

ПАВЛОВ
Максим Павлович

бакалавриат, Петрозаводский государственный университет
(Петрозаводск, Российская Федерация)
maksim_pavlov_2003@list.ru

СРАВНЕНИЕ МОДЕЛЕЙ YOLOV5 ДЛЯ ЗАДАЧИ ДЕТЕКЦИИ РАДУЖНОЙ ФОРЕЛИ В ВИДЕОПОТОКЕ

Научный руководитель:

Марахтанов Алексей Георгиевич

Статья поступила: 11.11.2022;

Принята к публикации: 20.11.2022;

Размещена в сети: 25.12.2022

Аннотация. Разнообразие моделей и параметров обучения затрудняет поиск оптимального решения задачи детекции объектов. В данной статье рассматриваются модели YOLOv5. Для каждой из них замеряются ключевые метрики обучения, и с помощью метода прямоугольников строится фронт Парето оптимальных решений. На его основании выбирается лучшая по точности и скорости детекции радужной форели модель. Автор приходит к выводу, что обучение моделей зависит не только от их архитектуры, аппаратных возможностей системы, но и от количества классов детекции.

Ключевые слова: нейронные сети, YOLOv5, детекция радужной форели, оптимальное множество, фронт Парето

Для цитирования: Павлов М. П. Сравнение моделей YOLOv5 для задачи детекции радужной форели в видеопотоке // StudArctic Forum. 2022. Т. 7, № 4. С. 35—40.

Радужная форель – один из многочисленных представителей семейства лососёвых. Её активно вылавливают в естественной среде обитания, выращивают в искусственных садках рыбных хозяйств. В последнее время технологии машинного обучения всё чаще используются в рыбопромышленной отрасли, а разработчики сталкиваются с ключевой задачей – детекцией объекта в видеопотоке. Существует достаточный набор методов решения данной задачи. Например, можно воспользоваться популярным алгоритмом YOLO (You Only Look Once) и одной из его последних реализаций YOLOv5 [Yu Yongcan] с широким разнообразием моделей. Каждая из них имеет свою собственную конфигурацию нейросети, но реализует один и тот же алгоритм - YOLO. Как следствие, специалисты задаются вопросом выбора модели для обучения нейросети и оптимальностью ее применения для конкретной задачи. Рассмотрим подробнее модели: YOLOv5n, YOLOv5s, YOLOv5sb, YOLOv5m, YOLOv5mb, YOLOv5l, YOLOv5lb и YOLOv5x для обнаружения радужной форели.

Все рассматриваемые модели обучались на одних и тех же данных в течение 100-300 эпох. Этого было достаточно для достижения сходимости процесса обучения. Все вычисления производились на видеокарте GeForce RTX 2060 с 6 Гб видеопамяти, CUDA 11.7 и Pytorch 1.12.1. В процессе обучения отсутствовали сторонние запущенные процессы. Модели обучались на предобученных весах yolov5. Набор данных состоит из 33 871 изображений рыбы, из которых 92,57% получены из открытых источников (ImageNet [Deng], Open Images [Kuznetsova], A Large-Scale Dataset for Fish Segmentation and Classification [Ulucan], Brackish Dataset [Pedersen], Fish Dataset [Fish Dataset]), а 7,43% собраны автором и его коллегами с помощью COCO Annotator [COCO Annotator]. Изображения, полученные из открытых источников, содержат фотографии разных видов рыб при различных условиях съемки. Среди изображений из собственного набора данных содержатся изображения радужной форели в условиях подводной и надводной съёмки. Каждый набор данных разделён на обучающую и тестовую выборки в пропорции 85% к 15%. В рамках данного исследования основное внимание уделяется точности детекции на собственных данных, поскольку получаемая в результате нейронная сеть предназначена для детекции форели в условиях, сходных с условиями, в которых были получены собственные данные. Также использовался autobach

YOLOv5, который при каждом обучении подбирал наиболее оптимальный размер batch size для моделей.

Таблица 1

Ключевые метрики обучения моделей YOLOv5 с разрешением получаемых изображений 640x640 пикселей

Название модели	Кол-во эпох	batch size	FPS, ms/img	FLOPS	Видео-память для детекции объектов	mAP 0.5:0.95, весь датасет	mAP 0.5:0.95, фотографии и под водой	mAP 0.5:0.95, фотографии и над водой
YOLOv5n	300	43	1.6	4.2	979 Мб	0.779	0.663	0.934
YOLOv5s	200	25	1.7	15.9	1027 Мб	0.832	0.675	0.963
YOLOv5s6	300	22	1.8	16.3	1065 Мб	0.827	0.704	0.974
YOLOv5m	200	13	4.0	48.2	1169 Мб	0.865	0.698	0.977
YOLOv5m6	300	12	4.1	49.3	1287 Мб	0.858	0.711	0.977
YOLOv5l	200	7	6.4	108.2	1379 Мб	0.873	0.703	0.981
YOLOv5l6	200	6	6.6	110.5	1689 Мб	0.848	0.713	0.982
YOLOv5x	200	3	12.5	203.8	1789 Мб	0.856	0.698	0.982

В ходе сравнения будут рассмотрены следующие метрики: количество миллисекунд на детекцию одного кадра - FPS; FLOPS; видеопамять; mAP 0.5 и mAP 0.5:0.95 на всём датасете, на фотографиях под водой, над водой. Для их подсчёта использовались методы, предложенные авторами YOLOv5 [YOLOv5 Documentation]. FPS замерялся при детекции объектов с помощью val.py с ключами «task speed» и «batch 32». Для подсчёта FLOPS использовалась python библиотека thor, количество параметров (весов нейросети) выводится при старте обучения для каждого слоя, в том числе и bias-ов. Видеопамять, необходимая нейросети для детекции объектов, замерялась при обнаружении уже обученной моделью одного набора изображений с помощью detect.py. Узнать занятую определенным процессом видеопамять можно с помощью nvidia-smi. Точность моделей подсчитывалась после обучения программой val.py с ключами «conf 0.001» и «iou 0.65». Для того, чтобы показатели mAP были подсчитаны на трёх множествах: всём датасете, фотографиях под водой и над водой, были созданы отдельные yaml файлы, указывающие val.py, на каких данных производить замеры в конкретных ситуациях.

Таблица 2

Ключевые метрики обучения моделей YOLOv5 с разрешением получаемых изображений 1280x1280 пикселей

Название модели	Кол-во эпох	batch size	FPS, ms/img	FLOPS	Видеопамять для детекции объектов	mAP 0.5:0.95, весь датасет	mAP 0.5:0.95, фотографии под водой	mAP 0.5:0.95, фотографии над водой
YOLOv5s6	200	5	6.9	16.3	1167 Мб	0.849	0.704	0.975
YOLOv5m6	100	2	16.0	49.3	1391 Мб	0.847	0.722	0.97

Занесём полученные метрики в таблицу (Таблица 1, Таблица 2). Заметим, что уже на этапе обучения YOLOv5x размер batch size снизился до 3. Это непосредственно влияет на количество подаваемых при обучении изображений, что делает обучение YOLOv5x6 на использованной в эксперименте видеокарте нецелесообразным: batch size был бы равным 1, а обучение заняло бы около двух-трёх недель для 200 эпох. Аналогичная ситуация с YOLOv5l6 при размере изображения 1280x1280 пикселей.

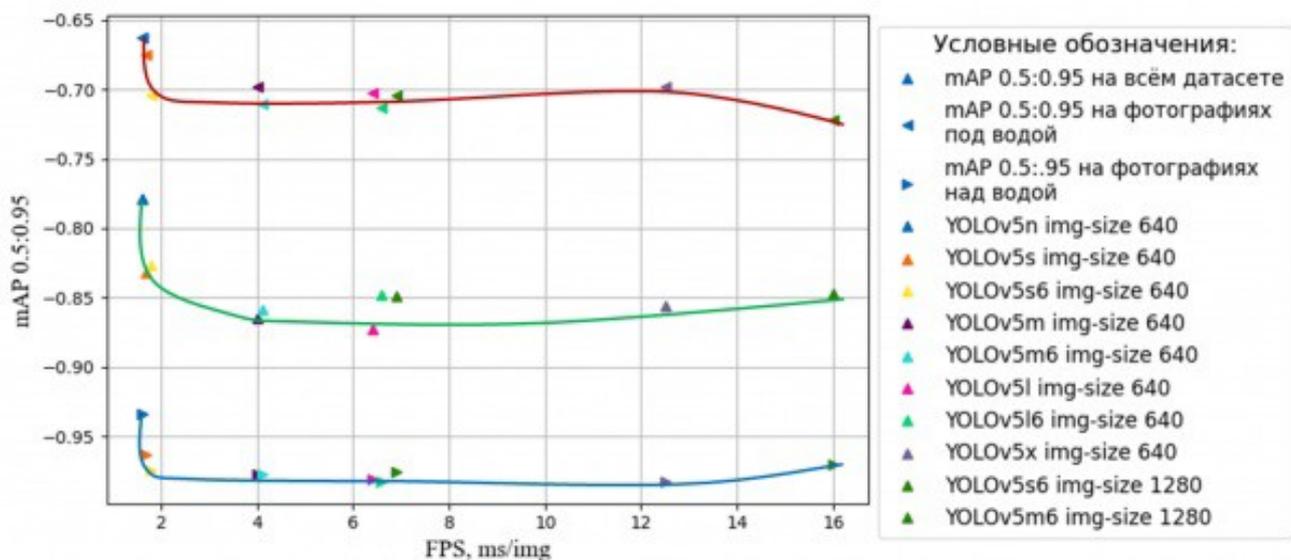


Рис. 1. Зависимость показателей mAP 0.5:0.95 от FPS детекции радужной форели для всех рассматриваемых моделей YOLOv5

Рассмотрим полученные данные с точки зрения оптимизационной задачи. Выделим математические множества, и, используя набор методов решения решений [Летов], найдём для каждого подходящее подмножество. Необходимо прийти к одной модели YOLOv5, оптимальной по двум критериям: частоте кадров детекции и точности обнаружения радужной форели. «Решение является оптимальным по Парето, если при замене его любым другим допустимым решением нельзя добиться улучшения значения хотя бы одного из критериев, не ухудшив при этом значения какого-то другого» [Муханов]. Сведём задачу к трём парето-оптимальным множествам, каждое из которых будет оптимальным решением для своего множества: mAP 0.5:0.95 на всём датасете, фотографиях под водой, над водой. Таким образом, задача будет проходить в плоскости решений, а не в n-мерном линейном пространстве, где n – число критериев оптимизации. Заметим, что если объединить точки каждого множества условными кривыми, то они будут вогнутыми вверх. Сделаем инверсию точек множества относительно оси Oy. Показатели mAP станут отрицательными (Рисунок 1), но после нахождения оптимальных подмножеств используем обратную инверсию и получим желаемый результат.

Таблица 3

Представление данных в виде сравнительной таблицы метода прямоугольников восстановления Паретовой границы

Модель	Координаты моделей в каждом из множеств в виде (mAP 0.5:0.95; FPS)	Знак
YOLOv5n	(-0.779; 1.6) (-0.663; 1.6) (-0.934; 1.6)	+++
YOLOv5s	(-0.832; 1.7) (-0.675; 1.7) (-0.963; 1.7)	+++
YOLOv5s6 img-size 640	(-0.827; 1.8) (-0.704; 1.8) (-0.974; 1.8)	-++
YOLOv5s6 img-size 1280	(-0.849; 6.9) (-0.704; 6.9) (-0.975; 6.9)	---
YOLOv5m	(-0.865; 4.0) (-0.698; 4.0) (-0.977; 4.0)	+ - +
YOLOv5m6 img-size 640	(-0.858; 4.1) (-0.711; 4.1) (-0.977; 4.1)	- + -
YOLOv5m6 img-size 1280	(-0.847; 16.0) (-0.722; 16.0) (-0.974; 16.0)	- + -
YOLOv5l	(-0.873; 6.4) (-0.703; 6.4) (-0.981; 6.4)	+ - +
YOLOv5l6	(-0.848; 6.6) (-0.713; 6.6) (-0.982; 6.6)	- + +
YOLOv5x	(-0.856; 12.5) (-0.698; 12.5) (-0.982; 12.5)	---

Для нахождения оптимального множества и фронта Парето используем метод прямоугольников, описанный в [Муханов: 13]. Зададим множество точек моделей в виде таблицы, состоящей из координаты модели в каждом из множеств и знака – входит ли она в оптимальное подмножество. Изначально последний столбец пустой и заполняется в ходе применения метода прямоугольников:

1. Просмотрим все оставшиеся точки. Если их первая координата больше текущей фиксированной точки, то исключаем данную модель (ставим знак «-»).
2. Выберем точку с минимальной второй координатой (отражающей миллисекунды на детекцию одного кадра). Зафиксируем значение первой координаты модели (mAP).
3. Возьмем точку с минимальным значением первой координаты, которые ещё не содержат знак «+» или «-». И исключаем все оставшиеся точки, у которых вторая координата больше.
4. Пока есть точки, не содержащие знак «+» или «-», продолжаем шаги 1-3.

После выполнения метода прямоугольников получим сравнительную таблицу (Таблица 3). На основании последнего столбца построим Паретову границу: проведём три кривых через точки моделей для каждого множества, содержащих два или более плюса (Рисунок 2). Полученные кривые являются решениями оптимальными по Парето: при замене любой модели на другую, не лежащую на кривой, нельзя добиться улучшения значения хотя бы одного из критериев, не ухудшив при этом значения какого-то другого. Заметим, что множество оптимальных решений состоит только из моделей, обученных на разрешении 640x640. Можно сделать вывод, что 6 Гб видеопамати недостаточно для эффективного обучения моделей YOLOv5 на большом разрешении входных изображений.

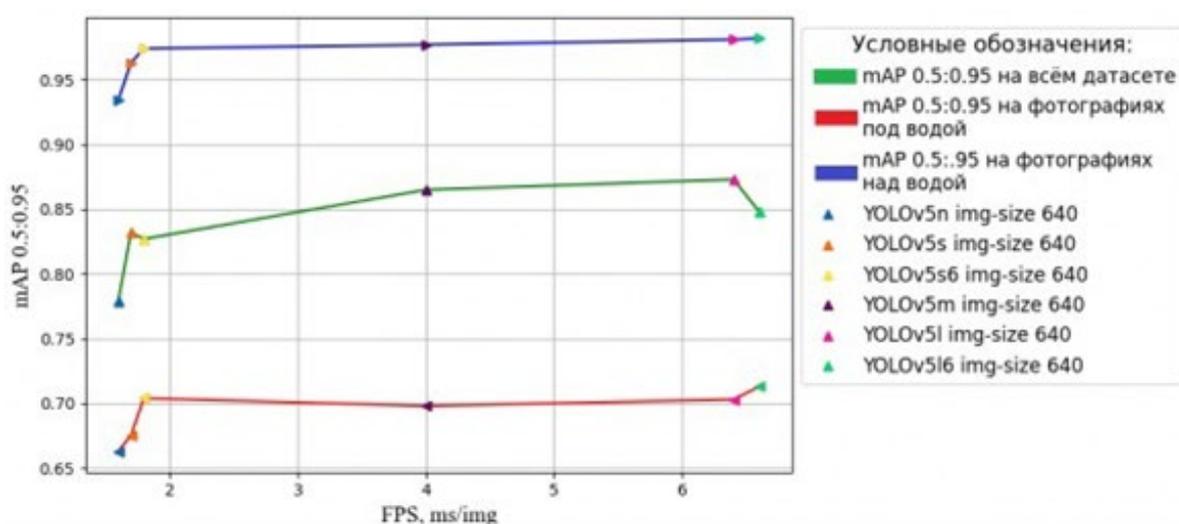


Рис. 2. Парето фронт оптимальных моделей YOLOv5 для детекции радужной форели

При детекции радужной форели одним из главных показателей является точность нейросети при детекции рыбы в воде. Надводные условия, как правило, приближены к «идеальным»: хорошее освещение, однотипный задний фон, рыбы перемещаются по одной траектории и т.д. Только под водой, в тяжелых условиях разной прозрачности воды, бликов и посторонних объектов, раскрывается модель, её понимание контура объектов.

Таблица 4

Сравнение изменений при переходе на большие модели YOLOv5, обученные на COCO датасете, с обученными на фотографиях радужной форели

Название модели	COCO датасет		Датасет радужной форели	
	mAP 05:095	FPS, ms/img	mAP 05:095	FPS, ms/img
YOLOv5n	27.6	1.9	0.779	1.6

YOLOv5s	37.1 (+34.42%)	2.1 (+10.53%)	0.832 (+ 6.80 %)	1.7 (+ 6.25%)
YOLOv5m	45.0 (+17.56%)	4.5 (+114.29%)	0.865 (+ 3.97 %)	4.0 (+135.29%)
YOLOv5l	48.7 (+08.22)	7.1 (+57.78%)	0.873 (+ 0.92 %)	6.4 (+60 %)
YOLOv5x	50.5 (+03.7%)	13.4 (+88.73%)	0.856 (- 1.95 %)	12.5 (+95.35%)

Сформулируем гипотезу - для детекции одного класса объектов (например, радужной форели) большие модели подходят хуже. Для её доказательства продолжим анализ метрик моделей YOLOv5. Сравним процентное соотношение изменения показателей mAP 0.5:0.95 и FPS при переходе к большим моделям YOLOv5 на датасете радужной форели и на COCO датасете [YOLOv5 Documentation]. Учтём, что COCO датасет включает в себя свыше 330 тысяч изображений и 80 классов объектов. Из сравнения показателей (Таблица 4) видно, что в задаче детекции радужной форели переход на большие модели YOLOv5 даёт прирост в mAP несколько раз меньше, чем на COCO. Показатели изменений FPS также отличаются.

Подводя итоги проделанных экспериментов, можно сделать вывод — одной из самых точных моделей при обнаружении радужной форели под водой является модель YOLOv5s6 с разрешением входного изображения 640x640 пикселей. Также данная модель обладает отличной скоростью детекции – 1.8 ms/img, что делает её наилучшим вариантом детекции радужной форели. Отметим, что все выводы оптимальных подмножеств моделей производились на экспериментальных данных, полученных в ходе обучения соответствующих моделей на видеокарте GeForce RTX 2060 с 6 Гб видеопамяти. На других видеоадаптерах или с большим числом классов объектов Парето-множества могут быть другими, но выводятся аналогично. Также показатели обучения моделей YOLOv5 зависят не только от архитектуры самих моделей, аппаратных возможностей системы, но и от самих данных для обучения. Каждая задача классификации объектов уникальна. И для раскрытия её в полной мере необходим тщательный поиск оптимальных решений. Из-за множества шумов в кадре и низкой прозрачности воды рассмотренный в статье вариант будет полезен другим разработчикам для принятия решений по сходным задачам детекции объектов.

СПИСОК ЛИТЕРАТУРЫ

- Летов А. М.* Динамика полета и управление. М.: Наука, 1969.
- Муханов К. А.* Алгоритмы многокритериальной оптимизации в задачах выбора параметров системы управления мехатронными модулями // Научный журнал «Студенческий форум». 2017. № 10. С. 12—15.
- Yu Yongcan.* Real-Time Underwater Maritime Object Detection in Side-Scan Sonar Images Based on Transformer-YOLOv5 // Remote Sensing. 2021. Т. 18, № 13. С. 1—28.
- Deng J. et al.* Imagenet : A large-scale hierarchical image database // 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009. С. 248—255.
- Kuznetsova A. et al.* The open images dataset v4 // International Journal of Computer Vision. 2020. Т. 128. №. 7. С. 1956—1981.
- Ulucan O., Karakaya D., Turkan M.* A Large-Scale Dataset for Fish Segmentation and Classification //2020 Innovations in Intelligent Systems and Applications Conference (ASYU). IEEE, 2020. С. 1—5.
- Pedersen M. et al.* Detection of marine animals in a new underwater dataset with varying visibility // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019. С. 18—26.
- Fish Dataset. URL: <https://public.roboflow.com/object-detection/fish> (дата обращения: 29.09.2022).
- YOLOv5 Documentation. URL: <https://docs.ultralytics.com/> (дата обращения: 29.09.2022).
- COCO Annotator. URL: <https://github.com/jsbroks/coco-annotator?ref=madewithvuejs.com> (дата обращения: 29.09.2022).

Maksim P. PAVLOV

bachelor's degree, Petrozavodsk State University
 (Petrozavodsk, Russian Federation)
maksim_pavlov_2003@list.ru

COMPARISON OF YOLOV5'S MODELS FOR THE RAINBOW TROUT DETECTION IN A VIDEO STREAM

Scientific adviser:

Alexey G. Marakhtanov
 Paper submitted on: 11/11/2022;
 Accepted on: 11/20/2022
 Published on: 12/25/2022.

Abstract. The diversity of models and training parameters makes it difficult to find an optimal solution for the object detection problem. In this article the models of YOLOv5 are considered. For each of them, the key learning metrics are measured and the Pareto front of optimal solutions is constructed using the rectangle method. Based on this, the best model in terms of accuracy and speed of rainbow trout detection is selected. The author concludes that the training of models depends not only on their architecture, the hardware capabilities of the system, but also on the number of detection classes.

Keywords: neural networks, YOLOv5, rainbow trout detection, optimal set, Pareto front

For citation: Pavlov M. P. Comparison of YOLOv5's models for the rainbow trout detection in a video stream. *StudArctic Forum*. 2022; 7(4): 35—40.

REFERENCES

- COCO Annotator. URL: <https://github.com/jsbroks/coco-annotator?ref=madewithvuejs.com> (accessed: 29.09.2022).
- Deng J. et al. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition, 2009, P. 248—255.
- Fish Dataset. URL: <https://public.roboflow.com/object-detection/fish> (accessed: 29 September 2022).
- Kuznetsova A. et al. The open images dataset v4. *International Journal of Computer Vision*, 2020, V.128, I.7, p. 1956-1981.
- Letov A. M. *Dinamika poleta i upravlenie [Flight dynamics and control]*. – Moscow, Nauka Publ., 1969. (in Russian)
- Muxanov K. A. Algoritmy` mnogokriterial`noj optimizacii v zadachax vy`bora parametrov sistemy` upravleniya mexatronny`mi modulyami [Multicriterion optimisation algorithms in mechatronic module control system parameter selection problems]. *Scientific journal "Student Forum"*, 2017, no.10, pp.12-15. (in Russian)
- Pedersen M. et al. Detection of marine animals in a new underwater dataset with varying visibility. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, p. 18-26.
- Ulucan O., Karakaya D., Turkan M. A Large-Scale Dataset for Fish Segmentation and Classification. *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2020, P. 1—5.
- YOLOv5 Documentation. Available at: <https://docs.ultralytics.com/> (accessed: 29 September 2022).
- Yu Yongcan, Zhao Jianhu, Gong Quanhua, Huang Chao, Zheng Gen, Ma Jinye Real-Time Underwater Maritime Object Detection in Side-Scan Sonar Images Based on Transformer-YOLOv5. *Remote Sensing*, 2021, V. 18, I. 13, P. 1—28.