International Symposium
"Navigating the Future of Traffic Management"
June 29 - July 3 2025, Athens, Greece

Training a YOLO-based model for speed limit sign recognition

Júlia Alves Porto[1], Dr. Apostolos Ziakopoulos[1], Daniel Felipe Lopez[2], Prof. George Yannis[1]

[1] National Technical University of Athens, Department of Transportation Planning and Engineering, Athens, Greece
[2] FRED Engineering, Rome, Italy

Introduction

Road characteristics are key to assessing and modelling road safety, influencing both crash outcomes and driver behavior. **Speed**, a core element of road design, is one of the five pillars of the Safe System approach (Finkel et al., 2020). While traffic sign recognition has a long research history (Eichner & Breckon, 2008; Torresen et al., 2004; Miyata, 2017), recent advances in object detection offer new opportunities for improvement.

Objective

The aim of this project is to train a You Only Look Once – **YOLO** model that recognizes **speed limit signs** from street-level imagery of the road.

Methodology - Training

Two different outsource and publicly available datasets were used to train the model. The first contains annotated images from an Italian project and includes 361 images captured by phone camera, with 720x1290 pixel-size, across 7 speed limit sign classes. The second one has annotated images of many traffic signs, from which the images containing only one speed limit sign were selected for the model training, resulting in 571 images with variable sizes, the most common being 400x300 pixels. **Four dataset configurations** (Table 1) were tested:

- Dataset 1 only;
- 400x300 images from Dataset 2 only;
- Both previous datasets resized to 640x640;
- Both original datasets in their original size, using YOLO default resizing.

Table 1: *Number of images per dataset*

Speed Limit	Dataset 1	Dataset 2	Dataset Resized	Dataset Grouped
5	0	27	27	30
20	7	0	7	7
30	151	25	177	183
40	25	205	235	301
50	118	24	146	153
60	15	29	45	64
70	23	18	44	49
80	2	130	133	175
90	20	58	78	92
100	0	55	56	76
Total	361	571	948	1130

Three versions of YOLO were used: **YOLOv5nu**, **YOLOv8n** and **YOLO11n**, the latest release at the time the research was conducted. We also experimented with, instead of relying solely on YOLO for both detection and classification of the speed limit signs, using the YOLO model only for the detection task and applying the cropped detected bounding-box as input to a pre-trained optical character recognition (OCR) model, **EasyOCR**, to read the numbers, therefore dividing the detection and classification tasks. With the combination of datasets, 24 models were trained, as presented by Table 2.

Table 2: *Trained models*

Model / Dataset	YOLOv5	YOLOv8	YOLOv11	YOLOv5 + OCR	YOLOv8 + OCR	YOLOv11 + OCR
Dataset 1	1	5	9	13	17	21
Dataset 2	2	6	10	14	18	22
Resized	3	7	11	15	19	23
Grouped	4	8	12	16	20	24

Methodology – Testing

The best models were later on tested on **video data** corresponding to **9 street-level recordings**, with approximately 10 minutes of duration each, 29.97 frames per second (fps) rate and 2304 x 1296 pixel-size. The video data was collected on 17 and 18 of Sepetember of 2024 and covers around **64 km of motorways** in Udine and Gorizia, Italy, and had manually coded speed limit information per 100m segments. For each video file, there is also a kml file with GPS coordinates per second. Figure 2 illustrates the video imagery.

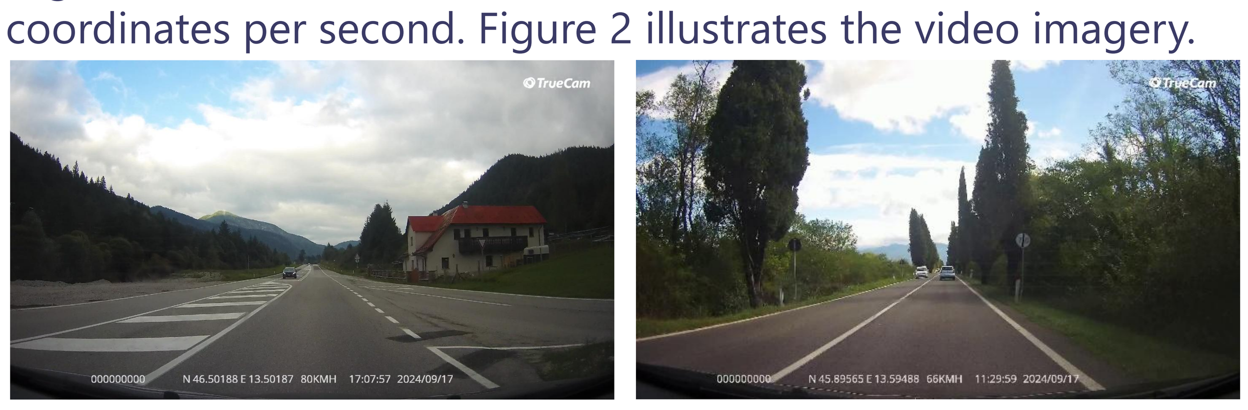

Figure 2: *Video imagery frame examples*

To extend detection to **cancelling speed limit signs**, selected frames from our video data were annotated using Roboflow. This new class was added to the detection task, using the best performing model as a basis. Due to class imbalance, cancelling sign images were cropped to match existing aspect ratios and augmented up to 3 times.

Results

Using only the YOLO model for both detection and classification tasks, results fell short of state-of-the-art benchmark, as can be seen in Table 3. Interestingly, **YOLOv5nu consistently outperformed the newer v8n and v11n models**.

In the two-stage YOLO + OCR pipeline, only the highest-confidence bounding box per image was passed to EasyOCR. Detection precision peaked at **0.983** for YOLOv5nu on Dataset 2 (Table 4). However, **overall classification performance was lower than expected**, likely due to **YOLO misclassifying similar shapes as speed limit signs**. OCR alone performed well on cropped ground truth signs, with accuracies above 90%—Dataset 1 reaching up to 0.976.

With the inclusion of speed limit cancellation signs, the **YOLOv5nu** model achieved **0.957** precision for standard speed limits and **0.969** for cancelling signs – promising results despite class imbalance (Table 5).

Table 3: *Results for YOLO-only models*

Dataset	YOLOv5	YOLOv8	YOLOv11
Dataset 1	0.713	0.685	0.241
Dataset 2	0.512	0.471	0.260
Resized	0.783	0.580	0.700
Grouped	0.768	0.701	0.567

Table 4: *Results for the YOLO + OCR pipeline*

Task	Detect	Classify	Detect	Classify	Detect	Classify	Classify
Model	YOLOv5	YOLOv5 + OCR	YOLOv8	YOLOv8 + OCR	YOLOv11	YOLOv11 + OCR	Only OCR*
Dataset 1	0.972	0.667	0.959	0.678	0.959	0.571	0.976
Dataset 2	0.983	0.462	0.979	0.667	0.978	0.818	0.938
Resized	0.979	0.708	0.973	0.715	0.967	0.655	0.938
Grouped	0.729	0.704	0.709	0.763	0.757	0.845	0.934

Table 5: *mAP results with cancel signs*

All	Class 0	Class 1
0.942	0.915	0.969
0.803	0.886	0.721
0.958	0.957	0.960
0.899	0.927	0.871

The best performing models were applied to the street-level videos using the following pipeline:

- Bottom-right mask – ROI = (0.2*h, h; 0.2*w, w);
- Gaussian blur with 5x5 kernel size;
- Slicing Aided Hyper Inference (SAHI) for input tiling (Akyon et al., 2021)

Discussion & Conclusion

To minimize false positives, only OCR labels appearing in three consecutive YOLO detections were accepted, with the most frequent label retained if multiple appeared. Detections were geocoded using video timestamps and GPS data (KML), enabling comparison with manually coded iRAP speed limits.

Cancelling signs were evaluated using a stricter filter: they had to appear for 30 consecutive frames. Also, speed limit signs were further restricted and required to appear for 5 consecutive frames. **Overall accuracy was of 0.94, while recall and F1-Score achieved, respectively, 0.73 and 0.82**. Figures 3 and 4 bring visual outcomes for the YOLOv5nu model trained with the resized dataset and cancellation signs.

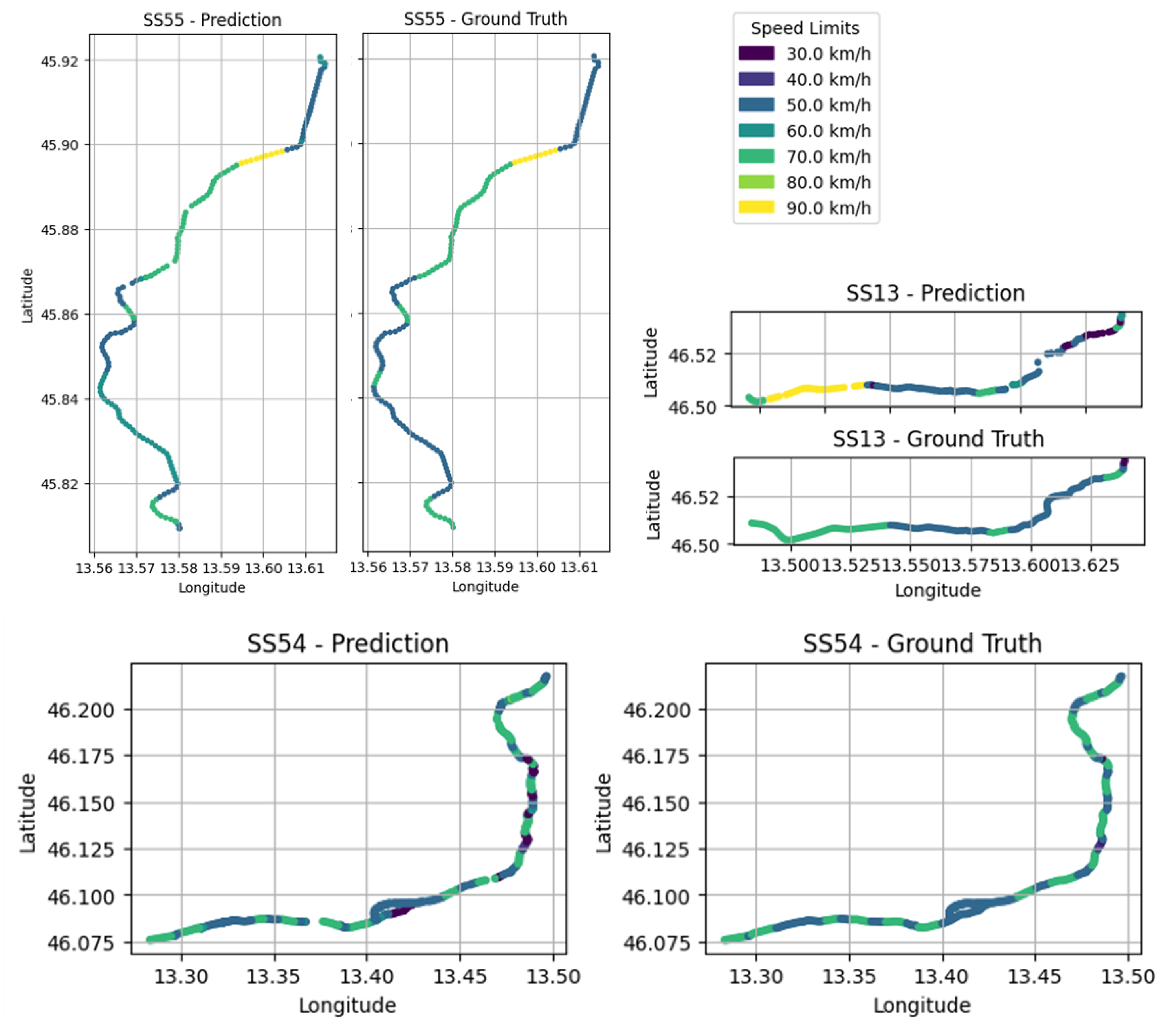

Figure 3: *Visual comparison of predicted vs ground-truth data*

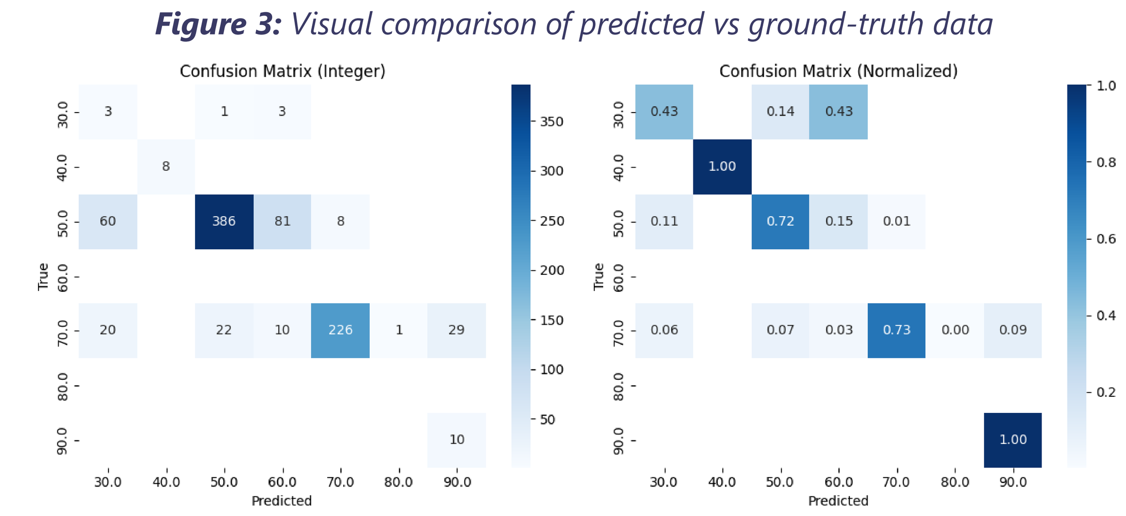

Figure 4: *Confusion matrices with raw (left) and normalized (right) values*

The proposed framework offers a **simple yet effective** pipeline for a key road safety task: identifying speed limits along road segments. The model can be further improved by addressing **temporary signs** and distinguishing them from the permanent ones. Future research should also aim at **optimizing computational efficiency**. The current YOLO + OCR pipeline runs at 0.3 ms per frame on a Tesla T4 GPU but slows to 170 ms on CPU.

Acknowledgments

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590.

IVORY

European Commission

Contact Information:
Júlia Alves Porto, PhD Candidate & Researcher NTUA
Department of Transportation Planning and Engineering
Email: julia_porto@mail.ntua.gr