BONUS PROJECT

Object Detection on Cityscapes

Student Name: Anuhya Kothapally

Student ID:011017141

Student Email: anuhyak@uark.edu

a. Train the object detector on the Cityscapes dataset

Followed the steps in the repository https://github.com/truongthanhdat/YOLO-V7-Cityscapes

1. Clone the project and its submodules
2. Setup the environment with all the dependencies
3. Setup the dataset by downloading from Cityscapes website after registering an account and then converting the annotations to a format that Yolo can understand.

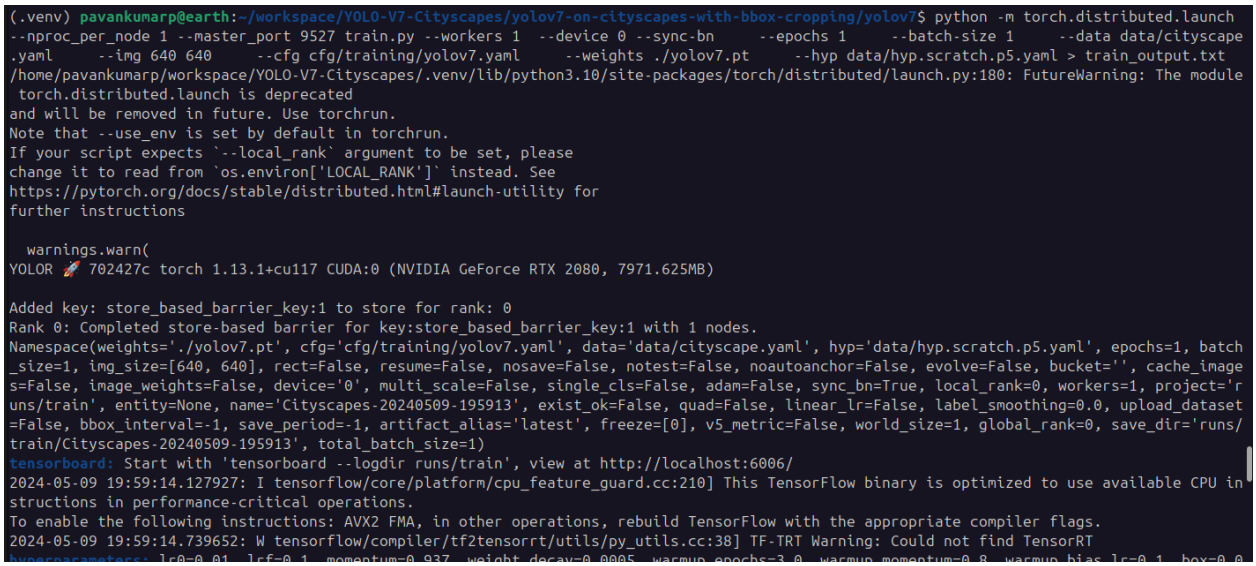
In order to train the model, there is train.py file within the repository that can be executed with various parameters.

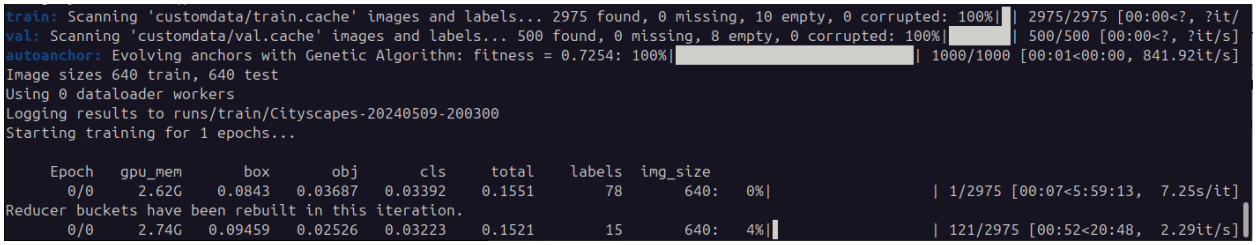
Following is a sample command when executed will start the training.

```
$ python -m torch.distributed.launch \
    --nproc_per_node 1 \
    --master_port 9527 \
    train.py \
    --workers 2 \
    --device 0 \
    --sync-bn \
    --epochs 100 \
    --batch-size 32 \
    --data data/cityscape.yaml \
    --img 640 640 \
    --cfg cfg/training/yolov7.yaml \
    --weights ./yolov7.pt \
    --hyp data/hyp.scratch.p5.yaml
```

However, due to GPU limitation I could not execute with a batch size of 32. I got an RTX-2080 GPU with 8GB total memory and executed with a batch size of 1. Each epoch took about 13 minutes to complete, ran the model for about 12 hours to execute see only 43 total epochs.

Here is the screen shot for the execution of train command with 1 epoch with a batch size of 1

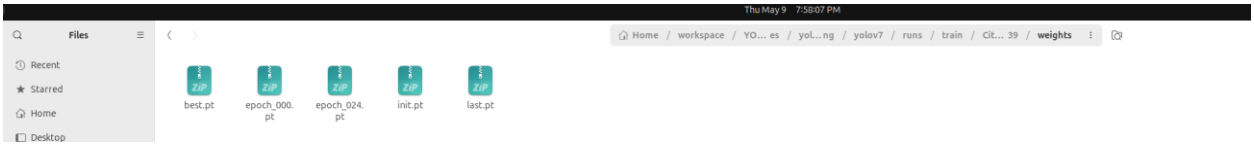




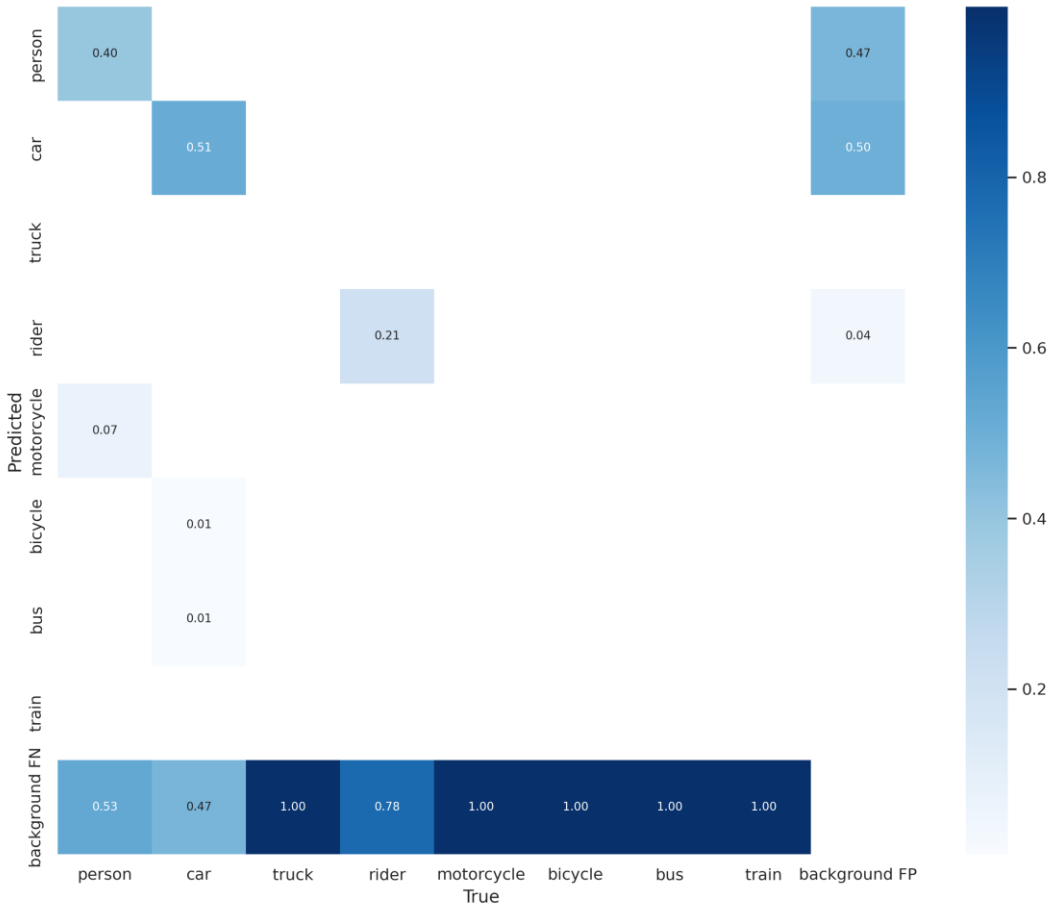
When the training is complete for 1 epoch.

```
     Epoch   gpu_mem       box       obj       cls     total    labels  img_size
       0/0     2.62G    0.0843   0.03687   0.03392    0.1551        78       640:   0%|                    | 1/2975 [00:07<5:59:13,  7.25s/it]
Reducer buckets have been rebuilt in this iteration.
       0/0     2.74G   0.06447   0.02391   0.01622    0.1046        40       640: 100%|████████████████████| 2975/2975 [18:17<00:00,  2.71it/s]
               Class      Images      Labels           P           R      mAP@.5  mAP@.5:.95: 100%|██████████| 250/250 [00:27<00:00,  9.00it/s]
1 epochs completed in 0.314 hours.

Images sizes do not match. This will causes images to be display incorrectly in the UI.
wandb:
wandb:
wandb: Run history:
wandb:      metrics/mAP_0.5 _
wandb: metrics/mAP_0.5:0.95 _
wandb:    metrics/precision _
wandb:       metrics/recall _
wandb:       train/box_loss _
wandb:       train/cls_loss _
wandb:       train/obj_loss _
wandb:         val/box_loss _
wandb:         val/cls_loss _
wandb:         val/obj_loss _
wandb:                x/lr0 _
wandb:                x/lr1 _
wandb:                x/lr2 _
wandb:
wandb: Run summary:
wandb:      metrics/mAP_0.5 0.16054
wandb: metrics/mAP_0.5:0.95 0.05574
wandb:    metrics/precision 0.67667
wandb:       metrics/recall 0.16185
wandb:       train/box_loss 0.06447
wandb:       train/cls_loss 0.01622
wandb:       train/obj_loss 0.02391
wandb:         val/box_loss 0.10797
wandb:         val/cls_loss 0.02275
wandb:         val/obj_loss 0.06018
wandb:                x/lr0 0.00333
wandb:                x/lr1 0.00333
wandb:                x/lr2 0.07001
wandb:
```

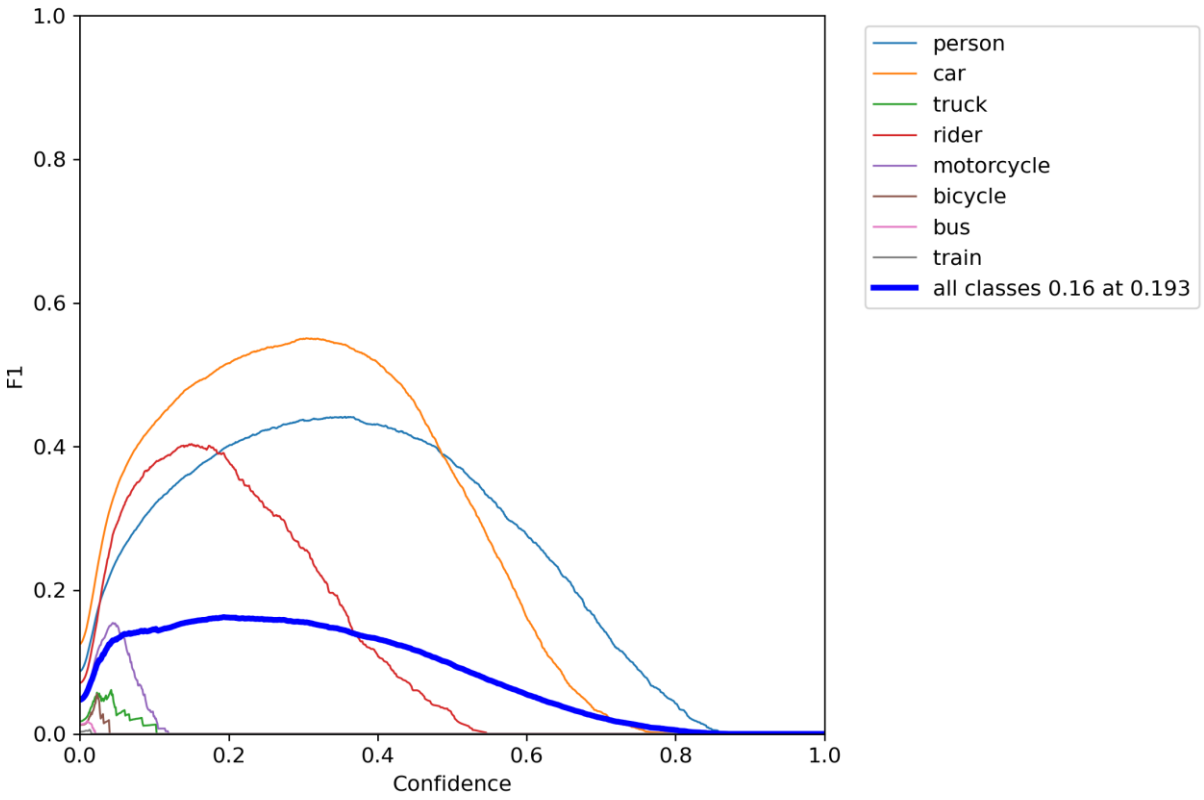
Sample checkpoints that got created.



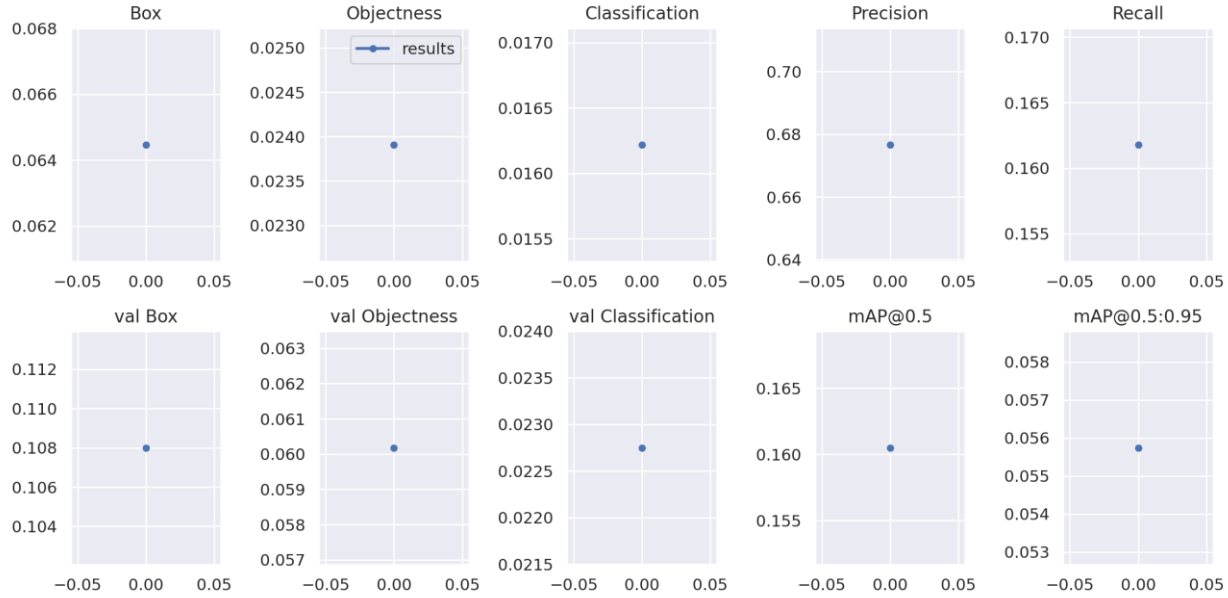
Confusion matrix for training log


person
car
truck
rider
motorcycle
bicycle
bus
train
background FN
Predicted
0.40
0.47
0.51
0.50
0.21
0.04
0.07
0.01
0.01
0.53
0.47
1.00
0.78
1.00
1.00
1.00
1.00
person
car
truck
rider
motorcycle
bicycle
bus
train
background FP
True
0.8
0.6
0.4
0.2

F1 curve


1.0
0.8
0.6
0.4
0.2
0.0
F1
0.0
0.2
0.4
0.6
0.8
1.0
Confidence
person
car
truck
rider
motorcycle
bicycle
bus
train
all classes 0.16 at 0.193

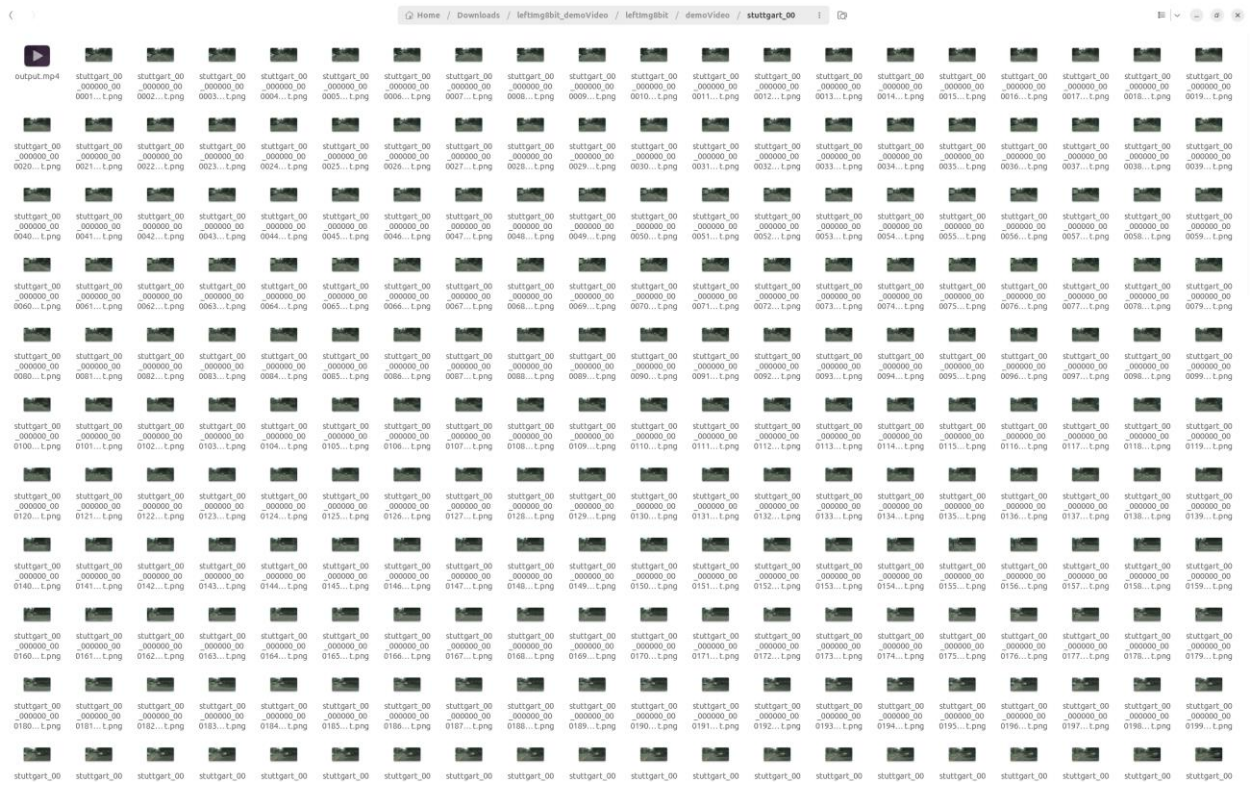
Results



B) Use your detector to detect objects on the videos provided in the Cityscapes dataset (leftImg8bit_demoVideo.zip). You can download the video from here: https://www.cityscapes-dataset.com/file-handling/?packageID=12

Downloaded the package and unzipped to notice that there three folders within the package.

Each containing several hundred png images

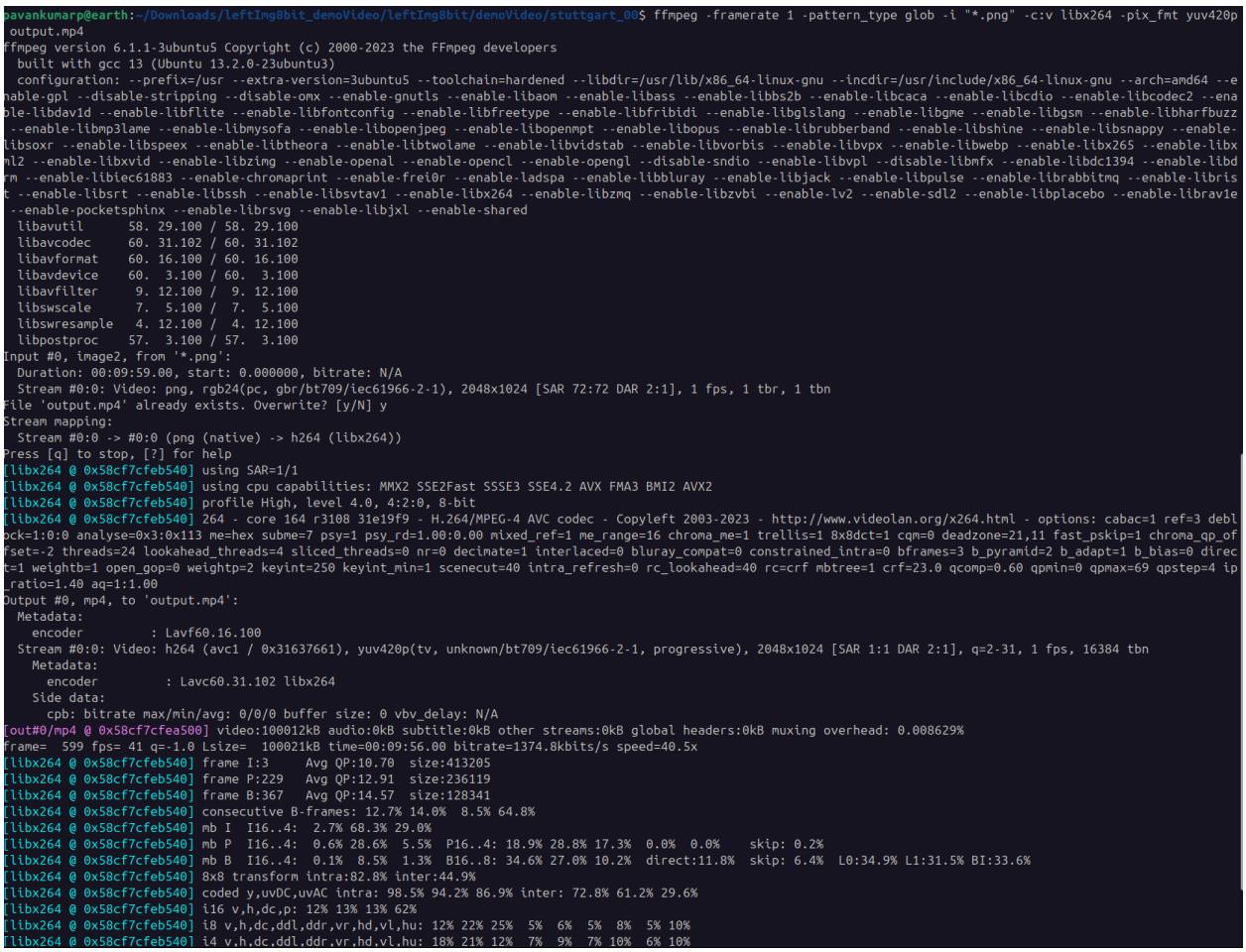


I have used ffmpeg to convert these images into a video. Framerate could be 1 or 25 fps depending on the preference.

Created one video for each folder, there by a total of 3 videos with names output.mp4, output1.mp4 and output2.mp4

Following is a sample execution.

ffmpeg -framerate 1 -pattern_type glob -i "*.png" -c:v libx264 -pix_fmt yuv420p output.mp4



Now I have provided these output.mp4, output1.mp4 and output2.mp4 videos as input to detect.py as input using the weights of yolov7_cityscapes.pt

python detect.py \

--weights yolov7_cityscapes.pt \

--conf 0.25 \

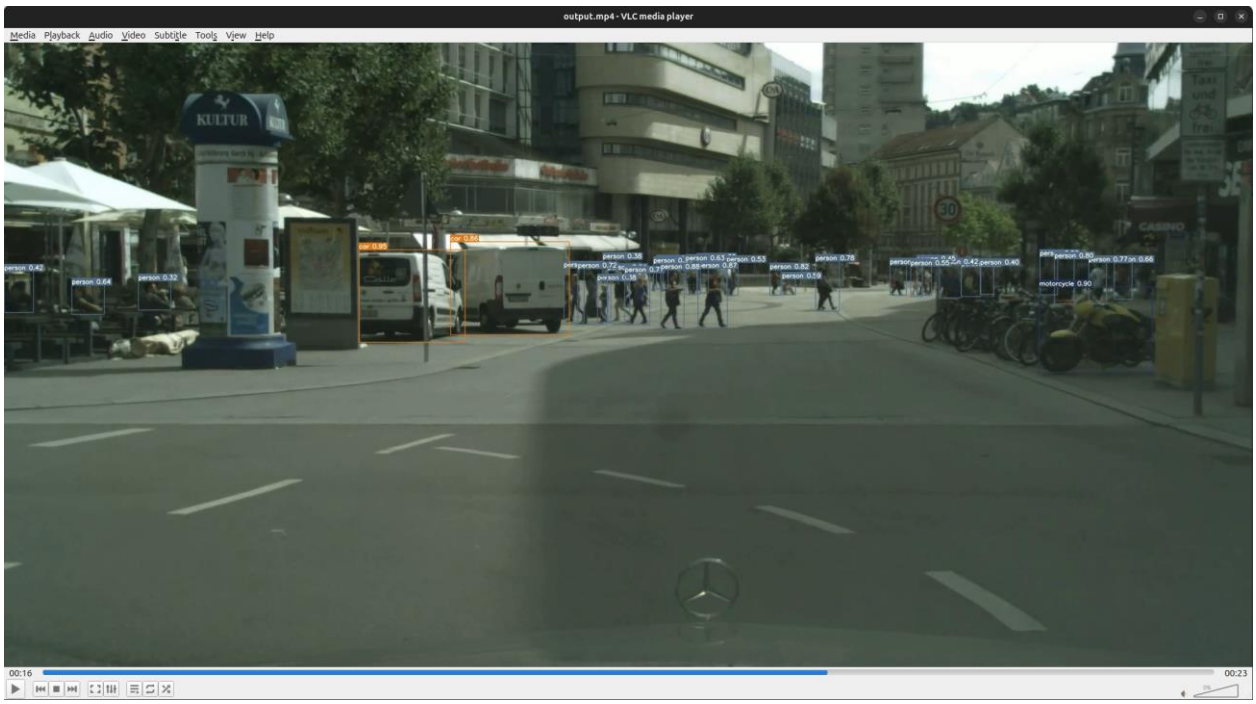
--img-size 640 \

--source output.mp4

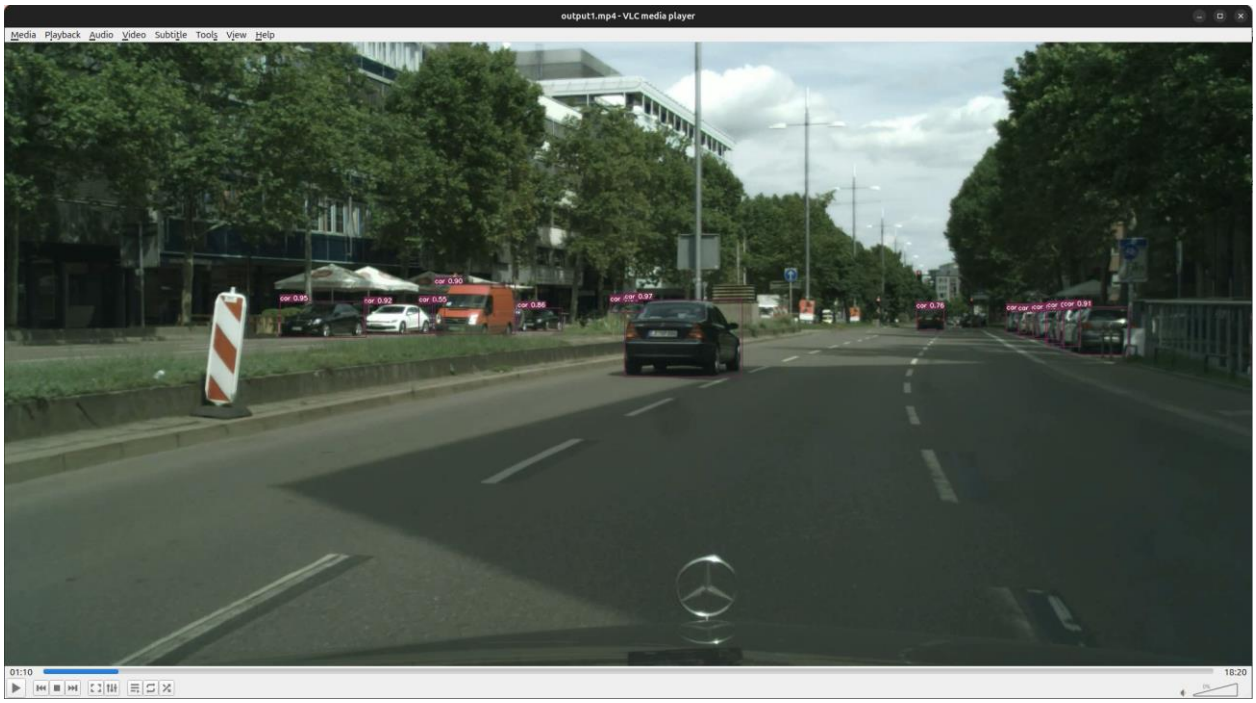
```
(.venv) pavankumarp@earth:~/workspace/YOLO-V7-Cityscapes/yolov7-on-cityscapes-with-bbox-cropping/yolov7$ python detect.py \
    --weights yolov7_cityscapes.pt \
    --conf 0.25 \
    --img-size 640 \
    --source output.mp4
Namespace(weights=['yolov7_cityscapes.pt'], source='output.mp4', img_size=640, conf_thres=0.25, iou_thres=0.45, device='', view_img=False, save_txt=False, save_conf=Fals
e, nosave=False, classes=None, agnostic_nms=False, augment=False, update=False, project='runs/detect', name='exp', exist_ok=False, no_trace=False, sr=False, sr_step=100)
/home/pavankumarp/workspace/YOLO-V7-Cityscapes/.venv/lib/python3.10/site-packages/huggingface_hub/file_download.py:1132: FutureWarning: `resume_download` is deprecated a
nd will be removed in version 1.0.0. Downloads always resume when possible. If you want to force a new download, use `force_download=True`.
  warnings.warn(
vqvae/diffusion_pytorch_model.safetensors not found
Loading pipeline components...: 100%|| 3/3 [00:00<00:00,  3.98it/s]
YOLOR 🚀 702427c torch 1.13.1+cu117 CUDA:0 (NVIDIA GeForce RTX 2080, 7971.625MB)

Fusing layers...
RepConv.fuse_repvgg_block
RepConv.fuse_repvgg_block
RepConv.fuse_repvgg_block
IDetect.fuse
/home/pavankumarp/workspace/YOLO-V7-Cityscapes/.venv/lib/python3.10/site-packages/torch/functional.py:504: UserWarning: torch.meshgrid: in an upcoming release, it will b
e required to pass the indexing argument. (Triggered internally at ../aten/src/ATen/native/TensorShape.cpp:3190.)
  return _VF.meshgrid(tensors, **kwargs)  # type: ignore[attr-defined]
Model Summary: 314 layers, 36524924 parameters, 6194944 gradients, 103.3 GFLOPS
 Convert model to Traced-model...
 traced_script_module saved!
 model is traced!

video 1/1 (1/599) /home/pavankumarp/workspace/YOLO-V7-Cityscapes/yolov7-on-cityscapes-with-bbox-cropping/yolov7/output.mp4: Cropped org img saved to: runs/detect/exp4/ou
tput_car_1_org.jpg
Cropped org img saved to: runs/detect/exp4/output_car_2_org.jpg
Cropped org img saved to: runs/detect/exp4/output_truck_0_org.jpg
Cropped org img saved to: runs/detect/exp4/output_car_3_org.jpg
Cropped org img saved to: runs/detect/exp4/output_car_4_org.jpg
Cropped org img saved to: runs/detect/exp4/output_car_5_org.jpg
11 persons, 6 cars, 1 truck, Done. (7.3ms) Inference, (1.6ms) NMS
video 1/1 (2/599) /home/pavankumarp/workspace/YOLO-V7-Cityscapes/yolov7-on-cityscapes-with-bbox-cropping/yolov7/output.mp4: Cropped org img saved to: runs/detect/exp4/ou
tput_car_0_org.jpg
Cropped org img saved to: runs/detect/exp4/output_car_1_org.jpg
Cropped org img saved to: runs/detect/exp4/output_car_2_org.jpg
Cropped org img saved to: runs/detect/exp4/output_car_3_org.jpg
Cropped org img saved to: runs/detect/exp4/output_car_4_org.jpg
Cropped org img saved to: runs/detect/exp4/output_truck_0_org.jpg
Cropped org img saved to: runs/detect/exp4/output_car_5_org.jpg
Cropped org img saved to: runs/detect/exp4/output_car_6_org.jpg
8 persons, 7 cars, 1 truck, Done. (9.2ms) Inference, (1.4ms) NMS
video 1/1 (3/599) /home/pavankumarp/workspace/YOLO-V7-Cityscapes/yolov7-on-cityscapes-with-bbox-cropping/yolov7/output.mp4: Cropped org img saved to: runs/detect/exp4/ou
tput_car_1_org.jpg
```

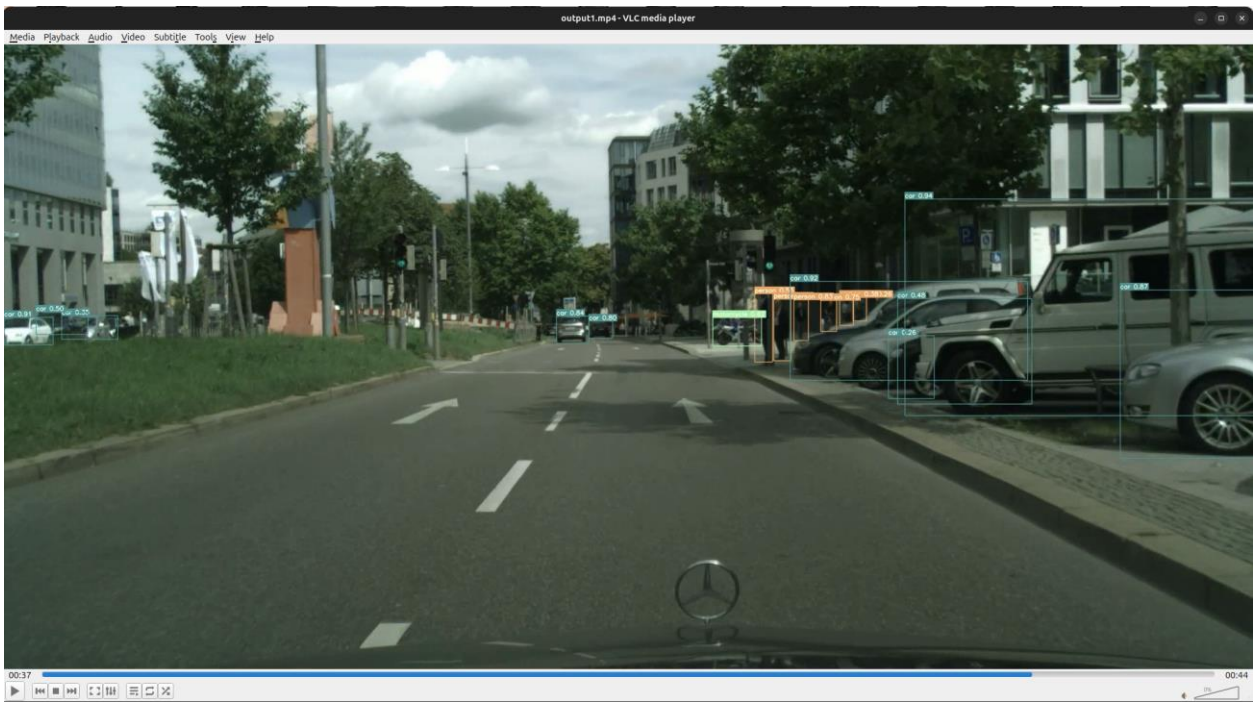
When inference is complete, a video is generated in runs/detect/exp{i}/ folder with bounding boxes

Output.mp4

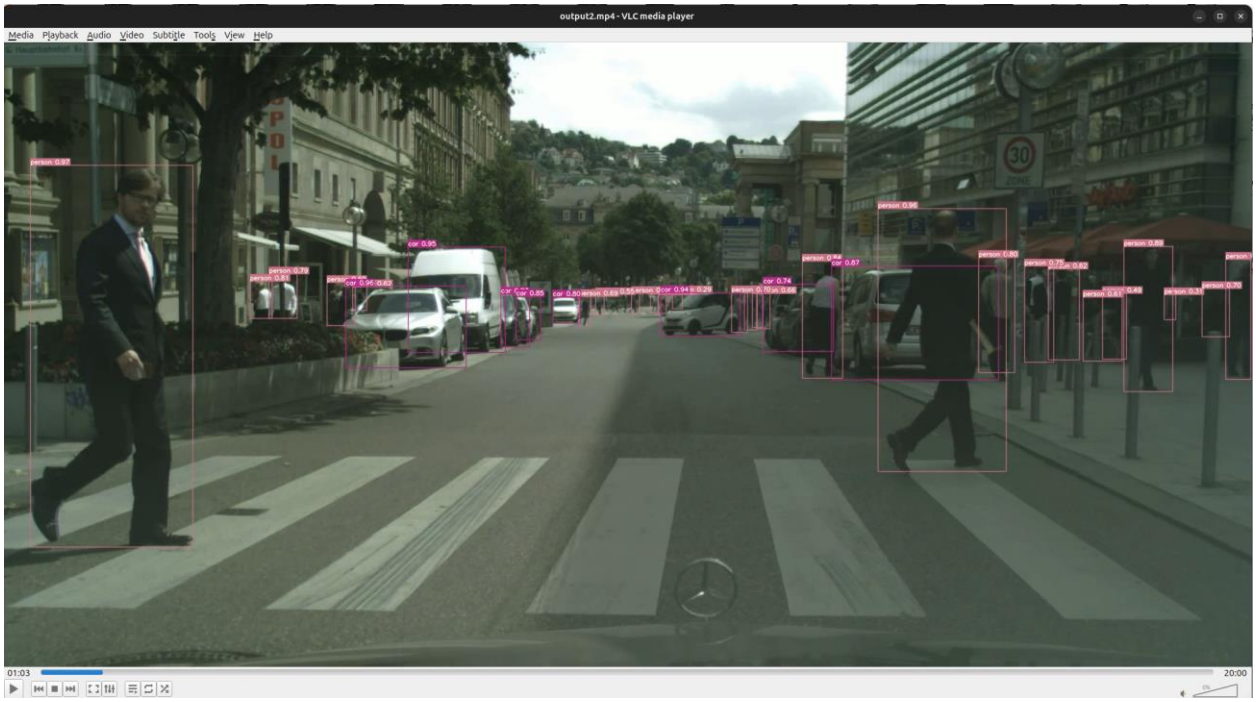

output.mp4 - VLC media player
Media Playback Audio Video Subtitle Tools View Help
KULTUR
00:16
00:23

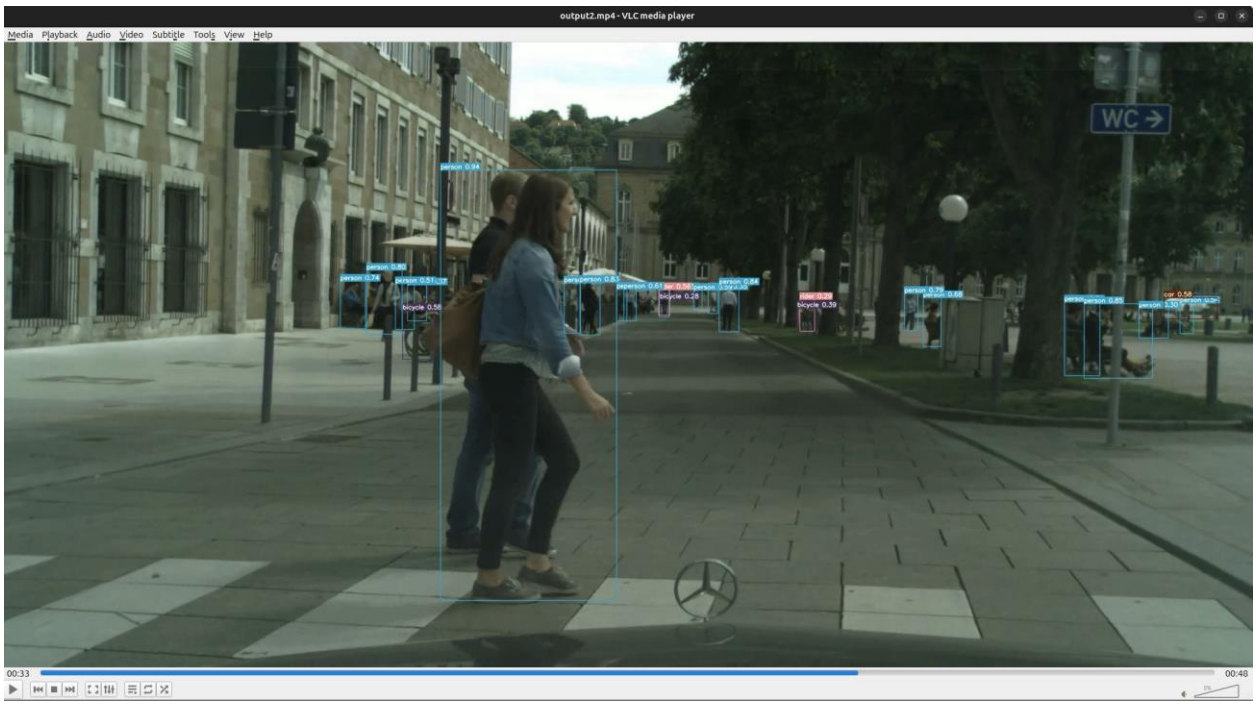
Output1.mp4


output1.mp4 - VLC media player
Media Playback Audio Video Subtitle Tools View Help
01:10
18:20


output1.mp4 - VLC media player
Media Playback Audio Video Subtitle Tools View Help
00:37
00:44

Output2.mp4


output2.mp4 - VLC media player
Media Playback Audio Video Subtitle Tools View Help
01:03
20:00


output2.mp4 - VLC media player
Media Playback Audio Video Subtitle Tools View Help
WC
00:33
00:48