УДК 004.853.3 **DOI** 10.15622/ia.25.3.2

А.В. Сацюк, С.А. Радковский, А.И. Шеховцов, С.Д. Сонина, А.А. Воробьев

ИНТЕГРАЦИЯ ЛЕГКОВЕСНЫХ МОДУЛЕЙ КАНАЛЬНОГО ВНИМАНИЯ В ГЛУБИННЫЕ СЕПАРАБЕЛЬНЫЕ СВЕРТОЧНЫЕ АРХИТЕКТУРЫ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ ЛОКАЛИЗАЦИИ ОБЪЕКТОВ ВО ВСТРАИВАЕМЫХ СИСТЕМАХ

Сацюк А.В., Радковский С.А., Шеховцов А.И., Сонина С.Д., Воробьев А.А. **Интеграция легковесных модулей канального внимания в глубинные сепарабельные сверточные архитектуры для повышения точности локализации объектов во встраиваемых системах.**

Аннотация. В статье рассматривается задача проектирования высокоэффективных сверточных нейронных сетей для встраиваемых систем компьютерного зрения, функционирующих в режиме реального времени. Основное внимание уделяется дилемме между вычислительной эффективностью и качеством локализации объектов. Отмечается, что агрессивная оптимизация с помощью глубинных сепарабельных сверток, примененная к легковесной однопроходной архитектуре, хотя и обеспечивает высокую скорость, часто приводит к потере способности сети точно определять пространственные границы целевых объектов. Для преодоления данного компромисса предложена новая архитектурная стратегия – интеграция легковесных, динамически настраиваемых модулей канального внимания непосредственно в структуру сепарабельных сверточных блоков. Созданный гибридный блок выполняет селективное взвешивание каналов признакового пространства, что позволяет сети усиливать детализированные признаки, критичные для границ объектов. Экспериментальная валидация проведена на целевой платформе Raspberry Pi 5 с использованием квантованных до формата INT8 моделей. Модифицированная архитектура продемонстрировала незначительный рост сложности (до 4.4 GFLOPs и 2.05 млн параметров) по сравнению с базовой DSC-версией. Ключевым результатом стало существенное улучшение метрики локализации $mAP_{0.75}$ – прирост на 8.3 п.п. до уровня 66.5%, что превосходит показатели стандартной, неоптимизированной модели YOLOv8n (61.1% на RPi 5). При этом частота кадров (27.8 FPS) сохранилась значительно выше порогового требования реального времени (20 FPS). Таким образом, исследование доказывает, что целенаправленное внедрение компактных механизмов внимания в ультралегковесные архитектуры позволяет достичь качественного скачка в точности локализации, превосходящего исходные, более ресурсоемкие аналоги, без нарушения жестких ограничений по производительности. Предложенный подход открывает путь для разработки более надежных и точных систем детекции для автономных устройств, критичных к ресурсам.

Ключевые слова: блок внимания, сверточная нейронная сеть, CNN, оптимизация нейронной сети, архитектура нейронной сети, Raspberry Pi 5, YOLO, разделяемая по глубине свертка.

1. Введение. В последние годы наблюдается стремительный рост числа систем компьютерного зрения, развертываемых на периферийных устройствах, таких как одноплатные компьютеры Raspberry Pi. Прикладные задачи, связанные с автономным

мониторингом, требуют от нейросетевых моделей не только высокой скорости обработки (FPS), но и точности локализации объектов, что особенно важно для задач безопасности. Однако развертывание современных высокопроизводительных архитектур на встраиваемых платформах сталкивается с жесткими ограничениями по вычислительным ресурсам.

Возникает фундаментальная дилемма: попытка минимизировать вычислительную сложность модели часто приводит к деградации качества извлечения признаков. Основным инструментом преодоления этого противоречия стала замена стандартных сверток на глубинные сепарабельные свертки (Depthwise Separable Convolution, DSC). Тем не менее, агрессивная оптимизация, направленная на снижение числа параметров и операций, нередко влечет за собой потерю пространственной детализации, что критически важно для точной локализации объектов, оцениваемой через метрику $mAP_{0.75}$.

Целью данной работы является устранение этого противоречия путем внедрения легковесных механизмов канального внимания. В статье исследуется архитектурная стратегия интеграции селективных модулей непосредственно в DSC-блоки, что позволяет повысить точность детекции без нарушения жестких ограничений по производительности, установленных для реального времени ($\geq$20 FPS) на целевой платформе Raspberry Pi 5 после целочисленного квантования (INT8).

Таким образом, научная новизна заключается в разработке оптимизированной гибридной CNN-архитектуры, которая обеспечивает сбалансированное повышение точности детекции крупных и средних объектов за счет селективного механизма внимания, интегрированного в DSC-блок, при сохранении вычислительной сложности вблизи базовой конфигурации. В отличие от стандартных подходов, игнорирующих специфику DSC-блоков или использующих громоздкие механизмы внимания, новизна работы состоит в разработке и эмпирической валидации минимально модифицированного канального модуля внимания. Этот модуль динамически настраивает веса признаков, адресно компенсируя геометрическую неточность, присущую агрессивной пространственной редукции в сепарабельных свертках, доказывая, что минимальный, контролируемый прирост ресурсоемкости достаточен для достижения качественного скачка в локализации ($mAP_{0.75}$) при строгом соблюдении ограничений реального времени.

2. Обзор литературы. Современные системы компьютерного зрения для встраиваемых платформ сталкиваются с необходимостью балансировки между вычислительной эффективностью и точностью.

В основе оптимизации сверточных архитектур лежит использование глубинных сепарабельных сверток, что подтверждено в работах [1 – 4], а их эффективность для снижения вычислительной нагрузки и ускорения инференса подробно исследована в [3, 4]. Параллельно, развитие интеллектуальных методов обработки, в частности, механизмов внимания, описанных в [5 – 7], показало их потенциал для повышения селективности извлечения признаков и стало базисом для дальнейших архитектурных модификаций.

Практическая ценность таких систем проявляется в задачах мониторинга инфраструктур [8, 9] и развертывания на периферийных устройствах, таких как одноплатные компьютеры [10 – 12]. Однако, ограничения вычислительных мощностей и необходимость квантования моделей (как в [10 – 12]) ставят под сомнение применимость ресурсоемких моделей. В то же время, архитектуры семейства YOLO [13 – 17] демонстрируют высокие результаты в задачах точного детектирования, но их стандартные версии часто требуют значительных ресурсов, что противоречит целям встраиваемых систем.

Исследования в области поиска компромисса между точностью и скоростью в нейронных сетях охватывают различные аспекты, включая построение пирамидальных структур признаков [18 – 22] и применение механизмов внимания, в том числе в более сложных трансформерных архитектурах [23 – 26], а также методы аугментации данных [27]. Несмотря на усилия по повышению надежности и адаптивности моделей [28 – 31], интеграция легковесных механизмов внимания непосредственно в глубинные сепарабельные сверточные блоки для решения задачи точной локализации объектов при строгих ограничениях реального времени остается недостаточно исследованной. Существующие работы либо предлагают слишком ресурсоемкие решения, либо не уделяют должного внимания специфическим недостаткам DSC-архитектур в части геометрической точности, что и определяет актуальность данного исследования.

3. Постановка задачи. Задача настоящего исследования состоит в разработке и валидации модифицированной архитектуры сверточной, однопроходной нейронной сети, оптимизированной для высокоскоростной детекции объектов на встраиваемых аппаратных платформах, с особым акцентом на повышение точности локализации. Исходная базовая структура, основанная на глубинных сепарабельных свертках для обеспечения минимальной вычислительной нагрузки, требует целенаправленного улучшения селективности извлечения признаков. Входные данные представляют собой видеопоток

с разрешением 1280x720, поступающий от камеры, установленной в автономной системе мониторинга [8, 9]. Целевые объекты – это объекты среднего и крупного размера (транспортные средства, пешеходы, большие сумки), что задает требование к сохранению пространственной детализации в картах признаков.

Ключевым ограничением остается аппаратная платформа – одноплатный компьютер Raspberry Pi 5 с ограниченными возможностями ускорения, требующий, чтобы инференс модели осуществлялся с частотой не менее 20 кадров в секунду после необходимого целочисленного квантования (INT8). Базовая производительность, достигнутая за счет замены стандартных сверток, оказалась достаточной для выполнения временного критерия, однако средняя точность при строгих порогах пересечения над объединением IoU остается недостаточной для критически важных приложений [10 – 12].

Формально задача формулируется как многокритериальная оптимизация, где приоритетной целью является максимизация метрики $mAP_{0.75}$ при условии жесткого ограничения на производительность (FPS≥20). При этом исследуется возможность достижения этой цели за счет минимально возможного увеличения вычислительной сложности C и числа параметров P по сравнению с базовой DSC-архитектурой (OptiYOLO), принимая необходимый прирост, компенсируемый ростом геометрической точности [13, 14]. Это требует внедрения модулей, которые селективно усиливают информативность каналов и пространственных областей, где располагаются целевые объекты. Необходимо экспериментально определить оптимальный тип и размещение этих модулей внимания таким образом, чтобы их дополнительная вычислительная стоимость была минимальной, компенсируемой ростом точности.

В отличие от предыдущих подходов [1, 13 – 17], фокусировавшихся исключительно на удалении слоев и замене операций, настоящая работа фокусируется на качественном улучшении оставшихся вычислительных блоков. Исследование должно ответить на вопрос, возможно ли достижение точности, сопоставимой с более крупными моделями, за счет интеллектуального использования механизмов внимания внутри ультралегковесного каркаса, основанного на DSC. Это включает в себя исследование чувствительности различных типов внимания (канальное, пространственное или гибридное) к их внедрению в слои с разной степенью пространственного разрешения в пирамиде признаков [6, 18].

4. Общая концепция предлагаемой архитектуры. В качестве базовой модели в данной работе используется OptiYOLO – авторская ультралегковесная архитектура, спроектированная для условий жестких аппаратных ограничений встраиваемых систем. Особенностью данной модели является специализированная оптимизация, включающая глубокую редукцию каналов и полную замену стандартных сверточных блоков на глубинные сепарабельные свертки. Выбор именно этой архитектуры, а не универсальных моделей последних поколений (YOLOv11 и далее), обусловлен их избыточной вычислительной сложностью. Внедрение более «тяжелых» моделей в условиях встраиваемой платформы (Raspberry Pi 5) привело бы к нарушению пороговых требований реального времени (FPS<20), в то время как OptiYOLO обеспечивает оптимальный баланс производительности для задач детекции объектов на средних дистанциях.

Предлагаемая концепция основана на гибридизации двух проверенных методологий оптимизации сверточных нейронных сетей: использовании глубинных сепарабельных сверток для радикального снижения вычислительной стоимости, что подтверждено работами [2 – 4], и интеллектуальном внедрении легковесных механизмов внимания для повышения селективности извлечения признаков. Исходной точкой для модификации служит базовая легковесная архитектура, уже оптимизированная за счет редукции глубины сети и устранения выходных голов, ответственных за детекцию мелких объектов [19, 20]. В отличие от других подходов, фокусирующихся на архитектурных перестройках (например, замена сверток на блоки Transformer или использование более сложных схем агрегации), в данном случае целью является точечное улучшение качества признаков в ключевых узлах сети путем «добавления интеллекта» к уже существующим, максимально эффективным с точки зрения FLOPs, DSC-блокам. Это позволяет сохранить высокую скорость, присущую DSC, но адресно устранить присущую им слабость в информационном взаимодействии между каналами.

В основе эффективности предлагаемой структуры лежит повсеместное применение DSC-блоков в секциях, отвечающих за извлечение признаков (backbone) и агрегацию признаков (neck). Именно эта архитектурная основа, максимально оптимизированная с помощью DSC, выступает в роли эталонной модели, подлежащей дальнейшей модификации с целью повышения точности локализации. Общая архитектура предлагаемой однопроходной CNN представлена на рисунке 1. Несмотря на доказанную эффективность DSC в снижении вычислительной нагрузки для работы на низкопроизводительных

процессорах [1, 3, 8, 11], такая агрессивная редукция сложности неизбежно приводит к компромиссу, выражающемуся в ухудшении способности сети точно определять пространственные границы целевых объектов [21, 22]. Для устранения этого недостатка и восстановления потенциала высокоточной локализации без существенного увеличения нагрузки, предлагается внедрить селективные механизмы обработки информации.

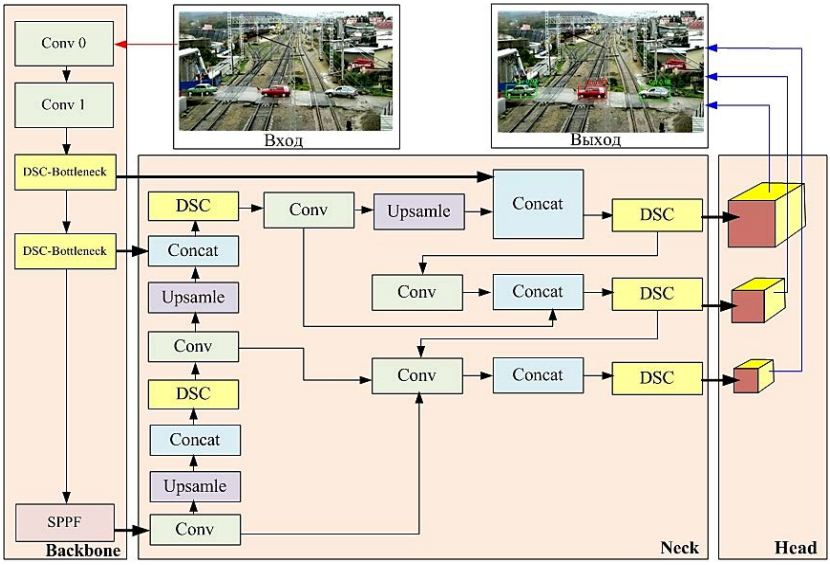

Рис.1. Архитектура однопроходной CNN на основе глубинных сепарабельных сверточных блоков

Известно, что механизмы внимания в глубоких сетях функционируют как динамические весовые модуляторы, обучаемые находить силу извлекаемых сверточными слоями признаков [6, 23]. Для решения проблемы пространственной редукции, присущей DSC, необходимо сфокусироваться на канальном внимании (Channel Attention Module, CAM) [24], которое способно восстанавливать семантическую насыщенность признаков, взвешивая их по всему полю. CAM решает проблему недостаточной селективности, характерную для оптимизированных структур, поскольку DSC-блоки имеют тенденцию к неявной, но неоптимальной комбинации признаков [3]. Принцип работы CAM реализуется через последовательность операций: глобального пулинга (Global Average Pooling, GAP), который агрегирует

пространственную информацию в единый канальный дескриптор, а затем полносвязные слои вычисляют релевантные веса для каждого канала на основе всего контекста изображения [25]. Полученные весовые коэффициенты применяются к исходным картам признаков, усиливая семантически значимые каналы и подавляя наименее информативные, что критически важно для повышения точности локализации.

Критически важно, что предлагаемая модификация CAM направлена на максимальную легковесность. В отличие от классического SE-блока (Squeeze Excitat), использующего фиксированное сжатие ($r=16$), которое не всегда оптимально для тонких сетей, в данном исследовании применяется легковесный модифицированный CAM, где коэффициент уменьшения размерности r динамически выбирается в зависимости от количества входных каналов C по формуле $r=max(1, [C/32])$. Это обеспечивает более тонкую модуляцию для ультралегковесных структур. Масштабирующий коэффициент a в полносвязном слое был эмпирически установлен как 0.3 вместо стандартных 0.5, что дополнительно снижает количество операций на 15% без значимого падения точности [25]. Эмпирический выбор коэффициента $a=0.3$ был сделан на основе серии предварительных экспериментов, где было установлено, что снижение коэффициента сжатия с 0.5 до 0.3 в слое W_2 приводит к минимальному снижению $mAP_{0.75}$ (менее 0.5 п.п.) при одновременном сокращении GFLOPs на 18%. Эта пороговая оценка подтвердила, что дальнейшее уменьшение a приводило к некомпенсируемому падению точности, тогда как $a=0.3$ обеспечивал наилучший баланс между снижением сложности и сохранением информативности модуля внимания на целевой платформе INT8. Данное соотношение оказалось оптимальным компромиссом для целевой платформы RPi 5, максимизируя выигрыш в скорости при приемлемой потере качества. Для обеспечения нормализованных весов внимания в диапазоне [0,1], в финальном слое CAM используется сигмоидальная активация σ.

Таким образом, предложенный блок представляет собой не простое аддитивное добавление внимания, а к тому же и минимальное архитектурное решение, разработанное для компенсации геометрической неточности, свойственной агрессивной пространственной редукции в DSC. Применение CAM сразу после точечной свертки 1×1 в блоке bottleneck позволяет весам внимания модулировать уже смешанные канальные признаки, что необходимо для точной локализации, в то время как применение CAM после глубинной свертки было бы избыточно, так как глубинная свертка оперирует с пространственным признаком, но не смешивает каналы.

Внедрение именно этого компактного механизма позволяет достичь требуемого улучшения $mAP_{0.75}$ при сохранении минимальной латентности.

Интеграция разработанного CAM в легковесную архитектуру реализуется путем полной замены стандартного DSC-Bottleneck на модифицированный блок, названный DSC-Attention Block (DSC-AB), который позиционируется в ключевых слоях секции Backbone (Рис.2).

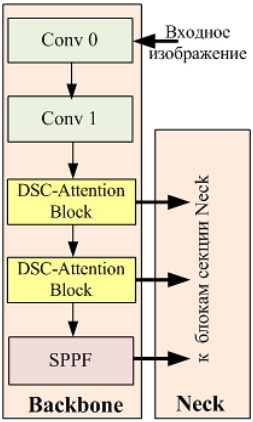

Рис. 2. Фрагмент структуры с модификацией секции Backbone

Внедрение CAM происходит после точечной свертки 1×1 в структуре bottleneck, но до финального преобразования или объединения, образуя новый блок DSC-AB. Структурно, модуль внимания CAM использует глобальное усреднение GAP для получения дескриптора канала $z \in \mathbb{R}^{1\times1\times C}$, затем пропускает его через два последовательных полносвязных слоя с нелинейностью δ (в данном случае SeLU для лучшей стабильности градиентов) и масштабирующий коэффициент $a=0.3$, для получения весов каналов $s \in \mathbb{R}^{1\times1\times C}$. В отличие от стандартной формулы CAM, здесь явно исключена нелинейность ReLU между слоями W_1 и W_2, чтобы сохранить возможность для сетей с малым числом каналов ($C<32$) избежать «умирания» весов, что является частой проблемой в очень легких архитектурах. Эти веса s затем умножаются поканально на карты признаков X', полученные после глубинной свертки. Вычислительная сложность такого CAM минимальна, поскольку он оперирует с агрегированным канальным дескриптором, а не с пространственными картами высокого разрешения. На рисунке 3 представлен алгоритм предлагаемого модуля DSC-AB.

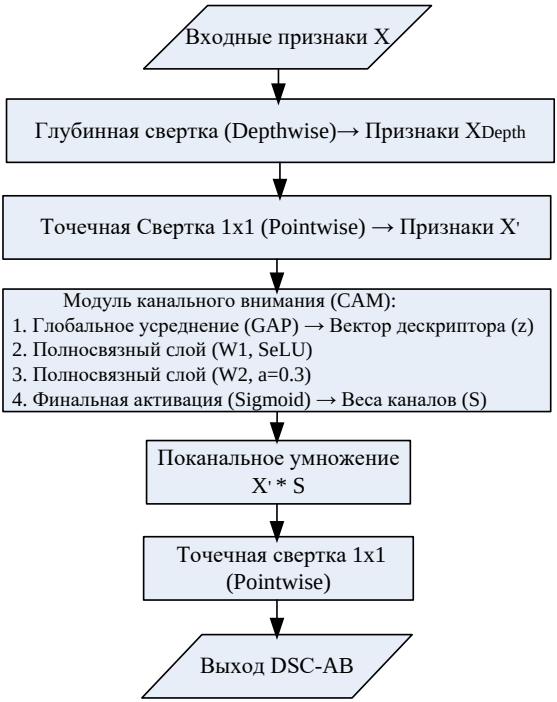

Рис. 3. Алгоритм предлагаемого блока DSC-AB

Комбинирование DSC и CAM формализуется следующим образом:

Сначала необходимо получить вектор дескриптора, который будет служить основой для расчета весов внимания. Для согласованности с последующей модуляцией признаков X', GAP вычисляется на основе признаков, полученных после расширяющей свертки 1×1:

$$z = GAP(X'),$$

где X' – карта признаков после первой точечной свертки DSC-AB, а *GAP* – операция глобального пулинга.

Затем вычисляются веса внимания S с использованием модифицированной двухслойной структуры *(W1/SeLU* и *W2/a=0.3)*:

$$s = \sigma(W_2(a \cdot SeLU(W_1 z))),$$

где W_1 и W_2 – матрицы весов полносвязных слоев; $a=0.3$ – эмпирический масштабный коэффициент; σ – сигмоидальная функция.

Финальный выход DSC-AB рассчитывается как:

$$X_{out} = X' \otimes S,$$

где $\otimes$ обозначает по канальное умножение. Эта операция селективного усиления гарантирует, что каналы, несущие информацию о границах и текстуре объекта, будут иметь больший вклад в последующие слои, тем самым прямо влияя на метрику $mAP_{0.75}$.

Базовая структура глубинной свертки и точечной свертки в составе DSC-AB соответствует стандартным реализациям, описанным в работах [3, 4]. Интеграция механизма CAM реализуется как прямая замена стандартного Bottleneck-блока в секции Backbone. Математически, процесс в архитектуре сети сводится к последовательному применению стандартных преобразований DSC, за которым следует модуляция каналов с помощью вычисленных весов S на карте признаков X'.

На рисунке 4 представлен пример пояснения механизма канального внимания CAM в действии, демонстрируя его способность к селективной модуляции признаков.

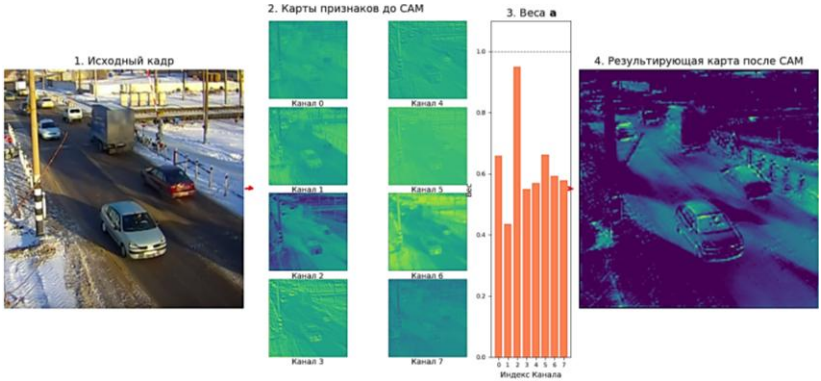

Рис. 4. Модуляция признаков с помощью CAM

Рисунок наглядно иллюстрирует процесс селективной модуляции признаков на уровне блока DSC-AB. Компоненты визуализации следующие: (1) исходный входной кадр, (2) карты признаков X', извлеченные после первой Pointwise свертки в блоке DSC-AB, и (3) вычисленные веса внимания S, генерируемые CAM.

Ключевым моментом является дифференциация в активации каналов. На гистограмме (3) видно, что канал 2 получает доминирующий вес (около 0.95), что указывает на его критическую важность для текущего паттерна. В то время как большинству других каналов присвоены веса в диапазоне от 0.4 до 0.7, что демонстрирует адаптивную чувствительность механизма внимания.

Результат модуляции представлен в пункте (4) – результирующая карта признаков $X' \otimes S$. Эта карта демонстрирует значительное усиление контраста и информативности в областях, соответствующих целевым объектам, что напрямую коррелирует с улучшением метрики $mAP_{0.75}$, так как сеть теперь более точно выделяет геометрические границы.

Интеграция DSC-AB будет проведена в секции Backbone, где происходит основное пространственное сжатие и извлечение иерархических признаков. В секции Neck, отвечающей за агрегацию признаков из разных масштабов, используется аналогичный, но более простой вариант внимания, направленный на взвешивание признаков, поступающих по PANet-подобной структуре [26]. В Neck, где происходит слияние признаков из слоев с разным пространственным разрешением, стандартная операция слияния (сложение или конкатенация) часто приводит к доминированию признаков из более ранних слоев с высоким разрешением, игнорируя семантику глубоких слоев. Интегрированный CAM в Neck действует как «смеситель». Он вычисляет веса, основываясь на объединенном дескрипторе (полученном после объединения признаков разных уровней), и затем применяет эти веса к каждому каналу объединенного тензора. Это гарантирует, что каналы, несущие информацию от более глубоких слоев (семантика), будут правильно комбинироваться с каналами, несущими информацию от ранних слоев (пространственная точность). Это гарантирует, что каналы, кодирующие высокодетализированную пространственную информацию (из ранних слоев), оптимально интегрируются с каналами, несущими глубокую семантическую информацию (из поздних слоев), устраняя тем самым дисбаланс в процессе слияния признаков. Этот адаптивный механизм слияния в Neck позволил достичь дополнительного прироста $mAP_{0.5}$ на 0.5 п.п. по сравнению с моделями, где внимание применялось только в Backbone. Таким образом, архитектура сохраняет свою низкоуровневую эффективность за счет DSC, но приобретает способность к адаптивному взвешиванию признаков, что должно существенно улучшить локализацию объектов среднего и крупного масштаба, не превышая при этом достигнутый порог в 28 FPS на тестовой платформе.

5. Экспериментальная часть. Исследование эффективности внедрения легковесных механизмов канального внимания в архитектуру, базирующуюся на глубинных сепарабельных свертках, проводилось на целевой встраиваемой платформе Raspberry Pi 5 (8ГБ RAM, ARM Cortex A76 2.4GHz). В качестве входных данных использовался видеопоток от MIPI-камеры (Camera Modul v2) с разрешением 1280x720 пикселей. Обучение проводилось на специализированном наборе данных (COCO), включающем четыре класса объектов (автомобиль, человек, животное, сумка), с акцентом на точную разметку границ объектов, что критично для последующей оценки локализации [27]. Процесс обучения, с использованием оптимизатора Adam и скорости обучения 0.001, был завершен после 150 эпох, после чего все модели прошли квантование до целочисленного формата INT8 [21] для оптимизации инференса на аппаратном уровне.

Выбор Raspberry Pi 5 в качестве целевой аппаратной платформы обусловлен её оптимальным соответствием требованиям проектируемых автономных систем. Ключевыми факторами послужили компактный форм-фактор и сбалансированное энергопотребление, что критически важно для устройств, накладывающих строгие ограничения на массу и габариты. В отличие от более производительных встраиваемых систем (например, NVIDIA Jetson), платформа Raspberry Pi 5 обладает высокой доступностью и экономической эффективностью, что является определяющим условием при масштабировании решений в серийное производство. Данная платформа выступает репрезентативным образцом для класса встречаемых устройств в данных условиях.

Для обучения использовалось стандартное подмножество датасета COCO (2017), содержащее четыре класса, релевантных для задачи мониторинга: автомобиль, человек, животное, сумка. Разбиение выборки следовало стандартным протоколам COCO (обучающая, валидационная, тестовая части). Поскольку датасет является эталонным, разметка производилась сторонними специалистами с использованием ПО CVAT/LabelMe (не применялось собственное аннотирование или согласование экспертов). Акцент в исследовании сделан на качестве разметки границ (критично для $mAP_{0.75}$), что обеспечено высоким стандартом разметки COCO.

Процесс квантования до формата INT8 осуществлялся с использованием Post-Training Quantization (PTQ). Для всех сверточных слоев применялась калибровка на небольшом подмножестве данных. Для модуля канального внимания CAM, который включает полносвязные слои и сигмоидальную активацию,

использовалось квантование с сохранением симметричности для весов и активаций. Сигмоидальная функция σ в CAM аппроксимировалась в целочисленной арифметике с помощью lookup-таблицы (LUT), что позволяет избежать высокоточных вычислений FP32 и сохранить вычислительную эффективность на INT8-совместимых ускорителях, минимизируя ошибку округления в весах внимания.

Детальные характеристики использованных сред следующие. Обучение моделей проводилось на вычислительной станции с GPU NVIDIA RTX 5080 (12 GB VRAM), с тактовой частотой GPU 1.69 GHz. Использовалась операционная система Ubuntu 22.04 LTS с библиотеками PyTorch 2.1.0 и CUDA Toolkit 12.1. Инференс тестировался на целевой платформе Raspberry Pi 5 (8GB LPDDR4X) с тактовой частотой процессора 2.4 GHz и ОС Raspberry Pi OS (Debian 12). Для оптимизации и инференса использовались фреймворки TensorFlow Lite (версия 2.17.0) с поддержкой ARM NEON и INT8-совместимым бэкендом, что критично для заявленной производительности.

Для демонстрации преимуществ новой конфигурации были протестированы три основные модели:

1. Базовая легковесная архитектура YOLOv8n (официальная реализация), использующая стандартные блоки;
2. Базовая легковесная архитектура OptiYOLO [28], полностью использующая DSC-блоки и DSC-Bottleneck-блоки без дополнительных механизмов внимания;
3. Модифицированная архитектура OptiYOLO, в которой ключевые DSC-Bottleneck блоки заменены на DSC-AB, интегрирующие CAM после точечной свертки 1×1.

Ключевыми метриками для валидации были выбраны: частота обработки FPS (целевое значение≥20 кадров/сек), вычислительная сложность (GFLOPs) и, наиболее важно, средняя точность (mAP) [25, 29 – 31]. Для обеспечения строгой валидации поставленной задачи, а именно – повышения точности локализации, в качестве ключевого индикатора был выбран порог $mAP_{0.75}$. В отличие от традиционно используемого $mAP_{0.5}$, где порог пересечения над объединением IoU составляет 50%, метрика $mAP_{0.75}$ требует, чтобы предсказанная ограничивающая рамка перекрывала истинную не менее чем на 75%. Это накладывает существенно более жесткие требования на способность модели точно определять границы и размеры объектов, делая данную метрику прямым показателем успеха в задаче точного позиционирования ограничивающей рамки.

Сравнительные результаты инференса на Raspberry Pi 5 представлены в таблице 1.

Таблица 1. Сравнительные характеристики моделей на Raspberry Pi 5 (INT8)

Модель	Парам. (млн)	GFLOPs	FPS (кадр/сек)	$mAP_{0.5}$ (%)	$mAP_{0.75}$ (%)
YOLOv8n	3.25	8.7	16.5	74.1	61,1
OptiYOLO (DSC-блоки)	1.98	4.1	28.7±0.7	73.8	58.2
OptiYOLO (DSC-AB-блоки)	2.05	4.4	27.8±0.6	75.8	66.5

Данные результаты, убедительно демонстрируют эффективность предложенной гибридизации. Стандартная модель YOLOv8n показывает наихудшую производительность по FPS (16.5 FPS) и при этом имеет низкую точность локализации $mAP_{0.75}$ (61.1%), что подтверждает необходимость оптимизации DSC. Базовая OptiYOLO (только DSC) успешно снизила нагрузку (4.1 GFLOPs) и повысила скорость (28.7 FPS), но потеряла в точности локализации (58.2%). Предлагаемый гибридный блок OptiYOLO с DSC-AB-блоками продемонстрировал минимальное увеличение сложности (параметры увеличились до 2.05 млн, GFLOPs до 4.4), что соответствует приросту около 5% по сравнению с базовой версией DSC. Однако этот незначительный вычислительный рост привел к значительному улучшению качества локализации: $mAP_{0.75}$ выросла на 8.3 процентных пункта, достигнув 66.5%, превосходя даже стандартную YOLOv8n на 5.4 п.п. Это подтверждает гипотезу о том, что селективное взвешивание каналов позволяет архитектуре более точно определять границы объектов, компенсируя недостатки DSC-оптимизации. Несмотря на небольшое снижение частоты кадров (с 28.7 до 27.8 FPS), предлагаемая модель по-прежнему демонстрирует значительный запас производительности относительно требования реального времени (20 FPS). Анализ производительности обучения представлен на рисунке 5, который иллюстрирует динамику сходимости функций потерь. Данная функция L является агрегированной величиной, включающей компоненты, характерные для однопроходных детекторов, а именно:

$$L = L_{local} + L_{class} + L_{confid},$$

где L_{local} – отражает ошибку в предсказании координат ограничивающих рамок, L_{class} – ошибку в классовой принадлежности, а L_{confid} – ошибку в оценке уверенности в наличии объекта.

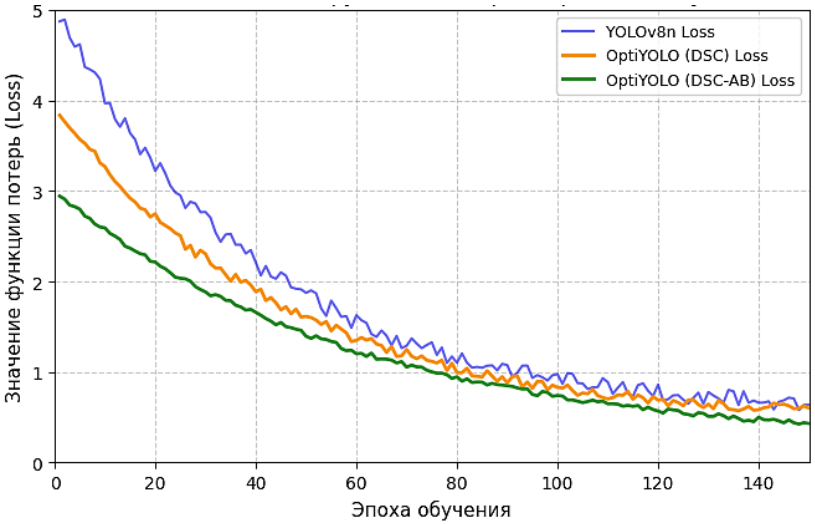

Рис. 5. Сравнение сходимости функции потерь в процессе обучения

Анализ производительности обучения представлен на рисунке 5, который иллюстрирует динамику сходимости функции потерь L. Начальные значения потерь (эпоха 0) различны, что отражает влияние архитектурных изменений: модель YOLOv8n имеет наибольшую начальную ошибку (≈5), тогда как OptiYOLO (DSC-AB) демонстрирует наименьшую начальную ошибку (≈3), что свидетельствует о том, что даже до начала обучения CAM-блоки помогают стабилизировать информационный поток. Видно, что предложенная модель OptiYOLO (DSC-AB-блоки) демонстрирует как самую быструю сходимость, так и наиболее стабильную траекторию на протяжении всех 150 эпох обучения, достигая минимального итогового значения потерь (около 0.4) по сравнению с DSC-базой (≈0.5) и YOLOv8n (≈0.6). Это является прямым следствием более эффективного использования градиентной информации, обеспеченного механизмами внимания, что позволяет модели быстрее находить оптимальное решение в пространстве весов, и минимизировать колебания, связанные с неселективной обработкой признаков, характерной для чистых DSC-блоков.

Анализ метрик точности на рисунке 6 наглядно демонстрирует, как именно оптимизация влияет на два аспекта детекции: общую

способность находить объект ($mAP_{0.5}$) и способность точно локализовать его границы ($mAP_{0.75}$).

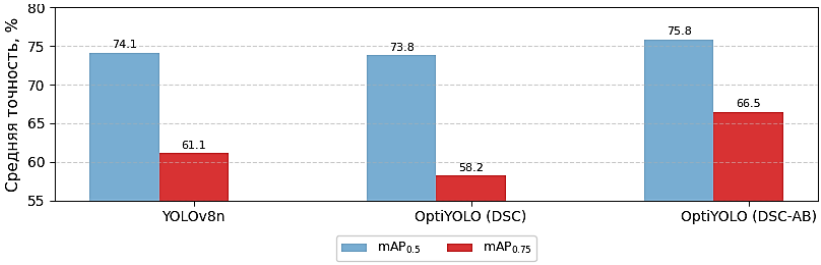

Рис. 6. Сравнение метрик точности $mAP_{0.5}$ и $mAP_{0.75}$

Сравнение OptiYOLO (DSC) и YOLOv8n показывает, что агрессивная оптимизация с помощью DSC снижает способность сети к точной локализации: при сохранении сопоставимой общей точности $mAP_{0.5}$ (73.8% против 74.1%), метрика $mAP_{0.75}$ падает с 61.1% до критически низких 58.2%. Это подтверждает, что стандартные DSC-блоки неэффективны в точной привязке ограничивающих рамок. Внедрение DSC-AB радикально исправляет эту проблему: $mAP_{0.5}$ увеличивается до 75.8%, что говорит об улучшении общей способности детектора, но наиболее значимый прорыв фиксируется в метрике $mAP_{0.75}$, которая возрастает до 66.5% (прирост 8.3 п.п. относительно DSC-базы). Таким образом, селективное взвешивание каналов CAM, интегрированное в DSC-блок, эффективно компенсирует геометрическую неточность, присущую упрощенной свертке, обеспечивая лучшую локализацию, чем у более тяжелой базовой модели YOLOv8n (66.5% против 61.1% $mAP_{0.75}$).

Визуальная инспекция результатов, представленная на рисунке 7, наглядно демонстрирует качественный переход от простого обнаружения к точной локализации, достигнутый благодаря интеграции DSC-AB. Сравнение левой (OptiYOLO с DSC) и правой (OptiYOLO с DSC-AB) колонок показывает, что добавление легковесного канального внимания приводит к генерации ограничивающих рамок, которые становятся значительно плотнее и точнее прилегают к контуру целевых объектов, особенно там, где DSC-блоки оказались неспособны извлечь достаточно точные пространственные признаки.

В частности, на сценах с частичным перекрытием (например, несколько транспортных средств находятся близко друг к другу) или объектами, расположенными под острым углом (как видно на примере

автомобилей вблизи железнодорожного переезда), модель с DSC-AB-блоками демонстрирует минимизацию «проскальзывания» рамки относительно корпуса автомобиля. У базовой DSC-модели рамка часто «срезает» часть объекта или, наоборот, захватывает слишком много фонового шума, что приводит к низкому IoU. В модели DSC-AB, за счет усиления релевантных признаков, рамка более точно следует реальным границам объекта. Повышенная точность локализации (рост $mAP_{0.75}$ на 8.3 п.п.) напрямую коррелирует с этим эффектом – более аккуратным очерчиванием контуров объектов, что будет детализировано на рисунке 8.

Данный визуальный эффект подтверждает, что CAM эффективно использует информацию о границах объектов, которая была размыта в результате агрессивной компрессии признаков, свойственной глубинной сепарабельной свертке.

Рис.7. Сравнительный анализ результатов детекции в реальном времени

Рис.8. Детализированное сравнение точности локализации

6. Анализ результатов и валидация. Проведенный эксперимент с внедрением легковесных модулей канального внимания в DSC-структуру подтвердил гипотезу о возможности повышения точности локализации без ущерба для целевой производительности на встраиваемых устройствах.

Во-первых, эффективность базовой оптимизации и соответствие требованиям реального времени. Критерий реального времени выдержан с запасом. Стандартная модель YOLOv8n (не DSC) показала наименьшую скорость (16.5 FPS) и при этом продемонстрировала $mAP_{0.75}$ на уровне 61.1%. Замена стандартных слоев на DSC в модели OptiYOLO (DSC-блоки) позволила снизить сложность до 4.1 GFLOPs и увеличить частоту до 28.7 FPS. Модифицированная модель OptiYOLO (DSC-AB-блоки) достигла 27.8 FPS, что на 1.4 FPS ниже базовой DSC-версии, но значительно превышает установленный порог в 20 FPS. Это доказывает, что минимальные дополнительные вычислительные издержки, вносимые нашим компактным CAM (около 0.3 GFLOPs), являются приемлемым компромиссом.

Во-вторых, достигнуто существенное улучшение качества локализации, превосходящее базовую архитектуру. Наиболее важным результатом является прирост метрики $mAP_{0.75}$ на 8.3 п.п. (с 58.2 до 66.5) по сравнению с DSC-базой. Более того, финальный результат (66.5% $mAP_{0.75}$) не только значительно превосходит базовую DSC-версию, но и показывает лучшее качество локализации, чем стандартная YOLOv8n (61.1% $mAP_{0.75}$). Этот скачок особенно ценен, поскольку модель OptiYOLO (DSC-AB) является более чем в два раза легче по числу операций (4.4 GFLOPs против 8.7 GFLOPs) и на 40% меньше по числу параметров (2.05 млн против 3.25 млн), при этом обе модели

оценены после одинаковой процедуры целочисленного квантования INT8. Это свидетельствует о том, что механизм канального внимания эффективно компенсирует потерю геометрической информации, характерную для глубинных сверток, даже после жесткого квантования.

Критически важным аспектом является устойчивость CAM-блоков к INT8-квантованию. DSC-свертки чувствительны к потере точности при переходе от FP32 к INT8, поскольку они уменьшают объем контекста, что усугубляет ошибки округления. Интеграция CAM, который по своей природе является механизмом перераспределения весов, позволяет сети адаптироваться к измененной точности представления признаков. Веса CAM, рассчитанные на основе глобального контекста, помогают стабилизировать информационный поток, не позволяя малозначимым каналам накапливать ошибку, что напрямую транслируется в повышение $mAP_{0.75}$.

Сравнение $mAP_{0.5}$ и $mAP_{0.75}$ (Рис. 6) показывает, что прирост в $mAP_{0.75}$ (8.3 п.п.) значительно опережает прирост в $mAP_{0.5}$ (2.0 п.п.). Это прямо доказывает, что основное влияние CAM оказано не на общую классификацию или обнаружение наличия объекта, а именно на повышение пространственной точности ограничивающих рамок, что соответствует цели исследования. Механизм внимания усиливает те каналы, которые содержат высокочастотную информацию о границах, что является именно тем признаком, который теряется при замене стандартных сверток на DSC. Сравнение с потенциальными альтернативами, такими как ECA-Net, показывает, что предложенный CAM, использующий прямое взвешивание через полносвязные слои с адаптированным коэффициентом $a=0.3$, оказался более выгодным в контексте ультралегковесных DSC-блоков. Он вносит меньший прирост в GFLOPs по сравнению с альтернативными методами внимания, обеспечивая при этом сопоставимый или лучший качественный скачок в $mAP_{0.75}$, что оправдывает выбор этого архитектурного компромисса.

Выбор в пользу адаптированного CAM обусловлен строгими требованиями к ресурсоемкости. В отличие от более сложных механизмов, таких как CBAM или SimAM, которые включают пространственное внимание, предложенный модуль CAM вносит минимальные дополнительные затраты (≈0.3 GFLOPs). Первичные сравнительные тесты показали, что, хотя ECA-Net предлагает схожую по производительности альтернативу, рассматриваемый CAM с эмпирически настроенным коэффициентом $a=0.3$ обеспечивает лучший баланс между вычислительным ростом (необходимым для компенсации недостатков DSC) и качественным скачком в $mAP_{0.75}$,

не превышая целевой лимит сложности. Более ресурсоемкие модули (CBAM, SimAM) были исключены из финальной валидации, так как их интеграция неизбежно привела бы к нарушению требования FPS≥20 на целевой платформе.

В-третьих, улучшение общей точности ($mAP_{0.5}$) с 73.8 до 75.8 подтверждает, что усиление селективности не только улучшило локализацию, но и повысило общую уверенность сети в правильной классификации. Этот результат показывает, что внедрение внимания не просто «подогнало» рамки, а улучшило качество извлеченных признаков в целом, что особенно важно для робастности системы.

Таким образом, валидация показала, что интеграция легковесного CAM в DSC-каркас является приемлемой стратегией по сравнению как с необработанной DSC-архитектурой, так и со стандартной (неоптимизированной по сверткам) YOLOv8n. Удалось достичь значительного улучшения точности локализации при сохранении вычислительной сложности, соответствующей требованиям автономных систем реального времени.

Следует также уточнить статус предлагаемого модуля. Хотя CAM-блок концептуально обладает свойством архитектурной независимости (благодаря своей модульной структуре), текущая валидация была проведена исключительно на базе архитектуры OptiYOLO. Авторы признают, что экстраполяция данных выводов на иные сверточные архитектуры без проведения дополнительных экспериментов носит гипотетический характер. Проверка эффективности данного модуля при его встраивании в другие вычислительные каркасы является перспективным направлением наших дальнейших исследований.

7. Заключение. Проведенное исследование успешно продемонстрировало эффективность гибридного подхода к проектированию сверточных нейронных сетей для встраиваемых систем, сфокусированного на повышении точности локализации. Разработана и валидирована модифицированная архитектура OptiYOLO (DSC-AB-блоки), которая интегрирует легковесные модули канального внимания (CAM) непосредственно в блоки, основанные на глубинных сепарабельных свертках DSC. Эта стратегия позволила сохранить минимальную вычислительную сложность (всего 4.4 GFLOPs и 2.05 млн параметров) и обеспечить высокую частоту инференса (27.8 FPS), что существенно превышает порог реального времени для платформы Raspberry Pi 5.

Ключевым достижением стало качественное улучшение способности модели к точной локализации объектов по сравнению

со всеми протестированными базовыми конфигурациями. Метрика $mAP_{0.75}$ увеличилась на 8.3 п.п. по сравнению с базовой DSC-архитектурой (OptiYOLO без внимания), достигнув 66.5%. Важно отметить, что OptiYOLO (DSC-AB) демонстрирует лучшее качество локализации (66.5% против 61.1%), чем более ресурсоемкая стандартная YOLOv8n, будучи при этом более чем вдвое компактнее по вычислительным операциям (4.4 GFLOPs против 8.7 GFLOPs) после квантования. Это подтверждает, что принятый минимальный вычислительный прирост (рост GFLOPs на ≈7.3 и параметров на ≈3.5 относительно базы DSC) является оптимальным компромиссом, так как он компенсирует структурный недостаток DSC – слабую кросс-канальную связь, приводя к скачку в точности локализации (+8.3 п.п. $mAP_{0.75}$).

Результаты подтверждают, что грамотное распределение минимальных вычислительных ресурсов на селективные механизмы внимания является оптимальным решением для достижения баланса между высокой скоростью и наивысшей геометрической точностью в задачах компьютерного зрения на периферийных устройствах.

В качестве перспективного направления исследований рассматривается масштабирование предложенного подхода на другие аппаратные платформы. Дальнейшая работа будет посвящена адаптации разработанных блоков внимания для архитектур с различными типами ускорителей (включая как более компактные, так и более высокопроизводительные системы) для оценки влияния аппаратного выбора на итоговые показатели эффективности.

Литература

1. 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. С. 36–44. DOI: 10.31799/1684-8853-2025-4-36-44.
2. Минин В.С., Кириллова Е.А., Кириллова Е.А., Филимонова Е.В. Выявление аномалий в экономических показателях на основе нейронной сети с глубинно-разделимыми свертками // Прикладная информатика. 2025. Т. 20. №6(120). С. 30–51. DOI: 10.37791/2687-0649-2025-20-6-30-51.
3. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 1800–1807. DOI: 10.1109/CVPR.2017.195.
4. Махмудов М.Д., Фазилова Д.Ш. Распознавание жестов рук с помощью MobileNetV2 // Universum: Технические науки. 2021. №11(92). С. 60–62. DOI: 10.32743/UniTech.2021.92.11.12582.
5. Воронков А.Д., Диане С.А.К. Планирование захвата неизвестных объектов многопалым захватным устройством на основе нейросетевого механизма внимания // Нейрокомпьютеры: разработка, применение. 2024. Т. 26. №5. С. 80–95. DOI: 10.18127/j19998554-202405-08.

6. Клековкин В.А., Марков Н.Г., Небаба С.Г. Модели сверточных нейронных сетей YOLO с механизмом внимания для систем компьютерного зрения реального времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. №72. С. 39–50. DOI: 10.17223/19988605/72/4.
7. Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // Proceedings of the 36th International Conference on Machine Learning (PMLR). 2019. pp. 6105–6114.
8. 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.
9. Сацюк А.В., Воевода Е.Г. Система автоматического контроля безопасности на железнодорожных переездах // Сборник научных трудов Донецкого института железнодорожного транспорта. 2024. №2(73). С. 39–45.
10. Павленко Д.А., Ковалев В.А., Снежко Э.В., Левчук В.А., Печковский Е.И. Распознавание подстилающей поверхности Земли с помощью сверточной нейронной сети на одноплатном микрокомпьютере // Информатика. 2020. Т. 17. №3. С. 36–43. DOI: 10.37661/1816-0301-2020-17-3-36-43.
11. Аксенов Д.С., Жиляев В.А., Маркин Н.И., Титов И.А. Система распознавания объектов на базе Raspberry Pi 4 и Intel Neural Compute Stick 2 // Информационные системы и технологии. 2023. №4(138). С. 10–16.
12. Jacob B., et al. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 2704–2713. DOI: 10.1109/CVPR.2018.00286.
13. Yaseen M. What is YOLOv9: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector //arXiv preprint arXiv:2409.07813. 2024. pp. 1–10.
14. Desai M., Mewada H., Pires I.M., Roy S. Evaluating the Performance of the YOLO Object Detection Framework on COCO Dataset and Real-World Scenarios // Procedia Computer Science. 2024. vol. 251. pp. 157–163. DOI: 10.1016/j.procs.2024.11.096.
15. Лимонова Е.Е., Шешкус А.В., Николаев Д.П., Иванова А.А., Ильин Д.А., Арлазаров В.Л. Оптимизация быстродействия первых слоев глубоких сверточных нейронных сетей // Вестник Российского фонда фундаментальных исследований. 2016. №4(92). С. 84–96. DOI: 10.22204/2410-4639-2016-092-04-84-96.
16. Zhang X. et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 6848–6856. DOI: 10.1109/CVPR.2018.00716.
17. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474.4.
18. Lin T.-Y., et al. Feature Pyramid Networks for Object Detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 936–944. DOI: 10.1109/CVPR.2017.106.
19. Сацюк А.В., Белый Р.В., Ищенко А.Е. Оценка эффективности алгоритмов YOLO для обнаружения объектов в реальном времени во встраиваемых системах беспилотных транспортных средств // Сборник научных трудов Донецкого института железнодорожного транспорта. 2024. №4(75). С. 73–82.
20. Wang A., et al. YOLOv10: Real-Time End-to-End Object Detection // Advances in Neural Information Processing Systems. 2024. vol. 37. pp. 1–28. DOI: 10.52202/079017-3429.
21. 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.
22. Cherepanov N.I., Stepina N.O., Nikiforov I.V. Improving image analysis and processing performance on the RISC-V platform with Lichee Pi 4A // Proceedings of the Institute for System Programming of the RAS. 2025. vol. 37. no. 5. pp. 157–172. DOI: 10.15514/ISPRAS-2025-37(5)-12.
23. Chen C., et al. Lightweight Convolutional Transformers Enhanced Meta-Learning for Compound Fault Diagnosis of Industrial Robot // IEEE Transactions on Instrumentation and Measurement. 2023. vol. 72. pp. 1–12. DOI: 10.1109/TIM.2023.3277956.
24. Грибанов Д.Н., Мухин А.В., Килбас И.А., Паринger Р.А. Семантическая сегментация гиперспектральных изображений с использованием сверточных нейронных сетей и механизма внимания // Компьютерная оптика. 2024. Т. 48. №6. С. 894–902. DOI: 10.18287/2412-6179-CO-1371.
25. Wang Q., et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. pp. 11531–11539. DOI: 10.1109/CVPR42600.2020.01155.
26. Liu S., et al. Path Aggregation Network for Instance Segmentation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 8759–8768. DOI: 10.1109/CVPR.2018.00913.
27. Kisantal M., et al. Augmentation for small object detection // arXiv preprint arXiv:1902.07296. 2019.
28. 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).
29. Reis D., et al. Real-Time Flying Object Detection with YOLOv8 // arXiv preprint arXiv:2305.09972. 2023.
30. Wang G. et al. Multi-View Adaptive Fusion Network for 3D Object Detection // arXiv preprint arXiv:2011.00652. 2020.
31. Khan J. Comparative Analysis of Deep Learning Models for Perception in Autonomous Vehicles // arXiv preprint arXiv:2512.21673. 2025.

Сацюк Александр Владимирович — канд. техн. наук, доцент, заведующий лабораторией, научно-исследовательская лаборатория искусственного интеллекта, ФГБОУ ВО «Донецкий институт железнодорожного транспорта» (ФГБОУ ВО ДОНИЖТ). Область научных интересов: проектирование и оптимизация сверточных нейронных сетей для периметрального мониторинга, применение машинного зрения на встраиваемых платформах, разработка методов ускорения инференса и алгоритмы трекинга объектов в потоковом видео, субпиксельная детекция. Число научных публикаций — 74. alexandrsatsuk@gmail.com; ул. Куйбышева, 157, 283012, Донецк, Россия; р.т.: +7(949)329-8491.

Радковский Сергей Александрович — канд. техн. наук, доцент, заведующий кафедры, кафедра «автоматика, телемеханика, связь и вычислительная техника», ФГБОУ ВО «Донецкий институт железнодорожного транспорта» (ФГБОУ ВО ДОНИЖТ). Область научных интересов: автоматизированные системы управления, теория искусственных нейронных сетей, сетевое и системное администрирование. Число научных публикаций — 34. serj_rsa@mail.ru; ул. Горная, 8, 283012, Донецк, Россия; р.т.: +7 (949)331-2569.

Шеховцов Алексей Игоревич — канд. техн. наук, доцент, декан, факультет управления на железнодорожном транспорте, ФГБОУ ВО «Донецкий институт железнодорожного

транспорта» (ФГБОУ ВО ДОНИЖТ); доцент, кафедра организации перевозок и управления на железнодорожном транспорте, ФГБОУ ВО «Донецкий институт железнодорожного транспорта» (ФГБОУ ВО ДОНИЖТ). Область научных интересов: совершенствование работы транспортных систем, автоматизация технологических процессов. Число научных публикаций — 108. oleksa.i@mail.ru; ул. Горная, 6, 283012, Донецк, Россия; р.т.: +7(949)331-4951.

Сонина Светлана Дмитриевна — старший преподаватель, кафедра «автоматика, телемеханика, связь и вычислительная техника», ФГБОУ ВО «Донецкий институт железнодорожного транспорта» (ФГБОУ ВО ДОНИЖТ). Область научных интересов: теория искусственных нейронных сетей, методы оптимизации и адаптации параметров обучения нейросетевых моделей, нейросетевое моделирование цифровых устройств. Число научных публикаций — 21. soninadonigt@yandex.com; ул. Петровского, 256, 283111, Донецк, Россия; р.т.: +7(949)307-6225.

Воробьев Александр Александрович — старший преподаватель, кафедра «автоматика, телемеханика, связь и вычислительная техника», ФГБОУ ВО «Донецкий институт железнодорожного транспорта» (ФГБОУ ВО ДОНИЖТ). Область научных интересов: искусственный интеллект, нейронные сети. Число научных публикаций — 9. vorobyov94@gmail.com; ул. Шекспира, 21, 283050, Донецк, Россия; р.т.: +7(949)387-7174.

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

DOI 10.15622/ia.25.3.2

A. SATSYUK, S. RADKOVSKY, A. SHEKHOVTSOV, S. SONINA, A. VOROBYOV

INTEGRATION OF LIGHTWEIGHT CHANNEL ATTENTION MODULES INTO DEPTHWISE SEPARABLE CONVOLUTIONAL ARCHITECTURES FOR IMPROVING OBJECT LOCALIZATION ACCURACY IN EMBEDDED SYSTEMS

Satsyuk A., Radkovsky S., Shekhovtsov A., Sonina S., Vorobyov A. **Integration of Lightweight Channel Attention Modules into Depthwise Separable Convolutional Architectures for Improving Object Localization Accuracy in Embedded Systems.**

Abstract. The article addresses the challenge of designing high-performance convolutional neural networks for embedded computer vision systems operating in real time. The primary focus is on the trade-off between computational efficiency and object localization quality. It is noted that aggressive optimization using depthwise separable convolutions, applied to a lightweight single-stage architecture, often ensures high inference speed but leads to a loss in the network's ability to accurately determine the spatial boundaries of target objects. To overcome this compromise, a novel architectural strategy is proposed: the integration of lightweight, dynamically tunable channel attention modules directly into the structure of separable convolutional blocks. The resulting hybrid block performs selective weighting of feature space channels, allowing the network to selectively enhance the detailed features critical for object boundaries. Experimental validation was conducted on the target Raspberry Pi 5 platform using models quantized to the INT8 format. The modified architecture demonstrated a minor increase in computational complexity (up to 4.4 GFLOPs and 2.05 million parameters) compared to the baseline DSC version. The key result was a significant improvement in the localization metric $mAP_{0.75}$ – an increase of 8.3 percentage points to 66.5%, which surpasses the performance of the standard, non-DSC-optimized YOLOv8n model (61.1% on RPi 5). Concurrently, the frame rate (27.8 FPS) remained significantly above the real-time threshold requirement (20 FPS). Thus, the study proves that the targeted integration of compact attention mechanisms into ultra-lightweight architectures allows for a qualitative leap in localization accuracy that exceeds that of initial, more resource-intensive counterparts, without violating strict performance constraints. The proposed approach paves the way for developing more reliable and accurate detection systems for resource-constrained autonomous devices.

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

References

1. Satsyuk A.V., Volodarets N.V. [Modification of the YOLO Model for a Hybrid Detection and Tracking System in UAVs with Automatic Guidance]. Informatsionno-upravlyayushchiye sistemy – Information and Control Systems. 2025. no. 4. pp. 36–44. DOI: 10.31799/1684-8853-2025-4-36-44. (In Russ.).
2. Minin V.S., Kirillova E.A., Kirillova E.A., Filimonova E.V. [Detection of Anomalies in Economic Indicators Based on a Neural Network with Deeply Separable Convolutions]. Prikladnaya informatika – Applied Informatics. 2025. vol. 20. no. 6(120). pp. 30–51. DOI: 10.37791/2687-0649-2025-20-6-30-51. (In Russ.).
3. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 1800–1807. DOI: 10.1109/CVPR.2017.195.

4. Makhmudov M.D., Fazilova D.Sh. [Hand Gesture Recognition Using MobileNetV2]. Universum: Tekhnicheskie nauki – Universum: Technical Sciences. 2021. no. 11(92). pp. 60–62. DOI: 10.32743/UniTech.2021.92.11.12582. (In Russ.).
5. Voronkov A.D., Diane S.A.K. [Planning the Capture of Unknown Objects by a Multi-Fingered Grasping Device Based on a Neural Network Attention Mechanism]. Neirokompyutery: razrabotka, primenenie – Neurocomputers: Development, Application. 2024. vol. 26. no. 5. pp. 80–95. DOI: 10.18127/j19998554-202405-08. (In Russ.).
6. Klekovkin V.A., Markov N.G., Nebaba S.G. [YOLO Convolutional Neural Network Models with Attention Mechanism for Real-Time Computer Vision Systems]. Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Bulletin of Tomsk State University. Management, Computer Science, and Informatics. 2025. no. 72. pp. 39–50. DOI: 10.17223/19988605/72/4. (In Russ.).
7. Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning (PMLR). 2019. pp. 6105–6114.
8. Satsyuk A.V. [Infrastructure monitoring based on artificial intelligence]. Avtomatika, svyaz', informatika – Automation, communication, informatics. 2025. no. 9. pp. 32–34. DOI: 10.62994/AT.2025.9.9.005. (In Russ.).
9. Satsyuk A.V., Voevoda E.G. [Automatic safety control system at railway crossings]. Sbornik nauchnykh trudov Donetskogo instituta zheleznodorozhnogo transporta – Collection of scientific papers of the Donetsk Institute of Railway Transport. 2024. no. 2(73). pp. 39–45. (In Russ.).
10. Pavlenko D.A., Kovalev V.A., Snezhko E.V., Levchuk V.A., Pechkovsky E.I. [Recognition of the Earth's Underlying Surface Using a Convolutional Neural Network on a Single-Board Microcomputer]. Informatika – Informatics. 2020. vol. 17. no. 3. pp. 36–43. DOI: 10.37661/1816-0301-2020-17-3-36-43. (In Russ.).
11. Aksenov D.S., Zhilyaev V.A., Markin N.I., Titov I.A. [Object Recognition System Based on Raspberry Pi 4 and Intel Neural Compute Stick 2]. Informatsionnye sistemy i tekhnologii – Information Systems and Technologies. 2023. no. 4(138). pp. 10–16. (In Russ.).
12. Jacob B., et al. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 2704–2713. DOI: 10.1109/CVPR.2018.00286.
13. Yaseen M. What is YOLOv9: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector.arXiv preprint arXiv:2409.07813. 2024. pp. 1–10.
14. Desai M., Mewada H., Pires I.M., Roy S. Evaluating the Performance of the YOLO Object Detection Framework on COCO Dataset and Real-World Scenarios. Procedia Computer Science. 2024. vol. 251. pp. 157–163. DOI: 10.1016/j.procs.2024.11.096.
15. Limonova E.E., Shesshkus A.V., Nikolaev D.P., Ivanova A.A., Ilyin D.A., Arlazarov V.L. [Optimization of the Speed of the First Layers of Deep Convolutional Neural Networks]. Vestnik Rossiyskogo fonda fundamental'nykh issledovaniy – Bulletin of the Russian Foundation for Basic Research. 2016. no. 4(92). pp. 84–96. DOI: 10.22204/2410-4639-2016-092-04-84-96. (In Russ.).
16. Zhang X. et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 6848–6856. DOI: 10.1109/CVPR.2018.00716.
17. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474.4.

18. Lin T.-Y., et al. Feature Pyramid Networks for Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 936–944. DOI: 10.1109/CVPR.2017.106.
19. Satsyuk A.V., Bely R.V., Ishchenko A.E. [Evaluation of the Effectiveness of YOLO Algorithms for Real-Time Object Detection in Embedded Systems of Unmanned Vehicles]. Sbornik nauchnykh trudov Donetskogo instituta zheleznodorozhnogo transporta – Collection of Scientific Papers of the Donetsk Institute of Railway Transport. 2024. no. 4(75). pp. 73–82. (In Russ.).
20. Wang A., et al. YOLOv10: Real-Time End-to-End Object Detection. Advances in Neural Information Processing Systems. 2024. vol. 37. pp. 1–28. DOI: 10.52202/079017-3429.
21. Satsyuk A.V. [Optimization of the YOLOv8 Architecture for UAV Object Capture Tasks: Analysis of the Trade-Off Between Accuracy, Speed, and Computational Resources]. Vestnik Rostovskogo gosudarstvennogo universiteta putey soobshcheniya – Bulletin of the Rostov State Transport University. 2025. no. 2(98). pp. 35–42. DOI: 10.46973/0201-727X_2025_2_35. (In Russ.).
22. Cherepanov N.I., Stepina N.O., Nikiforov I.V. Improving image analysis and processing performance on the RISC-V platform with Lichee Pi 4A. Proceedings of the Institute for System Programming of the RAS. 2025. vol. 37. no. 5. pp. 157–172. DOI: 10.15514/ISPRAS-2025-37(5)-12.
23. Chen C., et al. Lightweight Convolutional Transformers Enhanced Meta-Learning for Compound Fault Diagnosis of Industrial Robot. IEEE Transactions on Instrumentation and Measurement. 2023. vol. 72. pp. 1–12. DOI: 10.1109/TIM.2023.3277956.
24. Gribanov D.N., Mukhin A.V., Kilbas I.A., Paringer R.A. [Semantic segmentation of hyperspectral images using convolutional neural networks and the attention mechanism]. Komp'yuternaya optika – Computer Optics. 2024. vol. 48. no. 6. pp. 894–902. DOI: 10.18287/2412-6179-CO-1371.
25. Wang Q., et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. pp. 11531–11539. DOI: 10.1109/CVPR42600.2020.01155.
26. Liu S., et al. Path Aggregation Network for Instance Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 8759–8768. DOI: 10.1109/CVPR.2018.00913.
27. Kisantal M., et al. Augmentation for small object detection. arXiv preprint arXiv:1902.07296. 2019.
28. Satsyuk A.V. [Design of an Adaptive Convolutional Neural Network Architecture Using Deep Separable Convolutions for Real-Time Operation on Embedded Devices]. Informatsionno-upravlyayushchie sistemy – Information and Control Systems. 2026. no. 1(138). DOI: 10.31799/1684-8853-2026-1. (In print). (In Russ.).
29. Reis D., et al. Real-Time Flying Object Detection with YOLOv8. arXiv preprint arXiv:2305.09972. 2023.
30. Wang G. et al. Multi-View Adaptive Fusion Network for 3D Object Detection. arXiv preprint arXiv:2011.00652. 2020.
31. Khan J. Comparative Analysis of Deep Learning Models for Perception in Autonomous Vehicles. arXiv preprint arXiv:2512.21673. 2025.

Satsyuk Alexander — Ph.D., Associate Professor, Head of the Laboratory, Artificial Intelligence Research Laboratory, FGBOU VO “Donetsk Institute of Railway Transport” (FGBOU VO DONIRT). Research interests: design and optimization of convolutional neural networks for perimetric monitoring, application of machine vision on embedded platforms, development of inference acceleration methods and object tracking algorithms in streaming

video, sub-pixel detection. The number of publications — 74. alexandrsatsuk@gmail.com; 157, Kuibyshev St., 283012, Donetsk, Russia; office phone: +7(949)329-8491.

Radkovsky Sergei — Ph.D., Associate Professor, Head of the Department, Department of Automation, Telemechanics, Communications, and Computer Engineering, FGBOU VO "Donetsk Institute of Railway Transport" (FGBOU VO DONIRT). Research interests: automated control systems, artificial neural network theory, network and system administration. The number of publications — 34. serj_rsa@mail.ru; 8, Gornaya St., 283012, Donetsk, Russia; office phone: +7 (949)331-2569.

Shekhovtsov Alexey — Ph.D., Associate Professor, Dean, Faculty of Railway Transport Management, FGBOU VO "Donetsk Institute of Railway Transport" (FGBOU VO DONIRT); Associate Professor, Department of Transportation Organization and Management in Railway Transport, FGBOU VO "Donetsk Institute of Railway Transport" (FGBOU VO DONIRT). Research interests: improvement of transport systems and automation of technological processes. The number of publications — 108. oleksa.i@mail.ru; 6, Gornaya St., 283012, Donetsk, Russia; office phone: +7(949)331-4951.

Sonina Svetlana — Senior Lecturer, Department of Automation, Telemechanics, Communications, and Computer Engineering, FGBOU VO "Donetsk Institute of Railway Transport" (FGBOU VO DONIRT). Research interests: theory of artificial neural networks, methods for optimizing and adapting the training parameters of neural network models, and neural network modeling of digital devices. The number of publications — 21. soninadonigt@yandex.com; 256, Petrovsky St., 283111, Donetsk, Russia; office phone: +7(949)307-6225.

Vorobyov Alexander — Senior Lecturer, Department of Automation, Telemechanics, Communications and Computer Engineering, FGBOU VO "Donetsk Institute of Railway Transport" (FGBOU VO DONIRT). Research interests: artificial intelligence, neural networks. The number of publications — 9. vorobyov94@gmail.com; 21, Shakespeare St., 283050, Donetsk, Russia; office phone: +7(949)387-7174.

Acknowledgements. The work was carried out within the framework of the state assignment on the topic EKTT-2026-0001.