EVALUATING THE SIGNIFICANCE OF OUTDOOR ADVERTISING UTILIZING COMPUTER VISION - II

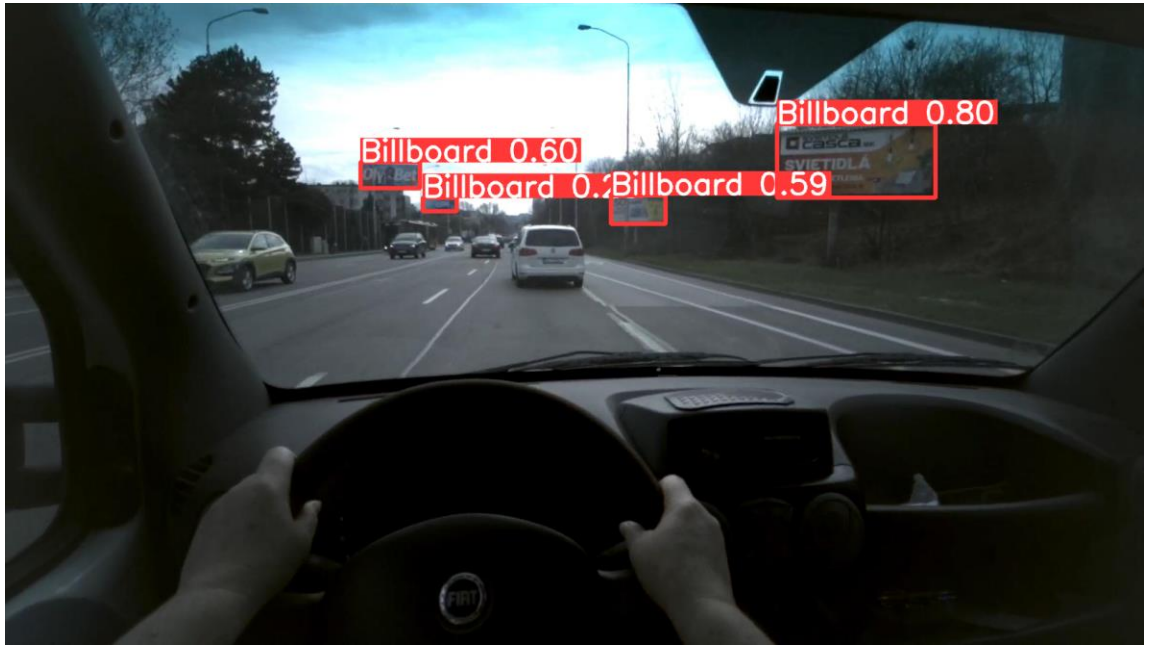
CARLOS ANDRES PIZARROSO TRONCOSO

SUPERVISOR: RNDR. ZUZANA BERGER HALADOVÁ, PHD

RECAP I: MOTIVATION

- Impact of roadside advertisements on drivers' attention.
- Development of a system that contributes to road safety.
- Application of Deep Learning technologies in Computer Vision.

RECAP II: PROJECT OBJECTIVES

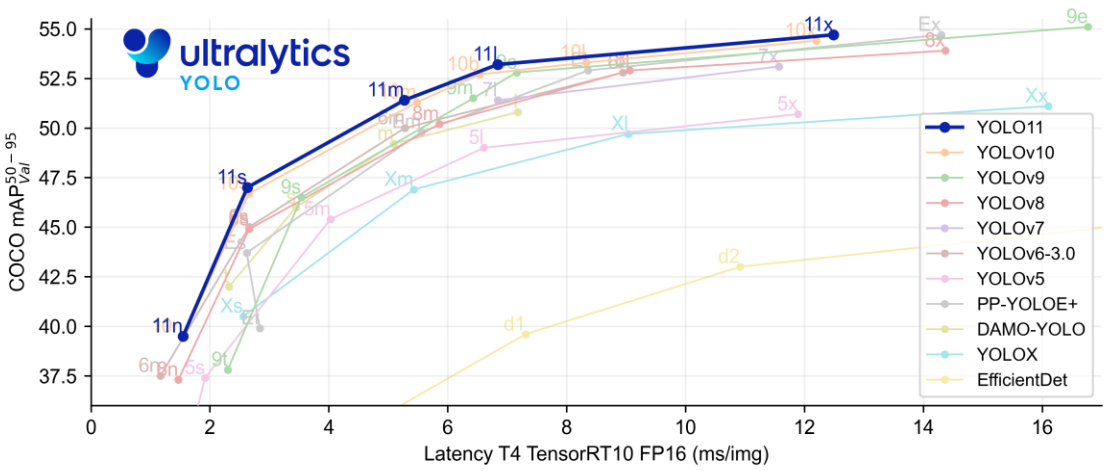


- To develop an automated system which evaluates road advertisement significance.
- To train an object detector which is capable of accurately detecting roadside billboards.
- To build a classification system which will be able to categorize drivers' gaze on billboards.

RECAP III: DATASETS FOR OBJECT DETECTION

- Mapillary Vistas Dataset
  - Large high resolution street object image dataset.
  - 25000 images available -> 9700 images used for object detector training.
  - An increase from 3000 to 9700 images.
- BillboardLamac Dataset
  - Smaller dataset focused on roadside billboards.
  - 1213 high resolution images used for fine-tuning the object detector.

RECAP IV: PRE-TRAINED MODELS



- Object Detection Tasks:
  - Comparison between robust models (YOLOv8) and the newest versions (YOLO11).
  - Training methods: Merging datasets (Seminar 1) vs base training with Mapillary Vistas and fine-tuning with BillboardLamac dataset (Seminar II).
- Classification Tasks:
  - Currently ***under development***: Trying different training strategies with YOLO and RESNET classification pretrained models.

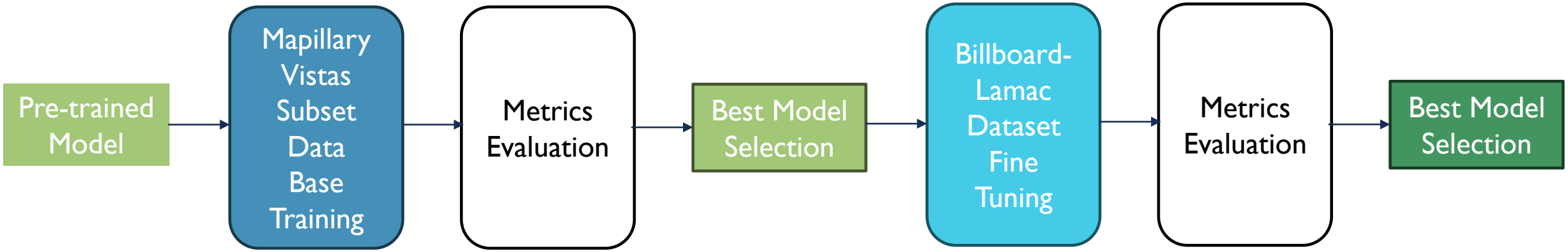
RESOURCES I

- Due to the limitations experienced previously, we stopped training in the cloud.
- Instead, the training of all models were performed using the DAI GPU Computing Servers.
- Due to its specifications, all trainings were done in Jupiter server.

RESOURCES II

- Jupiter GPU Computer Server:
  - CPU: AMD Ryzen 9 7950X (16 cores, 32 threads, 4.5–5.7 GHz)
  - RAM: 64 GB
  - SDD: 4 TB
  - GPU: 2 × NVIDIA GeForce RTX 4080 Super, each with16 GB of memory
- Due to its specifications, **parallel training** was possible!

TRAINING PROCESS



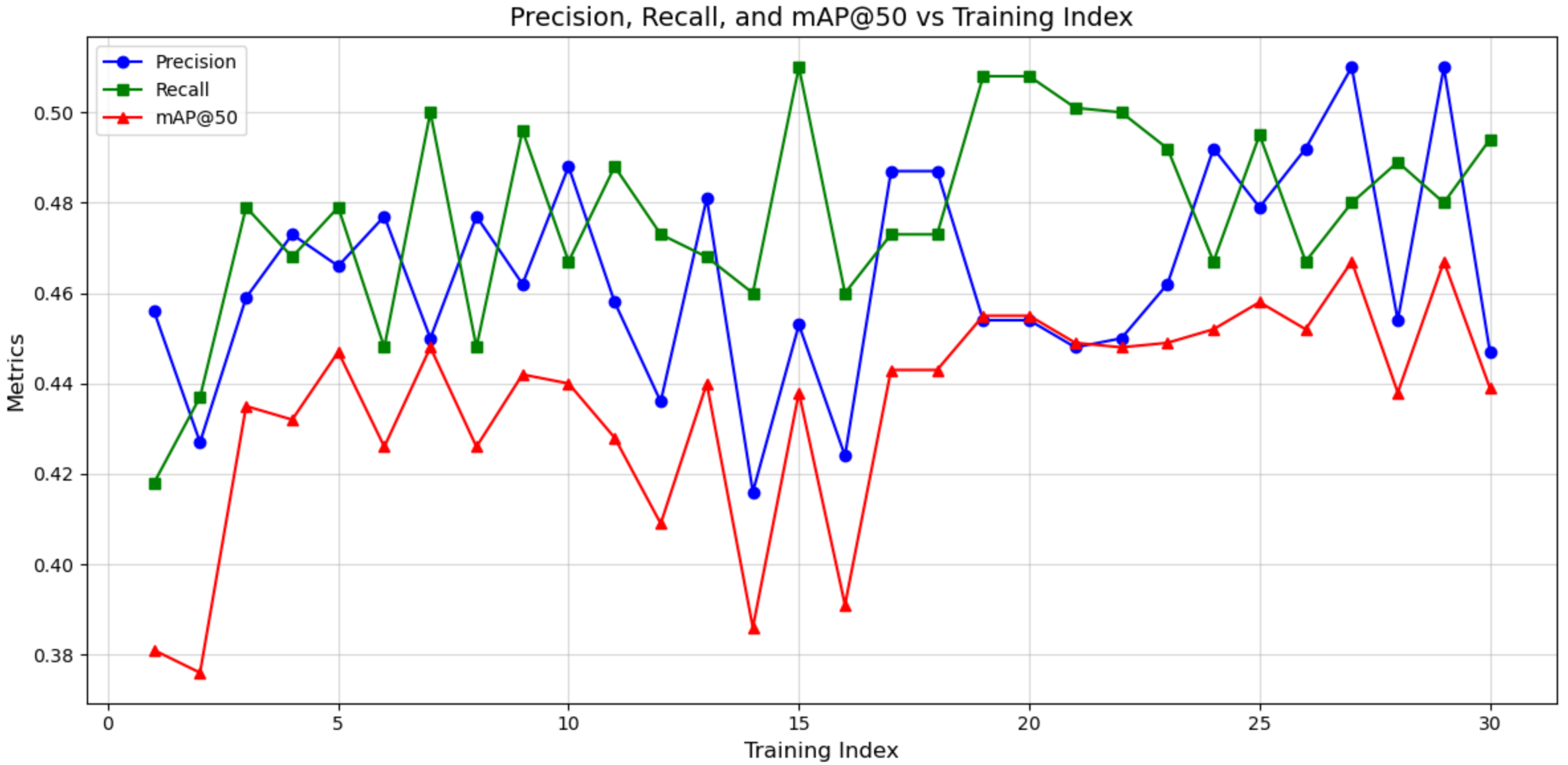
METRICS EVALUATION

- **Precision:** Measures how accurate the detector is when it identifies a billboard.
  - High precision means fewer false detections.
- **Recall:** Measures the detector's ability to find all billboards.
  - High recall means fewer missed detections.
- **mAP@50:** The average precision across all classes for detections with at least 50% overlap with the ground truth.
- **mAP@50-95:** Average precision across IoU thresholds ranging from 50% to 95%, calculated in steps of 5%.
- **Fitness:** A composite score balancing Precision, Recall, and mAP to represent overall model performance.

OBJECT DETECTOR TRAINING I: BASE TRAINING

- Training the object detector by using a pre-trained YOLOv8 model first on a subset of Mapillary Vistas Dataset.
- A total of 28 trainings trying different hyperparameters.
- Additional 2 trainings using the latest YOLO model (11 series) using the best set of hyperparameters found earlier.
- 2 best models selected, 1 per YOLO version.

Metric	19th Training (YOLOv8)	27th Training (YOLO11)
Precision	0.454	0.510
Recall	0.508	0.480
mAP@50	**0.455**	**0.467**
mAP@50-95	0.323	0.334
Fitness	0.337	0.347


Precision, Recall, and mAP@50 vs Training Index
Precision
Recall
mAP@50
Metrics
Training Index
0.38
0.40
0.42
0.44
0.46
0.48
0.50
0
5
10
15
20
25
30

OBJECT DETECTOR TRAINING I: BEST MODELS' HYPERPARAMETERS

Base Training YOLOv8 Model (19th)

- **Image size:** 1280x1280
- **Batch size:** 4
- **Epochs:** 100
- **Pre-trained YOLO model:** YOLOv8l (Large)
- **Data Augmentation:** flipud (0.5), fliplr (0.5), scale (0.5)
- **Optimizer:** Stochastic Gradient Descent (SGD)
- **Early Stopping:** Yes
- **Training Time:** 8 hours 40 minutes

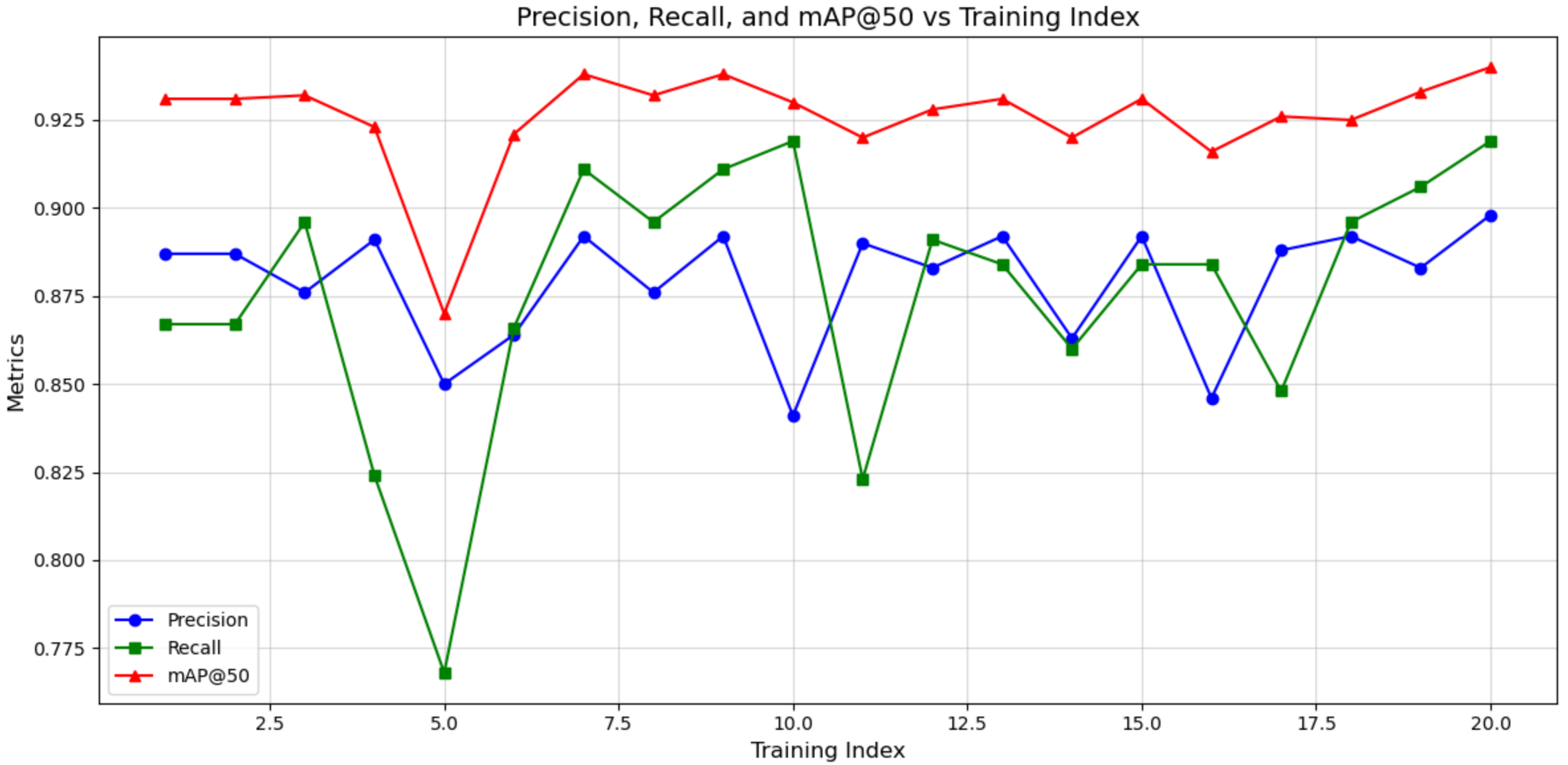
Base Training YOLO11 Model (27th)

- **Image size:** 1280x1280
- **Batch size:** 6
- **Epochs:** 150
- **Pre-trained YOLO model:** YOLO11l (Large)
- **Data Augmentation:** flipud (0.5), fliplr (0.5), scale (0.5)
- **Optimizer:** Stochastic Gradient Descent (SGD)
- **Early Stopping:** Yes
- **Training Time:** 13 hours

OBJECT DETECTOR TRAINING II: FINE-TUNING

Metric	3rd Training (YOLOv8)	10th Training (YOLO11)
Precision	0.892	0.898
Recall	0.884	0.919
mAP@50	**0.931**	**0.940**
mAP@50-95	0.770	0.794
Fitness	0.786	0.809

- The best models were fine-tuned using BillboardLamac dataset exclusively.
- A total of 20 trainings performed, 10 per each model.
- Similarly, different combination of hyperparameters tested.
- Best two models selected.


Precision, Recall, and mAP@50 vs Training Index
Metrics
Training Index
Precision
Recall
mAP@50
0.925
0.900
0.875
0.850
0.825
0.800
0.775
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

OBJECT DETECTOR TRAINING II: BEST FINE-TUNEDMODELS' HYPERPARAMETERS

Fine-Tuning YOLOv8-based Model (3rd)

- **Image Size:** 1280x1280
- **Epochs:** 100
- **Batch Size:** 4
- Rest of hyperparameters remained untouched (default)
- **Training time:** 1 hour 2 minutes

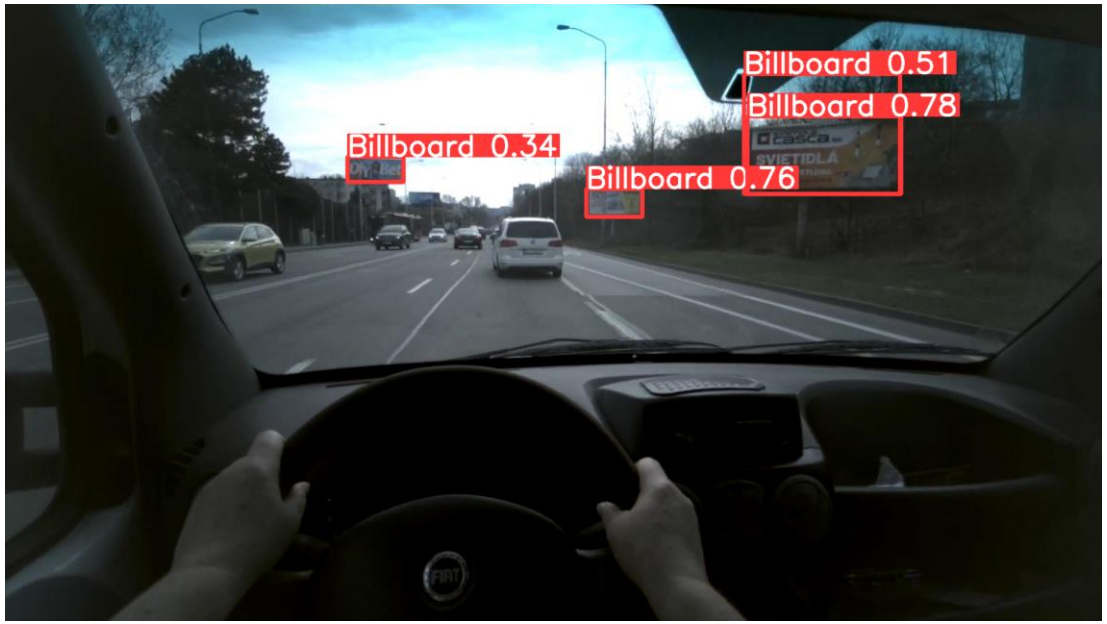
Fine-Tuning YOLO11-based Model (10th)

- **Image Size:** 1280x1280
- **Epochs:** 500 (Early Stopped Activated at 396 epochs)
- **Batch Size:** 6
- Rest of hyperparameters remained untouched (default)
- **Training time:** 4 hours 25 minutes

RESULTS I: RECAP OF THE BEST MODEL FROM PREVIOUS SEMESTER

Testing on Mapillary Vistas Images

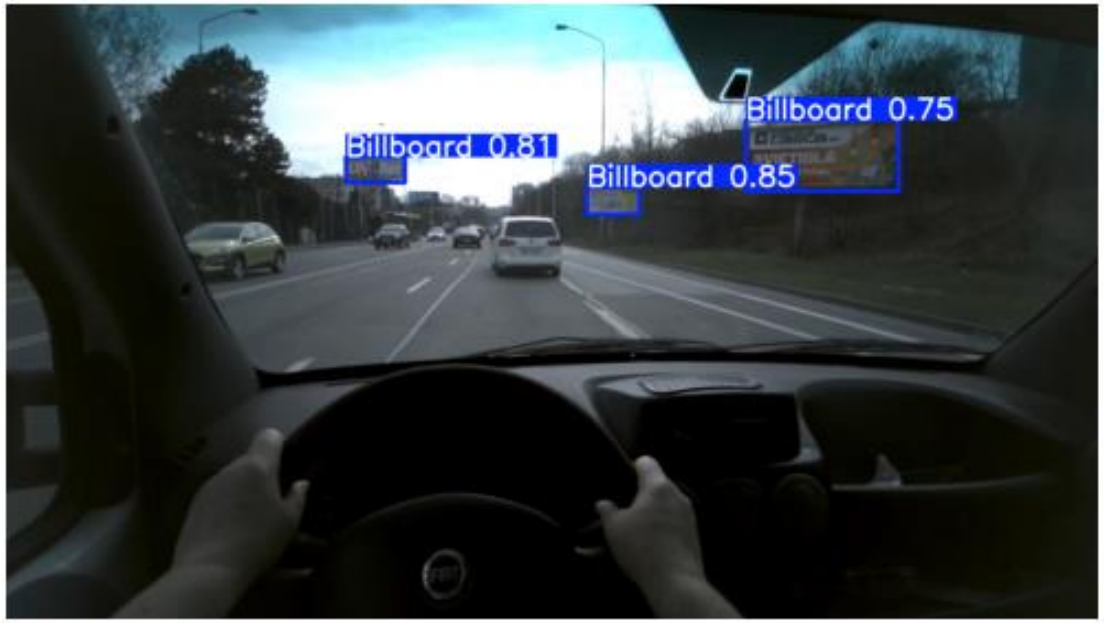
Testing on BillboardLamac Dataset Images



RESULTS: TESTING YOLOV8 BASE MODEL (WITHOUT FINE-TUNING)

Testing on Mapillary Vistas Images

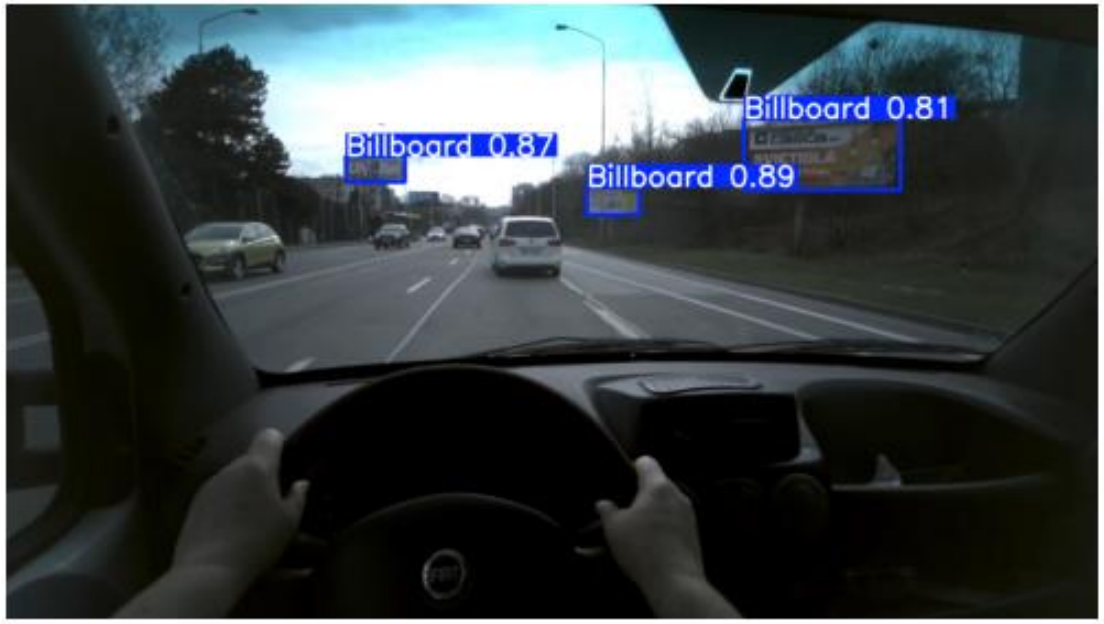
Testing on BillboardLamac Dataset Images



RESULTS: TESTING YOLO11 BASE MODEL (WITHOUT FINE-TUNING)

Testing on Mapillary Vistas Images

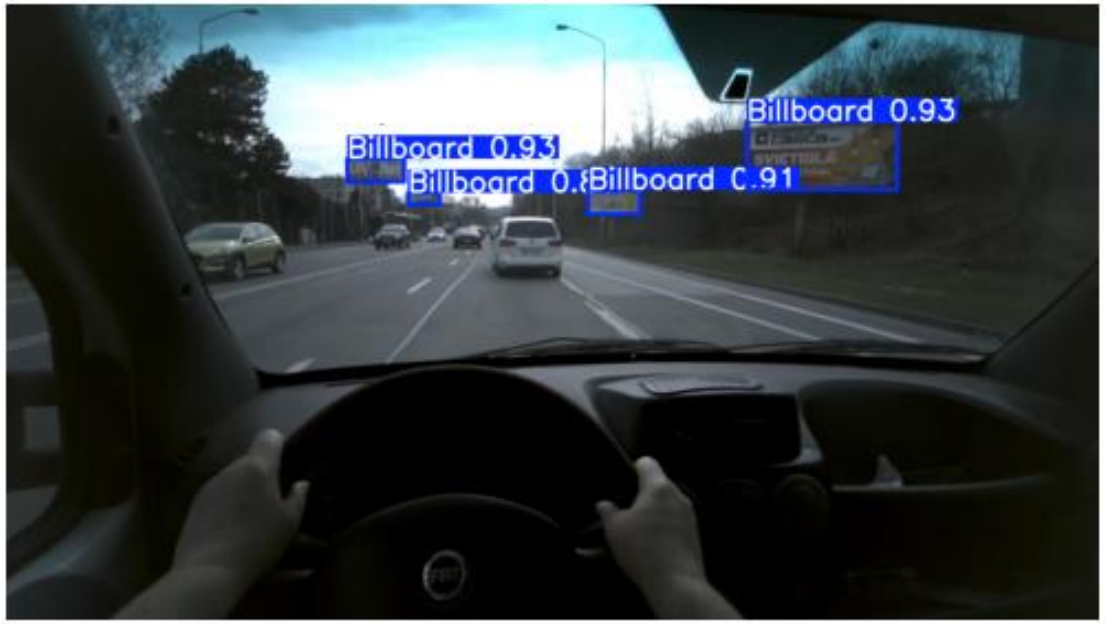
Testing on BillboardLamac Dataset Images



RESULTS: TESTING FINE-TUNED YOLOV8-BASED MODEL

Testing on Mapillary Vistas Images

Testing on BillboardLamac Dataset Images



RESULTS: TESTING FINE-TUNED YOLO11-BASED MODEL

Testing on Mapillary Vistas Images

Testing on BillboardLamac Dataset Images

CLASSIFICATION SYSTEM DEVELOPMENT

- BillboardLamac dataset includes a subset for the classification task.
- Data is divided into 4 predefined gaze duration classes: long, medium, none, and short.
- Each class has a set of images with IDs which link to additional information about billboards:
  - Billboard location in the image, size of the billboard, image in the billboard.

FURTHER STEPS

- **Object Detection Task:**
  - Review the best models.
  - Avoid overfitting.
- **Classification Task:**
  - Try different strategies to get the best information possible from billboards.
  - Train the classification models using YOLO and RESNET pretrained models, comparing them.

THANK YOU VERY MUCH FOR YOUR ATTENTION