Dataset :

Nature Dataset

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

Data Distribution :

Train dataset : 200 Images
Validation dataset : 600 Images

Classes :

1. Butterfly
2. Squirrel

Annotation Format: MS-COCO

Model Implementations :

In the previous modeling on other datasets, the best result was always DCN in a progressive way. So opting for DCN for training and testing. In the dataset, we have checked that most of the images are containing a single object and that also has been captured in many angles. Do deformability nature will help to find the best predictions.

Scheduler : x1 (12 epochs)

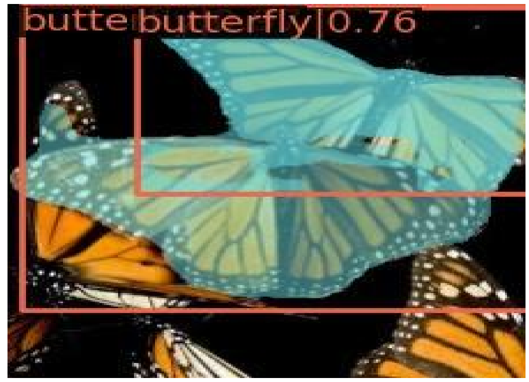
Results :

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.683
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.908
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.754
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 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.696
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.743
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.743
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.743
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 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.743
```

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.665
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.924
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.751
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 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.670
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.738
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.738
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.738
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 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.738
```

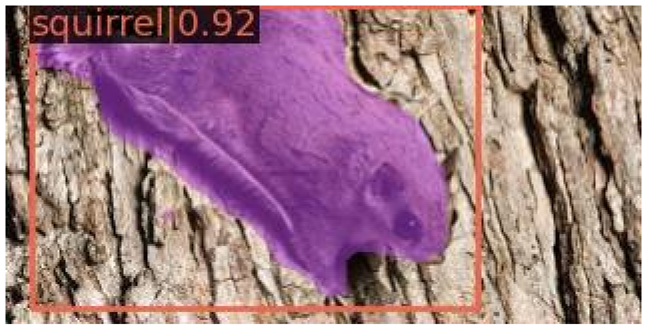
The results are quite decent, in most of the pictures the butterfly has been predicted very nicely. The wings have been covered very precisely in some cases.

But in some cases the body has been left untouched. Also when there are multiple objects then the model is also failing which is counted as a bad prediction.



In this single case, the 2 butterflies are also predicted as a squirrel which is a very bad outcome. And some images have extra masks in the background.

In the case of most of the squirrels, the prediction is very accurate. But the nails of them are slightly unmasked.



After getting decent results in DCN, now the next model implementation is with GCNet. Previously GCNet and DCN have been competing on a smaller scale, the reason behind this can be the DCN backbone inside the GCNet model.

Scheduler: x1 (12 epochs)

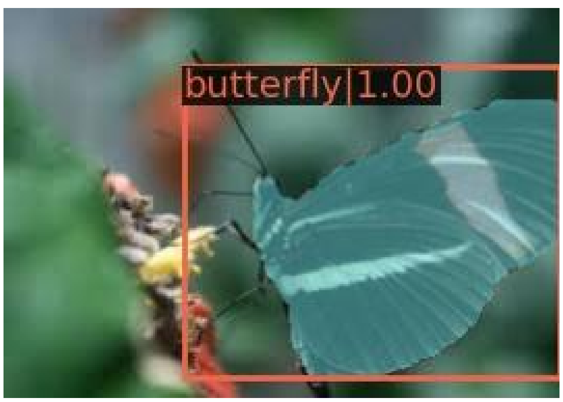
Results :

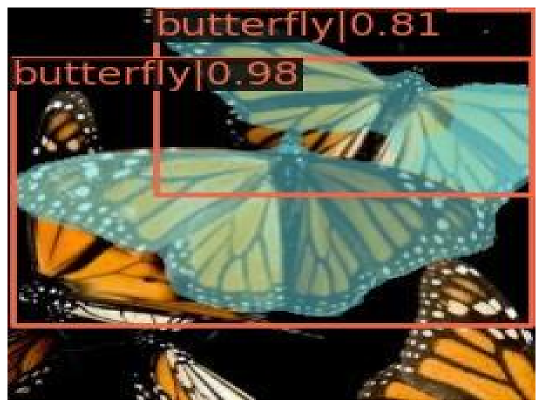
```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.764
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.936
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.858
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.770
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.815
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.815
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.815
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 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.815
```

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.739
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.932
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.825
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 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.751
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.786
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.786
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.786
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 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.786
```

As we can see the competing model has surpassed his competitor by a 10 percent increment in both bbox and segm mAP, thus we can say the predictions have improved quite a bit.

Primarily the butterfly predictions are most cases good as it was for DCN, but in some cases when the prediction went badly, the GCNet model improvised quite a bit. Though it couldn't surpass the multi-object detection anomaly.





The images where there were unwanted masks outside the object also have been decreased quite a bit, both for the butterfly and also for the squirrels.

The sensitive portions of the squirrels like nails and ears also have been masked quite well.

Still some of the prediction has not adapted the actual shape of the squirrel spreading the mask outside of the objects.

P.S. The GCNet model has rectified most of the eros created by the DCN model, still, it fails in some scenarios, so auto-labeling is still should not be practiced in this case, that is concrete.

To check if the other implementations upgrade or not, the next model chosen for this dataset implementation is QueryInst. As most images are having singular objects and the most important thing is that the objects are quite different in shape and color from the environment, so this model has high hope to gain increased performance.

Scheduler : x1 (12 epochs)

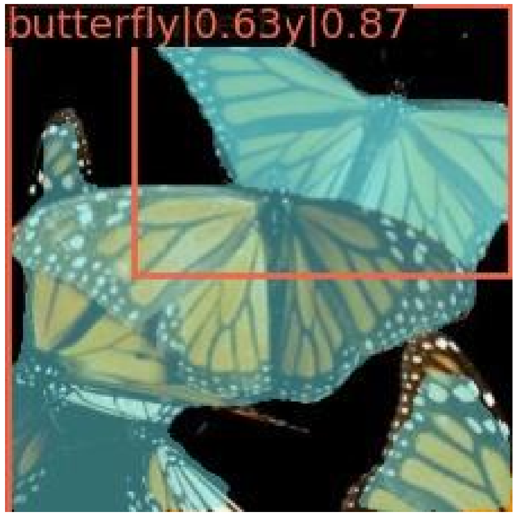
Results :

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.837
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.974
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.905
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 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.845
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.924
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.925
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.925
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 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.925


Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.791
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=1000 ] = 0.961
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=1000 ] = 0.902
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 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.811
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.863
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=300 ] = 0.863
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=1000 ] = 0.863
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 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.863
```

We can see that the results have been far better than the previous ones and that can be taken as a very nice prediction indeed. It has surpassed in every way from the previous models. The butterflies are super accurate.

We can see that the head parts were not having accurate masks but in this case,they are perfectly done by QueryInst. Some of the images containing multiple butterflies were not having the masks for all objects previously, but QueryInst also updated that.



Not only this, QueryInst also predicted correctly on those images where the butterflies were treated as squirrels by other models.

In Queryinst prediction most of the squirrels prediction surpassed the quality seen by other models. The nails and ears are perfectly predicted even predicted better.

In some cases of GCNet , the model prediction was giving masks outside of the object for GCNet , but QuertyInst also upgraded those and now, it can be treated as the best model , even if it can be taken in an auto-labeling process.

Conclusion :

After training on these 3 top models, we can find that QueryInst did a very nice job for this dataset. The perfection that QueryInst brought is quite like human-level / manual labeling. Also, in the note, the MaskRCNN model with Resnet101 backbone was implemented for this dataset, as the loss becomes NaN, so the model couldn't give any wise prediction and left the images unmasked, boxless, classless prediction, so, overfitting may happen some pieces were going in and ruin the prediction.

THE END

Performance Comparison Chart

Bounding Boxes (mAP)

Model / Dataset	Mask-RCNN	InstaBoost	Cascade Mask-RCNN on Swin Backbone	GCNet	QueryInst	Deformable Conv-Net
Thermal Dog Dataset	0.007	_	_	0.656	0.281	0.718
VHR-10 Dataset	_	0.661	0.716	_	_	0.537
TrashCan instance material Dataset	0.130	_	_	_	_	_
HardHat Dataset	_	_	0.199	0.392	_	0.389
Infrared Car-Person Dataset	_	_	0.332	0.457	0.118	0.462
Drone Gesture Control Dataset	0.710	_	_	0.709	0.738	0.751
Nature Dataset	NaN	_	_	0.764	0.837	0.683

Segmentations (mAP)

	Mask-RCNN	InstaBoost	Cascade Mask-RCNN with Swin Backbone	GCNet	QueryInst	Deformable Conv-Net
Thermal Dog Dataset	0.22	_	_	0.609	0.113	0.672
VHR-10 Dataset	_	0.661	0.657	_	_	0.498
TrashCan instance material Dataset	0.110	_	_	_	_	_
HardHat Dataset	_	_	0.184	0.382	_	0.369
Infrared Car-Person Dataset	_	_	0.286	0.413	0.097	0.418
Drone Gesture Control Dataset	0.693	_	_	0.656	0.690	0.712
Nature Dataset	NaN	_	_	0.739	0.791	0.665