

**АНАЛИЗ ЭФФЕКТИВНОСТИ АРХИТЕКТУР YOLO ДЛЯ ДЕТЕКЦИИ ОБЪЕКТОВ С БОРТА БПЛА
НА ПРИМЕРЕ ДАТАСЕТА VISDRONE**

Розанов Г.Б.

Автономная Некоммерческая Организация Высшего Образования «Университет Иннополис» г. Иннополис

В работе представлен комплексный анализ современных архитектур детекции объектов семейства YOLO (версии 10, 11 и 12) в контексте их применения на борту беспилотных летательных аппаратов (БПЛА). Особое внимание уделено исследованию компромисса между точностью детекции и вычислительной эффективностью, что является критически важным аспектом при разработке автономных систем компьютерного зрения. Экспериментальная часть исследования проведена на базе датасета VisDrone, включающего характерные для БПЛА сценарии съемки. Результаты демонстрируют, что YOLOv10 сохраняет свои конкурентные преимущества даже среди более новых версий, предлагая оптимальное соотношение метрик точности и требований к вычислительным ресурсам.

Ключевые слова: Нейронная сеть, Детекция изображений, Беспилотный летательный аппарат, Метрики производительности.

В условиях стремительного развития технологий автономных летательных аппаратов задача эффективной детекции объектов приобретает особую актуальность. Они решают задачи поиска пропавших людей, мониторинг городской среды и другие. Современные БПЛА требуют интеллектуальных систем обработки изображений, способных работать в режиме реального времени при ограниченных энергетических и вычислительных ресурсах. Семейство алгоритмов YOLO (You Only Look Once) уже несколько лет остается одним из наиболее востребованных решений для подобных задач благодаря своей архитектурной эффективности.

Однако постоянное появление новых версий YOLO (на момент исследования рассматривались 10, 11 и 12 поколения) ставит перед разработчиками закономерный вопрос о целесообразности перехода на более современные архитектуры. Особенно актуален этот вопрос для бортовых систем БПЛА, где каждый дополнительный процент точности должен быть тщательно взвешен против увеличения вычислительной нагрузки. Большинство работ (Wu et al., 2023; Zhang et al., 2022) ограничиваются анализом ранних версий (v3-v8) без учета специфики бортовых вычислений. Имеющиеся исследования часто фокусируются исключительно на метриках точности, игнорируя критически важные для БПЛА параметры — размер модели и энергоэффективность (Chen et al., 2023). [1],[2],[3],[4]

Таким образом цель работы - провести исследование применимости новых поколений модели YOLO для задач обработки изображений на борту БПЛА с учетом особенностей таких задач – низких вычислительных ресурсах.

Для проведения сравнительного анализа был выбран датасет VisDrone, содержащий более 10,000 изображений, полученных с БПЛА в различных условиях съемки. В датасете 10 классов – пешеход, человек, велосипедист, машина, минивэн, грузовик, трицикл, крытый трицикл, автобус, моторизованное транспортное средство. Данные разделены на тренировочную (70%), валидационную (15%) и тестовую (15%) выборки. Особую ценность данному датасету придает наличие объектов разного масштаба - от крупных транспортных средств до отдельных пешеходов, что хорошо отражает реальные условия работы бортовых систем детекции. Все эксперименты проводились на графическом процессоре Gigabyte RTX3060 с 12 ГБ видеопамяти GDDR6. Выбор именно этой конфигурации обусловлен ее распространенностью среди исследователей и разработчиков, что делает полученные результаты более репрезентативными. Подробные характеристики представлены в таблице 1. Все модели были взяты из открытого источника – Ultralytics, функция обучения стандартная – метод «train». Обучение контролировалось так же стандартными методами, предоставляемыми библиотекой Ultralytics, динамически изменялся шаг обучения с сохранением лучшей модели. Размер изображения приводился к 640px. Код обучения представлен в репозитории.[5]

Таблица 1

Архитектура	GA106
CUDA-ядра	3584
RT-ядра	28
Tensor-ядра	112
Базовая частота	1320
Объем видеопамяти	12Gb GDDR6
Шина памяти	192bit
Пропускная способность	360Gb/s

В исследовании учитывались следующие ключевые показатели:

- Точность детекции (mAP50, mAP50-95)
- Полнота обнаружения (Recall)
- Точность классификации (Precision)
- Временные характеристики (время инференса)
- Размеры моделей

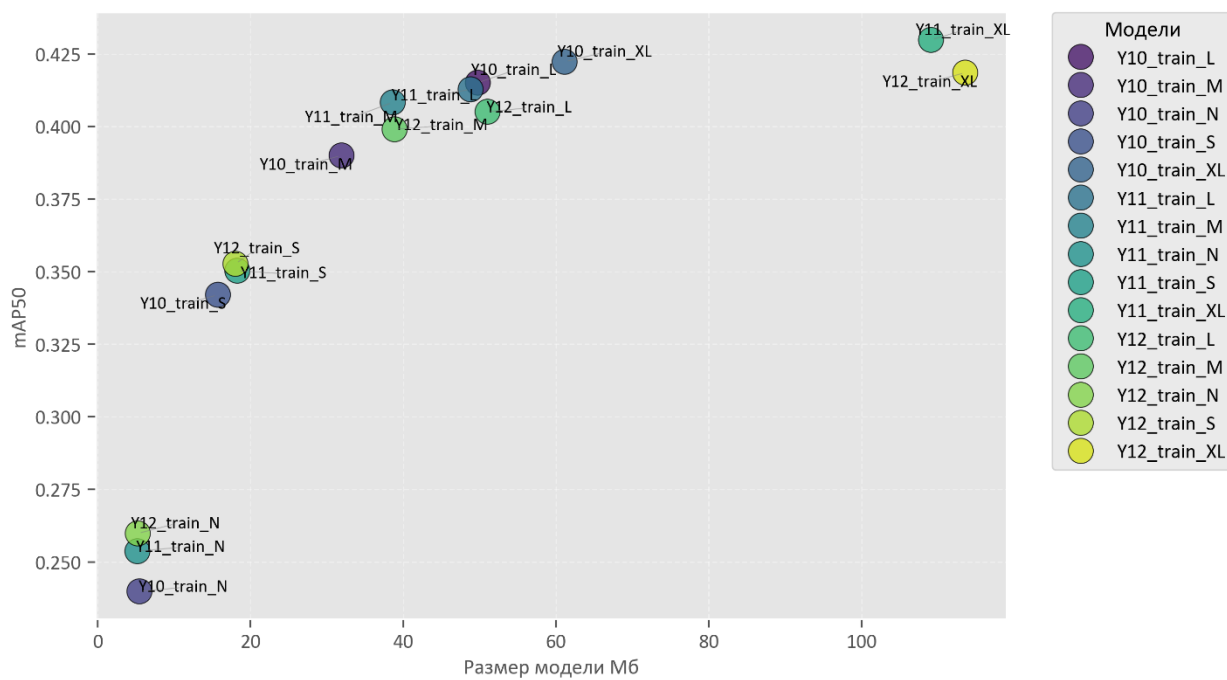


Рис 1. Корреляция метрики mAP50 и размера модели.

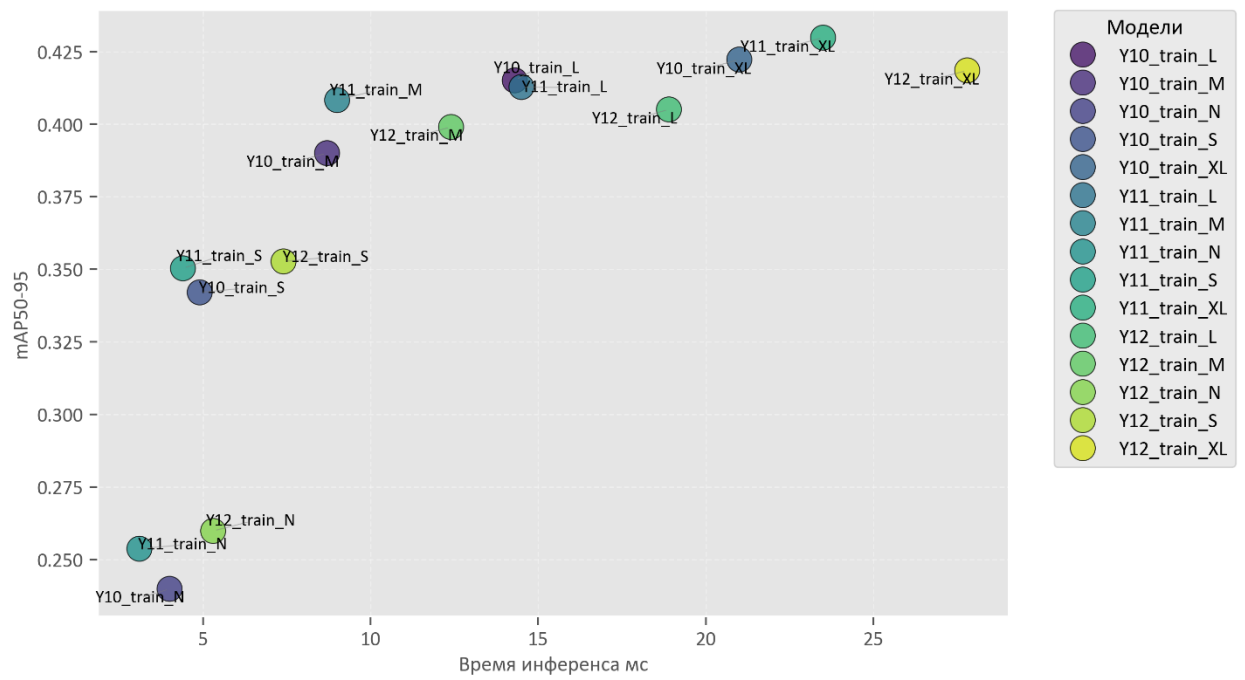


Рис 2. Корреляция метрики mAP50 и времени инференса.

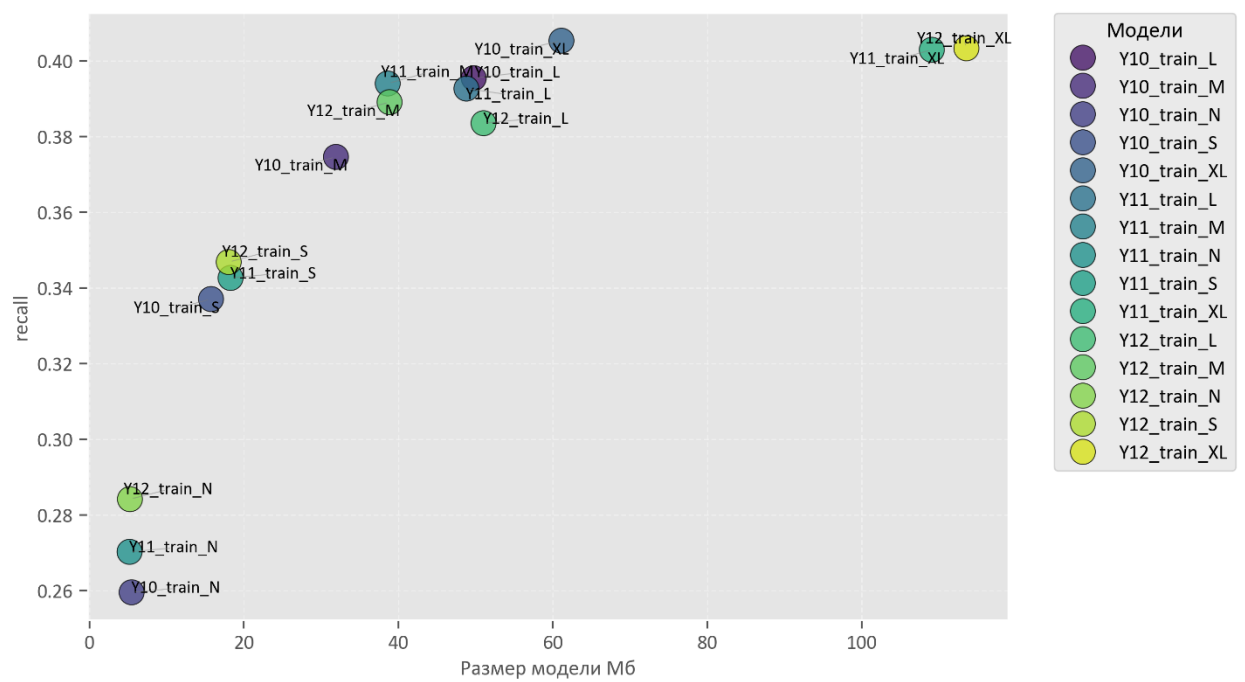


Рис.3 Корреляция метрики Recall и размера модели

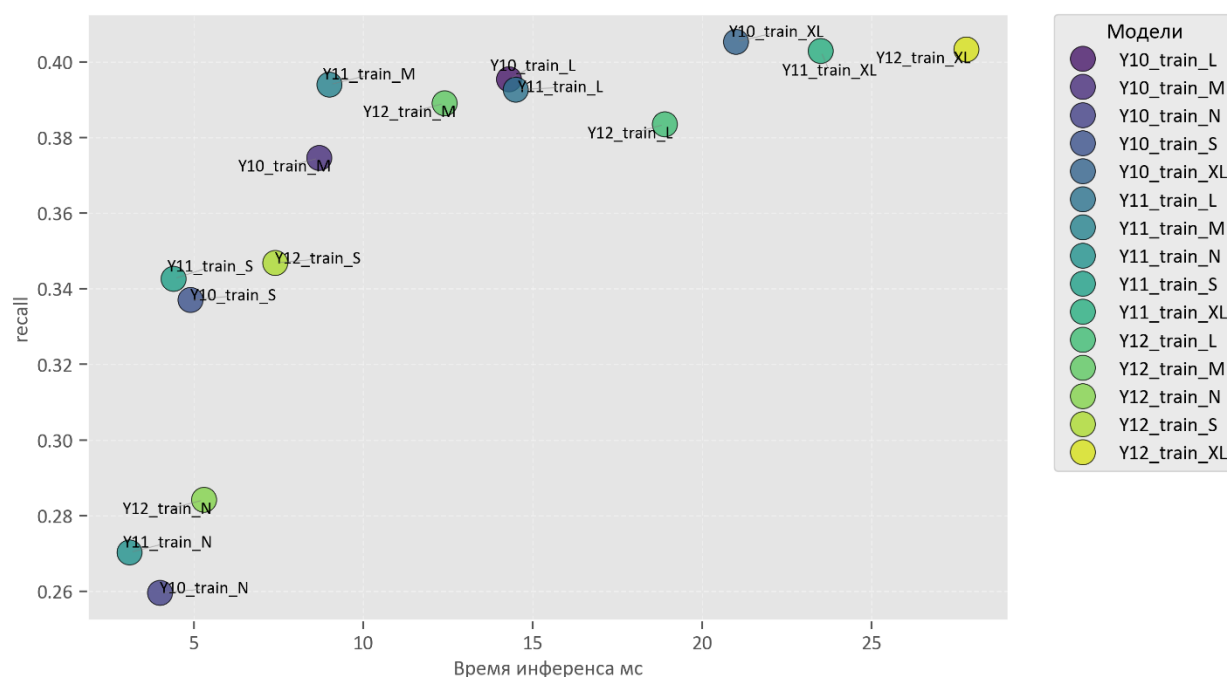


Рис. 4 Корреляция метрики Recall и времени инференса.

Таблица 2. Сравнительная таблица результатов обучения

Folder	Size_MB	metrics/precision	metrics/recall	metrics/mAP50	metrics/mAP50-95	inference_time
Y10_train_L	49.76	0.52777	0.39539	0.4151	0.25447	14.3
Y10_train_M	31.92	0.49965	0.37459	0.39003	0.23865	8.7
Y10_train_N	5.46	0.35353	0.25958	0.23987	0.13844	4.0
Y10_train_S	15.74	0.44926	0.33704	0.34204	0.20207	4.9
Y10_train_XL	61.12	0.52826	0.40529	0.42228	0.26196	21.0
Y11_train_L	48.81	0.52585	0.39264	0.41281	0.25443	14.5
Y11_train_M	38.62	0.5111	0.39398	0.40831	0.25041	9.0
Y11_train_N	5.19	0.36124	0.27027	0.25369	0.1469	3.1
Y11_train_S	18.27	0.46249	0.34262	0.35033	0.20903	4.4
Y11_train_XL	109.08	0.54107	0.40288	0.42983	0.26688	23.5
Y12_train_L	51.03	0.53371	0.38349	0.40509	0.24895	18.9
Y12_train_M	38.86	0.49712	0.38905	0.39904	0.24453	12.4
Y12_train_N	5.24	0.34996	0.28419	0.25983	0.15057	5.3
Y12_train_S	18.03	0.46108	0.34682	0.3527	0.21013	7.4
Y12_train_XL	113.56	0.5263	0.40328	0.41854	0.25909	27.8

Экспериментальные данные демонстрируют несколько важных закономерностей. Прежде всего для данного датасета YOLOv10 показывает сравнительные и в некоторых метриках даже лучшие показатели эффективности по сравнению с более новыми версиями. Таким образом при более чем в два раза большем размере и времени инференса модели YOLO v11 и v12 не показывают явного превосходства. Это связано в большей степени с размером датасета, но находясь в условиях задачи, когда большего позволить себе нельзя – создание датасета с БПЛА это дорогой и трудоемкий процесс, вывод об оптимальности именно 10-ого поколения достаточно жизнеспособный.

Похожие результаты получены для моделей семейства L. YOLOv10 показывает лучшие результаты при меньшем размере и времени инференса. Однако если сравнить семейство M, 10-ое поколение будет уступать 11-му и 12-му значительно, при условии, что новые поколения весят незначительно больше.

Интересные результаты получены при анализе динамики развития архитектур. Переход от Y10 к Y12 для моделей класса L сопровождался увеличением времени обработки на 32% при фактическом отсутствии роста точности. Аналогичная ситуация наблюдается и для XL-версий, где увеличение размера модели на 86% дало прирост mAP50 всего на 1.3%.

Полученные результаты ставят под сомнение целесообразность безусловного перехода на новые версии YOLO для задач бортовой обработки. В условиях, когда каждый дополнительный мегабайт модели и миллисекунда обработки критичны, YOLOv10 демонстрирует более сбалансированные характеристики.

На основании проведенного анализа можно сформулировать следующие рекомендации для разработчиков бортовых систем БПЛА:

- Для систем с жесткими ограничениями по ресурсам стоит рассматривать YOLOv10-M&L
- В большинстве практических случаев оптимальным выбором будет YOLOv10-XL
- Переход на YOLOv11/v12 оправдан только в системах с существенным запасом вычислительных ресурсов
- Использование версий Nano не рекомендуется из-за критически низкой точности

Проведенное исследование расширяет результаты предыдущих публикаций, впервые затрагивая новейшие версии модели YOLO и позволяет сделать несколько принципиальных выводов. Во-первых, YOLOv10 сохраняет лидерство по соотношению "точность-производительность" для задач бортовой детекции. Во-вторых, переход на более новые версии в большинстве случаев не оправдан из-за незначительного прироста точности при существенном увеличении требований к ресурсам. В-третьих, для различных сценариев использования оптимальными оказываются разные варианты YOLOv10: XL-версия для задач, требующих максимальной точности; L-версия как универсальное решение; S-версия для ресурсоограниченных систем.

Перспективным направлением дальнейших исследований видится разработка специализированных версий YOLO, оптимизированных именно для задач аэрофотосъемки, а не использование универсальных решений. Особый интерес представляет сочетание архитектурных преимуществ YOLOv10 с современными методами сжатия нейросетевых моделей.

Список использованной литературы

1. Wu et al. (2023) "Comparative Study of YOLO Variants for UAV-Based Object Detection"
2. Zhang et al. (2022) "YOLOv7 for UAV Applications: A Performance Benchmark"
3. Chen et al. (2023) "Edge-optimized YOLO for Real-time UAV Object Detection"
4. Wang & Li (2024) "Computational Efficiency Analysis of YOLO Architectures on Embedded GPUs"
5. <https://github.com/der-rozanov/YOLO012competition>