Dhruv Maheshwari 9001607595

CSCI 677 – Advanced Computer Vision

Assignment 5

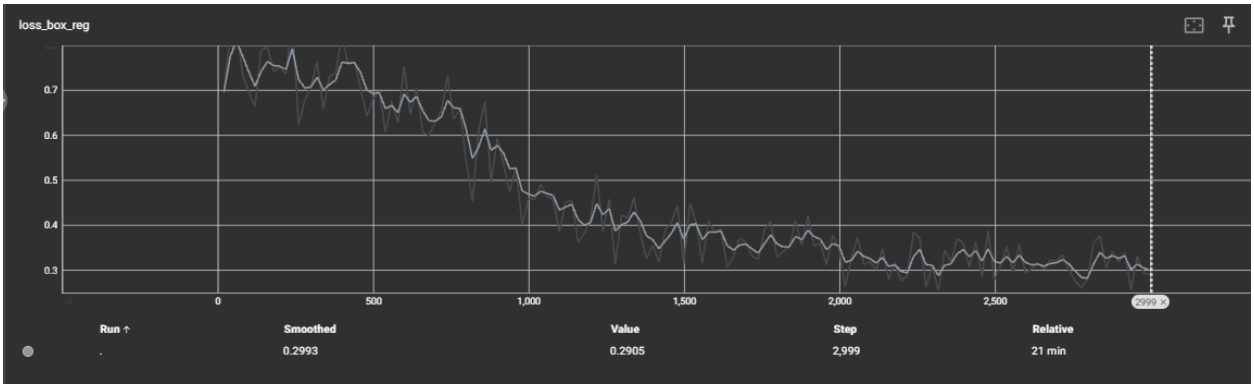
This assignment uses Detectron and DETR (Detection Transformer) for object detecttion.

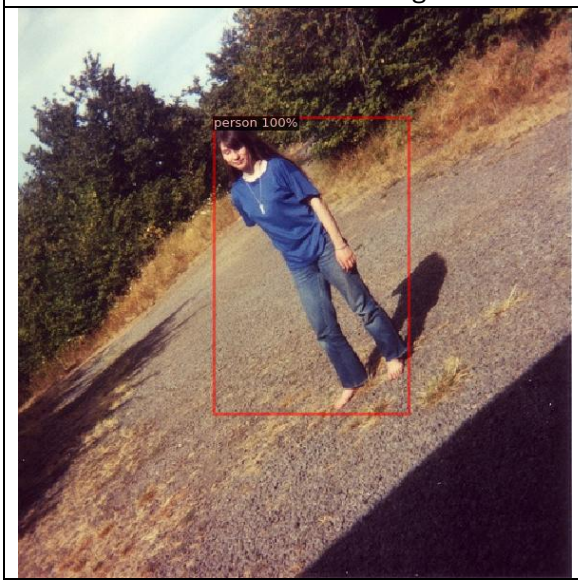
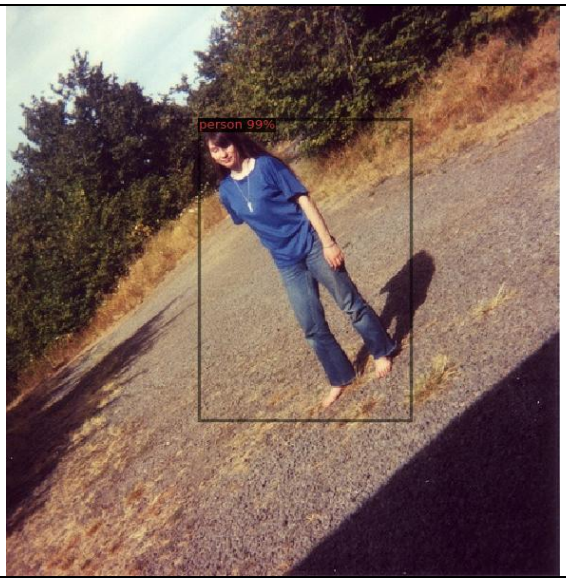
Detectron

(https://colab.research.google.com/drive/1R669tnd-KAyNg8oLYXDGp8jWqL9xxEKc?usp=sharing)

Model: COCO-Detection/faster_rcnn_R_50_FPN_3x.yaml
Base LR: 0.00025
Threshold: 0.7
Batch size: 2
Epochs: 3000

Loss:



Before fine tuning	Finetuned
	


bus 99%
person 100%
person 100%
person 92%
person 84%
person 99%
person 91%
bus 90%
person 94%
person 98%
person 99%
person 90%
person 95%
cycle 99%
person 86%
chair 94%
couch 91%
chair 84%
sofa 86%
person 89%
person 71%
person 96%
person 93%
person 96%
person 71%

AP50: 69.87272350926982

The fine tuned model is able to predict objects clearly, for example it is able to distinguish between a sofa and a couch, which a pretrained model isn't able to do. However, the confidence with which it predicts objects, is lower compared to the pretrained model by a slight margin.

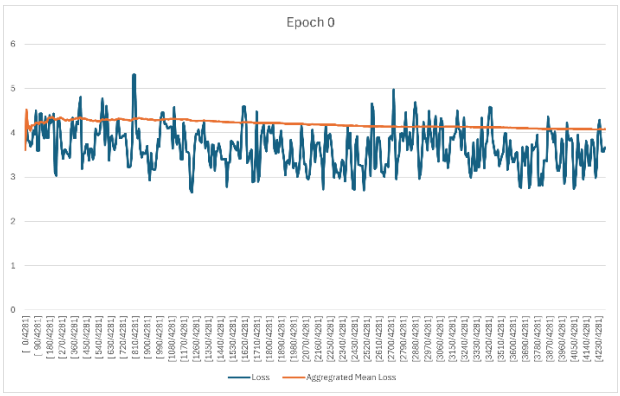
The pretrained model is able to find more objects within a scene, and also predicts multiple instances of the same object.

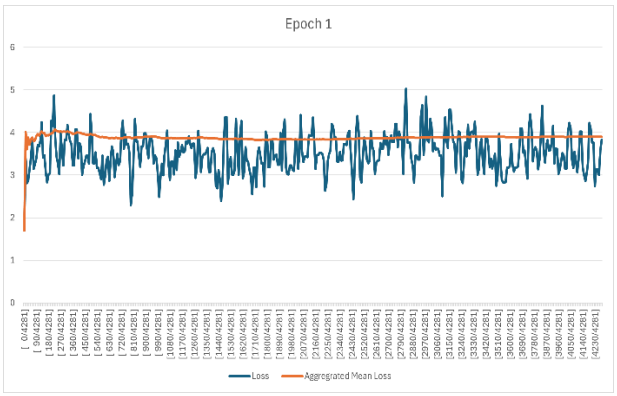
DETR

(https://colab.research.google.com/drive/1yNcDbwrWhw28sj1FPZ-dUpDYEX5mgj4M?usp=sharing)

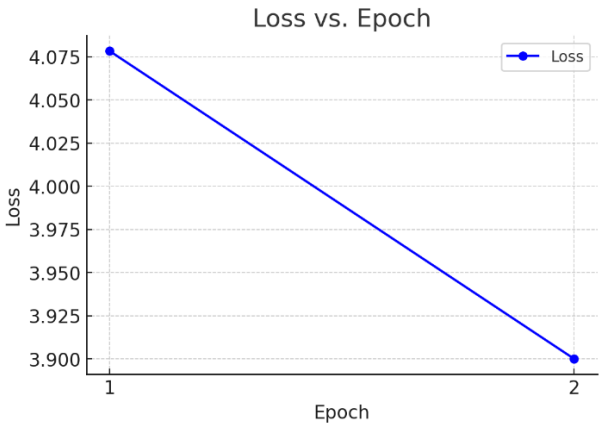
Model: detr_resnet50
LR: 0.00001, Backbone LR: 0.000001
Threshold: 0.7
Batch size: 4
Epochs: 2

Loss (mini batch for both epoch):





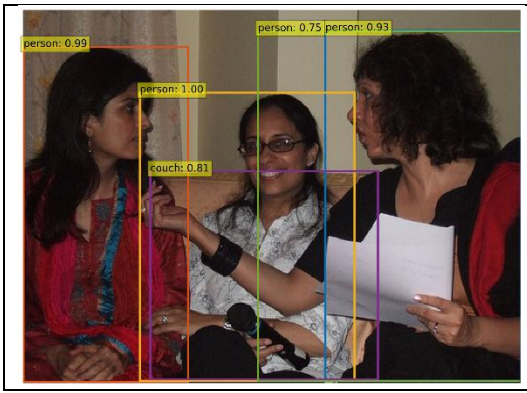
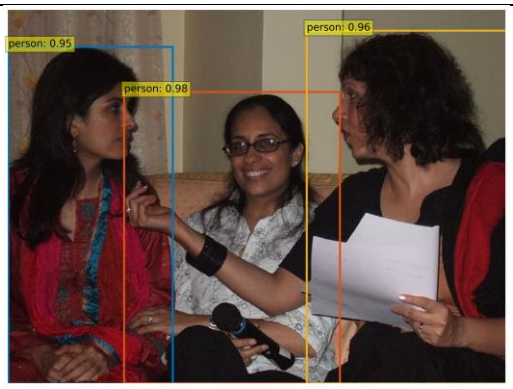
Loss:



Before fine tuning	Finetuned


person: 1.00
person: 1.00
person: 0.97
bus: 0.99
traffic
person: 1.00
traffic light: 0.78
person: 1.00
person: 0.92
person: 0.74
handbag: 0.98
bicycle: 1.00
person: 0.85
bus: 0.98
person: 0.99
person: 0.72
person: 0.88
person: 0.98
bicycle: 0.82
vase: 0.80
chair: 1.00
couch: 1.00
book: 0.80
couch: 0.81
book: 0.84
chair: 0.99
couch: 0.98
chair: 0.88

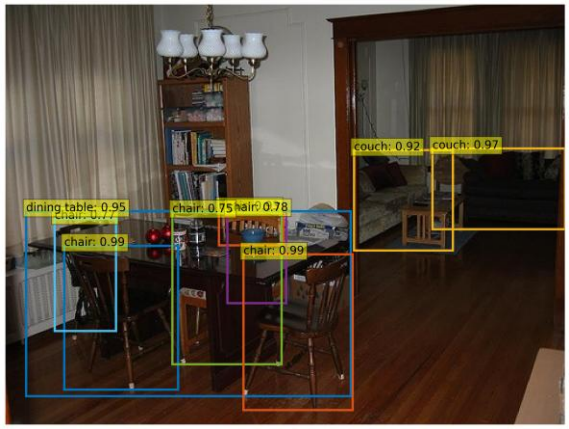

person: 0.99
person: 0.86
suitcase: 0.96
person: 0.94
person: 0.72
book: 0.92
book: 0.80
book: 0.87
couch: 0.95
couch: 0.99
laptop: 0.84
dining table: 0.94
chair: 1.00
cup: 0.90
chair: 1.00
chair: 0.90
couch: 0.92
couch: 0.97
dining table: 0.95
chair: 0.99
chair: 0.99
person: 1.00
bench: 0.72
person: 1.00
person: 0.99
person: 0.99

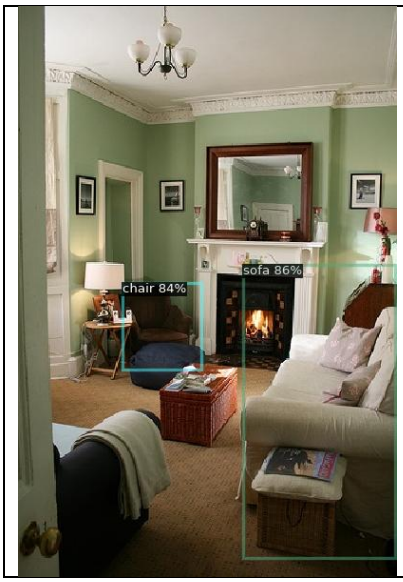
<table>
<tr>
<td> 
</td>
<td> 
</td>
</tr>
</table>

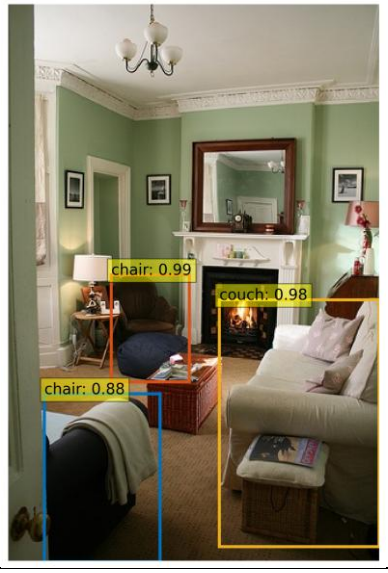
AP50: 0.783

The pretrained model is able to find more objects within a scene, and also predicts multiple instances of the same object.

<u>Comparing Detectron and DETR</u>

Detectron	DETR
AP50: 0.699	AP50: 0.783
	





DETR has better Average Precision. It is also able to find more objects and identify them, irrespective of they belong to same class or not. Also, it is able to give higher a confidence level while recognizing objects within a scene.