

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/truongthanhdai/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

```
(.venv) pavankumarp@earth:~/workspace/YOLO-V7-Cityscapes/yolov7-on-cityscapes-with-bbox-cropping/yolov7$ python -m torch.distributed.launch
--nproc_per_node 1 --master_port 9527 train.py --workers 1 --device 0 --sync-bn --epochs 1 --batch-size 1 --data data/cityscape
.yaml --img 640 640 --cfg cfg/training/yolov7.yaml --weights ./yolov7.pt --hyp data/hyp.scratch.p5.yaml > train_output.txt
/home/pavankumarp/workspace/YOLO-V7-Cityscapes/.venv/lib/python3.10/site-packages/torch/distributed/launch.py:180: FutureWarning: The module
torch.distributed.launch is deprecated
and will be removed in future. Use torchrun.
Note that --use_env is set by default in torchrun.
If your script expects '--local_rank' argument to be set, please
change it to read from 'os.environ['LOCAL_RANK']' instead. See
https://pytorch.org/docs/stable/distributed.html#launch-utility for
further instructions

warnings.warn(
YOLOR 🚀 702427c torch 1.13.1+cu117 CUDA:0 (NVIDIA GeForce RTX 2080, 7971.625MB)

Added key: store_based_barrier_key:1 to store for rank: 0
Rank 0: Completed store-based barrier for key:store_based_barrier_key:1 with 1 nodes.
Namespace(weights='./yolov7.pt', cfg='cfg/training/yolov7.yaml', data='data/cityscape.yaml', hyp='data/hyp.scratch.p5.yaml', epochs=1, batch
_size=1, img_size=[640, 640], rect=False, resume=False, nosave=False, notest=False, noautoanchor=False, evolve=False, bucket='', cache_image
s=False, image_weights=False, device='0', multi_scale=False, single_cls=False, adam=False, sync_bn=True, local_rank=0, workers=1, project='r
uns/train', entity=None, name='Cityscapes-20240509-195913', exist_ok=False, quad=False, linear_lr=False, label_smoothing=0.0, upload_dataset
=False, bbox_interval=-1, save_period=-1, artifact_alias='latest', freeze=[0], v5_metric=False, world_size=1, global_rank=0, save_dir='runs/
train/Cityscapes-20240509-195913', total_batch_size=1)
tensorboard: Start with 'tensorboard --logdir runs/train', view at http://localhost:6006/
2024-05-09 19:59:14.127927: I tensorflow/core/platform/cpu_feature_guard.cc:210] This TensorFlow binary is optimized to use available CPU in
structions in performance-critical operations.
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
2024-05-09 19:59:14.739652: W tensorflow/compiler/tf2tensorrt/utls/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
hyperparameters: lr=0.01, lrf=0.1, momentum=0.937, weight_decay=0.0005, warmup_epochs=3.0, warmup_momentum=0.8, warmup_bias_lr=0.1, box=0.0

train: Scanning 'customdata/train.cache' images and labels... 2975 found, 0 missing, 10 empty, 0 corrupted: 100%|██████████| 2975/2975 [00:00<?, ?it/
val: Scanning 'customdata/val.cache' images and labels... 500 found, 0 missing, 8 empty, 0 corrupted: 100%|██████████| 500/500 [00:00<?, ?it/s]
autoanchor: Evolving anchors with Genetic Algorithm: fitness = 0.7254: 100%|██████████| 1000/1000 [00:01<00:00, 841.92it/s]
Image sizes 640 train, 640 test
Using 0 dataloader workers
Logging results to runs/train/Cityscapes-20240509-200300
Starting training for 1 epochs...

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%|
Reducer buckets have been rebuilt in this iteration.
0/0      2.74G    0.09459 0.02526  0.03223  0.1521    15      640:    4%|
```

When the training is complete for 1 epoch.

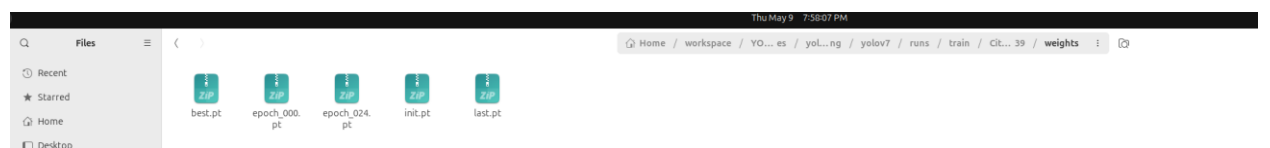
```

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
1 epochs completed in 0.314 hours.

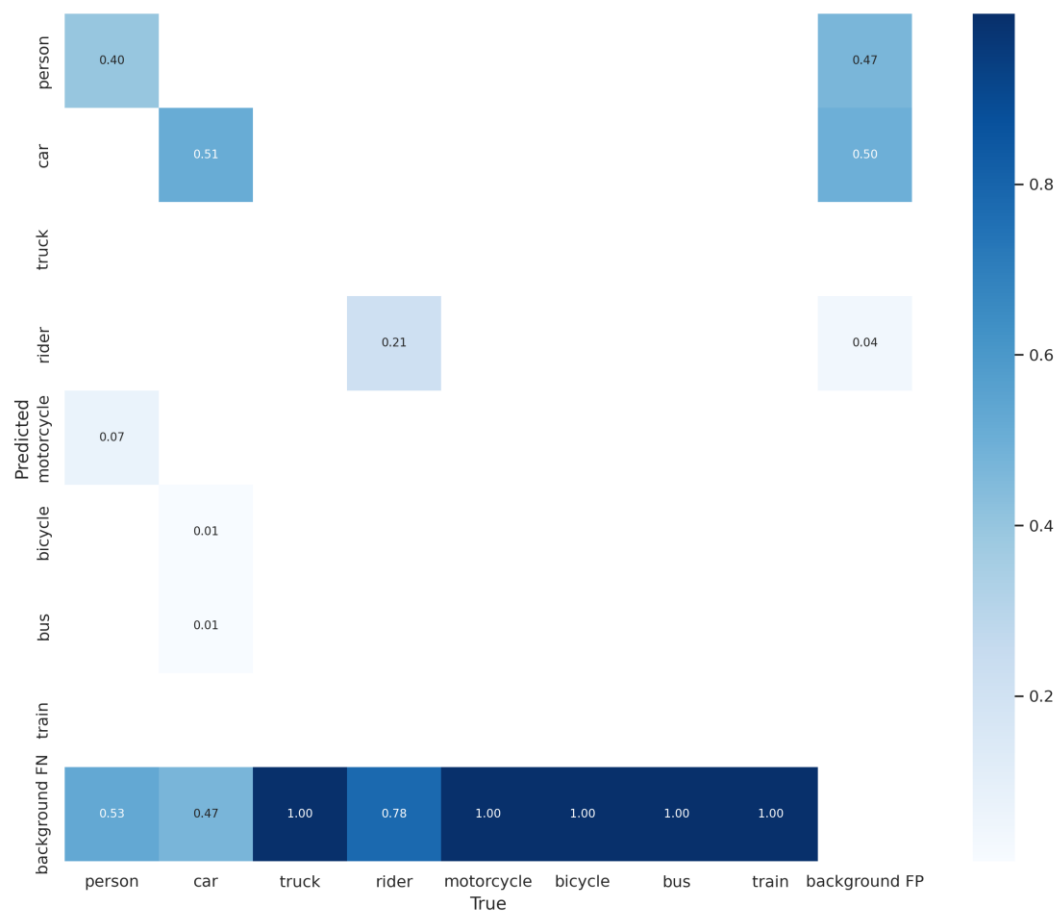
Images sizes do not match. This will causes images to be display incorrectly in the UI.
wandb:
wandb:
wandb: Run history:
wandb:   metrics/mAP_0.5 -
wandb: metrics/mAP_0.5:0.95 -
wandb:   metrics/precision -
wandb:   metrics/recall -
wandb:   train/box_loss -
wandb:   train/cls_loss -
wandb:   train/obj_loss -
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:

```

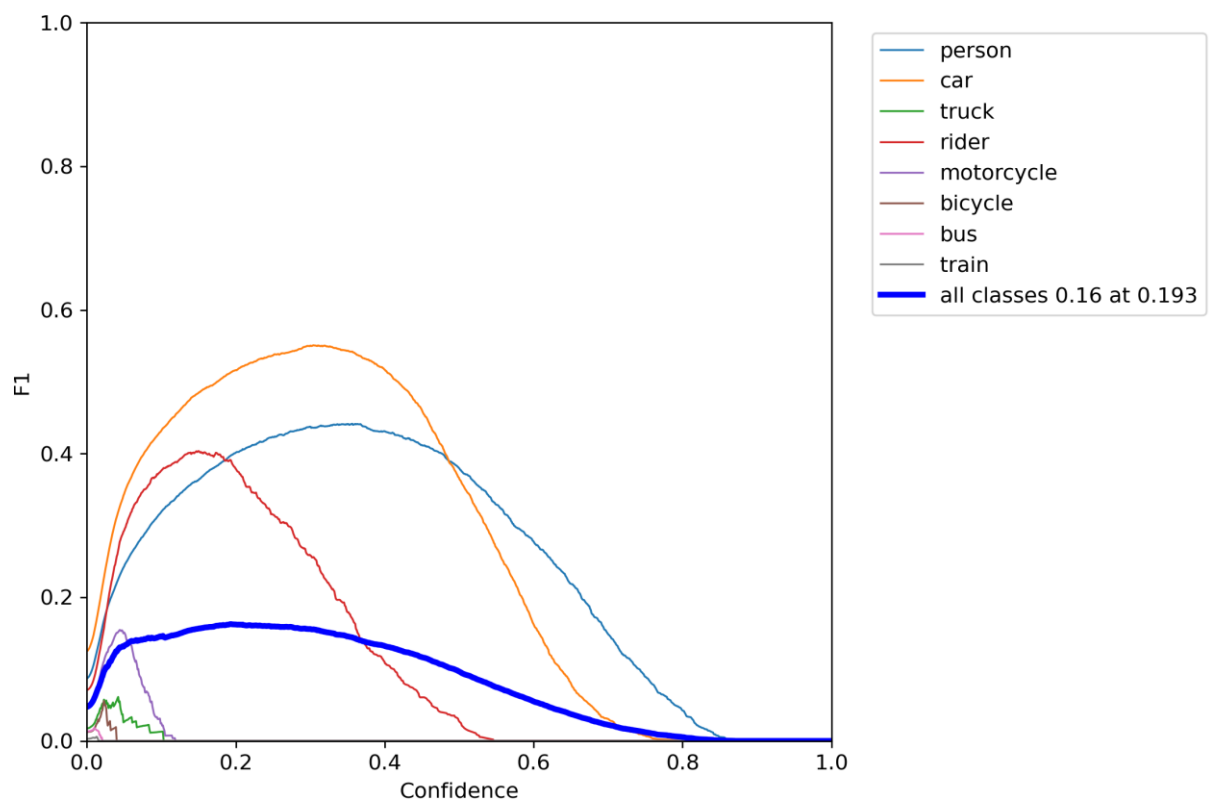
Sample checkpoints that got created.



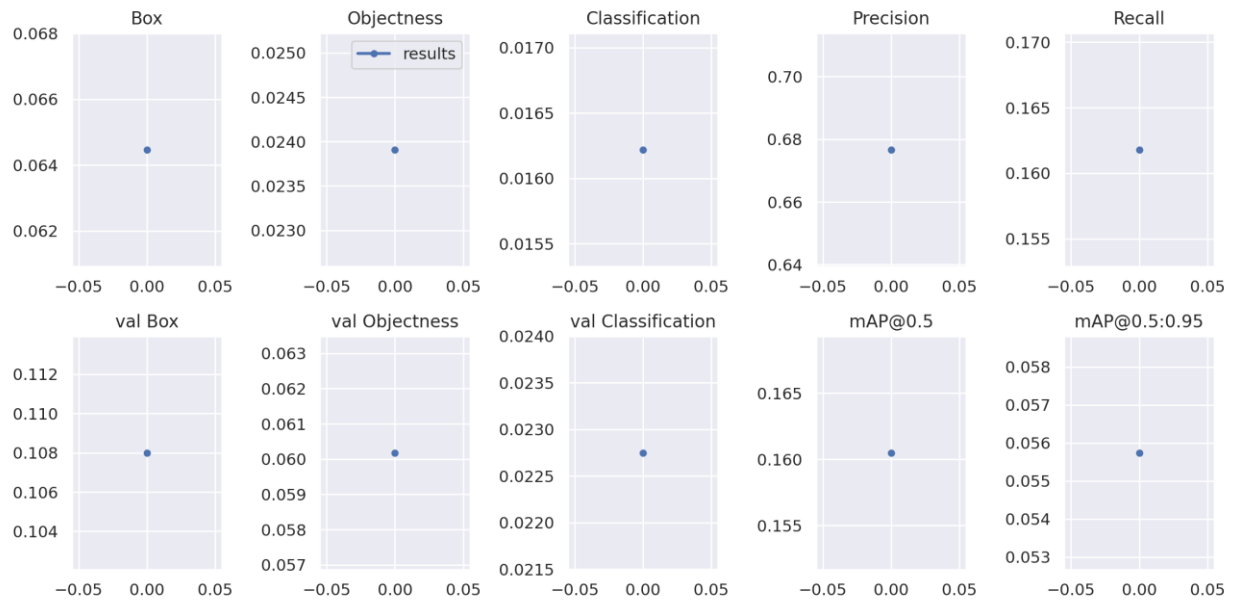
Confusion matrix for training log



F1 curve



Results

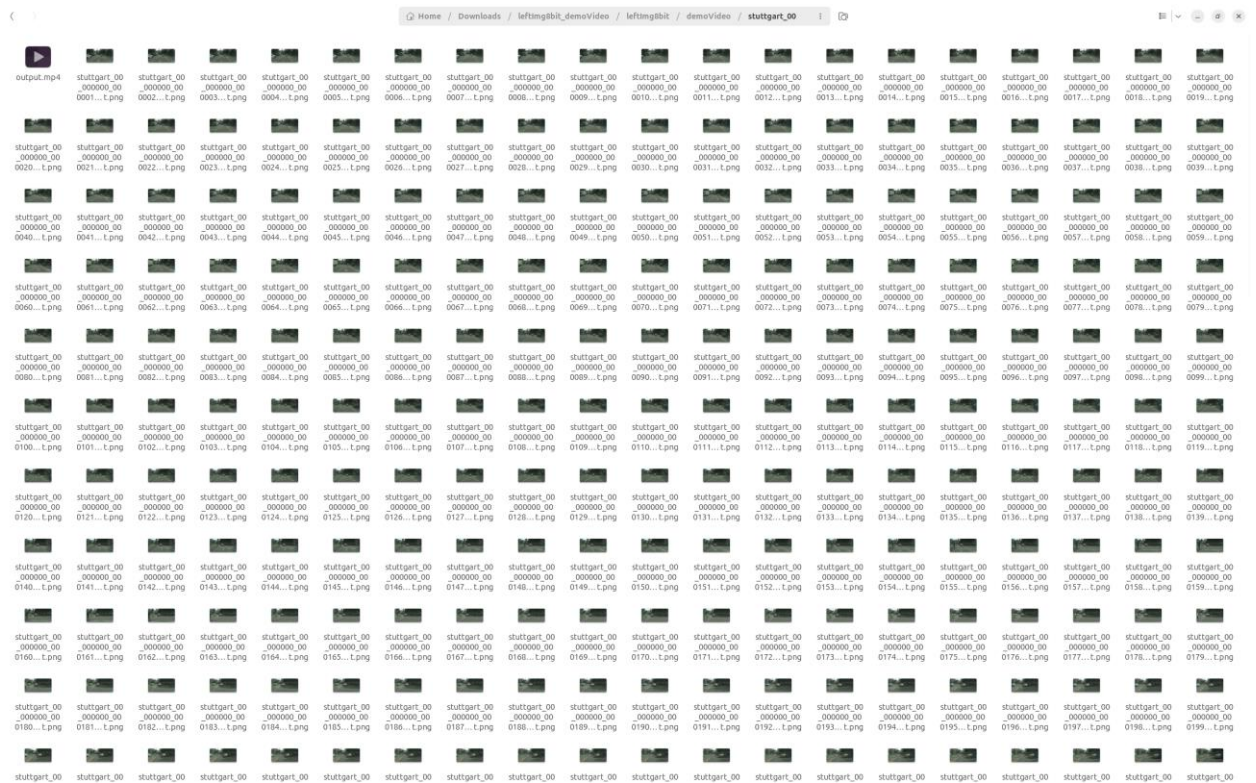


B) Use your detector to detect objects on the videos provided in the Cityscapes dataset (leftImg8bit_demoVideo.zip). You can download the video from here: <https://www.cityscapes-dataset.com/file-handling/?packageID=12>

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
```



```
(.venv) pavankunarp@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=False, 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/pavankunarp/workspace/YOLO-V7-Cityscapes/.venv/lib/python3.10/site-packages/huggingface_hub/file_download.py:1132: FutureWarning: 'resume_download' is deprecated and 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%
YOLOv7 792427c torch 1.13.1+cu117 CUDA:0 (NVIDIA GeForce RTX 2080, 7971.625MB)

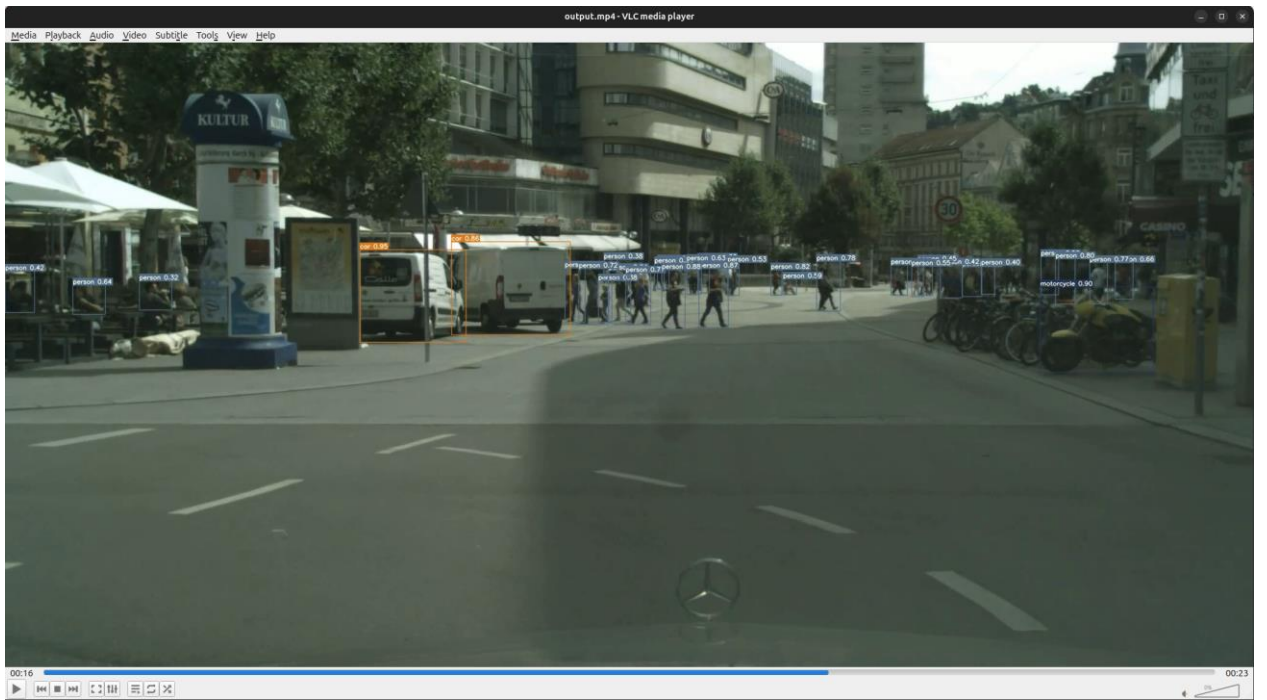
Fusing layers...
RepConv.fuse_repvgg_block
RepConv.fuse_repvgg_block
RepConv.fuse_repvgg_block
IDetect.fuse
/home/pavankunarp/workspace/YOLO-V7-Cityscapes/.venv/lib/python3.10/site-packages/torch/functional.py:584: UserWarning: torch.meshgrid: in an upcoming release, it will be required to pass the indexing 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/pavankunarp/workspace/YOLO-V7-Cityscapes/yolov7-on-cityscapes-with-bbox-cropping/yolov7/output.mp4: 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_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/pavankunarp/workspace/YOLO-V7-Cityscapes/yolov7-on-cityscapes-with-bbox-cropping/yolov7/output.mp4: Cropped org img saved to: runs/detect/exp4/output_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/pavankunarp/workspace/YOLO-V7-Cityscapes/yolov7-on-cityscapes-with-bbox-cropping/yolov7/output.mp4: Cropped org img saved to: runs/detect/exp4/output_car_1_org.jpg
```

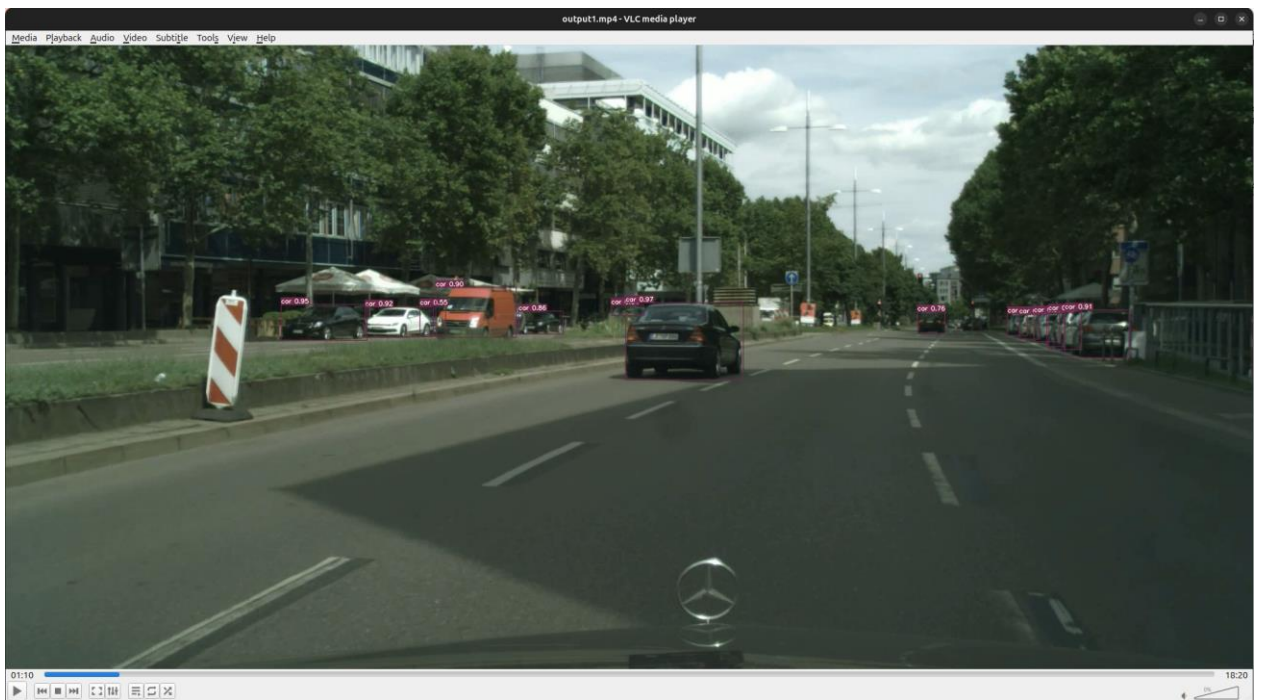
When inference is complete, a video is generated in runs/detect/exp{i}/ folder with bounding boxes

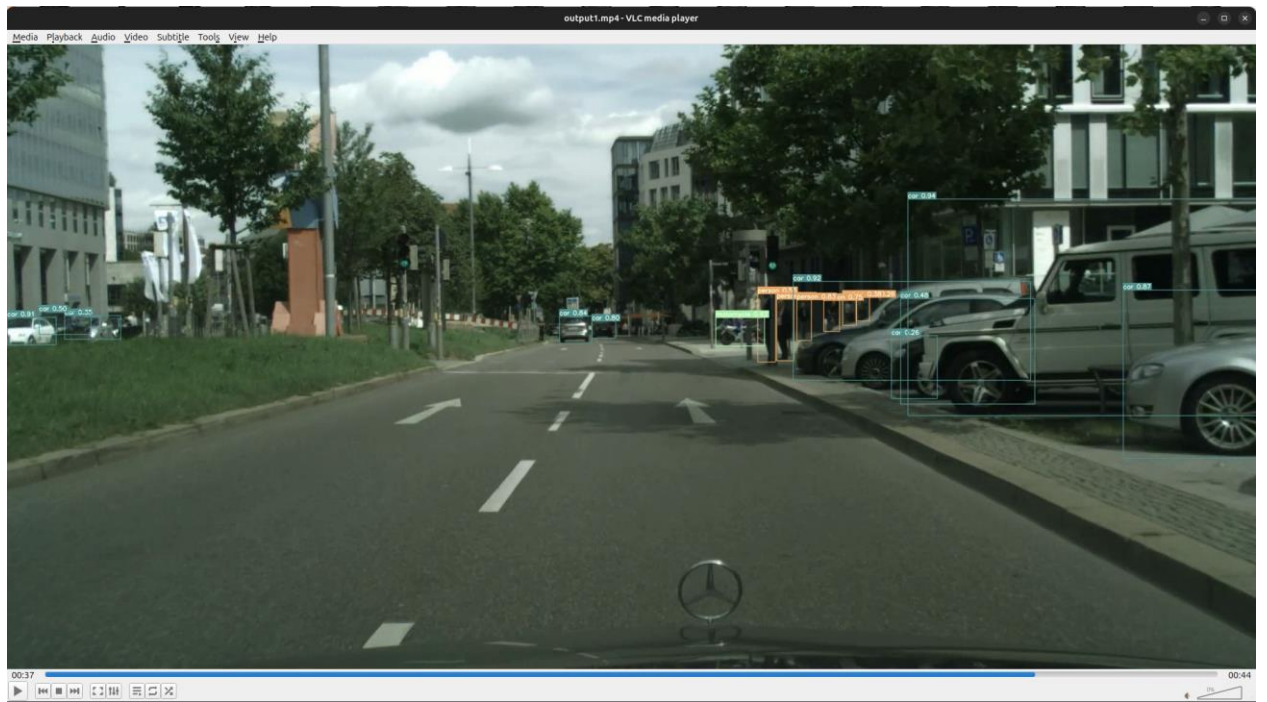
Output.mp4





Output1.mp4





Output2.mp4

