Це робиться для всіх регіонів. Далі вибираємо лише набір регіонів, для яких $IoU \geq 0.5$.

У статті [14] наведено детальний опис регіональної згорткової нейронної мережі та Fast R-CNN, Faster R-CNN, Region-Based Fully Convolutional Networks. Автори представляють огляд останніх оновлень у цій галузі та їх практичне застосування та класифікацію. Також порівнюють продуктивність і труднощі методів з точки зору швидкості, точності або простоти. В статті наведено, що швидкість роботи YOLO становить приблизно 21 ~ 155 кадрів в секунду, що є найшвидшим, але середня точність Mask R-CNN становить ~47,3, що перевершує всі інші методи.

Висновки

Таким чином можна зробити висновок, що використання нейронної мережі Mask R-CNN найкраще підходить для вирішення проблеми розпізнавання вільного місця на парковці з відеопотоку камери в реальному часі. Дана нейронна мережа має ряд переваг в порівнянні з такими як R-CNN, Fast R-CNN та YOLO. Опіраючись на різні дослідження даної нейронної мережі, можна з впевненістю сказати, що основною з її переваг є продуктивність та більша точність. В додаток до цього можна отримати оптимальний баланс між продуктивністю та точністю за рахунок правильного підбору налаштувань мережі.

References

1. Kvasnikov V., Ornatskyi D., Graf M., Shelukha M. Designing a computerized information processing system to build a movement trajectory of an unmanned aircraft vehicle. Eastern-European Journal of Enterprise Technologies. 2021. №1/9(109). P. 33–42.
2. Zhao Z., Zheng P., Xu S., Wu S. Object detection with deep learning: A review. Computer Vision and Pattern Recognition. 2019. №1. P. 1–21.
3. Adarsh P., Rathi P., Kumar M. YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. 6th International Conference on Advanced Computing and Communication Systems (ICACCS). 2020. P. 687–694.
4. Mao Q., Sun H., Liu Y., Jia R. Mini-YOLOv3: Real-Time Object Detector for Embedded Applications. IEEE Access. 2019. № 7. P. 133529–133538.
5. Xu Z., Jia R., Sun H. et al. Light-YOLOv3: fast method for detecting green mangoes in complex scenes using picking robots. Appl Intell. 2020. № 50. P. 4670–4687.
6. Martinez-Alpiste I., Golcarenenji G., Wang Q., Alcaraz-Calero J. A dynamic discarding technique to increase speed and preserve accuracy for YOLOv3. Neural Comput & Applic. 2021. № 33. P. 9961–9973.
7. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016. P. 779–788.
8. Meena S., Lasya N. et al. Hybrid Neural Network Architecture for Multi-Label Object Recognition using Feature Fusion. Procedia Computer Science. 2022. № 215. P. 78–90.
9. Wu M., Yue H., Wang J. Object detection based on RGC mask R-CNN. IET Image Processing. 2020. № 14. P. 1502–1508.
10. Levkivskyi V., Marchuk D., Lobanchykova N. Available parking places recognition system. CEUR Workshop Proceedings 4th Workshop for Young Scientists in Computer Science & Software Engineering Volume. 2022. № 3077. P. 123–134.
11. Girshick R., Donahue J., Darrell T., Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA. 2014. P. 580–587.
12. Ren S., He K., Girshick R., Sun R. Faster r-cnn: Towards real-time object detection with region offer networks. Computer Vision and Pattern Recognition. 2016. <https://doi.org/10.48550/arXiv.1506.01497>.
13. He K., Gkioxari G., Dollár P., Girshick R. IEEE International Conference on Computer Vision, ICCV. 2017. P. 2980–2988.
14. Bharati P., Pramanik A. Deep learning techniques—r-cnn to mask r-cnn: a survey. Computational Intelligence in Pattern Recognition, CIPR, Springer: Singapore. 2020. P. 657–668.