

В. О. КРОМКАЧ

аспірант кафедри комп'ютеризованих систем автоматичної
Національний університет «Львівська політехніка»
ORCID: 0009-0001-5608-5715

ІДЕНТИФІКАЦІЯ ОБ'ЄКТІВ СТАТИЧНИХ ЗОРОВИХ СЦЕН: НОВИЙ ПІДХІД

Ця наукова стаття пропонує новий підхід до ідентифікації об'єктів на статичних зорових сценах, що поєднує передові методи машинного навчання та обробки зображень. Завдяки двоетапній структурі запропонований метод ідентифікації забезпечує високу точність і швидкість виявлення об'єктів у різних умовах освітлення, спотворень та шумів. Переваги нового підходу включають високу продуктивність, ефективне використання ресурсів та масштабованість для різних галузей застосування.

Метою дослідження є підвищення точності і швидкості ідентифікації об'єктів на статичних зорових сценах, стійкості до різних видів шуму та спотворень. Дослідження передбачає оптимізацію алгоритмів обробки зображень і машинного навчання, які дозволяють ефективно виділяти та класифікувати об'єкти в умовах обмежених ресурсів та варіативності зорових сцен.

Методологія дослідження полягає у порівняльному аналізі існуючих методів ідентифікації об'єктів, включаючи як двоетапні, так і одноетапні підходи. Виявлено їхні переваги та обмеження. Проведено емпіричне дослідження та оцінку ефективності і точності різних підходів до виявлення та ідентифікації об'єктів.

Наукова новизна отриманих у роботі результатів полягає у поєднанні одноетапного та двоетапного виявлення об'єктів за допомогою RefineDet, що дозволяє покращити точність і швидкість виявлення об'єктів, порівняно з традиційними одноетапними методами.

Висновки. RefineDet є потужним та ефективним методом для виявлення об'єктів на зображеннях. Його унікальність полягає в комбінації переваг одноетапних та двоетапних підходів, що забезпечує високу точність та швидкість роботи. Це робить його придатним для використання в різних реальних додатках, де потрібна швидка і точна ідентифікація об'єктів.

Ключові слова: ідентифікація об'єктів, комп'ютерний зір, згорткові нейронні мережі, розпізнавання образів.

V. O. KROMKACH

Postgraduate Student at the Computerized Automation Systems Department
Lviv Polytechnic National University
ORCID: 0009-0001-5608-5715

THE ROLE OF COMPUTER VISION IN THE MODERN WORLD: ACHIEVEMENTS, CHALLENGES AND PROSPECTS

This research paper proposes a new approach to object identification in static visual scenes that combines advanced machine learning and image processing techniques. Due to the two-stage structure, the proposed identification method provides high accuracy and speed of object detection in various lighting, distortion and noise conditions. The advantages of the new approach include high performance, efficient use of resources and scalability for different application areas.

The purpose of the study is to increase the accuracy and speed of object identification in static visual scenes, resistance to various types of noise and distortions. The research involves the optimization of image processing and machine learning algorithms that allow for the efficient selection and classification of objects in conditions of limited resources and variability of visual scenes.

The methodology consists in a comparative analysis of existing object identification methods, including both two-stage and one-stage approaches. Their advantages and limitations are revealed. An empirical study and evaluation of the effectiveness and accuracy of various approaches to the detection and identification of objects was carried out.

The scientific novelty of the results obtained in the work is the combination of one-stage and two-stage detection of objects using RefineDet, which allows to improve the accuracy and speed of object detection, compared to traditional one-stage methods.

Conclusions. RefineDet is a powerful and efficient method for detecting objects in images. Its uniqueness lies in the combination of the advantages of one-stage and two-stage approaches, which ensures high accuracy and speed of work. This makes it suitable for use in various real-world applications where fast and accurate object identification is required.

Key words: object identification, computer vision, convolutional neural networks, pattern recognition.

Постановка проблеми

Швидкий розвиток науки й техніки в області обробки зображень розширив потенціал виявлення об'єктів за межі традиційних програм, таких як розпізнавання обличчя. Останні дослідження зосереджені на виявленні

історично значущих об'єктів, які мають велике значення для громад. Обробка зображень стала ключовою технологією для вдосконалення програм машинного навчання та штучного інтелекту (ШІ).

Метою виявлення об'єктів є ідентифікація та визначення місцезнаходження одного чи кількох значущих об'єктів у нерухомих даних або відео зображеннях. Це широкий термін, який охоплює різноманітні методи, які включають обробку зображень, класифікацію шаблонів [1], штучний інтелект і комп'ютерний зір. Він має широкий спектр можливих застосувань [2], включаючи запобігання дорожньо-транспортним пригодам, сповіщення про небезпечні товари у виробництві, моніторинг військових зон обмеженого доступу та складну взаємодію людини з комп'ютером. Через складність і різноманітність налаштувань програми виявлення кількох цілей у реальному світі важко знайти баланс між точністю та вартістю обробки [3].

Ідентифікація об'єктів, або розпізнавання, передбачає автоматичне виділення та класифікацію об'єктів. Автоматизація цього процесу вимагає ефективних методів машинного навчання, серед яких згорткові нейронні мережі (CNN) використовуються, в основному, через їхню ефективність у задачах розпізнавання об'єктів на основі зображень.

Процес ідентифікації складається з кількох етапів: попередня обробка, основна обробка та постобробка. Розробка алгоритму, який чудово ідентифікує об'єкти на зображеннях, вимагає оптимізації методу для кожного з цих етапів. Ідентифікація об'єктів є динамічною сферою, яка постійно викликає інтерес дослідників. Ефективна ідентифікація об'єктів на зображеннях передбачає зіставлення характеристик об'єктів, виявлених у візуальних даних. Алгоритм CNN користується особливою перевагою для розпізнавання об'єктів завдяки його надійній здатності ідентифікувати та класифікувати цифрові об'єкти. Однак залишаються проблеми, такі як виявлення малих об'єктів серед більших і робота зі складним фоном. Ці виклики стимулюють постійні інновації та вдосконалення для досягнення високоточних результатів.

Аналіз останніх досліджень і публікацій

У літературі про ідентифікацію об'єктів за допомогою згорткових нейронних мереж (CNN) часто розглядаються такі теми, як глибоке навчання, вилучення ознак і класифікація. Відтоді, як було представлено регіональну згортку (R-CNN) для точного виявлення та сегментації об'єктів, багато дослідників прийняли та вдосконалили R-CNN для підвищення точності ідентифікації об'єктів. Прогрес у дослідженнях CNN зосередився на розробці швидших алгоритмів із підвищеною точністю, що призвело до створення таких алгоритмів, як Fast R-CNN, Faster R-CNN і Mask R-CNN [4, 5].

CNN використовують методи згортання, щоб зібрати достатньо навчальних даних для ефективної класифікації об'єктів. Однак, великі навчальні набори даних можуть подовжити процес навчання та збільшити ризик переобладнання, коли алгоритм стає занадто пристосованим до навчальних даних і втрачає здатність до узагальнення. Запровадження R-CNN було відповіддю на пом'якшення проблем із надмірним оснащенням шляхом включення класифікації опорних векторних машин (SVM) у структуру CNN, хоча це доповнення також збільшило час обробки.

Щоб усунути недоліки R-CNN, зокрема залежність від модулів SVM, було розроблено Fast R-CNN, який замінив SVM функцією ймовірності Softmax, яка зазвичай використовується в CNN, спрощуючи процес. Подальший прогрес призвів до Faster R-CNN, який використовує техніку об'єднання регіонів інтересу (RoI), щоб забезпечити обробку для класифікації лише релевантних областей зображення, що підвищує ефективність [6, 7].

На основі архітектури Faster R-CNN було розроблено Mask R-CNN, додавши додатковий вхідний канал, відомий як RoI Align. Ця інновація не тільки покращує швидкість виявлення, але й розширює функціональні можливості, включаючи такі завдання, як визначення пози, поряд із традиційним виявленням об'єктів на одному зображенні. Ці ітераційні вдосконалення підкреслюють постійну еволюцію алгоритмів CNN для досягнення швидшої, точнішої та універсальної ідентифікації об'єктів.

Формулювання мети дослідження

Метою дослідження є підвищення точності і швидкості ідентифікації об'єктів на статичних зорових сценах, стійкості до різних видів шуму та спотворень. Дослідження передбачає оптимізацію алгоритмів обробки зображень і машинного навчання, які дозволяють ефективно виділяти та класифікувати об'єкти в умовах обмежених ресурсів та варіативності зорових сцен.

Виклад основного матеріалу дослідження

Виявляти об'єкти різного розміру на зображеннях із високою роздільною здатністю важко через проблеми, пов'язані з великою потребою в пам'яті та величезним обчислювальним навантаженням. Існуючі сучасні детектори добре працюють на зображеннях із низькою роздільною здатністю. Однак їх продуктивність значно обмежена на зображеннях високої роздільної здатності [8].

Аналіз загальних методів виявлення об'єктів на основі глибоких нейронних мереж включає чотири основні етапи: попередня обробка зображення, виділення ознак, класифікація і локалізація та постобробка.

Сирі зображення з набору даних не можна напряму подавати в нейронну мережу, тому на першому етапі попередньої обробки їх необхідно змінити до спеціальних розмірів і покращити якість, наприклад, підвищити

яскравість, насиченість кольорів і контрастність. Також для забезпечення варіативності навчальних даних застосовуються методи збільшення даних, такі як перевертання, обертання, масштабування, обрізка, зсув і додавання гауссового шуму. Крім того, генеративні змагальні мережі (GAN) можуть створювати нові зображення, що збагачують різноманітність вхідних даних.

Другий етап – виділення ознак – є ключовим для подальшого виявлення. Якість виділених ознак визначає максимальну можливу точність класифікації та локалізації об'єктів. Цей процес включає використання конволюційних шарів, які виділяють важливі деталі зображень.

На третьому етапі класифікації та локалізації використовується детектор, який відповідає за визначення кінцевих класів об'єктів та координат обмежувальних рамок. Детектор пропонує попередні області, що можуть містити об'єкти, і уточнює ці області шляхом класифікації та коригування координат.

Останній етап включає постобробку результатів виявлення для видалення слабких або надмірних детекцій. Метод неперекриття (Non-Maximum Suppression, NMS) є широко використовуваним підходом, який зберігає лише об'єкти з найвищими оцінками, видаляючи сусідні об'єкти з нижчими оцінками класифікації.

Для підвищення точності і швидкості ідентифікації об'єктів пропонується поєднання одноступінчастих і двоступінчастих детекторів.

Загалом існуючі детектори об'єктів поділяються на дві категорії: двоступеневий детектор, репрезентативний Faster R-CNN, який інтегрує виділення регіонів безпосередньо в CNN за допомогою RPN [9]. Інший одноступінчастий детектор, такий як YOLO, виконує ідентифікацію об'єктів за один прохід через мережу, що забезпечує високу швидкість [10] та SSD, що використовує багаторівневі ознаки для одночасного визначення та класифікації об'єктів [11].

Двоступеневі детектори мають високу точність локалізації та розпізнавання об'єктів, тоді як одноступінчасті детектори досягають високої швидкості висновку та тестування. Два ступені детекторів розділені шаром об'єднання ROI (регіон інтересів). У детекторі Faster R-CNN перший етап, який називається RPN, мережа пропозицій регіону, пропонує обмежувальні рамки потенційних об'єктів. На другому етапі мережа витягує функції за допомогою RoIPool з кожного блоку-кандидата та виконує класифікацію та регресію обмежувальної рамки.

Щоб повністю успадкувати переваги однокаскадних і двокаскадних детекторів, одночасно подолавши їх недоліки, запропоновано використання нового RefineDet (Refined Single-Shot Detection), який є сучасним методом об'єктного виявлення, що поєднує точність дворівневих методів (таких як Faster R-CNN) з ефективністю одноетапних підходів (як YOLO і SSD). Він складається з двох основних компонентів: однорівневого виявлення об'єктів та дворівневої регресії обмежувальних рамок.

Окрім покращення оригінальних стратегій зіставлення та генерування, розробка унікальних мережевих архітектур також є альтернативним методом виявлення малих об'єктів.

Основні етапи роботи RefineDet

1. Попереднє визначення та виділення об'єктів (Anchor Refinement Module – ARM):

– створення базових регіонів (anchors): на зображенні формуються попередні регіони, які можуть містити об'єкти;

– відбір важливих регіонів: ARM проводить грубу фільтрацію цих регіонів, видаляючи ті, що ймовірно не містять об'єктів;

– попередня регресія координат: ARM також виконує початкову регресію, тобто коригування координат обмежувальних рамок, щоб вони краще відповідали об'єктам.

2. Уточнення та класифікація об'єктів (Object Detection Module – ODM):

– класифікація об'єктів: ODM проводить детальну класифікацію об'єктів у відфільтрованих регіонах;

– уточнення регресії: ODM виконує остаточне уточнення координат обмежувальних рамок, забезпечуючи високу точність їх положення на зображенні.

Результати використання запропонованого методу представлені на рисунку 1.

Refinedet створює архітектуру зверху вниз і усуває дисбаланс з різних точок зору [12].

Переваги RefineDet

Висока точність: Завдяки дворівневій регресії, RefineDet забезпечує більш точне положення обмежувальних рамок, що підвищує загальну точність виявлення.

Ефективність: Поєднуючи одноетапне та двоетапне виявлення, RefineDet досягає балансу між швидкістю і точністю, що робить його ефективним для використання в реальних застосуваннях.

Зниження помилок фільтрації: Попередня фільтрація через ARM зменшує кількість помилкових детекцій, що знижує навантаження на подальші етапи обробки.

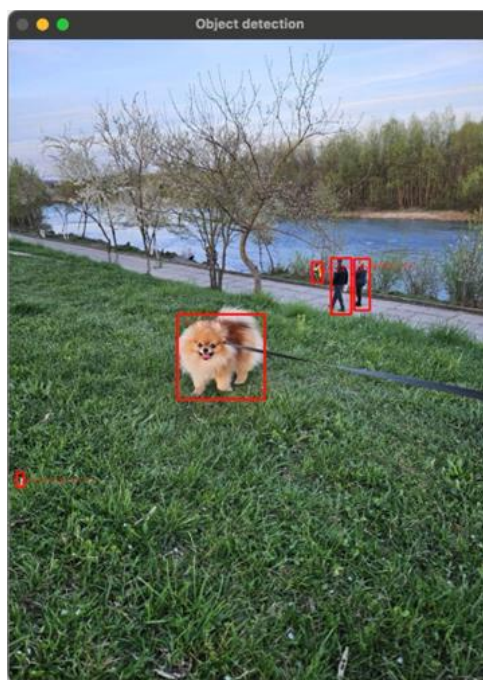


Рис. 1. Використання Refined Single-Shot Detection для об'єктного виявлення

Висновки

Ідентифікація об'єктів на статичних зорових сценах є однією з найважливіших задач у сфері комп'ютерного зору. Застосування цієї технології охоплює широкий спектр галузей, включаючи автоматизовані системи управління, медицину, робототехніку та безпеку. Проте існуючі методи мають свої обмеження, зокрема стосовно точності та швидкості виявлення об'єктів у складних умовах, на зображеннях високої роздільної здатності, виявлення малих об'єктів, тощо.

В результаті проведеного дослідження для ідентифікації об'єктів запропоновано використовувати RefineDet, який завдяки інноваційній структурі з ARM та ODM модулями забезпечує високу продуктивність і точність в задачах виявлення об'єктів на зображеннях. Обидва модулі тренуються одночасно, що дозволяє моделі поступово покращувати результати на основі зворотного зв'язку між модулями.

Отже, RefineDet є потужним та ефективним методом для виявлення об'єктів на зображеннях. Його унікальність полягає в комбінації переваг одноетапних та двоетапних підходів, що забезпечує високу точність та швидкість роботи. Це робить його придатним для використання в різних реальних додатках, де потрібна швидка і точна ідентифікація об'єктів.

Список використаної літератури

1. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, and C. Y. Fu. SSD: Single shot multibox detector, in Proc. Eur. Conf. Comput. Vis., Amsterdam, The Netherlands, Oct. 2016, pp. 11–14.
2. R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 23–28
3. Sana Pavan Kumar Reddy. Cusp Pixel Labelling Model for Objects Outline Using R-CNN. 2021. URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9667355> (дата звернення: 01.11.2024).
4. S. Ren, K. He, R. Girshick, J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intelligence, 39 (2017), pp. 1137-1149.
5. K. He G. Gkioxari P. Dollár R. Girshick. Mask R-CNN Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV) Venice, Italy 22–29 October 2017.
6. J. An, A. Mikhaylov, K. Kim. Machine learning approach in heterogeneous group of algorithms for transport safety-critical system. Appl. Sci., 10. 2020.
7. S.M. Abbas S.N. Singh Region-based object detection and classification using faster R-CNN Proceedings of the 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT) Ghaziabad, India 9–10 February 2018.
8. Jiaxu Leng, Ying Liu, Xinbo Gao, Senior Member. Selective region enlargement network for fast object detection in high resolution images. URL: <https://doi.org/10.1016/j.neucom.2021.08.015> (дата звернення: 01.11.2024).

9. S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149, Jun. 2017.
10. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified real-time object detection", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 779-788, Jun. 2016.
11. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, et al., "SSD: Single shot multibox detector" in *Computer Vision–ECCV*, Cham, Switzerland:Springer, pp. 21-37, 2016.
12. Kai Shuang, Zhiheng Lya, Jonathan Loo, Wentao Zhang. Scale-balanced loss for object detection. URL: <https://doi.org/10.1016/j.patcog.2021.107997> (дата звернення: 04.11.2024).

References

1. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, and C. Y. Fu. (2016) SSD: Single shot multibox detector, in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct., pp. 11–14.
2. R. Girshick, J. Donahue, T. Darrell, and J. Malik. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 23–28.
3. Sana Pavan Kumar Reddy (2021) Cusp Pixel Labelling Model for Objects Outline Using R-CNN. Retrieved from: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9667355> (accessed 01 November 2024).
4. S. Ren, K. He, R. Girshick, J. Sun. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 39, pp. 1137-1149.
5. K. He G. Gkioxari P. Dollár R. Girshick (2017) Mask R-CNN Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV) Venice, Italy 22–29 October.
6. J. An, A. Mikhaylov, K. Kim (2020) Machine learning approach in heterogeneous group of algorithms for transport safety-critical system. *Appl. Sci.*, 10.
7. S.M. Abbas S.N. (2018) Singh Region-based object detection and classification using faster R-CNN Proceedings of the 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT) Ghaziabad, India 9–10 February 2018.
8. Jiaxu Leng, Ying Liu, Xinbo Gao, Senior Member (2021) Selective region enlargement network for fast object detection in high resolution images. Retrieved from: <https://doi.org/10.1016/j.neucom.2021.08.015> (accessed 01 November 2024)
9. S. Ren, K. He, R. Girshick and J. Sun (2017) "Faster R-CNN: Towards real-time object detection with region proposal networks", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149.
10. J. Redmon, S. Divvala, R. Girshick and A. Farhadi (2016) "You only look once: Unified real-time object detection", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 779-788.
11. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, et al. (2016) "SSD: Single shot multibox detector" in *Computer Vision–ECCV*, Cham, Switzerland:Springer, pp. 21-37.
12. Kai Shuang, Zhiheng Lya, Jonathan Loo, Wentao Zhang (2021) Scale-balanced loss for object detection. Retrieved from: <https://doi.org/10.1016/j.patcog.2021.107997> (accessed 04 November 2024).