УДК 004.93:004.8

Яценко В. В., здобувач 4 курсу спеціальності
122 Комп'ютерні науки,
Ніколюк П. К., д-р фіз.-мат. наук, професор
кафедри інформаційних технологій

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ АЛГОРИТМІВ ДЕТЕКЦІЇ ОБ'ЄКТІВ РЕАЛЬНОГО ЧАСУ

Донецький національний університет імені Василя Стуса, м. Вінниця

Комп'ютерний зір – галузь штучного інтелекту, що займається автоматизованим аналізом візуальної інформації. Детекція об'єктів – ключове завдання цієї галузі, що полягає у автоматичному виявленні та просторовій локалізації об'єктів різних категорій у цифрових зображеннях чи відеопотоках [1].

Сучасний етап розвитку технологій характеризується масовим впровадженням систем автоматичного розпізнавання у повсякденне життя. Системи відеоспостереження у містах, автоматизовані лінії виробництва, допоміжні системи водіння автомобілів – усі ці застосування потребують надійного виявлення об'єктів у візуальних даних. В Україні детекція об'єктів набуває особливої гостроти через використання БПЛА у військовій сфері. Безпілотники потребують алгоритмів, здатних працювати в режимі реального часу за обмежених ресурсів, водночас завдання розвідки вимагають максимальної точності. Ця двоякість вимог – швидкість проти точності – визначає необхідність комплексного аналізу наявних підходів до детекції об'єктів.

У статті «Deep Learning for Generic Object Detection: A Survey» представлено комплексний огляд сучасних архітектур для детекції об'єктів на основі глибинного навчання [2]. Дослідження показує, що розвиток галузі рухається у двох напрямах: надточні моделі для критичних застосувань та ультрашвидкі детектори для периферійних пристроїв. Окрему увагу приділено гібридним архітектурам, що намагаються об'єднати переваги обох підходів через каскадні структури та адаптивне об'єднання ознак.

Традиційно алгоритми детекції поділяють на двоетапні та одноетапні детектори. Двоетапні детектори базуються на послідовній обробці зображення. Спочатку генеруються регіони-кандидати, далі кожна область класифікується та уточнюється. Такий підхід забезпечує високу точність локалізації, особливо для дрібних об'єктів. Faster R-CNN (Region-based Convolutional Neural Networks, регіональні згорткові нейронні мережі) є типовим представником цього класу. RetinaNet вдосконалив цей підхід через запровадження focal loss (фокусної функції втрат), що дає змогу мережі під час навчання приділяти більше уваги складним прикладам, покращуючи точність без зниження швидкості [3].

One-stage-детектори, тобто одноетапні алгоритми, здійснюють виявлення та класифікацію одночасно за один прохід, що суттєво прискорює обробку [4].

YOLO (You Only Look Once, дивишся лише раз) розглядає все завдання детекції як пряму регресію від пікселів зображення до координат та класів об'єктів,

що дає значний приріст у швидкості розпізнавання. Сучасні версії – YOLOv8 та YOLOv11 – демонструють значно кращий баланс між точністю та швидкістю завдяки оптимізованим опорним мережам і передовим методам аугментації даних під час навчання [5]. SSD (Single Shot Detector, детектор одиничного кадру) використовує передбачення на різних рівнях мережі для виявлення об'єктів різного масштабу.

Метою дослідження виступає аналіз сучасних методів детекції об'єктів з акцентом на співвідношення швидкості та точності для визначення оптимальних сценаріїв застосування різних підходів.

Під час дослідження необхідно обрати репрезентативні моделі, провести над ними тестування на стандартному датасеті з вимірюванням швидкості обробки, точності детекції та розміру моделей, проаналізувати отримані результати та сформулювати практичні рекомендації щодо вибору методу залежно від специфіки застосування.

Для об'єктивного порівняння було обрано шість моделей. Швидкі моделі – YOLOv8n та YOLOv11n – є найлегшими версіями сучасних YOLO – архітектур з розміром 5–6 MB, SSD300 – класичний одноетапний детектор розміром 136 MB, та MobileNet R-CNN – оптимізована версія Faster R-CNN для мобільних пристроїв розміром 74 MB. Точні моделі представлені Faster R-CNN з ResNet50 backbone розміром 160 MB – еталоном орієнтованого на точність підходу, та RetinaNet розміром 130 MB – детектором з focal loss.

Тестування проводилося на датасеті COCO2017, що містить 5 000 зображень з різноманітними об'єктами та сценами. Вимірювалися такі показники: швидкість обробки у кадрах на секунду (FPS), точність детекції за метриками mAP@0.5 (mean Average Precision при IoU=0.5) та mAP@0.5:0.95 (середнє значення mAP за різних порогів IoU від 0.5 до 0.95) та розмір моделі. Тестування виконувалося на обладнанні з графічним прискорювачем NVIDIA.

Результати тестування представлено на рис. 1. Зелені стовпці позначають швидкі моделі, сині – точні моделі, орієнтовані на максимальну точність. Червоні пунктирні лінії позначають критичні пороги: 15 FPS для відеоаналізу в реальному часі та 50 MB для розгортання на периферійних пристроях.

Швидкі моделі демонструють 53–67 FPS за точності 42–44 % mAP@0.5 та розмірі 5–136 MB, перевищуючи поріг у 15 кадрів у 3.5–4.5 рази. Точні моделі забезпечують 56–59 % mAP@0.5 (на 30 % вище), але працюють лише на 9 FPS при розмірі 130–160 MB.

Графік trade-off демонструє фундаментальний компроміс: збільшення точності у 1.3 раза вимагає зниження швидкості у 6–7 разів. MobileNet R-CNN займає проміжну позицію (26.6 FPS, 52.6 % mAP@0.5), оптимальну для обладнання середнього класу. Аналіз використання VRAM показує, що YOLO-моделі споживають 600–700 MB під час інференсу, що дає змогу розгортання на БПЛА з 4+ GB RAM після застосування квантизації INT8 та TensorRT оптимізації.

На основі отриманих результатів можна сформулювати чіткі рекомендації щодо вибору методу. YOLO – моделі оптимальні для реального часу на БПЛА та мобільних пристроях. Швидкість 50–70 FPS та розмір 5–6 MB дають змогу використання на embedded-системах. Точність 42–44 % mAP@0.5 є достатньою для задач моніторингу та навігації.

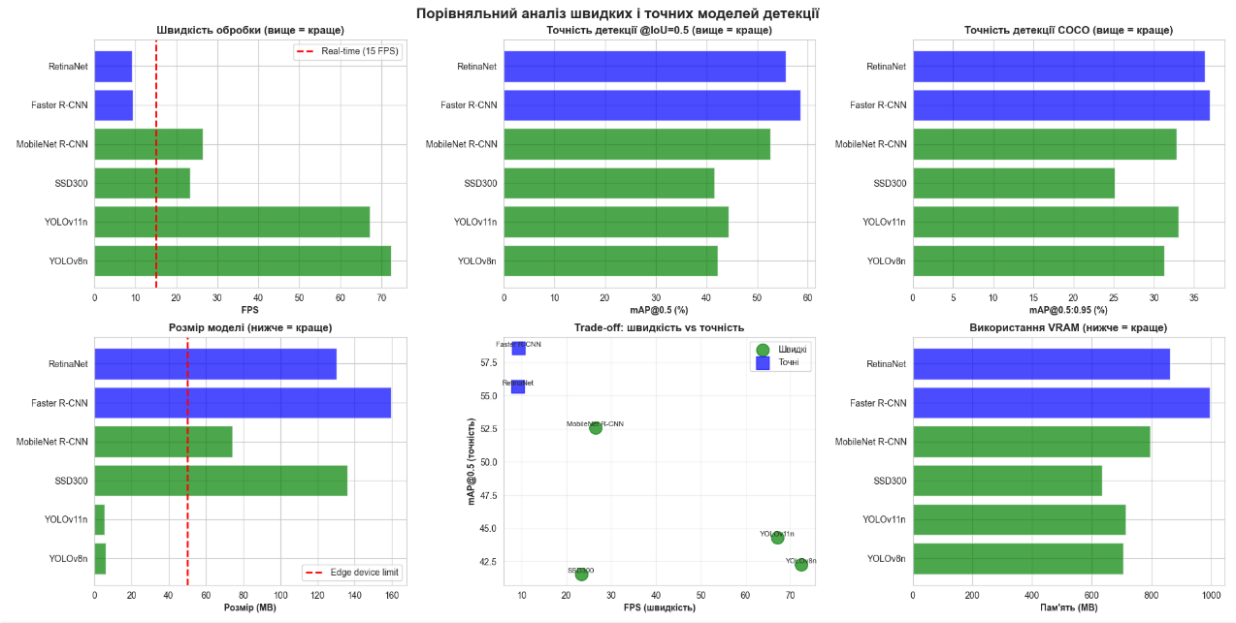


Рисунок 1 – Порівняльний аналіз продуктивності моделей детекції об'єктів

Faster R-CNN та RetinaNet доцільні для офлайн-аналізу: обробка відео, наукові дослідження, медична діагностика. Точність 56–59 % mAP@0.5 (на 30 % вище) виправдовує нижчу швидкість.

MobileNet R-CNN – проміжний варіант: 26.6 FPS, 52.6 % mAP@0.5. Підходить для відеоаналітики на обладнанні середнього класу.

Проведене дослідження підтверджує фундаментальний компроміс між швидкістю та точністю у методах детекції об'єктів. Одноетапні детектори забезпечують швидкість для реального часу, двоетапні демонструють вищу точність для критичних задач. Вибір методу визначається специфікою задачі та ресурсами.

Список використаних джерел

1. Computer Vision: Algorithms and Applications. URL: https://szeliski.org/Book/ (дата звернення: 02.12.2024).

2. Deep Learning for Generic Object Detection: A Survey. URL: https://arxiv.org/abs/1809.02165 (дата звернення: 02.12.2024).

3. Focal Loss for Dense Object Detection. URL: https://arxiv.org/abs/1708.02002 (дата звернення: 02.12.2024).

4. Object Detection in 20 Years: A Survey. URL: https://arxiv.org/abs/1905.05055 (дата звернення: 02.12.2024).

5. You Only Look Once: Unified, Real-Time Object Detection. URL: https://arxiv.org/abs/1506.02640 (дата звернення: 02.12.2024).