Training a YOLO-based model for speed limit sign recognition

Júlia Alves Porto
PhD Candidate
Department of Transportation Planning and Engineering
National Technical University of Athens, 5 Heroon Polytechniou Str., Athens GR-15773, Greece
Email: julia_porto@mail.ntua.gr

Apostolos Ziakopoulos
Senior Researcher
Department of Transportation Planning and Engineering
National Technical University of Athens, 5 Heroon Polytechniou Str., Athens GR-15773, Greece
Email: apziak@central.ntua.gr

Daniel Felipe Lopez
Transport Engineer and Data Scientist
FRED Engineering, 31 Via Treviso, Rome RM-00161, Italy
Email: daniel.lopez@fredengineering.onmicrosoft.com

George Yannis
Professor
Department of Transportation Planning and Engineering
National Technical University of Athens, 5 Heroon Polytechniou Str., Athens GR-15773, Greece
Email: geyannis@central.ntua.gr

Word Count: 3173 (excl. bibliography)

Submitted 06-03-2025

ABSTRACT
Road crashes are an endemic problem worldwide, and it is the number one cause of death for young people aged from 5 to 29 years old. Speed limit is a crucial information for assessing safety and safety requirements for a road segment. Following the International Road Assessment Programme (iRAP) methodology, it can be registered based on visual imagery input. In this research, we have trained three different versions of You Only Look Once (YOLO) models - YOLOv5nu, YOLOv8n and YOLO11n - to automatically identify and classify speed limit signs, using two public datasets. The best mean average precision achieved was of 0.783 so, to improve accuracy, we have retrained the YOLO models to identify the speed limit signs and classification was made using Optical Character Recognition (OCR) model. With this combination, the best mean average precision was set to 0.845, while standalone mean average precision for OCR reaches 0.976 when applied to ground truth cropped images. After the training, the model pipeline was tested on real video data imagery covering 64 km of northern Italy, from the provinces of Udine and Gorizia, and a coding pipeline was used to convert the frame-by-frame automated detection into timestamps, which were further associated with their respective geographic locations and results were compared with manually coded iRAP data. Overall precision of the model was of 89% for the test area, setting it close to state-of-the-art results. Future research steps include training the model to differentiate cancelling and temporary speed limit signs for a more flexible approach.
Keywords: Speed limit, iRAP, YOLO, OCR

INTRODUCTION

Road crashes are an endemic problem worldwide, and it is the number one cause of death for young people aged from 5 to 29 years old (WHO, 2023). On 2010, the United Nations implemented the Decade of Action for Road Safety, with the goal to halve the deaths caused by road crashes around the world. Although the goal was not achieved, there have been some improvements regarding road safety, and it is one of the United Nation's targets for the 2030 Agenda to promote safer roads.

To put road safety as a priority has brought strength to the concept that stakeholders can take a part in improving the transportation safety environment. That is a point of view sustained by the Safe System or Vision Zero approach, originally advocated by Swedish researchers (Tingvall and Haworth, 1999). Vision Zero sustains that the road environment can be built to mitigate consequences of human error when conducting a vehicle. It also sustains that road safety mustn't rely solely on post-factum occurrences, as is the case of crash data-based studies, and can be done proactively using proxy attributes for safety modelling, or even surrogate occurrences for crash data.

Among the attributes used for safety modelling, road characteristics play an important role. Horizontal road curvature, for example, can be used as input data for road safety proactive assessments and was found to have mixed effects on road safety (Wang et al., 2013). It is one of the 78 attributes used by the International Road Assessment Programme (iRAP) for road star rating (IRAP, 2024). Since many inputs are used for road safety modelling, the automation of their assessment can help to speed up further safety studies.

Speed plays an essential role in determining the road safety, and it is one of the five defining elements for a Safe System as defined by the Federal Highway Authority of the United States (Finkel et al., 2020). The project speed is an elemental part of road design, since most of the road's geometric parameters are based on it.

In this project, we train a model to recognize speed limit signs within a street-level image of the road. Automatic recognition of speed limit signs has already been attempted by other researchers, achieving precision of over 90% of overall correctness for small datasets (Eichner and Breckon 2008; Miyata 2017; Torresen et al., 2004). We believe that even higher results can be achieved using a state-of-the-art object detection model, the You Only Look Once (YOLO) (Jocher et al., 2023). Although research has been made with earlier versions of the YOLO model (Juanola 2019), the Ultralytics team has released a new version (YOLO11) and, as far as this author knows, it has not been tested for speed limit sign recognition yet. Also, we propose our own model by combining YOLO object detection with an Optical Character Recognition model. The models are trained and validated using two public datasets and tested on real video data collected from northern Italy.

METHODS

Model Training

Two different datasets were used to train the models, both publicly available on the Kaggle platform. The first one contains annotated mages of speed limits from an Italian project and contains 361 images of 7 different speed limit signs, captured by phone camera, of pixel-size 720x1280 (height x width). The second one has annotated images of traffic signs, including speed limits. For the images with one single speed limit sign, the class identification was manually relabeled to reflect the value of the speed limit, and that resulted in 571 images with variable sizes, the most common being 400 x 300 pixels (height x width).

Since the size of the image is an important input for the YOLO model and the two datasets had different sizes, we attempted four different strategies for training the model: using only dataset 1; using only the images with 400x300 size of dataset 2; resizing both datasets to 640x640 pixels (called 'Resized'); and using both datasets with their original image (called 'Grouped'). In the latter case, YOLO is automatically prepared to handle sizes up to 50% further than the 'assigned' image size when calling the model. Therefore, for each dataset we had the total number of images stated on Table 1.

TABLE 1 Training Datasets

Speed Limit	Dataset 1	Dataset 2	Resized	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

For each of the datasets, we trained a YOLO model for detection and classification, where each speed limit was labeled as a different class, and a YOLO model only for detection, where the speed limit signs were labeled as one single class, and later combined with an OCR model for classifying the sign according to the actual speed limit. The yolo models used were 'YOLOv5nu', 'YOLOv8n' and 'YOLO11n'. So, in total, we trained 24 different models, as summed up 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

Model Testing

Video data corresponds to 9 street-level recordings, with approximately 10 minutes of duration each. They all have 29.97 frames per second (fps) rate, 2304 x 1296 pixel-size (width x height) and were surveyed on the 17th and 18th of September of 2024. It covers 64 km of motorways from the provinces of Udine and Gorizia. On the video, there is a logo on the top-right corner, and in the bottom of the frame it states the geographical position in degrees, the velocity and the time of the recording in the format hh:mm:ss YYYY/MM/DD. There is also a number on the bottom left of the image, which appears to be a sequential number and was treated as noise, just like the portion of the car visible in the recording. Depending on the light conditions, the reflection of the vehicle is also visible. A random frame is shown on Figure 1 to illustrate the video imagery.

Additionally, for each video file, there was a correspondent kml file with the GPS longitude and latitude data for each second of the video.

Figure 1 Random frame from the video data

RESULTS

Surprisingly, when using only the YOLO model to detect and classify the speed limit signs, the YOLOv5nu presented the best performance on the training dataset, as can be seen by Table 3. However, when the best performing models were applied to the test dataset, it was evident that the model was overfitting for the 50 km/h speed limit class. Although the detection rate of the speed limit sign was satisfactory, predicting the correct speed limit class is essential for actual usability of the model. Therefore, the use of only YOLO detection model for speed limit classification was dicarded.

TABLE 3 Mean Average Precision results for the YOLO model

Model / 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

For the combination of YOLO with OCR, we recorded results for the validation dataset both for the detection task and for the overall classification task (Table 4). Again, YOLOv5 showed the highest mean average precision for object detection for most of the datasets used for training, however, higher precision values were found using the YOLO11 and YOLOv8 models combined with OCR. It was indeed surprising that the best OCR performances were not matched with the best detection performances. Furthermore, it is important to mind that, for our specific use, false positives are more dangerous than false negatives. Since we are using video data, a speed limit sign might be missed from a frame or two, but to have a wrong speed limit value input would imply a misclassification of the road segment.

TABLE 4 Mean Average Precision results for the YOLO + OCR pipeline

Task	Detection	Classification	Detection	Classification	Detection	Classification	Classification
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

* Results for ground truth bounding boxes

When the best model was applied to our test video dataset, the apparent low classification precision was identified as an outcome of the presence of many false positives in the YOLO detection model. Since the YOLO detection model was trained with only one class, it was overpredicting speed limit signs and misidentifying other signs and round objects as speed limit signs. However, when false positives were detected, usually there was no numerical digit that the OCR model could misidentify also, thus resulting in many of the detected YOLO objects being labeled with "None" and decreasing the overall metrics of the OCR model. Hence, for each dataset, the OCR model was processed for the cropped ground truth bounding box for each label, allowing to measure the standalone precision of the OCR model for a correctly detected speed limit sign.

When applying the model pipeline to real video data, some pre-processing and post-processing were applied to ensure better results:

- Pre-processing:
  - a mask was applied to the video data, so the model was processed only for the 80% "bottom-right" area;
  - Gaussian blur with a 5x5 kernel size was applied to smoothen the video resolution;
  - Slicing Aided Hyper Inference (SAHI) was used to divide the input frame in windows of the same size as the training dataset (Akyon et al. 2021).
- Post-processing:
  - a csv file was created with the prediction results, with the columns:
    - "Frame": the number of the current frame;
    - "x1", "y1", "x2", "y2": the bounding box corners for the YOLO detection output;
    - "timestamp": the second of the output video relative to the frame;
    - "OCR_Label";
    - "YOLO_Confidence" and "OCR_Confidence".
  - Rows were only added to the csv file if there was a YOLO bounding box predicted.
  - A speed limit change was considered valid if a valid "OCR_Label" (that is, with a value less than 150), was detected for 3 consecutive rows.
  - In case different labels were detected, the true speed limit was considered the most frequent label detected.

To allow comparison with the manually coded data, the output was geocoded using the reference kml videos, and the geopandas function 'ffill' was used to fill the speed limit information with the previous information until a change was recorded. Finally, data was smoothened to 100 meters segment, always keeping the highest computed speed limit value as the reference value for the segment, and speed limits of values equal or less than 30 were unified in a single class.

A visual comparison of the automatic pipeline results and the manually coded data can be seen on Figure 2. On Figure 3, the confusion matrices – with numerical and normalized values – are presented.

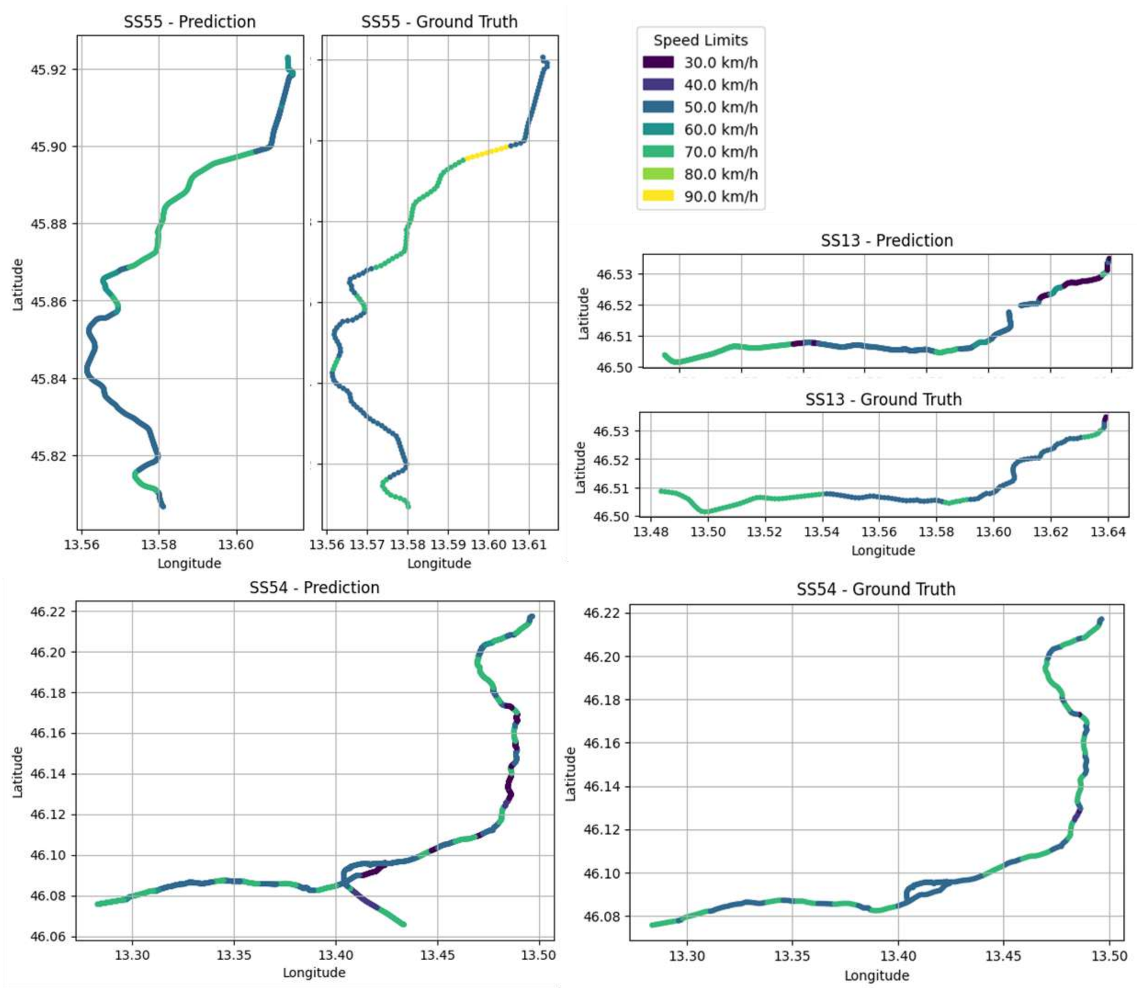

Figure 2 Visual comparison of predicted and ground truth speed limits, per road

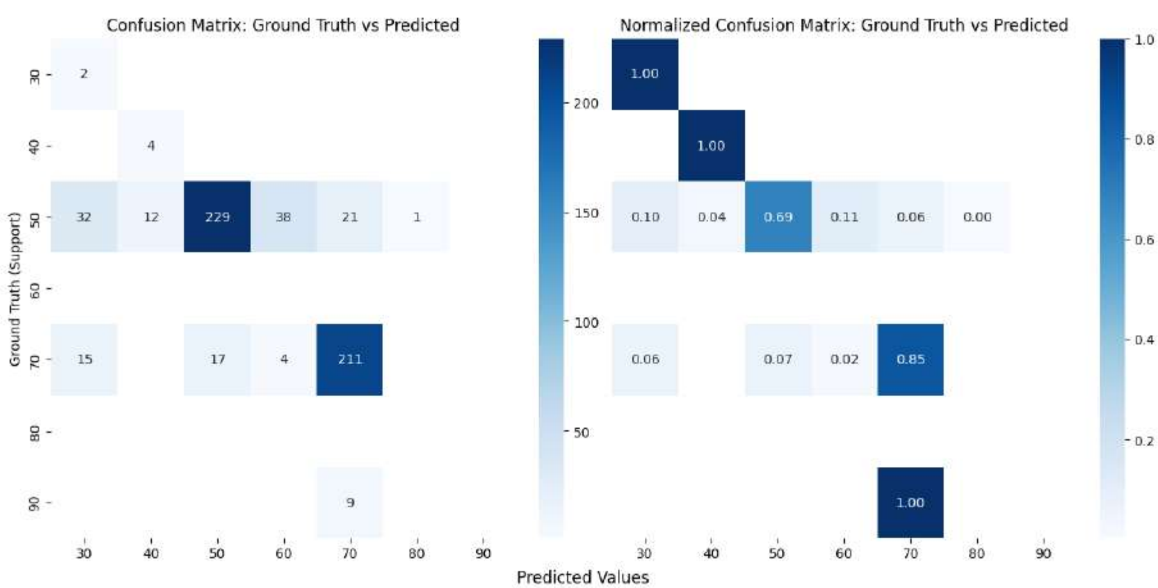

Figure 3 Confusion matrices of predicted and ground truth speed limit values

The model was run on video data using three different computer interpreters, depending on availability: Google Colab T4 and NVIDIA RTX 2080 GPU, with an approximate computing time of 0.6 second per frame; and Intel® Core™ Ultra 5 155U CPU, with an approximate computing time of 2.4 seconds per frame.

DISCUSSION

The overall precision, recall and f1-score of the proposed model compared to iRAP coding data were respectively of 0.89, 0.75 and 0.80. This shows great potential for combining state-of-the-art and freely available models to aid on iRAP coding. As of this moment, only one company is globally accredited for automated speed limit recording (iRAP, 2025).

The proposed method is also valuable for its generalizability by using a single class (“Speed limit sign”) as a detection target and an optical character recognition model for speed classification. Although other studies have achieved higher attribute-wise detection accuracy (Sanjeewani and Verma, 2021), their outcomes were for specific classes and may encounter an accuracy decrease when introducing new speed limit values. The difficulty in differentiating such similar objects can also be observed by the outcome of Jan et al. (2018), who proposed a convolutional neural network for identifying road attributes and, out of 5 mislabeled objects in their test dataset, 3 were speed limit signs incorrectly classified as one another.

As for the studies with focus solely on speed limit classification mentioned in the Introduction section (Eichner and Breckon 2008; Miyata 2017; Torresen et al., 2004), although results in this research didn’t reach the same level of accuracy, it should be stated that, in this research, final comparison is made with iRAP coding results, so it considers, for example, correctly recognized speed limit signs as false positives if they correspond to a temporary change in the roads settings as opposed to simply recognizing the image. Therefore, results aren’t directly comparable. It can be seen, though, that, if a sign is correctly detected, the OCR model has a precision rate comparable to these state-of-the-art methods.

Another interesting finding of the research was that, for most of the datasets used for training, the YOLOv5 outperformed the most recent versions, and that overall performance for detection for all models was very high, achieving over 95% precision for all datasets with the same input size. These findings corroborate the importance of standardizing input size for YOLO detection and that even older versions of YOLO are just as powerful and reliable as more recent ones.

Some limitations when defining the model have contributed to decrease accuracy overall. For example, canceling speed limit signs were not used for training, thus, there was no prediction in the model code to interprete those signs. The cancelling speed limit signs are circular signs with a white background, gray digits and a black diagonal stripe, and it cancels the previous limit and automatically sets it to the general speed limit of the road category. In the dataset, it is present in the location seen in Figure 4(a), and the model was not trained to recognize it.

Another limitation when comparing to iRAP coding is that temporary changes are not supposed to be accounted for. The model needs to be specifically trained to differentiate between permanent and temporary speed limit signs, either by their color or location (Figure 4b).

(a) (b)

Figure 4 Limitations of the model pipeline: (a) cancellation signs and (b) temporary signs

Also, the 60 km/h is not a common limit in Italy (Automobile Club d'Italia, 2025). It was observed that the 50 km/h was many times misinterpreted by the model as a 60 km/h sign, and that could easily be corrected manually, as a post-processing step. However, it is possible that 60 km/h exists, and, in other regions of the world, it is a more common sign. To keep the model generalizable, it was opted not to change this manually.

For future steps, a more detailed look will be taken over the limitations of this research, training the model specifically to differentiate cancellation and temporary speed limit signs. Other possible directions are: using the segmentation module of YOLO instead of the detection one to differentiate signs and background; annotate the speed limit signs to increase available datasets; revisit criteria for geographic location and comparison between the manual coding and the automated output; test different model architectures to reduce computational time.

CONCLUSIONS

Speed is well reported as a major factor in calculating road safety, being directly connected to the chances of one getting involved in a crash and the severity risk of possible crashes. It is the basis for all geometric attributes for the road to ensure security both for the drivers and for whomever might access the road area. It is also a main attribute for iRAP coding and star rating methodology.

In this research, public datasets were used to train a model for detection and classification of speed limit signs using YOLOv5nu, YOLOv8n, YOLO11n and a combination of those models with OCR. The best overall performing model (YOLOv11 + OCR) was tested in a real video dataset covering 64 km of northern Italy, and results were compared to manual iRAP labelling. An overall precision of 89% was found using the automated pipeline, and some of the mislabels can be attributed to limitations of the model rather than prediction errors.

Although the best performing overall model for classification was YOLOv11 + OCR (0.845), the highest speed limit sign detection was found with YOLOv5nu (0.983) rather than with YOLO11n (0.978). In future research, both models will be compared for the best fine-tuning of the results.

ACKNOWLEDGMENTS

This research is based on work carried out within the IVORY project. The project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101119590.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: J. Porto; data collection: D. Lopez; analysis and interpretation of results: J. Porto, D. Lopez, A. Ziakopoulos; draft manuscript preparation: J. Porto, A. Ziakopoulos; revision of the manuscript and supervision: G. Yannis. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. World Health Organization (WHO). *Global Status Report on Road Safety 2023*. 1st ed. Geneva. 2023.

2. Tingvall, Claes, and Narelle Haworth. Vision Zero - An Ethical Approach to Safety and Mobility. Presented at the 6th ITE International Conference Road Safety & Traffic Enforcement: Beyond 2000, Melbourne, 6-7 September 1999.

3. Wang, Chao, Mohammed A. Quddus, and Stephen G. Ison. The Effect of Traffic and Road Characteristics on Road Safety: A Review and Future Research Direction. *Safety Science* 57: 264–75. doi:10.1016/j.ssci.2013.02.012. 2013.

4. International Road Assessment Programme (iRAP). *iRAP Coding Manual Version 5.4 – Drive on Right Edition (English)*. 2024

5. Finkel, E., McCormick, C., Mitman, M., Abel, S., Clark, J. *Integrating the Safe System Approach with the Highway Safety Improvement Program: An Informational Report.* FHWA-SA-20-018. Fehr & Peers; Leidos, Inc.; U.S. Federal Highway Administration, Office of Safety. https://rosap.ntl.bts.gov/view/dot/58031. 2020. Accessed March 1, 2025.

6. Eichner, Marcin L., and Toby P. Breckon. Integrated Speed Limit Detection and Recognition from Real-Time Video. In 2008 IEEE Intelligent Vehicles Symposium, Eindhoven, Netherlands: IEEE, 626–31. doi:10.1109/IVS.2008.4621285. 2008.

7. Miyata, Shigeharu. Automatic Recognition of Speed Limits on Speed-Limit Signs by Using Machine Learning. *Journal of Imaging* 3(3): 25. doi:10.3390/jimaging3030025. 2017.

8. Torresen, J., J.W. Bakke, and L. Sekanina. Efficient Recognition of Speed Limit Signs. In Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749), Washington, WA, USA: IEEE, 652–56. doi:10.1109/ITSC.2004.1398978. 2004.

9. Jocher, Glenn, Jing Qiu, and Ayush Chaurasia. *Ultralytics YOLO*. https://github.com/ultralytics/ultralytics. 2023. Accessed March 1, 2025.

10. Juanola, Martí Sánchez. *Speed Traffic Sign Detection on the CARLA Simulator Using YOLO.* Master's thesis. Universitat Pompeu Fabra. 2019.

11. Akyon, Fatih Cagatay, Cemil Cengiz, Sinan Onur Altinuc, Devrim Cavusoglu, Kadir Sahin, and Ogulcan Eryuksel. *SAHI: A Lightweight Vision Library for Performing Large Scale Object Detection and Instance Segmentation*. Zenodo. doi:10.5281/zenodo.5718950. 2021.

12. Automobile Club d'Italia. *Everything You Need to Know About Driving in Italy: Road Rules, Tips and Useful Information.* https://www.italia.it/en/italy/things-to-do/tutto-quello-che-ce-da-sapere-per-guidare-in-italia-regole-stradali-consigli-e-informazioni-utili. Accessed March 1, 2025.

13. International Road Assessment Programme (iRAP). *iRAP Accreditation*. https://irap.org/accreditation/. Accessed March 6, 2025.

14. Sanjeewani, P., Verma, B. Single class detection-based deep learning approach for identification of road safety attributes. *Neural Computing and Applications*: Volume 33, Issue 15, 9691-9702. doi:10.1007/s00521-021-05734-z. 2021.

15. Jan, Z., Verma, B., Affum, J., Atabak, S., Moir, L. A convolutional neural network based deep learning technique for identifying road attributes. *2018 Internacional Conference on Image and Vision Computing New Zealand (ICVNZ)*:1-6. doi:10.1109/IVCNZ.2018.8634743. 2018.