

IMAGE TAGGING AND ROAD OBJECT DETECTION

Guide : Prof. Anoop M. Namboodiri

Mentor : Sangeeth Reddy

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



Agenda

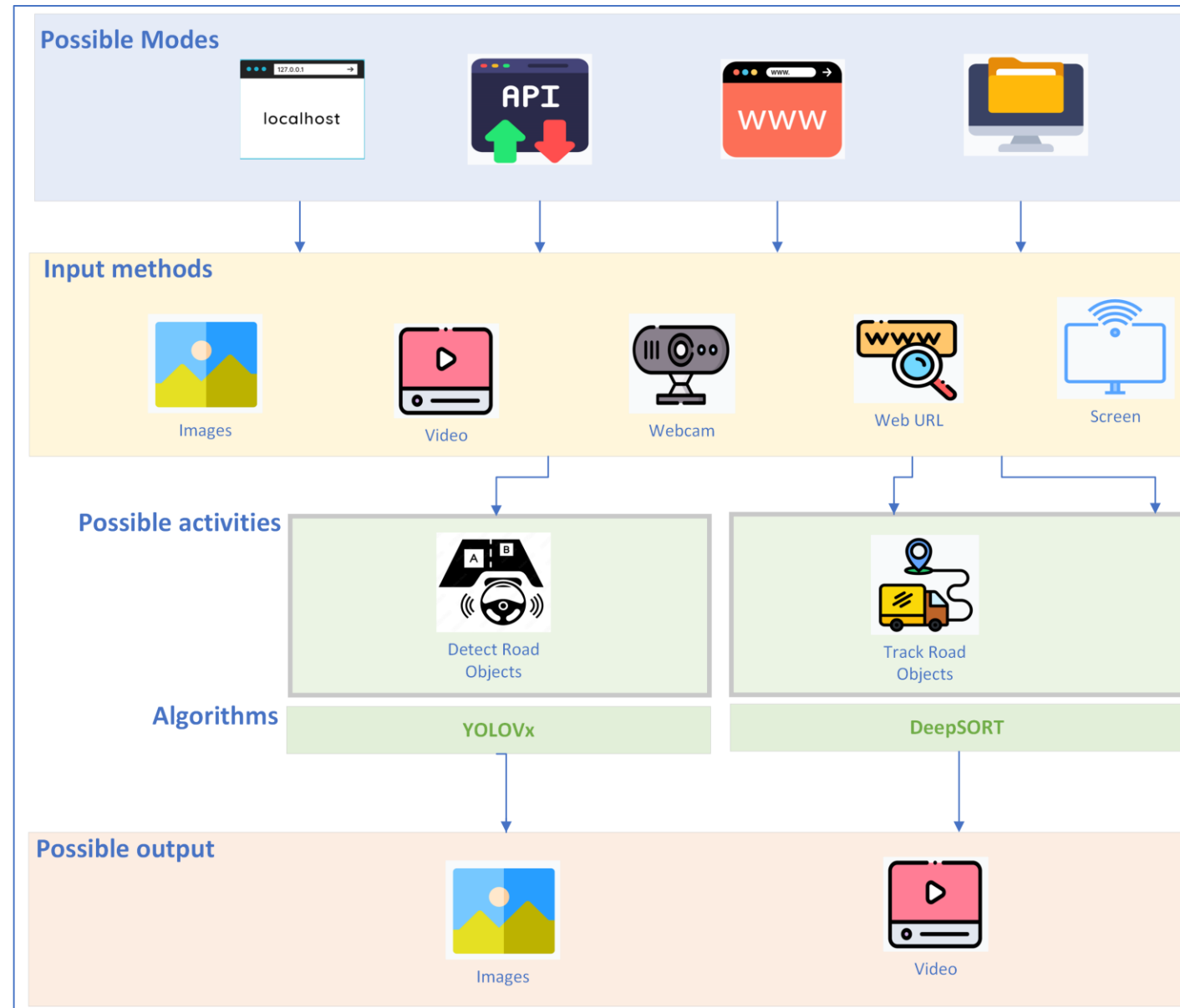
- Objective
- Methodology and design considerations
- Outcome in stages with final outcome
- Challenges
- Application Demo



Objective

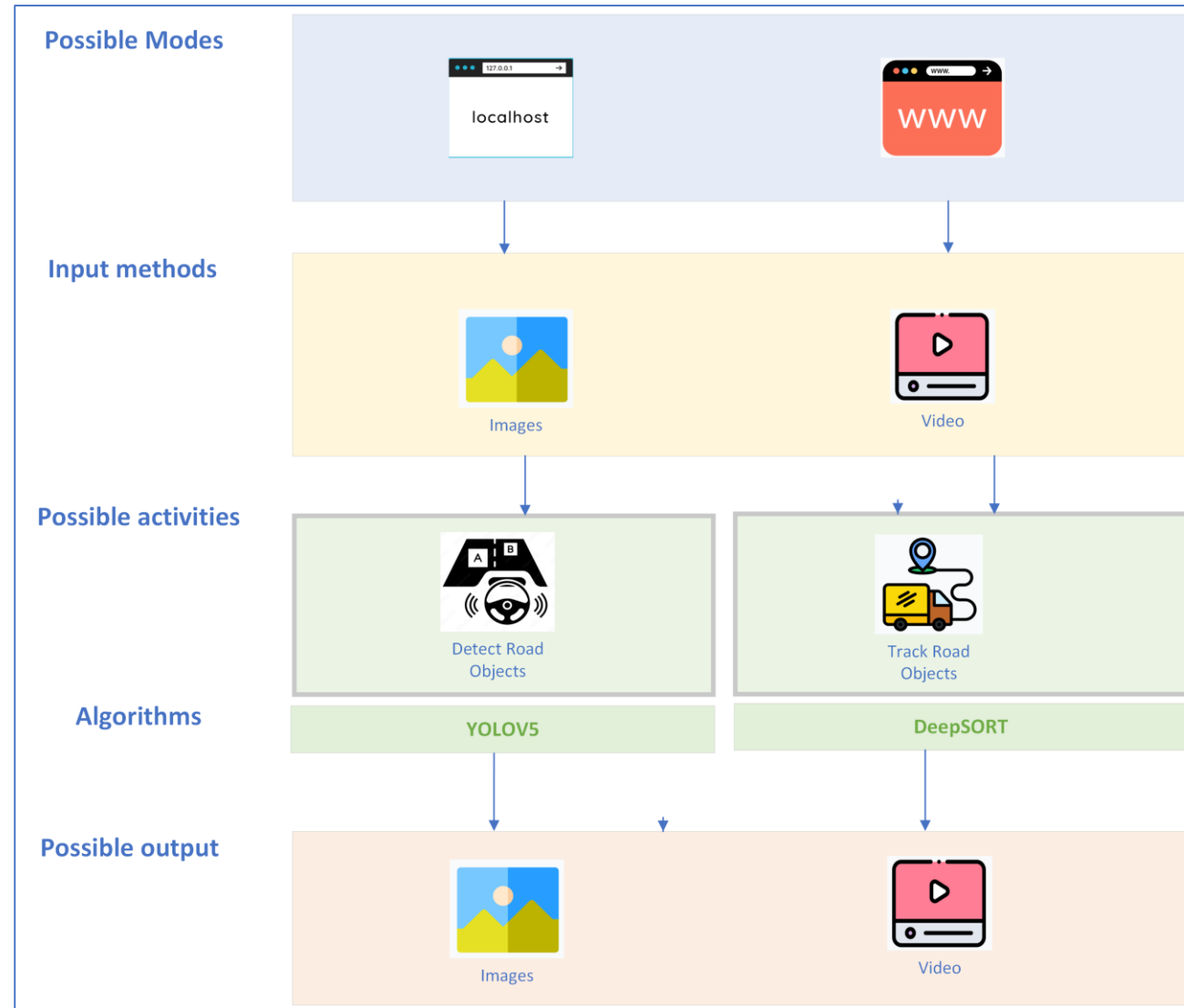
Build an application to detect multiple objects, tag and tracking in a video.

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

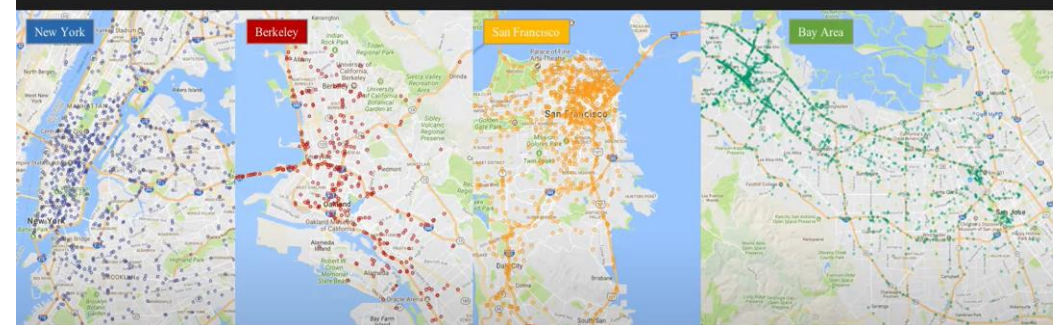
Scene Diversity : Scene = City , Tunnel , Highway , Parking , Residential

Time of the day = Dusk , Daytime , night

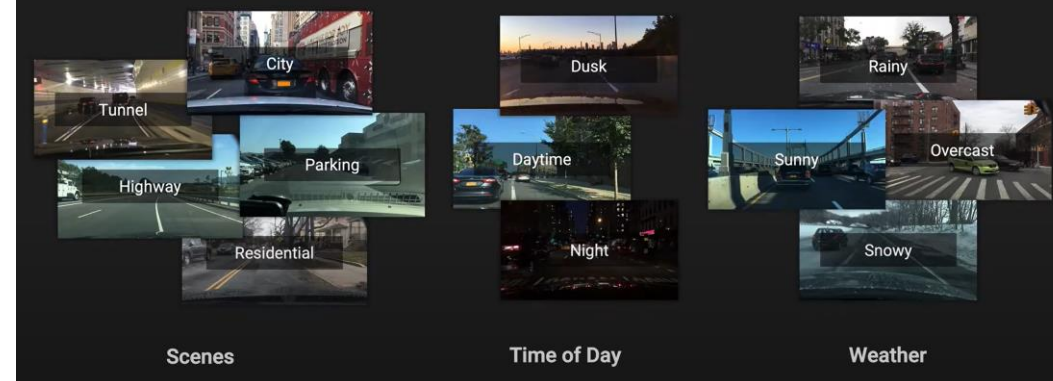
Weather = Rainy , Sunny , Overcast , Snowy

There are >10 objects per image with pixel annotation and tracking

Geography Diversity



Scene Diversity



Outcome in stages

Stage 1 – Understand the data



Original Image

```
bdd100k_labels_val.json  X
1  [
2    {
3      "name": "000d4f89-3bcbe37a.jpg",
4      "attributes": {
5        "weather": "overcast",
6        "scene": "city street",
7        "timeofday": "daytime"
8      },
9      "timestamp": 10000,
10     "labels": [
11       {
12         "category": "traffic sign",
13         "attributes": {
14           "occluded": false,
15           "truncated": false,
16           "trafficLightColor": "none"
17         },
18         "manualShape": true,
19         "manualAttributes": true,
20         "box2d": {
21           "x1": 1000.698742,
22           "y1": 281.992415,
23           "x2": 1040.626872,
24           "y2": 326.91156
25         },
26         "id": 0
27       },
28       {
29         "category": "traffic sign",
30         "attributes": {
```

BDD100k dataset bounding box JSON format

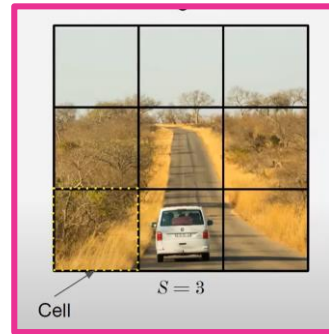
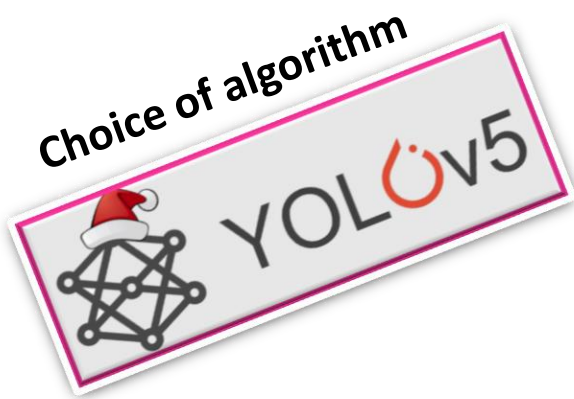
One object
complete details
in an image

Object class

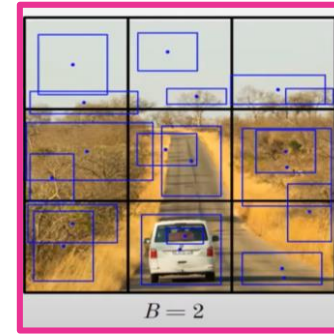
Bounding box
dimensions

Outcome in stages

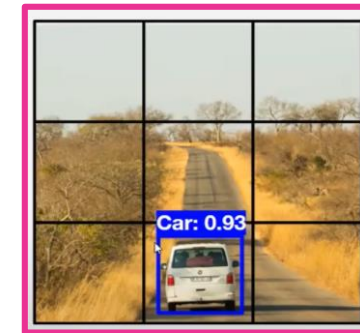
Stage 2 – Identify the object detection algorithm



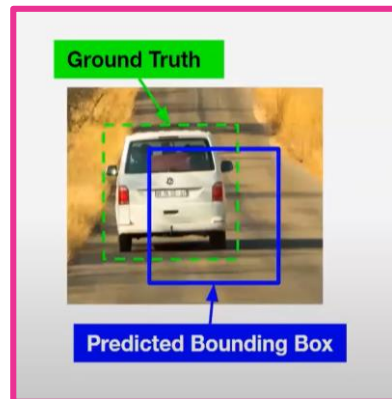
Divide the image to cells



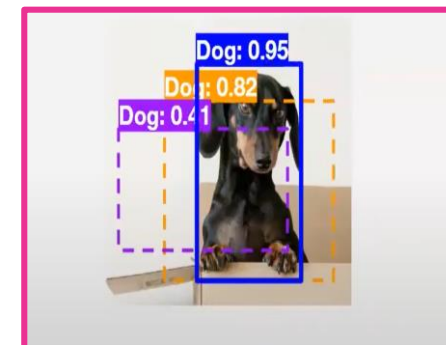
Predicting bounding boxes within each cell



Selection of bounding box which is above confidence threshold



Intersection over Union

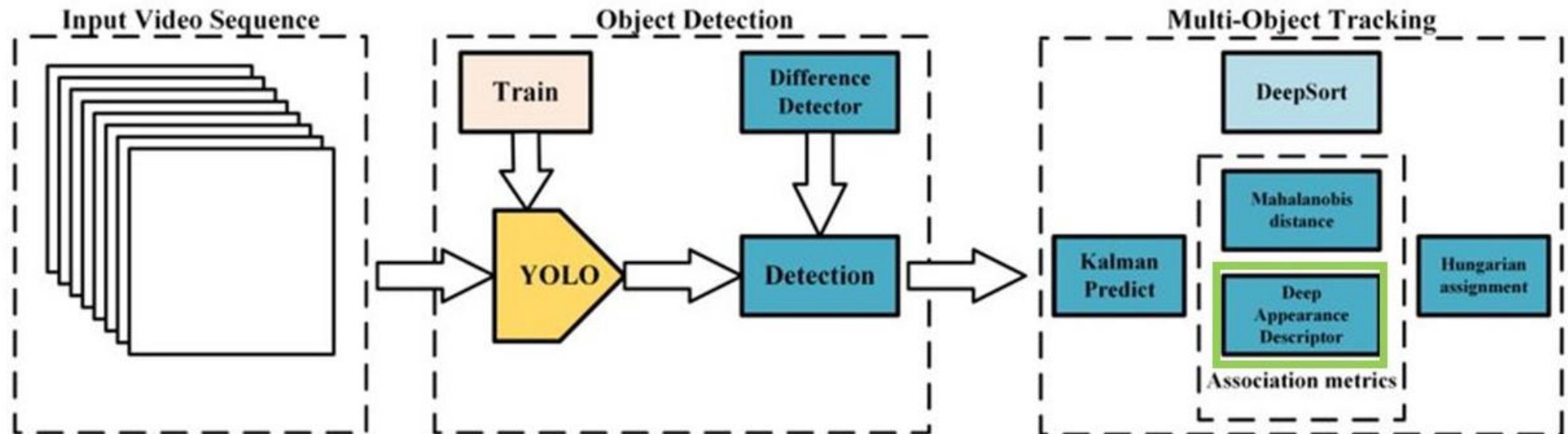


Non max suppression

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



BDD100k dataset bounding box JSON format

```
1 [
2   {
3     "name": "000d4f89-3bcbe37a.jpg",
4     "attributes": {
5       "weather": "overcast",
6       "scene": "city street",
7       "timeofday": "daytime"
8     },
9     "timestamp": 10000,
10    "labels": [
11      {
12        "category": "traffic sign",
13        "attributes": {
14          "occluded": false,
15          "truncated": false,
16          "trafficLightColor": "none"
17        },
18        "manualShape": true,
19        "manualAttributes": true,
20        "box2d": {
21          "x1": 1000.698742,
22          "y1": 281.992415,
23          "x2": 1040.626872,
24          "y2": 326.91156
25        },
26        "id": 0
27      },
28      {
29        "category": "traffic sign",
30        "attributes": {
```

One object complete details in an image

Object class

Bounding box dimensions

bounding box x1,y1,x2,y2 normalized

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

000d4f89-3bcbe37a - Notepad

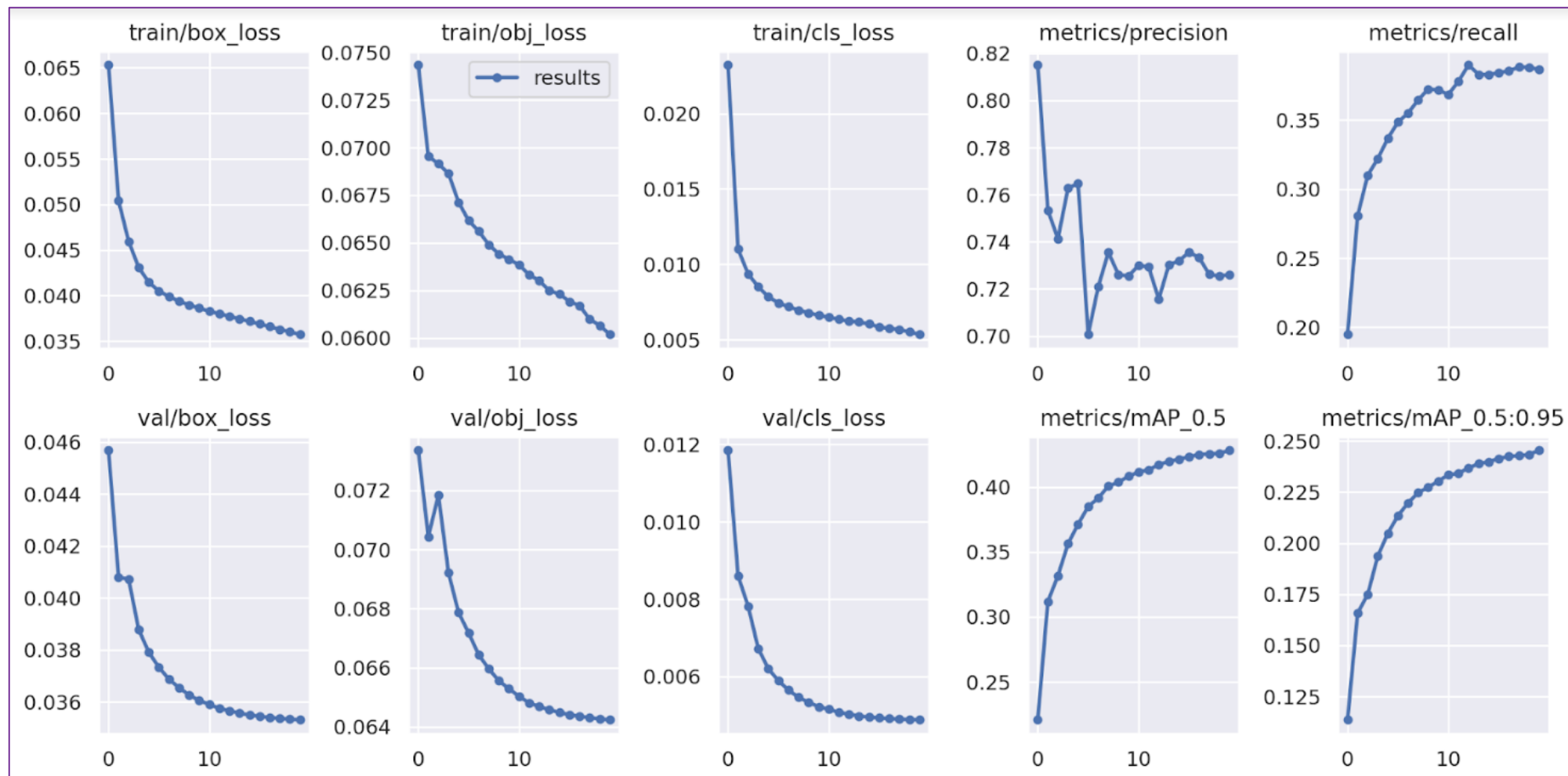
File Edit Format View Help

```
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 0.47204071666666667 0.03202805625 0.025714280555555608
2 0.015361844921875001 0.4674488784722222 0.030723689843750002 0.060612231944444435
2 0.46025830133564527 0.5259084653235868 0.05424690826620946 0.05630693620272922
```

Class Label

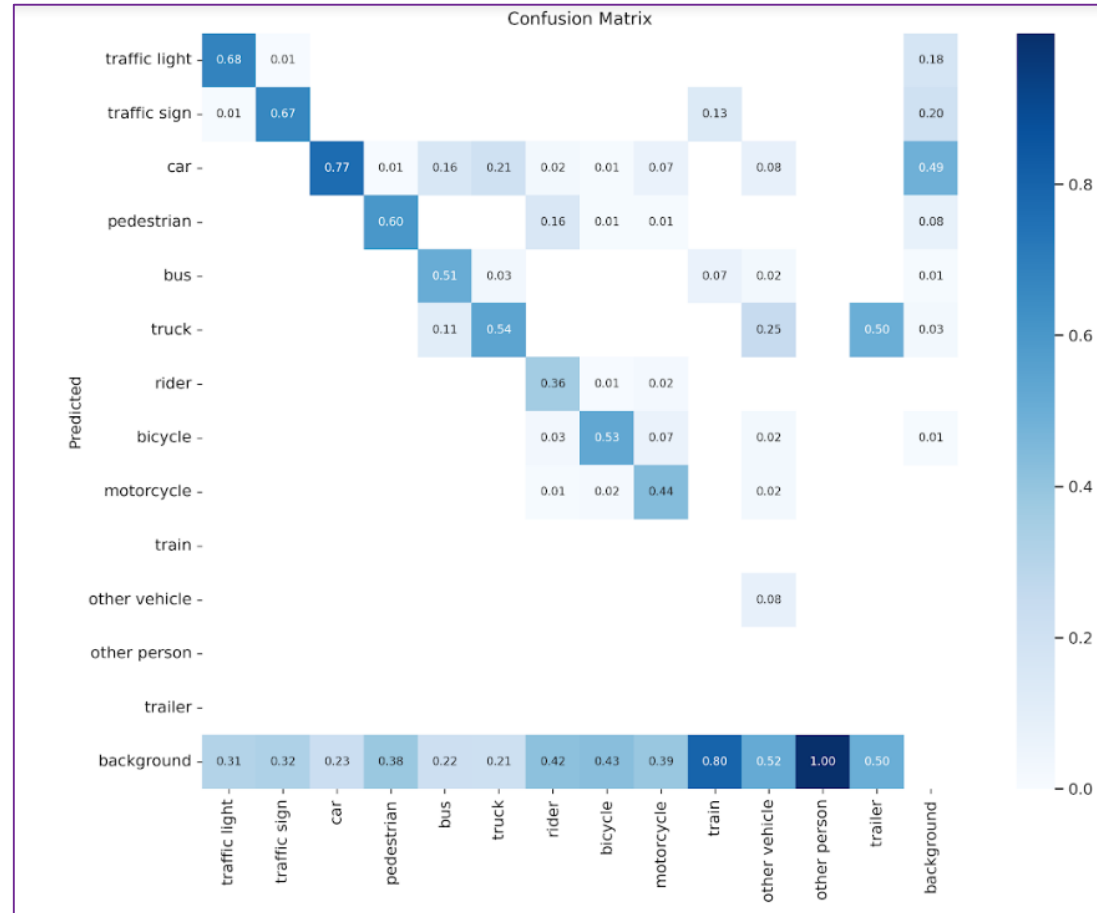
Actual Outcomes

Preliminary results



Actual Outcomes

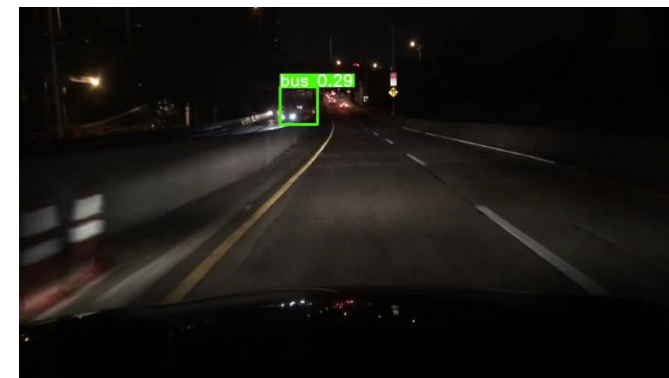
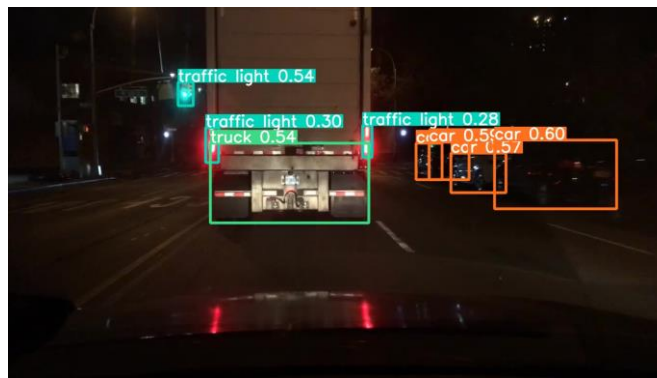
Preliminary results – Confusion Matrix



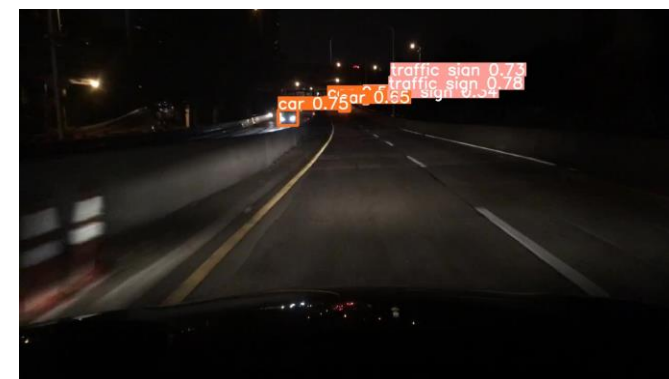
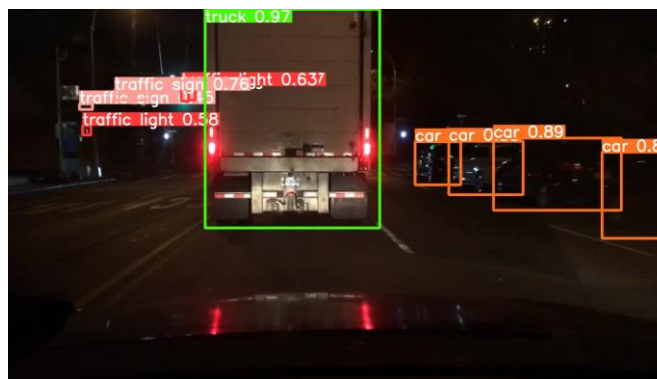
Actual Outcomes

YOLOv5 Object Detection

Pre-trained
COCO
dataset

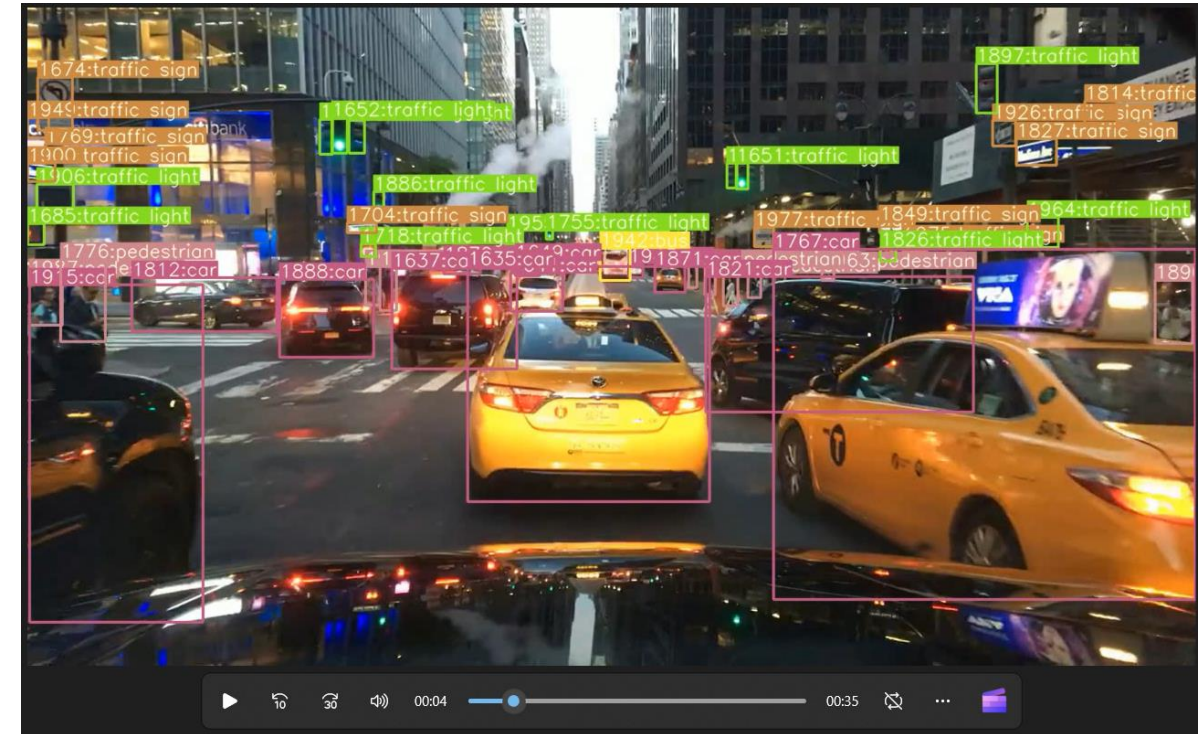
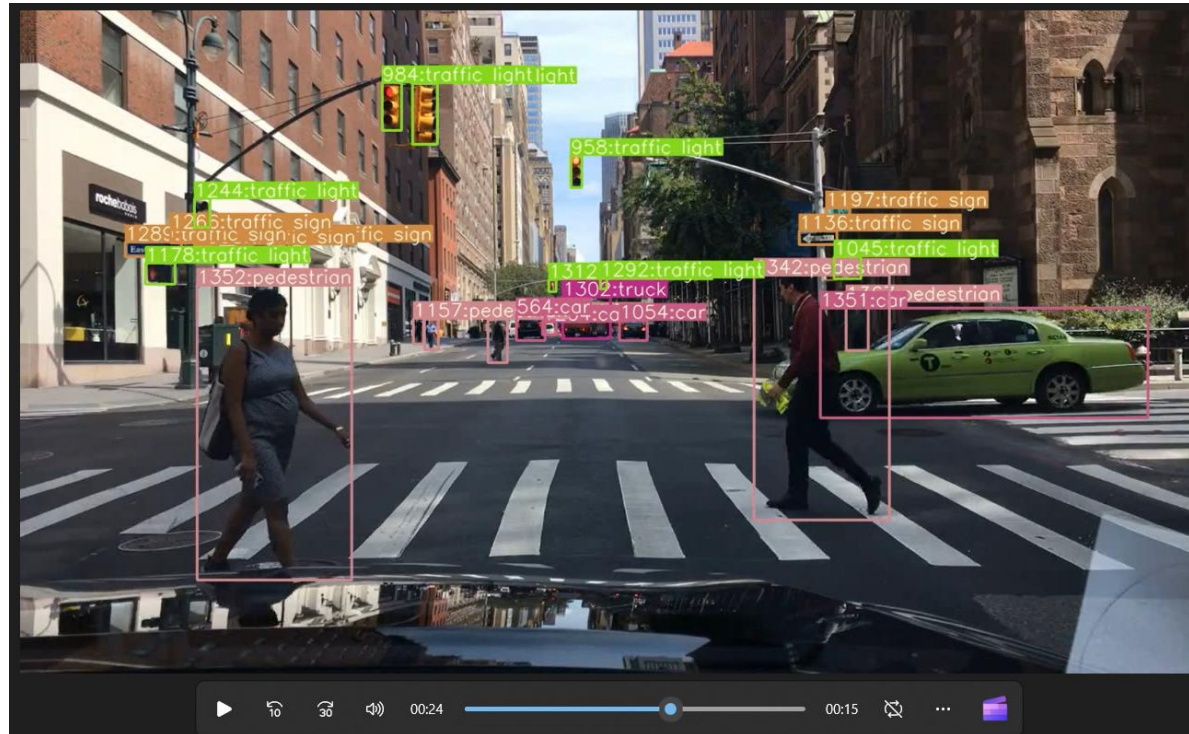


After
BDD100K
dataset
training



Outcome in stages

Stage 5 – Process videos for object tracking



Actual Outcomes

YOLOv5 + Deep SORT tracker output



Tracked GIF via YOLOv5 + DeepSORT

Outcome in stages

Stage 6 – Develop streamlit application and deploy onto Streamlit Cloud

The screenshot shows the 'User Configurations' sidebar on the left with the following settings:

- Select input type: ☒ Image
- Select IoU threshold: 0.50 (slider from 0.50 to 1.00)
- Select confidence threshold: 0.10 (slider from 0.10 to 1.00)

The main content area is titled 'IIIT - Multi Object Detection' and includes a note: 'Select options left-handed menu bar.' Below this is an 'Upload An Image' section with a drag-and-drop area that says 'Drag and drop file here' and 'Limit 200MB per file • PNG, JPEG, JPG', along with a 'Browse files' button.

Application for Image detection

The screenshot shows the 'User Configurations' sidebar on the left with the following settings:

- Select input type: ☒ Video
- Select IoU threshold: 0.50 (slider from 0.50 to 1.00)
- Select confidence threshold: 0.10 (slider from 0.10 to 1.00)
- Objects of interest on road scene: Choose an option (dropdown menu)
- ☐ Select all
- Do we need to track objects?: ☒ Yes

The main content area is titled 'IIIT - Multi Object Detection' and includes a note: 'Select options left-handed menu bar.' Below this is an 'Upload Video' section with a drag-and-drop area that says 'Drag and drop file here' and 'Limit 200MB per file • MP4, MPEG, MOV', along with a 'Browse files' button.

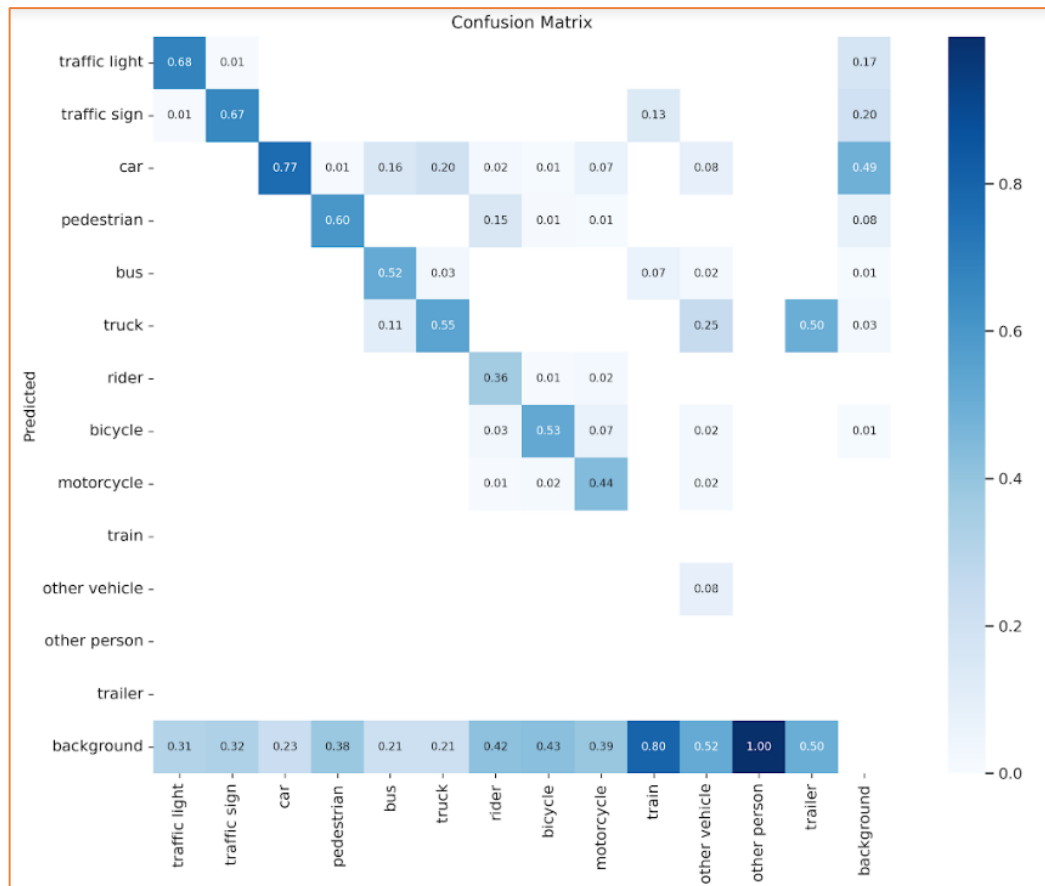
Application for Video detection and tracking

Challenges

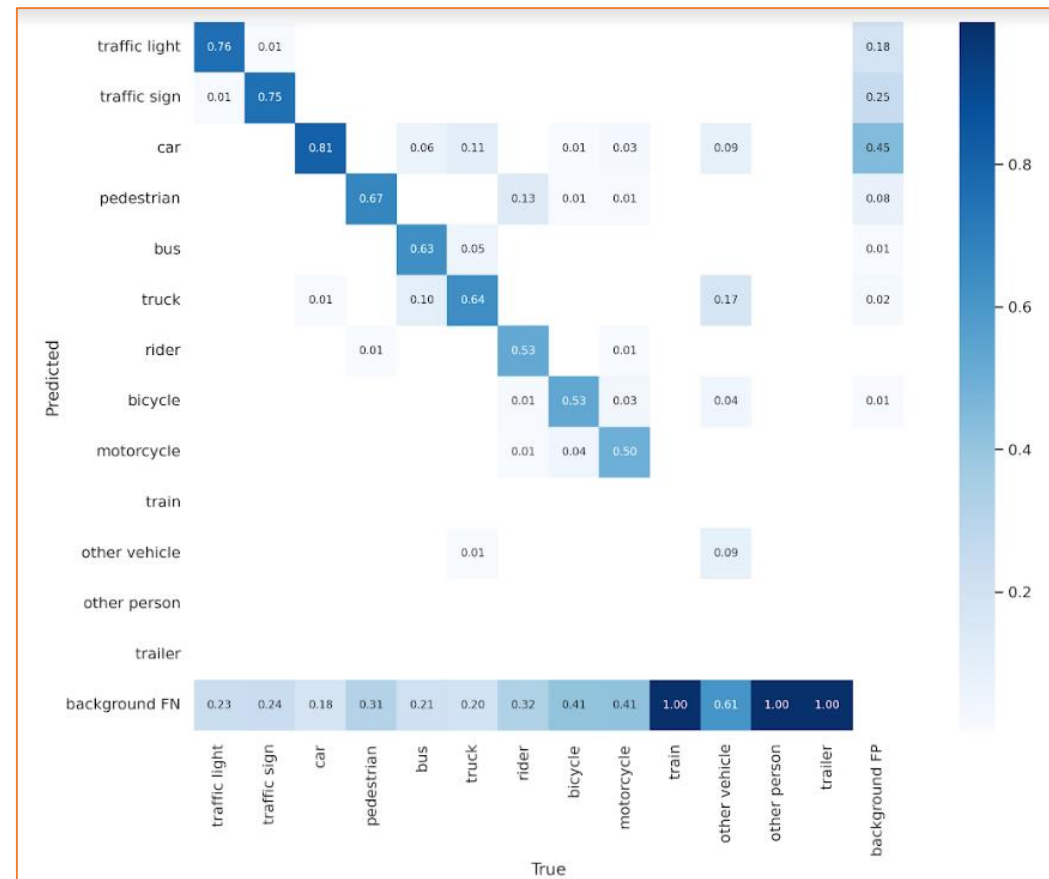
1. Missing labels for some of the images of BDD100k data.
2. Downloading of BDD100k videos were time consuming.
3. Resource constraint – GPU causing disconnection of runtime in colab.
4. Sharing GIT Code repositories
5. Understanding cloud constraints
6. Understanding installation of packages in streamlit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

Stretch Goals

YOLOv5 vs YOLOv7



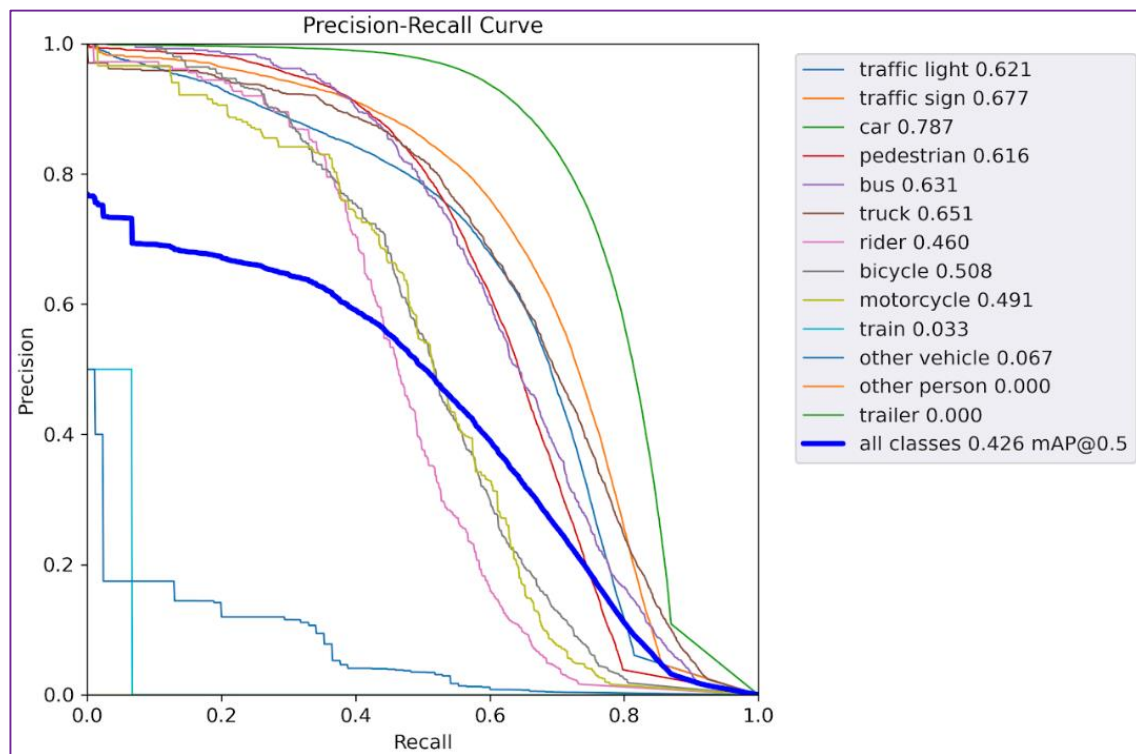
V5



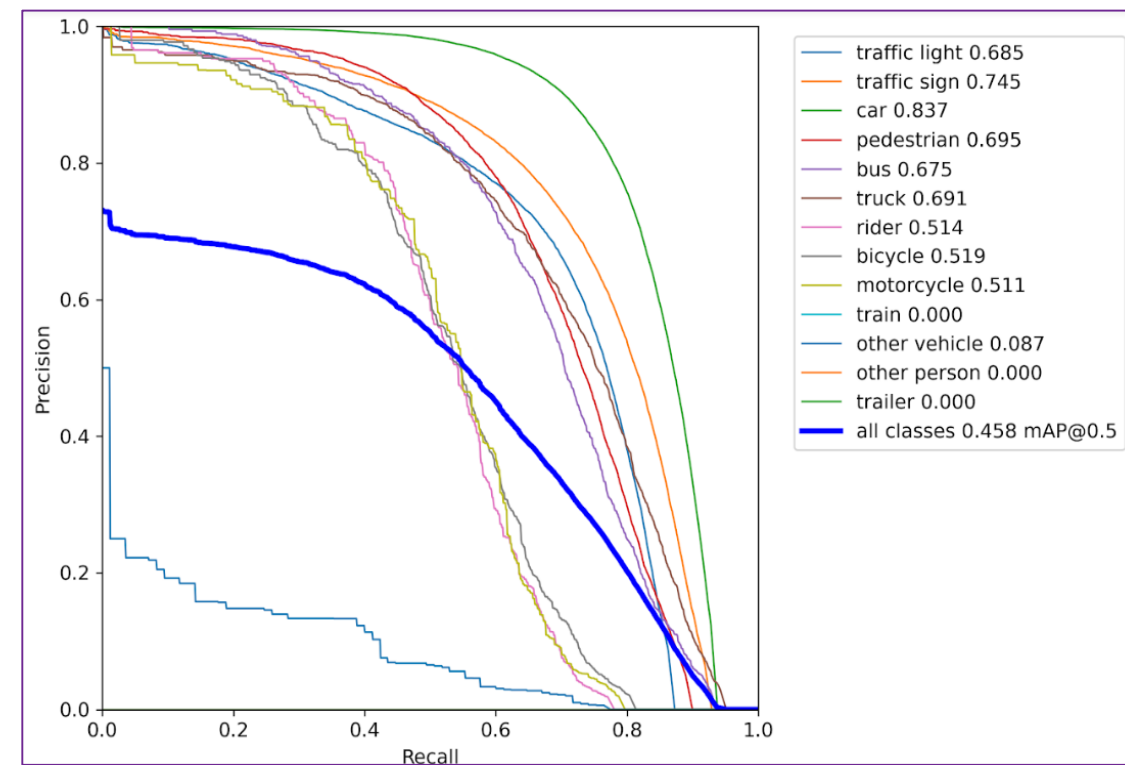
V7

Stretch Goals

YOLOv5 vs YOLOv7



V5



V7

Stretch Goals

YOLOv5 vs YOLOv7

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

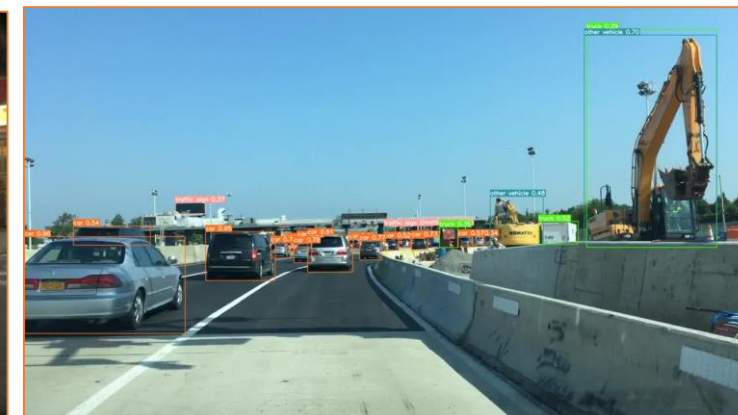
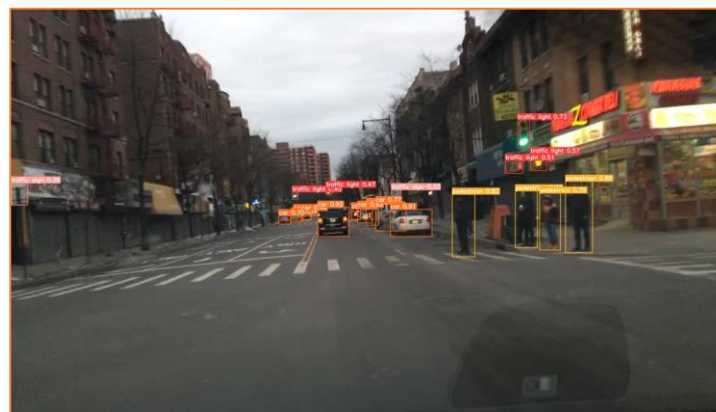
Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95:
all	10000	186033	0.742	0.422	0.458	0.258
traffic light	10000	26884	0.708	0.668	0.685	0.258
traffic sign	10000	34724	0.727	0.701	0.745	0.396
car	10000	102837	0.833	0.76	0.837	0.522
pedestrian	10000	13425	0.756	0.617	0.695	0.344
bus	10000	1660	0.752	0.586	0.675	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

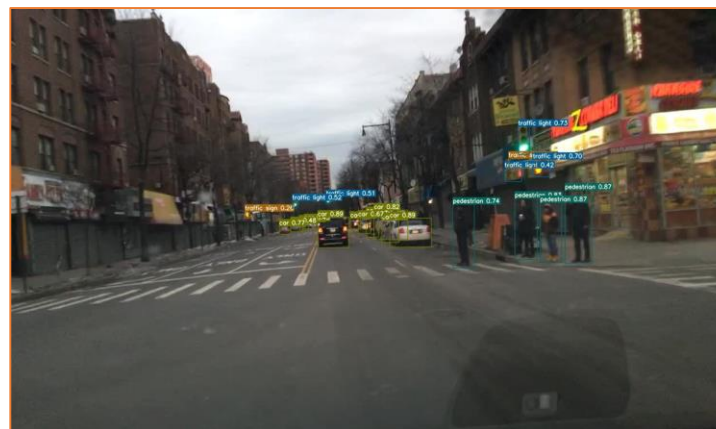
Stretch Goals

YOLOv5 vs YOLOv7

V5



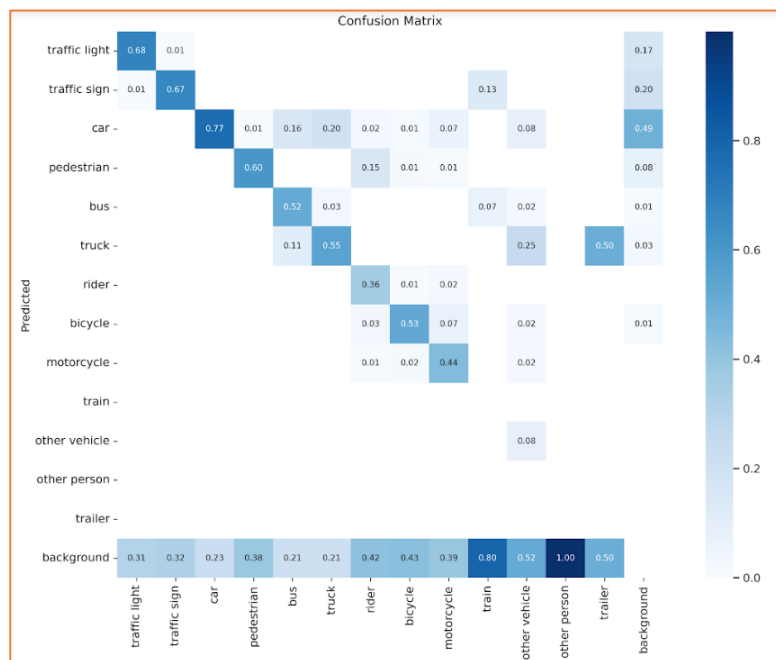
V7



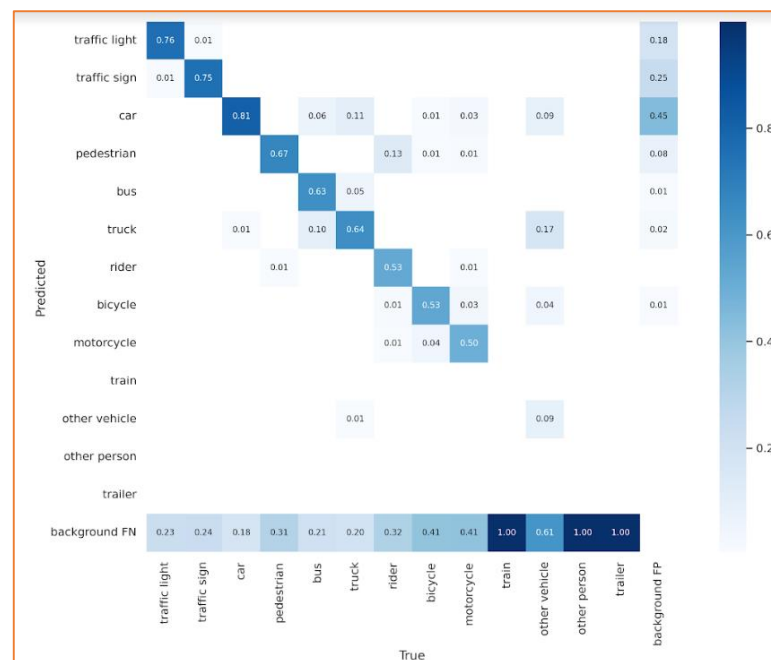
Stretch Goals

YOLOv5 vs YOLOv7 vs YOLOv8

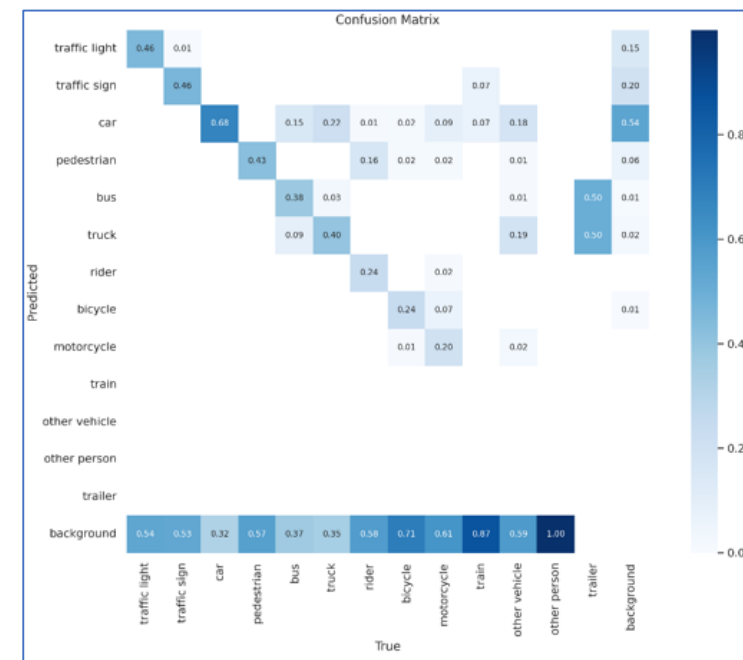
(Trained v8 nano model architecture with all BDD100k data for 20 epochs)



V5



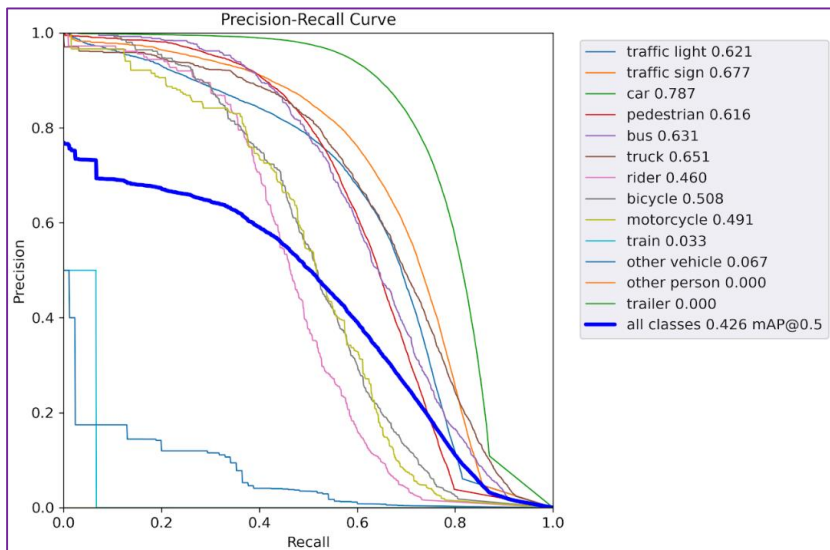
V7



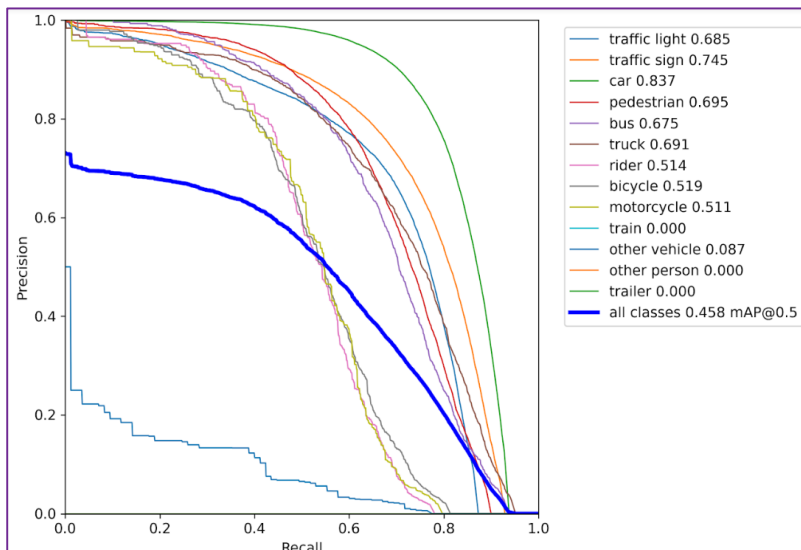
V8

Stretch Goals

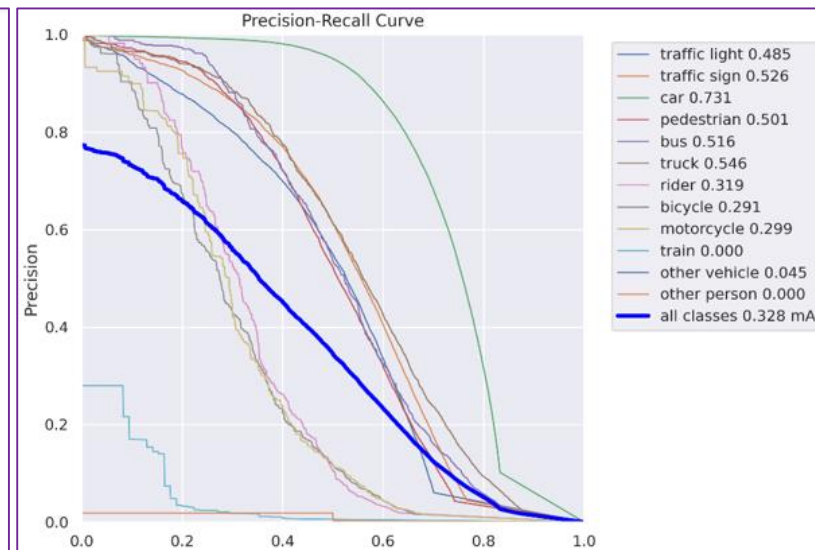
YOLOv5 vs YOLOv7 vs YOLOv8



V5



V7



V8

Stretch Goals

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

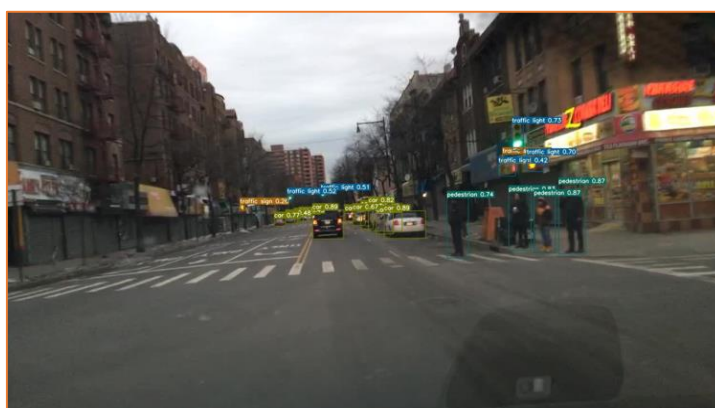
V7

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

Stretch Goals

Explored YOLOv8



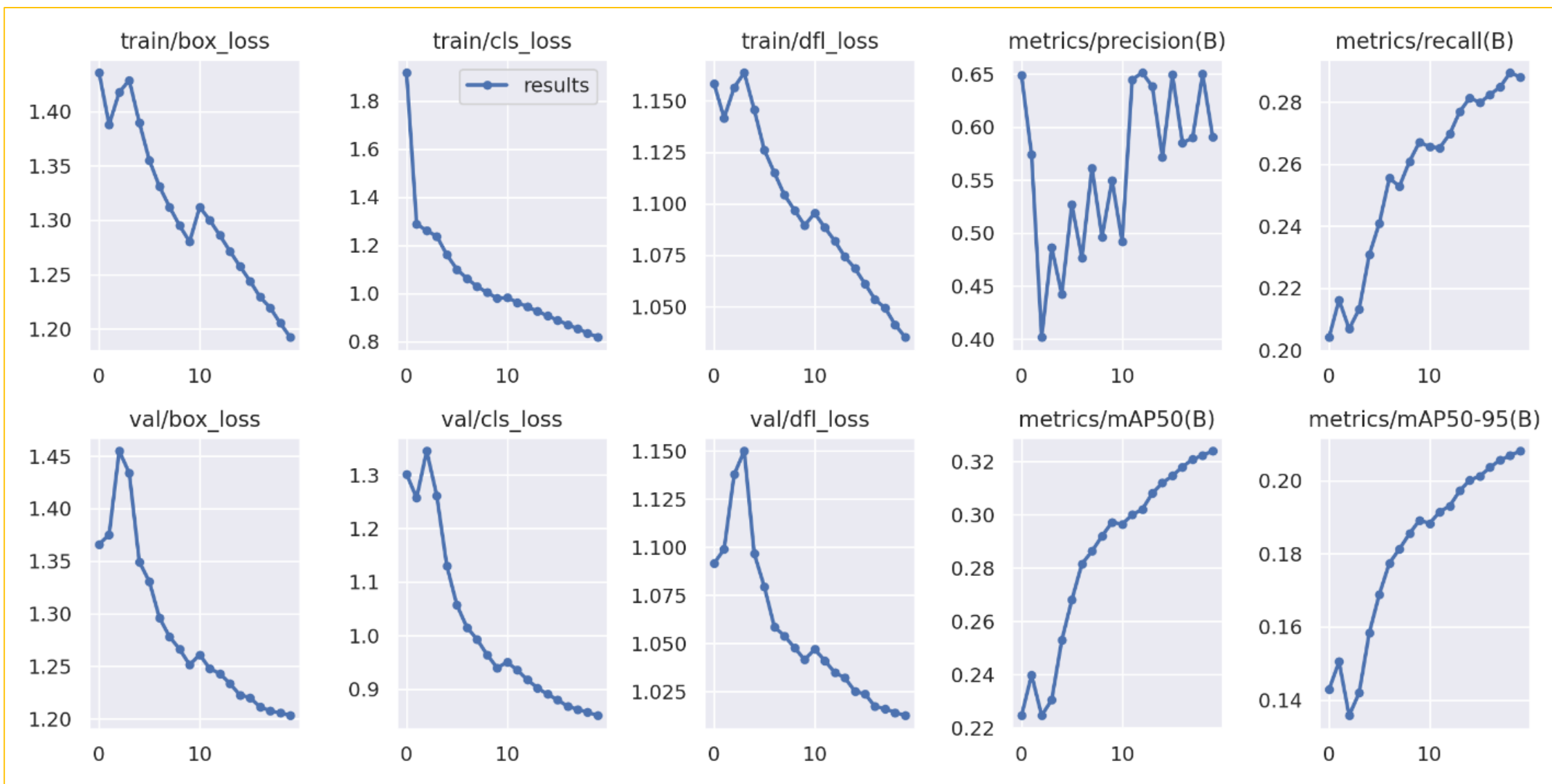
V5

V7

V8

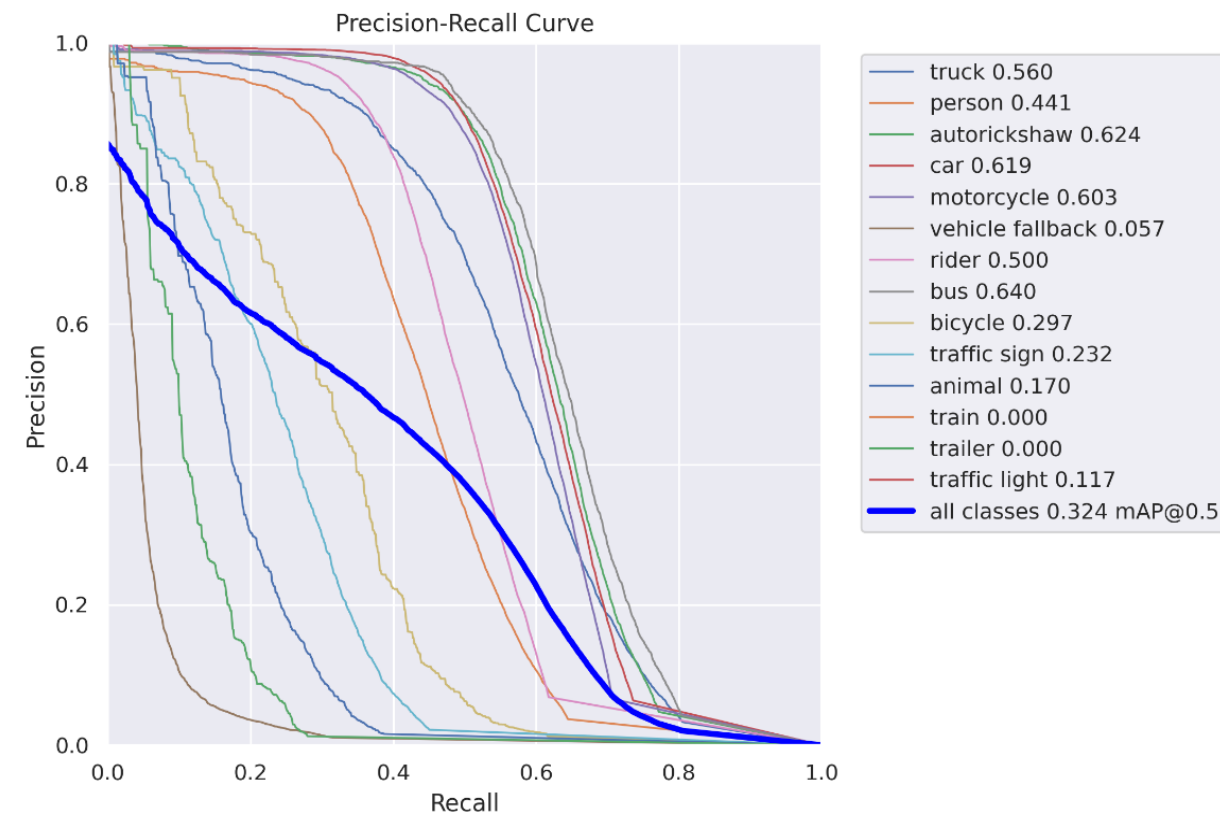
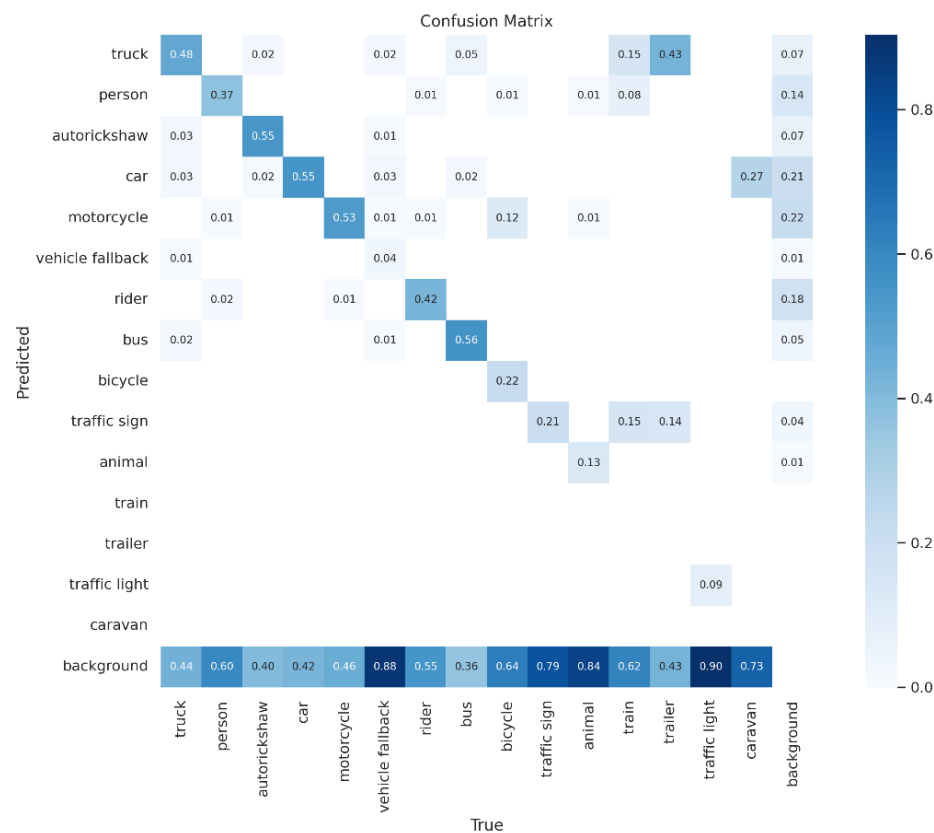
Actual Outcomes

YOLOv8 – Trained on IDD dataset



Actual Outcomes

Confusion Matrix & P-R Curves



Actual Outcomes

Validation Metrics

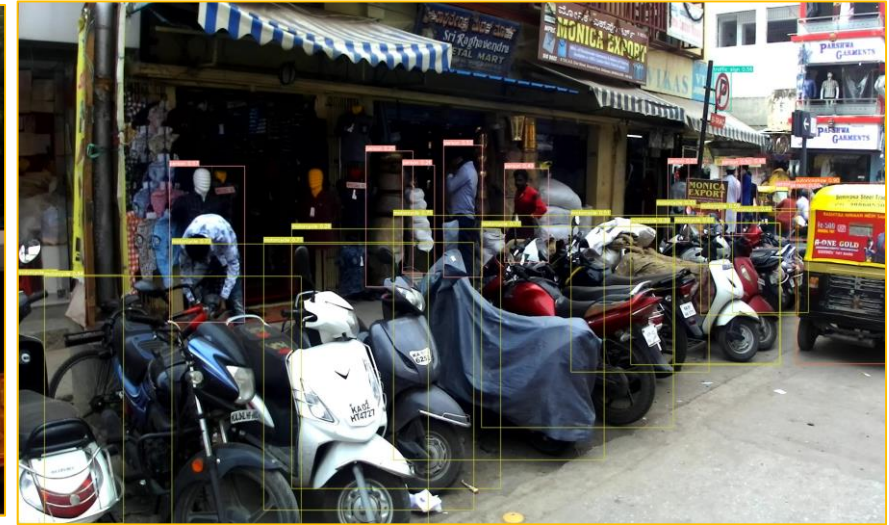
Class	Images	Instances	Box(P	R	mAP50	mAP50-95) :
all	10224	126004	0.59	0.288	0.324	0.208
truck	10224	7075	0.673	0.508	0.561	0.4
person	10224	18070	0.692	0.382	0.442	0.244
autorickshaw	10224	7781	0.745	0.568	0.624	0.445
car	10224	24831	0.751	0.557	0.619	0.432
motorcycle	10224	25484	0.765	0.544	0.602	0.357
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

Actual Outcomes

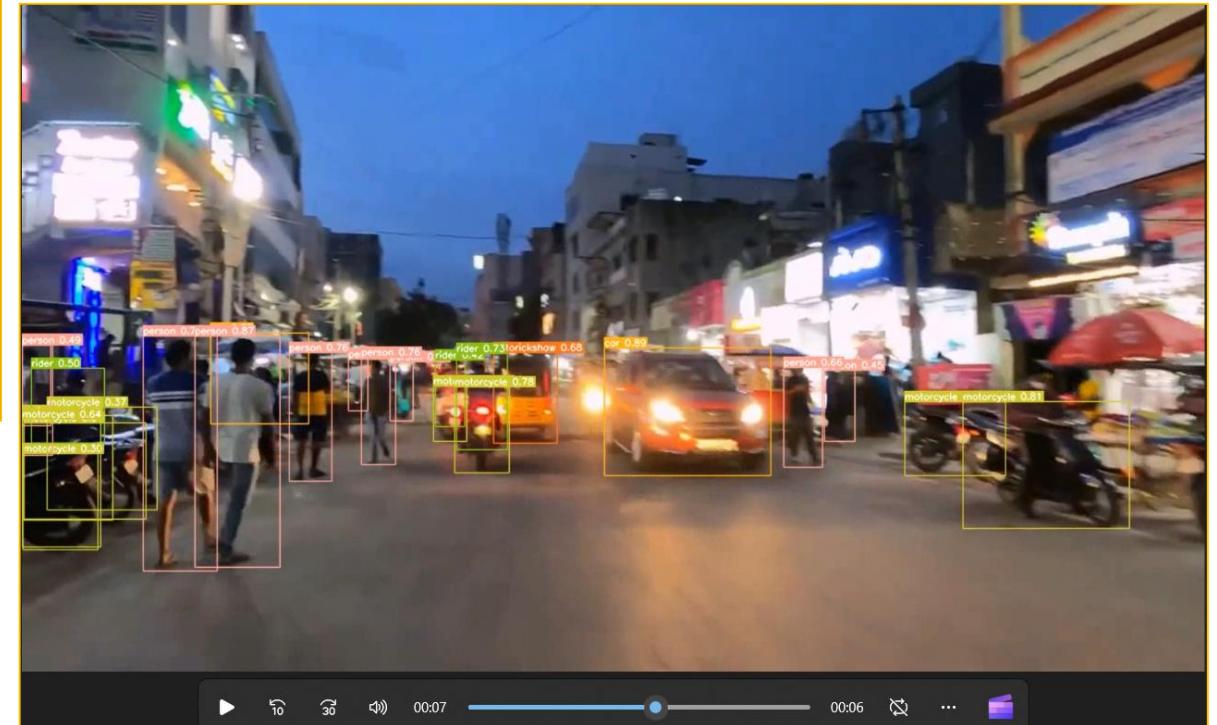
YOLOv8 Object Detection



BDD100K
dataset
training



YOLOv8 Object Detection - Videos



Stretch Goals

YOLOv8 –Semantic segmentation on IDD dataset (IN PROGRESS)

☐ Completed:

- Explored semantic segmentation using YOLO-v8.
- Downloaded dataset.
- Understood the label file format.

☐ In Progress:

- Script to convert JSON label files to YOLO understandable text file format.

☐ Future Steps:

- Training
- Review Train and validation metrics
- Review image and video results.

Application Demo

