

## Выводы по тестовому заданию

В рамках предварительного тестирования были исследованы три подхода к детекции людей на видео с плотной толпой (результаты экспериментов доступны в *notebooks/experiments.ipynb*). На основе замеров скорости, точности и визуального анализа сформировано сравнение: (замеры делались на GPU T4x2 в Kaggle)

Таблица 1 - сравнение трех подходов решения задачи детекции

	Скорость (CPU)	Людей в кадре	Плюсы	Минусы
YOLOv8n	~0.02 с/кадр	16	Очень быстрый, лёгкий	Не видит мелких людей вдали
SAHI + YOLOv8m	~0.99 с/кадр	30	Находит мелких людей, лучше recall	Дублирует рамки из-за наложения тайлов
SAHI + RT-DETR	~2.44 с/кадр	37	Максимальный охват, меньше ложных пропусков	Очень медленный, дублирование остаётся

В зависимости от требований к быстродействию и точности можно выбрать подходящий вариант:

- 1) Выбор в пользу real-time - YOLOv8
- 2) Выбор в пользу offline анализа - SAHI + DETR

## Анализ качества

Какие проблемы решений остались:

### 1. Ложные срабатывания

Модель иногда интерпретирует статичные объекты (картины, вывески, фонари) как людей.

Повышение порога уверенности (`conf_threshold`) снижает количество ложных детекций, но одновременно ухудшает recall для мелких фигур в толпе.

### 2. Дублирование bounding box'ов

При использовании SAHI один и тот же человек может быть обнаружен в нескольких соседних тайлах, что приводит к наложению рамок. Это особенно заметно на границах тайлов.

### 3. Ограниченная чувствительность к мелким объектам

Даже в режиме SAHI часть силуэтов в глубине кадра не распознаётся

из-за крайне малого размера ( $< 15$  пикселей).

### **Возможные решения**

1. **Постобработка с адаптивным NMS**

Использовать Non-Maximum Suppression с пониженным порогом IoU (например, 0.3) для подавления дублей от SAHI.

2. **Фильтрация по геометрии**

Отсеивать bounding box'ы с площадью  $< 400$  пикселей ( $20 \times 20$ ) — это уберёт большинство ложных срабатываний на статичных объектах.

3. **Адаптивный порог уверенности**

Ввести зависимость `conf_threshold`: для мелких объектов — ниже порог, для крупных — выше.

4. **Предобработка видео**

Применить апскейлинг через Real-ESRGAN перед инференсом, чтобы улучшить разрешимость мелких лиц/фигур.

5. **Дообучение модели**

Дообучение на специализированном датасете (например собственных размеченных кадрах) позволило бы:

- снизить число ложных срабатываний,
- улучшить детекцию мелких объектов,
- адаптировать модель под специфику сцен (улица, метро, концерт).