

Обработка датасетов, полученных с БАС, с помощью нейронных сетей

Команда "Квадрицепс"

Наш подход

YOLOv11m + High-Res + Тайлинг

1

Выбор YOLOv11m

Лучший компромисс между **точностью, скоростью** и потреблением VRAM для сценариев с высоким разрешением и тайлингом.

2

Стабильность обучения

Быстрый экспериментальный цикл благодаря экосистеме Ultralytics, обеспечивающий **стабильное и надёжное обучение** модели.

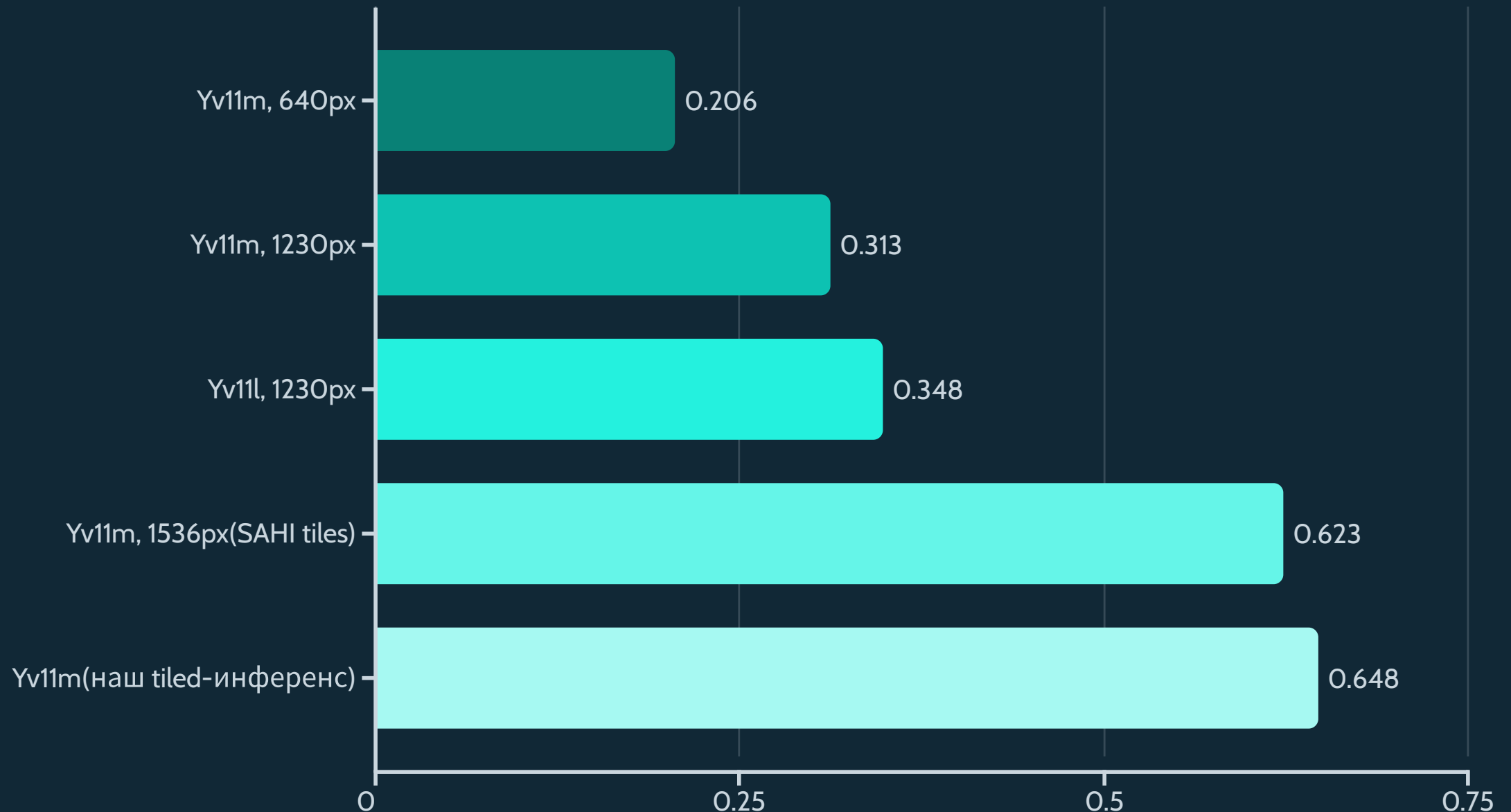
3

Масштабирование разрешения

Основной прирост метрики достигнут за счёт увеличения разрешения изображений и применения **плиточной обработки**, а не за счёт увеличения размера модели.

Мы сфокусировались на увеличении пиксельного размера объекта для детектора, чтобы он мог видеть больше деталей, что оказалось эффективнее простого "накачивания" модели.

Динамика роста метрики (абляция)



Ключевым фактором успеха стало использование tiled-инференса. Вместо использования самой "тяжёлой" сети, мы сосредоточились на том, чтобы "приблизить" человека к детектору, увеличивая детализацию на пиксельном уровне.

Альтернативные подходы и уроки

Мы исследовали несколько других архитектур, но они не показали требуемой производительности или были слишком сложны для быстрого внедрения в рамках хакатона.

Faster R-CNN ResNet50-FPNv2

Метрика: 0.4807

- Медленные итерации обучения.
- Требовалась глубокая настройка якорей и порогов для микрообъектов.
- Плохо масштабируется на больших кадрах без тайлинга.

TPH-YOLOv5++

Метрика: 0.5042

- Решение "из коробки" без глубокого тюнинга уступило.
- Дополнительные накладные расходы.
- Более сложный контур тренировки.

Мы сознательно "срезали углы": где требуется много тюнинга, мы не успели бы на хакатоне.

Оптимизация обучения на RTX 4090

Для достижения максимальной эффективности и скорости обучения на нашей рабочей станции с RTX 4090, мы применили ряд передовых техник:

1

AMP + TF32

Автоматическая смешанная точность и формат TF32 для ускорения вычислений и снижения потребления памяти.

2

Channels_Last

Использование формата channels_last для свёрток для повышения пропускной способности GPU.

3

Умеренные аугментации

Применялись только умеренные аугментации (Affine, цвет/контраст, лёгкий blur/noise), чтобы не "ломать" мелкие цели.

4

Стабильный батч

При `imgsz 1536` батч 8–12 (24 GB VRAM) работал стабильно, при необходимости использовался градиентный аккумулятор.

Это позволило нам добиться **стабильной скорости итераций** и провести множество экспериментов с различными конфигурациями и гиперпараметрами плиточной обработки в сжатые сроки.

Финальный инференс: Инженерное превосходство

Наше решение спроектировано для быстрой и точной обработки гигапиксельных кадров с очень мелкими объектами. Ниже приведены ключевые элементы, обеспечивающие его скорость и надёжность:

- **Warm-up** двух форм для устранения "холодного старта".
- **Coarse-окна** для резкого сокращения числа тайлов, обрабатываемых моделью.
- **Параллельная нарезка и декодирование** с использованием ThreadPoolExecutor.
- **Пакетный инференс тайлов** с MAX_TILES_PER_BATCH=4 для ровной загрузки GPU.
- **FP16 + channels_last** для снижения потребления памяти и ускорения свёрток.
- **Точное и дешёвое слияние** с torchvision NMS и отсечкой <4px.
- **Детальное профилирование** с CUDA-событиями для постоянной оптимизации.

Это не просто запуск предсказания, а **многоступенчатый конвейер**, оптимизированный для максимальной эффективности.

Итоговая архитектура пайплайна

1 Обучение модели

YOLOv11m (1 класс "person") в режиме высокого разрешения.

2 Пайплайн инференса

Coarse-обработка → формирование окон →
пакетная обработка тайлов → декодирование →
NMS → вывод результатов.

3 Настройка гиперпараметров

Отдельный grid-search для оптимизации гиперпараметров инференса под метрику организаторов.

4 Единый API

Использование единой функции `predict(images)` для удобства сабмита и продакшена.

5 Финальные гиперпараметры инференса

`TILE_SIZE=1536`, `OVERLAP=0.20`, `COARSE_SHORT_SIDE=1280`, `COARSE_CONF=0.10`, `COARSE_EXPAND=1.8`,
`MAX_TILES=64`, `NMS_IOU=0.55`, `MIN_WH_PIX=4`

Вся система спроектирована для **лёгкого развёртывания** и прозрачной работы, обеспечивая высокую точность и скорость.

Планы на будущее

Если нам будет предоставлено дополнительное время, мы планируем дальнейшее совершенствование решения:

- **Умный Coarse:** Интеграция heatmap/segment-подсказок для минимизации пустых тайлов.
- **Лёгкая ТТА:** (Test Time Augmentation: flip/rotate) + NMS/WBF для повышения точности инференса без переобучения.
- **Pseudo-labeling:** Применение semi-supervised обучения на неразмеченных данных с полётов.
- **Видео-трекинг:** Интеграция ByteTrack для уменьшения ложных срабатываний и пропусков во времени на видеопотоке.
- **Экспорт модели:** Оптимизация для ONNX/TensorRT для развёртывания на edge-GPU и серверах.

Ценность для поисково-спасательных операций



Прирост метрики

Значительное улучшение метрики mAP с 0.2064 до **0.6479** относительно базовой YOLOv11m.



Прозрачный пайплайн

Легко масштабируется для обработки потоков данных с дронов в реальном времени.



Быстрая итерация

Детальные профили и возможность быстрой настройки гиперпараметров обеспечивают **предсказуемую эксплуатацию**.

Наш пайплайн **готов к продакшену**, предоставляя высокоэффективное решение для поиска пропавших людей.

Контакты команды Квадрицепс

Благодарим за внимание! Будем рады ответить на ваши вопросы.



Клещенок Максим Андреевич

@S37483920441



Сахабутдинов Рустам Ринадович

@Rustam_Sahabutdinov