Drone Gesture Control Dataset

Dataset :

Flight Gesture Dataset

Link:
https://drive.google.com/drive/folders/1-8fjsykf-WSSuKLO92T6NXem2_CzCJll?usp=sharing

Data Distribution :

Train dataset : 100 Images
Validation dataset : 220 Images

Classes :

1. Flight
2. Hand
3. Hand-Land
4. Land
5. Takeoff
6. TakeoffHand

Annotation Format: MS-COCO

Model Implementations :

Primarily the GCNet model has been taken to perform on the dataset as it checks through every single detail and then predicts the masks precisely. The sparsity of the model may accurately cover the torso and as we have previously seen on the other datasets, it has always proven itself as a good prediction model.

Scheduler : x1 (12 epochs)

Results :

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.709
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.804
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.786
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.730
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.658
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.834
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.834
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.834
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.788
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.913

Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.656
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.811
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.786
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.658
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.721
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.768
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.768
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.768
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.723
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.885
```

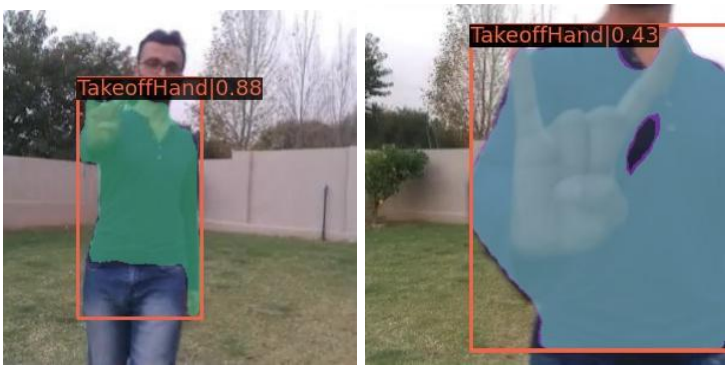
Looking at the prediction it can be told that most of the segmentation masks are moreover accurate, only a few mistakes are there. All the flight class and land masks have been predicted very precisely.

Most of the masks of the hand class are predicted perfectly. The mask near the fingers is somehow tweaking a bit, but mostly it is perfect. In some cases, the masks are worse and in some cases, the class has been predicted wrong, due to the likeliness of the finger gestures.

Double wrapping of masks is troublesome for this job.

Leaving 1-2 images, all the images of takeoff class have been predicted successfully. Those images with anomalies only happen to cover the head portion of the man and should be treated as a bad prediction.

Also the takeoff class has been well predicted, but in some cases, there are mask overlaps. Give attention to the images below:



Also, the most abrupt prediction is that every single image of the “hand-land” class has been predicted as either flighttakeoffhand or hand, which is an indication of bad predictive modeling.

Updating to **MRCNN** , gave slightly better results as it holds a higher place in the coco-test-dev leaderboard.

Scheduler: x1(12 epochs)

Results :

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.710
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.845
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.823
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.744
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.617
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.827
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.827
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.827
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.819
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.844
```

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.693
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.851
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.832
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.706
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.691
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.813
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.813
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.813
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.806
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.875
```

The results are upgraded for the box but better for the mask ones with Mask-RCNN.
All the masks are giving better predictions than they gave in GCNet.

Still, the boxes are overlapping in most of the cases, which is a new anomaly.

The classes with similarities are also getting affected can be seen in the pictures above. But these are in very few numbers which is a betterment of the previous model. Although the land class has been predicted correctly in all cases which were also done by GCNet.

Further up-gradation was with QueryInst which was taken as the classes could be predicted correctly using this algo-structure.

Scheduler: x3(36 epochs)

Results :

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.738
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.834
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.815
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.750
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.691
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.915
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.915
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.915
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.901
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.945
```

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.690
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.841
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.822
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.690
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.741
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.845
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.845
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.845
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.839
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.898
```

QueryInst not only upgraded the results far better but also upgraded the class predicted well than in previous cases.

Every single image of class except hand and hand land has been predicted accordingly.

In previous cases where hand-land classes were never predicted at all or in very low numbers, but in this scenario, it is present in broad cases and that is an up-gradation of this process. Also, the mask near the finger has been updated nicely, so that's also a rectification of previous models.

The below pictures are of classes: hand, hand, hand-land.

At last, we adopted DCN as it was the best model tested through previous cases.

Still, in this case, most of the data are symmetrical and share no complex spatial transformation, this might be a failure but the mask predictions of this model range very high, so can predict best.

Scheduler: x1(12 epochs)

Results :

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.751
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.862
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.831
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.774
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.676
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.877
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.877
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.877
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.862
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.915
```

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.712
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.869
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.848
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.706
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.731
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.825
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.825
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.825
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.814
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.888
```

It can be seen that all the DCN has surpassed all the results. All the images of class except hand land and “take offhand” are predicted correctly. Improvisation of the hand can be seen in this upgrade. And the percentage of handling classes have increased which is also a good sign of this test.

Though the classes are often predicted similar for “hand-land” and “flighttakeoffhand” in these previous models, in this case, have they have been lowered in a very big number. Previously the number was more than ½ of images, now it’s greater than 80 percent and which is a good prediction.

Conclusion :

The most challenging task in this dataset was to correctly identify the classes and modeling through the SOTA models let us reach a very good prediction. Primarily GCNet is an offset of the DCN model, still, the prediction went bad for it, telling that overfitting is present in that case. So, a simpler model can reach better than the model in higher ranking the scoreboard section is found working with this dataset. As we go from below the mask were also rectified simultaneously, so basically improvisation is completely present here.

THE END

Nature Dataset