

ПРИМЕНЕНИЕ МОДЕЛЕЙ СЕМЕЙСТВА YOLO ДЛЯ ДЕТЕКЦИИ МАЛЫХ ОБЪЕКТОВ

А. В. Каптуров¹⁾, О. О. Колб²⁾, Д. А. Гутлыев³⁾

¹⁾Объединенный институт проблем информатики НАН Беларуси, Минск, Беларусь, ultraluxe25@mail.ru

²⁾Белорусский государственный университет, Беларусь, Минск, mwkolb1@gmail.com

³⁾Белорусский государственный университет, Беларусь, Минск, daykanot@gmail.com

Рассмотрена проблема детекции малых объектов в компьютерном зрении, обусловленная её важностью для множества практических задач, а также значительными трудностями, возникающими при обработке малых объектов современными моделями. Анализируются определения и особенности детекции малых объектов, рассматриваются основные методы обучения, а также различные алгоритмы, используемые для этой задачи. Проведено тестирование моделей на реальных данных.

Ключевые слова: компьютерное зрение; детекция; YOLO; малые объекты; метрики; дроны; нейронные сети; обучение моделей.

APPLICATION OF YOLO FAMILY MODELS FOR DETECTION OF SMALL OBJECTS

A. V. Kapturov¹⁾, A. A. Kolb²⁾, D. A. Gutlyyev³⁾

¹⁾United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus, ultraluxe25@mail.ru

²⁾Belarussian state university, Belarus, Minsk, mwkolb1@gmail.com

³⁾Belarussian state university, Belarus, Minsk, daykanot@gmail.com

The problem of small object detection in computer vision is considered, due to its importance for many practical tasks, as well as significant difficulties encountered in processing small objects by modern models. Definitions and peculiarities of small object detection are analyzed, the main training methods are considered, as well as various algorithms used for this task. Testing of models on real data is carried out.

Keywords: computer vision; detection; YOLO; small objects; metrics; drones; neural networks; model training.

Детекция объектов одна из самых сложных и важных задач современного компьютерного зрения. Детекция малых объектов – это одно из направлений детекции, которое имеет большое теоретическое и практиче-

ское значение в различных сценариях, таких как наблюдение, обнаружение беспилотников, обнаружение пешеходов на дороге, обнаружение знаков дорожного движения для беспилотного транспорта и др. Несмотря на значительные результаты в общей детекции объектов, в детекции малых объектов наблюдаются значительные трудности. Даже у самых передовых моделей детекции обнаруживается большой разрыв на метриках при детекции обычных объектов и малых объектов. [1].

Для решения проблемы детекции малых объектов были предложены различные работы, которые можно разделить на шесть групп: ориентированные на выборку (sample-oriented) методы, учитывающие масштаб (scale-aware) методы, основанные на внимании (attention-based) методы, методы имитации признаков (feature-imitation), методы моделирования контекста (feature-imitation), методы фокусирования и детекции (focus-and-detect) [2].

Актуальность детекции малых объектов возрастает в поисково-спасательных операциях, особенно при использовании беспилотных летательных аппаратов. К примеру, в 2017 году Береговая охрана США участвовала в 15 951 операции, в ходе которых погибло 618 человек. Одной из самых трудных задач является поиск и локализация пропавших людей и объектов. Современные технологии и искусственный интеллект, могут существенно улучшить этот процесс.

Теоретические основы

В данной работе рассмотрен набор данных «Aerial-drone Floating Objects» (AFO) [3], созданный для детекции малых объектов в морской среде. Набор содержит 3647 изображений с почти 40 000 отмеченными объектами и доступен для академического использования, что способствует развитию более эффективных систем для поисково-спасательных операций.

В наборе данных AFO более 99% объектов имеют площадь менее 1% от общей площади изображения. Также присутствуют множество изображений с высокой плотностью объектов — более 30% изображений содержат более 20 экземпляров объектов. Датасет AFO включает 6 классов надводных объектов: люди в воде (human), доски для виндсерфинга или сапбординга (wind/sup-board), лодки (boat), плавучие буи (buoy), парусные лодки (sailboat) и каяки (kayak).

Таким образом, AFO может быть использован для разработки и проверки методов обнаружения мелких объектов [4].

Для обучения модели детекции малых объектов были выбраны модели YOLO11 small и YOLO11 medium компании Ultralytics[5].

Метрики и скорость работы моделей YOLO11.

Модель	Размер (пиксели)	mAPval 50-95	Скорость CPU ONNX (мс)	Скорость T4 TensorRT10 (мс)	Параметры (М)	FLOPs (B)
YOLO11n	640	39.5	56.1 ± 0.8	1.5 ± 0.0	2.6	6.5
YOLO11s	640	47.0	90.0 ± 1.2	2.5 ± 0.0	9.4	21.5
YOLO11m	640	51.5	183.2 ± 2.0	4.7 ± 0.1	20.1	68.0
YOLO11l	640	53.4	238.6 ± 1.4	6.2 ± 0.1	25.3	86.9
YOLO11x	640	54.7	462.8 ± 6.7	11.3 ± 0.2	56.9	194.9

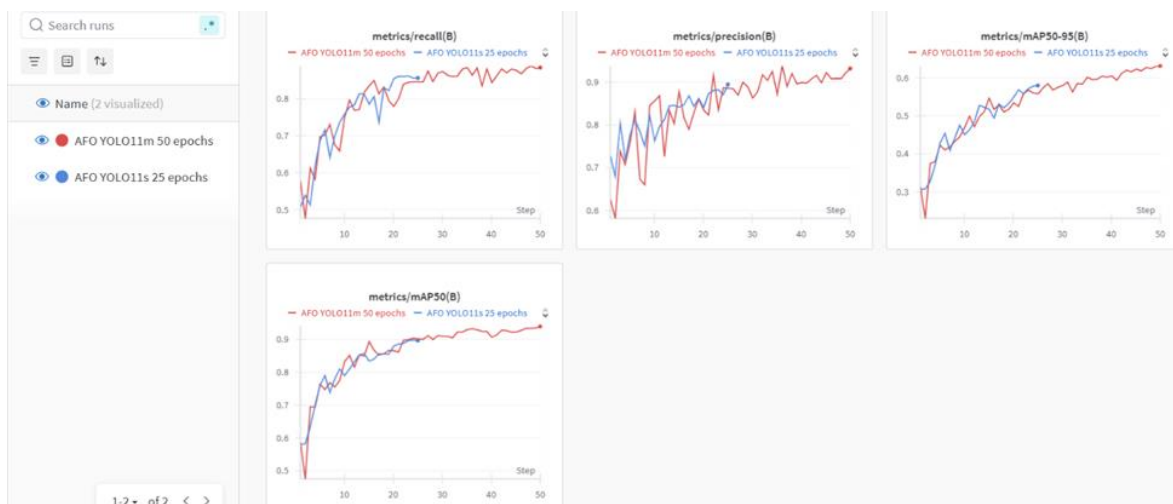
Дадим краткое описание таблицы. Названия моделей из семейства YOLO11 (n — nano, s — small, m — medium, l — large, x — extra large). Разрешение входного изображения (640x640 пикселей). mAPval 50-95 - это средняя точность (mean Average Precision) в диапазоне IoU от 0.5 до 0.95. Скорость CPU ONNX (мс) – это время обработки на CPU с использованием формата ONNX (в миллисекундах). Скорость T4 TensorRT10 (мс) – это время обработки на GPU NVIDIA T4 с использованием TensorRT10 (в миллисекундах). Параметры (М) – это количество параметров модели в миллионах. FLOPs (B) – это количество операций с плавающей точкой в миллиардах.

Практические результаты

Модель YOLO11s обучалась на 25 эпохах, тогда как YOLO11m на 50 эпохах. Мониторинг метрик производился с помощью TensorBoard во время обучения и с помощью W&V по результатам обучения обеих моделей.

Рисунок состоит из четырех графиков, которые показывают изменение ключевых метрик в процессе обучения моделей YOLO11 на наборе данных AFO. В эксперименте сравниваются AFO_YOLO11M_50_epoch (красная линия) — обучение в течение 50 эпох и AFO_YOLO11S_25_epoch (синяя линия) — обучение в течение 25 эпох.

Каждый график отображает одну из метрик в зависимости от шага обучения (step). Верхний левый график - метрика recall (полнота). Верхний посередине график - метрика precision (точность). Нижний график - метрика mAP0.5 (средняя точность при IoU = 0.5). Верхний правый график - метрика mAP50-95 (средняя точность, усредненная для IoU от 0.5 до 0.95). Рассмотрим отдельно каждый график подробнее.



Метрики производительности YOLO11s и YOLO11m на наборе данных AFO

Recall (верхний левый график): обе модели показывают рост полноты с увеличением числа шагов, модель с 50 эпохами (красная линия) достигает более высокого значения (примерно 0.9) по сравнению с моделью на 25 эпох (синяя линия, около 0.85). Более длительное обучение позволяет модели лучше находить объекты, особенно мелкие, что важно для набора AFO, где 99% объектов имеют площадь менее 1% изображения.

Precision (график посередине): точность обеих моделей примерно одинаковая и достигает 0.9 на 25-й эпохе.

Колебания могут быть связаны с тем, что модель пытается найти баланс между метриками precision и recall, особенно в условиях сложного набора данных с мелкими объектами.

mAP50 (нижний график): метрика mAP50 (средняя точность при IoU = 0.5) растет для обеих моделей. Модель с 50 эпохами достигает значения около 0.95, что значительно выше, чем у модели с 25 эпохами (около 0.9).

mAP50-95 (верхний график справа): усредненная метрика mAP для диапазона IoU от 0.5 до 0.95 также выше у модели с 50 эпохами (около 0.65) по сравнению с 25 эпохами (около 0.55). Эта метрика более строгая, так как учитывает точность локализации объектов на разных уровнях IoU. Разница между моделями подтверждает, что дополнительное обучение улучшает не только обнаружение, но и точность локализации.

Заключение

Модель, обученная в течение 50 эпох, демонстрирует лучшие результаты по всем метрикам по сравнению с моделью, обученной 25 эпохами. Это ожидаемо, поскольку большее количество эпох позволяет модели лучше адаптироваться к сложным данным, особенно при работе с мел-

кими объектами, как в наборе AFO. Метрики для модели с 50 эпохами показывают меньшую волатильность, особенно метрика precision, что свидетельствует о более стабильном обучении.

Несмотря на хорошие результаты по большинству метрик, метрика mAP50-95 остается относительно низкой (около 0.55–0.65, даже для 50 эпох). Это может быть связано с трудностями классификации мелких объектов и их отличия от фона, как отмечается в статье [1]. Оба эксперимента показывают улучшение метрик с увеличением числа эпох, что подтверждает эффективность архитектуры YOLO11 для детекции малых объектов, в частности для обнаружения объектов в морской среде.

Для дальнейшего улучшения производительности стоит увеличить количество эпох (например, до 75 или 100) и проанализировать, продолжат ли метрики расти. Также стоит рассмотреть применение методов аугментации данных или внедрение механизмов внимания для повышения точности классификации мелких объектов. Поскольку метрика mAP50-95 остается низкой, полезно исследовать, какие классы объектов вызывают наибольшие ошибки, и сосредоточиться на улучшении их обнаружения.

Библиографические ссылки

1. You Only Look Once: Unified, Real-Time Object Detection [Electronic resource] / J. Redmon [et al.] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. 2016. Mode of access: <https://ieeexplore.ieee.org/document/7780460> (date of access: 27.03.2025).
2. Cheng G., Yuan X., Yao X., Yan K., Zeng Q., Xie X. Towards Large-Scale Small Object Detection: Survey and Benchmarks // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023. Т. 45. №. 11. С. 13467-13488.
3. AFO - Aerial dataset of floating objects [Электронный ресурс] / Режим доступа: <https://www.kaggle.com/datasets/jangsienicajzkowy/afo-aerial-dataset-of-floating-objects/code>. (дата доступа 27.03.2025)
4. Gqsienica-Józkowy, J., Knapik, M., & Cyganek, B. (2021). An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance. Integrated Computer-Aided Engineering, 1–15.
5. Ultralytics YOLO11 Modes [Электронный ресурс] / Режим доступа: <https://docs.ultralytics.com/modes/>. (дата доступа 27.03.2025)