УДК 004.93’12

НЕЙРОСЕТЕВЫЕ АРХИТЕКТУРЫ ДЕТЕКЦИИ ОБЪЕКТОВ НА ОСНОВЕ ОРИЕНТИРОВАННЫХ ОГРАНИЧИТЕЛЬНЫХ РАМОК

О. О. Колб, М. М. Лукашевич

Белорусский государственный университет,
Минск, Беларусь, kolbaa@bsu.by, LukashevichMM@bsu.by

В работе исследуются алгоритмы глубокого обучения для задачи обнаружения объектов на изображениях с акцентом на архитектуры YOLOv8, YOLOv11 и YOLOv12 с поддержкой ориентированных ограничительных рамок (OBB). Проведён сравнительный анализ их производительности, точности и способности к обобщению на медицинском датасете. Результаты представлены в виде мета-таблиц и визуализаций, а также проанализированы по метрикам. Проведено тестирование моделей на реальных данных.

Ключевые слова: компьютерное зрение; детекция; YOLO; метрики, медицинские данные; нейронные сети; обучение моделей.

NEURAL NETWORK ARCHITECTURES FOR OBJECT DETECTION BASED ON ORIENTED BOUNDING BOXES

A. A. Kolb, M. M. Lukashevich

Belarusian State University,
Minsk, Belarus, kolbaa@bsu.by, LukashevichMM@bsu.by

The paper investigates deep learning algorithms for object detection in images, with a focus on the YOLOv8, YOLOv11, and YOLOv12 architectures with support for oriented bounding boxes (OBBs). A comparative analysis of their performance, accuracy, and generalisation ability on a medical dataset is conducted. The results are presented in the form of meta-tables and visualisations, and analysed according to metrics. The models are tested on real data.

Keywords: computer vision; detection; YOLO; metrics; medical data; neural networks; model training.

1. Введение

Современное общество всё глубже уходит в цифровое пространство, где обработка визуальных данных становится неотъемлемой частью как научных исследований, так и практических решений в индустрии. Обнаружение объектов на изображениях является центральной задачей компьютерного зрения, лежащей в основе таких направлений, как автономные транспортные системы, медицинская диагностика, контроль качества производства, мониторинг инфраструктур и обеспечение безопасности.

Эволюция методов детекции прошла путь от медленных двухступенчатых алгоритмов к быстрым и эффективным одностадийным моделям, среди которых особое место занимает архитектура YOLO (You Only Look Once) [1]. Однако, классическая прямоугольная локализация, взятая за основу в YOLO, обладает ограничениями и не позволяет корректно описывать наклонные или вытянутые объекты. Это критично в задачах компьютерного зрения.

В ответ на эти вызовы развивается подход с использованием ориентированных ограничительных рамок (OBB), обеспечивающих более высокую точность локализации в сложных и насыщенных сценах. Новейшие модификации YOLO — версии v8, v11 и v12 — включают поддержку OBB, предоставляя возможности для построения универсальных и при этом высокоточных моделей.

Таким образом, исследование эффективности данных архитектур и проведение их сравнительного анализа на прикладных датасетах представляют собой актуальную научную и практическую задачу, на которую настоящая работа предлагает систематизированное и экспериментально подтверждённое решение.

2. Теоретические основы

Архитектура эксперимента опиралась на два ключевых программных компонента: библиотеку Ultralytics и инструменты из экосистемы scikit-learn [2]. Первая — это современный, гибкий фреймворк, разработанный вокруг моделей YOLO [3], включая последние версии с поддержкой ориентированных ограничительных рамок (OBB). Благодаря унифицированному API и встроенной поддержке аннотаций, визуализаций и экспорта моделей, Ultralytics стал ядром реализации.

Вторая составляющая — sklearn — использовалась для дополнительного анализа метрик, построения диаграмм, расчёта корреляций между показателями и генерации сравнительных таблиц. Визуализация ошибок, подсчёт precision/recall, F1 и построение confusion matrix были реализованы с её помощью.

Эксперименты проводились на локальной машине с поддержкой CUDA и параллельно в Colab и Kaggle, где всё окружение и зависимости автоматически инициализировались при помощи встроенных скриптов. Все результаты сохранялись в структурированных каталогах, обеспечивая удобную репликацию и анализ.

Для систематизации экспериментов и наглядного представления всех результатов использовалась платформа Weights & Biases (W&B). Через автоматическую интеграцию с библиотекой Ultralytics логировались основные метрики, графики обучения, визуализации предсказаний и сравнения

между архитектурами. Это позволило выявлять закономерности, отлавливать неустойчивое поведение моделей и быстро переключаться между сессиями даже при сбоях в обучении. W&B выступила в роли аналитической среды и системы восстановления при повторных запусках.

В качестве исследуемого датасета использовался набор данных «DentalXRAY: рентгенограмма зубов и заболеваний». DentalXRAY [4] — это медицинский датасет, включающий 644 изображения рентгеновских панорам челюсти детей. Разрешение варьируется от 1500x1000 до 2000x1000 пикселей. Каждое изображение сопровождается аннотациями OBB для описания различных стоматологических аномалий. Применение — обучение моделей поддержки диагностики в клинической практике.

Классы представляют собой диагнозы: от Decayed Tooth и Root Canal Treatment до Enamel Hypoplasia и Abnormal Eruption of Teeth. Баланс классов умеренно смещён: от нескольких экземпляров (например, Congenitally Missing Teeth) до сотен (Caries Restoration, Decayed Tooth). Типичный снимок — синий или тёмно-серый фон с белыми контурами зубов, минимальный шум, но высокое разнообразие случаев.

Для обучения модели детекции малых объектов были выбраны модели YOLOv8 (n-l), YOLOv11 (n-l), YOLOv12 (n-l) [5, 6, 7]. Их характеристики представлены в табл. 1–2.

Таблица 1

Метрики и скорость работы моделей YOLO8

Модель	Размер (пиксели)	mAPval 50-95	Скорость CPU ONNX (мс)	Скорость T4 TensorRT 10 (мс)	Параметры (M)	FLOPs (B)
YOLOv8n-obb	1024	78,0	204,77	3,57	3,1	23,3
YOLOv8s-obb	1024	79,5	424,88	4,07	11,4	76,3
YOLOv8m-obb	1024	80,5	763,48	7,61	26,4	208,6
YOLOv8l-obb	1024	80,7	1278,42	11,83	44,5	433,8
YOLOv8x-obb	1024	81,36	1759,10	13,23	69,5	676,7

Таблица 2

Метрики и скорость работы моделей YOLO11

Модель	Размер (пиксели)	mAPval 50-95	Скорость CPU ONNX (мс)	Скорость T4 TensorRT10 (мс)	Параметры (M)	FLOPs (B)
YOLO11n-obb	1024	78,4	117,6 ± 0,8	4,4 ± 0,0	2,7	17,2
YOLO11s-obb	1024	79,5	219,4 ± 4,0	5,1 ± 0,0	9,7	57,5
YOLO11m-obb	1024	80,9	562,8 ± 2,9	10,1 ± 0,4	20,9	183,5
YOLO11l-obb	1024	81,0	712,5 ± 5,0	13,5 ± 0,6	26,2	232,0
YOLO11x-obb	1024	81,3	1408,6 ± 7,7	28,6 ± 1,0	58,8	520,2

Дадим краткое описание этих таблиц. Названия моделей из семейства YOLO (n — nano, s — small, m — medium, l — large, x — extra large). Разрешение входного изображения (1024 пикселей). mAPval 50-95 – это средняя точность (mean Average Precision) в диапазоне IoU от 0,5 до 0,95 (график динамики этой метрики представлен также на рис. 1). 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 (M) – это количество параметров модели в миллионах. FLOPs (B) – это количество операций с плавающей точкой в миллиардах.

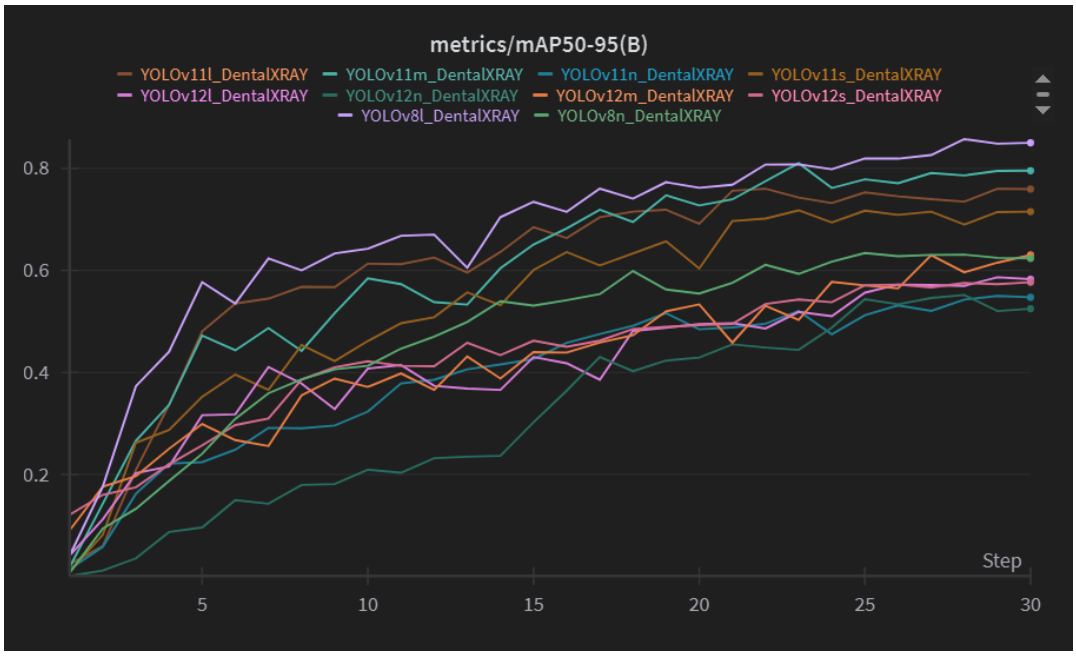

Рис. 1. График mAP50-95 для всех моделей на датасете DentalXRAY

3. Практические результаты

Модели семейства YOLOv8 продемонстрировали наивысшее качество на медицинском датасете DentalXRAY (рис. 2), опередив другие архитектуры по всем ключевым метрикам, перечисленным ниже.

- YOLOv8l достигла лучших показателей: mAP@0.5 = 0,9666, mAP@0.5:0.95 = 0,8503, precision ≈ 0,90, recall ≈ 0,90. Это делает модель особенно пригодной для задач точной медицинской локализации патологий.
- YOLOv8m предложила удачный баланс: mAP@0.5:0.95 = 0,8324 при отличной полноте (recall ≈ 0,88) и высоком precision. Такая модель предпочтительна для интеграции в системы с ограниченным ресурсом.

- YOLOv8s и YOLOv8n, несмотря на упрощённую архитектуру, достигли mAP@0.5 выше 0,78, что делает их применимыми в лёгких клинических системах, особенно при необходимости экономии вычислительных ресурсов.

Модели данной серии стабильно сходились при обучении и демонстрировали устойчивость к различиям в плотности, освещённости и форме объектов.

Рис. 2. Показатели YOLOv8 на датасете DentalXRAY

Архитектура YOLOv11 (рис. 3) показала конкурентоспособные результаты, особенно на вариантах m и l.

- YOLOv11m: mAP@0.5 = 0,9334, mAP@0.5:0.95 = 0,7956. Отличная полнота (recall ≈ 0,89) при высоком качестве локализации делает модель сопоставимой с YOLOv8m.
- YOLOv11l: чуть более низкий mAP@0.5:0.95 (0,7594), но стабильный recall ≈ 0,90, что важно для задач раннего скрининга.
- YOLOv11s — облегчённая модель, показала приемлемое качество (mAP@0.5:0.86), особенно в задачах с ограниченными ресурсами.
- YOLOv11n несколько отстаёт по всем метрикам, но остаётся применимым вариантом при строгих ограничениях.

YOLOv11 подтвердил свою устойчивость в условиях переменных сцен, сохранив высокую полноту даже на слабовыраженных патологиях. Визуальные ошибки локализации минимальны у моделей l и m.

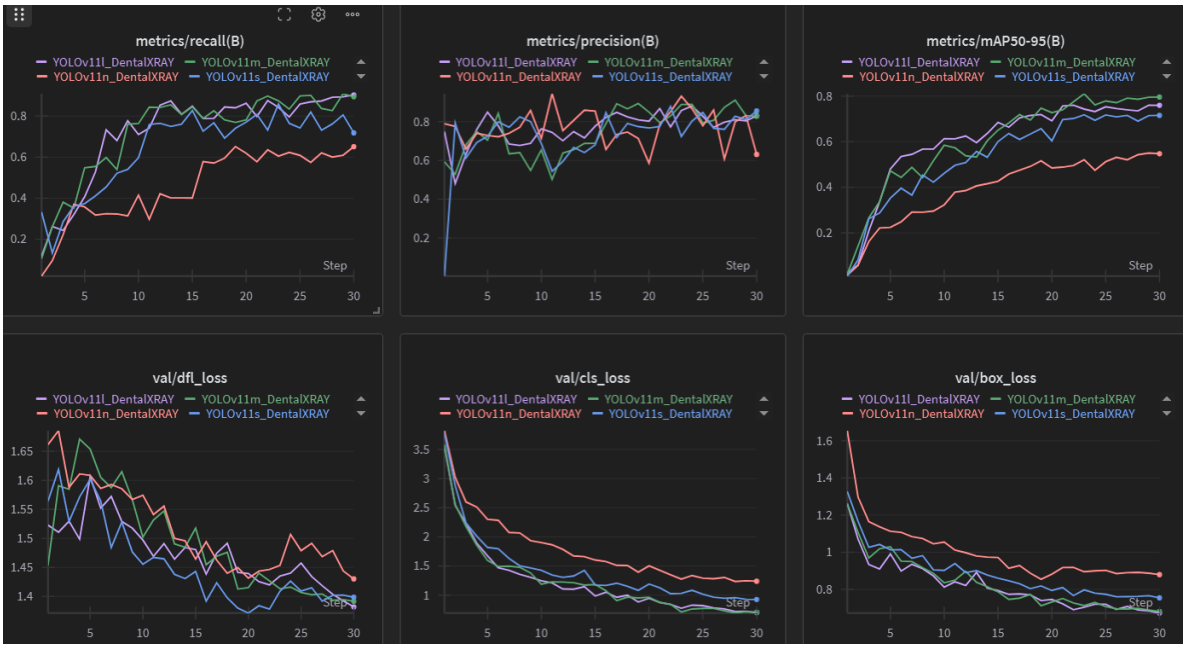

Рис. 3. Показатели YOLOv11 на датасете DentalXRAY

Модели семейства YOLOv12 демонстрируют слабые результаты на медицинском датасете DentalXRAY (рис. 4), что особенно заметно на фоне отсутствия предобученных весов.

- YOLOv12m и YOLOv12l показали наилучшие результаты: mAP@0.5:0.95 = 0,6302 и 0,5828 соответственно. Высокие значения precision (> 0,92) и recall (~0,87) говорят о хорошей способности к генерализации даже при обучении «с нуля».

Рис. 4. Показатели YOLOv12 на датасете DentalXRAY

- YOLOv12s и YOLOv12n уступают по метрикам mAP и точности, что делает их менее предпочтительными при высокой плотности классов. При этом recall остаётся достаточно высоким (~0,87), что делает их приемлемыми для задач, где важна полнота, например, при первичном медицинском скрининге.

Несмотря на отсутствие предобученных моделей, YOLOv12 быстро сходится в обучении и демонстрирует хорошую способность к генерализации даже на взятом наборе медицинских данных. Ошибки локализации чаще всего возникают на классах с наименьшим представлением.

4. Заключение

Медицинский датасет DentalXRAY, включающий разнообразные клинические случаи с рентгенограммами, показал себя как весьма чувствительный индикатор способности моделей к детекции на слабо контрастных и неструктурированных сценах. При сравнении всех трёх семейств моделей наблюдаются следующие ключевые тренды.

YOLOv8 — самая сбалансированная по всем метрикам серия: YOLOv8m: mAP@0.5:0.95 ≈ 0,83, mAP@0.5 = 0,88, Precision и Recall на уровне ~0,85. Даже компактные модели (n, s) обеспечивают приемлемую полноту и точность. Отлично справляется с большинством классов заболеваний, включая редкие — стабильность достигается благодаря зрелой архитектуре и наличию предобученных весов.

YOLOv11 — демонстрирует высокую полноту (recall) на всем диапазоне моделей: YOLOv11m: mAP@0.5:0.95 = 0,8045, recall доходит до 0,86. При этом точность (precision) немного ниже, что говорит о склонности к ложным срабатываниям — важный аспект для клинических применений, где избыточная чувствительность может быть оправдана. YOLOv11 — хороший компромисс между скоростью и качеством, особенно в условиях ограниченного числа аннотаций.

YOLOv12 — несмотря на отсутствие предобученных весов, показывает удивительно конкурентные результаты: YOLOv12l: mAP@0.5:0.95 = = 0,6, mAP@0.5 = 0,9, precision ≈ 0,84, recall ≈ 0,85. Архитектура демонстрирует быстрое схождение и хорошую адаптацию к медицинским изображениям даже «с нуля». Наименьшие версии (n, s) всё же заметно уступают по точности, хотя остаются рабочими для задач с акцентом на recall.

Сравнительный анализ показал, что универсальной модели для всех случаев не существует. Лёгкие конфигурации выигрывают в скорости и устойчивости к переобучению, тогда как более тяжёлые архитектуры приносят значимые преимущества лишь при высокой визуальной сложности и богатом контексте. Иными словами, эффективность определяется не масштабом модели сам по себе, а её соответствием конкретной задаче и характеру данных.

Результаты также показали, что использование ориентированных ограничительных рамок (OBB) заметно повышает точность локализации в насыщенных и сложных сценах. Несмотря на необходимость дополнительных усилий при подготовке данных и оценке, этот подход обеспечивает более адекватное описание объектов, особенно в условиях поворота, перекрытия и высокой плотности.

Ожидаемо, увеличение числа эпох положительно влияет на адаптацию моделей к сложным данным, так как уже на 30 эпохах данные показывают меньшую волатильность, что свидетельствует о более стабильном обучении. Также стоит учесть то, что несмотря на общие улучшения, метрика mAP50-95 остаётся относительно низкой (около 0,65–0,75 для большинства моделей), что указывает на трудности в классификации объектов малочисленных классов.

Следовательно, для повышения производительности целесообразно увеличить число эпох (например, до 50–100), протестировать различные стратегии аугментации данных и оптимизировать гиперпараметры, уделяя особое внимание классам, вызывающим наибольшие ошибки. Такой целенаправленный анализ позволит повысить полноту и точность детекции.

Таким образом, данная работа демонстрирует потенциал моделей архитектуры YOLO для использования в подобных задачах и обозначила главные условия их успешного применения — качество данных, корректность аннотаций и соответствие модели целям практического использования.

Библиографические ссылки

1. You Only Look Once: Unified, Real-Time Object Detection / J. Redmon [et al.] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. 2016. URL: https://ieeexplore.ieee.org/document/7780460 (date of access: 04.06.2025).

2. Scikit-learn: machine learning in Python / F. Pedregosa [et al.] // Journal of Machine Learning Research. 2011. Vol. 12. P. 2825–2830.

3. Ultralytics YOLO // GitHub. URL: https://github.com/ultralytics/ultralytics (date of access: 04.06.2025).

4. DentalXRAY Dataset // Kaggle. URL: https://www.kaggle.com/datasets/wanmugui/childrens-dental-panoramic-x-ray-dataset. (date of access: 04.06.2025).

5. Ultralytics YOLO Models. URL: https://docs.ultralytics.com/ru/models/ (date of access: 04.06.2025).

6. *Khanam R.* YOLOv11: An Overview of the Key Architectural Enhancements // arXiv preprint arXiv:2410.17725. 2024. URL: https://arxiv.org/abs/2410.17725 (date of access: 04.06.2025).

7. *Tian Y.* YOLOv12: Attention-Centric Real-Time Object Detectors // YOLOv12 Official Site. 2025. URL: https://yolov12.com (date of access: 04.06.2025).