

# Отчёт - VISION\_YOLO\_BLIP

---

Модель реализована в двух вариантах на питоне VISION\_YOLO\_BLIP.ipynb и на сервисе Roboflow, позволяющем размечать и запускать/разворачивать на предобученные YOLO (1-12) моделях тестовые картинки и видео. Разметка как в первом, так и во втором случае сделана на Roboflow как и аугментация. Код VISION\_YOLO\_BLIP.ipynb генерирует видео (output\_yolo\_blip.mp4) длиной 30 сек с размеченными распознанными объектами и выводом BLIP.

Проект распознавания предметов на столе. Запустить requirements.txt. Данные надо распаковать в файле Video.v9i.yolov11.zip в корневую директорию и запустить файл \*.ipynb.

Данные были размечены через систему Roboflow. train: ../train/images val: ../valid/images test: ../test/images

nc: 13 names: ['basket', 'dish with food', 'fork', 'glass empty', 'glass plate', 'glass with tea', 'knife', 'paper', 'redpaper', 'spoon', 'sugar bag', 'table', 'tea pot']

roboflow: workspace: aalex11 project: video-ytrrz version: 9 license: CC BY 4.0  
url: <https://universe.roboflow.com/aalex11/video-ytrrz/dataset/9>

Код интегрирует метод YOLO и большую модель изображений и видео BLIP, которая генерирует подсказки к распознанным объектам на видео к каждому датафрейму типа ("[Frame 200] YOLO: fork, glass plate, glass empty, dish with food, knife, table, tea pot, paper, redpaper, basket, glass with tea, spoon | BLIP: a person is sitting at a table with a plate of food")

Описание вывода:

Тренировочные метрики (обновляются каждый эпизод обучения): box\_loss — ошибка предсказания границ объектов (боксов). Чем меньше, тем лучше.

cls\_loss — ошибка классификации объектов. Тоже должно снижаться.

dfl\_loss — "Distribution Focal Loss", метрика точности по координатам бокса. Чем ниже, тем точнее.

GPU\_mem — использованная видеопамять.

Instances — сколько объектов обрабатывалось в эпоху.

Метрика	Значение	Интерпретация
Precision	0.788	Модель корректно определяет 78.8% объектов (из всех предсказанных).
Recall	0.806	Модель находит 80.6% всех реальных объектов.
mAP@0.5	0.851	Средняя точность при IoU=0.5 — 85.1% (хороший результат).
mAP@0.5:0.95	0.673	При строгих условиях (разные IoU) точность падает до 67.3% (проблемы с точностью локализации).

Метрика	Значение	Интерпретация
F1-score	0.797	Баланс между Precision и Recall — 79.7% (нормально, но можно улучшить).

## Проблемные классы и рекомендации:

### 1-й knife (нож) — критические проблемы

- Precision = 0, Recall = 0, mAP = 0

Проблема: Модель вообще не детектирует ножи.

Причины:

- Очень мало данных (всего 1 объект в выборке).
- Возможно, ножи слишком мелкие или перекрыты другими объектами.  
Рекомендации:
- Добавить больше изображений с ножами (разные ракурсы, размеры, фоны).
- Проверить разметку (нет ли ошибок в аннотациях).
- Использовать аугментации (масштабирование, повороты).
- 

### 2-й sugar bag (пакет сахара) — низкий Recall

- Recall = 0.408 (модель пропускает 59.2% сахарных пакетов).

Проблема: Модель плохо находит сахарные пакеты.

Причины:

- Возможно, пакеты сливаются с фоном (сливаются контуры).
- Недостаточно примеров (8 объектов — мало).  
Рекомендации:
- Добавить больше изображений с сахарными пакетами.
- Использовать контрастную аугментацию (изменение яркости, насыщенности).
- Попробовать увеличить разрешение входного изображения.
- 

### 3-й glass plate (стеклянная тарелка) — низкий Recall и mAP

- Recall = 0.5, mAP@0.5:0.95 = 0.59

Проблема:\*\* Модель плохо детектирует стеклянные тарелки.

Причины:

- Стекло прозрачное → сложно определить границы.

- Возможно, тарелки частично перекрыты другими объектами.  
Рекомендации:
- Добавить больше примеров с тарелками (разные углы, освещение).
- Использовать аугментации с изменением контраста.
- Попробовать другую модель (например, с лучшей детекцией прозрачных объектов).

4-й table (стол) — низкий Precision и Recall\*

- Precision = 0.384, Recall = 0.767

Проблема: Модель часто ошибается (много ложных срабатываний).

Причины:

- Столы занимают большую часть кадра → сложно определить границы.
- Возможно, в разметке неточные bounding boxes.  
Рекомендации:
- Улучшить разметку (более точные bbox).
- Использовать аугментации с обрезкой (чтобы модель училась детектировать части стола).

---

Хорошо работающие классы:

basket, dish with food, glass empty', paper, redpaper, tea pot`

- Precision и Recall близки к 1, mAP высокий → модель уверенно их детектирует.

---

Общие рекомендации по улучшению модели:

1-й Увеличить датасет (особенно для knife, sugar bag, glass plate).

2-й Проверить разметку (нет ли ошибок в аннотациях).

3-й Применить аугментации (изменение яркости, контраста, размытие).

4-й Настроить гиперпараметры (например, уменьшить confidence threshold, если много пропусков).

5-й Попробовать другую архитектуру (если проблема с прозрачными объектами).

Вывод:

Модель показывает хороший mAP@0.5 (0.851), но есть проблемы с Recall и детекцией сложных объектов (knife, sugar bag, glass plate).

Рекомендуется:

- Увеличить данные для слабых классов.
- Проверить разметку.
- Оптимизировать аугментации.
- Возможно, попробовать другую модель (например, YOLOv8 или EfficientDet).

На модели Roboflow YOLO 11 mAP@50 88.3% Precision 76.4% Recall 79.6%

Анализ метрик детекции (mAP@50, Precision, Recall) для Roboflow

1-й mAP@50 (mean Average Precision при IoU=0.5) — 88.3% Хороший результат

- Модель точно локализует объекты (при IoU  $\geq 0.5$ ).
- 88.3% — значит, что в среднем модель корректно находит и классифицирует объекты в 88.3% случаев.
- Сравнение с benchmarks:
  - 90% – отличный результат (например, YOLOv8 на COCO).
  - \*85-90% – очень хорошо.
  - <80% – требует доработки.

2-й Precision (Точность) — 76.4%

- Есть немного ложных срабатываний
- Из всех предсказанных объектов 76.4% действительно правильные.
- Остальные 23.6% – ложные обнаружения (FP, False Positives).
- Как улучшить?
  - Увеличить порог confidence (например, с 0.25  $\rightarrow$  0.4).
  - Добавить больше данных для классов с низким Precision.
  - Использовать аугментации (размытие, шум) для уменьшения переобучения.

3-й Recall (Полнота) — 79.6%

Модель пропускает ~20% объектов

- Находит 79.6% всех реальных объектов, но 20.4% пропускает (FN, False Negatives).

- Как улучшить?
- Уменьшить порог confidence (если модель слишком консервативна).
- Добавить примеры сложных случаев (маленькие, перекрытые объекты - что имеет место в данном случае на множестве казров).
- Попробовать другую модель (если текущая плохо находит мелкие объекты).
- 

---

#### Рекомендации для улучшения модели

Если цель — максимизировать точность (Precision):

Увеличить confidence threshold (например, до 0.4-0.5). (??)      Добавить больше данных для классов с высоким FP (ложные срабатывания).

Если цель — максимизировать полноту (Recall):      Уменьшить confidence threshold (например, до 0.1-0.2).

Улучшить детекцию мелких объектов (изменить anchor boxes, увеличить разрешение).

Общие улучшения:

Оптимизировать разметку (проверить аннотации на ошибки).

Использовать аугментации (размытие, изменение освещения).

Попробовать другую модель (например, YOLOv8, если текущая — YOLOv5).

---

Вывод \*mAP@50 = 88.3% → модель работает очень хорошо в целом.

Precision = 76.4% → есть ложные срабатывания (можно улучшить).

Recall = 79.6% → модель пропускает ~20% объектов (можно доработать).

Оптимальная стратегия:

- Если важнее точность (меньше ложных детекций) → повысить confidence threshold.
- Если важнее полнота (меньше пропусков) → понизить confidence threshold + улучшить детекцию мелких объектов.
- Если нужно и то, и другое → добавить данные + дообучить модель.