

НАВЧАННЯ МОДИФІКОВАНОЇ МОДЕЛІ YOLOV8 ДЛЯ ДЕТЕКЦІЇ ТРАНСПОРТНИХ ЗАСОБІВ НА КОМБІНОВАНОМУ ДАТАСЕТІ

Вінницький національний технічний університет

Анотація

У роботі представлено результати навчання модифікованої моделі YOLOv8-nano з модулями C3Ghost-SVAM на комбінованому датасеті для детекції транспортних засобів з камер відеоспостереження. Сформовано датасет обсягом 12000 зображень шляхом поєднання власного датасету (2000 зображень зі складними умовами зйомки) та публічного датасету DETRAC (10000 зображень). Експериментально підтверджено, що запропонована модифікація забезпечує зменшення кількості параметрів на 23.3% та обчислювальної складності на 9.9% при незначному зниженні точності (mAP@0.50:0.95 знизився лише на 1.2%).

Ключові слова: YOLOv8, детекція об'єктів, глибоке навчання, Ghost-згортки, SVAM, датасет DETRAC, транспортні засоби.

Abstract

The paper presents the results of training a modified YOLOv8-nano model with C3Ghost-SVAM modules on a combined dataset for vehicle detection from surveillance cameras. A dataset of 11,400 images was formed by combining a custom dataset (2,000 images with challenging conditions) and the public DETRAC dataset (10,000 images). Experimental results confirm that the proposed modification reduces the number of parameters by 23.3% and computational complexity by 9.9% with only a slight decrease in accuracy (mAP@0.50:0.95 decreased by only 1.2%).

Keywords: YOLOv8, object detection, deep learning, Ghost convolutions, SVAM, DETRAC dataset, vehicles.

Вступ

Детекція транспортних засобів є базовою задачею для інтелектуальних транспортних систем (ІТС), систем автоматичної фіксації порушень ПДР та аналізу трафіку. Сучасні методи детекції на основі глибоких нейронних мереж, зокрема сімейство YOLO (You Only Look Once), демонструють високу точність та швидкість обробки [1]. Однак розгортання таких моделей на периферійних пристроях (Edge) з обмеженими обчислювальними ресурсами залишається актуальною проблемою.

Додатковою проблемою є обмежена кількість спеціалізованих датасетів для навчання детекторів транспортних засобів у складних умовах зйомки: нічний час, погана погода (дощ, туман), низька роздільна здатність камер. Існуючі публічні датасети (KITTI, COCO) переважно містять зображення з добрими умовами освітлення та якості [2].

Метою роботи є формування комбінованого датасету та навчання модифікованої моделі YOLOv8-nano з модулями C3Ghost-SVAM для детекції транспортних засобів з камер відеоспостереження в різноманітних умовах.

Формування комбінованого датасету

Для навчання модифікованої моделі YOLOv8 з модулями C3Ghost-SVAM сформовано комбінований набір даних, що складається з двох джерел:

Власний датасет містить 2000 зображень, отриманих з відео камер спостереження за дорожнім рухом з різних джерел (YouTube, публічні архіви відеозаписів). Зображення характеризуються різноманітними умовами зйомки:

- час доби: день (60%), сутінки (20%), ніч (20%);
- погодні умови: ясно (50%), хмарно (30%), дощ (15%), туман (5%);
- ракурс камери: фронтальний, бічний, діагональний;
- щільність трафіку: низька, середня, висока.

Датасет DETRAC (Detection and Tracking Benchmark for Autonomous Vehicles) – публічний набір даних для детекції та трекінгу транспортних засобів [3], з якого відібрано 10000 зображень. DETRAC містить відеозаписи з камер спостереження за дорожнім рухом у Китаї та забезпечує додаткове різноманіття умов освітлення, типів транспортних засобів та сцен.

Характеристики комбінованого датасету наведено в таблиці 1.

Таблиця 1 – Характеристики комбінованого датасету

Параметр	Значення
Загальна кількість зображень	12000
– власний датасет	2000
– DETRAC	10000
Кількість анотованих об'єктів	~85000
Середня кількість об'єктів на зображенні	7.3
Роздільна здатність	640×640
Формат анотацій	YOLO
Розподіл (train/val/test)	70/20/10 %

Анотування виконано у форматі YOLO з використанням інструменту CVAT (Computer Vision Annotation Tool). Для кожного зображення визначено обмежувальні прямокутники транспортних засобів з класом «vehicle». Розподіл даних здійснено у співвідношенні 70:20:10 для тренувальної, валідаційної та тестової вибірок.

Архітектура модифікованої моделі

Базовою моделлю обрано YOLOv8-nano як найлегшу версію сімейства YOLOv8, що забезпечує оптимальний баланс між швидкістю та точністю для Edge-пристроїв [4]. Модифікація архітектури включає:

1. Заміна стандартних Bottleneck-блоків на GhostBottleneck у шарах P5/32 backbone та PAN-шляху. Ghost-згортки генерують частину feature maps через лінійні операції замість повних згорток, що зменшує кількість параметрів та FLOPS [5].
2. Інтеграція модуля уваги CBAM (Convolutional Block Attention Module) після конкатенації в CSP-блоках. CBAM послідовно застосовує каналну та просторову увагу для підсилення важливих ознак.

Структуру модифікованої моделі наведено в таблиці 2.

Таблиця 2 – Структура модифікованої нейронної мережі

Шар	Тип	Канали	Особливості
0-2	Conv + C2f	64→128	P1→P2
3-6	Conv + C2f	256→512	P3/8, P4/16
7-9	Conv + C3Ghost + SPPF	1024	P5/32
10-15	Upsample + Concat + C2f	512→256	FPN шлях
16-21	GhostConv + C3Ghost_CBAM	512→1024	PAN шлях
Detect	Detect Head	nc=1	P3, P4, P5

Результати навчання

Навчання проводилося в середовищі PyCharm з використанням бібліотеки Ultralytics YOLOv8 та фреймворку PyTorch 2.0. Апаратна платформа: NVIDIA GeForce RTX 4090 (24 GB VRAM), Intel Xeon W-2245. Гіперпараметри: 100 епох, batch size = 16, optimizer = SGD, learning rate = 0.01 з косинусним відпалом.

Для відстеження експериментів використовувалася платформа Weights & Biases, що забезпечила візуалізацію градієнтів, розподілу ваг та порівняння різних конфігурацій моделей.

Результати навчання на комбінованому датасеті наведено в таблиці 3.

Таблиця 3 – Результати навчання детектора на комбінованому датасеті

Модель	mAP@0.50:0.95	mAP@0.50	Params, М	GFLOPS
YOLOv8-nano (baseline)	0.715	0.915	3.0	8.1
Запропонована модель	0.706	0.908	2.3	7.3
<i>Різниця</i>	-1.2%	-0.7%	-23.3%	-9.9%

Запропонована конфігурація демонструє незначне зниження якості порівняно з базовою моделлю: mAP@0.50:0.95 зменшується з 0.715 до 0.706 (різниця -1.2%). Однак цей результат досягається при суттєвому скороченні обчислювальної складності: кількість параметрів зменшена на 23.3% (з 3.0 до 2.3 млн), а GFLOPS – на 9.9% (з 8.1 до 7.3).

Після навчання модель оптимізована для Edge-пристроїв шляхом конвертації в TensorRT з INT8-квантуванням. Для калібрування масштабів активацій використовувався репрезентативний датасет (200 зображень з тренувальної вибірки). Розмір квантованої моделі зменшився в 4 рази порівняно з FP32-версією.

Висновки

Сформовано комбінований датасет обсягом 11400 зображень для навчання детектора транспортних засобів шляхом поєднання власного датасету (1400 зображень зі складними умовами: ніч, дощ, туман, низька якість) та публічного датасету DETRAC (10000 зображень).

Навчено модифіковану модель YOLOv8-nano з модулями C3Ghost-CBAM. Результати підтверджують ефективність запропонованого підходу: зниження якості становить лише 1.2% (mAP@0.50:0.95: 0.715→0.706), тоді як кількість параметрів зменшена на 23.3%, а обчислювальна складність – на 9.9%.

Запропонований підхід забезпечує оптимальний баланс між точністю та обчислювальною ефективністю, що є важливим для розгортання моделей на Edge-пристроях з обмеженими ресурсами (NVIDIA Jetson Nano, мобільні пристрої).

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Jocher G., Chaurasia A., Qiu J. Ultralytics YOLOv8. 2023. URL: <https://github.com/ultralytics/ultralytics>.
2. Geiger A., Lenz P., Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite // Proceedings of CVPR. 2012. P. 3354–3361.
3. Wen L. et al. UA-DETRAC: A benchmark suite for multi-object detection and tracking // Computer Vision and Image Understanding. 2020. Vol. 193. Art. 102907.
4. Романець В., Бісікало О. Виявлення об'єктів дорожнього руху з камер відеоспостереження // Herald of Khmelnytskyi National University. Technical Sciences. 2025. Т. 355, № 4. С. 491–497.
5. Han K. et al. GhostNet: More features from cheap operations // Proceedings of CVPR. 2020. P. 1580–1589.

Романець Владислав Олександрович – аспірант групи 126-22а, факультет інтелектуальних інформаційних технологій та автоматизації, ВНТУ, Вінниця, e-mail: romanets.vlad@gmail.com

Бісікало Олег Володимирович – д-р техн. наук, професор, завідувач кафедри автоматизації та інтелектуальних інформаційних технологій, ВНТУ, м. Вінниця, e-mail: obisikalo@vntu.edu.ua

Romanets Vladyslav O. – post-graduate student, group 126-22a, faculty of intellectual information technologies and automation, VNTU, Vinnytsia, e-mail: romanets.vlad@gmail.com

Bisikalo Oleg V. – Doctor of Technical Sciences, Professor of the Department of Automation and Intelligent Information Technologies, VNTU, Vinnytsia, e-mail: obisikalo@vntu.edu.ua