Article

Not peer-reviewed version

Deep Learning-Based Real-Time Railway Obstruction Detection

K. Divya Bharathi , Moksha Odlamaani * , Orsu Shashi Kumar , Lavanya Pyla , Shanmukhi Rayala

Posted Date: 27 February 2026

doi: 10.20944/preprints202602.1804.v1

Keywords: railway obstruction detection; deep learning; computer vision; object detection; YOLO; real-time monitoring

Article

Deep Learning-Based Real-Time Railway Obstruction Detection

K. Divya Bharathi, Moksha Odlamaani *, Orsu Shashi Kumar, Lavanya Pyla and Shanmukhi Rayala

Department of Artificial Intelligence and Machine Learning Malla Reddy University, Hyderabad, India

* Correspondence: moksha2422@gmail.com

Abstract

The railway sector plays a vital role in passenger transportation, economic growth, and large-scale urban development. Despite continuous improvements in signaling and control systems, railway safety remains a major concern due to frequent incidents caused by track obstructions such as unauthorized human intrusions, animals, vehicles, and miscellaneous foreign objects. Existing railway monitoring systems rely heavily on manual inspection, physical barriers, and rule-based alerts, which suffer from delayed response times, limited coverage, high maintenance costs, and reduced reliability under challenging conditions including low illumination, adverse weather, and crowded environments. This paper presents a deep learning-based real-time railway obstruction detection framework utilizing advanced image analysis and object detection techniques. Multiple state-of-the-art neural network architectures, including SSD, Faster R-CNN, RetinaNet, YOLOv3, YOLOv7, YOLOv8, YOLOv5S6, YOLOv5X6, and YOLOv9, are implemented and systematically evaluated for detecting and localizing railway obstructions. The models are trained and tested using a hybrid dataset composed of custom railway images and publicly available datasets to ensure robustness across diverse operational scenarios. Performance is assessed using standard metrics such as Mean Average Precision (mAP), precision, recall, and inference speed. Experimental results demonstrate that YOLO-based architectures, particularly YOLOv8 and YOLOv9, achieve superior detection accuracy while maintaining real-time processing capability. These findings highlight the effectiveness of YOLO-derived models as reliable solutions for enhancing railway safety and supporting intelligent transportation systems.

Keywords: railway obstruction detection; deep learning; computer vision; object detection; YOLO; real-time monitoring

I. Introduction

Railway transportation remains one of the most reliable and widely used modes of transport for both passengers and freight worldwide. It facilitates large-scale trade, industrial growth, and human mobility, all of which are essential to a country's economic development. Despite continuous advancements in railway infrastructure, signaling systems, and operational procedures, railway safety continues to face significant challenges, particularly due to unforeseen obstacles present on railway tracks.

Unauthorized human intrusions, animals crossing the tracks, stalled vehicles at level crossings, rolling stones, garbage, fallen trees, and other foreign objects represent common causes of railway obstructions. These incidents pose serious safety risks and have been associated with derailments, collisions, service disruptions, and loss of life. Railway safety studies indicate that track-level intrusions and delayed obstacle detection contribute to a substantial proportion of railway accidents, emphasizing the urgent need for automated and intelligent monitoring solutions.

Traditional railway monitoring systems primarily rely on closed-circuit television (CCTV) cameras, physical fencing, occasional patrolling, and manual surveillance. While these approaches provide basic situational awareness, they suffer from several inherent limitations. Continuous monitoring across extensive railway networks is labor-intensive, prone to human error, and affected by operator fatigue. CCTV-based systems often fail to generate timely alerts and require constant human supervision. Moreover, conventional rulebased and image processing techniques perform poorly in realworld railway environments characterized by low illumination, adverse weather conditions, shadows, occlusions, and complex backgrounds.

With rapid advancements in artificial intelligence, particularly in deep learning and computer vision, automated object recognition systems have become critical tools for enhancing security and surveillance through video analysis. Modern machine learning algorithms can automatically extract complex visual features from raw images without manual feature engineering, enabling accurate identification of multiple objects within individual images or video sequences. These techniques have demonstrated exceptional performance across diverse domains, including autonomous vehicles, traffic management, industrial inspection, and video surveillance systems.

Several state-of-the-art deep learning models, including SSD, Faster R-CNN, RetinaNet, and various YOLO architectures, have been widely adopted for object detection tasks. Two-stage detectors such as Faster R-CNN achieve high detection accuracy through region proposal generation followed by classification, whereas single-stage detectors like SSD and YOLO prioritize inference speed, making them suitable for real-time applications. Recent versions of the YOLO framework have significantly improved detection precision, robustness to environmental variations, and computational efficiency, positioning them as strong candidates for safety-critical applications.

Real-time operational efficiency is a crucial requirement for railway safety systems. Delayed detection of track obstructions can result in severe consequences due to long braking distances and the need for rapid response during train operation. Therefore, an effective railway obstruction detection system must balance high detection accuracy with fast computational performance while operating reliably across diverse environmental conditions.

This study addresses these challenges by developing an automated railway obstruction detection system based on advanced deep neural networks. The proposed approach incorporates multiple state-of-the-art object detection techniques and presents a comprehensive comparative evaluation of SSD, Faster R-CNN, RetinaNet, and several YOLO variants, including YOLOv3, YOLOv7, YOLOv8, YOLOv5S6, YOLOv5X6, and YOLOv9. The models are trained and evaluated using a combination of custom railway imagery and publicly available datasets, ensuring robustness, reliability, and broad applicability across real-world railway environments.

II. Literature Review

Over the past few decades, scholars have focused more on improving railroad security because there's been an increasing occurrence of incidents involving unapproved entries into tracks and obstacles at this level. Existing methods like periodic visual checks and algorithm-driven analysis fall short in managing extensive rail systems efficiently. Consequently, scientists now extensively investigate machine learning and neural network methodologies for enhancing diagnostic precision and operational speed.

Early studies focused primarily on conventional methods like background removal, feature extraction through edges, and tracking movements to identify intruders near railway tracks. Although demonstrating limited effectiveness under controlled conditions, these techniques faced significant hurdles upon implementation due to varying environmental factors such as fluctuating light sources, temperature shifts, shadows, and obstruction of objects. The constraint led to the adoption of algorithms specifically crafted for extracting features automatically from data.

Kumar and Harsha conducted an extensive study on employing machine vision techniques to identify flaws within railway systems while also safeguarding against unauthorized entries. A study demonstrated that advanced machine-learning techniques excelled over traditional approaches when optimizing image-related tasks, thereby enhancing accuracy and robustness significantly. The authors highlighted the importance of large datasets and quick responses when conducting their studies; however, they noted that much recent research is concentrated exclusively on a specific kind of artificial intelligence model, making it difficult to find suitable methods for practical applications.

Numerous scholars investigated two-step object recognition architectures like Fast R-CNN in applications related to rail safety inspections. The algorithms produce candidate regions first before performing object identification, leading to excellent recognition rates. Nevertheless, owing to high computational demands, dual-stage recognizers frequently fall short of meeting timely requirements, especially within stringent railroad settings demanding swift action responses.

Introducing RetinaNet aimed at overcoming issues like uneven classes and tiny objects through its incorporation of a focal loss mechanism. Research employing RetinaNet technology has demonstrated enhanced object recognition for detecting tiny and partly obscured items like rocks and dirt along railroad routes. Although RetinaNet possesses several benefits, it struggles with average processing speeds, potentially hindering its application in high-speed rail surveillance applications.

The authors of Baldega et al. Techniques for anomaly identification were explored through employing machine-learning algorithms on datasets comprising rail infrastructure sensors such as vibrations and sounds. Their strategy showcased successful outcomes for maintaining infrastructure integrity and pinpointing faults efficiently. Nevertheless, sensors alone cannot interpret visuals; they frequently necessitate intricate setup, adjustment, and upkeep, rendering extensive implementation prohibitively costly and difficult.

Lee et al. Proposed an integrated system using both laser range finders and cameras to detect obstacles within railroad passages. Using a fusion method improved both precision during location tasks under dim lighting scenarios significantly. Despite requiring costly equipment and intricate coordination methods, these setups severely hinder broad adoption in largescale rail infrastructures due to high costs and complexity.

New developments within the YOLO series enhance not only the precision but also accelerate model execution efficiency substantially. These newer models include advanced neural networks, refined features for better detection, and more efficient learning techniques in their architecture design. The YOLO model is excellently adapted for use in live monitoring systems like detecting obstacles on railways due to recent improvements. Despite limited comparative analyses of different versions of You Only Look Once algorithms alongside general object recognition systems for specific rail applications, numerous investigations exist on this topic.

A substantial portion of current studies focus exclusively on a particular type of detector, limited sets of intrusion types, or laboratory-based experiments; however, recent publications highlight the potential of deep learning technology for enhancing railroad security.

III. System Architecture

The proposed railway obstruction detection system follows a modular and hierarchical architecture designed to ensure scalability, robustness, and real-time operational efficiency. The system architecture decomposes the overall detection process into a sequence of interconnected functional modules, each responsible for a specific stage of data acquisition, processing, detection, and decision-making. Such a structured design enables efficient handling of large-scale video data while maintaining low latency and high detection accuracy. Furthermore, the modular organization facilitates seamless integration of advanced deep learning models and allows future extensions or upgrades without requiring significant modifications to the overall framework. The complete workflow of the proposed system is illustrated in Figure 1.

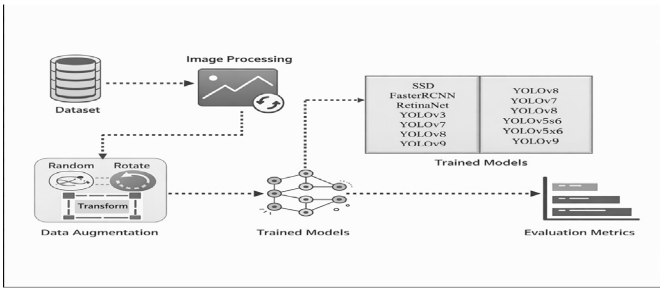

Figure 1. Overall architecture of the proposed deep learning-based railway obstruction detection system.

A. Data Acquisition Module

The data acquisition module is responsible for capturing continuous video streams and image frames from surveillance cameras installed along railway tracks, level crossings, and other high-risk zones. The cameras operate under diverse environmental conditions, including varying illumination levels, weather variations, and partial occlusions. The incoming video streams are segmented into individual frames at predefined time intervals to support both real-time inference and offline model training.

B. Preprocessing Module

The preprocessing module prepares the acquired frames for efficient model training and inference. Each frame is resized to a fixed input resolution to ensure compatibility with deep learning models and to reduce computational complexity. Pixel normalization is applied to standardize input values, and color space conversion is performed to transform images from BGR to RGB format. In addition, basic noise reduction techniques are employed to improve visual quality under low-light and adverse weather conditions.

C. Data Augmentation Module

To improve model generalization and address dataset imbalance, a data augmentation module is incorporated during the training phase. This module applies transformations such as random rotation, horizontal flipping, scaling, translation, and brightness adjustment. These operations increase the diversity of training samples and enable the models to learn robust feature representations, thereby improving detection performance across varying railway environments.

D. Detection Module

The detection module constitutes the core of the proposed system and employs deep learning-based object detection algorithms to identify potential obstructions on railway tracks. Multiple state-of-the-art models, including SSD, Faster RCNN, RetinaNet, and several YOLO variants, are trained and evaluated within this module. Single-stage detectors such as YOLO offer high inference speed suitable for real-time deployment, whereas two-stage detectors such as Faster RCNN provide higher localization accuracy. The module outputs bounding boxes, object class labels, and confidence scores for each detected entity.

E. Decision and Visualization Module

The decision and visualization module analyzes the outputs generated by the detection module to assess the severity of detected intrusions. Confidence thresholds are applied to suppress false positives, and detected objects are categorized based on predefined risk levels. A real-time visualization interface is developed using the Flask framework to display detection results, overlay bounding boxes, and generate alert notifications for railway authorities. This module enables continuous monitoring and supports rapid response to potential safety threats.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2026 doi:10.20944/preprints202602.1804.v1

F. System Scalability and Extensibility

The modular architecture of the proposed system allows easy replacement or integration of new object detection models without disrupting the overall workflow. Future extensions may include the deployment of edge computing devices for low-latency processing, multi-camera data fusion to improve spatial coverage, and integration with IoT-based sensors to enhance system reliability and scalability.

Overall, the proposed architecture provides an efficient and flexible framework for real-time railway obstruction detection, achieving high detection accuracy, low latency, and adaptability to evolving railway safety requirements.

IV. Methodology

A. Dataset Collection

A hybrid dataset is constructed by combining custom railway track images with publicly available datasets such as COCO. The dataset includes multiple obstruction categories such as humans, animals, vehicles, stones, debris, and other foreign objects commonly found on railway tracks. Images are captured under diverse environmental conditions, including varying illumination, weather scenarios, and camera viewpoints, to improve model generalization in real-world deployments.

B. Data Preprocessing and Augmentation

All input images are resized to a uniform resolution of 640 × 640 pixels to ensure compatibility across different detection architectures. Preprocessing steps include pixel normalization and color space conversion. To enhance dataset diversity and reduce overfitting, data augmentation techniques such as random rotation, horizontal flipping, scaling, and brightness adjustment are applied during training.

C. Model Selection and Description

To evaluate performance trade-offs between detection accuracy and real-time inference, multiple deep learning–based object detection models are selected, including both singlestage and two-stage architectures.

Single-stage detectors such as **SSD** and **YOLO variants** perform object localization and classification in a single forward pass. SSD employs a convolutional backbone network with anchor-based default bounding boxes generated at multiple feature scales and jointly optimizes classification and localization loss functions. YOLO models utilize lightweight backbone networks and directly regress bounding box coordinates and class probabilities from feature maps, enabling high inference speed suitable for real-time applications.

Two-stage detectors such as **Faster R-CNN** first generate candidate object regions using a Region Proposal Network (RPN), followed by classification and bounding box refinement. This architecture typically uses a ResNet-based backbone and anchor-based region proposals, providing higher detection accuracy at the cost of increased computational complexity.

All selected models are trained on the same dataset using identical preprocessing and augmentation strategies to ensure a fair and consistent comparative evaluation.

D. Training Configuration

Training is performed on GPU-accelerated systems to reduce computational time. Adaptive learning rate optimization is employed to stabilize convergence, and early stopping is used to mitigate overfitting. Model-specific classification and bounding box regression loss functions are applied according to the respective detection architectures. Model checkpoints are periodically saved for subsequent evaluation.

 doi:10.20944/preprints202602.1804.v1

V. Performance Evaluation Metrics

The performance of the proposed railway obstruction detection system is assessed using standard object detection metrics. Precision and Recall are computed as follows:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Mean Average Precision (mAP) is used to measure detection accuracy across all object classes at predefined Intersection over Union (IoU) thresholds. Inference speed, measured in frames per second (FPS), is also reported to evaluate the realtime applicability of the proposed system.

A. Experimental Setup

All experiments had been carried out on a GPU-multiplied machine equipped with an NVIDIA RTX-collection GPU, an Intel-primarily based processor, and enough device memory. Inference speed is reported in frames per 2nd (FPS) and serves as an indicator of give up-to-end detection latency beneath same testing conditions for all fashions.

B. Quantitative Performance Analysis

The consequences indicate that YOLO-primarily based detectors continuously outperform SSD and faster R-CNN in phrases of inference pace at the same time as maintaining aggressive detection accuracy. despite the fact that quicker R-CNN achieves fantastically excessive precision and take into account, its low inference velocity limits its applicability in actual-time railway monitoring eventualities in which rapid response is critical.

VI. Results and Discussion

This section presents the experimental results and a detailed discussion of the performance of various deep learning–based object detection models for railway obstruction detection. The models are evaluated using detection accuracy, precision, recall, and real-time inference speed to analyze their applicability in safety-oriented railway monitoring scenarios.

Table 1. Performance Comparison of Object Detection Models.

Model	mAP (%)	Precision	Recall	FPS
SSD	78.4	0.79	0.76	25
Faster R-CNN	85.6	0.87	0.83	7
RetinaNet	83.2	0.84	0.81	15
YOLOv7	88.9	0.90	0.87	45
YOLOv8	91.3	0.92	0.90	55
YOLOv9	**93.1**	**0.94**	**0.92**	**60**

Single-degree detectors such as YOLOv7, YOLOv8, and YOLOv9 exhibit a good balance among accuracy and computational efficiency. among those, YOLOv9 achieves the best imply common Precision of ninety three.

A. Qualitative Analysis and Visual Results

To further assess detection performance, qualitative analysis is conducted on real-time railway surveillance images. Figure 2 shows a sample input image obtained from the monitoring camera feed, containing background elements commonly present in railway environments.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2026 doi:10.20944/preprints202602.1804.v1

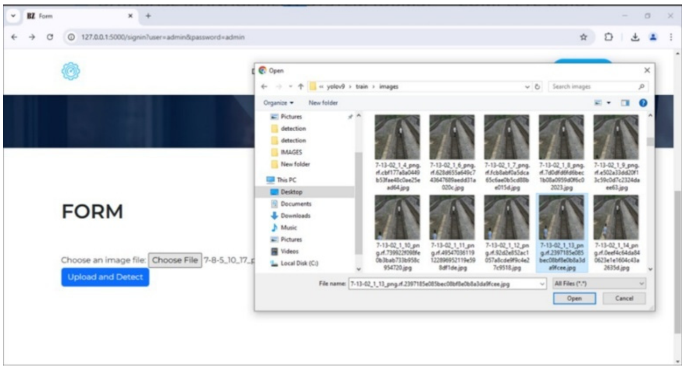

Figure 2. Sample input image selected from the railway surveillance feed.

Figure 3 illustrates human intrusion detection on railway tracks. The model accurately localizes human presence with high confidence despite background complexity. Reliable human detection is critical, as unauthorized pedestrian intrusions are a major contributor to railway safety incidents.

Figure 4 presents detection results for non-human obstructions such as animals, vehicles, and debris. The model successfully identifies multiple object classes within the same frame, demonstrating robustness and generalization across diverse obstruction types.

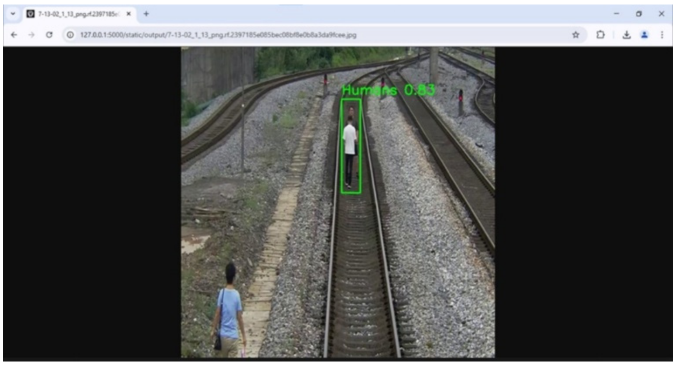

Figure 3. Human intrusion detection on railway tracks.

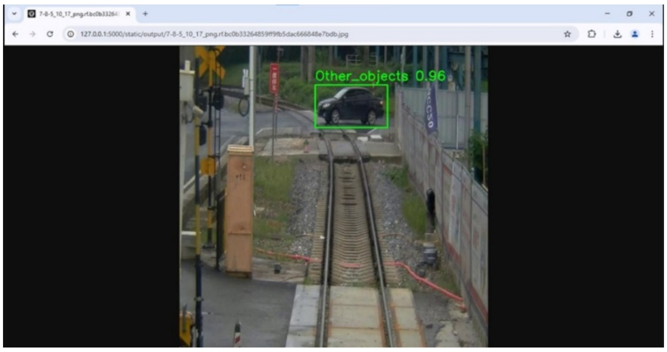

Figure 4. Detection of non-human obstructions including animals, vehicles, and debris.

B. Discussion

The experimental results highlight the importance of selecting detection architectures that balance accuracy and inference speed for railway safety applications. Although SSD and RetinaNet offer respectable detection accuracy, their performance deteriorates when small objects, occlusion, or crowded backdrops are present. Faster R-CNN's large processing overhead limits its real-time usefulness despite its outstanding accuracy. On the other hand, under the same hardware settings, YOLOv8 and YOLOv9 consistently provide higher accuracy with reduced latency. According to the observed performance, YOLO-based designs are a good fit for real-time monitoring systems where prompt obstacle identification is essential. However, real deployment performance may differ according on ambient factors and hardware design, suggesting the necessity for additional optimization and field-level assessment. Overall, the findings show the useful benefits of single-stage detectors in real-time situations and confirm the efficacy of deep learning-based object detection for railway obstruction detection.

VII. Conclusion

The document introduced an extensive deep-learning-driven approach aimed at detecting obstacles on railways instantly while focusing on improving train operation security, visibility about situations during travel, and preventing accidents within rail networks. An innovative method employs sophisticated image analysis tools for real-time identification of numerous types of obstacles, encompassing people, wildlife, motorized equipment, and other non-human entities across various settings characterized by fluctuating light levels, climatic shifts, and changing backgrounds.

An extensive examination into various cutting-edge methods for detecting objects across different systems was carried out to assess how these techniques balance aspects such as precision in identifying targets, processing time required by computers during recognition tasks, and resource consumption levels. Experimental findings indicate that single-stage YOLO models generally surpass both SSD and Faster R-CNN by offering superior real-time capabilities without sacrificing detection precision. Specifically, models like YOLOv8 and YOLOv9 demonstrate outstanding Mean Average Precision along with efficient latency in GPU configurations, making them ideal choices for real-time railway surveillance tasks requiring swift responses.

Additionally, incorporating live data displays and warning systems enhances the usability of our suggested model through quick detection of risky conditions and efficient response actions. In conclusion, these results confirm the efficacy of using deep-learning-based object recognition in detecting obstacles on railways efficiently during live operations, highlighting the significance of choosing model configurations that strike an optimal equilibrium between precision and performance requirements for ensuring safe operation of advanced traffic management infrastructures.

VIII. Future Work

Despite showcasing impressive results, considerable avenues still need exploration in further studies and improvements. An enhancement could come through investigating transformers for object recognition tasks, potentially leading to better overall comprehension and more precise detections within intricate and busy train environments. Moreover, implementing this system in edge computing nodes might decrease dependency on central computation, cut down delays significantly, and allow for immediate site-specific data analytics at distant train stations.

Further research might concentrate on integrating multiple camera systems for enhanced surveillance, minimizing gaps in view, and enhancing overall reliability when monitoring extensive rail networks over extended distances. Combining various data points like IoT devices, train velocity info, and environmental monitors might significantly improve risk analysis capabilities while allowing for proactive safety notifications.

Ultimately, extensive field tests on an industrial scale alongside ongoing data accumulation is crucial for confirming how well the new approach functions in practical settings over time and ensuring its lasting effectiveness and dependability.

References

1. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

 doi:10.20944/preprints202602.1804.v1

2. J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv preprint arXiv:1804.02767, 2018.
3. S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
4. W. Liu et al., “SSD: Single Shot MultiBox Detector,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016, pp. 21–37.
5. T.-Y. Lin et al., “Focal Loss for Dense Object Detection,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988.
6. C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,” arXiv preprint arXiv:2207.02696, 2022.
7. G. Jocher et al., “Ultralytics YOLOv8,” arXiv preprint arXiv:2301.00506, 2023.
8. C.-Y. Wang et al., “YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information,” arXiv preprint arXiv:2402.13616, 2024.
9. T.-Y. Lin et al., “Microsoft COCO: Common Objects in Context,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014, pp. 740–755.
10. A. Kumar and S. Harsha, “Machine Vision-Based Railway Defect Detection: A Survey,” *IEEE Transactions on Intelligent Transportation Systems*, 2024.
11. Z. Zhang et al., “Real-Time Railway Intrusion Detection Using Deep Learning,” *IEEE Access*, vol. 9, 2021.
12. P. Bałdyga et al., “Machine Learning Methods for Railway Infrastructure Monitoring,” *Sensors*, vol. 21, no. 3, 2021.
13. T. Lei et al., “Obstacle Detection in Railway Tunnels Using LiDAR and Vision Fusion,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3675–3685, 2021.
14. R. Girshick, “Fast R-CNN,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448.
15. S. Chen et al., “Deep Learning for Intelligent Transportation Systems: A Survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 4555–4571, 2021.