

ИССЛЕДОВАНИЕ ПРИМЕНЕНИЯ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ YOLOv8 И YOLO11 ПРИ ДЕТЕКТИРОВАНИИ ТРАНСПОРТНЫХ СРЕДСТВ НА АЭРОФОТОСНИМКАХ

Ли Цзюмин¹, Мамонова Т.Е.²

¹ ТПУ, ИШИНЭС, ОИС, аспирант, гр. А4-36, e-mail: czyumin1@tpu.ru

² ТПУ, ИШИНЭС, ОИС, доцент, e-mail: stepte@tpu.ru

Аннотация

В данной работе выполнено детектирование транспортных средств на аэрофотоснимках с помощью модели сверточной нейронной сети YOLOv8 и YOLO11. При обучении данных моделей были проведены ряд экспериментов с различными гиперпараметрами размера партии (*batch*) с использованием наборов данных CARPK. В результате исследований показано, что модели YOLOv8n и YOLO11n имеют высокие показатели по используемым метрикам оценки точности и имеют способность точно обнаруживать и определять транспортные средства на аэрофотоснимках. Сравнительный анализ результатов показал, что обе модели имеют близкие результаты по используемым метрикам оценки точности, а по скорости вычислений модель YOLO11n превосходит модель YOLOv8n.

Ключевые слова: детектирование объектов, сверточная нейронная сеть, YOLOv8, YOLO11, CARPK

Введение

В настоящее время детектирование объектов на базе аэрофотоснимков имеет широкие перспективы применения в области контроля и безопасности транспортного потока. Задача детектирования объектов является важной частью применения технических устройств, на базе которых в качестве платформы в сочетании с технологией компьютерного зрения можно быстро воспринимать окружающую обстановку и сократить потребление человеческого ресурса. В связи с ограничением аппаратных ресурсов оборудования подобных технических устройств, для повышения скорости детектирования и обеспечения его точности применяются облегченные методы глубокого обучения, представленные сверточными нейронными сетями (CNN).

Методы детектирования объектов на основе CNN делятся на две группы:

1) Двухэтапные алгоритмы детектирования на основе генерации регионов, представленные R-CNN и Faster-RCNN.

2) Одноэтапные алгоритмы детектирования на основе регрессии, представленные серией YOLO (you only look once) [1].

Одноэтапные методы детектирования имеют более высокую скорость определения объекта и сквозную производительность по сравнению с двухэтапными методами. В связи с этим в настоящее время данные методы используются в качестве основного объекта исследования при детектировании объектов с применением результатов аэрофотосъемки.

Алгоритм серии YOLO с момента своего появления в 2015 году занял важное место в области детектирования объектов в реальном времени благодаря эффективной скорости и высокой точности. YOLOv8 был выпущен компанией Ultralytics в 2023 году, в отличие от предыдущих версий, в YOLOv8 используется оптимизированный компромисс между точностью и скоростью, что делает его идеальным для различных приложений [4]. YOLO11 является последней версией серии, которая улучшает способность модели к извлечению признаков и повышает точность детектирования за счет внедрения архитектурных оптимизаций, таких как новые модули C3k2 и C2PSA, сохраняя при этом легкие характеристики и оптимизированную скорость вывода.

Таким образом, в представляемой работе для детектирования транспортных средств на аэрофотоснимках на малых высотах используются модели сверточной нейронной сети YOLOv8 и YOLO11. Обучения данных моделей выполнены в наборе данных CARPK.

Описание архитектуры моделей YOLOv8 и YOLO11

В разделе Backbone YOLOv8 используется модуль C2F, который имеет меньшее количество параметров и лучшую способность извлечения признаков, что повышает точность обнаружения и эффективность вычислений. Модель YOLOv8 достигает наилучшей производительности в различных эталонных наборах данных. Например, модель YOLOv8n достигает mAP (средняя точность) 37,3 на наборе данных COCO.

YOLO11 является последней версией семейства YOLO, разработанной командой Ultralytics в США и Испании. Модель YOLO11 отличается от предыдущих версий YOLO, в том числе измененной архитектурой модели, улучшением извлечения признаков и оптимизацией методов обучения. По сравнению с моделью YOLOv8, в YOLO11 модуль C2F заменен на C3K2, что позволяет использовать различные размеры сверточных ядер и адаптироваться к более сложным задачам и требованиям многомасштабного обнаружения. За модулем SPPF был добавлен модуль C2PSA, целью которого является улучшение извлечения признаков за счет добавления механизма многоголового внимания.

Для обеспечения гибкого применения на аппаратных устройствах в различных сценариях, модели YOLOv8 и YOLO11 имеют разные размеры: *nano(n)*, *small(s)*, *medium(m)*, *large(l)* и *extra-large(x)*. Из-за наличия разноплановых параметров модели, высоких требований к вычислениям и хранению данных, высокой потребности в памяти компьютера и необходимости ускорения с помощью высокопроизводительной видеокарты для использования всех преимуществ моделей с большим количеством параметров, например, YOLO11m, YOLO11l и YOLO11x, их трудно развернуть на устройствах с ограниченной производительностью и вычислительными ресурсами.

Согласно данным из официальной документации Ultralytics, указанной в [2], соответствующие среднее значение средней точности *mAP*, количество параметров *Parameters*, вычислительная сложность *GFLOPs* и скорость *Speed* моделей YOLOv8n и YOLO11n приведены в таблице 1. Как видно из данных в таблице, YOLO11n имеет более высокую точность обнаружения по сравнению mAP с YOLOv8n, при использовании меньшего количества параметров и меньшего числа FLOP. На CPU YOLOv11n показывает более высокую скорость работы, на GPU он работает медленнее, чем YOLOv8n.

Таблица 1. Сравнение ключевых показателей моделей YOLOv8n и YOLO11n

Модель	mAP50-90,%	Parameters, M	GFLOP,s	Speed(CPU), ms	Speed(GPU), ms
YOLOv8n	37,3	3,2	8,7	80,4	0,99
YOLO11n	39,5	2,6	6,5	56,1	1,5

Эксперимент и результат

Для тестирования производительности моделей YOLOv8n и YOLO11n при обнаружении транспортных средств на аэрофотоснимках используется набор данных CARPK (Car Parking Lot Dataset), представленный в источнике [3]. Данный набор является крупномасштабным набором данных для обнаружения и подсчета автомобилей, собранный с высоты 40 м в условиях парковок с разнообразными сценариями (освещение, плотность автомобилей, углы обзора). Он содержит 1 448 изображений и информацию о 89 777 автомобилях в различных сценах для 4 различных парковок. Для обучения набор данных делится на обучающий, валидационный и тестовый набор по соотношению 7:1:2.

В таблице 2 показаны ключевые параметры для обучения модели. Для обучения нейронной сети были выбраны следующие параметры: размер входных изображений (*imgsz*) 640, оптимизатор (*optimizer*) SGD, импульс (*momentum*) 0,937 и начальная скорость обучения (*lr0*) 0,01. Размером партии (*batchsize*) является количество обучающих примеров, которые используются в одной итерации, он может влиять на производительность модели [4]. Для

выбора подходящего размера партии необходимо учитывать ограничения памяти, динамику обучения и требуемую производительность обобщения. Таким образом, для исследования влияния размера партии (*batchsize*) на эффективности каждой модели, при обучении моделей проведен ряд экспериментов с изменением данного параметра.

Таблица 2. Параметры для обучения моделей

	epochs	batchsize	imgsz	optimizer	momentum	lr0
YOLOv8n	100	8	640	SGD	0.937	0.01
	100	16	640	SGD	0.937	0.01
YOLO11n	100	8	640	SGD	0.937	0.01
	100	16	640	SGD	0.937	0.01

Далее будут рассмотрены метрики оценки, которые оценивают точность обнаружения моделей YOLOv8n и YOLO11, как точность (P), отзыв (R), средняя точность (AP) и среднее значение средней точности (mAP):

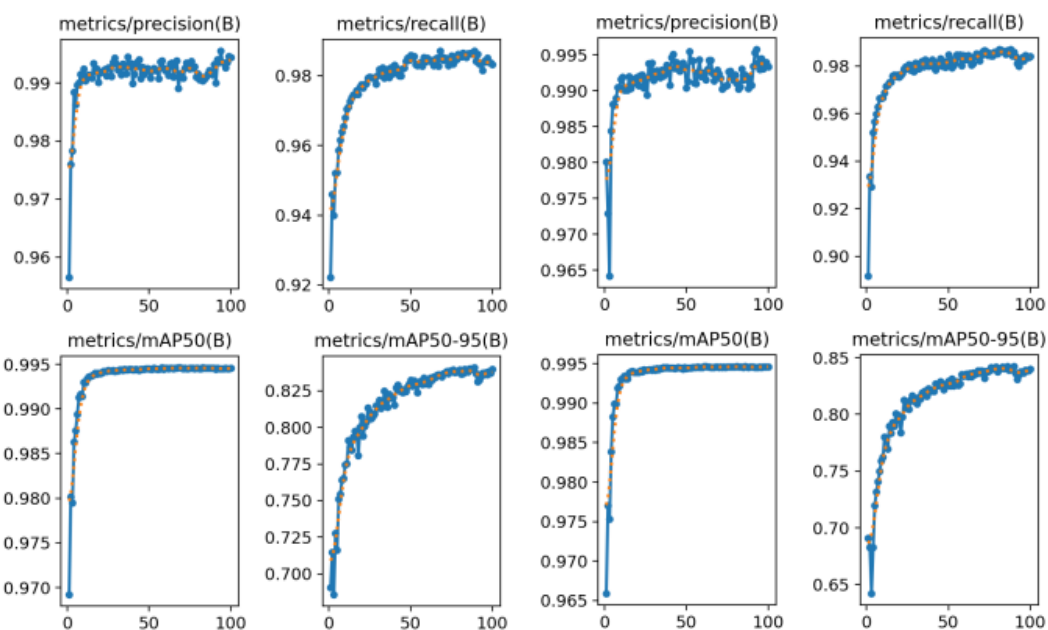
1. Точность (P): определяет долю истинно положительных прогнозов среди всех положительных прогнозов, оценивая способность модели избегать ложных срабатываний. Она используется для оценки точности предсказаний модели.

2. Отзыв (R): рассчитывает долю истинных положительных результатов среди всех реальных положительных результатов, оценивая способность модели обнаруживать все экземпляры класса. Он используется для оценки способности модели находить полную цель.

3. Средняя точность (AP): представляет собой площадь под кривой точность-отзыв и используется для измерения общей эффективности модели при различных пороговых значениях.

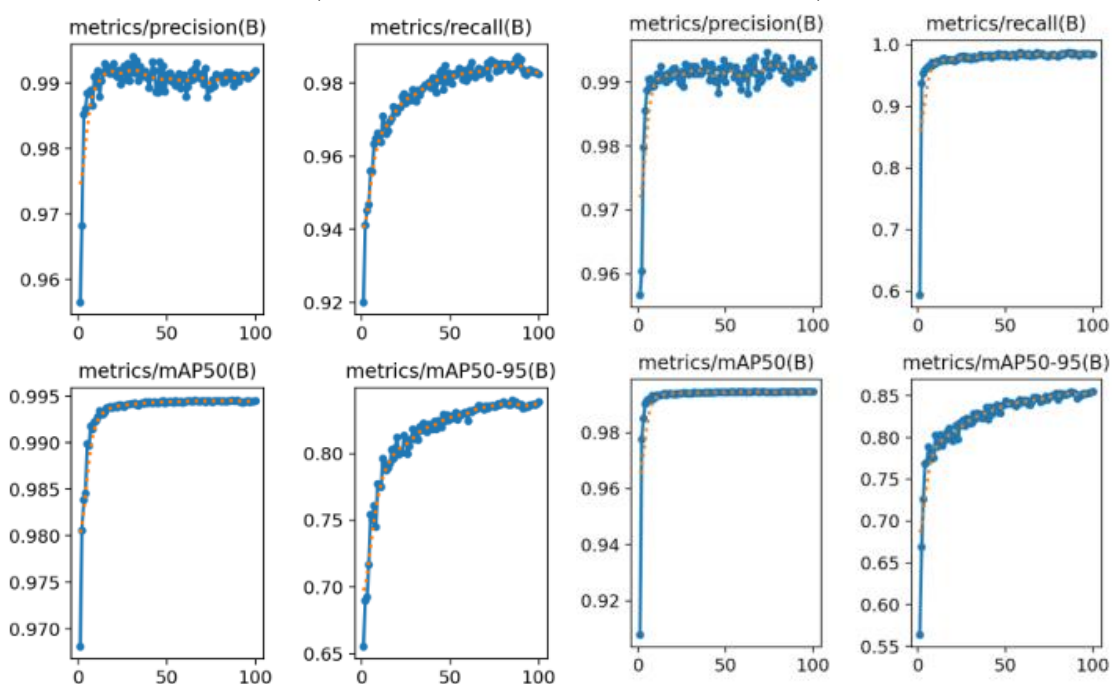
4. Среднее значение средней точности (mAP): используется для комплексной оценки эффективности модели в сценах обнаружения объектов с несколькими классами. mAP50 - средняя точность, рассчитанная при пороге intersection over union (IoU), равном 0,50.

После обучения моделей получены графики изменений показателей моделей YOLOv8n и YOLO11n, показанные на рисунке 1. Кривые Precision и Recall круто поднимаются, а затем сглаживаются, что показывает способность модели поддерживать высокую точность при высокой Recall, что отражает хорошую производительность модели. Кривые metrics/mAP50(B) и metrics/mAP50-90 (B) показывают, что модель сохраняет лучшую эффективность обнаружения при различных пороговых значениях. Сравнивая изменения в кривых metrics/mAP50 (B) и metrics/mAP50 (B) при разных размерах партии, можно увидеть, что в начале обучения показатели оценки растут быстрее, но в конце они сходятся к одному уровню.



а)

б)



в)

г)

Рис. 1. Показатели оценки для моделей при различных значениях размера партии:

а) YOLOv8n, batch = 8, б) YOLOv8n, batch = 16,

в) YOLO11n, batch = 8, г) YOLO11n, batch=16

Значения показатели оценки валидации моделей, приведены в таблице 3.

Таблица 3. Показатели оценки моделей.

	batch	P, %	R, %	mAP50, %	mAP50-95, %	Parameters(M)	FLOPs	FPS
YOLOv8n	8	99,2	98,8	99,5	84,3	3,2	8,7G	12
	16	99,2	98,6	99,5	84,2	3,2	8,7G	12
YOLO11n	8	99,1	98,5	99,45	84,00	2,58	6,3G	16
	16	99,3	98,4	99,50	85,50	2,58	6,3G	16

Исходя из результатов точности (P), отзыва (R) и среднего значения средней точности (mAP) приведенные в таблице 4, было выявлено, что модели YOLOv8n и YOLO11n модель обладает высокой точностью обнаружения. В результате проведения сравнения показателей точности (P) и среднего значения средней точности (mAP50, mAP50-95) при различных размерах партии видно, что модель YOLO11n показала лучшие результаты при значении размера партии batch = 16. Результат сравнения показателей количества параметров, вычислительной сложности (FLOP) и количества кадров в секунду (FPS) показал, что YOLO11n имеет более высокую скорость обнаружения. Таким образом, модель YOLO11n лучше модели YOLOv8n в задаче обнаружения транспортных средств на маловысотных аэрофотоснимках.

Далее на тестовом наборе было выполнено тестирование модели и YOLO11n, обученных при значении размера партии batch=16. Результаты показателей оценки на тестовом наборе, представленные в таблице 4 показали, что модель YOLO11n имеют высокую точность на тестовом наборе. Можно сделать вывод, что модели обладают значительной обобщающей способностью.

Таблица 4. Показатели оценки моделей на тестовом наборе

	P, %	R, %	mAP0.5, %	mAP0.5-0.95, %
YOLO11n	99,20	98,90	99,50	84.9

На рисунке 2 показаны результаты визуализации на тестовом наборе с использованием обученной модели YOLO11n. На основе Python и Tkinter реализована система обнаружения транспортных средств, которая поддерживает функции чтения изображений из папок, запуска модели YOLO11 для обнаружения транспортных средств и вывода количества обнаруженных транспортных средств. Видно, что все автомобили обнаружены успешно на изображении с различными сценами.

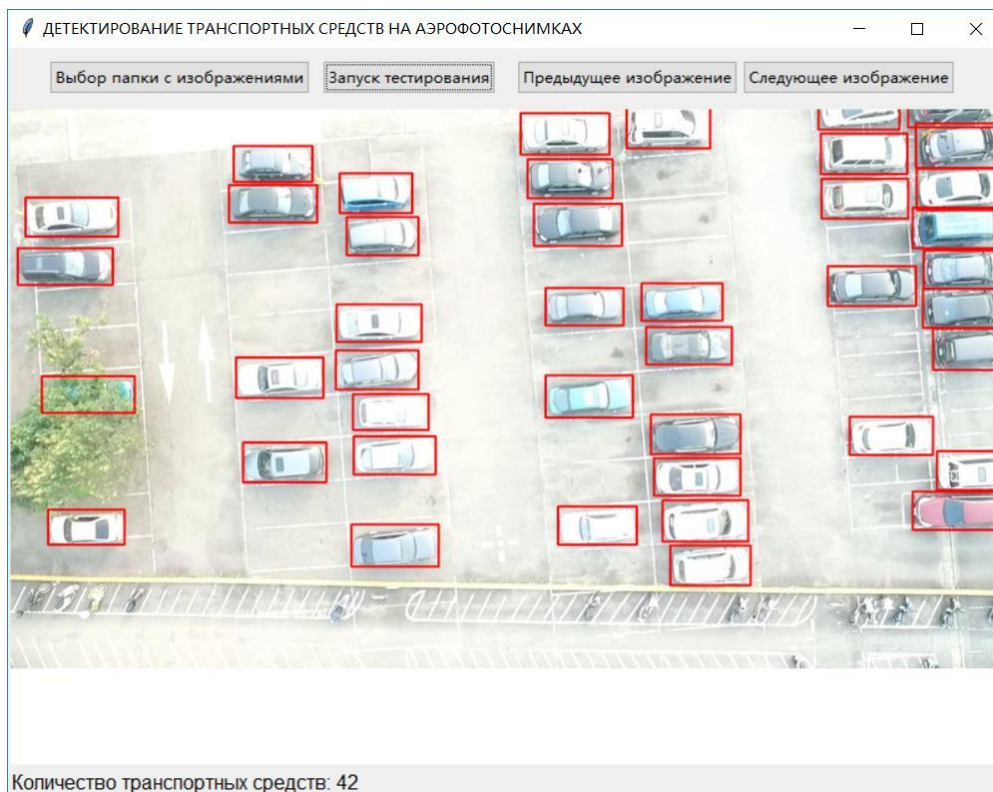


Рис. 2. Результаты визуализации с использованием модели YOLO11

Заключение

В данной работе реализовано исследование моделей сверточных нейронных сетей YOLOv8n и YOLO11n для задач обнаружения транспортных средств на аэрофотоснимках, выполненных на расстоянии 40 м. Результат оценки и визуализации показывают, что данные модели обладают достаточной точностью в задаче обнаружения транспортных средств на аэрофотоснимках. В задаче обнаружения транспортных средств на маловысотных аэрофотоснимках YOLOv8n не уступает YOLO11n по точности обнаружения, однако показатели скорости показывают, что YOLOv8n имеет более низкую скорость обнаружения, чем YOLO11n. Полученный результат по скорости вычислений подтверждает эффективность нейронной сети YOLO11n для быстрого и точного обнаружения объектов. Можно сказать, что модель YOLO11n обеспечивает более сильную техническую поддержку для интеллектуальных систем мониторинга дорожного движения с помощью технических устройств и аппаратных средств, работающих на низких расстояниях от земли.

Список использованных источников

1. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection // Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR). – 2016. – P. 779-788.
2. Ultralytics, «Object Detection Models - YOLO11 Documentation» [Электронный ресурс]. – URL: docs.ultralytics.com/tasks/detect/#models/ (дата обращения: 05.03.2025)
3. Car Parking Lot Dataset (CARPK). [Электронный ресурс]. – URL: lafi.github.io/LPN/ (дата обращения: 05.03.2025)
4. Размер партии. [Электронный ресурс]. – URL: ultralytics.com/ru/glossary/batch-size (дата обращения: 05.03.2025)
5. Ultralytics YOLOv8. [Электронный ресурс]. – URL: docs.ultralytics.com/ru/models/yolov8/ (дата обращения: 05.03.2025)