

ОЦЕНКА ВЛИЯНИЯ ПРЕДОБРАБОТКИ ИЗОБРАЖЕНИЙ НА ПРОИЗВОДИТЕЛЬНОСТЬ И ТОЧНОСТЬ ДЕТЕКЦИИ ОБЪЕКТОВ В SLAMOT-СИСТЕМАХ

И.И. ЛЕВОНЕНКО^{1,2}, А.Д. РОБАЧЕВСКИЙ¹, Г. ВАН³, В.Ю. ЦВЕТКОВ¹

1 – Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

2 – ОАО «Пеленг», Республика Беларусь

3 – Национальный университет оборонных технологий, КНР

Поступила в редакцию 12 апреля 2026

Аннотация. Рассматривается задача экспериментальной оценки влияния методов предобработки изображений на производительность и точность детекции объектов в SLAMOT-системах. С целью выявления оптимальных стратегий улучшения входных данных для мобильных робототехнических платформ, проведено сравнительное исследование эффективности различных алгоритмов предобработки, включая повышение резкости на основе нерезкого маскирования, адаптивную коррекцию контраста CLAHE с гамма-коррекцией, интерполяционное увеличение разрешения и попатчевую обработку (tiling). Экспериментальная оценка выполнена на базе легковесной одностадийной архитектуры YOLOv8n, адаптированной для детекции транспортных средств на аэрофотоснимках посредством программного маппинга классов. В рамках стандартизированного протокола бенчмаркинга осуществляется расчет метрик точности ($mAP@0,5$; *Precision*; *Recall*; *F1-мера*), временных характеристик инференса, а также оценка потребления оперативной и видеопамати. Исследование направлено на определение компромиссов между качеством детекции, вычислительной сложностью и скоростью обработки в условиях ограниченных аппаратных ресурсов бортовых вычислительных систем БЛА.

Ключевые слова: детекция объектов, компьютерное зрение, обработка изображений, SLAMOT.

Введение

В условиях стремительного развития автономных робототехнических систем и беспилотных летательных аппаратов методы компьютерного зрения перестали быть вспомогательным инструментом и превратились в фундаментальный компонент обеспечения восприятия окружающей среды. Ключевую роль в этом процессе играют задачи одновременной локализации, картографирования и трекинга объектов (SLAMOT, Simultaneous Localization and Mapping with Object Tracking), где надежная детекция объектов напрямую влияет на точность построения карты, устойчивость оценки траектории и качество долгосрочного сопровождения динамических целей [1,2]. Однако эффективность нейросетевых детекторов в реальных сценариях существенно зависит не только от архитектуры модели, но и от качества входных визуальных данных, что делает проблему предобработки изображений критически важной для практического развертывания автономных систем.

Использование нейронных сетей в SLAMOT-системах позволяет отказаться от жестко заданных алгоритмических правил и опираться на обученные модели, способные адаптироваться к динамическим, частично наблюдаемым и слабоструктурированным условиям реального мира [3]. Методы предобработки изображений, включая повышение резкости (sharpening), адаптивную коррекцию контраста (CLAHE, Contrast Limited Adaptive Histogram Equalization), гамма-коррекцию и попатчевую обработку (tiling), способны существенно улучшить визуальные признаки, используемые нейронной сетью для принятия решений [4,5]. Однако каждый метод

вносит дополнительные вычислительные затраты и может по-разному влиять на метрики качества детекции, такие как $mAP@0,5$, $Precision$, $Recall$ и $F1$ -мера, а также на показатели производительности: медианное время инференса, потребление оперативной и видеопамати, стабильность кадровой частоты [6].

В настоящей работе проводится систематическая оценка влияния методов предобработки на производительность и точность легковесной архитектуры детекции YOLOv8n [7,8]. Выбор данной модели обусловлен ее широким применением в мобильной робототехнике благодаря оптимальному балансу между скоростью инференса и качеством детекции. Для выбранной архитектуры анализируется изменение метрик точности ($mAP@0,5$; $Precision@0,5$; $Recall@0,5$; $F1$ -мера) и производительности (латентность инференса, использование RAM/VRAM, общее время обработки кадра) при применении четырех конфигураций предобработки: базовой (без обработки), попатчевой обработки, комбинации CLAHE с гамма-коррекцией и повышения резкости на основе нерезкого маскирования. Эксперименты выполняются на наборе данных UAVDT (Unmanned Aerial Vehicle Drone Target Detection and Tracking Video Dataset), характерном для задач аэрофотосъемки и мониторинга с БЛА (Беспилотный Летательный Аппарат), что позволяет оценить устойчивость методов в условиях малого размера объектов, изменчивого освещения и динамических сцен [9].

Целью работы является определение возможностей повышения эффективности нейросетевого детектирования объектов в SLAMOT-системах БЛА с помощью различных методов предобработки изображений: повышения резкости, коррекции контраста и попатчевой обработки.

Количественные метрики оценки эффективности детекции

Для объективной оценки эффективности детекторов объектов в работе используется стандартный набор метрик, принятый в сообществе компьютерного зрения. В основе всех количественных показателей лежит процедура сопоставления предсказанных ограничивающих рамок с истинными аннотациями с использованием метрики пересечения по объединению (IoU , Intersection over Union), которая вычисляется по формуле

$$IoU = \frac{Area(B_{pred} \cap B_{gt})}{Area(B_{pred} \cup B_{gt})}, \quad (1)$$

где B_{pred} и B_{gt} обозначают предсказанную и истинную ограничивающие рамки соответственно.

Детекция считается корректной, если значение IoU между предсказанной и истинной рамкой превышает порог 0,5 и класс объекта определен верно. В противном случае срабатывание классифицируется как ложное положительное, а пропущенный истинный объект учитывается как ложное отрицательное срабатывание.

Метрика $Precision$ (точность) отражает долю корректных детекций среди всех сработавших детекторов и вычисляется как отношение истинно положительных срабатываний к сумме истинно положительных и ложно положительных

$$Precision = \frac{TP}{TP + FP}, \quad (2)$$

где TP (True Positive) – число верно обнаруженных объектов; FP (False Positive) – число ложных детекций.

Высокое значение $Precision$ характеризует способность модели избегать ложных срабатываний, что критически важно для навигационных систем, где ошибочная детекция препятствия может привести к некорректному планированию траектории.

Метрика $Recall$ (полнота) показывает, какая доля истинных объектов была обнаружена моделью, и рассчитывается как отношение истинно положительных срабатываний к сумме истинно положительных и ложно отрицательных

$$Recall = \frac{TP}{TP + FN}, \quad (3)$$

где TP (True Positive) – количество корректно обнаруженных объектов; FN (False Negative) – количество пропущенных объектов, которые присутствуют на изображении.

В задачах SLAMOT высокий $Recall$ необходим для построения плотной семантической карты и обеспечения надежного долгосрочного трекинга объектов.

Поскольку максимизация $Precision$ часто приводит к снижению $Recall$ и наоборот, для сводной оценки используется $F1$ -мера – гармоническое среднее между двумя этими величинами, вычисляемое по формуле

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, \quad (4)$$

Данная метрика достигает максимального значения только при сбалансированно высоких показателях точности и полноты, что делает ее удобной для сравнения моделей в условиях компромисса между типами ошибок.

Ключевой интегральной метрикой в работе выступает $mAP@0,5$ (mean Average Precision при пороге $IoU=0,5$). Ее расчет выполняется путем построения кривой зависимости $Precision$ от $Recall$ для каждого класса объектов при последовательном изменении порога уверенности модели. Площадь под этой кривой определяет значение Average Precision (AP) для данного класса. В классической интерпретации используется 11-точечная интерполяция

$$AP = \frac{1}{11} \sum_{t \in \{0; 0,1; \dots; 1,0\}} \max_{\tilde{r} \geq t} p(\tilde{r}), \quad (5)$$

где $p(\tilde{r})$ обозначает максимальное значение $Precision$ при уровне полноты не ниже $\tilde{r}$. Итоговое значение mAP получается усреднением AP по всем N категориям объектов в датасете

$$mAP@0,5 = \frac{1}{N} \sum_{i=1}^N AP_i. \quad (6)$$

Порог локализации 0,5 означает, что предсказанная рамка считается верной при перекрытии с истинной не менее чем на 50 %, что представляет собой компромиссное требование, допускающее умеренные погрешности позиционирования в динамических сценах. Значение $mAP@0,5$ чувствительно как к качеству классификации, так и к точности локализации, что делает его универсальным индикатором общей эффективности детектора в составе SLAMOT-системы.

В рамках данного бенчмарка все перечисленные метрики рассчитываются автоматически по результатам инференса на тестовой выборке UAVDT с использованием аннотаций в формате COCO (Common Objects in Context), что обеспечивает воспроизводимость и сопоставимость результатов для различных конфигураций моделей и методов предобработки.

Экспериментальная оценка влияния методов предобработки изображений на производительность и точность детекции объектов в SLAMOT-системах

Экспериментальная оценка эффективности методов предобработки изображений в конвейере SLAMOT выполнена на базе легковесной одностадийной модели детекции YOLOv8n. Выбор данной архитектуры обусловлен ее широким распространением в мобильных робототехнических системах: безякорный механизм детекции в сочетании с оптимизированной архитектурой CSPDarknet (Cross Stage Partial Darknet) обеспечивает высокую скорость инференса при сохранении конкурентоспособной точности распознавания.

Модели семейства YOLO использует веса, предварительно обученные на датасете COCO, и адаптированы для работы с целевыми классами UAVDT (автомобиль, грузовик, автобус) посредством программного маппинга категорий на этапе инференса. Соответствие классов установлено следующим образом: car (COCO класс 2 → UAVDT класс 1), truck (COCO класс 7 →

UAVDT класс 2), bus (COCO класс 5 → UAVDT класс 3). Выбор именно этих категорий обусловлен спецификой задач SLAMOT в городской среде, где транспортные средства являются основными динамическими объектами, требующими отслеживания для построения семантических карт и планирования безопасных траекторий.

Сравнительный анализ осуществляется по совокупности измеряемых характеристик. Время инференса характеризует быстродействие модели при обработке единичного кадра и фиксируется с разделением на этапы предобработки изображения и непосредственного прямого прохода нейронной сети. Потребление памяти отражает пиковый объем оперативной (RAM) и видеопамати (VRAM), необходимый для выполнения инференса, и измеряется консервативным методом – фиксируется максимальное потребление в течение всего цикла обработки. Размер модели и количество обучаемых параметров определяют требования к хранению весовых коэффициентов и вычислительную сложность архитектуры.

Непосредственное тестирование проводилось в рамках стандартизированного протокола, обеспечивающего воспроизводимость результатов. Порог уверенности устанавливался на уровне 0,2, что ниже стандартного значения 0,5 и позволяет детектировать объекты с умеренной уверенностью модели, что особенно важно для задач SLAMOT, где пропуск объекта может иметь более серьезные последствия, чем ложное срабатывание. Такой выбор порога обусловлен спецификой аэрофотосъемки с БЛА, где объекты часто имеют малый размер, частично перекрыты или находятся в сложных условиях освещения.

Перед замерами выполнялась фаза стабилизации, включавшая 10 итераций прогона модели на репрезентативном входном тензоре, что позволяло минимизировать влияние кэширования, динамической компиляции графа вычислений на итоговые замеры времени.

Характеристики временной эффективности включали медианное время инференса как робастную оценку центральной тенденции, 5-й и 95-й перцентили для оценки диапазона типичных задержек, медианное абсолютное отклонение как меру стабильности производительности, а также минимальное и максимальное время выполнения для выявления экстремальных случаев. Для методов предобработки (повышение резкости, интерполяционное увеличение разрешения в 1,5 раза, адаптивная коррекция контраста CLAHE с гамма-коррекцией, попачевая обработка) отдельно фиксировалось время выполнения алгоритмов улучшения изображения, что позволяло оценить их вклад в общую задержку обработки кадра.

При использовании метода tiling исходное изображение разбивалось на перекрывающиеся патчи размером 640×640 пикселей, что соответствует стандартному размеру входа моделей семейства YOLOv8, с перекрытием 128 пикселей по каждой оси для предотвращения потери объектов, попадающих на границы разделов. После выполнения детекции на каждом патче результаты со всех фрагментов объединялись в единое пространство координат с применением нежесткого не-максимального подавления с порогом $IoU=0,7$. Выбор повышенного порога подавления дубликатов обусловлен спецификой попачевой обработки: соседние патчи могут содержать частично перекрывающиеся детекции одного и того же объекта, и стандартный порог 0,5 мог бы привести к излишне агрессивному удалению валидных предсказаний, тогда как значение 0,7 позволяет сохранить детекции частично перекрывающихся объектов, обеспечивая баланс между устранением дубликатов и полнотой обнаружения. Визуализация результатов осуществлялась путем отрисовки только ограничивающих рамок без текстовых подписей классов и значений уверенности, что соответствовало требованиям к чистоте экспериментальной оценки без визуальных артефактов и исключало влияние затрат на рендеринг текста на замеры производительности.

Метод повышения резкости (sharpening), реализованный в данном исследовании, основан на технике нерезкого маскирования. Алгоритм применяет гауссово размытие к исходному изображению с ядром размером 9×9 пикселей и стандартным отклонением 10,0 для получения низкочастотной компоненты. Затем осуществляется взвешенное сложение исходного изображения и размытой версии с коэффициентами 1,5 и −0,5 соответственно. Данная операция эквивалентна добавлению к исходному изображению высокочастотной компоненты, полученной путем вычитания размытой версии из оригинала, что приводит к усилению контраста на границах объектов. Применительно к задаче детекции транспортных средств на аэрофотоснимках UAVDT, метод направлен на компенсацию размытия, вызванного движением

беспилотной платформы и вибрацией камеры, улучшая тем самым видимость мелкоразмерных объектов для сверточной нейронной сети.

На рис. 1 приведены примеры изображений из тестовой выборки UAVDT, использованные для бенчмаркинга нейросетевых моделей.



Рис. 1. Примеры кадров из тестовой выборки для бенчмаркинга детекторов

Несмотря на то, что на рис. 1 в качестве примера представлены лишь 4 кадра, они отражают визуальное разнообразие сцен (различные ракурсы, условия освещения, плотность объектов). Экспериментальная оценка выполнялась для 300 изображений, отобранных из полного набора данных UAVDT, содержащего более 4000 кадров.

На рис. 2 представлены результаты детектирования объектов с помощью модели YOLOv8n без предварительной обработки входных изображений.



Рис. 2. Детекция объектов моделью YOLOv8n без предобработки

На рис. 3 представлены результаты детектирования объектов с помощью модели YOLOv8n, с предварительной обработкой входных изображений методом поплатчевой обработки.

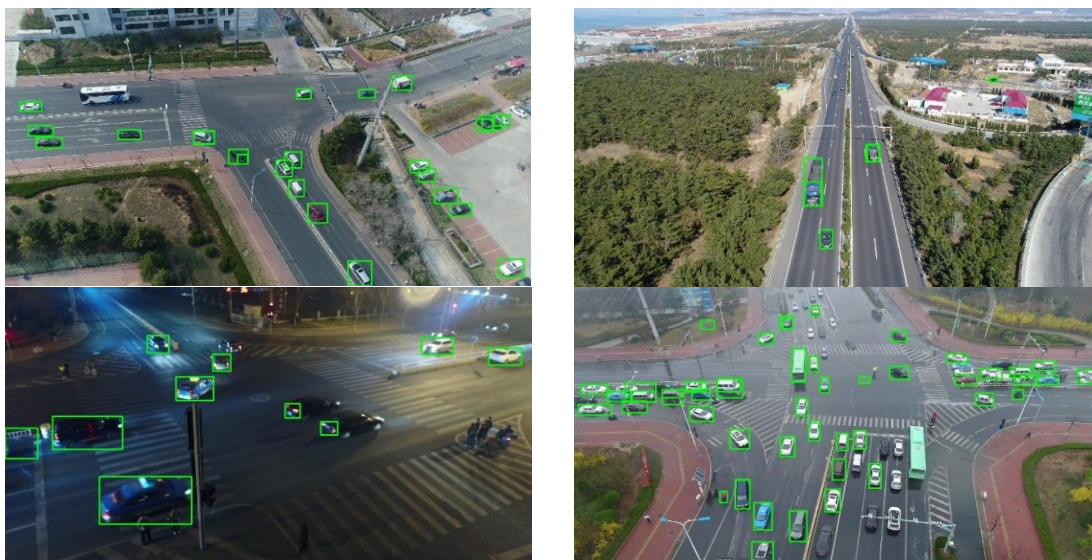


Рис. 3. Детекция объектов моделью YOLOv8n с предобработкой методом tiling

Как следует из представленных на рис. 3 изображений, применение стратегии поплатчевой обработки привело к существенному увеличению количества корректно локализованных объектов, особенно малоразмерных, что выразилось в значительном росте числа истинно положительных срабатываний. Данный эффект имеет критическое значение для задач SLAMOT, где максимальная полнота восприятия динамических объектов является приоритетным требованием для обеспечения безопасности навигации и построения плотной семантической карты.

На рис. 4 представлены результаты детектирования объектов с помощью модели YOLOv8n, с предварительной обработкой входных изображений на основе методов CLAHE и гамма-коррекции.



Рис. 4. Детекция объектов моделью YOLOv8n с предобработкой методами CLAHE и гамма-коррекции

Использование комбинации методов CLAHE и гамма-коррекции обусловило общее снижение количества генерируемых детекций. Несмотря на наблюдаемую частичную фильтрацию отдельных типов ложных срабатываний, данный процесс сопровождался возникновением новых ошибок классификации и локализации. Такое неоднозначное влияние свидетельствует о том, что глобальное улучшение контраста не всегда способствует улучшению экстракции признаков нейронной сетью в условиях аэрофотосъемки и может приводить к потере информативных текстурных особенностей, необходимых для надежной детекции.

На рис. 5 представлены результаты детектирования объектов с помощью модели YOLOv8n, с предварительной обработкой входных изображений на основе метода повышения резкости (sharpening).



Рис. 5. Детекция объектов моделью YOLOv8n с предобработкой методом sharpening

Метод повышения резкости на основе нерезкого маскирования продемонстрировал эффективность в сценариях с динамической съемкой, характеризующихся значительным движением камеры и сопутствующим размытием объектов. Применение данного метода позволило компенсировать деградацию высокочастотных компонентов изображения, вызванную вибрацией беспилотной платформы и изменением точки съемки в процессе полета. Усиление контраста на границах объектов способствовало более надежному извлечению признаков сверточной нейронной сетью, что привело к снижению количества частичных ошибок детекции, в частности, ложноположительных срабатываний на текстурированных участках фона и ложноотрицательных результатов для малоконтрастных транспортных средств.

В табл. 1 приведены результаты сравнительной оценки модели YOLOv8n для четырех конфигураций: без предобработки, попятчевая обработка, комбинированный метод CLAHE с гамма-коррекцией и sharpening.

Табл. 1. Характеристики детектирования объектов с помощью YOLOv8n

Конфигурация	Среднее время инференса, с	Потребление RAM, МБ	$mAP@0,5$	$Precision@0,5$	$Recall@0,5$	$F1@0,5$
Без предобработки	0,0159	671,77	4,03	57,10	0,17	0,34
Tiling	0,0833	704,35	4,76	50,93	0,29	0,58
CLAHE + гамма-коррекция	0,0305	687,36	3,6	55,21	0,14	0,27
Sharpening	0,0177	682,63	4,66	55,27	0,17	0,34

Представленные данные демонстрируют низкие значения $mAP@0,5$ для всех конфигураций, что обусловлено существенным доменным смещением между обучающей выборкой COCO и тестовым датасетом UAVDT. Метод попачечной обработки показал наибольшую эффективность, увеличив $Recall@0,5$ до 0,29 и $F1$ до 0,58 относительно базового уровня, однако ценой пятикратного роста времени инференса. Наблюдаемый дисбаланс между высоким $Precision$ (57–59 %) и низким $Recall$ (0,14–0,29) отражает консервативное поведение YOLOv8n: модель надежно детектирует лишь наиболее очевидные объекты, пропуская малоразмерные транспортные средства, характерные для аэрофотосъемки. Tiling частично компенсирует это ограничение за счет сохранения пространственного разрешения, критичного для детекции мелких объектов.

Применение комбинации методов CLANE и гамма-коррекции привело к системной деградации всех метрик качества детекции: $mAP@0,5$ снизился на 11 % (с 4,03 до 3,6), $Recall$ упал на 18 % (с 0,17 до 0,14), а интегральная метрика $F1$ показала наибольшее снижение среди всех методов – на 21 % (с 0,34 до 0,27). При этом время инференса увеличилось в 1,9 раза (с 0,0159 с до 0,0305 с), что обусловлено вычислительной сложностью алгоритма адаптивной нормализации гистограммы, требующего локальной обработки каждого тайла изображения размером 8×8 пикселей. Наблюдаемое ухудшение характеристик объясняется фундаментальным несоответствием между принципами работы CLANE и особенностями предобученной модели YOLOv8n: агрессивное повышение локального контраста, особенно в однородных областях (небо, дорожное покрытие, водная поверхность), приводит к генерации артефактов в виде ложных границ и текстурного шума, которые интерпретируются сверточной сетью как потенциальные объекты детекции, увеличивая количество ложноположительных срабатываний.

Метод повышения резкости (sharpening) продемонстрировал промежуточные результаты: при незначительном увеличении времени инференса на 11 % (с 0,0159 с до 0,0177 с) и минимальном росте потребления памяти (на 1,6 %) удалось достичь прироста $mAP@0,5$ на 16 % относительно базовой конфигурации (с 4,03 до 4,66). При этом метрики $Precision$ (55,27 %) и $Recall$ (0,17) остались на уровне, близком к базовому, что свидетельствует о способности метода sharpening улучшать общую точность детекции без существенного влияния на полноту обнаружения. Умеренное усиление высокочастотных компонентов изображения способствует лучшей локализации границ объектов, компенсируя размытие, вызванное движением камеры, однако не обеспечивает такого прироста $Recall$, как метод tiling, поскольку не решает проблему недостаточного разрешения для малоразмерных объектов.

Заключение

На основе сравнительного бенчмаркинга конфигураций нейросетевого детектора определена оптимальная конфигурация пайплайна предобработки, обеспечивающая максимальную устойчивость детекции при минимальных вычислительных затратах. Показано, что методы предобработки изображений оказывают существенное влияние на производительность и точность детекции объектов в SLAMOT-системах, однако характер этого влияния неоднозначен и зависит от специфики применяемого алгоритма.

Экспериментальные результаты свидетельствуют о низких абсолютных значениях $mAP@0,5$ (3,6–4,76) для всех конфигураций вследствие доменного смещения между COCO и UAVDT, что подтверждает необходимость дообучения модели. Тем не менее, методы предобработки позволяют оптимизировать отдельные аспекты детекции: Tiling обеспечил наилучшую полноту ($Recall@0,5 = 0,29$; $F1 = 0,58$) с приростом 70 %, однако пятикратное увеличение времени инференса ограничивает его применимость в системах реального времени. Sharpening продемонстрировал наиболее сбалансированный результат (прирост $mAP@0,5$ на 16 % при затратах времени +11 %), что делает его перспективным для бортовых платформ, тогда как CLANE с гамма-коррекцией привел к незначительной системной деградации метрик.

Таким образом, модель YOLOv8 продемонстрировала свою пригодность для интеграции в SLAMOT-системы благодаря высокой скорости инференса (15,9 мс на кадр в базовой конфигурации) и умеренному потреблению памяти, однако для достижения требуемого уровня точности в задачах аэрофотосъемки необходима комбинация различных методов предобработки с последующим дообучением на репрезентативной выборке UAVDT.

Благодарности. Работа выполнена при поддержке Белорусского республиканского фонда фундаментальных исследований (БРФФИ) и Национального фонда естественных наук Китая (НФЕНК) в рамках проекта T25КИ-054 «Технология многокамерной локализации и построения карты открытого пространства».

EVALUATION OF THE IMPACT OF IMAGE PREPROCESSING ON PERFORMANCE AND ACCURACY OF OBJECT DETECTION IN SLAMOT SYSTEMS

I.I. LIAVONENKA, A.D. ROBACHEVSKIY, WANG GUOYAN, V.YU. TSVIATKOU

Abstract. This paper presents an experimental evaluation of the impact of image preprocessing methods on the performance and accuracy of object detection in SLAMOT (Simultaneous Localization, Mapping, and Object Tracking) systems. To identify optimal strategies for enhancing input data for mobile robotic platforms, a comparative study was conducted on the effectiveness of various preprocessing algorithms, including unsharp masking-based sharpening, CLAHE (Contrast Limited Adaptive Histogram Equalization) with gamma correction, interpolation-based resolution upscaling, and patch-based processing (tiling). The experimental evaluation was performed using the lightweight single-stage YOLOv8n architecture, adapted for vehicle detection in aerial imagery through programmatic class mapping. Within a standardized benchmarking protocol, accuracy metrics ($mAP@0.5$; *Precision*; *Recall*; *F1-score*), inference timing characteristics, and RAM/VRAM consumption were computed. The study aims to determine trade-offs between detection quality, computational complexity, and processing speed under the hardware resource constraints of onboard computing systems.

Keywords: object detection, computer vision, image processing, SLAMOT.

Список литературы

1. Cadena C., Carlone L., Carrillo H. et al. // IEEE Transactions on Robotics. 2016. Vol. 32, No. 6. P. 1309–1332.
2. Placed J.A., Strader J., Carrillo H. et al. // IEEE Transactions on Robotics. 2023. Vol. 39, No. 5. P. 3403–3438.
3. Bochkovskiy A., Wang C.Y., Liao H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection // arXiv preprint arXiv:2004.10934. 2020.
4. Zuiderveld K. Contrast Limited Adaptive Histogram Equalization // Graphics Gems IV. Academic Press Professional, Inc. 1994. P. 474–485.
5. Gonzalez R.C., Woods R.E. Digital Image Processing. 4th ed. Pearson, 2018.
6. Everingham M., Eslami S.M.A., Van Gool L. et al. // International Journal of Computer Vision. 2015. Vol. 111, No. 1. P. 98–136.
7. Jocher G., Chaurasia A., Qiu J. Ultralytics YOLOv8 // GitHub Repository. 2023. URL: <https://github.com/ultralytics/ultralytics>
8. Liu W., Anguelov D., Erhan D. et al. SSD: Single Shot MultiBox Detector // Proc. of ECCV. 2016. P. 21–37.
9. Du D., Qi Y., Yu H. et al. The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking // arXiv preprint arXiv:1804.00518. 2018.