IMAGE TAGGING AND ROAD OBJECT DETECTION

Guide : Prof. Anoop M. Namboodiri

Mentor : Sangeeth Reddy

- Richard Deepak
- Rahul Juluru
- Sridhar Deshpande
- Kshama Pandey

5

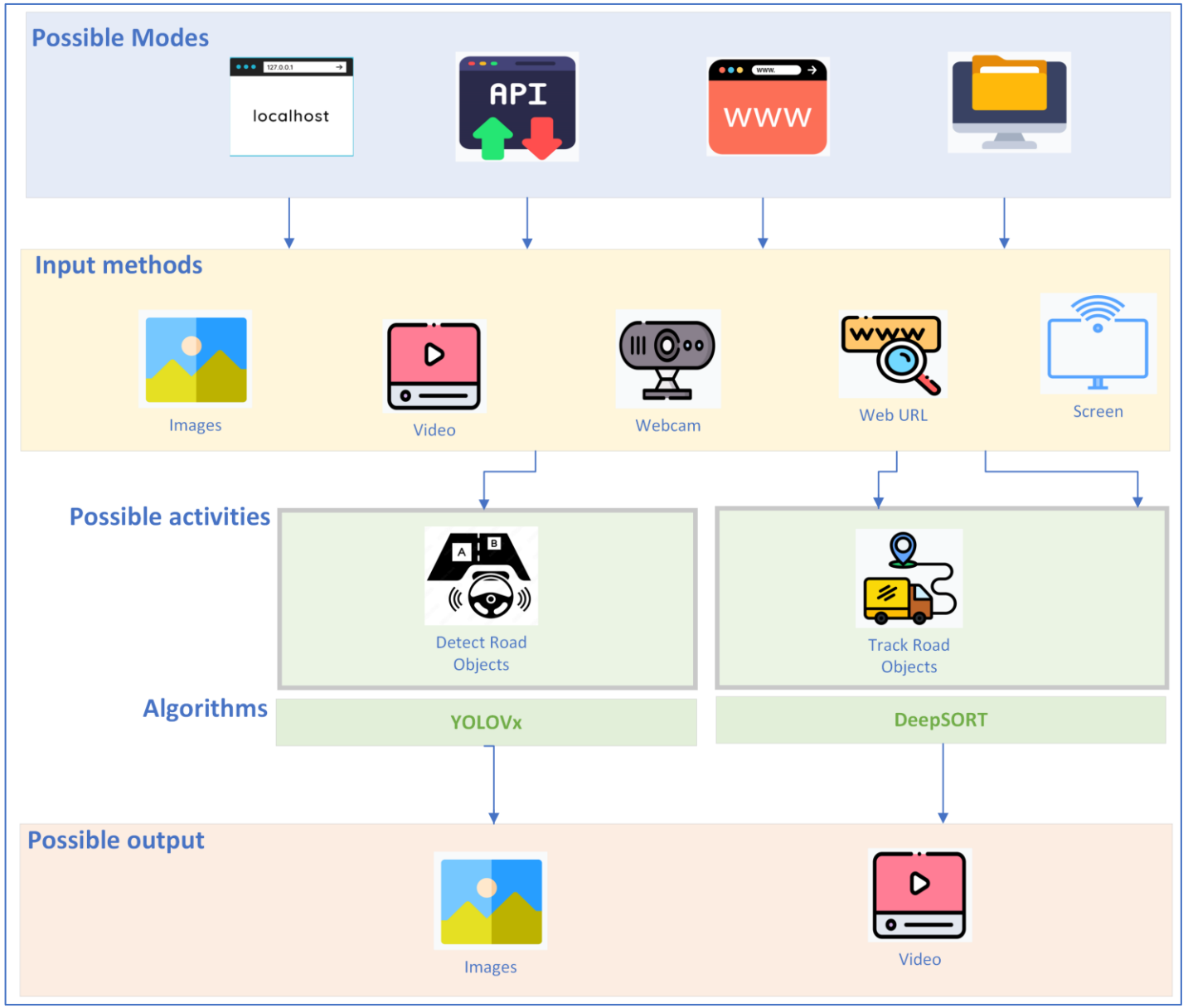
Agenda

- Objective .
- Methodology and design considerations.
- Outcome in stages.
- Challenges.
- Application Demo.
- Going Beyond.
- Summary.

Objective

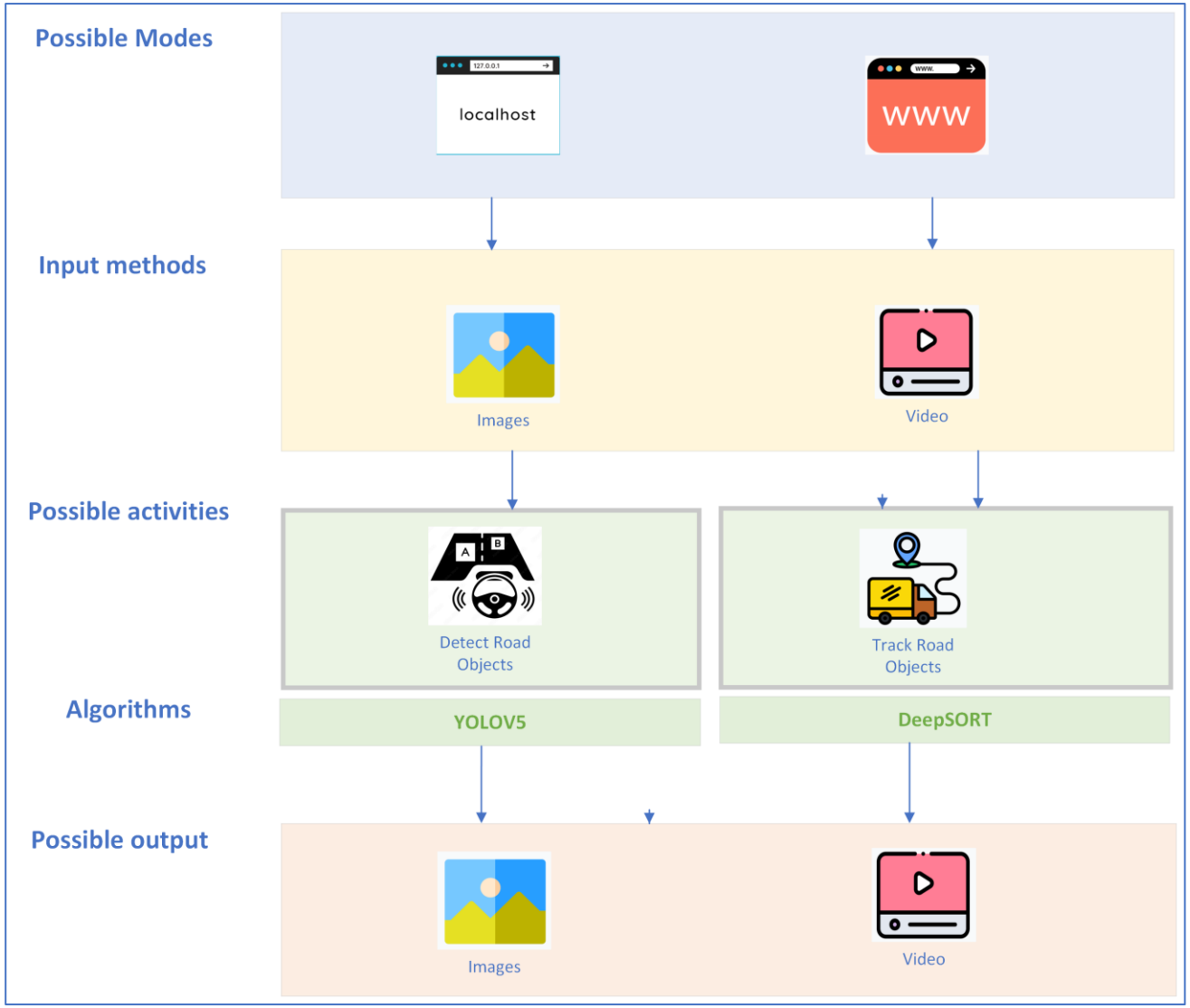
Build an application to detect multiple objects, tag and tracking in a video.

Methodology and design considerations



Methodology and design considerations

Project Scope



Outcome in stages

Understand the data

Learn about BDD100K dataset and download the images from website. Load these images with labels in Google Drive for further processing

Identify the object detection algorithm

Understand the repository with detailed analysis on the way algorithm works and its efficiency. We have chosen YOLOv5 for this purpose and understood steps for download its repository

Identify the tracking algorithm

We have chosen Deep SORT algorithm as our tracking algorithm as it has greater efficiency and support community.

Process images & videos for object detection

Train , validate and test the BDD100K images & videos with YOLOv5 algorithm. Obtain the final model weights generated from the epoch which can be the base for object detection.

Process tracking algorithm for videos

Test the BDD100K video samples downloaded form website to understand the tracking efficiency after downloading their repo and run the required files. Customize the solution to import essential packages / libraries.

Build the streamlit application and deploy the solution

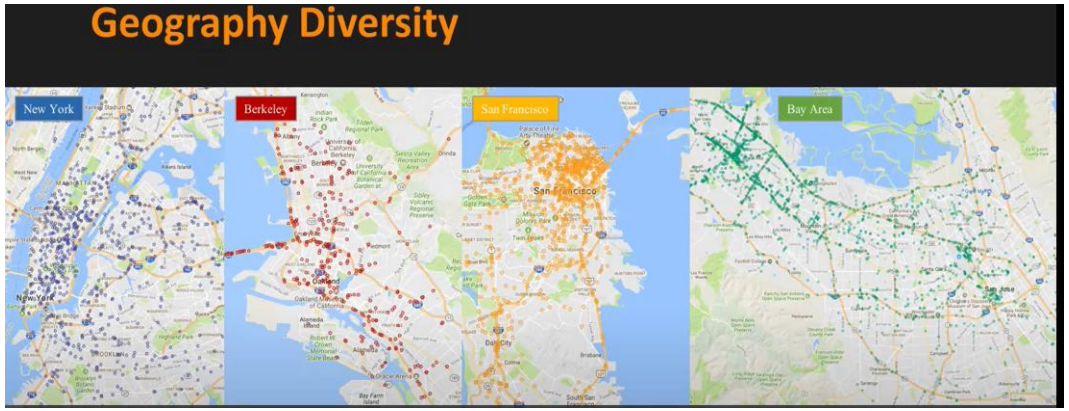
Develop the streamlit application to use the model weights for object detection and tracking. Deploy the entire solution to Streamlit Cloud after pushing entire solution to GIT repository.

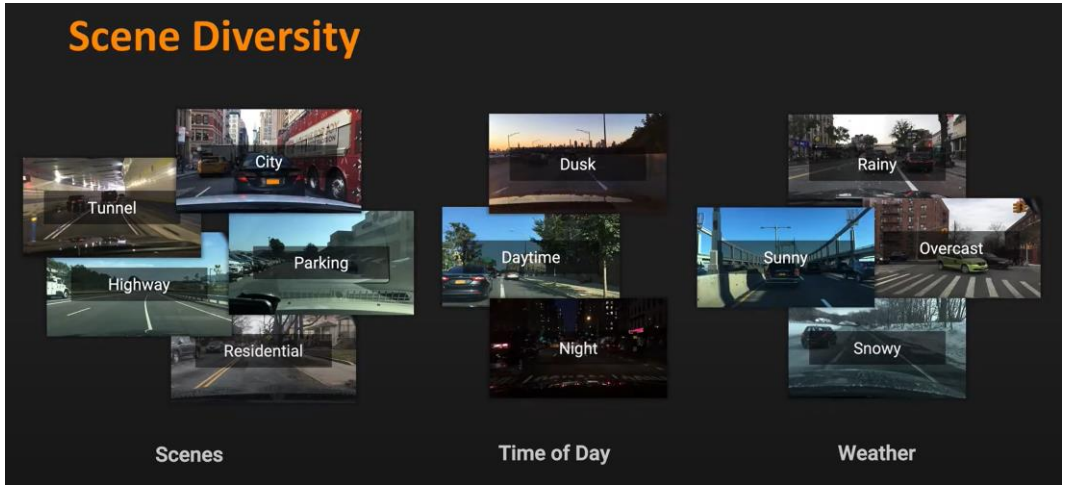
Outcome in stages

Stage 1 – Understand the data

BDD100K Dataset

- Consists of 100,000 videos.
  Each video is about 40 seconds long, 720p, and 30 fps
- Geographically diverse
- Scene Diversity : Scene = City , Tunnel , Highway , Parking , Residential
- Time of the day = Dusk , Daytime , night
- Weather = Rainy , Sunny , Overcast , Snowy
- There are >10 objects per image with pixel annotation and tracking





Outcome in stages

Stage 1 – Understand the data

Original Image

BDD100k dataset bounding box JSON format

dd100k_labe...ges_val.json*

```
[
    {
        "name": "000d4f89-3bcbe37a.jpg",
        "attributes": {
            "weather": "overcast",
            "scene": "city street",
            "timeofday": "daytime"
        },
        "timestamp": 10000,
        "labels": [
            {
                "category": "traffic sign",
                "attributes": {
                    "occluded": false,
                    "truncated": false,
                    "trafficLightColor": "none"
                },
                "manualShape": true,
                "manualAttributes": true,
                "box2d": {
                    "x1": 1000.698742,
                    "y1": 281.992415,
                    "x2": 1040.626872,
                    "y2": 326.91156
                },
                "id": 0
            },
            {
                "category": "traffic sign",
                "attributes": {
```

Image name

One object complete details in an image

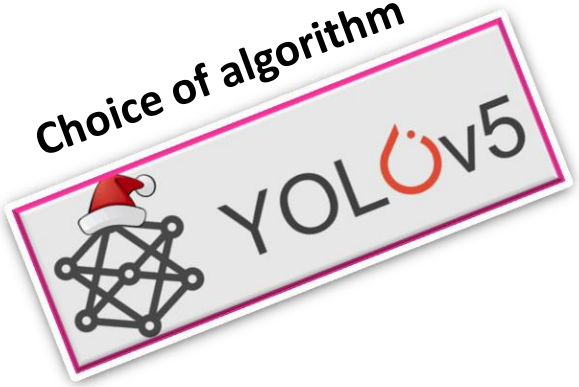
Object class

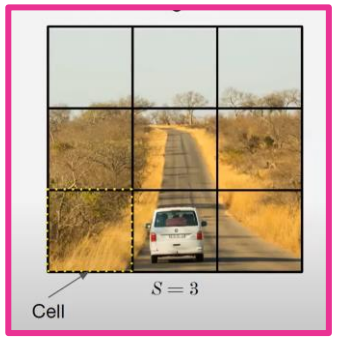
Bounding box dimensions

Outcome in stages

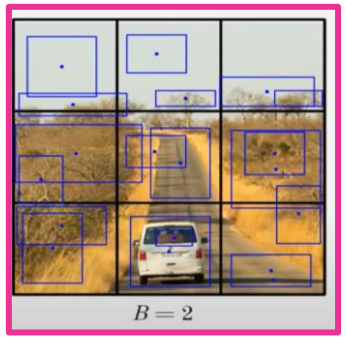
Stage 2 – Identify the object detection algorithm

Choice of algorithm

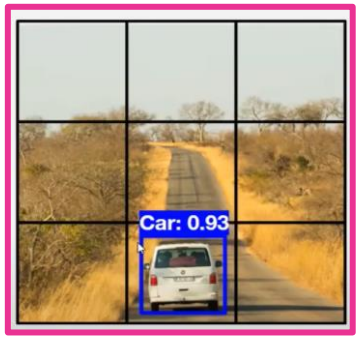




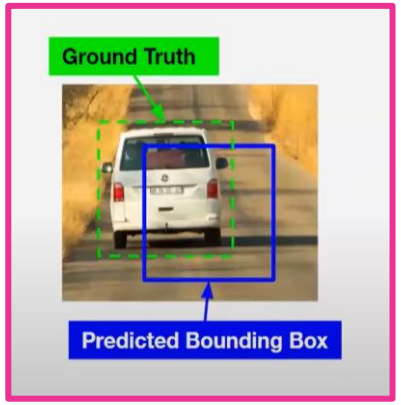
Divide the image to cells



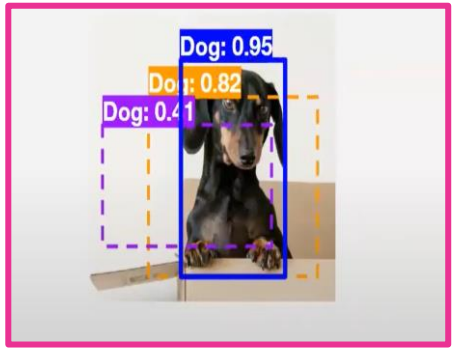
Predicting bounding boxes within each cell



Selection of bounding box which is above confidence threshold



Intersection over Union



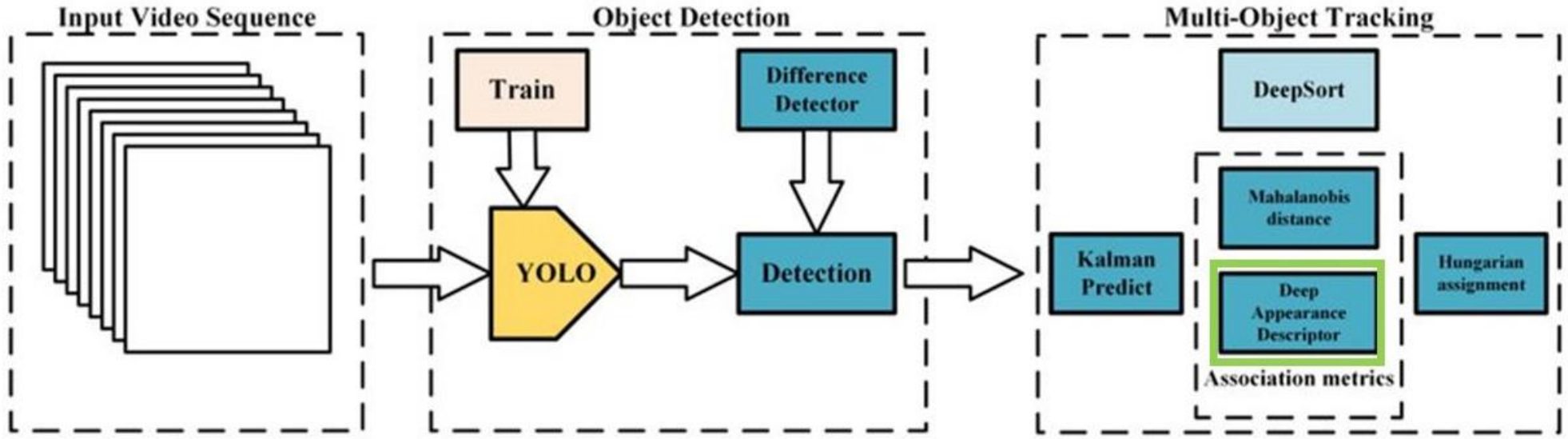
Non max suppression

Outcome in stages

Stage 3 – Identify the tracking algorithm

Choice of algorithm

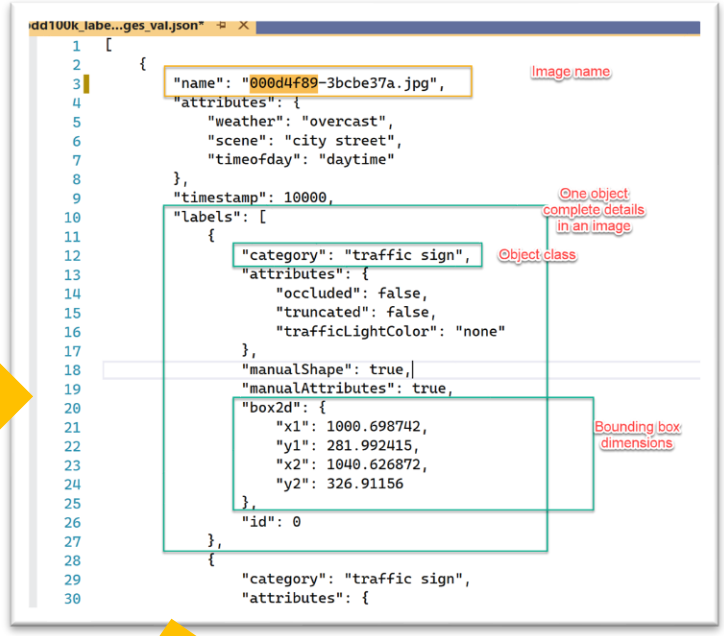
DeepSORT



Outcome in stages

Stage 4 – Process images and videos for object detection

Original Image




```
[
    {
        "name": "000d4f89-3bcbe37a.jpg",
        "attributes": {
            "weather": "overcast",
            "scene": "city street",
            "timeofday": "daytime"
        },
        "timestamp": 10000,
        "labels": [
            {
                "category": "traffic sign",
                "attributes": {
                    "occluded": false,
                    "truncated": false,
                    "trafficLightColor": "none"
                },
                "manualShape": true,
                "manualAttributes": true,
                "box2d": {
                    "x1": 1000.698742,
                    "y1": 281.992415,
                    "x2": 1040.626872,
                    "y2": 326.91156
                },
                "id": 0
            },
            {
                "category": "traffic sign",
                "attributes": {
```

BDD100k dataset bounding box JSON format

000d4f89-3bcbe37a - Notepad

File Edit Format View Help

bounding box x1, y1, x2, y2 normalized

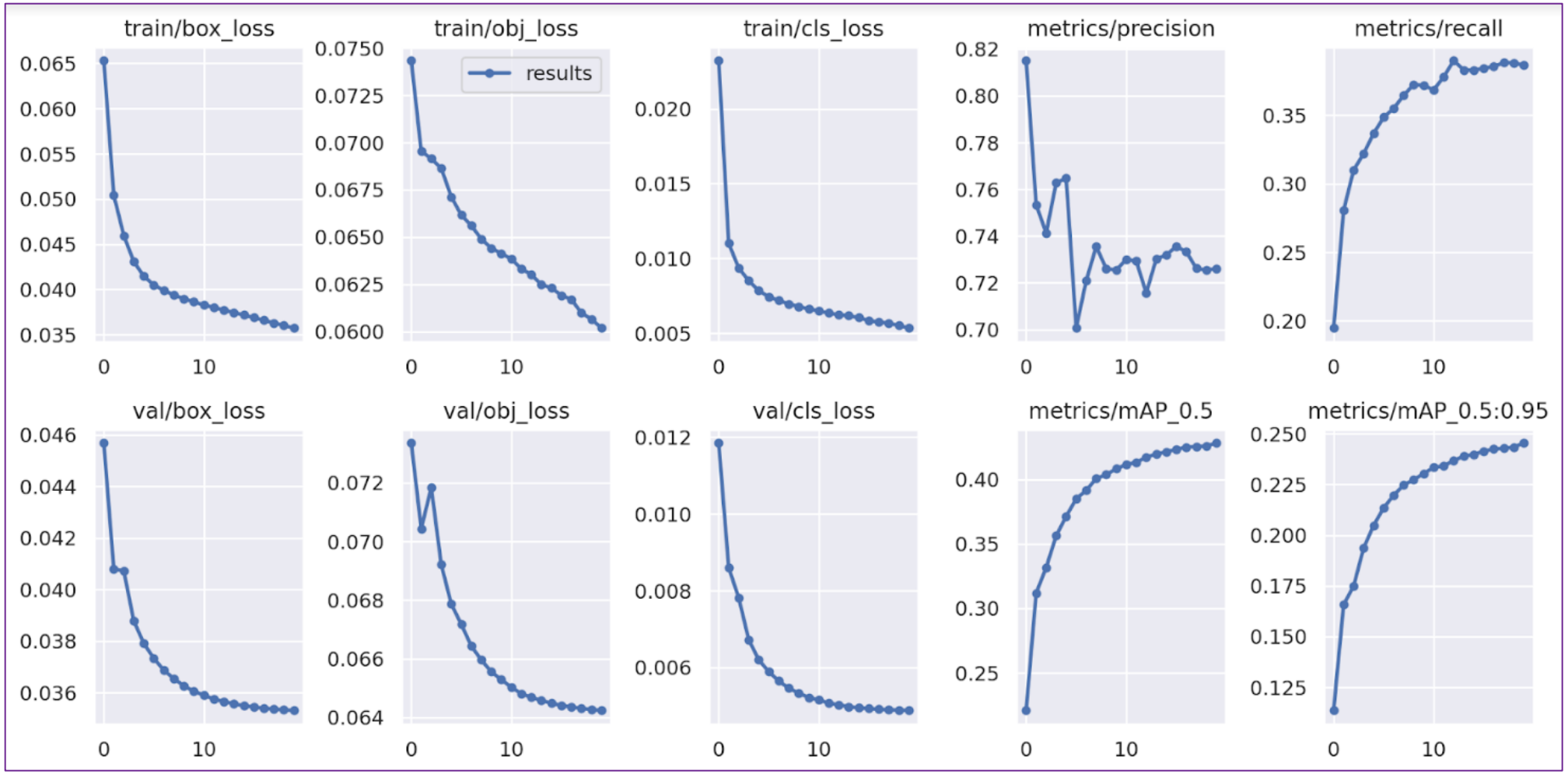
```
2 0.32104250507812504 0.49959173124999995 0.03926019609374998 0.0440816236111111
2 0.2588552611822023 0.4901953158935303 0.062384705760595335 0.05467654567594956
2 0.23528997304687502 0.49408152638888886 0.06612243515625002 0.06244896666666667
2 0.1738167703125 0.48760738138567267 0.1456759875 0.07154148499356758
2 0.058102510156249995 0.4720407166666667 0.03202805625 0.025714280555555608
2 0.015361844921875001 0.4674488784722222 0.030723689843750002 0.060612231944444435
2 0.46025830133564527 0.5259084653235868 0.05424690826620946 0.05630693620272922
```

Class Label

Converted to TXT file with required class and Bbox with mapping data and labels

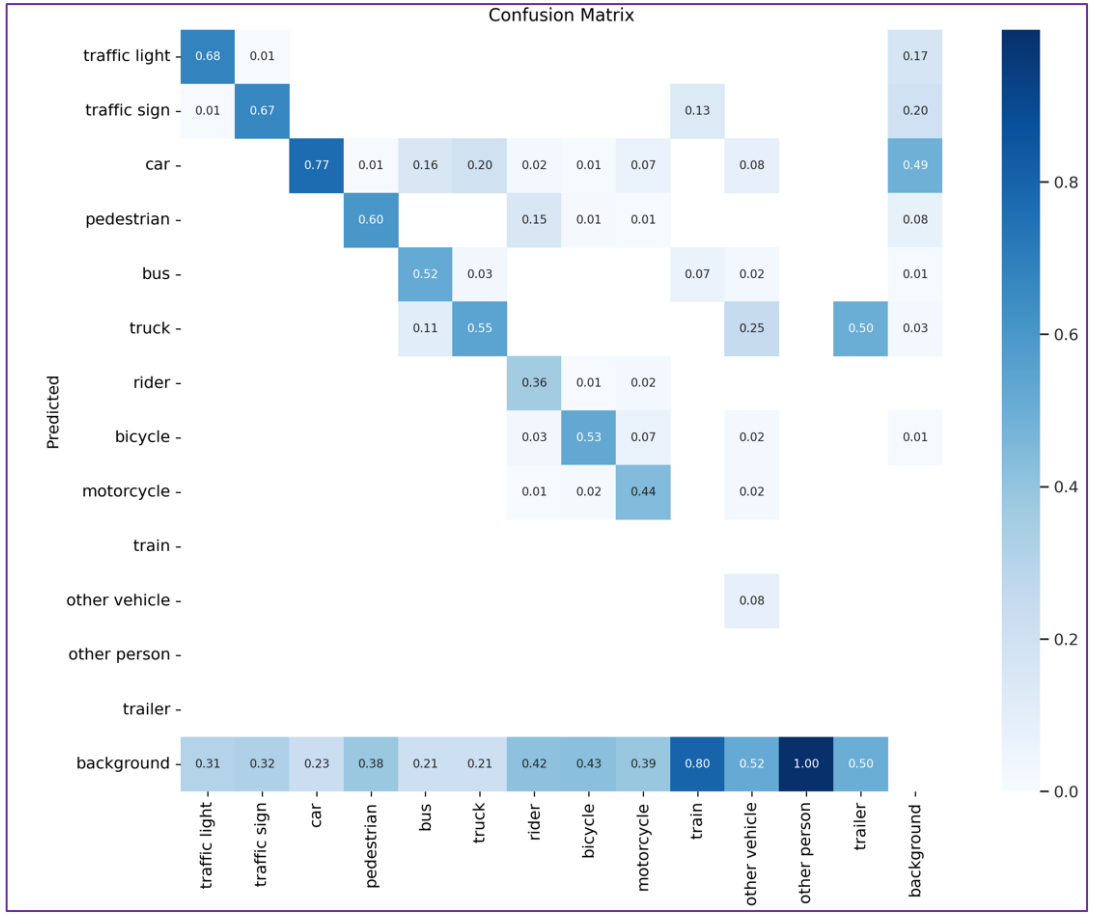
YOLO-v5

Training Metrics



YOLO-v5

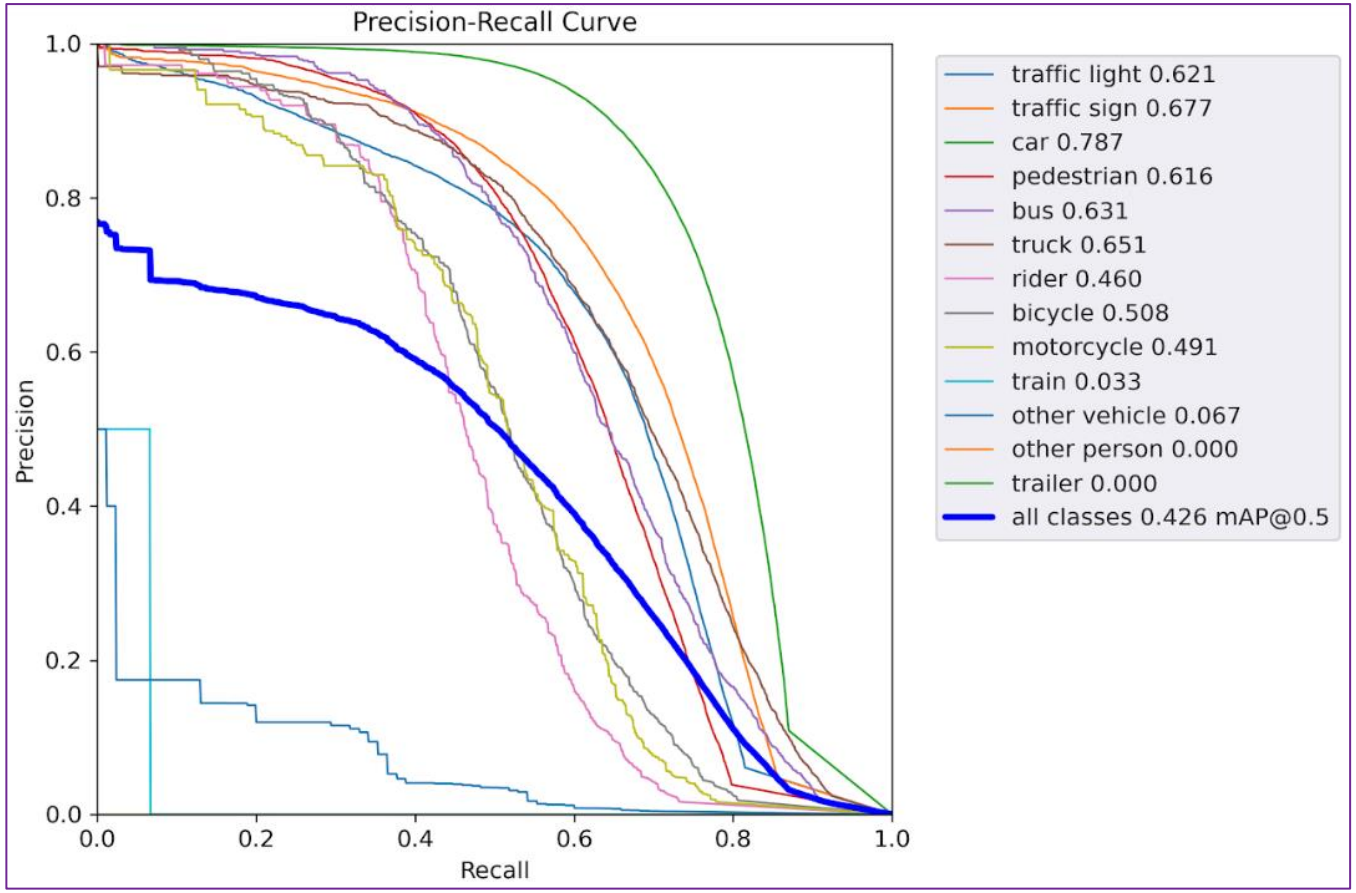
Confusion Matrix & validation results



Class	Images	Instances	P	R	mAP50	mAP50-95:
all	10000	186033	0.727	0.387	0.426	0.244
traffic light	10000	26884	0.676	0.601	0.621	0.234
traffic sign	10000	34724	0.728	0.623	0.677	0.36
car	10000	102837	0.796	0.722	0.787	0.505
pedestrian	10000	13425	0.719	0.553	0.616	0.307
bus	10000	1660	0.698	0.551	0.631	0.488
truck	10000	4243	0.691	0.594	0.651	0.477
rider	10000	658	0.733	0.388	0.46	0.231
bicycle	10000	1039	0.601	0.477	0.508	0.247
motorcycle	10000	460	0.667	0.446	0.491	0.244
train	10000	15	1	0	0.0326	0.0261
other vehicle	10000	85	0.143	0.0706	0.0672	0.0507
other person	10000	1	1	0	0	0
trailer	10000	2	1	0	0	0

YOLO-v5

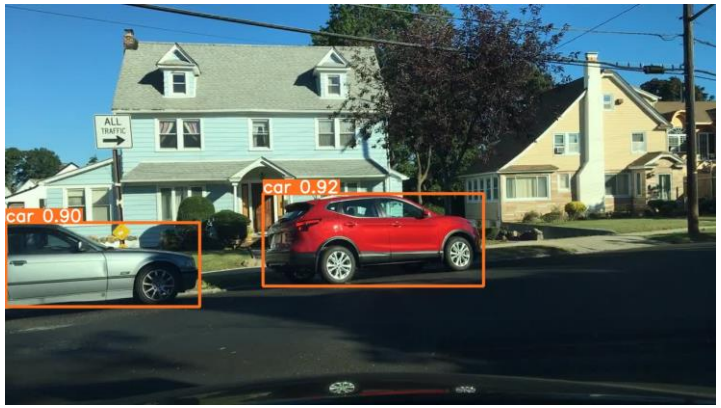
PR Curves

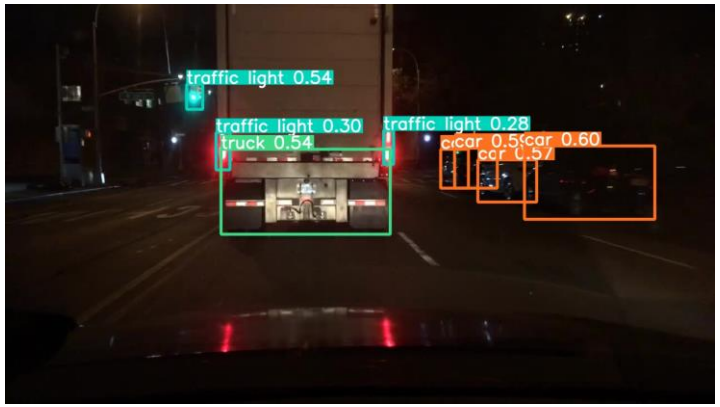


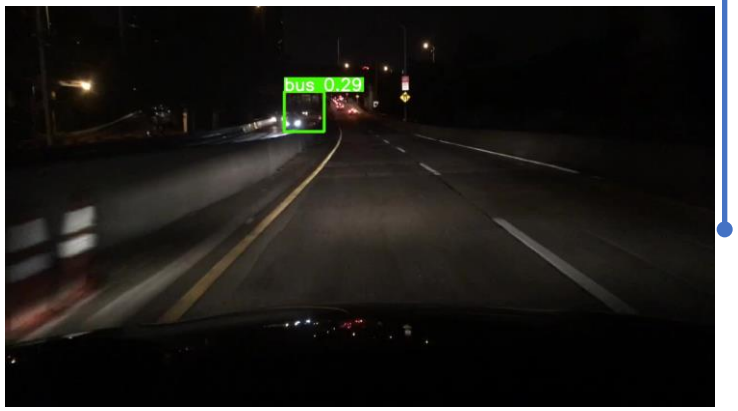
YOLO-v5

Trained model vs Pretrained model

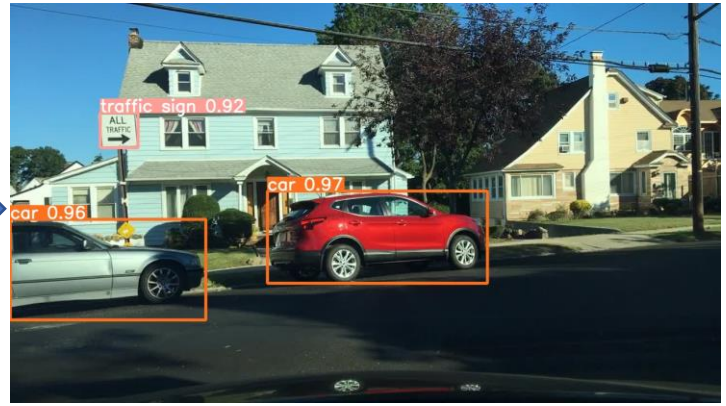
Pre-trained (COCO dataset)

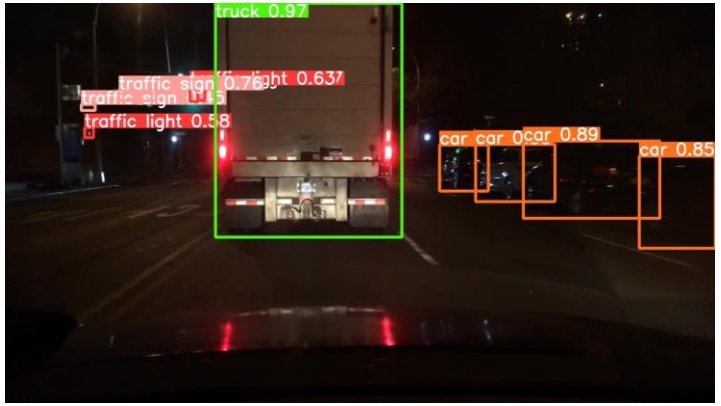






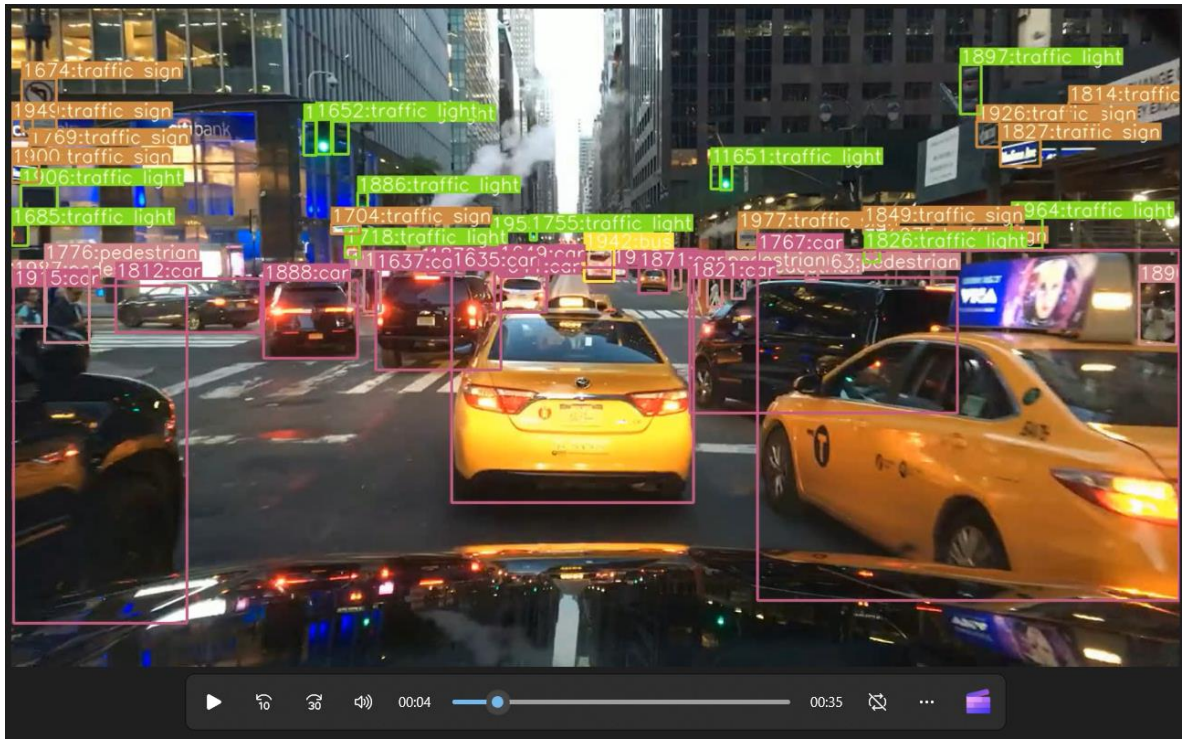
Trained model (BDD100K dataset)





Outcome in stages

Stage 5 – Processed videos for object tracking



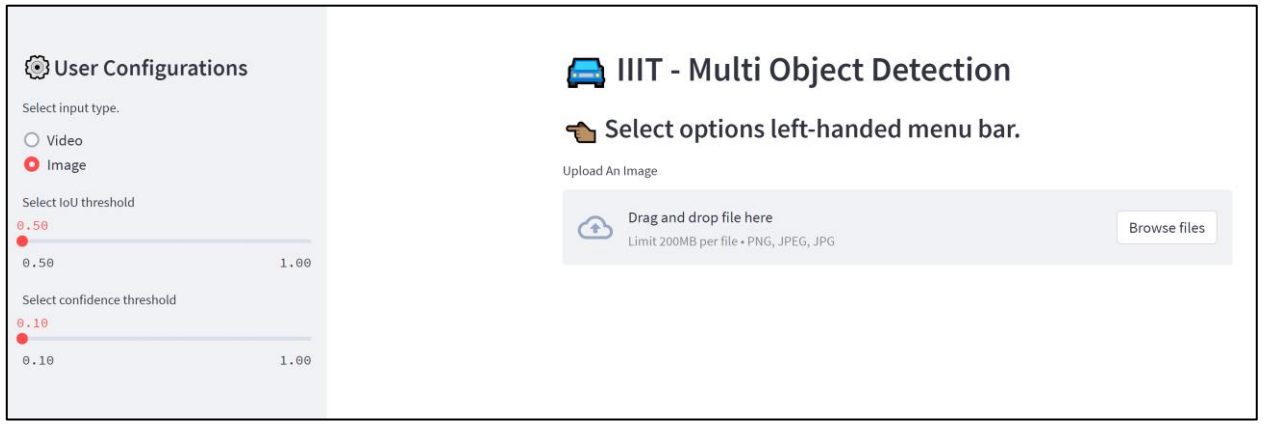
YOLO-v5

Detection + Deep SORT tracker output

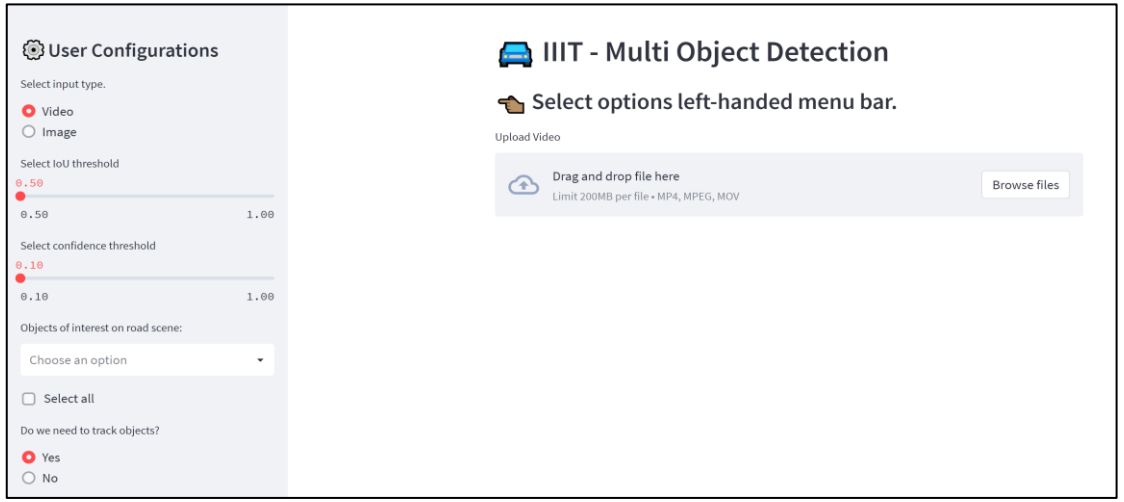
Tracked GIF via YOLOv5 + DeepSORT

Outcome in stages

Stage 6 – Develop streamlit application and deploy onto Streamlit Cloud



Application for Image detection



Application for Video detection and tracking

Application Demo

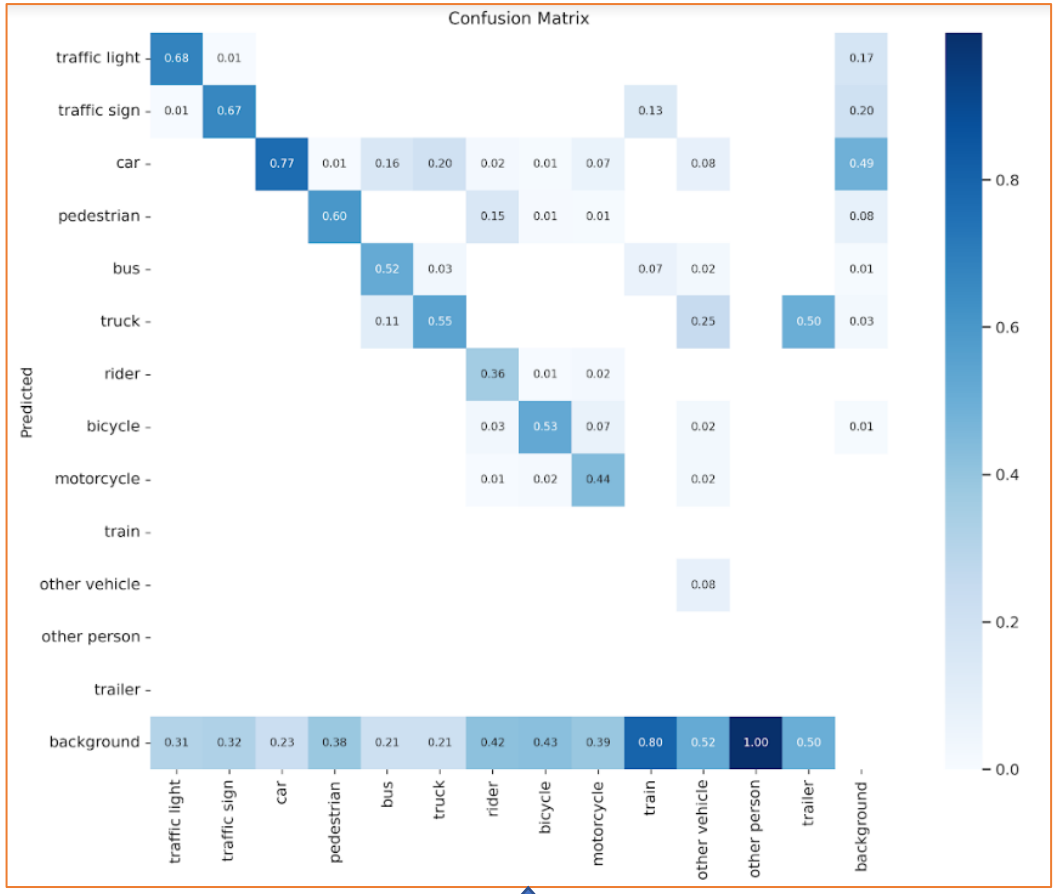
Challenges

1. Missing labels for some of the images of BDD100k data.
2. Downloading of BDD100k videos were time consuming.
3. Resource constraint – GPU causing disconnection of runtime in colab.
4. Understanding cloud constraints
5. Understanding installation of packages in streamlit cloud.
6. How to store the models so as to be loaded in the cloud environment.
7. Understanding Codec formats compatible with browsers.
8. Resource constraints while testing the web-UI locally.

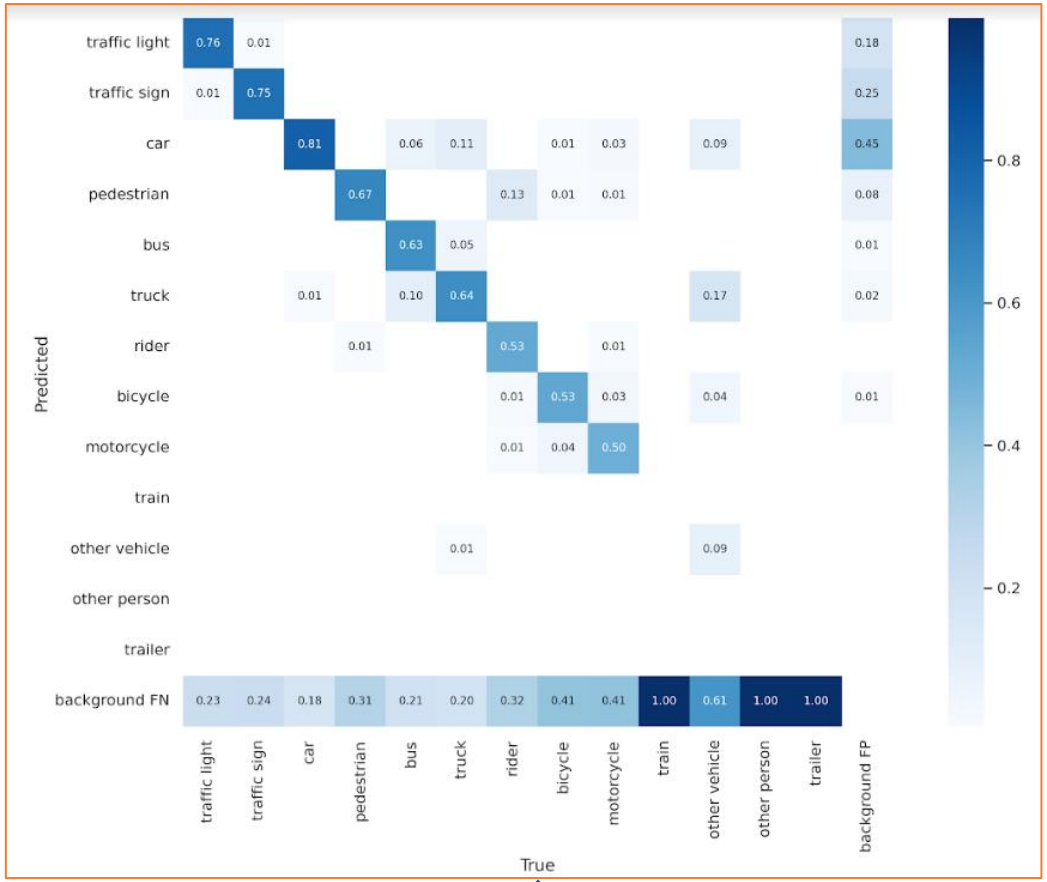
Going Beyond

1.YOLO-V7

YOLOv5 vs YOLOv7



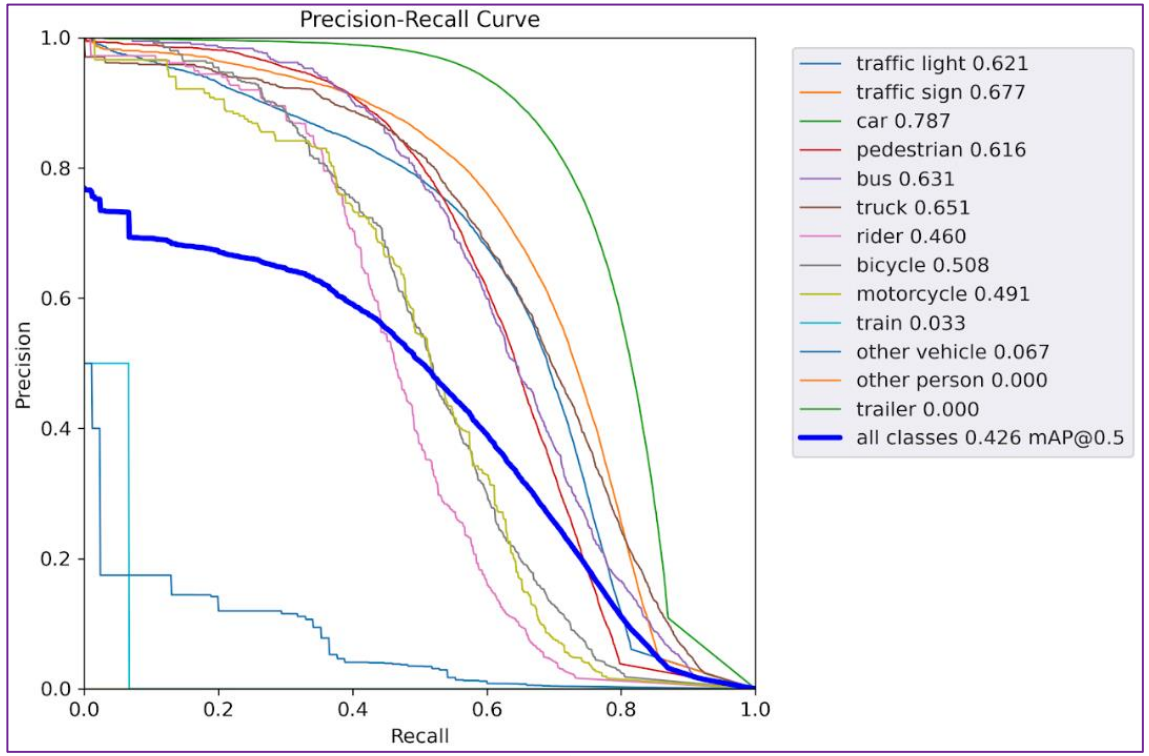
V5



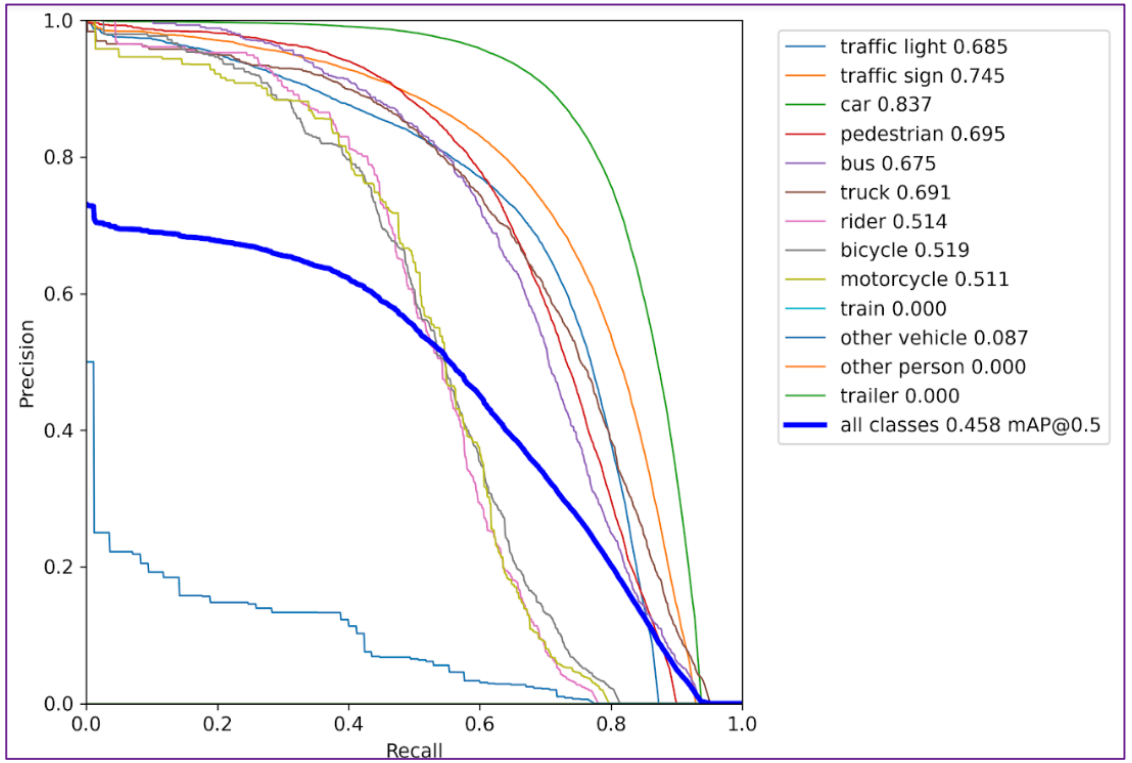
V7

1.YOLO-V7

YOLOv5 vs YOLOv7



V5



V7

1.YOLO-V7

Class	Images	Instances	P	R	mAP50	mAP50-95:
all	10000	186033	0.727	0.387	0.426	0.244
traffic light	10000	26884	0.676	0.601	0.621	0.234
traffic sign	10000	34724	0.728	0.623	0.677	0.36
car	10000	102837	0.796	0.722	0.787	0.505
pedestrian	10000	13425	0.719	0.553	0.616	0.307
bus	10000	1660	0.698	0.551	0.631	0.488
truck	10000	4243	0.691	0.594	0.651	0.477
rider	10000	658	0.733	0.388	0.46	0.231
bicycle	10000	1039	0.601	0.477	0.508	0.247
motorcycle	10000	460	0.667	0.446	0.491	0.244
train	10000	15	1	0	0.0326	0.0261
other vehicle	10000	85	0.143	0.0706	0.0672	0.0507
other person	10000	1	1	0	0	0
trailer	10000	2	1	0	0	0

V5

Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95:
all	10000	186033	0.742	0.422	0.458	0.258
traffic light	10000	26884	0.708	0.668	0.685	0.258
traffic sign	10000	34724	0.727	0.701	0.745	0.396
car	10000	102837	0.833	0.76	0.837	0.522
pedestrian	10000	13425	0.756	0.617	0.695	0.344
bus	10000	1660	0.752	0.586	0.675	0.517
truck	10000	4243	0.704	0.63	0.691	0.505
rider	10000	658	0.611	0.491	0.514	0.257
bicycle	10000	1039	0.611	0.497	0.519	0.249
motorcycle	10000	460	0.73	0.457	0.511	0.251
train	10000	15	1	0	0	0
other vehicle	10000	85	0.218	0.0824	0.0873	0.0519
other person	10000	1	1	0	0	0
trailer	10000	2	1	0	0	0

V7

1.YOLO-V7

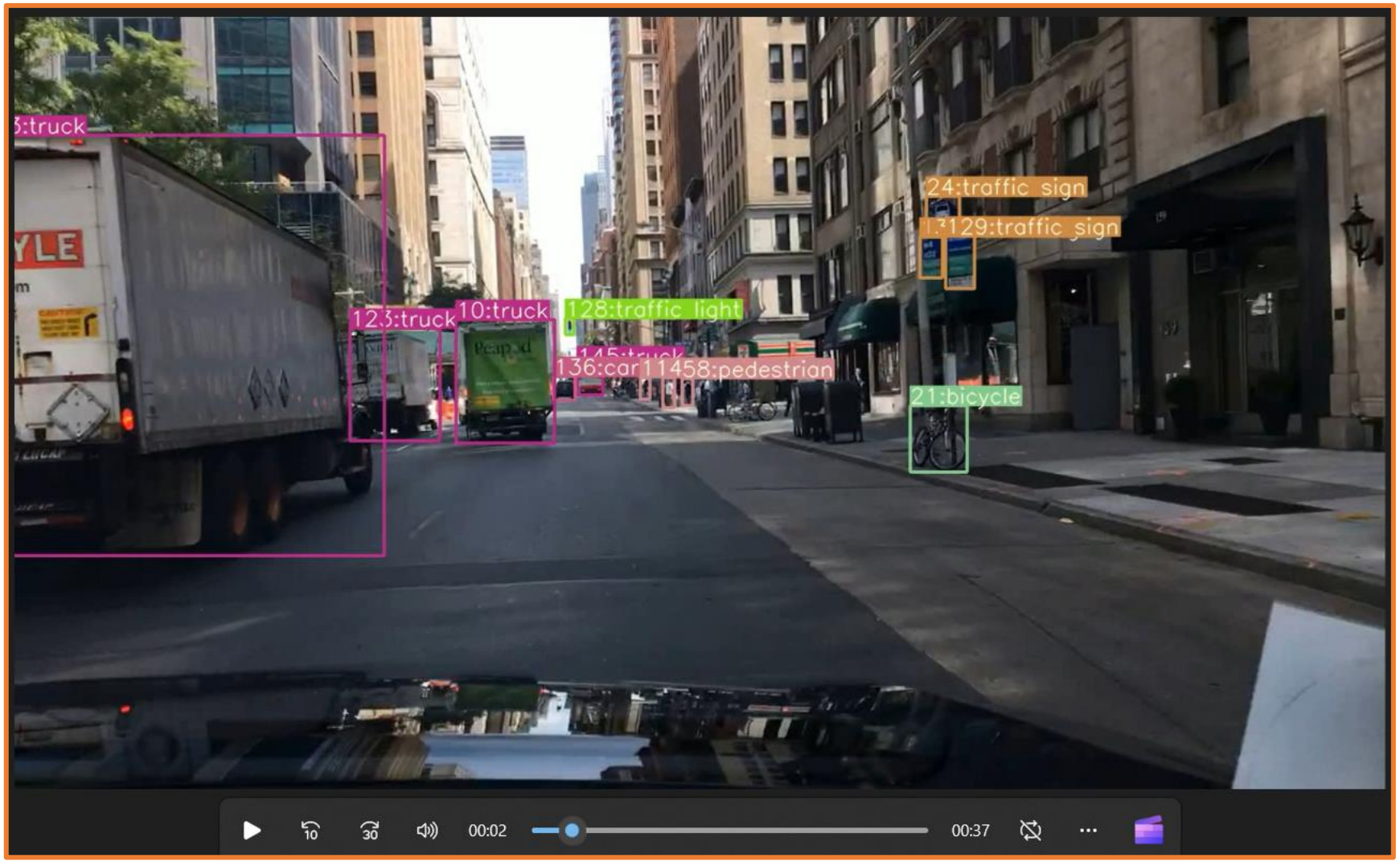
YOLOv5 vs YOLOv7

V5

V7

2.YOLO-V7 – Detection + Tracker

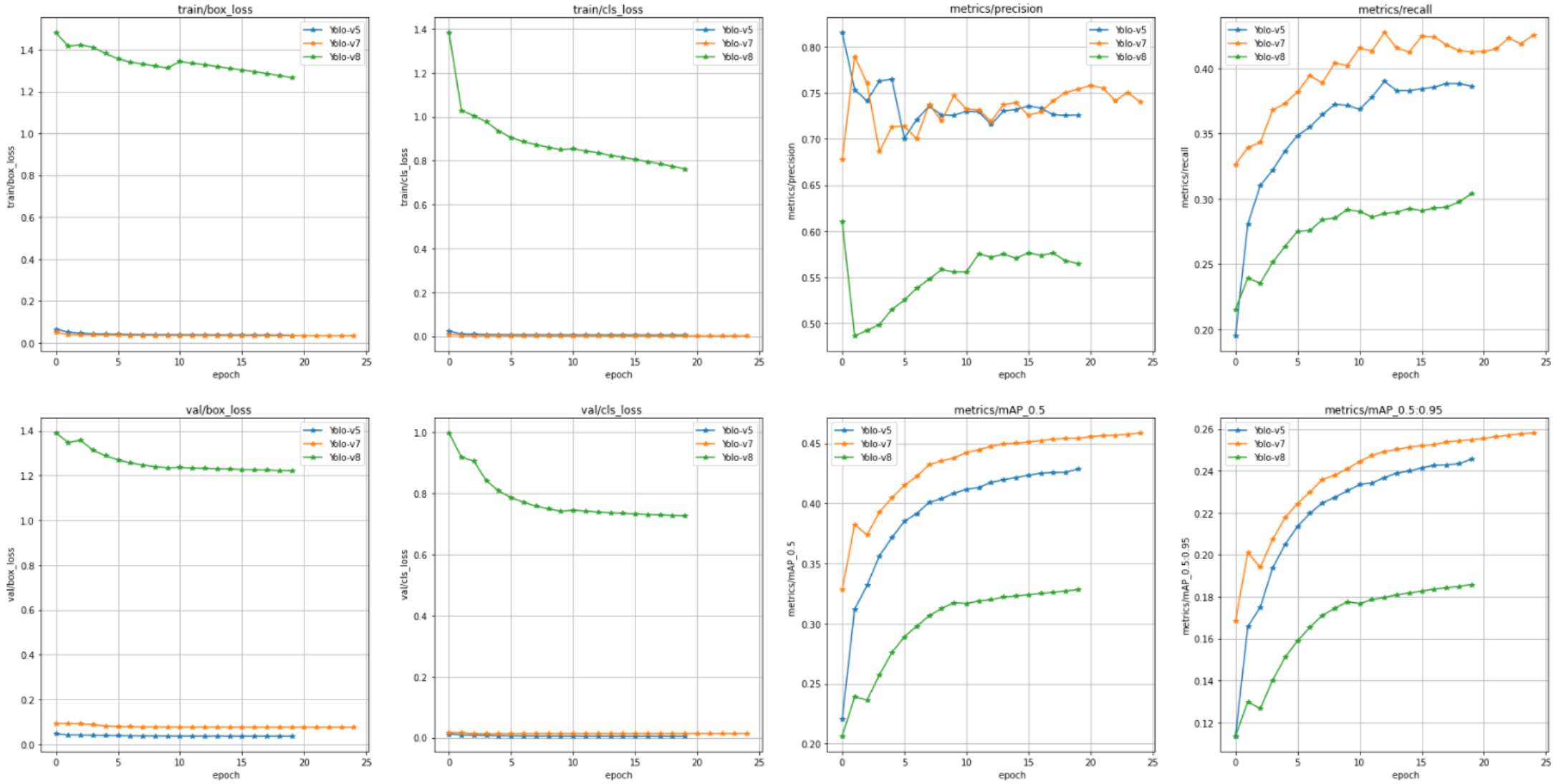
Results:



3.YOLO-V8

YOLOv5 vs YOLOv7 vs YOLOv8

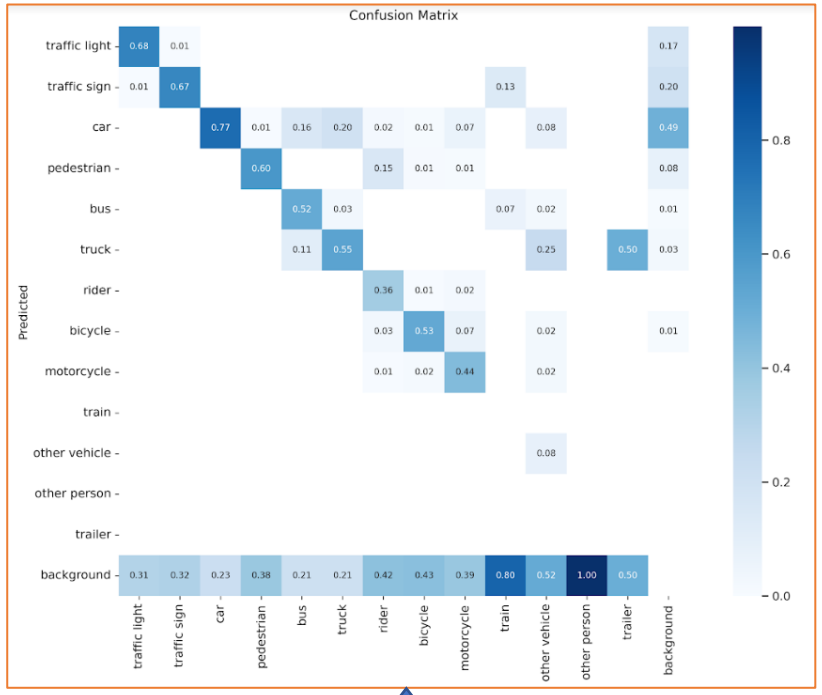
(Trained v8 nano model architecture with all BDD100k data for 20 epochs)



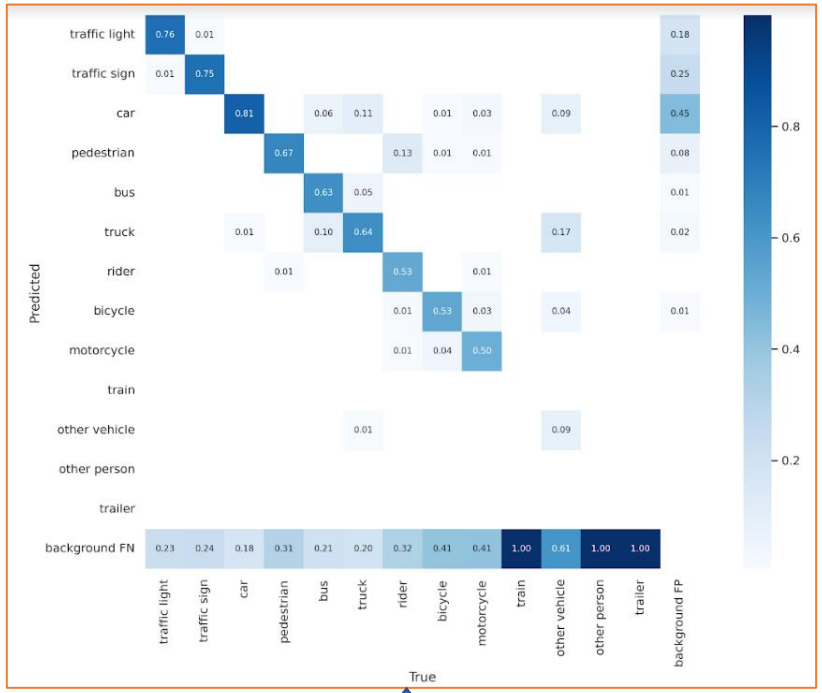
3.YOLO-V8

YOLOv5 vs YOLOv7 vs YOLOv8

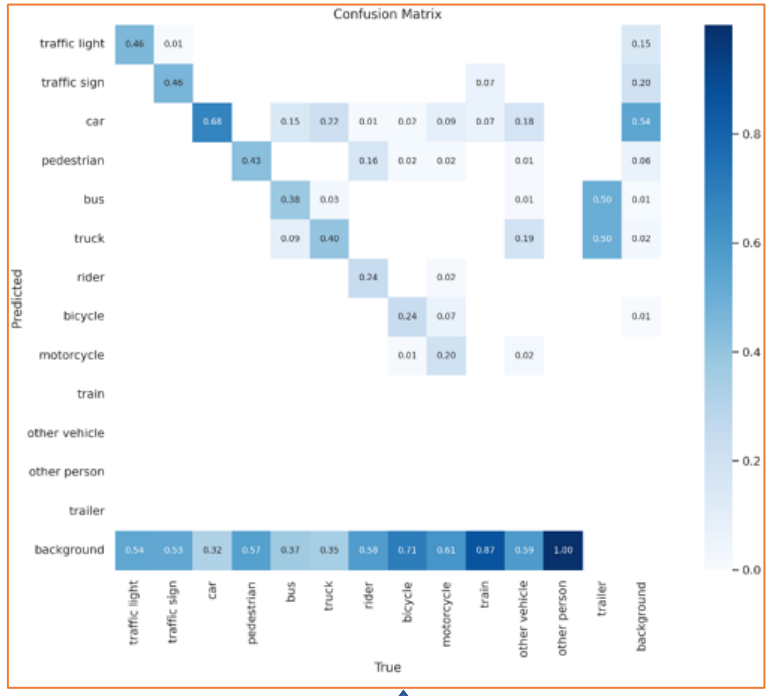
Confusion Matrix



V5



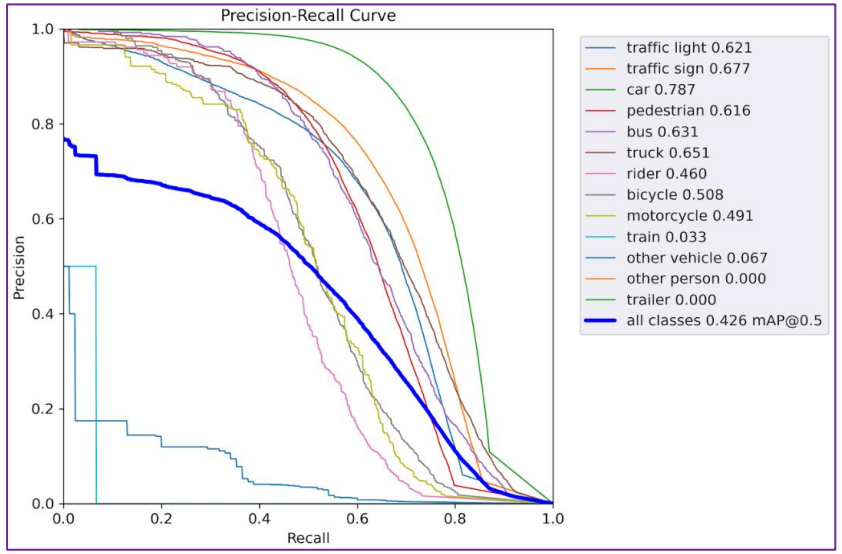
V7



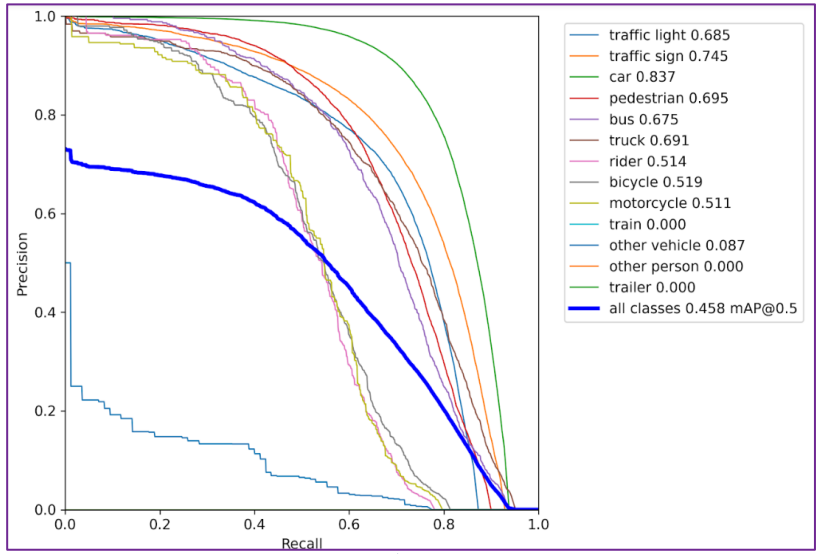
V8

3.YOLO-V8

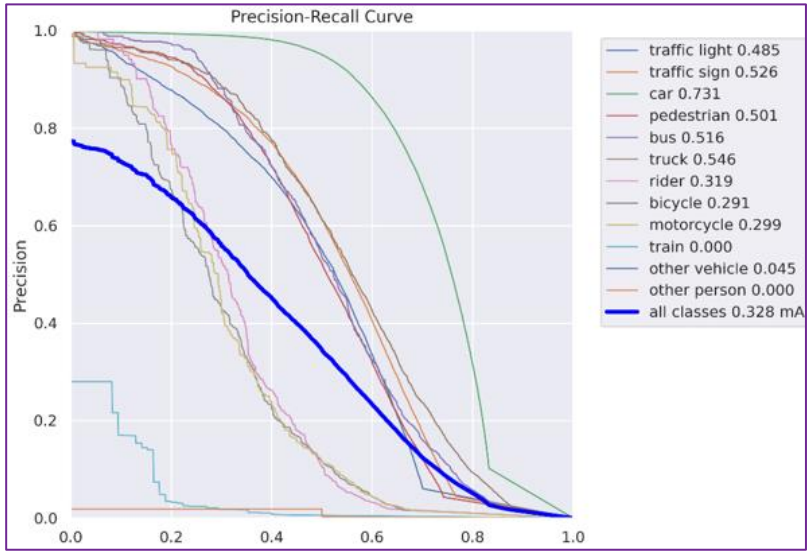
YOLOv5 vs YOLOv7 vs YOLOv8



V5



V7



V8

3.YOLO-V8

YOLOv5 vs YOLOv7 vs YOLOv8

Class	Images	Instances	P	R	mAP50	mAP50-95:
all	10000	186033	0.727	0.387	0.426	0.244
traffic light	10000	26884	0.676	0.601	0.621	0.234
traffic sign	10000	34724	0.728	0.623	0.677	0.36
car	10000	102837	0.796	0.722	0.787	0.505
pedestrian	10000	13425	0.719	0.553	0.616	0.307
bus	10000	1660	0.698	0.551	0.631	0.488
truck	10000	4243	0.691	0.594	0.651	0.477
rider	10000	658	0.733	0.388	0.46	0.231
bicycle	10000	1039	0.601	0.477	0.508	0.247
motorcycle	10000	460	0.667	0.446	0.491	0.244
train	10000	15	1	0	0.0326	0.0261
other vehicle	10000	85	0.143	0.0706	0.0672	0.0507
other person	10000	1	1	0	0	0
trailer	10000	2	1	0	0	0

V5

Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95:
all	10000	186033	0.742	0.422	0.458	0.258
traffic light	10000	26884	0.708	0.668	0.685	0.258
traffic sign	10000	34724	0.727	0.701	0.745	0.396
car	10000	102837	0.833	0.76	0.837	0.522
pedestrian	10000	13425	0.756	0.617	0.695	0.344
bus	10000	1660	0.752	0.586	0.675	0.517
truck	10000	4243	0.704	0.63	0.691	0.505
rider	10000	658	0.611	0.491	0.514	0.257
bicycle	10000	1039	0.611	0.497	0.519	0.249
motorcycle	10000	460	0.73	0.457	0.511	0.251
train	10000	15	1	0	0	0
other vehicle	10000	85	0.218	0.0824	0.0873	0.0519
other person	10000	1	1	0	0	0
trailer	10000	2	1	0	0	0

V7

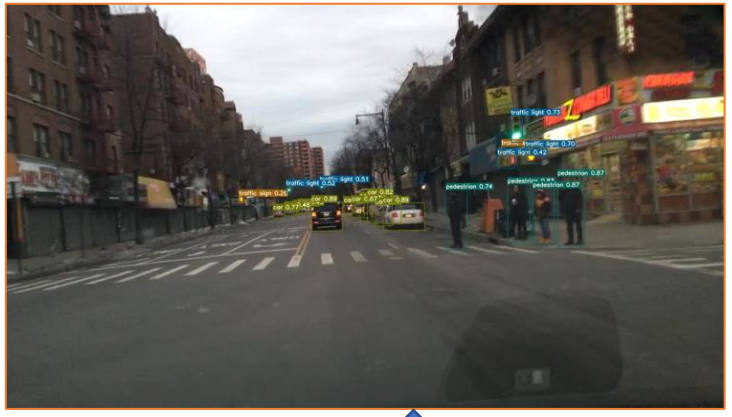
Class	Images	Instances	Box(P	R	mAP50	mAP50-95):
all	10000	186033	0.565	0.303	0.328	0.186
traffic light	10000	26884	0.597	0.473	0.486	0.178
traffic sign	10000	34724	0.633	0.49	0.526	0.274
car	10000	102837	0.696	0.694	0.732	0.466
pedestrian	10000	13425	0.598	0.462	0.502	0.239
bus	10000	1660	0.579	0.48	0.516	0.398
truck	10000	4243	0.621	0.496	0.546	0.395
rider	10000	658	0.52	0.299	0.317	0.155
bicycle	10000	1039	0.48	0.281	0.292	0.143
motorcycle	10000	460	0.545	0.257	0.299	0.139
train	10000	15	0	0	0	0
other vehicle	10000	85	0.0815	0.0118	0.0447	0.0264
other person	10000	1	1	0	0	0
trailer	10000	2	1	0	0.00988	0.00593

V8

3.YOLO-V8

YOLOv5 vs YOLOv7 vs YOLOv8

V5

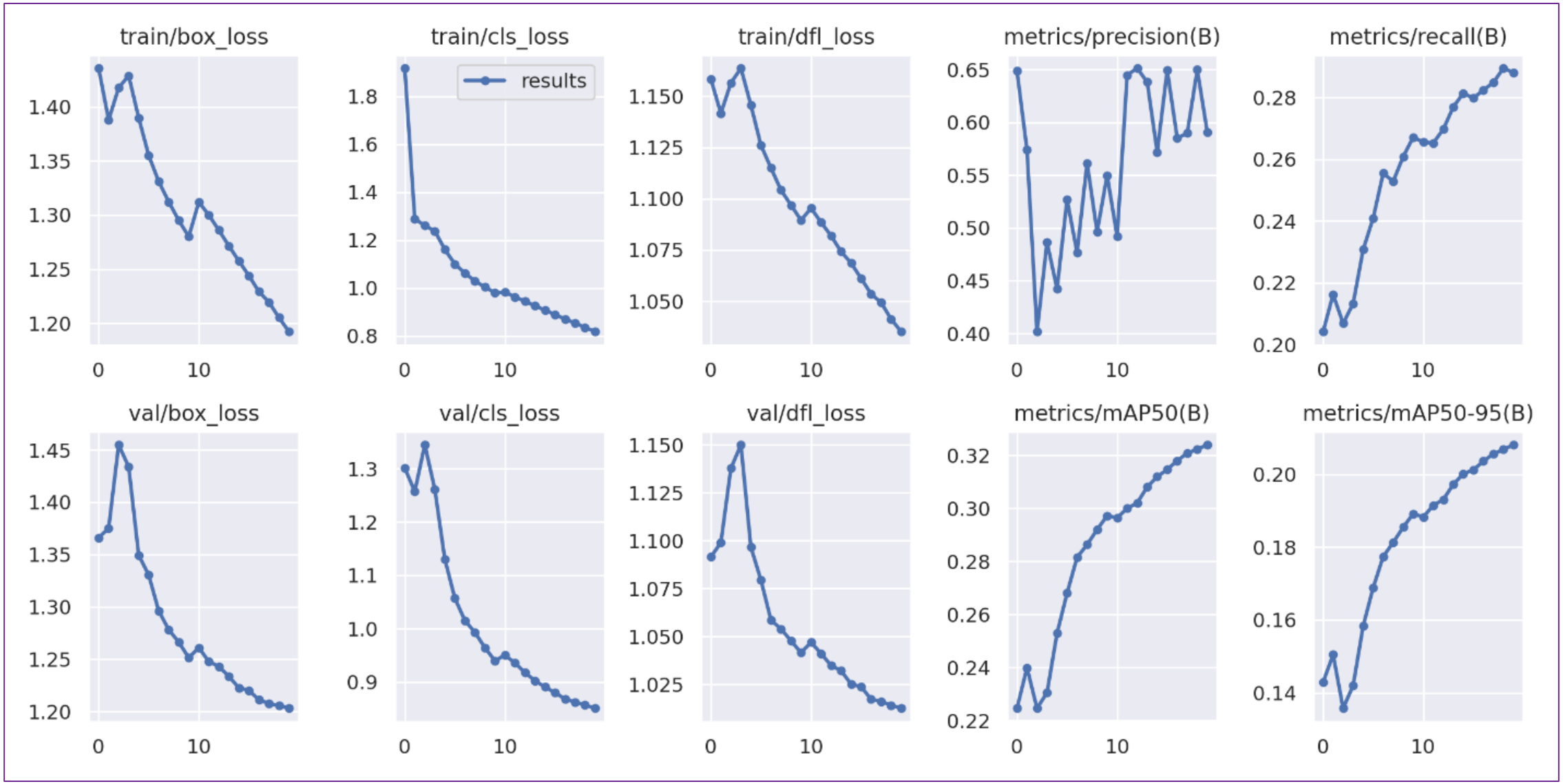


V7

V8

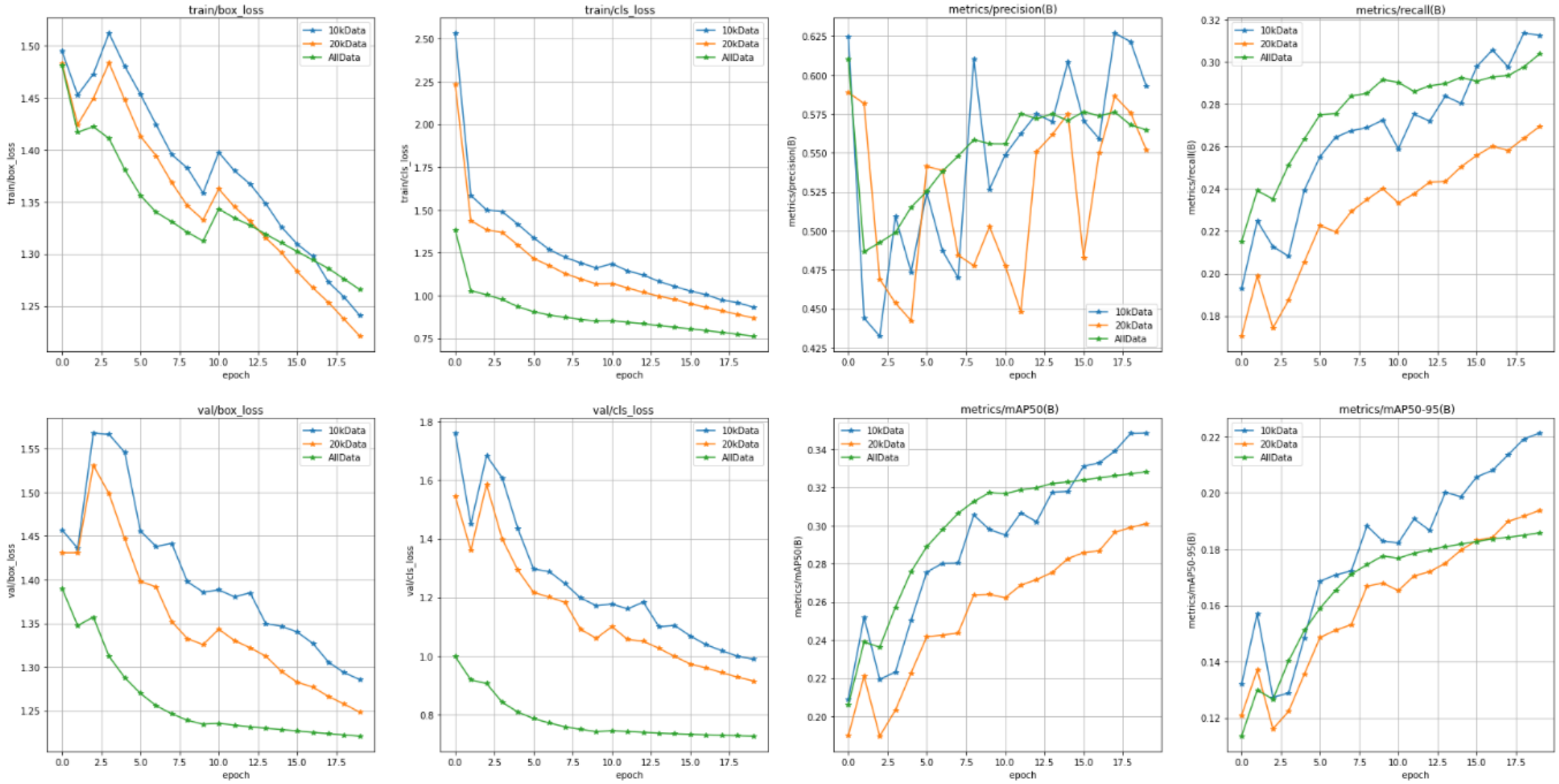
4.YOLO-V8 : IDD Dataset

Train Metrics



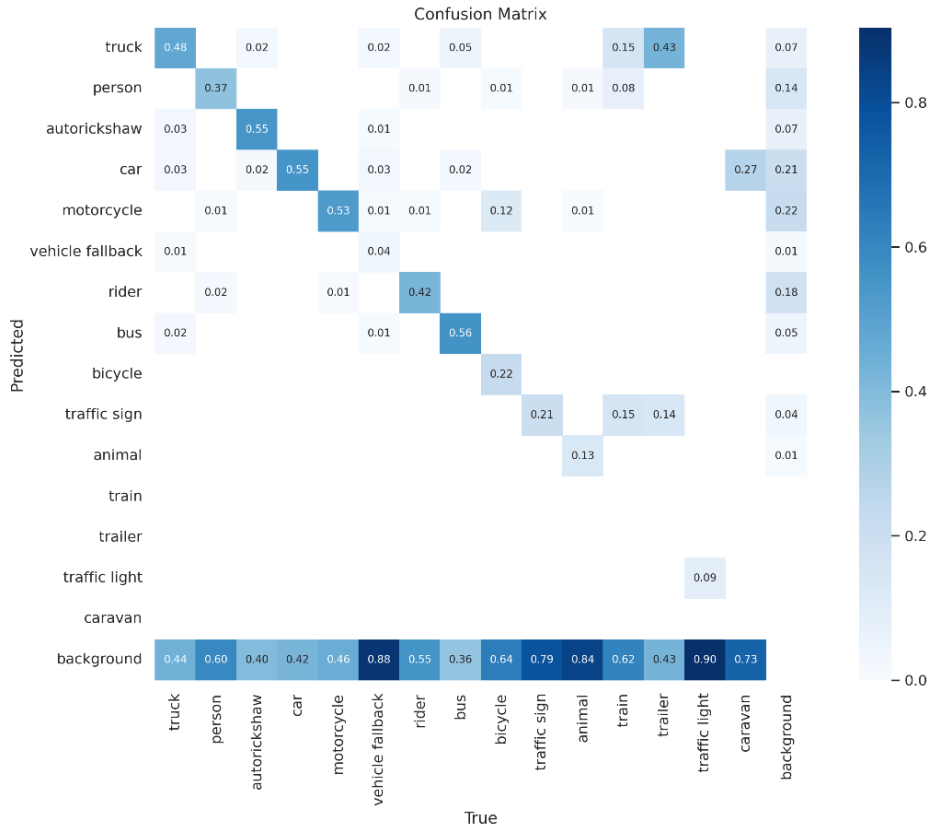
4.YOLO-V8 : IDD Dataset

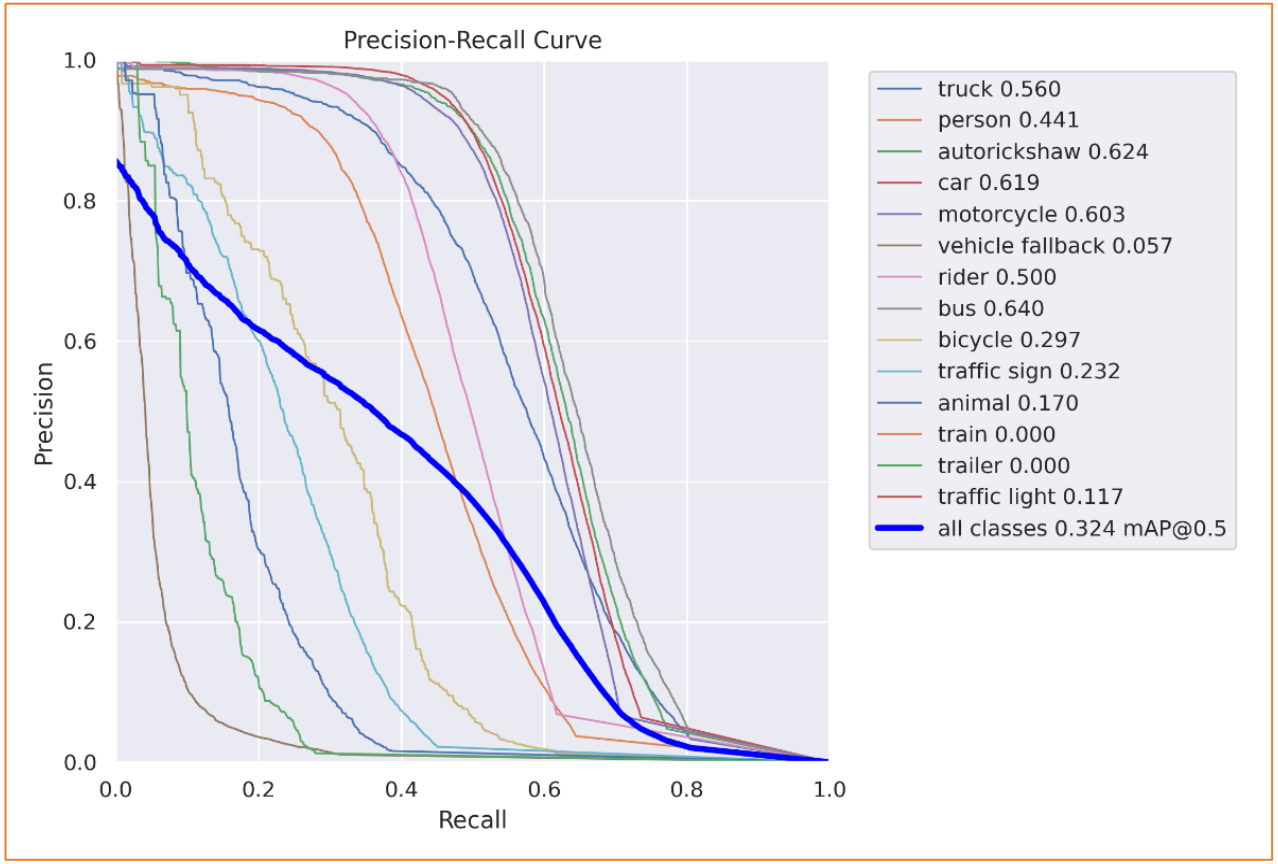
Train Metrics comparison against different Data sizes



4.YOLO-V8 : IDD Dataset

YOLOv8 – Trained on IDD dataset
[Confusion Matrix & P-R Curves]





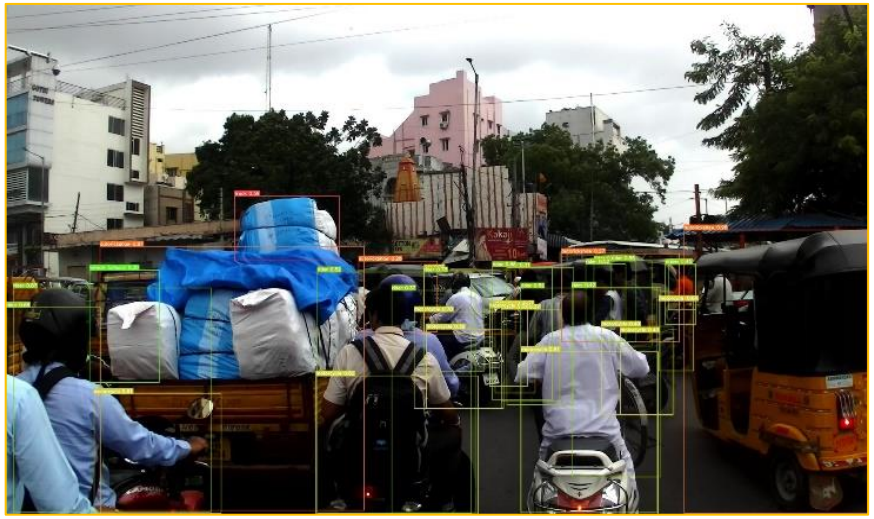
4.YOLO-V8 : IDD Dataset

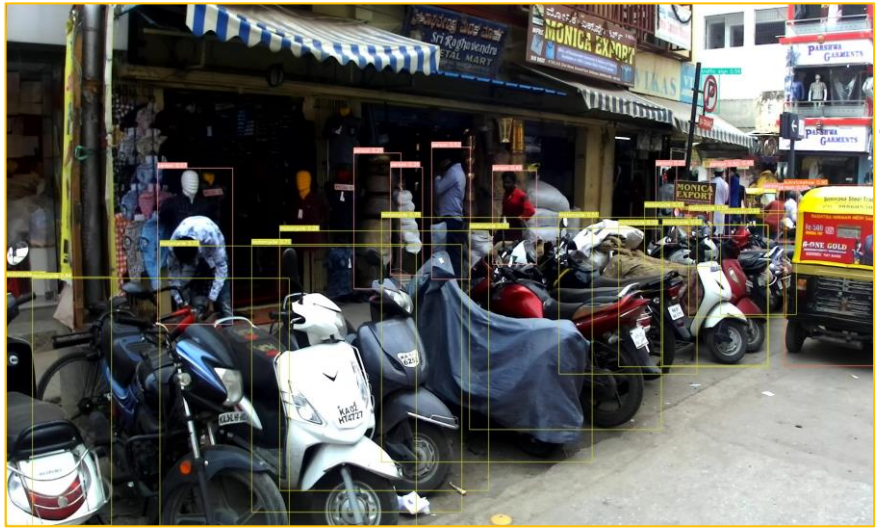
[Validation Metrics]

Class	Images	Instances	Box(P	R	mAP50	mAP50-95):
all	10224	126004	0.59	0.288	0.324	0.208
truck	10224	7075	0.673	0.508	0.561	0.4
person	10224	18070	0.692	0.382	0.442	0.244
autorickshaw	10224	7781	0.745	0.568	0.624	0.445
car	10224	24831	0.751	0.557	0.619	0.432
motorcycle	10224	25484	0.765	0.544	0.602	0.357
vehicle fallback	10224	6078	0.457	0.0439	0.0576	0.0313
rider	10224	24510	0.736	0.434	0.5	0.286
bus	10224	4910	0.728	0.589	0.64	0.491
bicycle	10224	569	0.616	0.254	0.297	0.175
traffic sign	10224	4287	0.57	0.213	0.233	0.121
animal	10224	1460	0.588	0.138	0.171	0.0831
train	10224	13	1	0	0	0
trailer	10224	7	0	0	0	0
traffic light	10224	918	0.533	0.0937	0.116	0.0579
caravan	10224	11	0	0	0	0

4.YOLO-V8 : IDD Dataset

[Object Detection – Images]

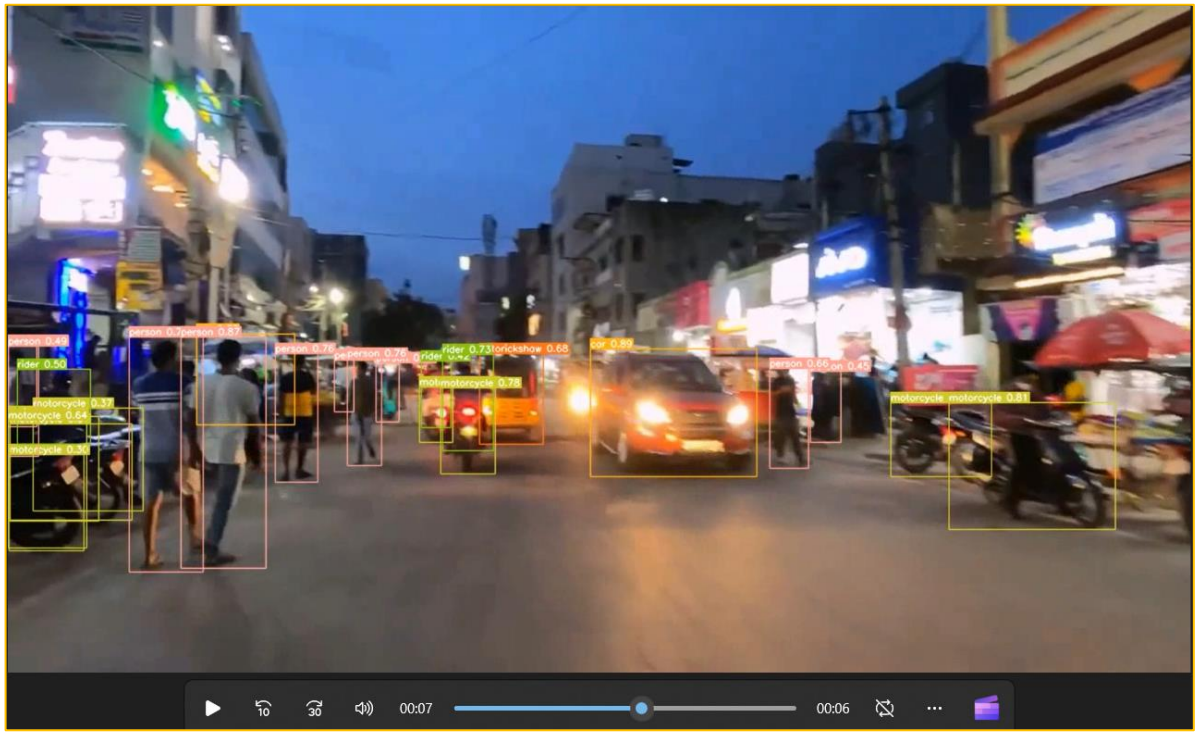




BDD100K dataset training

4.YOLO-V8 : IDD Dataset

[Object Detection – Videos]



5.YOLO-V8 : IDD Dataset – Semantic Segmentation

Preparing Labels

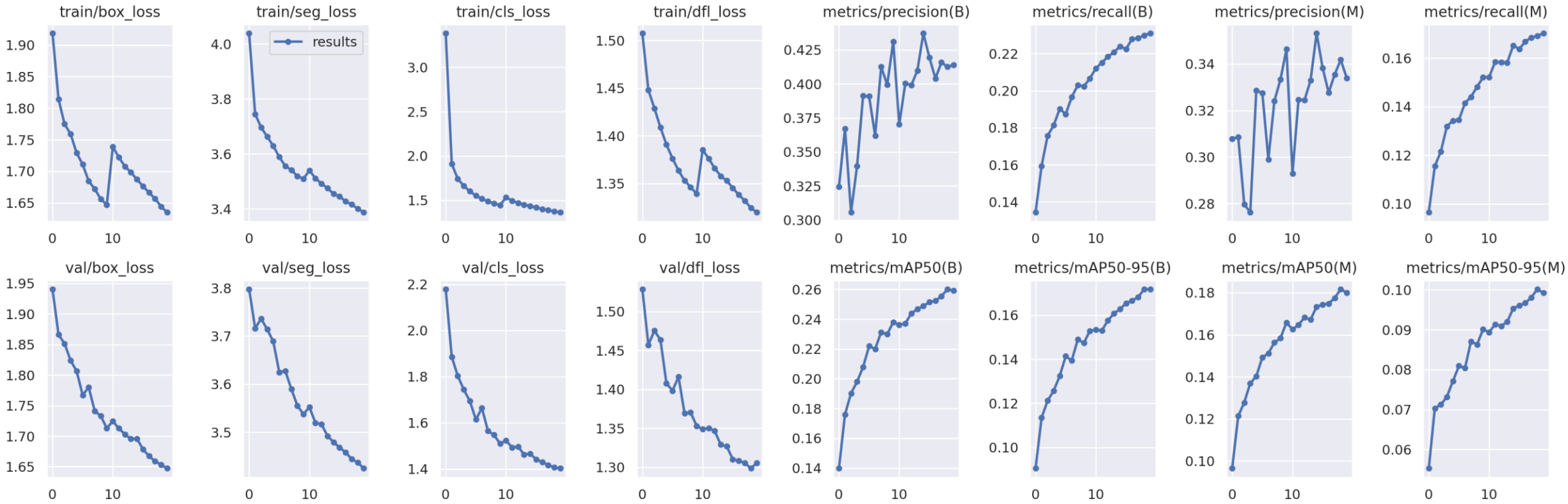

```
{
    "imgHeight": 1080,
    "imgWidth": 1920,
    "objects": [
        {
            "date": "14-Jun-2019 17:26:53",
            "deleted": 0,
            "draw": true,
            "id": 0,
            "label": "road",
            "polygon": [
                [
                    1094.3204419889503,
                    572.817679558011
                ],
                [
                    1372.375690607735,
                    553.7237569060774
                ],
                [
                    1919.0,
                    699.2307692307692
                ],
                [
                    1919.0,
                    708.4615384615385
                ],
```



```
0 0.5699585635359116 0.5303867403314917 0.714779005524862 0.512707182
1 0.6084944751381216 0.5414364640883979 0.5991712707182321 0.550276243
1 0.0 0.7585470085470085 0.055288461538461536 0.7318376068376068 0.108
1 0.6015625 0.5373931623931624 0.5835336538461537 0.5405982905982906 0
1 0.7160220994475138 0.5138121546961326 0.9994791666666667 0.593370165
2 0.2535911602209945 0.5558011049723757 0.2548342541436464 0.560220994
3 0.0 0.24640883977900555 0.0 0.0 0.9988259668508288 0.001104972375690
1 0.0037292817679558015 0.7624309392265194 0.04848066298342542 0.73149
4 0.6899171270718232 0.5314917127071824 0.6843232044198896 0.531491712
2 0.6116022099447513 0.5259668508287293 0.6140883977900553 0.529281767
2 0.15290055248618784 0.5878453038674034 0.1522790055248619 0.58563535
5 0.6681629834254145 0.4375690607734807 0.6681629834254145 0.432044198
6 0.619682320441989 0.4751381215469614 0.6221685082872929 0.4751381215
6 0.6631906077348066 0.5116022099447515 0.6607044198895029 0.511602209
7 0.5065607734806631 0.4243093922651934 0.5065607734806631 0.365745856
8 0.5363950276243094 0.36243093922651937 0.5525552486187846 0.36022099
6 0.5792817679558012 0.5226519337016575 0.5780386740331491 0.519337016
```

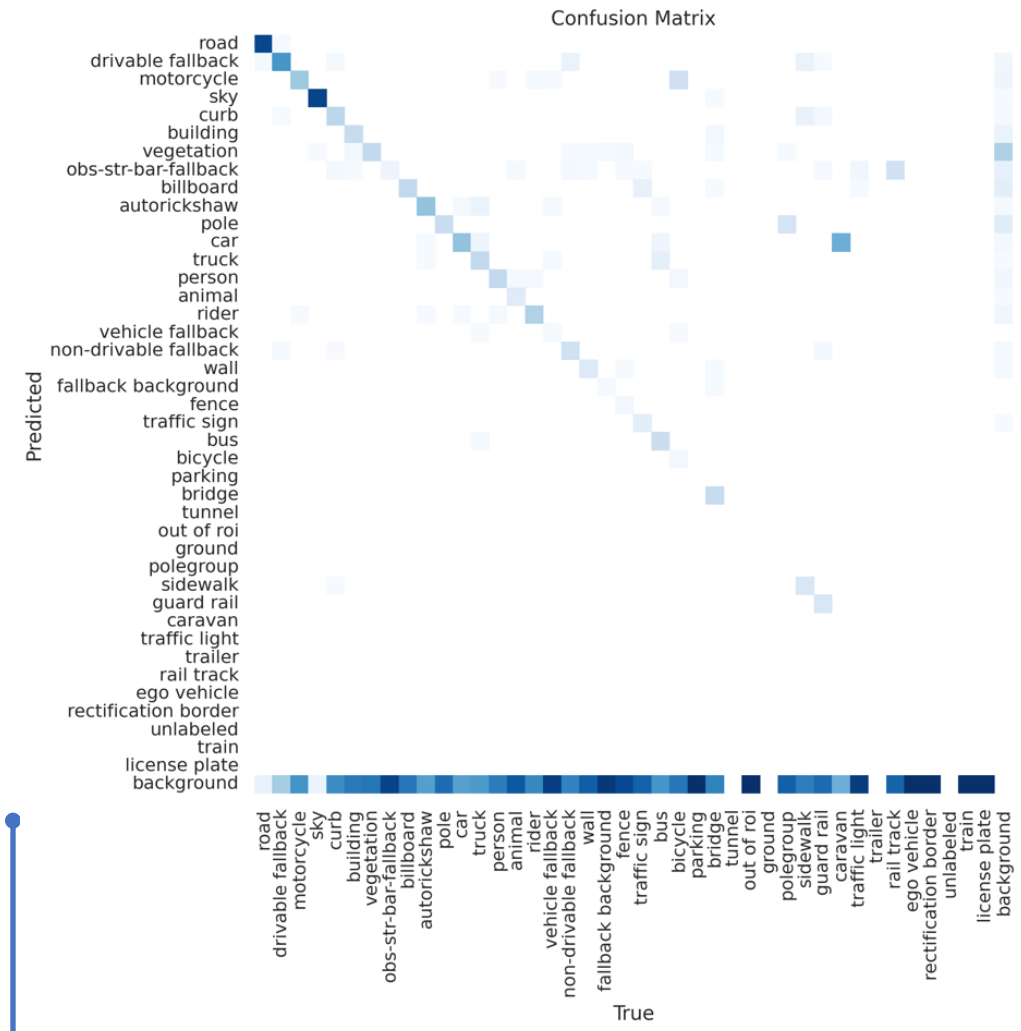
5.YOLO-V8 : IDD Dataset – Semantic Segmentation

[Train Metrics]

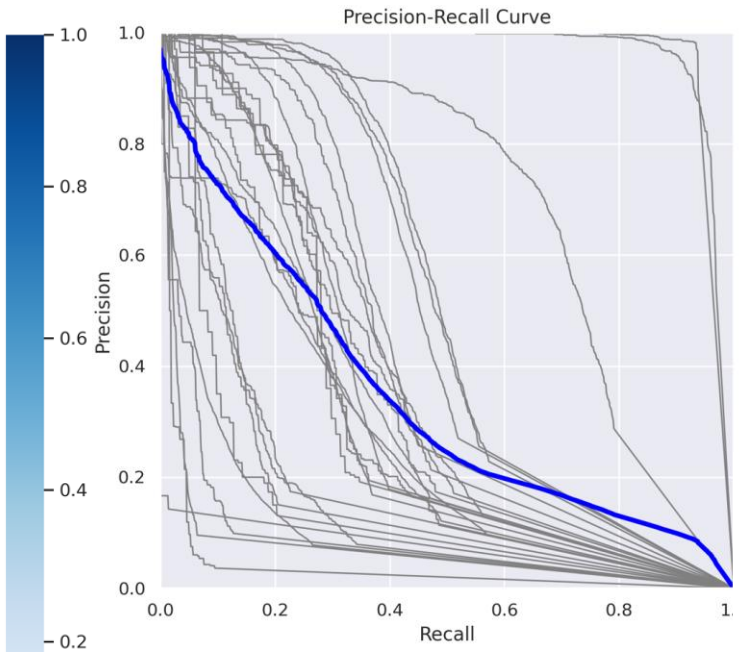


5.YOLO-V8 : IDD Dataset – Semantic Segmentation

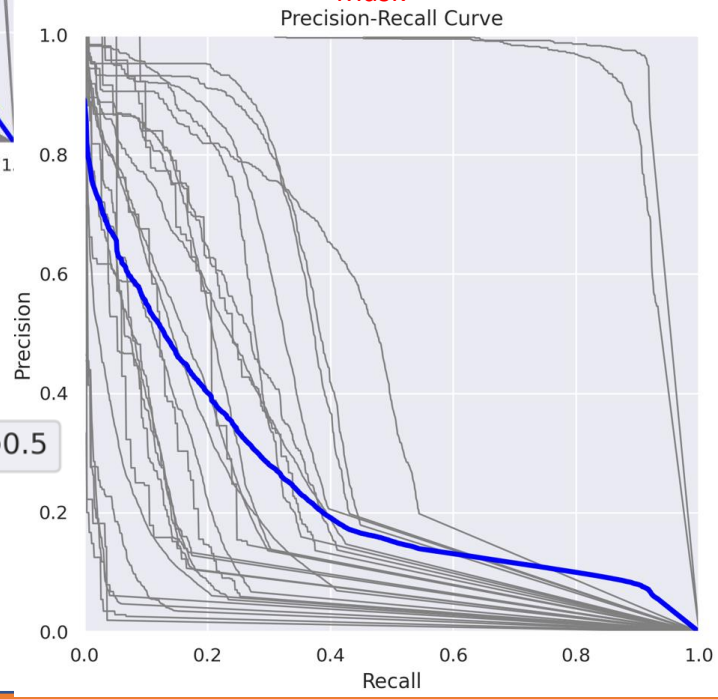
[Confusion Matrix & P-R Curves]



Box



Mask



5.YOLO-V8 : IDD Dataset – Semantic Segmentation

[Validation Metrics]

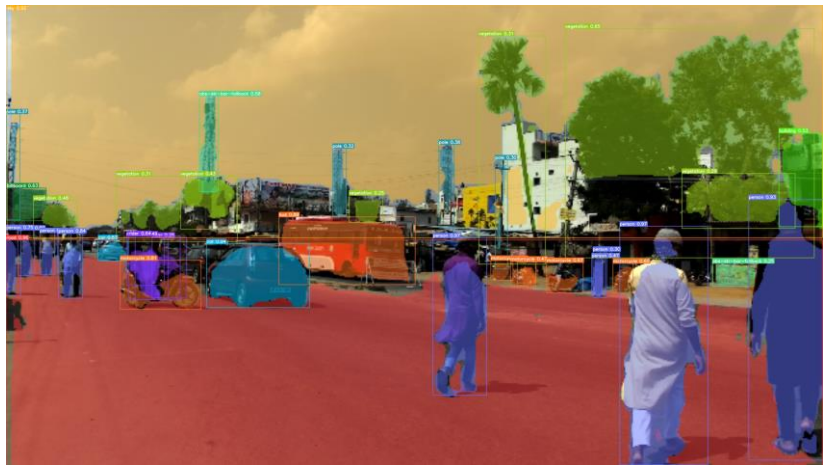
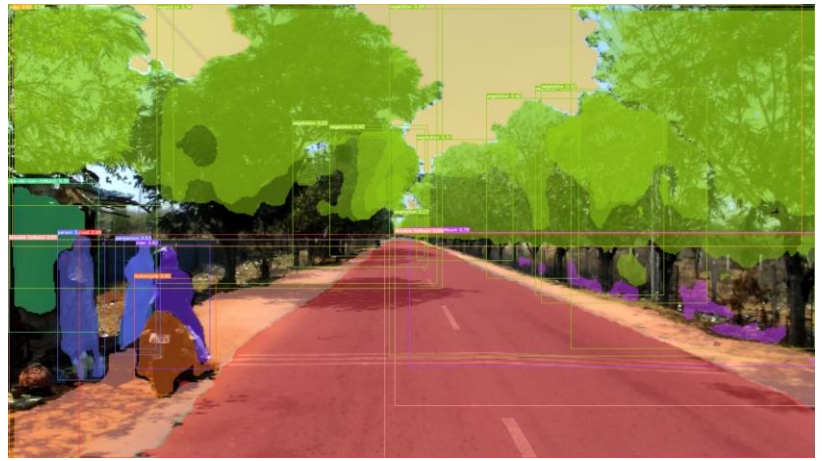
Class	Images	Instances	Box(P	R	mAP50	mAP50-95)	Mask(P	R	mAP50	mAP50-95):
all	1053	101654	0.00526	0.0166	0.00317	0.00134	0.00329	0.0116	0.00191	0.000697
person	1053	1079	0.0004	0.0241	0.000205	6.65e-05	1.54e-05	0.000927	7.72e-06	6.95e-06
bicycle	1053	2018	0.000334	0.00149	0.00018	1.8e-05	0.000111	0.000496	5.98e-05	1.79e-05
car	1053	6688	0.039	0.342	0.0336	0.012	0.0286	0.251	0.0208	0.00633
motorcycle	1053	1063	0	0	0	0	0	0	0	0
airplane	1053	1259	0.0235	0.00635	0.0119	0.00489	0.00293	0.000794	0.00149	0.000298
bus	1053	3818	0.0173	0.0503	0.00916	0.0034	0.0072	0.021	0.00369	0.00114
train	1053	15201	0.0122	0.00257	0.00625	0.00213	0.00251	0.000526	0.00127	0.000379
truck	1053	22848	0.0378	0.0418	0.0199	0.00834	0.0235	0.0259	0.012	0.0039
boat	1053	7335	0.0123	0.00354	0.00618	0.00278	0.00947	0.00273	0.00475	0.00188
traffic light	1053	2437	0.00295	0.0242	0.00151	0.000849	0.0022	0.0181	0.00112	0.00046
fire hydrant	1053	13106	0.000746	7.63e-05	0.000429	4.29e-05	0	0	0	0
stop sign	1053	3556	0.00301	0.000562	0.00152	0.000381	0.00301	0.000562	0.00152	0.00038
parking meter	1053	1043	0.0101	0.00671	0.00508	0.00421	0.00862	0.00575	0.00435	0.00262
bench	1053	5285	0.0103	0.0235	0.00549	0.00203	0.00582	0.0132	0.00297	0.00102
bird	1053	692	0.0227	0.0766	0.0148	0.00757	0.0231	0.078	0.0139	0.0066
cat	1053	5654	0	0	0	0	0	0	0	0
dog	1053	2665	0.00133	0.000375	0.000698	0.000349	0.00398	0.00113	0.00202	0.000536
horse	1053	1174	0.000454	0.000852	0.000341	0.000205	0.000454	0.000852	0.000341	3.41e-05
sheep	1053	1509	0	0	0	0	0	0	0	0
cow	1053	265	0	0	0	0	0	0	0	0
elephant	1053	649	0	0	0	0	0	0	0	0
bear	1053	889	0	0	0	0	0	0	0	0
zebra	1053	328	0	0	0	0	0	0	0	0
giraffe	1053	133	0	0	0	0	0	0	0	0
backpack	1053	3	0	0	0	0	0	0	0	0
umbrella	1053	121	0.000433	0.00826	0.000221	0.000177	0.000433	0.00826	0.000221	0.000177
tie	1053	22	0	0	0	0	0	0	0	0
frisbee	1053	151	0	0	0	0	0	0	0	0
skis	1053	237	0	0	0	0	0	0	0	0
snowboard	1053	300	0	0	0	0	0	0	0	0
sports ball	1053	12	0	0	0	0	0	0	0	0
kite	1053	103	0	0	0	0	0	0	0	0
baseball glove	1053	5	0	0	0	0	0	0	0	0
skateboard	1053	2	0	0	0	0	0	0	0	0
surfboard	1053	2	0	0	0	0	0	0	0	0
bottle	1053	1	0	0	0	0	0	0	0	0
wine glass	1053	1	0	0	0	0	0	0	0	0

5.YOLO-V8 : IDD Dataset – Semantic Segmentation

[Semantic Segmentation – Images]





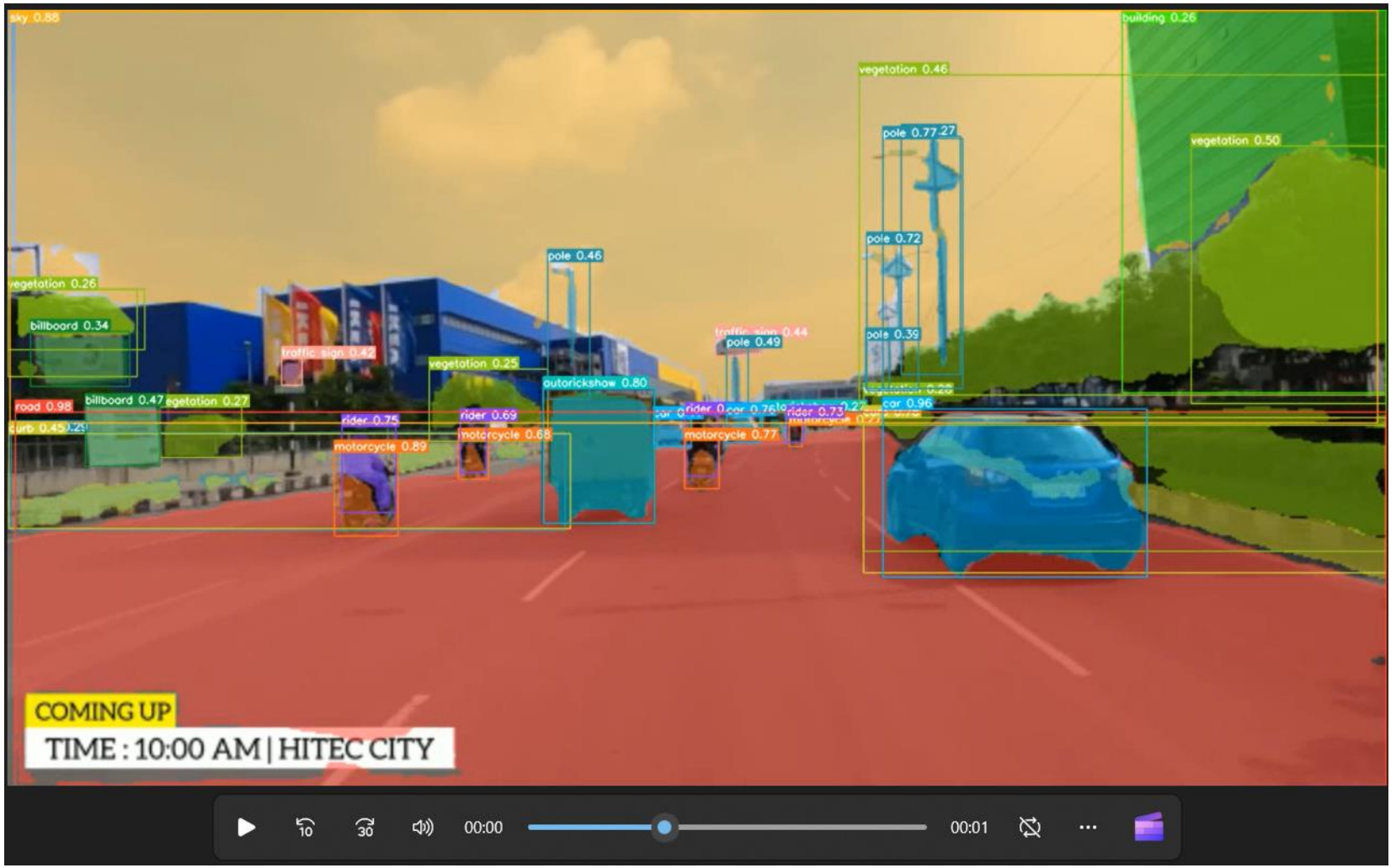



5.YOLO-V8 : IDD Dataset – Semantic Segmentation

[Semantic Segmentation – Videos]

Pre-trained COCO dataset

After BDD100K dataset training



Summary:

	YOLO-V5 [BDD100k]	YOLO-V7 [BDD100k]	YOLO-V8 [BDD100k]	YOLO-V8 [IDD Detection]	YOLO-V8 [IDD Seg]
Pretrained model	Yolov5m	yolov7	yolov8n	Yolov8n	yolov8n-seg
# Images for training (Train:Test:Val)	~70% of 100k (70:20:10)	~70% of 100k (70:20:10)	~70% of 100k (70:20:10)	~68% of 46588 (70:20:10)	~70% of 10k* (70:20:10) * -10k PNG images were not used (~20GB) due to resource constraint
# Classes	13	13	13	15	41
# Batch Size	64	32	32	32	32
Image Size	640	640	640	640	640

Summary:

	YOLO-V5 [BDD100k]	YOLO-V7 [BDD100k]	YOLO-V8 [BDD100k]	YOLO-V8 [IDD Detection]	YOLO-V8 [IDD Seg]
# epochs	20	25	20	20	20
Train time per epoch	~00:27:40	~00:25:20	~00:08:00	~00:04:00	~00:12:40
map@0.5 [box]	Train: 0.426 Valid: 0.426	Train: 0.459 Valid: 0.458	Train: 0.328 Valid: 0.328	Train: 0.324 Valid: 0.324	Train: 0.260 Valid: 0.260
map@0.5 [mask]	NA	NA	NA	NA	Train: 0.182 Valid: 0.180

GPU used for training - NVIDIA A100-SXM4-40GB, 40536MiB)

Thank You