

## Сравнительный анализ нейросетевых моделей для задачи обнаружения беспилотных летательных аппаратов

С.В. Вычегжанин<sup>1</sup>, А.Г. Татарина<sup>1</sup>, Р.Е. Мышкин<sup>2</sup>

<sup>1</sup> Вятский государственный университет, 610000, Россия, г. Киров, ул. Московская, д. 36;

<sup>2</sup> АО «Научно-исследовательский институт средств вычислительной техники»,  
610025, Россия, г. Киров, ул. Мельничная, д. 31

### Аннотация

В статье проведен сравнительный анализ моделей YOLO и вариантов их тонкой настройки для задачи обнаружения беспилотных летательных аппаратов в реальном времени. Представлены оценки качества и вычислительной производительности различных семейств моделей от YOLOv3 до YOLO12 и версий от нано- до сверхбольшой. Проанализированы их преимущества и недостатки на основе широкого спектра мер, включая точность, полноту, F1-меру, среднюю точность, время вывода и размер модели. Исследованы варианты тонкой настройки предварительно обученных на корпусе Common Objects in Context моделей при разных разрешениях входных изображений и скорости обучения. Выполнен анализ ошибок обнаружения беспилотных летательных аппаратов малых размеров в сложных условиях с использованием специализированных наборов данных. Исследование показало, что семейство YOLO11 является оптимальным по совокупности оцениваемых параметров, обеспечивая баланс между качеством и вычислительной производительностью. Наиболее высокие оценки качества обнаружения беспилотных летательных аппаратов на тестовом наборе данных были получены с использованием тонко настроенной модели yolo11m с увеличенным разрешением входных изображений и достигли значений  $F1=0,970$ ,  $AP50=0,981$ ,  $AP50-95=0,765$ .

**Ключевые слова:** обработка изображений, обнаружение объектов, нейросетевые алгоритмы, глубокое обучение, трансферное обучение.

**Цитирование:** Вычегжанин, С.В. Сравнительный анализ нейросетевых моделей для задачи обнаружения беспилотных летательных аппаратов / С.В. Вычегжанин, А.Г. Татарина, Р.Е. Мышкин // Компьютерная оптика. – 2026. – Т. 50, № 2. – 1728. – DOI: 10.18287/COJ1728.

**Citation:** Vychezhzhanin SV, Tatarinova AG, Myshkin RE. Comparative analysis of neural network models for detecting unmanned aerial vehicles. Computer Optics 2026; 50(2): 1728. DOI: 10.18287/COJ1728.

### Введение

Беспилотные летательные аппараты (БПЛА), или дроны, стали неотъемлемой частью современной жизни благодаря своей универсальности, низкой стоимости и способности выполнять задачи в труднодоступных или опасных для человека условиях. Дроны широко используются в различных гражданских и коммерческих приложениях, таких как доставка товаров [1], наблюдение за окружающей средой [2], сельскохозяйственные работы [3], предотвращение стихийных бедствий [4], фотосъемка [5], мониторинг инфраструктуры [6].

Однако широкое распространение дронов сопровождается ростом инцидентов, связанных с их неправомерным использованием. Дроны способны переносить взрывчатые вещества или химическое оружие и могут быть запущены в критически важные зоны или места скопления людей с целью причинения вреда [4]. Дроны применяются для отправки запрещенных веществ и предметов, таких как мобильные телефоны, в тюрьмы [7]. Шпионские дроны используются для сбора конфиденциальной информации [4]. Эти инциденты подчеркивают необходимость разработки эффективных систем обнаружения и борьбы с дронами, способных функционировать в реальном времени и обеспечивать высокую точность и надежность.

Существует четыре метода обнаружения дронов. Радиолокационный метод позволяет обнаруживать дроны на значительных расстояниях, особенно в условиях плохой видимости, но может быть неэффективен при обнаружении небольших и низколетящих объектов, а также дорогостоящим в эксплуатации. Радиочастотный мониторинг позволяет обнаруживать сигналы каналов управления дронами, однако не способен обнаружить автономные дроны или дроны, использующие нестандартные частоты. Акустический метод обнаруживает дроны по их характерным звуковым сигнатурам и может определять звуки, излучаемые дронами, на расстояниях, превышающих дистанции достоверного визуального обнаружения, однако его эффективность снижается для обнаружения дронов в шумной среде или на большом расстоянии. Визуальный метод с использованием систем видеонаблюдения обеспечивает недорогое обнаружение и точное отслеживание дронов, но имеет ограниченную дальность обнаружения и чувствителен к погодным условиям и освещению.

Визуальное обнаружение дронов на основе компьютерного зрения представляет собой многообещающий подход. Использование данных с видеокamer позволяет обнаруживать дроны в режиме реального времени. Преимуществами компьютерного зрения являются возможность интеграции с существующей инфраструктурой видеонаблюдения и относительная доступность оборудования по сравнению, например, с передовыми радиолокационными системами. Системы компьютерного зрения могут использовать различные типы камер, включая тепловизионные и мультиспектральные, что позволяет обнаруживать дроны в различных условиях.

Сверточные нейронные сети являются одними из наиболее значимых моделей для обнаружения и категоризации объектов на изображениях и в видеопоследовательностях. Популярным представителем моделей данного типа является YOLO (You Only Look Once). Он был специально разработан для преодоления проблем, связанных со скоростью вывода, при сохранении конкурентоспособной точности. Это достигается путем одновременного выполнения определения ограничивающей рамки и классификации объектов на одном этапе. Исследования показывают, что YOLO эффективно работает в реальном времени и способен обнаруживать небольшие объекты, такие как дроны, с высокой точностью. Однако, несмотря на прогресс в области глубокого обучения, задачи обнаружения дронов остаются сложными из-за малого размера объектов, их высокой скорости перемещения и вариативности фона. Поэтому дальнейшее исследование моделей YOLO для специфических задач обнаружения дронов представляет собой актуальную научно-практическую задачу.

Целью работы является экспериментальное исследование разных версий моделей YOLO и вариантов их тонкой настройки для обнаружения объектов на изображениях. Объектами интереса являются дроны, т.е. решается задача одноклассового обнаружения.

Статья построена следующим образом. В параграфе 1 представлен обзор работ по данной тематике. В параграфе 2 приводится описание набора данных, используемого для обучения и тестирования нейросетевых моделей. В параграфе 3 описан метод решения задачи обнаружения БПЛА. В параграфе 4 перечислены модели YOLO, используемые в экспериментах. В параграфе 5 приведены программные и аппаратные настройки. В параграфе 6 описаны меры, используемые для оценки производительности моделей. В параграфе 7 проанализированы результаты экспериментов. Заключение содержит выводы.

### *1. Обзор публикаций по тематике работы*

Многочисленные работы по разработке и совершенствованию методов обнаружения дронов обуславливают высокую степень важности исследований в данном направлении. Среди методов визуального обнаружения выделяют традиционные алгоритмы машинного обучения [8, 9], сиамскую сеть [10 – 12], трансформер [13, 14], YOLO [15 – 19]. Эти методы в основном применяются в естественных и инфракрасных сценах.

Исследования показывают, что методы на основе традиционных алгоритмов машинного обучения применимы только для обнаружения объектов одного типа в простых наборах данных. Методы на основе сиамских сетей хорошо подходят для отслеживания одной цели в реальном времени. Методы на основе трансформера эффективны в обработке сложных сцен и захвате долгосрочных зависимостей. Методы на основе YOLO характеризуются балансом скорости и точности, подходящими для обнаружения и отслеживания объектов в реальном времени. В настоящей работе выбор был сделан в пользу YOLO вследствие ее лучших характеристик для работы в реальном времени.

Архитектура YOLO в настоящее время имеет 12 версий, которые активно исследуются научным сообществом. Широкое разнообразие и неоднозначность поведения моделей YOLO в разных условиях требует большого объема экспериментальной работы для выбора оптимальной модели перед практическим применением. Существует ряд работ по анализу YOLO для идентификации различных объектов и событий, в том числе дронов [15 – 19]. В статье [15] анализируются 6 архитектур YOLO с v5 по v11 применительно к задаче обнаружения дефектов дорожного покрытия на изображениях, полученных с БПЛА. Авторы отмечают, что YOLOv5-l и v9-c обеспечивают наилучшую точность, а YOLOv11-n является наиболее эффективной по совокупности точности, размера модели и вычислительной сложности. В исследовании [16] проводится сравнение десяти моделей YOLOv5 и пяти моделей YOLOv8 при решении задачи обнаружения средств индивидуальной защиты на строительных площадках. Наиболее точной оказалась модель YOLOv8-m, достигшая  $mAP50 = 92\%$ .

В статье [17] представлен комплексный сравнительный анализ эффективности моделей YOLOv5, YOLOv8 и YOLOv8 DeepSORT в области обнаружения и отслеживания БПЛА. Авторы показали, что модели YOLOv8 демонстрируют более высокую точность и частоту кадров в секунду (FPS) по сравнению с YOLOv5. В работе [18] анализируется YOLOv9 в сравнении с версиями v3, v4, v5 и v8 для обнаружения БПЛА. По результатам экспериментов данная сеть превосходит по точности все предыдущие версии, достигая  $mAP50 = 95,7\%$ .

Исследование [19] сравнивает версии YOLOv5 и YOLOv7 для обнаружения небольших БПЛА. Результаты показывают, что YOLOv5 лучше работает на изображениях с коррекцией цвета, а YOLOv7 – на изображениях с исходными цветами. Анализ различных расстояний показывает, что YOLOv5 демонстрирует хорошую точность обнаружения на близком, среднем и дальнем расстоянии, особенно на сложных фонах, а YOLOv7, наоборот, показывает низкую точность обнаружения в сценариях с дальними расстояниями и сложными фонами.

Вклад настоящей статьи состоит в следующем: 1) выполнено комплексное сравнение семейств моделей от YOLOv3 до YOLO12 и версий от нано- до сверхбольшой; 2) проведена оценка качества и вычислительной производительности моделей YOLO для обнаружения БПЛА на изображениях; 3) проанализированы преимущества и недостатки моделей на основе широкого спектра мер, включая точность, полноту, время вывода, время обучения и размер модели; 4) сформирован тестовый набор данных с реальными условиями видеонаблюдения, предполагающих наличие в кадре птиц, деревьев и объектов городской инфраструктуры; 5) проведено исследование, направленное на повышение точности работы моделей в режиме реального времени.

## 2. Набор данных

В качестве обучающего набора данных использовался DUT Anti-UAV [20], содержащий изображения с разрешением от  $240 \times 160$  до  $5616 \times 3744$  пкс с одним классом объектов «дрон» в сложных сценариях обнаружения. Набор данных состоит из 10000 изображений, разделенных на обучающую, проверочную и тестовую выборки.

Дополнительно в данной работе был сформирован небольшой тестовый набор данных Custom, содержащий 100 случайно выбранных изображений с дронами из тестовой выборки DUT Anti-UAV и 100 сделанных авторами фотографий городского ландшафта с разрешением от  $1024 \times 768$  до  $1920 \times 1080$  пкс. Этот набор создан с целью оценки качества работы моделей, обученных на общедоступном наборе данных, в реальных условиях видеонаблюдения, когда в кадре могут быть птицы, деревья и различные объекты городской инфраструктуры: здания, опоры линий электропередач, строительные краны и пр. Распределение изображений и дронов в наборах данных приведено в табл. 1.

Табл. 1. Количество изображений и дронов в наборах данных

Выборка	DUT Anti-UAV		Custom	
	Количество изображений	Количество дронов	Количество изображений	Количество дронов
Обучающая	5200	5243	–	–
Проверочная	2600	2621	–	–
Тестовая	2200	2245	200	102

На рис. 1 представлено распределение площадей рамок, ограничивающих дроны, на изображениях в DUT Anti-UAV. В наборе данных преобладают изображения с малоразмерными дронами площадью в интервале (400; 900] ( $\sim 25 \times 25$ ) пкс.

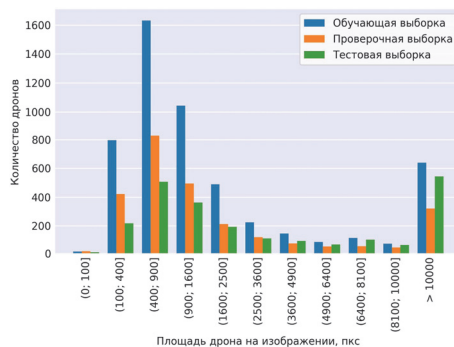


Рис. 1. Распределение площадей дронов на изображениях в DUT Anti-UAV

## 3. Метод обнаружения объектов

Модели семейства YOLO предоставляются с предобученными весами, полученными на наборе данных COCO [21]. Данный набор содержит разметку для 80 классов объектов, включая классы «самолёт» (airplane) и «птица» (bird), объекты которых могут быть визуально схожи с БПЛА. Это способствует эффективной инициализации модели при тонкой настройке (fine-tuning) на специализированном наборе данных с изображениями БПЛА, которая актуальна для задачи одноклассового обнаружения объектов [22, 23] и позволяет избежать ресурсоемкого обучения с нуля.

Важным аспектом является выбор гиперпараметров тонкой настройки моделей, а также параметров предобработки и аугментации данных. Значения некоторых гиперпараметров, таких как размер батча, зависят от количества весов обучаемой модели и объема доступной видеопамати GPU, и в большей степени оказывают влияние на время обучения. Другие гиперпараметры, например, такие как скорость обучения (learning rate) и алгоритм оптимизации, управляют изменением весов модели при обучении.

В библиотеке Ultralytics скорость обучения по умолчанию равна 0,01 [24]. Учитывая наличие похожих классов на этапе предобучения, предлагается снизить скорость обучения при тонкой настройке до 0,0001, чтобы уменьшить влияние эффекта «катастрофического забывания» на формирующие полезные признаки веса, полученные на этапе предобучения.

Наиболее трудные случаи обнаружения БПЛА представляют объекты небольших размеров, поэтому предлагается увеличить параметр, отвечающий за масштабирование входных изображений с разрешения 640×640 до 1024×1024, что увеличит время обучения и вывода, а также объем требуемой видеопамяти, но улучшит детекцию деталей на изображениях.

Параметры аугментации данных, связанные с расширением цветового пространства и геометрическими преобразованиями, будут использоваться со значениями по умолчанию, так как эксперименты в этом направлении не показали прироста в качестве обнаружения.

Таким образом, далее экспериментально исследуются следующие варианты тонкой настройки на специализированном наборе данных для обнаружения БПЛА моделей семейства YOLO, предобученных на наборе данных COCO:

- 1) тонкая настройка по умолчанию ( $FT_{\text{default}}$ );
- 2) тонкая настройка с уменьшенной скоростью обучения и увеличенным масштабированием входных изображений ( $FT_{\text{custom}}$ ).

#### 4. Модели

Экспериментальное исследование проведено для 20 версий моделей из 7 семейств YOLO, перечисленных в табл. 2 и имеющих предварительно обученные веса на наборе данных COCO. Суффикс в названии модели указывает на ее версию: «n» – нано, «m» – средняя, «x» – сверхбольшая, «t» – крошечная, «e» – расширенная.

Табл. 2. Семейства и версии моделей YOLO

Семейство	Версия		
	Нано	Средняя	Сверхбольшая
YOLOv3u	YOLOv3u-tiny	–	YOLOv3u
YOLOv5u	YOLOv5un	YOLOv5um	YOLOv5ux
YOLOv8	YOLOv8n	YOLOv8m	YOLOv8x
YOLOv9	YOLOv9t	YOLOv9m	YOLOv9e
YOLOv10	YOLOv10n	YOLOv10m	YOLOv10x
YOLO11	YOLO11n	YOLO11m	YOLO11x
YOLO12	YOLO12n	YOLO12m	YOLO12x

Реализация алгоритмов взята из библиотеки Ultralytics [24], предоставляющей исследователям и программистам модели YOLO и функции по их обучению, валидации, выводу и экспорту. Следует отметить, что данная библиотека не поддерживает модели YOLOv1, YOLOv2, YOLOv4 и YOLOv7, а для модели YOLOv6 не предоставляет предобученные на COCO веса, поэтому эти модели не использовались в экспериментах.

#### 5. Программные и аппаратные настройки

Обучение моделей выполнялось в облачной вычислительной среде на платформе Kaggle на двух графических ускорителях NVidia Tesla T4 16 Gb с CUDA 12.6, Python 3.10.12 и Ultralytics 8.3.98. При обучении всех версий YOLO использовались значения параметров, приведенные в табл. 3. Размер батча принимался равным 8, 16 или 64 при параметрах по умолчанию, а при улучшенных параметрах 4 – 8, 16 и 32 соответственно, учитывая версию модели и максимальную загрузку памяти графического ускорителя. Значения параметров аугментации данных принимались по умолчанию:  $hsv\_h=0,015$ ,  $hsv\_s=0,7$ ,  $hsv\_v=0,4$ ,  $degrees=0,0$ ,  $translate=0,1$ ,  $scale=0,5$ ,  $shear=0,0$ ,  $perspective=0,0$ ,  $flipud=0,0$ ,  $fliplr=0,5$ ,  $bgr=0,0$ ,  $mosaic=1,0$ ,  $mixup=0,0$ ,  $copy\_paste=0,0$ .

Табл. 3. Значения параметров обучения

Параметр	Значение	
	$FT_{\text{default}}$	$FT_{\text{custom}}$
Количество эпох	50	50
Оптимизатор	AdamW	AdamW
Размер батча	64 для $n$ , $t$ , 16 для $m$ , 8 для $x$ , $e$ версий	32 для $n$ , $t$ , 16 для $m$ , 4 – 8 для $x$ , $e$ версий
Размер изображения	(640, 640)	(1024, 1024)
Начальная скорость обучения и финальная доля скорости обучения	(0,01, 0,01)	(0,0001, 0,0001)
Аугментация данных	по умолчанию	по умолчанию

#### 6. Оценка производительности

Оценка моделей YOLO и вариантов тонкой настройки осуществлялась с помощью мер качества, вычислительной производительности и размера. К мерам качества относятся:

- точность (precision) – определяет долю объектов, действительно принадлежащих данному классу, относительно всех объектов, которые модель отнесла к этому классу;

- полнота (recall) – определяет, какую долю всех экземпляров объектов данного класса модель предсказала верно;
- F1-мера (F1) – представляет собой гармоническое среднее между точностью и полнотой;
- средняя точность (AP) – определяется как площадь под кривой precision – recall; обычно вычисляется при значении порога перекрытия между реальной и предсказанной ограничивающей рамкой (Intersection over Union, IoU), равного 0,5 (AP50) или варьирующегося от 0,5 до 0,95 с шагом 0,05 (AP50-95).

Вычислительная производительность определяется временем вывода модели, т.е. временем, которое требуется модели для обработки входных данных и генерации предсказаний, и количеством кадров в секунду (FPS). Мерами размера являются занимаемый моделью объем дискового пространства и количество параметров модели.

## 7. Анализ результатов

Оценки качества, вычислительной производительности и размера моделей представлены в Приложении А (табл. 4 и 5). Значения AP50, AP50-95 и F1-меры приведены для модели YOLO:

- 1) предобученной на наборе COCO оригинальной (Base), относящей дрон к классу «самолет»;
- 2) дообученной с параметрами по умолчанию (FT<sub>default</sub>);
- 3) дообученной с уменьшенной скоростью обучения и увеличенным масштабированием входных изображений (FT<sub>custom</sub>).

### 7.1. Анализ качества

Результаты, представленные в Приложении А (табл. 4 и 5), показывают, что уменьшение скорости обучения и увеличение масштаба изображения дает ощутимый прирост качества обнаружения объектов небольших размеров: на тестовой выборке DUT Anti-UAV на 3,4–12 п.п. по F1-мере и 7,0–14,5 п.п. по AP50, на корпусе Custom на 1,3–14,8 п.п. по F1-мере и 4,3–15,9 п.п. по AP50. В сравнении с оригинальной моделью на тестовой выборке DUT Anti-UAV среднее значение F1-меры после FT<sub>default</sub> повысилось на 36,0 п.п., а после FT<sub>custom</sub> – на 42,3 п.п., на корпусе Custom – на 36,2 и 44,6 п.п. соответственно.

Рис. 2, построенный для наилучшего по качеству варианта тонкой настройки FT<sub>custom</sub>, демонстрирует, что на тестовой выборке DUT Anti-UAV самые высокие оценки AP50=0,982 и AP50-95=0,775 получила yolov9e, а наименьшие оценки AP50=0,897 и AP50-95=0,608 у yolov3u-tiny. На корпусе Custom результаты аналогичны. Значительный разрыв между значениями AP50 и AP50-95 свидетельствует о том, что модели испытывают трудности с обработкой дронов разных размеров при более высоких пороговых значениях IoU, что является потенциальным направлением улучшения алгоритмов для данной задачи.

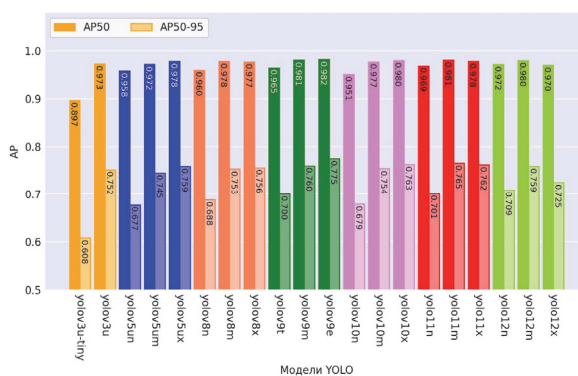


Рис. 2. Средняя точность для варианта FT<sub>custom</sub>

Рис. 3 демонстрирует компромисс между точностью и полнотой на тестовой выборке DUT Anti-UAV с учетом размера моделей. Наномодели наиболее сильно выделяются из общей массы, по обоим мерам уступая средним и сверхбольшим моделям. Лидером оказалась yolov11m, достигая F1-меры, равной 0,970. За ней следуют yolov8x и yolov9e со значением F1-меры, равным 0,968.

Анализ результатов двух вариантов тонкой настройки FT<sub>default</sub> и FT<sub>custom</sub> (рис. 4) показывает, что увеличение масштаба входного изображения увеличивает время обучения моделей: сверхбольших в 2,5 раза (3,3 раза для yolov12x), средних в 2,2 раза (3,3 раза для yolov12m), нано в 2 раза (2,2 раза для yolov12n). При этом повышается F1-мера для моделей: сверхбольших на 4,5–12,0 п.п., средних на 3,4–5,3 п.п., нано на 5,2–12,0 п.п. на тестовой выборке DUT Anti-UAV. Также повышается средняя точность AP50-95 моделей: сверхбольших на 8,2–16,5 п.п., средних на 7,0–10,3 п.п., нано на 7,7–14,5 п.п.

Таким образом, увеличение разрешения входных изображений повышает точность обнаружения малых объектов в сложных случаях, однако требует повышенных вычислительных затрат как в процессе обучения, так и в процессе вывода модели.

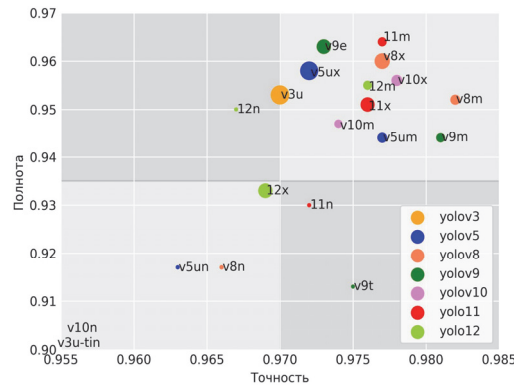


Рис. 3. Точность и полнота с учетом размера моделей

7.2. Анализ вычислительной производительности и размера

На рис. 5 представлено время вывода и размер моделей. Самая быстрая модель yolov3u-tiny имеет время вывода 8,5 мс (FPS = 118) при обработке изображений с разрешением 1024×1024, а самая медленная yolov3u – 117,8 мс (FPS = 8). Сверхбольшие модели оказались слишком тяжелыми для работы в реальном времени на графическом ускорителе NVidia Tesla T4, достигая значений FPS от 8 до 14. Средняя скорость вывода, вычисленная для каждого семейства YOLO, показывает, что YOLOv10 является самым энергоэффективным (40,8 мс) и в среднем занимает наименьший объем дискового пространства (33 Мб), а YOLOv3 – самым энергозатратным (63,2 мс) и тяжелым (110,8 Мб).

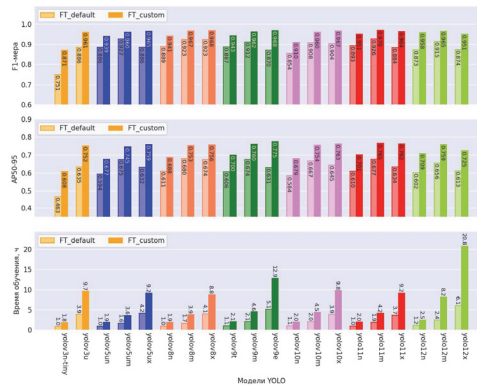


Рис. 4. F1-мера, AP50-95 и время обучения для вариантов FT<sub>default</sub> и FT<sub>custom</sub>

Таким образом, увеличение разрешения входных изображений повышает точность обнаружения малых объектов в сложных случаях, однако требует повышенных вычислительных затрат как в процессе обучения, так и в процессе вывода модели.

7.2. Анализ вычислительной производительности и размера

На рис. 5 представлено время вывода и размер моделей. Самая быстрая модель yolov3u-tiny имеет время вывода 8,5 мс (FPS = 118) при обработке изображений с разрешением 1024×1024, а самая медленная yolov3u – 117,8 мс (FPS = 8). Сверхбольшие модели оказались слишком тяжелыми для работы в реальном времени на графическом ускорителе NVidia Tesla T4, достигая значений FPS от 8 до 14. Средняя скорость вывода, вычисленная для каждого семейства YOLO, показывает, что YOLOv10 является самым энергоэффективным (40,8 мс) и в среднем занимает наименьший объем дискового пространства (33 Мб), а YOLOv3 – самым энергозатратным (63,2 мс) и тяжелым (110,8 Мб).

7.3. Анализ ошибок обнаружения объектов

После тонкой настройки как с параметрами по умолчанию, так и при сниженной скорости обучения и увеличенном разрешении входных изображений модели демонстрируют высокие показатели мер качества на проверочной и тестовой выборках набора данных DUT Anti-UAV, описанного в параграфе 2. При этом наилучший баланс качества и вычислительных затрат показывают модели среднего размера, например, тонко настроенная модель yolov11m.

Однако использование моделей в условиях, приближенных к реальным, часто связано с тем, что на вход поступают изображения без объектов БПЛА, но с объектами, визуально похожими на них – например, птицами, элементами фасадов зданий, облаками и т.д. В результате этого часто возникают ложные срабатывания, при

которых модель ошибочно обнаруживает БПЛА. Для проверки качества детекции моделей в таких ситуациях подготовлен тестовый набор данных Custom, описанный в параграфе 2. Полученные оценки качества детекции показывают, что тонкая настройка с подобранными параметрами позволяет получить модели, более устойчивые к ложным срабатываниям. На рис. 6 приведены примеры сложных изображений, на которых модель yolo11m, тонко настроенная при параметрах по умолчанию, ложно обнаруживает БПЛА (рис. 6а), а модель yolo11m, тонко обученная при параметрах, приведённых в параграфах 3 и 5, демонстрирует большую устойчивость к ложным срабатываниям (рис. 6б).

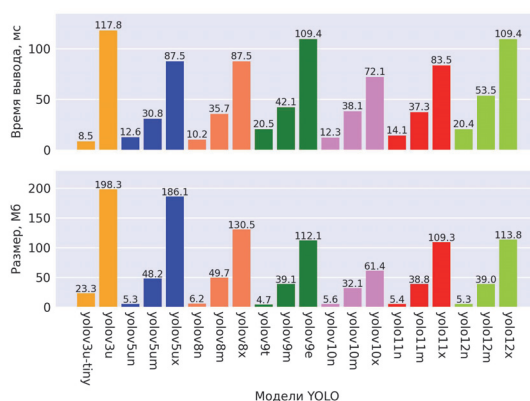


Рис. 5. Время вывода и размер моделей



Рис. 6. Примеры работы модели, тонко настроенной (а) при параметрах по умолчанию, (б) при подобранных параметрах

#### 7.4. Анализ общей производительности

Модели YOLO11 обеспечивают наилучший компромисс между качеством, скоростью вывода, объемом занимаемого дискового пространства и временем обучения. Они имеют надежную производительность при различных пороговых значениях IoU и размерах объектов. Это объясняется использованием в их архитектуре модуля перекрестной частичной сети с внутренним вниманием (C2PSA) для сохранения контекстной информации, что приводит к улучшению сходимости и общей производительности. Модели YOLO12, имея сопоставимый размер с моделями YOLO11, могут обеспечить более высокую точность, обусловленную сложными архитектурными решениями, включая модуль внимания области (Area Attention Module), остаточные эффективные агрегационные сети (R-ELAN) и алгоритм FlashAttention, но значительно снижают вычислительную производительность.

Наномодели демонстрируют достаточно высокий уровень точности обнаружения дронов, имея минимальное время вывода и размеры. Они являются лучшим решением для низкопроизводительных и экономичных систем видеоналиктики. Такие модели способны эффективно работать в реальном времени на CPU, не требуя высокопроизводительных графических ускорителей.

#### Заключение

Проведено комплексное сравнительное исследование производительности различных алгоритмов YOLO и двух вариантов их тонкой настройки для решения задачи обнаружения БПЛА. Продемонстрированы сильные и

слабые стороны каждой версии и семейства YOLO на основе анализа качества, вычислительной производительности и размера с использованием набора мер, таких как точность, полнота, F1-мера, средняя точность, время вывода, FPS, количество параметров и размер модели. Оптимальный баланс по совокупности оцениваемых параметров имеет семейство YOLO11. Наиболее инновационное и сложное по архитектуре семейство YOLO12 продемонстрировало сопоставимую точность, но низкую вычислительную производительность. С использованием сформированного в настоящей работе набора изображений с реальными условиями видеонаблюдения показано, что уменьшение скорости обучения и увеличение разрешения входных изображений при тонкой настройке моделей позволяет существенно повысить качество обнаружения небольших объектов в сложных условиях.

### References

- [1] Benarbia T, Kyamakya K. A literature review of drone-based package delivery logistics systems and their implementation feasibility. *Sustainability* 2022; 14(1): 360. DOI: 10.3390/su14010360.
- [2] Alsalam BHY, Morton K, Campbell D, Gonzalez F. Autonomous UAV with vision based on-board decision making for remote sensing and precision agriculture. 2017 IEEE Aerospace Conference, Big Sky, MT, USA, 4–11 March 2017: 1-12. DOI: 10.1109/AERO.2017.7943593.
- [3] Reedha R, Dericquebourg E, Canals R, Hafiane A. Transformer neural network for weed and crop classification of high resolution UAV images. *Remote Sens* 2022; 14(3): 592. DOI: 10.3390/rs14030592.
- [4] Chamola V, Kotes P, Agarwal A, Gupta N, Guizani M. A comprehensive review of unmanned aerial vehicle attacks and neutralization techniques. *Ad Hoc Netw* 2021; 111: 102324. DOI: 10.1016/j.adhoc.2020.102324.
- [5] Coveney S, Roberts K. Lightweight UAV digital elevation models and orthoimagery for environmental applications: Data accuracy evaluation and potential for river flood risk modelling. *Int J Remote Sens* 2017; 38(8-10): 3159-3180. DOI: 10.1080/01431161.2017.1292074.
- [6] Agyemang IO, Zhang X, Adjei-Mensah I, Acheampong D, Fiasam LD, Sey C, Yussif SB, Effah D. Automated visionbased structural health inspection and assessment for post-construction civil infrastructure. *Autom Constr* 2023; 156(2): 105153. DOI: 10.1016/j.autcon.2023.105153.
- [7] Turkmen Z, Kuloglu M. A new era for drug trafficking: Drones. *Forensic Sci. Addict. Res.* 2018; 2(1): 114-118. DOI: 10.31031/FSAR.2018.02.000539.
- [8] Mahdavi F, Rajabi R. Drone detection using convolutional neural networks. 2020 6th Iranian Conf on Signal Processing and Intelligent Systems (ICSPIS) 2020: 1-5. DOI: 10.1109/ICSPIS51611.2020.9349620.
- [9] Kassab M, Zitar RA, Barbaresco F, Seghrouchni AEF. Drone detection with improved precision in traditional machine learning and less complexity in single shot detectors. *IEEE Trans Aerosp Electron Syst* 2024; 60(4): 3847-3859. DOI: 10.1109/TAES.2024.3368991.
- [10] Cheng F, Liang Z, Peng G, Liu S, Li S, Ji M. An anti-UAV long-term tracking method with hybrid attention mechanism and hierarchical discriminator. *Sensors* 2022; 22(10): 3701. DOI: 10.3390/s22103701.
- [11] Huang B, Li J, Chen J, Wang G, Zhao J, Xu T. Anti-UAV410: A thermal infrared benchmark and customized scheme for tracking drones in the wild. *IEEE Trans Pattern Anal Mach Intell* 2023; 46(5): 2852-2865. DOI: 10.1109/TPAMI.2023.3335338.
- [12] Xie X, Xi J, Yang X, Lu R, Xia W. STTrack: Spatio-temporal-focused siamese network for infrared UAV tracking. *Drones* 2023; 7(5): 296. DOI: 10.3390/drones7050296.
- [13] Yu Q, Ma Y, He J, Yang D, Zhang T. A unified transformer-based tracker for anti-UAV tracking. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2023: 3036-3046. DOI: 10.1109/CVPRW59228.2023.00305.
- [14] Tong X, Zuo Z, Su S, Wei J, Sun X, Wu P, Zhao Z. ST-Trans: Spatial-temporal transformer for infrared small target detection in sequential images. *IEEE Trans Geosci Remote Sens* 2024; 62: 5001819. DOI: 10.1109/TGRS.2024.3355947.
- [15] Yang Z, Lan X, Wang H. Comparative analysis of YOLO series algorithms for UAV-based highway distress inspection: performance and application insights. *Sensors* 2025; 25(5): 1475. DOI: 10.3390/s25051475.
- [16] Elesawy A, Mohammed Abdelkader E, Osman H. A detailed comparative analysis of you only look once-based architectures for the detection of personal protective equipment on construction sites. *Eng* 2024; 5(1): 347-366. DOI: 10.3390/eng5010019.
- [17] Siewongsuwan K, Angsuseranee N, Wongkamchang P, Phasinam K. Comparative analysis of UAV detection and tracking performance: Evaluating YOLOv5, YOLOv8, and YOLOv8 DeepSORT for enhancing anti-UAV systems. *Edelweiss Appl Sci Tech* 2024; 8(5): 708-726. DOI: 10.55214/25768484.v8i5.1737.
- [18] Hakani R, Rawat A. Edge computing-driven real-time drone detection using YOLOv9 and NVIDIA Jetson Nano. *Drones* 2024; 8(11): 680. DOI: 10.3390/drones8110680.
- [19] Dewangan V, Saxena A, Thakur R, Tripathi S. Application of image processing techniques for UAV detection using deep learning and distance-wise analysis. *Drones* 2023; 7(3): 174. DOI: 10.3390/drones7030174.
- [20] Zhao J, Zhang J, Li D, Wang D. Vision-based anti-UAV detection and tracking. *IEEE Trans Intell Transp Syst* 2022; 23(12): 25323-25334. DOI: 10.1109/TITS.2022.3177627.
- [21] Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: Common objects in context. In Book: Fleet D, Pajdla T, Schiele B, Tuytelaars T, eds. *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V*. Cham: Springer International Publishing Switzerland; 2014: 740-755. DOI: 10.1007/978-3-319-10602-1\_48.
- [22] Zoph B, Ghiasi G, Lin T-Y, Cui Y, Liu H, Cubuk ED, Le QV. Rethinking pre-training and self-training. *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems 2020*: 3833-3845.
- [23] Li Y, Zhang H, Zhang Y. Rethinking training from scratch for object detection. *arXiv Preprint*. 2021. Source: <<https://arxiv.org/abs/2106.03112>>. DOI: 10.48550/arXiv.2106.03112.
- [24] Ultralytics. 2025. Source: <<https://github.com/ultralytics>>

## Приложение А

Табл. 4. Оценки средней точности

Модель	DUT Anti-UAV						Custom					
	AP50			AP50-95			AP50			AP50-95		
	Base	FT <sub>def</sub>	FT <sub>cust</sub>	Base	FT <sub>def</sub>	FT <sub>cust</sub>	Base	FT <sub>def</sub>	FT <sub>cust</sub>	Base	FT <sub>def</sub>	FT <sub>cust</sub>
yolov3u-tiny	0,331	0,726	0,897	0,192	0,463	0,608	0,224	0,573	0,716	0,122	0,329	0,448
yolov3u	0,593	0,911	0,973	0,389	0,635	0,752	0,434	0,815	0,815	0,270	0,560	0,641
yolov5un	0,415	0,902	0,958	0,262	0,594	0,677	0,311	0,827	0,885	0,184	0,499	0,589
yolov5um	0,555	<b>0,953</b>	0,972	0,374	0,675	0,745	0,475	<b>0,902</b>	0,927	0,291	0,641	0,684
yolov5ux	0,628	0,916	0,978	0,418	0,632	0,759	<b>0,512</b>	0,883	0,971	0,300	0,623	0,725
yolov8n	0,407	0,910	0,960	0,256	0,611	0,688	0,304	0,778	0,889	0,153	0,469	0,602
yolov8m	0,573	0,950	0,978	0,383	<b>0,680</b>	0,753	0,458	0,899	0,954	0,275	0,641	0,719
yolov8x	0,628	0,946	0,977	0,426	0,674	0,756	0,488	<b>0,902</b>	0,970	0,293	<b>0,655</b>	0,727
yolov9t	0,386	0,903	0,965	0,241	0,606	0,700	0,264	0,750	0,896	0,146	0,467	0,626
yolov9m	0,584	0,944	0,981	0,393	0,674	0,760	0,434	0,877	0,954	0,279	0,631	0,698
yolov9e	0,600	0,907	<b>0,982</b>	0,404	0,631	<b>0,775</b>	0,488	0,858	0,962	<b>0,302</b>	0,613	<b>0,735</b>
yolov10n	0,397	0,888	0,951	0,248	0,584	0,679	0,259	0,704	0,820	0,146	0,450	0,553
yolov10m	0,570	0,943	0,977	0,392	0,667	0,754	0,417	0,892	0,953	0,260	0,630	0,711
yolov10x	0,619	0,935	0,980	0,415	0,645	0,763	0,500	0,873	0,959	0,283	0,612	0,703
yolov1n	0,411	0,911	0,969	0,255	0,610	0,701	0,343	0,750	0,908	0,186	0,469	0,624
yolov1m	0,615	<b>0,953</b>	0,981	0,418	0,677	0,765	0,476	0,878	0,953	0,291	0,637	0,728
yolov1x	<b>0,644</b>	0,916	0,978	<b>0,434</b>	0,634	0,762	0,463	0,859	<b>0,978</b>	0,282	0,619	0,722
yolov12n	0,445	0,902	0,972	0,279	0,602	0,709	0,295	0,754	0,812	0,151	0,479	0,569
yolov12m	0,601	0,947	0,980	0,405	0,656	0,759	0,473	0,884	0,948	0,287	0,630	0,716
yolov12x	0,594	0,910	0,970	0,412	0,613	0,725	0,475	0,865	0,952	0,292	0,612	0,695

Табл. 5. Оценки F1-меры, вычислительной производительности и размера

Модель	F1-мера						Время вывода, мс		FPS, кадр/с		Кол-во парам., М	Размер, Мб
	DUT Anti-UAV			Custom			640	1024	640	1024		
	Base	FT <sub>def</sub>	FT <sub>cust</sub>	Base	FT <sub>def</sub>	FT <sub>cust</sub>						
yolov3u-tiny	0,341	0,751	0,871	0,173	0,610	0,671	4,6	8,5	217	118	12,1	23,3
yolov3u	0,570	0,886	0,961	0,438	0,778	0,791	52,6	117,8	19	8	103,7	198,3
yolov5un	0,431	0,886	0,939	0,305	0,772	0,797	10,0	12,6	100	79	2,6	5,3
yolov5um	0,536	<b>0,927</b>	0,960	0,429	0,810	0,878	12,2	30,8	82	32	25,1	48,2
yolov5ux	0,607	0,886	0,965	0,495	<b>0,815</b>	0,891	46,7	87,5	21	11	97,2	186,1
yolov8n	0,416	0,889	0,941	0,283	0,776	0,807	9,3	10,2	108	98	3,2	6,2
yolov8m	0,557	0,923	0,967	0,418	0,789	0,897	13,8	35,7	73	28	25,9	49,7
yolov8x	0,608	0,923	0,968	0,448	0,812	0,899	44,3	87,5	23	11	68,2	130,5
yolov9t	0,400	0,887	0,943	0,321	0,705	0,819	19,2	20,5	52	49	2,0	4,7
yolov9m	0,571	0,912	0,962	0,432	0,806	0,885	16,2	42,1	62	24	20,1	39,1
yolov9e	0,589	0,870	0,968	0,441	0,811	0,888	43,9	109,4	23	9	58,1	112,1
yolov10n	0,423	0,854	0,910	0,295	0,681	0,778	12,8	12,3	78	81	2,3	5,6
yolov10m	0,572	0,908	0,960	0,414	0,783	0,905	14,6	38,1	69	26	15,4	32,1
yolov10x	0,617	0,904	0,967	<b>0,540</b>	0,783	0,885	33,0	72,1	30	14	29,5	61,4
yolov1n	0,448	0,893	0,951	0,393	0,692	0,838	12,5	14,1	80	71	2,6	5,4
yolov1m	0,605	0,926	<b>0,970</b>	0,483	0,780	0,869	15,4	37,3	65	27	20,1	38,8
yolov1x	<b>0,638</b>	0,884	0,964	0,479	0,785	<b>0,933</b>	40,2	83,5	25	12	56,9	109,3
yolov12n	0,462	0,873	0,958	0,327	0,716	0,802	19,7	20,4	51	49	2,6	5,3
yolov12m	0,602	0,915	0,965	0,462	0,773	0,874	19,5	53,5	51	19	20,2	39,0
yolov12x	0,585	0,874	0,951	0,443	0,778	0,839	50,1	109,4	20	9	59,1	113,8

## Сведения об авторах

**Вычегжанин Сергей Владимирович**, 1988 года рождения, в 2021 году защитил диссертацию на соискание степени кандидата технических наук по специальности 05.13.17 «Теоретические основы информатики». В настоящий момент работает в должности старшего научного сотрудника на кафедре прикладной математики и информатики Вятского государственного университета. Сфера научных интересов: компьютерное зрение, обработка изображений, нейронные сети. E-mail: [vychegzhaninsv@gmail.com](mailto:vychegzhaninsv@gmail.com)

**Татаринова Александра Геннадьевна**, 1988 года рождения, в 2019 году защитила диссертацию на соискание степени кандидата технических наук по специальности 05.13.17 «Теоретические основы информатики». В настоящий момент работает в должности доцента кафедры прикладной математики и информатики Вятского государственного университета. Сфера научных интересов: компьютерное зрение, обучение нейронных сетей. E-mail: [tatarinova.alexg@gmail.com](mailto:tatarinova.alexg@gmail.com)

**Мышкин Роман Евгеньевич**, зам. генерального директора АО «НИИ СВТ» по научно-техническому развитию. E-mail: [myshkin.r@niisvt.ru](mailto:myshkin.r@niisvt.ru)

---

*Поступила в редакцию 06 мая 2025 г. Окончательный вариант – 03 сентября 2025 г.*

---

---

# Comparative analysis of neural network models for detecting unmanned aerial vehicles

S.V. Vychezhzhanin<sup>1</sup>, A.G. Tatarinova<sup>1</sup>, R.E. Myshkin<sup>2</sup>

<sup>1</sup> Vyatka State University, Ul. Moskovskaya 36, Kirov, 610000, Russia;

<sup>2</sup> JSC "SRI of Computer Engineering", Ul. Melnichnaya 31, Kirov, 610025, Russia

## Abstract

The paper presents a comparative analysis of YOLO models and their fine-tuning options for the task of real-time unmanned aerial vehicles detection. The quality and computational performance of different model families from YOLOv3 to YOLO12 and versions ranging from nano to extra-large are estimated. Their advantages and disadvantages are analyzed based on a wide range of measures, including precision, recall, F1-measure, average precision, inference time, and model size. Fine-tuning options for pre-trained models trained on the Common Objects in Context dataset are investigated for different input image resolutions and learning rates. We analyze detection errors of small unmanned aerial vehicles in complex environments using specialized datasets. The study showed that the YOLO11 family is the optimal in terms of the set of evaluated parameters, providing a balance between quality and computational performance. The highest quality scores for unmanned aerial vehicles detection on the test dataset were obtained using the fine-tuned yolo11m model with increased input image resolution and achieved values of F1=0.970, AP50=0.981, AP50-95=0.765.

**Keywords:** image processing, object detection, neural network algorithms, deep learning, transfer learning.

**Citation:** Vychezhzhanin SV, Tatarinova AG, Myshkin RE. Comparative analysis of neural network models for detecting unmanned aerial vehicles. *Computer Optics* 2026; 50(2): 1728. DOI: 10.18287/COJ1728.

---

## About authors

**Vychezhzhanin Sergey Vladimirovich** (b. 1988) defended his dissertation for the degree of Candidate of Technical Sciences in 2021, specializing in 05.13.17 "Theoretical Foundations of Informatics." At present, he is a senior researcher at the Department of Applied Mathematics and Informatics at Vyatka State University. Research interests are computer vision, image processing, neural networks. E-mail: [vychezhzhaninsv@gmail.com](mailto:vychezhzhaninsv@gmail.com)

**Tatarinova Alexandra Gennadievna** (b. 1988) defended her dissertation for the degree of Candidate of Technical Sciences in 2019, specializing in 05.13.17 "Theoretical Foundations of Informatics." She currently holds the position of docent at the Department of Applied Mathematics and Informatics at Vyatka State University. Her research interests include computer vision and neural network training. E-mail: [tatarinova.alexg@gmail.com](mailto:tatarinova.alexg@gmail.com)

**Myshkin Roman Evgenyevich** Deputy CEO for Research and Technical Development, JSC "SRI of Computer Engineering". E-mail: [myshkin.r@niisvt.ru](mailto:myshkin.r@niisvt.ru)

---

*Received May 06, 2025. The final version – September 03, 2025.*

---