Joint CSCE Construction Specialty & CRC Conference 2025
Conférence conjointe spécialisée en construction de la SCGC et CRC-2025

Montreal, Quebec
July 28-31, 2025 / *28-31 juillet 2025*

Developing a Deep Learning Model Framework to Automate Speed Limit Sign Detection and Classification for the iRAP Elements

Zakariya Gadi, Rebecca Dziedzic, and Luis Amador

Building, Civil, and Environmental Engineering, Concordia University, 1515 Ste-Catherine, Montreal, Quebec H3G 2W1, Canada

Abstract: The International Road Assessment Program (iRAP) is a global initiative aimed at enhancing road safety through systematic evaluations of road conditions and infrastructure. Traditionally reliant on manual visual assessments, recent advancements are shifting toward automating these processes to improve efficiency and scalability. This study presents a deep learning (DL) framework utilizing the YOLOv8 (You Only Look Once, version 8) model to automate the detection and classification of speed limit signs in real time, a critical component of iRAP assessments. The main technical contributions include the development of a customized, real-time detection framework, creating a labeled dataset of 724 images across 14 distinct speed limit classes, and a comparative evaluation demonstrating YOLOv8's superior performance over YOLOv5. The proposed model achieved an F1 score of 0.82 and a mean average precision (mAP) of 0.71 over Intersection over Union (IoU) thresholds from 0.5 to 0.95, outperforming YOLOv5 (F1 score 0.78, mAP 0.64). These results demonstrate robust detection performance under diverse conditions, supporting the feasibility of vision-based automation within infrastructure safety assessment processes. The framework offers a scalable and efficient solution for automating key components of road safety assessments, minimizing reliance on manual evaluations, and advancing intelligent transportation systems. Its fast, one-stage detection enables real-time assessment, significantly reducing processing time and allowing quicker identification of speed limits and potential hazards, with meaningful implications for strengthening iRAP's operational capacity and supporting data-driven road safety improvements in the domain.

1. INTRODUCTION

Roads are essential to the economic and social fabric. Ensuring road safety is paramount for the efficiency and sustainability of the transportation system. Despite efforts to improve road safety worldwide, statistics continue showing a high incidence of fatal and serious road accidents due to safety shortcomings of the road environment. Canada reports nearly 1,800 road traffic deaths each year, with a fatality rate of 5.2 per 100,000 people and over 150,000 non-fatal road injuries annually, imposing a substantial burden on its healthcare system (Transport Canada 2021).

Assessing road safety enables authorities to identify risks and prioritize improvements effectively. The International Road Assessment Program (iRAP) is a widely adopted tool used across various regions to evaluate road safety systematically. The iRAP (2021a) utilizes scores to qualify safety deficiencies and a star rating to facilitate its interpretation, ranging from 5 stars (the safest road) to 1 star (a very unsafe road). The scores are built through several attributes, including fixed objects along the road, the distance between them, the areas defined by horizontal signage, and the traffic flow density. The traditional methods of collecting data for the iRAP attributes, primarily involving manual surveys and in-person assessments, present several challenges. These conventional methods are time-consuming and require significant human resources. Moreover, they are prone to human error, leading to inconsistencies in the collected data, and can substantially impact any decisions made based on this information. Furthermore, the costs associated with deploying survey teams and the logistical challenges of accessing remote or dangerous areas make these methods less feasible (Leduc 2008). Another significant drawback is the limited frequency of data collection due to the high costs and time required, leading to outdated or insufficient data. This can severely impact road safety interventions, making them less effective and potentially increasing the risk for road users. According to the Canadian infrastructure report (Lei et al. 2020), 71% of municipalities collect data on their roads at least once every five years. The low data collection rate is primarily due to the complexity of manual detection methods (Lei et al. 2020). Due to these constraints, over 30% of the roads in developing countries have not been assessed in the last five years (World Bank 2020).

This research aims to develop a robust framework to automate detecting and classifying speed limit signs for iRAP real-time assessments. The proposed framework utilizes the YOLOv8 one-stage detection model, optimized explicitly for real-time object detection tasks with faster inference speeds than previous models, such as two-stage object detection (Mask R-CNN) by (Tabernik and Skočaj 2019).

2. LITERATURE REVIEW

According to Abduljabbar et al. (2019), AI-based road element assessment has supported governments and agencies in maintaining accurate road infrastructure databases, which is essential for iRAP scoring. Several studies target critical iRAP elements, such as traffic signs, lane markings, pedestrian crossings, and roadside barriers, due to their direct impact on road safety. Advanced methods like YOLO, CNNs, and semantic segmentation models are widely used for automated detection (Mehta et al. 2019) (Tabernik and Skočaj 2019). UAVs (Unmanned aerial vehicles) and satellite imagery further improve the detection of lane separations and pedestrian crossings (Mahmud et al. 2021) (Brkić et al. 2023). While detailed features like roadside barriers require LiDAR for spatial accuracy (Brkić et al. 2022). more complex elements, such as obstacles and intersections, demand deep learning models capable of handling occlusions (Brkić et al. 2023) (Brkić et al. 2022). These targets often require multi-task learning models or deep segmentation techniques capable of handling complex scenarios and occlusions. In contrast, straightforward elements like simple lane markings, stop signs, and basic road delineations are consistently more straightforward to detect (Tabernik and Skočaj 2019). These features exhibit minimal variation in appearance and are well-suited for standard computer vision methods like CNNs and traditional image segmentation models. Additionally, automated frameworks employing RGB cameras or satellite imagery further enhance the efficiency of detecting these simpler features, making them reliable candidates for automation in large-scale road assessments (Lin et al. 2023).

The datasets used in the reviewed studies for iRAP-related road element detection exhibit notable variation in sources, the number of images for training and testing, pixel sizes, and image resolutions. Satellite imagery datasets, such as those from Pleiades Neo 3, provide a resolution of 512×512 pixels, with 2602 images (Brkić et al. 2023) and 5846 images (Miler et al. 2023). UAV-based datasets, such as those captured using a Mavic 2 UAV, originally feature resolutions of 5472×3648 pixels with 300 images (Mahmud et al. 2021). In traffic sign detection studies, the Belgium Traffic Sign Dataset (BTSD) (Mehta et al. 2019) comprises 7277 images with a standard size of 64×64 pixels, allowing for efficient multi-class classification across 62 categories. The DFG (consulting company) traffic-sign dataset (Tabernik and Skočaj 2019), collected via vehicle-mounted cameras, consists of 8697 images, offering resolutions ranging from 1920×1080 to 720×576 pixels, with variations reflecting urban and rural road conditions. In studies leveraging RGB-D cameras, images were captured at a resolution of 1280×720 pixels (Lin et al. 2023). LiDAR datasets (Brkić et al. 2022), used for roadside severity-object detection, consisting of 5000 images of road cross sections, provide orthophotos with spatial resolutions of 1 cm × 1 cm, rendered into images sized at 2500×4000 pixels. These datasets are divided into 75%-25% splits for training and validation purposes. Challenges across these datasets include lighting inconsistencies, occlusions (e.g., from vegetation or vehicles), and imbalances in class representation, particularly for rare road features. To enhance the effectiveness of datasets, researchers recommend strategies such as data augmentation (e.g., cropping, rotation, brightness adjustment), synthetic data generation to address rare classes, and integrating multi-source data to improve diversity and coverage. These measures balance computational efficiency with data quality for robust and scalable iRAP assessments.

The reviewed studies applied diverse deep learning models for road element detection and iRAP assessments, including CNNs, YOLO, YOLACT++, U-Net, Mask R-CNN, DeepLab V3+, and Random Forest (RF) models. YOLO was widely used for real-time detection due to its high accuracy in large-scale data, achieving up to 98.8% accuracy for school zones and 85.1% precision for roadside severity object detection from LiDAR images (Brkić et al. 2023) (Brkić et al. 2022). In pavement distress detection using street-view maps, YOLOv3 achieved a mean average precision of 88.37%, demonstrating its capability in dynamic and large-scale urban environments. Nonetheless, environmental elements' inconsistencies in lighting and occlusions posed significant challenges (Lei et al. 2020). U-Net outperformed RF and XGBoost in road segmentation tasks using high-resolution satellite imagery (Miler et al. 2023), while DeepLab V3+ with a ResNet-50 backbone excelled in UAV-based road segmentation, despite challenges with complex backgrounds (Mahmud et al. 2021). CNNs were primarily used for traffic sign detection and classification tasks. In a study using the Belgium Traffic Sign Dataset (Mehta et al. 2019), CNNs optimized with the Adam algorithm achieved competitive results in classifying 62 categories of traffic signs. However, the model faced difficulties recognizing occluded and highly variable signs, limiting its real-world applicability. A more advanced CNN-based framework (Mask R-CNN) was employed for large-scale traffic sign detection across 200 categories (Tabernik and Skočaj 2019). It achieved under 3% error rates, showcasing its ability to handle high intra-category variation. Despite its accuracy, the computational cost and reduced performance on small or partially visible signs remain limitations. In addition (Kačan et al. 2020) proposed a multi-task learning framework for iRAP attribute classification that demonstrated efficiency by recognizing all attributes in a single forward pass. The model effectively handled class imbalance issues with a macro-F1 scoring of 61.5%. However, its dependency on extensive annotated data and reduced adaptability for new attributes posed challenges. Another approach by (Lin et al. 2023) utilizing RGB-D cameras combined depth and color data using the YOLACT ++ model to achieve accurate 3D quantification of pavement features (potholes and cracks). The model's imprecise segmentation of crack edges highlighted limitations in handling fine details, particularly in noisy environments.

The proposed YOLOv8-based framework enhances road safety by enabling real-time assessment. Its fast, one-stage detection reduces processing time, allowing quicker identification of speed limits and potential hazards. With high accuracy and robustness against varying conditions, it ensures reliable data collection, leading to faster and more efficient road safety evaluations. Deployable on-edge devices streamline assessments, enabling authorities to implement safety measures more swiftly, ultimately reducing accidents and improving road conditions.

3. METHODOLOGY

The research aims to automate the assessment of road environments for iRAP by using the DL technique. Deep learning models, such as object detection, have demonstrated promising results in various object detection tasks. This study focuses on adapting these models to identify and assess road environments, contributing to more reliable and faster road safety evaluations. Figure 1 illustrates the methodology employed to accomplish the proposed framework for automating the identification of speed limit signs based on iRAP coding manual guidelines. The process begins with collecting a comprehensive dataset from various channels, including public repositories. Following data acquisition, data selection is performed to identify the most relevant subsets for training, followed by a preprocessing phase, where the data undergoes resizing, normalization, and manual labeling to ensure it is in an optimal format for model training. Subsequently, model selection progresses to determine the most appropriate deep-learning technique. Depending on the type of road elements being analyzed and the desired outputs, this could involve object detection techniques. Once the dataset has been prepared, the selected DL model is trained on the labeled dataset. This ensures that a sufficiently large and diverse dataset is utilized to improve generalization and robustness. The performance of the trained model is then evaluated by comparing the predicted road elements with the ground truth, using metrics such as mAP and the IoU. A minimum accuracy threshold of 70% is set as the target for model performance. Should the model fail to meet this threshold, further optimization involves adjusting the hyperparameters or applying additional data augmentation techniques to improve the model's accuracy. This iterative process of tuning and retraining continues until the model achieves the desired performance level. Once the model meets or exceeds the performance criteria, it is deemed ready for deployment in road safety assessments, contributing to the automation of iRAP evaluation processes. A performance threshold of 70% was set for the model to be deemed acceptable in this study. Despite extensive efforts to find references or benchmarks in the literature to justify this threshold, no specific information was available. Therefore, the threshold was chosen based on observed performance levels in previous studies to align closely with typical metrics reported in similar applications (Brkić et al. 2022) (Kačan et al. 2020) (Miler et al. 2023). This ensures the model's performance is consistent with established standards, even without direct references.

3.1. Identification of Road Elements and Definition Based on iRAP Classification

The first step in the methodology involves identifying and defining the specific target elements of the road that are to be automated. In this research, the speed limit signs were selected for automation. Once the desired road element is selected, the next step is to refer to the iRAP coding manual (iRAP 2021b). This manual provides a detailed classification framework and instructions on categorizing each segment in the road elements, enabling precise categorization based on standardized criteria. Following this, a comprehensive dataset collection process is undertaken.

3.2. Data Collection, Selection, and Preprocessing

In this research, data collection is specifically designed to support network-wide road safety audit assessments and the calibration of DL algorithms. The DL model training dataset is primarily sourced from publicly available online repositories. For the object detection model (YOLOv8), three repositories of street-view images are utilized to ensure diversity and robustness.

These images are collected from sources available on Kaggle, including (Hemateja 2021), (Maranhão 2019), and (Sichkar 2019). The image resolutions vary, including 1360 × 800, 1920 × 1080, and 300 × 400 pixels. These repositories contain diverse road environment images. These datasets contain various classes of traffic signs, with approximately 58 classes, each comprising around 120 images from multiple European regions. For this study, only images of speed limit signs were extracted, resulting in 724 images used. The YOLO Transfer Learning Guide supports the choice of 724 images for training a YOLOv8 model (Restack 2024). The images include 14 classes of speed limit signs intended to control speeds on roads. The network already has a robust understanding of general object features when using a pre-trained YOLOv8 model, such as one initialized with weights from the COCO (Common Objects in Context) dataset with 1.5 million object instances. This allows the model to adapt effectively to task-specific datasets with significantly fewer images than training from scratch.

The images contain objects other than the speed limit signs, such as shadows, varying daylight conditions, lighting scenarios, trees, traffic signs, cars, and other background elements. This significant interference from background objects challenges the detection of speed limit signs. However, it makes the model more robust against noise. The resulting dataset reflects actual engineering conditions more closely, making it more valuable for iRAP data collection.

The images are street-view and ready to be fed to the model. No additional preprocessing or editing, such as data augmentation, cropping, or color conversion, was performed on the images. The only modification was manually labeling the speed limit signs in the images using the YOLO label tool. This manual process ensured accurate bounding boxes and annotations for each instance in all images, which were then saved in the correct YOLO format. After labeling, the images and their corresponding annotations were split and placed into two sets: training and validation folders, along with writing the (classes.txt) and (data.yaml) files were the preprocessing steps conducted on the dataset. This allowed for a straightforward implementation of the YOLOv8 model for object detection and classification. The training set consisted of 599 images, which were used to teach the model to detect and classify speed limit signs. The validation set comprising 125 images was implemented to evaluate the model's performance during training and adjust parameters to prevent overfitting.

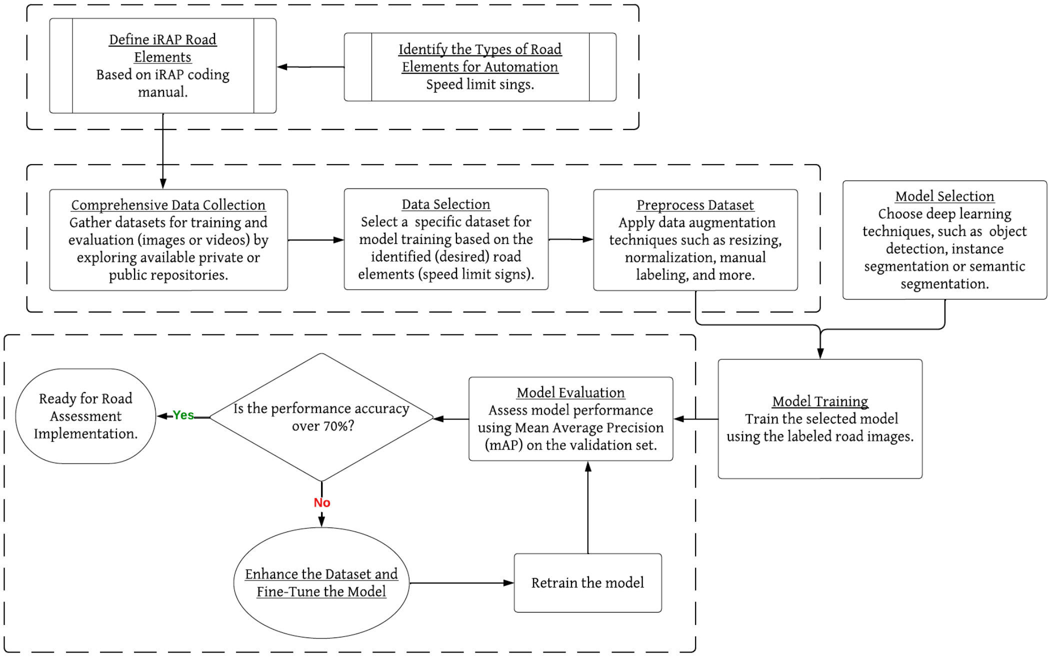

Figure 1: High-level proposed approach

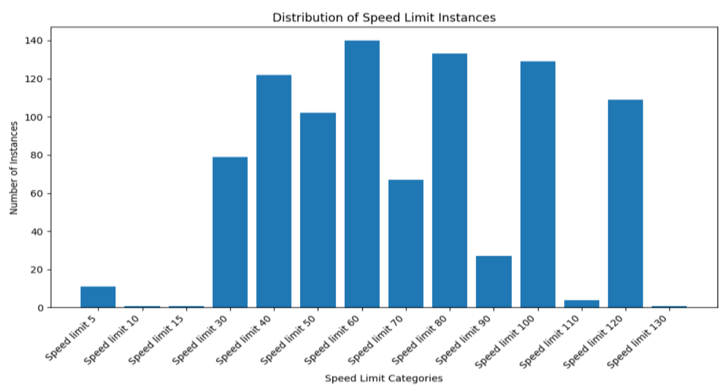

Figure 2: Instances of occurrence in the dataset

The instances vary significantly, as shown in Figure 2, each speed limit's occurrence in the training dataset. The distribution of instances indicates a significant emphasis on speed limits between 30 km/h and 100 km/h, which are common in urban and suburban areas. Lower and higher speed limits have fewer instances, reflecting their relatively less frequent use. This imbalance can affect model performance, as it may become biased towards the more frequent classes, leading to poor generalization of the less frequent classes. However, with limited data available for the minority classes, a pre-training model and fine-tuning it on a specific dataset can help improve the model's performance.

3.3. Model Selection

Deep learning for image object detection for iRAP assessments is essential when distinguishing between different instances of the same object category. It identifies the object type and differentiates between individual instances, making it suitable for detecting objects like speed limit signs (Hafiz and Bhat 2020). The YOLOv8 Large variant significantly improves detection accuracy, particularly under challenging conditions such as occlusions, diverse lighting, and varied backgrounds, while providing enhanced precision in distinguishing similar-looking speed limit signs. Its robust feature extraction capabilities ensure reliable, high-confidence detections suitable for real-time road safety applications. Released in 2023, YOLOv8 extends the capabilities of its predecessors by incorporating advanced features, such as a decoupled head for improved localization and classification and anchor-free detection for better adaptability to diverse object sizes. These innovations make it particularly effective for detecting small and precise objects like speed limit signs. Furthermore, YOLOv8 utilizes a more efficient backbone and optimized model architecture, enabling faster inference and training times while maintaining state-of-the-art results in object detection tasks. Its ability to balance speed and precision makes it an excellent choice for this application, ensuring accurate detection without compromising efficiency (Zhu et al. 2024).

3.4. Model training

Training DL models is crucial in adapting a generic technique to a specific task, such as detecting iRAP elements. Model training aims to optimize the model's ability to recognize patterns within the data by adjusting its internal parameters (weights and biases) through iterative learning (Goodfellow et al. 2016). The YOLOv8 model has been trained on a Google Colab cloud server using a Tesla GPU with CUDA version 12.1, which provides high computational capabilities. The experimental framework includes PyTorch 2.3.1 for foundational DL, with the YOLOv8 algorithm implemented in Python to perform object detection. The methodology for training the YOLOv8 model to detect and classify speed limit signs involved several key steps. The first step was to install the necessary libraries, import and verify these libraries (this setup allowed for efficient debugging, version control integration, and an enhanced coding environment tailored for machine learning workflows), and culminate in training the model using a pre-configured dataset on the Google Colab platform, which allows writing and executing Python code in a Jupyter Notebook environment. Table 1 illustrates the model training configuration. Using a pre-trained model and a well-defined dataset configuration helps accelerate the training process and improve the model's performance. It is worth noting that the 100 epochs were completed in 1.012 hours. By combining these parameters, the command effectively sets up the YOLOv8 model to learn from the dataset, adjusting its weights and biases over 100 epochs to improve its ability to detect and recognize speed limit signs in images of 640x640 pixels.

Table 1: Model training configuration

YOLO Variant	Epoch	optimizer	batch	Image size	Initial learning rate	Weight decay
YOLOv8-L	100	AdamW	16	640	0.01	0.0005

3.5. Model Evaluation

In the model evaluation phase, it is crucial to assess the performance of YOLOv8 to ensure it meets road element detection and classification requirements in iRAP automation. The primary metrics considered at this stage are precision, recall, global F1-Score, and mean Average Precision (mAP). These metrics provide insights into models' ability to generalize to unseen data, measuring their predictions' accuracy, consistency, and reliability. Precision and recall, along with the F1-score, provide a balanced view of how well the models identify relevant instances while avoiding false positives and are appropriate for object detection tasks (Goodfellow et al. 2016).

[1] $\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$ [2] $\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$

Where (TP) represents true positives, (FP) false positives, and (FN) false negatives.

The mAP is a standard metric in object detection that evaluates the precision-recall trade-off at different Intersections over Union (IoU) thresholds. It summarizes the model's performance across various IoU thresholds, providing a comprehensive measure of detection accuracy. The mAP 0.5-0.95 averages the precision over multiple IoU thresholds from 0.5 to 0.95. The AP is calculated as the area under the precision-recall curve, and the mAP is obtained by averaging the APs for all categories (Li et al. 2023).

[3] $\text{AP} = \int_0^1 \text{P dR}$ [4] $\text{mAP} = \sum_{i=1}^{N} \frac{\text{APi}}{\text{N}}$ [5] $\text{F1} - \text{score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$

Where AP represents the average precision score for each class defined by the area under the Precision-Recall curve; P precision; and R recall; mAP mean average precision score across all classes; N number of classes; and APi average precision for the i-th class.

The F1 Score is a harmonic mean and crucial metric that balances precision and recall, providing a single measure of a model's accuracy. It is beneficial in cases with imbalanced datasets, where one class significantly outnumbers the other. Its value ranges from 0 to 1, with 1 indicating perfect precision and recall. This makes the F1 score ideal for applications where both false positives and false negatives have critical consequences. The F1 score is calculated for each epoch by combining the overall precision and recall computed across all classes in the validation set.

4. RESULTS

4.1. Training and Validation Losses

Training and validation losses are crucial for assessing a machine-learning training process and generalization capabilities. Training losses provide insights into how well the model performs on the training data. Validation loss is critical for assessing the model's ability to generalize. A model that performs well on the training data but poorly on the validation data is likely overfitting (Goodfellow et al. 2016). Figure 3 (left) illustrates the training and validation box losses for the YOLOv8 model through 100 epochs to detect speed limit signs. The training box loss, represented by the blue curve, consistently decreases from an initial value of approximately 1.0 to around 0.38 by the final epoch. This steady decline indicates that the model effectively learns to localize objects within the training dataset, suggesting a stable optimization process with no significant fluctuations. In contrast, the validation box loss, represented by the red curve, also starts at a similar value but decreases gradually, stabilizing at approximately 0.7 by the end of training. The slower decline and higher final value of the validation loss suggest a degree of overfitting as training progresses. A notable divergence between the training and validation loss curves occurs at epoch 3, where the gap between the two begins to widen, potentially indicating the onset of overfitting. Despite this divergence, the relatively low and stable validation loss reflects the model's ability to generalize well to unseen data.

The training classification loss in Figure 3 (right) starts at approximately 4.8 and rapidly decreases within the initial epochs, stabilizing at a value near 0.5 by the end of training. This indicates that the model effectively minimizes classification errors for the training dataset by optimizing its parameters. In contrast, the validation classification loss begins significantly higher, at around 21, and decreases more gradually throughout the training process, eventually stabilizing at a value near 1.0 by epoch 100. The large gap between training and validation losses during the early epochs can be attributed to the model adjusting its parameters to fit the training data during this stage. In contrast, its ability to generalize to unseen data remains underdeveloped. By epoch 3, the divergence between the two curves becomes apparent, and while the validation loss shows improvement over time, the persistent gap reflects some level of overfitting.

4.2. Performance metrics

In object detection, achieving high model performance is challenging due to the task's complexity, which involves classifying objects and accurately localizing them within an image. The results indicate that epoch 73 exhibited the highest overall performance, significantly outperforming epoch 20, which initially appeared optimal based solely on loss stabilization but demonstrated lower practical validation metrics. Epoch 100 showed signs of overfitting, as indicated by a decline in validation performance compared to epoch 73. Thus, epoch 73 was selected as the optimal checkpoint for its superior precision, recall, and mAP, prioritizing empirical validation performance over loss curve stabilization.

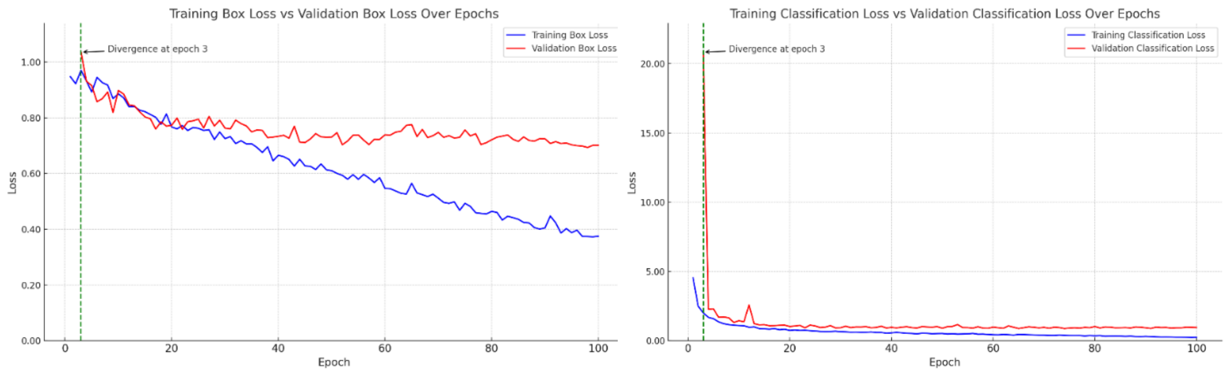

Figure 3: Left: Comparison of training and validation box losses and Right: Comparison of training and validation classification losses

Table 2: YOLOv8 validation metric results

Epoch	Precision	Recall	mAP@50	mAP@50-95	F1-Score
20	65%	81%	71%	58%	72%
73 (selected)	74%	92%	86%	71%	82%
100	71%	88%	81%	68%	78%

4.3. Performance Comparison with Previous YOLO Version

It is essential to benchmark the performance of YOLOv8 against earlier versions of the YOLO model, particularly YOLOv5, to validate the model selection. As shown in Table 3, the proposed YOLOv8-based approach outperforms YOLOv5 across key evaluation metrics. This performance gain can be attributed to several improvements introduced in YOLOv8. First, its redesigned architecture, featuring a fully decoupled head and anchor-free detection, enhances object localization and classification accuracy. Second, YOLOv8 has significantly more parameters (approximately six times more than YOLOv5), giving it a greater capacity to learn complex visual features, which is especially important for identifying detailed traffic signs. Third, its deeper network and refined modules allow for better feature extraction, improving the detection of small or partially occluded signs. Fourth, the anchor-free mechanism simplifies the training process and increases detection robustness across varying object shapes and sizes. Finally, despite being trained over fewer epochs (73 vs. 96), YOLOv8 achieves a higher mAP, reflecting more efficient optimization and better gradient flow. These advancements collectively justify the selection of YOLOv8 for this study.

Table 3: YOLOv5 validation metric results

Epoch	Precision	Recall	mAP@50	mAP@50-95	F1-Score
96	69%	89%	80%	64%	78%

4.4. Image Visualization

In the image visualization analysis, most speed limit signs were accurately identified, with clear bounding boxes around the signs and appropriate labels indicating the detected speed limit and the model's confidence score. For example, in Figure 4 images labeled (A), (B), and (C), the 100, 120, and 80 km/h signs were correctly localized and classified with high confidence. However, the model failed to detect the speed limit sign in the image labeled (E), where the actual speed limit was 80 km/h. This omission highlights a gap in the model's detection capability, suggesting a need for further improvement.

5. DISCUSSION

This study aimed to develop a deep learning framework using YOLOv8 to automate detecting and classifying speed limit signs for iRAP evaluations. It provided a scalable, efficient, and reliable alternative, addressing the limitations of traditional, manual road safety assessments. The dataset utilized in this study featured challenging environments, including varying lighting, occlusions, and diverse background elements such as trees, shadows, and vehicles. Despite these complexities, YOLOv8 demonstrated reliable detection performance, benefiting from its ability to focus exclusively on the pixels within the bounding box. This capability minimizes the influence of irrelevant background features, enabling accurate object detection and classification. The model's real-time inference capability and low computational requirements further underscore its utility. The training process was completed within 1.012 hours on a modest computing platform, emphasizing its efficiency. This makes YOLOv8 an accessible tool for widespread adoption, even in resource-constrained environments where high-performance hardware might not be available. An essential aspect of this work was the dataset preparation and labeling. The manual labeling process, while time-consuming, was critical to the model's success. Each speed limit sign was meticulously annotated to ensure precise bounding boxes and class definitions. High-quality labels are indispensable for supervised learning tasks as they directly influence the model’s ability to generalize and perform accurately. However, the study encountered challenges related to the validation set. Missing instances and unrepresented classes in the validation set negatively impacted the model’s reliability and vision. This issue highlights the importance of a comprehensive dataset that includes all potential classes and sufficient examples for each class. A robust validation set is vital for accurately evaluating the model’s performance and ensuring its applicability. Addressing this gap through better dataset coverage and augmentation strategies will be key to improving future iterations of the framework. Regarding other studies, the same objective can be reached with different techniques, such as instance segmentation models like Mask R-CNN (Tabernik and Skočaj 2019). While these models excel in pixel-level segmentation and provide precise boundary detection, they are computationally intensive and have slower inference times compared to YOLOv8. Mask R-CNN is particularly effective for detecting and analyzing overlapping or irregularly shaped objects, making it well-suited for scenarios where such features are prominent. However, the increased resource demands and slower processing speed may limit its applicability for real-time applications in resource-constrained environments, giving YOLOv8 a practical advantage for iRAP-related tasks.

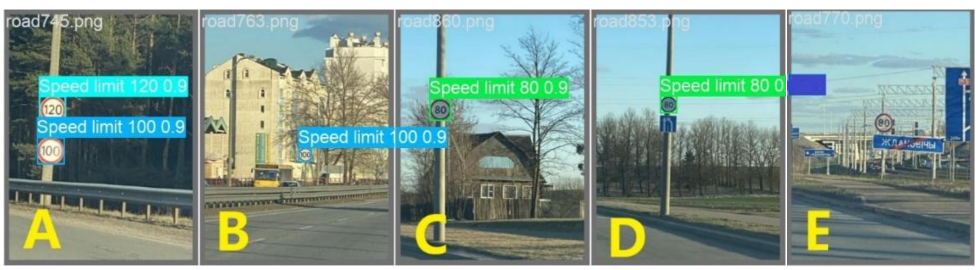

Figure 4: YOLOv8 detection and classification

6. CONCLUSION

The YOLOv8 model demonstrated effective object detection and classification capabilities, accurately identifying most speed limit signs with high confidence and outperforming the YOLOv5. The minor instances of lower confidence, misclassification, or missed detections suggest areas for further refinement. Overall, the results indicate that YOLOv8 provides a promising solution for automating speed limit sign detection in iRAP assessments. While the model performed effectively under diverse conditions, further refinement of the dataset and additional strategies to address class imbalance and missed instances will enhance its robustness and scalability for broader deployment.

REFERENCES

Brkić, I., Miler, M., Ševrović, M., and Medak, D. 2022. Automatic roadside feature detection based on LiDAR road cross section images. *Sensors*, 22 (15): 5510.

Brkić, I., Ševrović, M., Medak, D., and Miler, M. 2023. Utilizing high resolution satellite imagery for automated road infrastructure safety assessments. *Sensors*, 23 (9): 4405. MDPI.
Goodfellow, I., Bengio, Y., and Courville, A. 2016. *Deep learning*. MIT press, Cambridge, Massachusetts, USA.
Hafiz, A.M., and Bhat, G.M. 2020. A survey on instance segmentation: state of the art. *International Journal of Multimedia Information Retrieval*, 9 (3): 171–189.
Hemateja, A. 2021. Traffic Sign Dataset - Classification. Available from https://www.kaggle.com/datasets/ahemateja19bec1025/traffic-sign-dataset-classification. [accessed 16 July 2024].
iRAP. 2021a. International Road Assessment Program Specifications and Manuals: Star Rating and Investment Plan Manual.
iRAP. 2021b. iRAP Specification, Manuals and Guides: Project Planning Manual.
Kačan, M., Oršić, M., Šegvić, S., and Ševrović, M. 2020. Multi-task learning for iRAP attribute classification and road safety assessment. *In 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC).* IEEE. pp. 1–6.
Leduc, G. 2008. Road traffic data: Collection methods and applications. Working Papers on Energy, Transport and Climate Change, 1 (55): 1–55.
Lei, X., Liu, C., Li, L., and Wang, G. 2020. Automated Pavement Distress Detection and Deterioration Analysis Using Street View Map. *IEEE Access,* 8: 76163–76172.
Li, J., Yuan, C., and Wang, X. 2023. Real-time instance-level detection of asphalt pavement distress combining space-to-depth (SPD) YOLO and omni-scale network (OSNet). *Automation in Construction*, 155: 105062.
Lin, W., Li, X., Han, H., Yu, Q., and Cho, Y.-H. 2023. A novel approach for pavement distress detection and quantification using RGB-D camera and deep learning algorithm. *Construction and Building Materials*, 407: 133593.
Mahmud, M.N., Osman, M.K., Ismail, A.P., Ahmad, F., Ahmad, K.A., and Ibrahim, A. 2021. Road image segmentation using unmanned aerial vehicle images and DeepLab V3+ semantic segmentation model. *In 2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*. IEEE. pp. 176–181.
Maranhão, A. 2019. Road Sign Detection. Available from https://www.kaggle.com/datasets/andrewmvd/road-sign-detection. [accessed 16 July 2024].
Mehta, S., Paunwala, C., and Vaidya, B. 2019. CNN based traffic sign classification using Adam optimizer. *In International conference on intelligent computing and control systems (ICCS)*. IEEE. pp. 1293–1298.
Miler, M., Sevrović, M., and Medak, D. 2023. Analysis of Machine Learning Algorithms Performances for Road Segmentation on Very High-Resolution Satellite Imagery as Support of Road Infrastructure Assessment. *International Multidisciplinary Scientific GeoConference: SGEM,* 23 (2.1): 121–128.
Restack. 2024. Yolov8 Transfer Learning Guide. Available from https://www.restack.io/p/yolov8-transfer-learning-answer-cat-ai. [accessed 10 December 2024].
Sichkar, V. 2019. Traffic Signs Dataset in YOLO format. Available from https://www.kaggle.com/datasets/valentynsichkar/traffic-signs-dataset-in-yolo-format. [accessed 16 July 2024].
Tabernik, D., and Skočaj, D. 2019. Deep learning for large-scale traffic-sign detection and recognition. *IEEE transactions on intelligent transportation systems*, 21 (4): 1427–1440.
Transport Canada. 2021. Transportation in Canada.
World Bank. 2020. Guide for Road Safety Opportunities and Challenges: Low and Middle Income Country Profiles.
Yin, C., Tang, J., Yuan, T., Xu, Z., and Wang, Y. 2021. Bridging the gap between semantic segmentation and instance segmentation. *IEEE Transactions on Multimedia*, 24: 4183–4196.
Zhu, J., Wu, Y., and Ma, T. 2024. Multi-Object Detection for Daily Road Maintenance Inspection With UAV Based on Improved YOLOv8. *IEEE Transactions on Intelligent Transportation Systems*.