

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

**СРАВНЕНИЕ РЕЗУЛЬТАТИВНОСТИ РАЗЛИЧНЫХ ВЕРСИЙ
МОДЕЛЕЙ YOLO В ЗАДАЧАХ ОБНАРУЖЕНИЯ ОБЪЕКТОВ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 451 группы
направления 09.03.04 — Программная инженерия
факультета КНиИТ
Ланкина Ильи Сергеевича

Научный руководитель

к. ф.-м. н., доцент, завкафедрой _____

И. А. Батраева

Заведующий кафедрой

к. ф.-м. н., доцент _____

С. В. Миронов

Саратов 2025

ВВЕДЕНИЕ

Актуальность темы. В биологии существует большое количество исследований, требующих отслеживания передвижения животных для выявления закономерностей в их поведении. Для решения такой задачи чаще всего используется метод непосредственного наблюдения человеком за интересующим животным. Особенно трудны для человека случаи, когда животные плотно расположены или когда они похожи, но относятся к разным видам. В таких случаях обычно используют специальные метки, например, краской. Даже так эта задача требует много времени, поэтому биологи хотят использовать нейросети для ускорения решения.

Одной из задач компьютерного зрения является задача обнаружения объектов, похожих по внешнему виду. Но эта задача обычно не является приоритетной, и чаще пытаются добиться высоких метрик на наборах данных с множеством хорошо различимых объектов. Так за последние два года вышло целых пять версий семейства нейросетей YOLO с указанием лишь их метрик на одном наборе данных, содержащем только хорошо различимые объекты. YOLO считается одним из лидеров в задачах компьютерного зрения, и, несмотря на то, что упомянутые пять версий все относятся к одному семейству, они имеют разную структуру и требования к обучению.

Коллеги из МГУ в рамках своих исследований попросили сравнить эффективности решения таких задач при помощи последних моделей YOLO на предоставленном датасете для выбора наиболее подходящей модели. В настоящей работе будут рассматриваться подходы к задачам классификации и обнаружения объектов, популярные модели, решающие эти задачи, последние версии YOLO и их сравнение друг с другом.

Цель бакалаврской работы — сравнить результативность YOLOv8–12 на задачах обнаружения объектов на подготовленном наборе данных.

Поставленная цель определила **следующие задачи:**

- изучить структуру и принцип работы CNN, одностадийных и двухстадийных детекторов;
- изучить основные различия между версиями YOLO;
- изучить влияние набора данных на эффективность обученных нейронных сетей;
- реализовать модели, выбранные для исследования;

- проанализировать работу и сравнить модели;
- подвести итоги исследования и сделать выводы на основе полученных результатов.

Методологические основы решения задачи обнаружения с помощью семейства моделей YOLO представлены в работах Редмона Д., Вана Ч. и Вана А..

Теоретическая значимость бакалаврской работы. Заключается в проведении сравнения моделей семейства имеющих значительно отличающиеся архитектуры, разработанные разными командами разработчиков независимо, но при этом относящихся к одному семейству и имеющих малое количество исследований подтверждающих повышение результативности от старых версий к новым. Это дает дополнительную информацию о качестве передовых архитектур относительно друг друга.

Практическая значимость бакалаврской работы. Заключается в изучении результативности передовых моделей для решения задачи обнаружения объектов на специфических данных, таких как плотно расположенные объекты, и обобщение объектов похожих на один или несколько точно идентифицируемых классов в отдельный класс.

Структура и объём работы. Бакалаврская работа состоит из введения, 2-ух разделов, заключения, списка использованных источников и 3-ех приложений. Общий объем работы — 52 страницы, из них 40 страниц — основное содержание, включая 24 рисунка и 2 таблицы, список использованных источников информации — 27 наименований.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Обзор существующих методов решения задачи обнаружения объектов» посвящен обзору эволюции подходов к решению поставленной задачи.

В начале раздела описывается общая постановка задачи обнаружения объектов для лучшего понимания того, что задача обнаружения тесно связана с задачей классификации объектов.

Во втором подразделе описываются популярные методы решения задачи классификации.

Прежде всего описывается постановка задачи классификации в случае компьютерного зрения. Затем описываются сверточные нейронные сети, как наиболее популярный и эффективный способ решения этой задачи на данный момент. Вводятся необходимые понятия и описываются основные составляющие таких нейронных сетей. Это включает в себя принцип действия сверточных слоев, слоев субдискретизации, полносвязной нейронной сети, а также методов оптимизации, сокращения количества параметров и повышения качества обучения моделей.

В этом же подразделе затрагивается вопрос влияния набора данных на качество итоговой модели. В частности описаны зависимость от выбора непосредственных классов, размера выборки и объема видеопамати, используемой на одну итерацию обучения модели.

В конце этого подраздела описан общий принцип работы и архитектура сверточных нейронных сетей, а также архитектуры одних из наиболее популярных моделей-классификаторов, VGG16 и VGG19. На их примере видно, что нет однозначного первенства между моделями-классификаторами, а также, что отличия между моделями обычно заключаются в порядке, количестве, параметрах и составе последовательных групп сверточных слоев и слоев субдискретизации. Это дает общее понимание того, что выбор оптимальной модели для конкретной задачи классификации может отличаться.

В третьем подразделе описываются известные методы решения задачи обнаружения.

В начале подраздела подробнее описывается связь задач классификации и обнаружения, где указано, что общая идея большинства существующих моделей-детекторов основана на том, чтобы предлагать нейросети-классификатору

части изображения и, основываясь на результате классификации, делать выводы о наличии в предложенной области объекта и ее оптимальном выборе.

Дальше описываются известные подходы к определению позиций объектов относящихся к заданному классу. Описаны одностадийные и двухстадийные детекторы, их различия и некоторые наиболее важные представители каждого типа.

Среди одностадийных детекторов рассматриваются метод скользящего окна, заключающийся в наиболее очевидном и простом переборе всех возможных позиций с их последующей классификацией, и методе DPM (англ. deformable part-based model), который в отличие от других рассмотренных моделей не использует сверточные нейросети и делает выбор в пользу более простых и быстрых моделей по упрощенному до гистограммы направленным градиентам изображению для достижения сравнительно высокой точности с возможностью обработки изображений в режиме реального времени.

Из двухстадийных детекторов были рассмотрены R-CNN, fast R-CNN, faster R-CNN. Показаны различия в принципе работы и архитектуре каждого из них. Выявлены основные преимущества и недостатки. Закономерно, что наиболее продвинутой и тяжеловесной faster R-CNN имеет преимущество во всем кроме сложности своей структуры. Ее идея заключается в использовании еще одной нейросети для улучшения качества предлагаемых областей поиска, основываясь на обучающей выборке.

Затем идет описание способов сравнения качества предсказаний моделей в целом и относительно друг друга. Приводятся основные метрики, такие как функции потерь, точность, полнота, средняя точность. Описаны различия функций потерь и их влияние на обучение модели. Вводятся необходимые понятия и приводятся графики зависимостей метрик для лучшего понимания методов вычисления и связи ключевых метрик mAP50 и mAP50–95 с качеством модели в целом и другими метриками.

В конце подраздела описывается семейство моделей YOLO. Сначала даются необходимые исторические справки и общие сведения о первой версии модели YOLO, выпущенной в 2015 году. Описаны ее преимущества по скорости и точности предсказаний в режиме реального времени по отношению к другим существовавшим методам. Подробно описан общий принцип работы моделей этого семейства с подробным пояснением, почему скорость обработки изоб-

ражений настолько выше чем у аналогов, а затем приведен список основных изменений между версиями, подробно останавливаясь на версиях 8–12. Приведена также архитектура первой YOLO и YOLOv12, взятые из оригинальных исследований этих моделей. Это дает наглядное представление о количестве архитектурных изменений по сравнению с первой версией.

Второй раздел «Сравнительный анализ обученных моделей и их применение на практике» посвящен сравнению качества предсказаний выбранных моделей как на валидационных данных, так и на данных относящимся к отдельным подзадачам.

В начале раздела описывается цель исследований и обоснование для сравнения моделей вместо однозначного выбора последней версии. Поднимается проблема недостаточной исследованности последних моделей YOLO в силу их быстрого выхода, различности архитектур и процедур предобработки данных и общей цели повышения качества на одном конкретном наборе данных, а также факт того, что разработчики моделей YOLO не ставили перед собой задачу исследования качества моделей на специфических наборах данных, работу на которых и сравнивает данная работа.

Во втором подразделе описывается подготовленный набор данных. Обосновано его качество взаимодействием со специалистами, знакомыми с предметной областью набора данных. Описано общее количество изображений подготовленных для обучения, выбранные классы с приведением иллюстраций представителей каждого.

Далее в этом подразделе приведен код, использованный для подготовки и предобработки изображений для повышения качества обучения. Большой акцент сделан на том, что в отличие от оригинальных исследований предобработка изображений одинакова для всех моделей. Описаны функции преобразования изображения и соответствующих меток объектов, включая поворот на угол, отражение по вертикали и горизонтали, изменение насыщенности, удаление небольших прямоугольных областей из изображения. Описан алгоритм случайной генерации данных преобразований. Данный алгоритм позволяет генерировать случайные искажения на наборе данных, при этом он зависит только от одного числа — количество итоговых изображений, полученных из одного. Этот алгоритм также обеспечивает сохранность оригинальных данных путем создания первого искажения как отсутствие искажений. Последующие копии

изображения будут случайно преобразованы для того, чтобы модель не могла легко найти закономерность преобразований, а искала признаки классов. Описанные преобразования изображений помогают моделям искать более общие признаки классов, что повышает качество предсказаний.

В третьем подразделе описываются взятые подверсии YOLOv8–12 и обоснование релевантности их сравнения. Модели были выбраны с учетом возможности их тренировать локально и объективности их сравнения. У YOLO с 8-ой версии обычно выпускается пять подверсий в зависимости от возможностей GPU: «nano», «small», «medium», «large», «extra». Эти подверсии обычно добавляются к названию модели как n, s, m, l, x соответственно. Выбраны были «medium» подверсии, так как они полностью соответствуют заявленной структуре и отличаются от более больших лишь количеством дополнительных обработок изображения и точностью меньшей относительно них менее чем на 10%.

В этом же разделе приведен общий вид кода обучения моделей на примере модели YOLOv8m.

Четвертый подраздел посвящен непосредственному сравнению обучения моделей по метрикам mAP50 и mAP50–95 (рис. 1 и 2). Обоснование сравнения только этих двух метрик заключается в больших различиях между архитектурами обученных моделей и показательности этих двух.

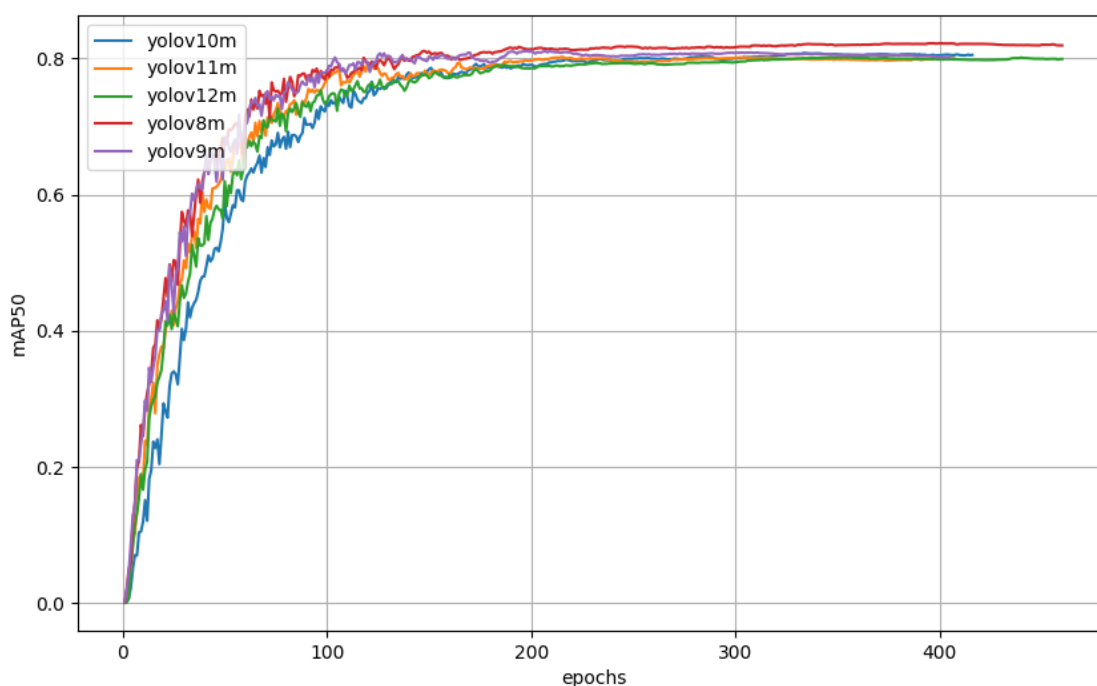


Рисунок 1 – График зависимости метрики mAP50 моделей от эпохи обучения

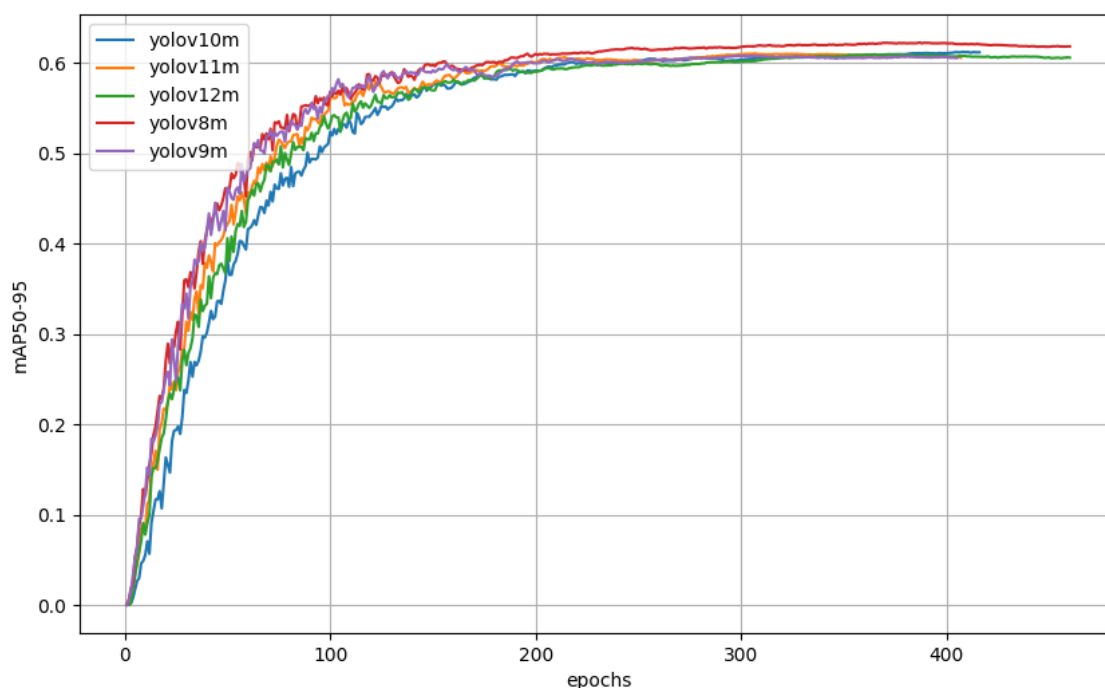


Рисунок 2 – График зависимости метрики mAP50–95 моделей от эпохи обучения

Также в этом подразделе происходит сравнению полученных результатов с оригинальными исследованиями (рис. 3 и 4). На графиках у большинства версий 5 точек, соответствующих подверсиям(n, s, m, l, x).

Показатели производительности

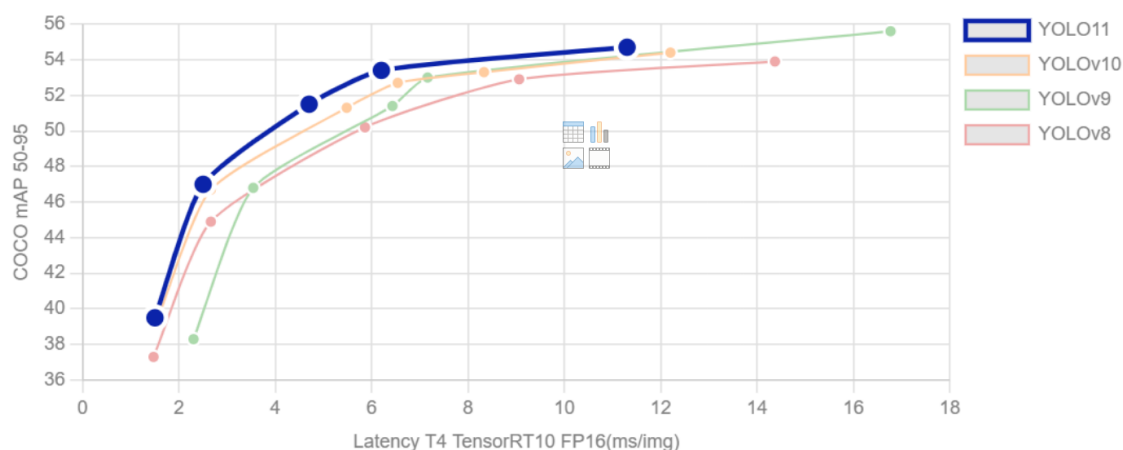


Рисунок 3 – График метрики mAP50–95 и производительности для подверсий моделей YOLO (с сайта Ultralytics)

Полученные результаты на данном наборе данных не соответствуют результатам оригинальных исследований, что подтверждает неоднозначность первенства среди данных моделей. В этом конкретном случае все модели, кроме 8-ой показали почти одинаковый результат, но 8-ая имела конечное преимуще-

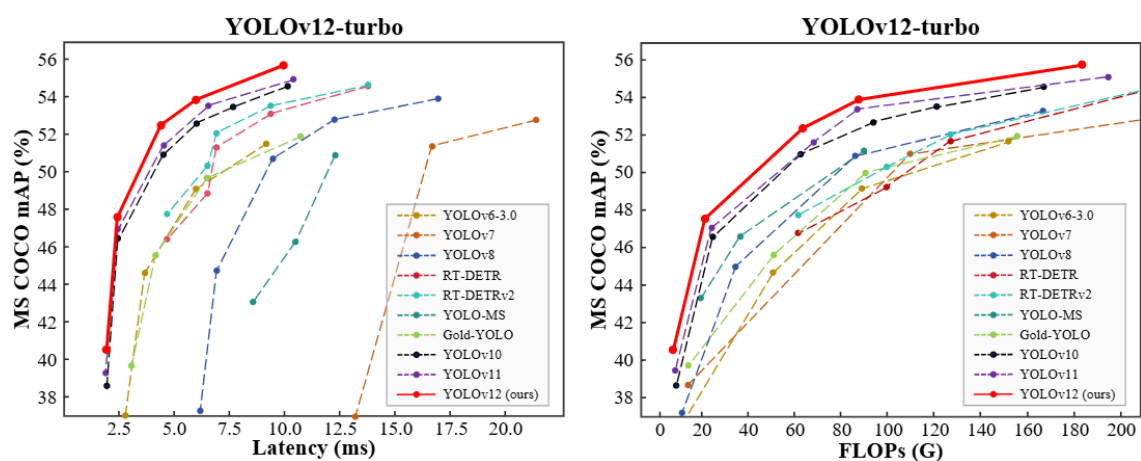


Рисунок 4 – График метрики mAP50–95 и производительности для подверсий моделей YOLO (из репозитория YOLOv12)

ство сравнимое с заявленными в оригинальных исследованиях между соседними версиями моделей.

Пятый подраздел описывает полученные результаты на двух целевых поднаборах данных, в одном из которых представители определяемых классов располагаются вплотную друг к другу и частично перекрываются (рис. 5), а в другом находятся представители абстрактного класса, который похож на определяемые, но не полностью (рис. 6).

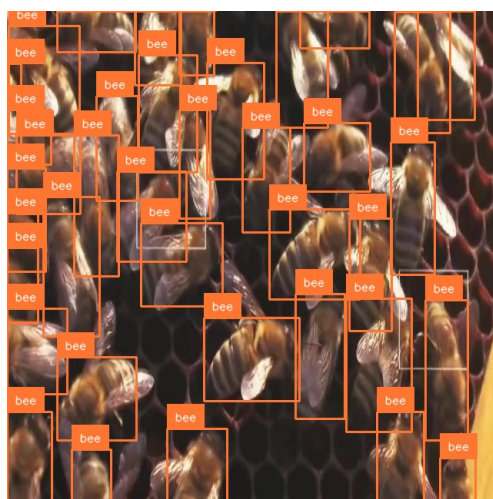
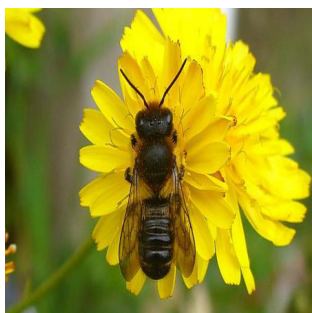


Рисунок 5 – Пример плотного расположения пчел и их разметки моделью YOLOv8m

Результаты были вынесены в таблицы 1 и 2.

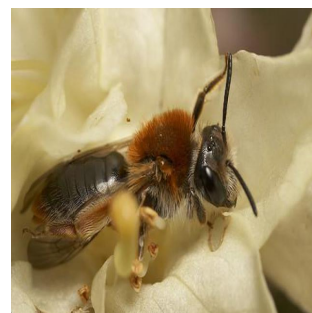
В последнем подразделе подытожив с ранее посчитанными во время обучения метриками сделан вывод, что 8-ая версия лучше других справилась почти со всеми заданиями. 11-ая лучше обобщила более абстрактный объект, а во всех остальных метриках была в топ-3 с небольшой разницей. 12-ая была в топ-



(a)



(b)



(c)

Рисунок 6 – Примеры пчелы неопределенного вида

Таблица 1 – Метрики моделей на подзадаче обобщения «других» пчел

Модель	mAP50	mAP50–95
YOLOv8m	0.839	0.689
YOLOv9m	0.827	0.677
YOLOv10m	0.840	0.689
YOLOv11m	0.871	0.718
YOLOv12m	0.840	0.699

Таблица 2 – Метрики моделей на подзадаче плотного расположения объектов

Модель	mAP50	mAP50–95
YOLOv8m	0.640	0.528
YOLOv9m	0.608	0.498
YOLOv10m	0.577	0.476
YOLOv11m	0.625	0.512
YOLOv12m	0.626	0.516

3 во всех. Остальные две модели были менее стабильны и показали хороший результат только в основной задаче обнаружения.

Таким образом для биологических задач, где представлены похожие классы, рекомендуется обучать 8-ую версию. В том числе в задачах с плотным месторасположением объектов. А в случае, где нужно обобщать часть классов до более абстрактного лучше использовать 11-ую, так как при лучших результатах в этой задаче, в остальных случаях она тоже проявила себя лишь немногим хуже 8-ой.

ЗАКЛЮЧЕНИЕ

Таким образом, несмотря на заявленное улучшение между последними моделями YOLO, в силу их различных структур и недостаточность исследований на разных наборах данных и подзадачах обнаружения объектов, следует, что нет однозначного первенства среди них. Для каждой задачи нужно исследовать и сравнивать эффективность разных версий и выбирать лучшую.

Дополнительно можно уточнить, что не смотря на то, что данная работа показала большую точность на 8-ой версии YOLO, чем на более новых, как для всего набора данных, так и для случая с плотным размещением объектов, это не значит, что это верно для других наборов данных. Вопрос выбора наилучшей модели для конкретной задачи на конкретных наборах данных необходимо исследовать отдельно. Данная работа лишь подтвердила возможность преимущества по точности более ранних версий перед новыми.

Основные источники информации:

- Redmon, J. You only look once: Unified, real-time object detection / J. Redmon, S. K. Divvala, R. B. Girshick, A. Farhadi // CoRR.— 2015.
- YOLOv8 in PyTorch [Электронный ресурс]. — URL: <https://github.com/autogyro/yolo-V8> (Дата обращения 28.05.2025). Загл. с экр. Яз. англ.
- Wang, C. Yolov9: Learning what you want to learn using programmable gradient information // Computer Vision- ECCV 2024- 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part XXXI / Ed. by A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, G. Varol. — Vol. 15089 of Lecture Notes in Computer Science. — Springer, 2024. — Pp. 1 — 21.
- Wang, A. Yolov10: Real-time end-to-end object detection // Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10–15, 2024 / Ed. by A. Globersons, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. M. Tomczak, C. Zhang.— 2024.
- Ultralytics YOLO11 [Электронный ресурс]. — URL: <https://github.com/ultralytics/ultralytics> (Дата обращения 28.05.2025). Загл. с экр. Яз. англ.
- YOLOv12: Attention-Centric Real-Time Object Detectors [Электронный ре-

- сурс]. — URL: <https://github.com/sunsmarterjie/yolov12> (Дата обращения 28.05.2025). Загл. с экр. Яз. англ
- Ultralytics YOLO11 [Электронный ресурс]. —
URL: <https://docs.ultralytics.com/ru/models/yolo11/>
(Дата обращения 28.05.2025). Загл. с экр. Яз. англ.
- COCO Common Objects in Context [Электронный ресурс]. —
URL: <https://cocodataset.org/>
(Дата обращения 28.05.2025). Загл. с экр. Яз. англ.
- Explore the Roboflow Universe [Электронный ресурс]. —
URL: <https://universe.roboflow.com/>
(Дата обращения 28.05.2025). Загл. с экр. Яз. англ.