THE END

Aerial-Maritime Dataset

Link:
https://drive.google.com/drive/folders/1yNvUuvPTxuB0umoMwqAAIJcAI3oLpWTx?usp=sharing

Data Distribution :

Train dataset : 32 Images
Validation dataset : 370 Images

Classes :

1. movable-objects
2. boat
3. car
4. dock
5. jetski
6. lift

Annotation Format: MS-COCO

Model Implementations :

The first model opted for ResNest as it gave nice accuracy on the previous dataset. Also, it is a model having 0.46+ bbox accuracy on the coco-mini Val and coco test-dev.

Variant : cascade_rcnn_s101_fpn_syncbn-backbone+head_mstrain-range_1x

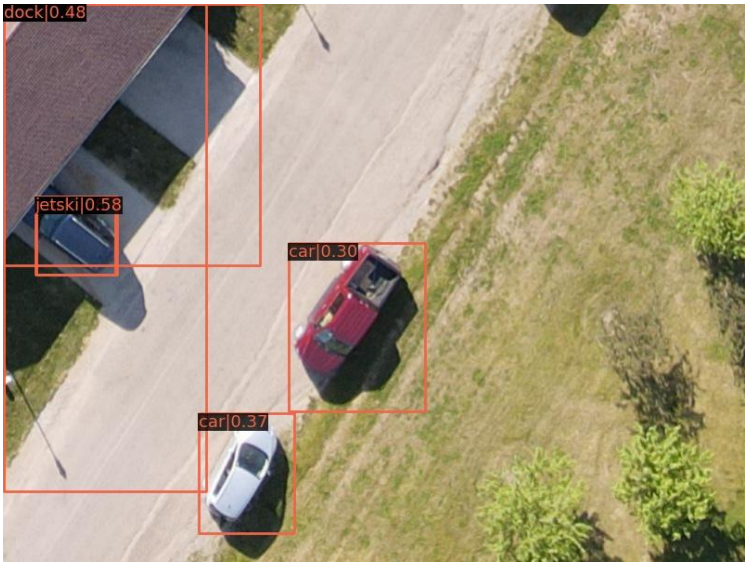
Scheduler: x1 (12 epochs)

Results :

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.229
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.466
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.197
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.100
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.258
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.390
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.390
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.390
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.211
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.428
```

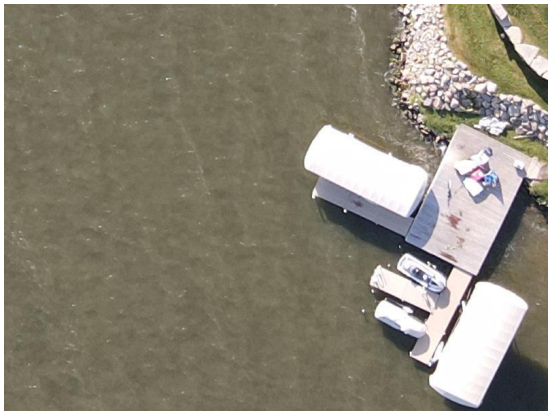
We can see the results are very poor. Looking at the predictions, one will have to try harder to find accurate and nice predictions.

Somehow some of the cars have been predicted nicely. But other objects have been left untouched.



Also, some cars have been predicted by different classes.

Some of the images haven't been inferred and left raw.



Also, the jet skis have been annotated as docks, But some of the lifts have been annotated well and that is the only positive thing in this prediction.

In the next step GFL model has been opted for training.

Variant : gfl_x101_32x4d_fpn_dconv_c4-c5_mstrain_2x

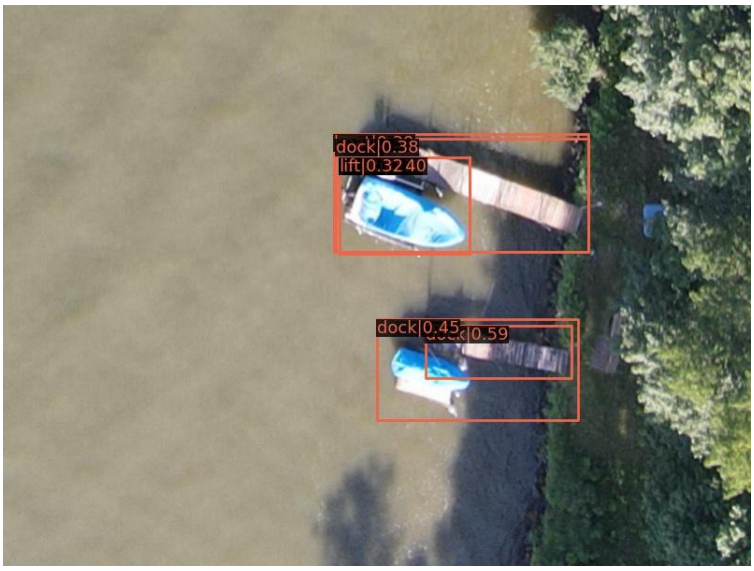
Scheduler: x2 (24 epochs)

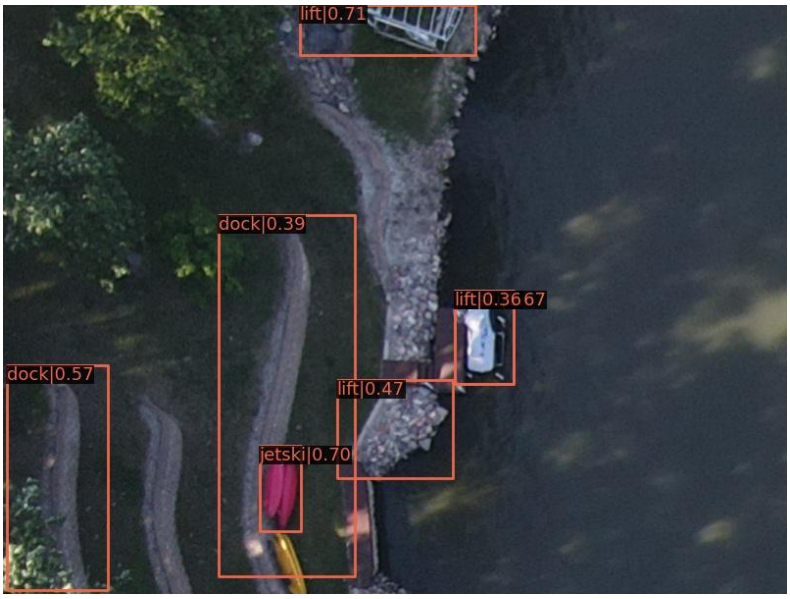
Results :

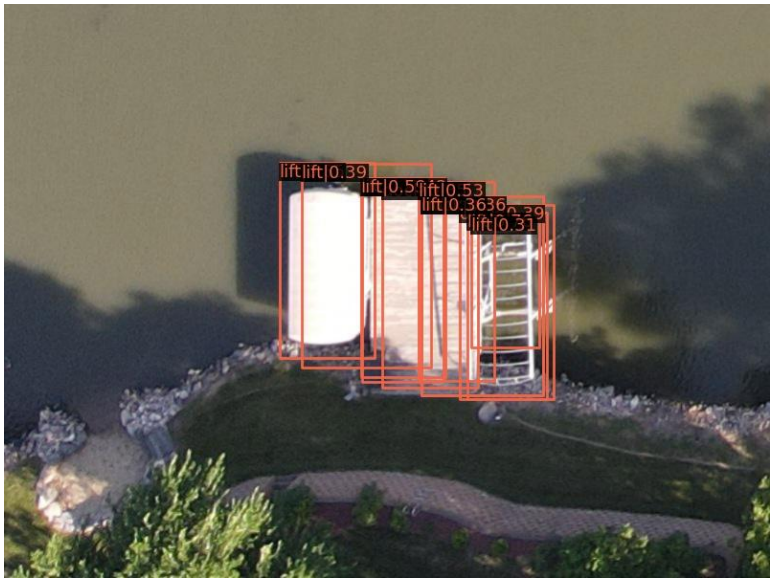
```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.240
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.466
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.212
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.099
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.278
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.484
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.484
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.484
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.330
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.537
```

The results have been updated a bit, the new model came with various errors.

Most of the jetskis have been treated as lifts. Also, the ground has been treated as a dock and there are multiple overlaps on some objects, appropriately on every 1 object per 5.







Failing on ResNest the next opted model was DETR as it gave the best performance in the previous case.

Variant : deformable_detr_twostage_refine_r50_16x2_50e

Scheduler: 50 Epochs

Results :

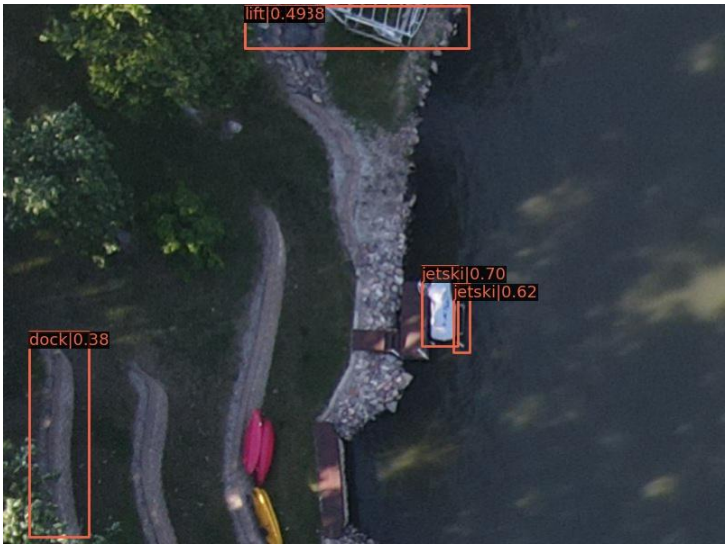
```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.243
Average Precision  (AP) @[ IoU=0.50       | area=   all | maxDets=1000 ] = 0.469
Average Precision  (AP) @[ IoU=0.75       | area=   all | maxDets=1000 ] = 0.213
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.001
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.126
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.258
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.474
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.474
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.474
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.055
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.333
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.502
```

The results have improved a bit, but the accuracy is still very low, where most of the prediction is bad, having multiple overlaps and the classes in some cases (which is rectified a bit) are wrong.

But somehow the cars are predicted well, also the jet skis are predicted well in most cases.

The prediction of lifts and docks also have been optimized a bit.

But some of the rocky ground parts have been predicted as docks which is also a drawback.



In the next step VFNet has been taken to check if the model is doing any anomaly or the data is itself a bad one for training.

Variant : vfnet_x101_64x4d_fpn_mdconv_c3-c5_mstrain_2x

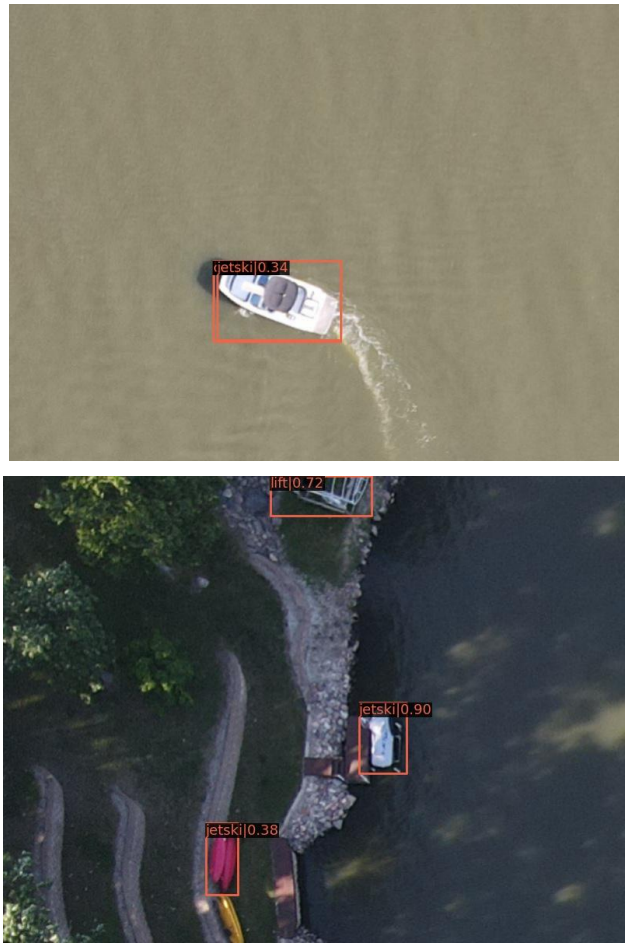
Scheduler: 50 Epochs

Results :

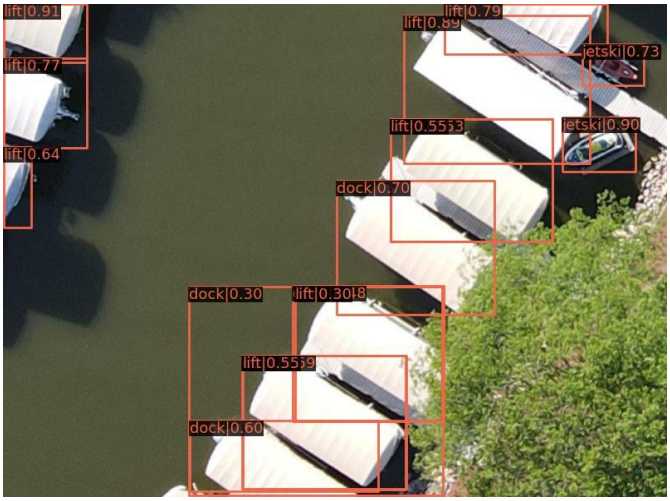
```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.265
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.515
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.244
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.141
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.294
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.477
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.477
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.477
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.002
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.303
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.552
```

The results have improved a bit, but the results are still very poor. But some of the errors have been rectified a bit.

The jet ski predictions are having quite nice, but some overlaps are present.



On the other hand the lift prediction is better than the previous model's performance. Though there are some overlaps of bboxes, they are less than the previous case.



But the most huge drawback is the prediction of the dock, where most of the rocky side near the dock has been predicted as docks, so overall the prediction is improved, but it's far away from the accurate results.

Conclusion: as the dataset has been trained on a small scale of data, additionally on the data which are far more deviated than in the test set, that is the reason to have such bad prediction over the SOTA models. Not only one, but 3 models failed similarly on this dataset, indicating that data split or data distribution or class distribution in the data haven't performed very well, so the few-shot method failed readily.

Performance Comparison Chart

Bounding Boxes (mAP)

Model / Dataset	Sparse-RCNN	VFNet	ResNest	DETR	GFL
Chess Dataset	0.179	0.658	0.676	0.681	0.693
Aerial Maritime Dataset	_	0.265	0.229	0.243	0.240