



Чемпионат - Омская область

Детекция человеческих силуэтов на снимках  
лесного массива, полученных с помощью  
БПЛА

Соколов Д.И.



# Задача

В рамках чемпионата была поставлена задача детекции координат объектов (людей) на фотографиях. Это “классическая” задача детекции объектов на изображениях. В данном случае она осложнена тем, что в тренировочной выборке изображений с объектами крайне мало (меньше 100 шт.)

# Используемое ПО и библиотеки

- Python
- Pandas
- YOLOv5
- mmdetection



YOLOv5 была выбрана в качестве “бейслайна” так обладает высокой скоростью обучения и инференса, а также дает хорошее качество детекции.

# Данные

Пример изображения из  
тренировочной выборки (для  
наглядности, вырезана часть и  
круги заменены на квадраты)





# Данные

Данных было очень мало (~ 100 изображений с объектами), решено было отделить 20 % на валидацию и добавить часть изображений без объектов к тренировочной и валидационной выборкам.

В дальнейшем изображения из валидации были добавлены в тренировочную выборку, т.к. локальная валидация не коррелировалась с лидербордом.

Дальше были написан код для конвертации кругом в прямоугольники, визуализации результатов, конвертации датафреймов в txt метки и обратно.

Основной проблемой было определение порога для детекции объектов (confidence threshold) для получения наилучшей метрикой, т.к. она не отражала результаты увиденные глазами, при просмотре изображений с результатами детекции.



# Решение

После нескольких экспериментов были использованы следующие параметры для обучения:

```
python -m torch.distributed.run --nproc_per_node 2 train.py --img 1920 --batch 6 --epochs 90 --data  
../data/dataset.yaml --weights yolov5m.pt --hyp hyp.scratch-med.yaml --device 0,1 --cache disk
```

и инференса:

```
python detect.py --augment --img 1920 --conf 0.25 --iou 0.45 --source  
'../input/detection-of-human-silhouettes-in-forest-images/test_dataset_test/test' --weights  
runs/train/$exp_last/weights/best.pt --save-txt --save-conf --half
```

Лучший результат на public leaderboard: 0.497166 - модель yolov5m обученная 90 эпох на изображениях 1920\*1080, с порогом 0.44



# Решение

Следует отметить, что yolov5m обученная 45 эпох на изображениях 1280\*720 дает результат лишь немного хуже на 0.01-0.25.

В дальнейшем я использовал библиотеку mmdetection (модели Faster R-CNN и Cascade R-CNN, однако мне не удалось улучшить метрику на лидерборде. ~ 0.45 для обеих моделей обученных 60 эпох.

В конце я попробовал библиотеку WBF  
<https://github.com/ZFTurbo/Weighted-Boxes-Fusion>  
подобрав iou\_thr получил + 0.1 на public leaderboard.



# Модели

Ссылки на веса и блокноты обучения и инференса моделей:

<https://github.com/dimka11/human-silhouettes-in-forest/blob/main/README.md>






# Планы на будущее

Главной проблемой было очень малое количество изображений с людьми. Для получения хороших результатов таких изображений должно быть на порядок больше.

Возможно использование двух-стадийных детекторов (Cascade R-CNN) позволит достичь лучших результатов.

Также не очень понятна метрика в данном соревновании она слабо коррелировала с результатами детекции увиденными глазами.



Спасибо за внимание!

Telegram: @dimkoss11