

**Авраменко С.Є., аспірант спеціальності 124 Системний аналіз**  
**Науковий керівник: Желдак Т.А. к.т.н., доцент кафедри системного аналізу та управління**  
*(Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна)*

## **ОГЛЯД ПІДХОДІВ КОМП'ЮТЕРНОГО ЗОРУ ДЛЯ ВИЯВЛЕННЯ ОБ'ЄКТІВ У 2D ПРОСТОРІ У ЗАСТОСУВАННІ ДО АВТОНОМНОГО ВОДІННЯ**

Прагнення до автономного водіння впливає з його потенціалу для підвищення безпеки, ефективності та заходів економії часу, з оцінками глобальних економічних втрат у розмірі 600 мільярдів доларів США через дорожньо-транспортні пригоди [1]. Алгоритми виявлення об'єктів на основі глибокого навчання відіграють ключову роль у сприйнятті навколишнього середовища, ідентифікації дорожніх знаків, класифікації перешкод і точному визначенні місцезнаходження пішоходів і транспортних засобів для систем безпілотних транспортних засобів. Автономні транспортні засоби мають перспективу скоротити кількість смертельних випадків і травм, спричинених людськими помилками під час ДТП, що є основною причиною смертності на дорогах.

Звичайні підходи до виділення характеристик зображення включають такі методи, як локальний бінарний шаблон (LBP) і гістограма орієнтованих градієнтів (HOG). Хоча LBP був швидким і придатним для менш потужних чіпів на той час, йому бракує можливості запропонувати вичерпний опис графічних функцій [2]. Навпаки, HOG об'єднує кілька комірок в один блок, забезпечуючи краще збереження як геометричної, так і оптичної інформації про зображення, але вимагає більших обчислювальних ресурсів.

Традиційним методам виявлення, що покладаються на розроблені вручну екстрактори ознак і класифікатори, не вистачало точності та ефективності, досягнутої зараз завдяки глибокому навчанню. Методи глибокого навчання для виявлення об'єктів можна класифікувати на дві основні категорії: двоетапні та одноетапні підходи. Двоетапний підхід передбачає створення блоків-кандидатів і розпізнавання об'єктів у цих блоках окремо, тоді як одноетапний підхід безпосередньо дає результати виявлення.

Серія R-CNN мереж вперше запровадила двоетапні методи виявлення об'єктів, використовуючи вибіркового пошук для створення кадрів виявлення-кандидата та глибокі згорткові нейронні мережі для виділення ознак, значно підвищуючи точність виявлення об'єктів [3]. Однак недоліком цього методу був повільний вивід через окреме вилучення характеристик глибини для кожного кадру. Щоб вирішити цю проблему, наступні вдосконалення, такі як SPPNet і Fast R-CNN, були спрямовані на вилучення повних функцій зображення та використання об'єднання регіонів інтересу (RoI) для масштабування можливостей кадру-кандидата, дозволяючи спільне навчання мережі. Faster R-CNN додатково уніфікував генерацію вікон кандидатів і класифікацію / регресію в єдину мережу, що призвело до спільного навчання.

Серія YOLO зробила революцію в одноетапному виявленні об'єктів шляхом безпосереднього розподілу зображень на підобласті та прогнозування ймовірності об'єкта, класу та зміщення положення для кожного [4]. Його простота та швидкість проклали шлях до наступних версій, таких як YOLOv2, v3, v4 та v5, у кожній з яких представлено вдосконалення для підвищення точності та швидкості виявлення. YOLOv2 покращив точність шляхом включення пакетної нормалізації та повної згортки. YOLOv3 представив Darknet-53 як магістральну мережу та багатомасштабне прогнозування. YOLOv4 вдосконалив процес виявлення для окремих графічних карт, тоді як YOLOv5 запропонував різні розміри детекторів цілей для різноманітних програм [5], [6].

Виявлення об'єктів у комп'ютерному зорі передбачає ідентифікацію та класифікацію об'єктів у зображенні, часто використовуючи згорточні нейронні мережі (CNN). Однак нещодавній інтерес до трансформерів, які довели свою ефективність у захопленні довгострокових залежностей у послідовних даних, таких як обробка природної мови (NLP), стимулював дослідження їх потенціалу в задачах комп'ютерного зору. Нещодавно дослідники досліджували наскрізні методи прогнозного виявлення, такі як DETR (detection transformer), передбачаючи кадр виявлення безпосередньо для кожного об'єкта [7]. Незважаючи на успіх DETR, він страждає повільною збіжністю та проблемами з виявленням дрібних об'єктів. Підхід Deformable DETR був спрямований на покращення продуктивності DETR [8]: Deformable DETR представив локальний модуль уваги на основі розрідженості для швидшої збіжності та кращого виявлення малого масштабу, тоді як Dai запропонував згорнуту піраміду, що деформується, і попереднє навчання без вчителя, щоб підвищити швидкість і точність DETR [9].

Таким чином, викладена проблема виявлення об'єктів отриманих з камер та інших сенсорів є актуальною. З удосконаленням штучного інтелекту, ML і DL системи виявлення об'єктів стають все більш важливими, вимагаючи швидких і точних механізмів виявлення для розпізнавання, класифікації та локалізації об'єктів на зображеннях. Її рішення для задач автономних транспортних засобів, що дозволяє їм ідентифікувати перешкоди, запобігати аваріям і забезпечувати безпечну навігацію буде розглянуто в наступних публікаціях.

#### Список використаних джерел:

- [1] Q. Chen, "A Review on Object Detection Algorithms for Autonomous Driving," Aug. 2023.
- [2] Y. Xu, L. Peng, and G.-Y. Li, "Multi modal registration of structural features and mutual information of medical image," *Future Gener. Comput. Syst.*, vol. 93, pp. 499–505, Apr. 2019, doi: 10.1016/j.future.2018.09.059.
- [3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 2980–2988. doi: 10.1109/ICCV.2017.322.
- [4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.
- [5] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement." arXiv, Apr. 08, 2018. doi: 10.48550/arXiv.1804.02767.
- [6] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv, Apr. 22, 2020. doi: 10.48550/arXiv.2004.10934.
- [7] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers." arXiv, May 28, 2020. doi: 10.48550/arXiv.2005.12872.
- [8] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable Transformers for End-to-End Object Detection." arXiv, Mar. 17, 2021. doi: 10.48550/arXiv.2010.04159.
- [9] Z. Dai, B. Cai, Y. Lin, and J. Chen, "UP-DETR: Unsupervised Pre-training for Object Detection with Transformers," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–11, 2022, doi: 10.1109/TPAMI.2022.3216514.