

# CV\_Intern\_Visionlab\_IITD\_Assignment\_Sep\_2024

This report outlines the experiments and results of the assignment given by IITD. It involved the evaluation, analysis and fine tuning of the DINO object detection model on a pedestrian dataset containing 200 images.

## Ground Truth Visualization





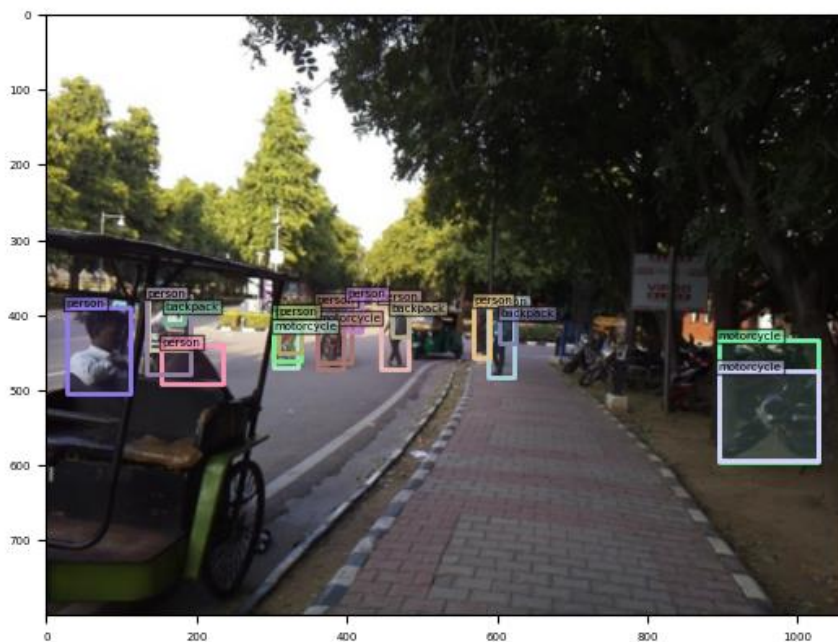
## Inference and Evaluation on pre-trained weights

Bounding box AP values with pre-trained model:

IoU metric: bbox						
Average Precision	(AP)	@[ IoU=0.50:0.95	area= all	maxDets=100	] =	0.484
Average Precision	(AP)	@[ IoU=0.50	area= all	maxDets=100	] =	0.845
Average Precision	(AP)	@[ IoU=0.75	area= all	maxDets=100	] =	0.505
Average Precision	(AP)	@[ IoU=0.50:0.95	area= small	maxDets=100	] =	0.406
Average Precision	(AP)	@[ IoU=0.50:0.95	area=medium	maxDets=100	] =	0.590
Average Precision	(AP)	@[ IoU=0.50:0.95	area= large	maxDets=100	] =	0.795
Average Recall	(AR)	@[ IoU=0.50:0.95	area= all	maxDets= 1	] =	0.100
Average Recall	(AR)	@[ IoU=0.50:0.95	area= all	maxDets= 10	] =	0.492
Average Recall	(AR)	@[ IoU=0.50:0.95	area= all	maxDets=100	] =	0.603
Average Recall	(AR)	@[ IoU=0.50:0.95	area= small	maxDets=100	] =	0.545
Average Recall	(AR)	@[ IoU=0.50:0.95	area=medium	maxDets=100	] =	0.687
Average Recall	(AR)	@[ IoU=0.50:0.95	area= large	maxDets=100	] =	0.836

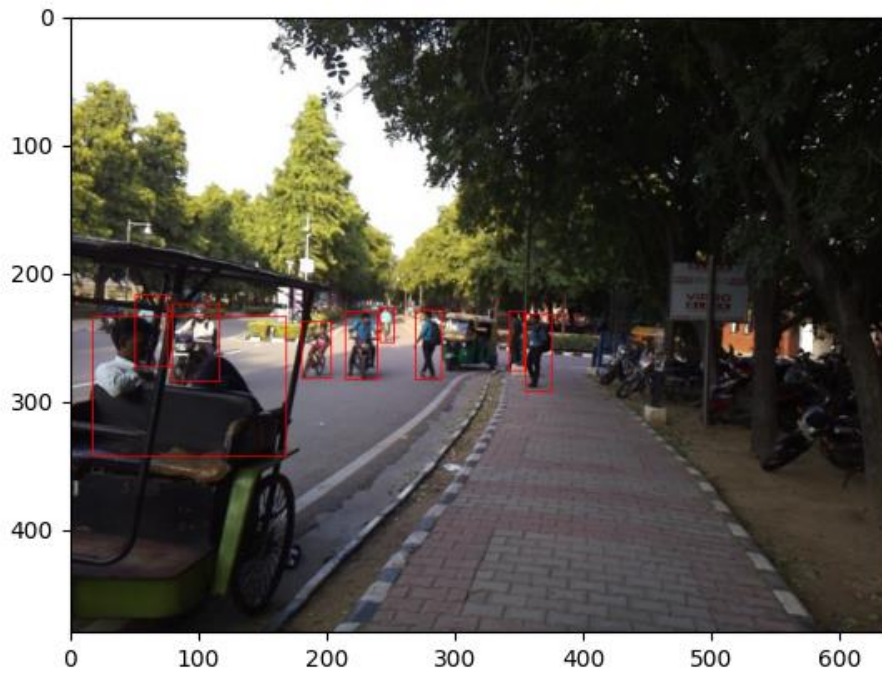
## Analysis

Inference with pre-trained model (checkpoint0033\_4scale.pth):

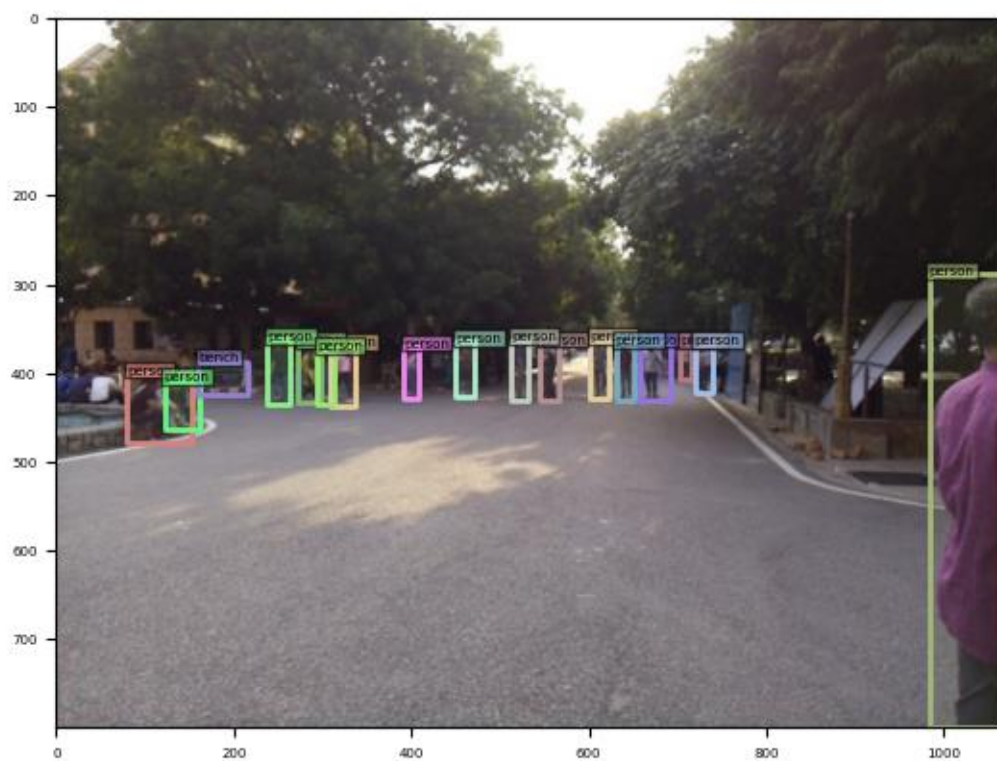




Ground truth:



- Results are fairly consistent with the ground truth. Larger objects are detected with more average precision while smaller objects with lesser AP.
- False negatives for people who are sitting and are farther away (see middle left)



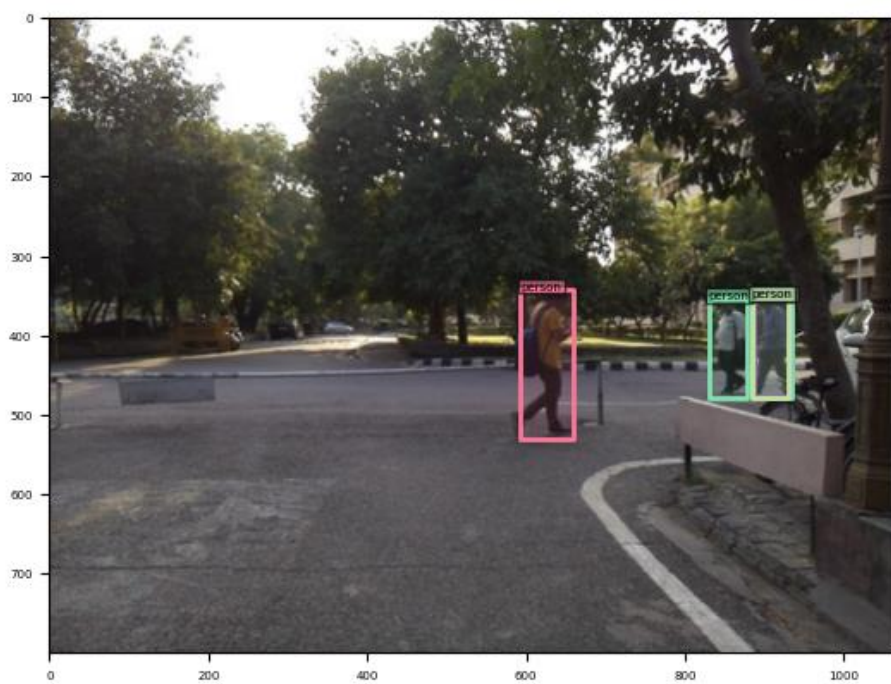
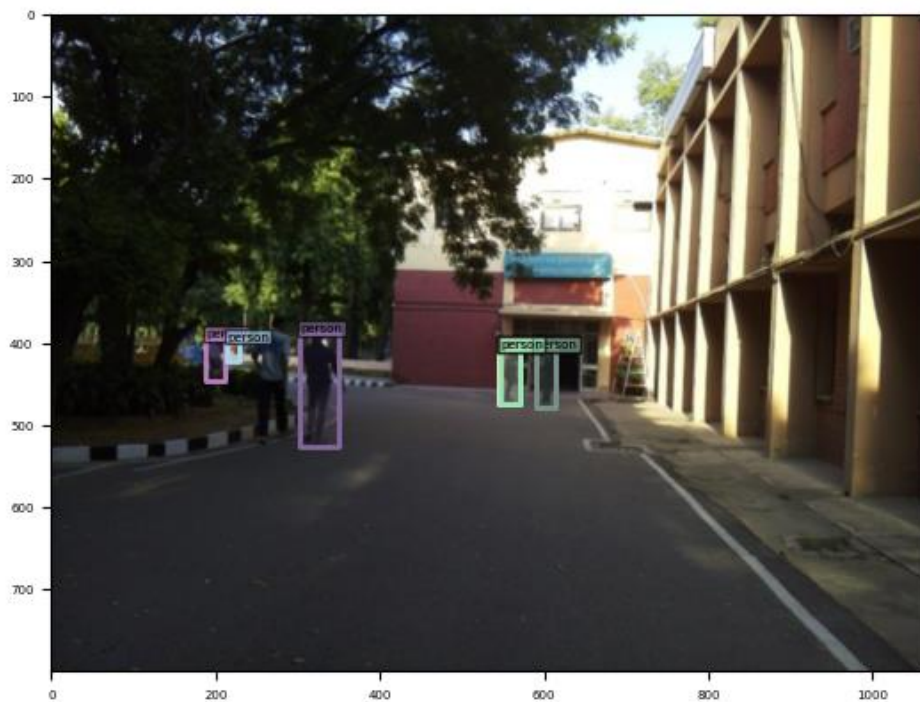
## Inference and Evaluation on fine-tuned weights

### 1. ATTEMPT 1

- Pre-trained model: checkpoint0033\_4scale.pth
- lr=0.00025
- epochs=30

```
IoU metric: bbox
Average Precision (AP) @[ IoU=0.50:0.95 | area= all | maxDets=100 ] = 0.305
Average Precision (AP) @[ IoU=0.50 | area= all | maxDets=100 ] = 0.628
Average Precision (AP) @[ IoU=0.75 | area= all | maxDets=100 ] = 0.248
Average Precision (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.247
Average Precision (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.447
Average Precision (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.252
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets= 1 ] = 0.076
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets= 10 ] = 0.368
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets=100 ] = 0.548
Average Recall (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.480
Average Recall (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.680
Average Recall (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.571
```

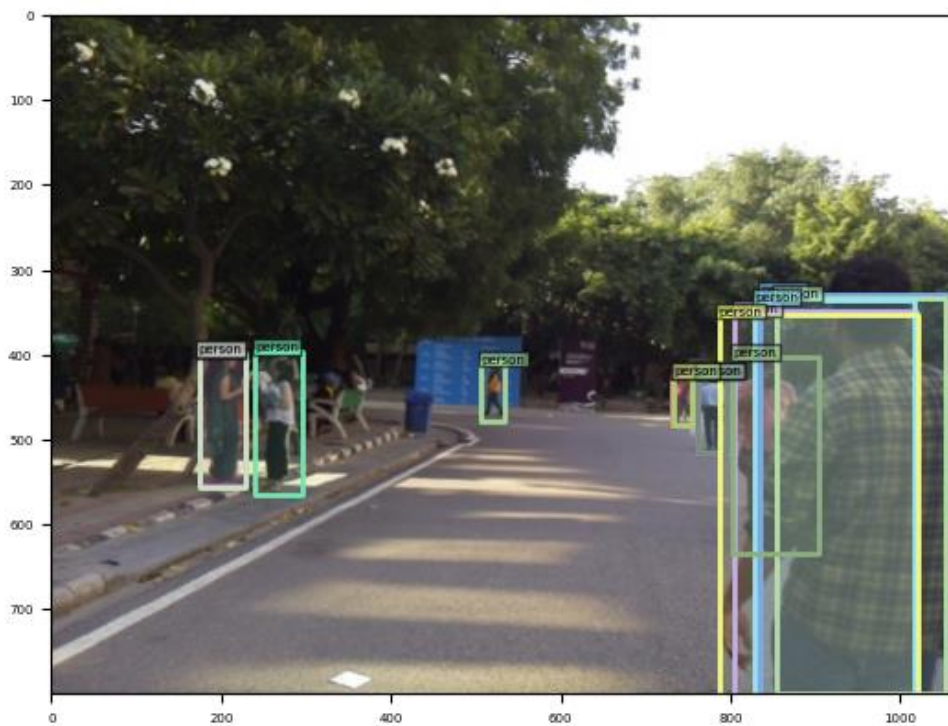






## Observations:

- Lots of false positives (overlapping bounding boxes) for small and large objects. Possibly, due to IOU threshold value.





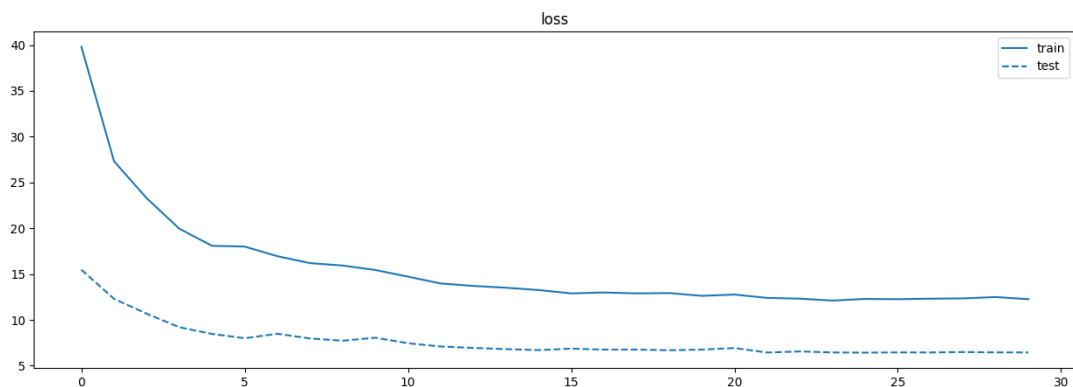
- Low sample size but occlusion handled well in this case:



- False negative for sitting people



Loss plot:



## 2. ATTEMPT 2

- Pre-trained model: checkpoint0011\_4scale.pth
- lr=0.0001
- epochs=12

```
IoU metric: bbox
Average Precision (AP) @[ IoU=0.50:0.95 | area= all | maxDets=100 ] = 0.874
Average Precision (AP) @[ IoU=0.50 | area= all | maxDets=100 ] = 0.891
Average Precision (AP) @[ IoU=0.75 | area= all | maxDets=100 ] = 0.717
Average Precision (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.621
Average Precision (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.919
Average Precision (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.825
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets= 1 ] = 0.806
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets= 10 ] = 0.755
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets=100 ] = 0.747
Average Recall (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.669
Average Recall (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.794
Average Recall (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.629
Training time 0:12:54
```

Due to time constraints (practical exam on 23<sup>rd</sup> Sept), I wasn't able to experiment more. If I could, I would've experimented by doing the following:

- If I had access to more GPU memory, I would have increased batch size (I used batch size of 2)
- Since the dataset is smaller, I would have run it for more epochs (I ran it for 30 epochs)
- In DETR based models, num\_queries generally refers to the number of object queries used in the model. It correlates to the number of objects the model can detect in a single image. Default value is 900. I would've tried with a smaller value (200-500)
- Since it's a small dataset which I'm running for more epochs, I would slightly increase weight decay to prevent overfitting. Weight decay adds a penalty term to the loss function that is proportional to the sum of the squared weights in the model.