УДК 004.853.3

ЛЕГКОВЕСНОЕ ВНИМАНИЕ ДЛЯ ТОЧНОЙ ЛОКАЛИЗАЦИИ В DSC-CNN ВО ВСТРАИВАЕМЫХ СИСТЕМАХ

LIGHTWEIGHT ATTENTION FOR ACCURATE LOCALIZATION IN DSC-CNNS FOR EMBEDDED SYSTEMS

Сацюк Александр Владимирович
кандидат технических наук, доцент,
заведующий научно-исследовательской лабораторией искусственного интеллекта,
Донецкий институт железнодорожного транспорта
alexandrsatsuk@gmail.com

Аннотация. Развертывание систем компьютерного зрения на периферийных устройствах для мониторинга железнодорожной инфраструктуры требует баланса между скоростью и точностью локализации. Данная работа предлагает оптимизированную архитектуру, основанную на глубинных сепарабельных свертках (DSC), дополненную легковесным модулем канального внимания (CAM) – DSC-AB. Интеграция CAM компенсирует присущую DSC потерю геометрической информации, что критически важно для высоких порогов IoU. Эксперименты на Raspberry Pi 5 показали, что DSC-AB обеспечивает прирост mAP0.75 на 8.3 п.п. по сравнению с базовой DSC-моделью, сохраняя при этом производительность выше 27 FPS после квантования INT8.

Ключевые слова: блок внимания, сверточная нейронная сеть, CNN, оптимизация нейронной сети, архитектура нейронной сети, Raspberry Pi 5, YOLO, depthwise separable convolution.

Satsyuk Alexander Vladimirovich
Candidate of Technical Sciences,
Associate Professor,
Head of the Artificial Intelligence Research Laboratory,
Donetsk Institute of Railway Transport
alexandrsatsuk@gmail.com

Annotation. Deploying computer vision systems on edge devices for railway infrastructure monitoring demands a balance between speed and localization accuracy. This work proposes an optimized architecture based on Depthwise Separable Convolutions (DSC), augmented with a lightweight Channel Attention Module (CAM) – DSC-AB. CAM integration compensates for the geometric information loss inherent in DSC, which is critical for high IoU thresholds. Experiments on Raspberry Pi 5 show that DSC-AB yields an mAP0.75 gain of 8.3 pp over the baseline DSC model while maintaining performance above 27 FPS after INT8 quantization.

Keywords: attention block, convolutional neural network, CNN, neural network optimization, neural network architecture, Raspberry Pi 5, YOLO, depthwise separable convolution.

Поддержка исследований: Работа выполнена в рамках государственного задания по теме ЕКТТ-2026-0001

Развертывание систем компьютерного зрения для инфраструктуры железнодорожного транспорта России на периферийных устройствах требует нахождения тонкого баланса между вычислительной эффективностью и качеством извлечения признаков [1, 2]. Глубинные сепарабельные свертки (DSC) [3, 4], используемые в архитектурах MobileNet [5], снижают число операций в 8–9 раз по сравнению с обычными свертками, что критически важно для встраиваемых платформ [6]. Однако эта оптимизация, нацеленная на минимизацию FLOPs, часто ведет к потере способности сети точно локализовывать объекты, что проявляется в падении метрики mAP0.75. Для устранения этого недостатка, не увеличивая существенно нагрузку, предлагается интегрировать легковесные механизмы внимания. Механизмы внимания динамически взвешивают важность признаков, усиливая релевантные области.

Цель работы – оптимизировать архитектуру на основе DSC путем внедрения селективных модулей внимания для повышения точности локализации (mAP0.75) при сохранении высокой частоты кадров на Raspberry Pi 5 после INT8 квантования.

Задача состоит в разработке модифицированной сверточной сети для высокоскоростной детекции на встраиваемых платформах с акцентом на точность локализации. Исходная структура на базе DSC обеспечивает низкую вычислительную нагрузку [3], но требует улучшения селективности. Модель должна работать со скоростью не менее 20 FPS на Raspberry Pi 5 после квантования до INT8. Формально, необходимо минимизировать сложность C и число параметров P, сохраняя FPS ≥ 20, и максимизировать mAP0.5. Основной фокус смещается с удаления слоев на качественное улучшение существующих вычислительных блоков за счет внедрения внимания.

Концепция архитектуры основана на гибридизации DSC для снижения стоимости и легковесного канального внимания (CAM) для повышения селективности. Базовая модель (рис. 1) использует DSC-блоки в секциях извлечения (backbone) и агрегации (neck), что обеспечивает скорость, но ухудшает локализацию.

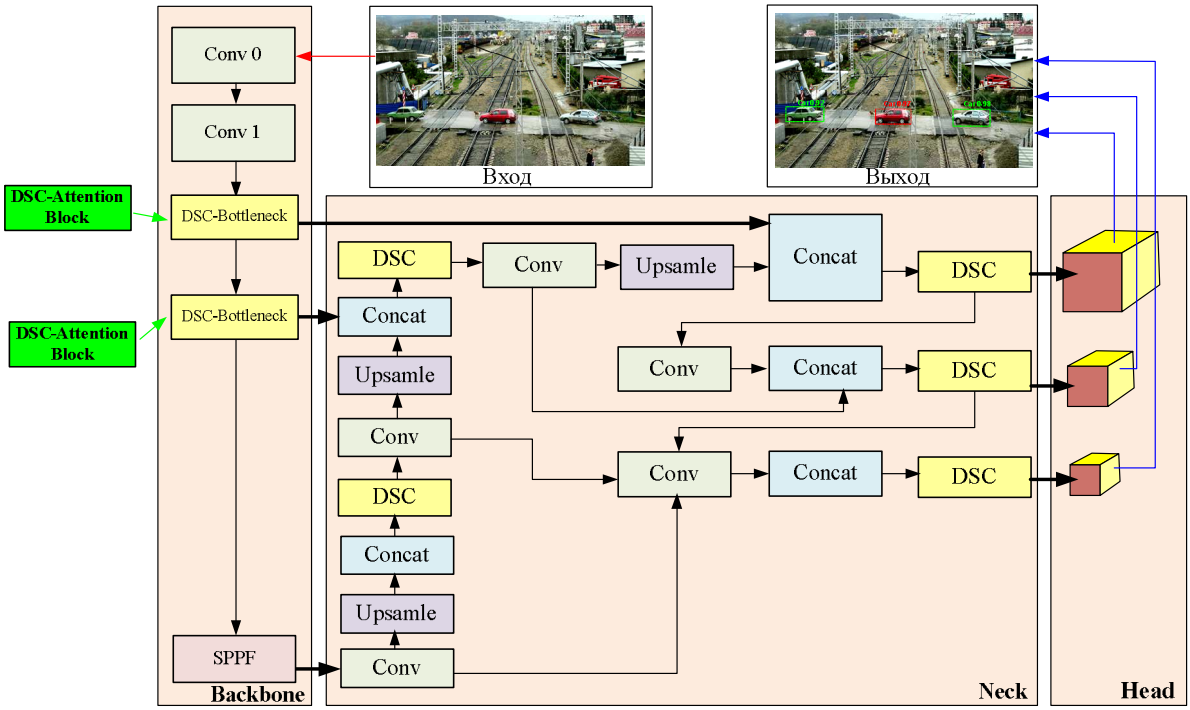

Рисунок 1– Архитектура однопроходной CNN на основе глубинных сепарабельных сверточных блоков

Для восстановления точности вводится CAM, который взвешивает важность каналов [7]. Канальное внимание CAM функционирует путем агрегирования пространственной информации через глобальное усреднение для получения дескриптора, последующего вычисления весов важности каналов с помощью двух полносвязных слоев и применения этих весов для селективной модуляции исходных карт признаков. Для ультралегковесных структур предложен модифицированный CAM с динамическим коэффициентом сжатия r = max(1,[C/32]) и эмпирически выбранным масштабирующим коэффициентом a = 0.3 во втором полносвязном слое, что снижает сложность на 15% по сравнению со стандартным SE-блоком. Этот модуль внедряется в блок Backbone сразу после точечной свертки 1×1 в DSC-блоке, формируя DSC-Attention Block (DSC-AB) (рис. 1 зеленые блоки), поскольку именно после смешивания каналов требуется модуляция для точной локализации. CAM в секции Neck адаптируется для взвешивания признаков, поступающих из разных масштабов PANet-подобной структуры, обеспечивая оптимальное слияние семантики и пространственной точности.

Исследование проводилось на Raspberry Pi 5 с использованием видеопотока 1280x720. Обучение завершалось после 150 эпох, после чего модели квантовались до INT8. Тестировались три конфигурации: стандартная YOLOv8n, OptiYOLO с чистыми DSC-блоками и OptiYOLO с DSC-AB блоками (табл. 1). Ключевыми метриками были FPS, GFLOPs и mAP0.75, требующая высокой точности локализации.

Таблица 1 – Сравнение праметров моделей

Модель	Params (млн)	GFLOPs	FPS	mAP0.5 (%)	mAP0.75 (%)
YOLOv8n	3.25	8.7	16.5	74.1	61.1
OptiYOLO (DSC)	1.98	4.1	28.7	73.8	58.2
OptiYOLO (DSC-AB)	2.05	4.4	27.8	75.8	66.5

Модель с DSC-AB показала минимальное увеличение сложности (4.4 GFLOPs), но привела к росту mAP0.75 на 8.3 п.п. по сравнению с чистой DSC-версией, достигнув 66.5%, что превышает результат стандартной YOLOv8n (61.1%) при более чем двукратной экономии GFLOPs. Сходимость функции потерь на графике показала, что модель DSC-AB сходится быстрее и к более низкому значению потерь, что отражает более эффективное обучение за счет внимания. Анализ точности подтвердил, что прирост mAP0.75 (8.3 п.п.) значительно превосходит прирост mAP0.5 (2.0 п.п.), доказывая, что CAM сфокусированно улучшает именно геометрическую точность. Визуальная инспекция подтвердила, что рамки, предсказанные моделью с DSC-AB, плотнее прилегают к контурам объектов, минимизируя ошибки позиционирования.

Валидация (рис. 2) подтвердила, что интеграция легковесного CAM в DSC-структуру успешно повышает точность локализации, сохраняя требования реального времени (27.8 FPS). Увеличение на

mAP0.75 8.3 п.п. до 66.5% доказывает, что механизм внимания эффективно компенсирует потерю геометрической информации в DSC-блоках. При этом модель OptiYOLO (DSC-AB) превосходит по локализации более тяжелую YOLOv8n, будучи компактнее по операциям. Устойчивость CAM к INT8-квантованию также подтверждена, поскольку адаптивное взвешивание стабилизирует информационный поток. Грамотное распределение минимальных вычислительных ресурсов на селективные механизмы внимания является оптимальным решением для достижения баланса между скоростью и геометрической точностью на периферийных устройствах.

Рисунок 2 – Сравнение результатов детекции в реальном времени

Таким образом, проведенное исследование успешно продемонстрировало, что интеграция легковесных модулей канального внимания в архитектуру на основе глубинных сепарабельных сверток позволяет достичь значительного улучшения точности локализации критически важных объектов железнодорожной инфраструктуры, увеличив метрику mAP0.75 на 8.3 п.п. по сравнению с базовой DSC-моделью, при этом сохраняя высокую производительность (27.8 FPS) на встраиваемой платформе Raspberry Pi 5. Этот гибридный подход целенаправленно компенсирует недостаток кросс-канальной связи в DSC, обеспечивая наилучший баланс между вычислительной эффективностью и геометрической точностью детекции, что необходимо для робастного мониторинга.

Список литературы:

1. Сацюк А.В. Оптимизация архитектуры YOLOv8 для задач захвата объекта БПЛА: анализ компромисса между точностью, скоростью и вычислительными ресурсами // Вестник Ростовского государственного университета путей сообщения. – 2025. – № 2(98). – С. 35–42. – DOI: 10.46973/0201-727X_2025_2_35. – EDN LCAOOM.
2. Сацюк А.В. Мониторинг инфраструктуры на основе искусственного интеллекта // Автоматика, связь, информатика. – 2025. – № 9. – С. 32–34. – DOI: 10.62994/AT.2025.9.9.005.
3. Chollet F. Xception: Deep learning with depthwise separable convolutions // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – P. 1251–1258.
4. Сацюк А.В., Володарец Н.В. Модификация модели YOLO для гибридной системы детекции и трекинга в БПЛА с автоматическим наведением // Информационно-управляющие системы. – 2025. – № 4(137). – С. 36–44. – DOI: 10.31799/1684-8853-2025-4-36-44.
5. Mobilenetv2: Inverted residuals and linear bottlenecks / M. Sandler [et al.] // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – P. 4510–4520.
6. Сацюк А.В. Проектирование адаптивной архитектуры сверточной нейронной сети с использованием глубинных сепарабельных сверток для работы в реальном времени на встраиваемых устройствах // Информационно-управляющие системы. – 2026. – № 1(138). – DOI: 10.31799/1684-8853-2026-1 (в печати).

7. Hu J., Shen L., Sun G. Squeeze-and-excitation networks // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – C. 7132–7141.

List of references:

1. Satsyuk A.V. Optimization of the YOLOv8 architecture for UAV object capture tasks: analysis of the trade-off between accuracy, speed, and computing resources // Bulletin of the Rostov State Transport University. – 2025. – № 2 (98). – P. 35–42. – DOI: 10.46973/0201-727X_2025_2_35. – EDN LCAOOM.
2. Satsyuk A.V. Infrastructure monitoring based on artificial intelligence // Automation, communication, informatics. – 2025. – № 9. – P. 32–34. – DOI: 10.62994/AT.2025.9.9.005.
3. Chollet F. Xception: Deep learning with depthwise separable convolutions // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – P. 1251–1258.
4. Satsyuk A.V., Volodarets N.V. Modification of the YOLO model for a hybrid detection and tracking system in a UAV with automatic guidance // Information and Control Systems. – 2025. – № 4(137). – P. 36–44. – DOI: 10.31799/1684-8853-2025-4-36-44.
5. Mobilenetv2: Inverted residuals and linear bottlenecks / M. Sandler [et al.] // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – P. 4510–4520.
6. Satsyuk A.V. Design of an adaptive architecture of a convolutional neural network using depthwise separable convolutions for real-time operation on embedded devices // Information and control systems. – 2026. – № 1(138). – DOI: 10.31799/1684-8853-2026-1 (in press).
7. Hu J., Shen L., Sun G. Squeeze-and-excitation networks // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – P. 7132–7141.