

<https://doi.org/10.31891/2307-5732-2025-357-77>

УДК 004.8

ТЕЛІШЕВСЬКИЙ ПЕТРО

Національний університет «Львівська Політехніка»

<https://orcid.org/0009-0008-8328-0373>email: petro.a.telishevskiy@lpnu.ua**МАГЕРОВСЬКИЙ ДМИТРО**

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0002-0640-7827>email: Dmytro.v.maherovskyi@lpnu.ua**БОЙКО НАТАЛІЯ**

Національний університет «Львівська Політехніка»

<https://orcid.org/0000-0002-6962-9363>email: nataliya.i.boyko@lpnu.ua

СЕГМЕНТАЦІЙНЕ РОЗПІЗНАВАННЯ СКУПЧЕНЬ МЕХАНІЗОВАНИХ ОБ'ЄКТІВ НА ЗОБРАЖЕННІ МЕТОДАМИ ДВОЕТАПНОЇ СЕГМЕНТАЦІЇ НА БАЗІ R-CNN

У статті розглядається важлива задача сегментаційного розпізнавання скупчень механізованих об'єктів на зображеннях з використанням методів сегментації на основі R-CNN. Скупчення транспортних засобів або інших механізованих об'єктів створюють значні проблеми в різних галузях, від управління заторами на дорогах та аналізу аварій до цілеспрямованих операцій із застосуванням безпілотників. Точна ідентифікація меж об'єктів та оцінка їхньої кількості є важливими для ефективного прийняття рішень у таких сценаріях. Авторі пропонують рішення, засноване на застосуванні моделей нейронних мереж Mask R-CNN та Cascade Mask R-CNN. Дослідження включає широкий огляд останніх досліджень, які застосовували архітектури Mask R-CNN і ResNet-50 для задач виявлення об'єктів, висвітлюючи їхні сильні сторони та обмеження, особливо в умовах змінної видимості.

В експериментальному розділі результати демонструють перевагу каскадної архітектури Cascade mask R-CNN над архітектурою Mask R-CNN, що в більшості випадків забезпечує меншу похибку. Однак, в окремих випадках Mask R-CNN показав кращу продуктивність або був на рівні із Cascade mask R-CNN. Головним недоліком каскадної структури залишається складність в обчисленнях та вимоги до апаратного забезпечення, що в свою чергу створює багато проблем.

Майбутня робота може бути зосереджена на оптимізації виведення моделі

Ключові слова: R-CNN, Mask R-CNN, Cascade Mask R-CNN, mean Average Precision, глибоке навчання

TELISHEVSKYI PETRO, MAHEROVSKYI DMYTRO, BOYKO NATALIYA

Lviv Polytechnic National University

SEGMENTATION RECOGNITION OF CLUSTERS OF MECHANIZED OBJECTS IN AN IMAGE USING TWO-STAGE SEGMENTATION METHODS BASED ON R-CNN

The paper addresses the critical task of segmentation recognition of clusters of mechanized objects in images using two-stage segmentation methods based on R-CNN architectures. Clusters of vehicles or other mechanized objects pose significant challenges in various fields, from traffic congestion management and accident analysis to targeted operations involving UAVs. Precise identification of object boundaries and estimation of the number of objects are essential for effective decision-making in such scenarios. The authors propose a solution based on the application of Mask R-CNN and Cascade Mask R-CNN. The research includes an extensive review of recent studies that applied Mask R-CNN and ResNet-50 architectures for aerial and object detection tasks, highlighting their strengths and limitations, particularly under varying visibility conditions.

In the experimental part, the results demonstrate that the cascade R-CNN mask outperforms the R-CNN mask in general, achieving higher precision across most categories. However, in specific cases at high IoU thresholds, Mask R-CNN showed competitive or superior performance. Despite Cascade Mask R-CNN's overall advantages, including improved segmentation accuracy due to its cascade structure refining detection at multiple IoU thresholds, computational complexity and hardware requirements remain significant challenges.

Future work may focus on optimizing model inference speed and robustness under adverse environmental conditions to facilitate broader real-world deployment.

Keywords: R-CNN, Mask R-CNN, Cascade Mask R-CNN, mean Average Precision, deep learning

Стаття надійшла до редакції / Received 06.07.2025

Прийнята до друку / Accepted 15.08.2025

Постановка проблеми

Скупчення механізованих об'єктів можуть становити проблему для виконання різного роду завдань. Починаючи від розпізнавання заторів та вирішення проблеми їх обходу, скупчення механізованих об'єктів можуть становити проблему, або навіть небезпеку, у випадку дорожньо-транспортних пригод на швидкісних трасах (завал автомобілів, що спричиняє подальші аварії). Також скупчення механізованих об'єктів ймовірно є вигідною ціллю для скиду некерованих об'єктів із безпілотників, або їх позначення на контурних картах, або що. У випадку із описаними вище проблемами, вкрай необхідно визначати чіткі границі механізованих об'єктів, точки їх дотикання. Необхідністю також є підрахування кількості таких механізованих об'єктів (для прикладу для оцінки

об'єму скупчення сили, або розуміння потенційної кількості постраждалих, абощо). Для вирішення проблеми сегментаційного розпізнавання механізованих об'єктів запропоновано використання методів та засобів на основі моделей штучного інтелекту. Від так відео буде по кадру поділене на окремі зображення, які буде проаналізовано, отримані результати опрацьовано системою. Надалі аналітику, або оператору буде надано звіт згідно із тими механізованими об'єктами що є на зображенні. Для найбільш ефективного вирішення задачі сегментаційного розпізнавання буде використано методи на базі R-CNN.

Аналіз досліджень та публікацій

У статті [1] описано використання глибокої нейронної мережі Mask R-CNN для виявлення транспортних засобів на знімках з безпілотників (UAV). Автори проаналізували 282 зображення з дронів DJI Mavic PRO, ановані через VGG Image Annotator. Після до навчання моделі та оновлення ваг, вдалося досягти результату розпізнавання у 93 та 92 відсотки відповідно для двох найкращих моделей (ResNet-50 та MS COCO). Також до переваг методу Mask R-CNN можна віднести можливість маскувати об'єкти, а не просто будувати навколо них обмежувальні рамки, що безпосередньо впливає на можливість максимально наближено оцінювати ступінь пошкодження транспортних об'єктів. Недоліком методу Mask R-CNN є висока вимога до продуктивності системи, у випадку динамічного розпізнавання фреймрейт моделі може бути нижчим за такий у камери. Також недоліком дослідження цих авторів є відсутність тестування в умовах недостатньої видимості - дослідження проводили у сонячну погоду, проте не проводили у сутінках, в умовах дощу і/або туману.

Приклад використання ResNet також можна знайти у роботі авторів Шмідт та Кафа [2]. Попри те, що об'єктом дослідження були статичні об'єкти (будинки та їх дахи) - система на основі Mask R-CNN продемонструвала схожі результати із розпізнаванням транспорту: більше 90% дахів будинків на території походження набору даних у Німеччині було не лише розпізнано, а й визначено їх границі. Найкращий показник, як і в результаті із дослідженням наведеним вище, продемонструвала модель ResNet.

ResNet-50 — це глибока згортоква нейронна мережа (CNN) із 50 шарами, яка використовує концепцію залишкового навчання (residual learning). Ця архітектура була представлена в 2015 році дослідниками Kaiming He, Xiangyu Zhang, Shaoqing Ren та Jian Sun у статті "Deep Residual Learning for Image Recognition". Ключовою особливістю ResNet є залишкові (residual) з'єднання, або skip connections, які дозволяють сигналу обходити один або кілька шарів, полегшуючи навчання дуже глибоких мереж. Це допомагає уникнути проблеми зникаючого градієнта, яка часто виникає при навчанні глибоких мереж. Ключовою перевагою даної моделі є висока точність: ResNet-50 досягнув найкращого показника помилок 20,47% і найкращого показника помилок 5,25% на наборі даних ImageNet. Складністю у застосуванні моделі є високі вимоги до точності гіперпараметрів [3].

Як висновок, можна відзначити, що Mask R-CNN, використовуючи модель ResNet-50, демонструє виняткові результати для досліджень що спираються на зйомку з повітря. Незалежні дослідження, що проводилися у різних країнах світу, на різних об'єктах, спільно демонструють високу точність роботи, водночас із можливістю розпізнавання не лише об'єктів дослідження але і їх меж. Це важлива особливість підходу що досліджується, дозволяє визначати не лише кількість транспортних об'єктів та їх розташування, але і стан цих об'єктів - це є великою перевагою не лише для цивільного, але і для інших видів застосування. Можливість визначати наближену "складність" транспортної ситуації що відбулася, дозволить ефективно використовувати наявні ресурси, проте необхідно зважати на складність обчислень, та можливу затримку розпізнавання через високі вимоги до продуктивності апаратного забезпечення. Також варто зазначити, що на популярних безпілотників типу "Мавік" обчислювальної потужності може не вистачити для задачі розпізнавання, що унеможливило автономну роботу даного безпілотників. Найкращі результати система показувала після до навчання нейронної мережі, враховуючи складність підбору гіперпараметрів для ResNet-50, це є одним із потенційних ризиків застосування моделі машинного навчання яку необхідно враховувати.

Виклад основного матеріалу

У цьому дослідженні ми поекспериментуємо з методами, які використовують R-CNN. Оскільки нам цікаво визначати на зображенні групу об'єктів. Для цього будемо проводити сегментацію і використаємо дві моделі нейронної мережі, а саме Mask R-CNN та Cascade Mask R-CNN.

Mask R-CNN є розширенням fast R-CNN з додатковою гілкою, яка генерує маску об'єкта [5]. Це дозволяє точніше виявляти межі об'єктів, але вимагає складнішого просторового планування. Модель використовує втрати від багатозадачності, включаючи класифікацію, коробкову регресію та генерацію маски. Вони обчислюються за допомогою рівняння.

$$L = L_{cls} + L_{box} + L_{mask},$$

де L_{cls} – classification loss, L_{box} – bounding-box loss, L_{mask} – mask generation loss.

Відмінність від інших систем полягає в тому, що маска не конкурує між класами, що дозволяє генерувати маски для кожного класу окремо. На рис.1 буде зображено Структуру Mask R-CNN.

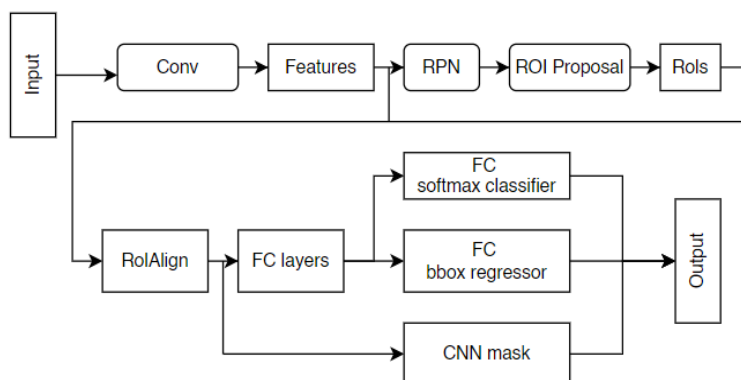


Рис. 1. Візуалізація структури Mask R-CNN

Zhaowei Cai та Nuno Vasconcelos у своїй роботі досліджують, що окремі класифікатори часто перенастроюються, а продуктивність IoU відрізняється від навчання та виводу. Тому вони запропонували створити каскадну структуру на основі Mask R-CNN. Ідея полягає в тому, що регіон генерується за допомогою найнижчого детектора з порогом IoU 0,5. Результат використовується для наступного детектора, де значення IoU дорівнює 0.6. Потім на основі відповідного детектора навчається детектор зі значенням 0,7 IoU для отримання остаточного вихідного значення.

Можна сказати, що модель каскадної маски R-CNN має 4 кроки. На рис. 2 показано візуалізацію структури каскадної маски R-CNN.

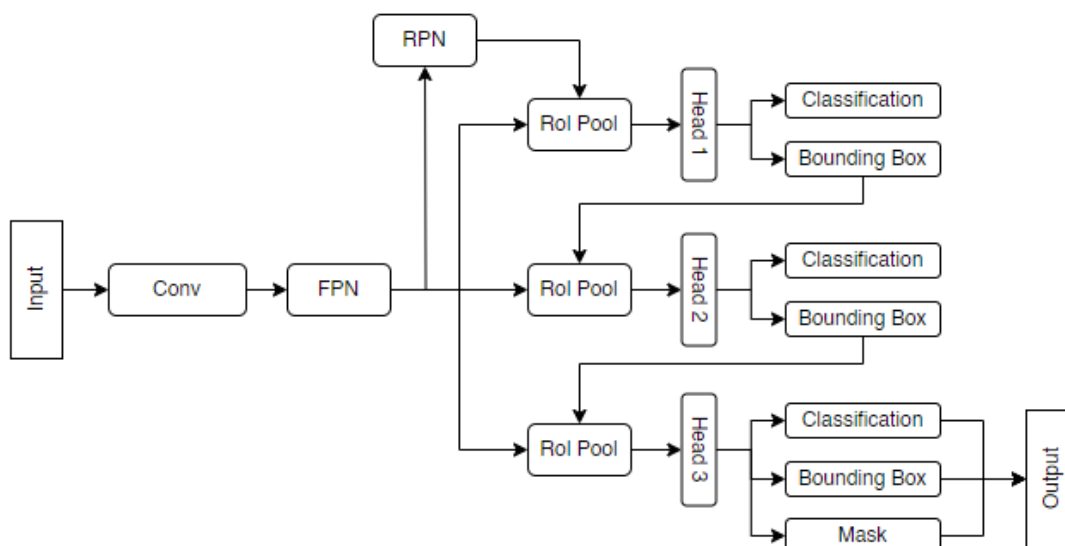


Рис. 2. Візуалізація структури Cascade mask R-CNN

Результати дослідження

Проведемо кілька експериментів з використанням наведених вище моделей: Mask R-CNN та Cascade Mask R-CNN. Як механізовані об'єкти було вирішено використати автомобілі, а для того ускладнити задачу буду визначати об'єкти на автотранспорті, а саме пошкодження. Було взято набір даних Car Damage Detection [5] де у нас є 6 класів пошкодження автотранспорту.

Давайте поговоримо, що буде використовуватися як метрика оцінки для оцінювання моделі, що тут представленні – це mean Average Precision. Дана метрика оцінки буває двох варіантів, а саме, mAP50 та mAP50-95. Різниця полягає в тому що для mAP50 – це середня точність при значенні IoU ≥ 0.5 , а для mAP50-95 це становить середня точність, яка формується усередненим значенням IoU 0.5 – 0.95, де крок становить 0.05.

Уже згадали про показник IoU що формує mAP, це показник перетину двох областей, а саме, передбачуваної маски із істинною маскою. Повна назва IoU – Intersection over Union. Однак, крім цього показника для розрахунку mean Average Precision використовується Precision та Recall. Де для Precision – це число правильно передбачуваних об'єктів до усіх передбачень, а Recall це число правильно передбачуваних об'єктів до усіх реальних об'єктів. Після чого розраховується Average Precision, дане значення буде формуватися від площі кривої Precision-Recall.

Маючи усі значення можна розрахувати mAP50. Спершу ми перевіряємо якщо IoU буде більше рівне 0.5 то передбачення вважається правдивим і тоді рахуємо Average Precision(AP), після чого рахуємо середнє значення для AP по всіх класах за формулою:

$$mAP50 = \frac{1}{N} \sum_{i=1}^N AP_i,$$

Де AP_i – це середня точність для класу i ; N – кількість класів. Отож відомо як формується mAP50 і mAP50-95, то ж переглянемо таблицю 1 з результатними проведених експериментів.

Таблиця 1

Показники mAP50 та mAP50-95 по усім класах пошкодження

	Mask R-CNN		Cascade mask R-CNN	
	mAP50	mAP50-95	mAP50	mAP50-95
усереднене	0.735	0.563	0.780	0.596

У таблиці 2 будерозглянуто значення по кожному із класів.

Таблиця 2

Значення mAP50 та mAP50-95 для типу пошкоджень

	Mask R-CNN		Cascade mask R-CNN	
	mAP50	mAP50-95	mAP50	mAP50-95
glass shatter	0.964	0.908	0.975	0.927
tire flat	0.942	0.929	0.961	0.918
lamp broken	0.853	0.696	0.884	0.741
scratch	0.64	0.352	0.703	0.421
dent	0.604	0.342	0.682	0.359
crack	0.388	0.15	0.476	0.212

Для кращого розуміння результатів були побудовані стовпчасті діаграми. Дані діаграми будуть представлені на рис. 3.

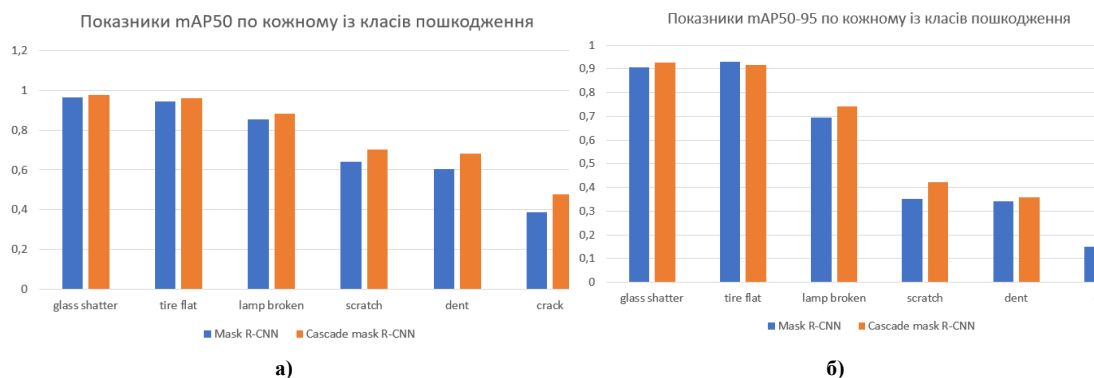


Рис. 3. Стовпчаста діаграма по метрикам оцінки методів: а) – для Map50; б) – для Map50-95

Якщо ми спочатку подивимося на таблиці, то можемо помітити, що каскадна маска R-CNN має перевагу над маскою R-CNN. Однак, діаграми кажуть що по деяким класам таких як tire flat для mAP50-95 перевага в більш точній сегментації є у Mask R-CNN.

Але в загальному у показники що по mAP50, що по mAP50-95 кращі у Cascade mask R-CNN, що є очікувано, тому що дана модель нейронної мережі була створена, щоб покращити результати Mask R-CNN.

Висновок

У цьому дослідженні вивчалось використання методів сегментації, а саме Mask R-CNN та Cascade mask R-CNN, для розпізнавання кластерів механізованих об'єктів на зображеннях. За допомогою обширної експериментальної оцінки з використанням набору даних Car Damage Detection продемонстровано ефективність цих моделей у виявленні та сегментації різних типів пошкоджень транспортних засобів. Результати показують, що Cascade Mask R-CNN забезпечує вищу точність сегментації в більшості випадків порівняно з Mask R-CNN, про що свідчить покращення метрик mAP50 та mAP50-95. Однак Mask R-CNN показав дещо кращі результати в деяких категоріях при більш високих порогах точності, що свідчить про те, що при виборі моделі слід враховувати пріоритети конкретного завдання.

Тим не менш, такі проблеми, як високі обчислювальні вимоги та чутливість моделі до умов навколишнього середовища, залишаються значними. Подальша робота має бути зосереджена на оптимізації обчислюваних вимог та вирішення впливу умов навколишнього середовища.

Література

1. Bilecik Şeyh Edebali University, Turkey, Yayla, R., Albayrak, E., Bilecik Şeyh Edebali University, Turkey, Yüzgeç, U., & Bilecik Şeyh Edebali University, Turkey. (2022). Vehicle Detection from Unmanned Aerial Images with Deep Mask R-CNN. *Computer Science Journal of Moldova*, 30(2 (89)), 148–169. <https://doi.org/10.56415/csjm.v30.09>
2. Musyarofah, Schmidt, V., & Kada, M. (2020). Object detection of aerial image using mask-region convolutional neural network (mask R-CNN). *IOP Conference Series: Earth and Environmental Science*, 500(1), 012090. <https://doi.org/10.1088/1755-1315/500/1/012090>
3. Boesch, G. (2023, Листопад 14). *Deep Residual Networks (ResNet, ResNet-50)*. Viso.Ai. <https://viso.ai/deep-learning/resnet-residual-neural-network/>
4. Bi, X., Hu, J., Xiao, B., Li, W., & Gao, X. (2023). IEMask R-CNN: Information-Enhanced Mask R-CNN. *IEEE Transactions on Big Data*, 9(2), 688–700. <https://doi.org/10.1109/TBDATA.2022.3187413>
5. Wang, X., Li, W., & Wu, Z. (2023). CarDD: A New Dataset for Vision-Based Car Damage Detection. *IEEE Transactions on Intelligent Transportation Systems*, 24(7), 7202–7214. <https://doi.org/10.1109/TITS.2023.3258480>

References

1. Bilecik Şeyh Edebali University, Turkey, Yayla, R., Albayrak, E., Bilecik Şeyh Edebali University, Turkey, Yüzgeç, U., & Bilecik Şeyh Edebali University, Turkey. (2022). Vehicle Detection from Unmanned Aerial Images with Deep Mask R-CNN. *Computer Science Journal of Moldova*, 30(2 (89)), 148–169. <https://doi.org/10.56415/csjm.v30.09>
2. Musyarofah, Schmidt, V., & Kada, M. (2020). Object detection of aerial image using mask-region convolutional neural network (mask R-CNN). *IOP Conference Series: Earth and Environmental Science*, 500(1), 012090. <https://doi.org/10.1088/1755-1315/500/1/012090>
3. Boesch, G. (2023, November 14). *Deep Residual Networks (ResNet, ResNet-50)*. Viso.Ai. <https://viso.ai/deep-learning/resnet-residual-neural-network/>
4. Bi, X., Hu, J., Xiao, B., Li, W., & Gao, X. (2023). IEMask R-CNN: Information-Enhanced Mask R-CNN. *IEEE Transactions on Big Data*, 9(2), 688–700. <https://doi.org/10.1109/TBDATA.2022.3187413>
5. Wang, X., Li, W., & Wu, Z. (2023). CarDD: A New Dataset for Vision-Based Car Damage Detection. *IEEE Transactions on Intelligent Transportation Systems*, 24(7), 7202–7214. <https://doi.org/10.1109/TITS.2023.3258480>