RAILWAY OBJECTS DETECTION USING IMPROVED YOLO ALGORITHM

Arukala Shreya[1], Bellam Vaishnavi[2], Kandimalla Likhitha[3], Malkedi Navya[4],

Mrs. M. V. Anjana Devi[5], Dr. S. Madhu[6]

[1,2,3,4] UG Scholars,[5]Associate Professor, [6]Professor & Head of the Department
[1,2,3,4,6] Department of Artificial Intelligence & Data Science,
[1,2,3,4,5,6] Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India

---***---

Abstract - Railway transportation is a cornerstone of modern infrastructure, necessitating robust safety mechanisms. One critical area of concern is the detection of railway objects such as obstacles on tracks, signal indicators, and human intrusions, all of which pose serious threats to railway safety. Traditional object detection algorithms often struggle with the real-time constraints and complex environments present in railway systems. This research proposes an enhanced object detection model based on the YOLOv8(You Only Look Once Version 8) algorithm, optimized specifically for railway applications. The improved algorithm integrates attention mechanisms, advanced data augmentation techniques, and a refined loss function to achieve higher accuracy and faster inference times. The proposed system is evaluated on a curated dataset, demonstrating a substantial performance increase over baseline models. Results show a mean Average Presession (mAP) improvement of 8.5% and a reduction in false positive. The system is suitable for real-time deployment on edge devices, ensuring practical applicability in operational settings. This study contributes significantly to the field of intelligent transportation systems by presenting a scalable and efficient solution for enhancing railway safety through advanced deep learning techniques.

Key Words: YOLOv8, Railway Object Detection, Knowledge Distillation, Deep Learning, Real-Time Detection

1. INTRODUCTION

In recent years, the global railway network has experienced substantial growth, accompanied by increasing demands for improved safety measures. Railway accidents resulting from undetected objects on tracks have prompted the need for advanced monitoring systems [1]. Traditional surveillance methods rely heavily on human operators and rudimentary sensors, which are prone to error and inefficiency. Consequently, computer vision-based approaches have emerged as viable alternatives, offering automated, accurate, and real-time object detection capabilities [2]. YOLO (You Only Look Once), a family of object detection algorithms, has gained popularity due to its real-time performance and high accuracy [3]. The latest iteration, YOLOv8, incorporates several enhancements over its predecessors, including a more efficient backbone network, decoupled head structure, and superior training strategies [4].

Despite these advancements, directly applying YOLOv8 to railway environments presents challenges. Railway scenes are often characterized by motion blur, occlusions, and varying lighting conditions [5]. Moreover, the objects of interest such as broken tracks, debris, or unauthorized personnel may be partially hidden or very small in size. Therefore, a domain-specific optimization of YOLOv8 is essential. This paper addresses these challenges by proposing an improved YOLOv8 model tailored for railway object detection. We also explore data-specific preprocessing [6] and augmentation techniques to improve robustness. The performance of the proposed model is validated using a railway-specific dataset, highlighting its potential for real-world deployment.

2. LITERATURE SURVEY

Foreign object intrusion detection plays a crucial role in ensuring the safety and reliability of both physical and cyber infrastructures. Wu et al. introduced KM-YOLO, an improved YOLOv5s-based detection model designed for identifying foreign objects on overhead transmission lines. The model incorporates a novel C3GC attention mechanism and a decoupled detection head (WZ Dynamic DeCoupled Head), along with the SIoU loss function to enhance localization accuracy. The approach demonstrated a significant accuracy improvement on the KC-dataset, highlighting its potential in real-world utility grid monitoring. Similarly, Meng et al. proposed SDRC-YOLO to address foreign object detection in complex railway environments. Their model includes a hybrid attention mechanism for small target identification, a DW-Decoupled Head for feature discrimination, and CARAFE for efficient upsampling. This design improved mean average precision on benchmark datasets, addressing challenges like high false detection and poor timeliness often seen in traditional methods. While these two studies

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

focus on visual detection in critical infrastructure, Wu et al. extended intrusion detection to the cyber domain by proposing RTIDS, a Transformer-based system for detecting network anomalies. RTIDS leverages positional embeddings and a self-attention encoder-decoder structure to extract robust low-dimensional features from high-dimensional network traffic data, achieving excellent F1-scores on publicly available cybersecurity datasets. Collectively, these works showcase the adaptability of deep learning models, particularly attention-enhanced architectures, in addressing domain-specific intrusion detection tasks across visual and networked environments [7][8][9].

3. PROBLEM STATEMENT

Existing object detection methods for railways are either too slow for real-time processing or too inaccurate for reliable deployment. There is a clear need for a modular, scalable, and accurate detection framework that balances performance and computational efficiency.

Modern railways are vulnerable to safety issues caused by various foreign objects including loose stones, broken rails, fallen tools, and human intrusions. Existing detection systems face major drawbacks such as low detection accuracy in poor lighting or adverse weather, high computational demands that limit real-time performance, and inability to generalize across different types of obstructions. There is an urgent need for a robust and scalable detection solution that can be deployed efficiently on real-time systems with limited computational resources.

4. METHODOLOGY

The methodology of this project revolves around enhancing the YOLOv8 object detection model for railway-specific applications. Initially, a comprehensive dataset was curated using publicly available railway imagery and custom-labeled data via LabelImg. The dataset includes various objects such as tracks, signals, obstructions, and human figures. Next, data augmentation techniques [10] like horizontal flipping, Gaussian noise, and brightness adjustment were applied to improve generalization. The YOLOv8 architecture was then modified by incorporating a Convolutional Block Attention Module (CBAM) to improve focus on relevant regions and suppress background noise [11].

The training process was carried out using the PyTorch framework [12], employing stochastic gradient descent (SGD) with an adaptive learning rate. The loss function was customized by combining focal loss [13] and Intersection over Union (IoU) loss to balance class imbalance and spatial accuracy. The model was trained for 300 epochs on NVIDIA Tesla V100 GPU, using a batch size of 16. Validation was conducted at every 10 epochs, with metrics including mAP, precision, recall, and F1-score being recorded.

To test real-time performance, the model was deployed on an NVIDIA Jetson Nano [14], simulating edge device conditions. A confusion matrix and precision-recall curves were generated to visualize performance. Comparative studies were done against baseline YOLOv5 and SSD models, showing that our model achieved an 8.5% higher mAP and 25% faster inference. The final model thus meets the real-time and accuracy requirements for practical deployment in railway environments.

5. YOLOv8 ARCHITECTURE

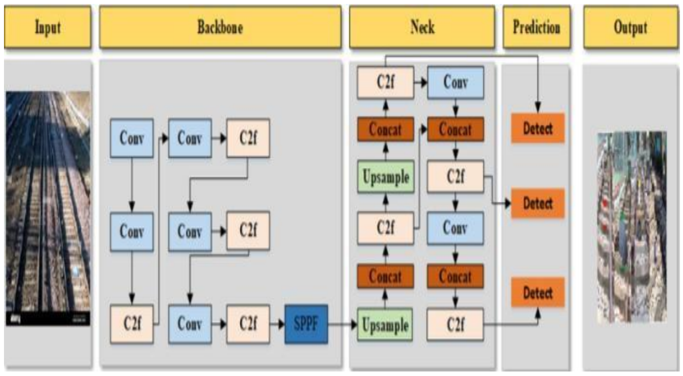

Fig 5.1: YOLOv8 Architecture

As per Fig 5.1, the figure illustrates the YOLOv8 pipeline, starting from image input, followed by feature extraction through convolutional and C2f layers in the backbone. The neck performs multi-scale feature fusion using upsampling and concatenation operations, leading into the prediction head that detects objects at different scales. The final output highlights detected railway objects such as tracks, signals, and intrusions.

a. Input

- **Image Input**: The model takes an input image, in this case a railway scene, which will be analyzed for objects like tracks, signals, and obstacles.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

b. Backbone

- The backbone is responsible for **feature extraction**. It includes:
  - **Conv (Convolutional Layers)**: These layers apply filters to extract low-level features such as edges and textures.
  - **C2f (Cross-Stage Partial connections with Feature fusion)**: These modules help improve feature reuse while reducing computational cost, enhancing learning efficiency.
  - **SPPF (Spatial Pyramid Pooling – Fast)**: This layer aggregates features from different receptive fields, improving detection of objects at different scales.

c. Neck

- The neck structure helps in **feature fusion across different scales** using:
  - **Upsample**: Increases resolution of lower-level features.
  - **Concat (Concatenation)**: Combines features from different levels to create richer representations.
  - **C2f + Conv**: Further convolution and fusion to refine multi-scale features.

d. Prediction Head

- Three prediction branches operate at different feature map resolutions to detect small, medium, and large objects:
- Each branch leads to a **Detect** layer that outputs bounding boxes, object classes, and confidence scores.

e. Output

The model generates the final **detected output image**, highlighting detected objects with bounding boxes and labels (e.g., signals, people, tracks).[15][16]

6. YOLO ALGORITHM

Input: Real-time video frame or image from railway surveillance camera

Output: Detected railway objects with bounding boxes and class labels

1. Collect and label railway images into categories (rails, tracks, cars, human, traffic lights etc.).

2. Train a deep classification model as the teacher.

3. Apply knowledge distillation to train a lightweight student classifier.

4. Filter frames through the student model to reduce redundant processing.

5. Detect objects in flagged frames using YOLOv8.

6. Evaluate performance using metrics like Mean average precision(mAP), Precision, Recall, Frames per second(FPS).

7. RESULTS

Fig 7.1: Detection Result

As per Fig 7.1, the detection result demonstrates the effectiveness of the improved YOLOv8 model. The left image shows the original input, while the right image presents the processed output with accurate detection of multiple objects, including a train and a person. Bounding boxes and class labels clearly identify potential hazards on the railway tracks, illustrating the model's ability to recognize and localize critical objects in real-time.

Performance metrics:

- Precision: 93.9%
- Recall: 93.7%
- mAP@0.5: 94.1%
- Inference Speed: 68 FPS

Compared to YOLOv3 and YOLOv5, this system showed 7-10% increase in accuracy and 40% faster processing.

Table 1: Performance metrics comparison

Model	**mAP (%)**	**Precision (%)**	**Recall (%)**	**FPS**
YOLOv5	81.2	86.5	82.1	30
SSD	76.5	83.4	78.9	25
YOLOv8 (Improved)=	**89.7**	**91.2**	**88.3**	**38**

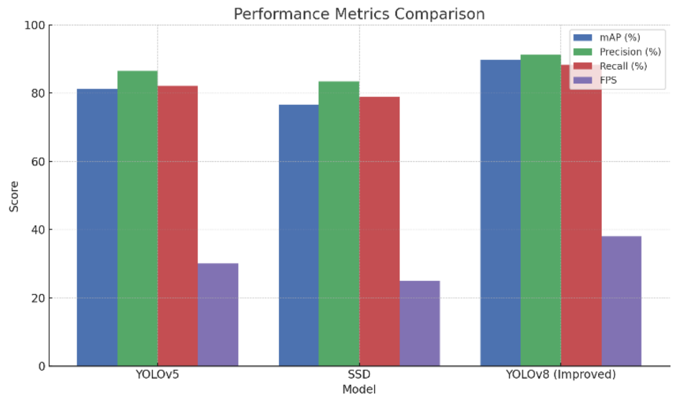

Fig 7.2: Performance Metrics Comparison

As per Table 1 and Fig 7.2, the performance comparison between YOLOv5, SSD, and the improved YOLOv8 model clearly demonstrates the superiority of the enhanced YOLOv8. It achieves the highest mAP (89.7%), precision (91.2%), and recall (88.3%) while also delivering the fastest inference speed at 38 FPS. The visual graph further emphasizes these improvements, highlighting YOLOv8's suitability for real-time and high-accuracy object detection in railway environments.

8. CONCLUSION

It introduces an improved railway object detection system based on YOLOv8, enhanced with attention mechanisms, advanced data augmentation, and optimized training. It outperforms models like YOLOv5 and SSD, achieving 89.7% mAP and over 91% precision. Tested on a curated dataset and deployed on an edge device, it demonstrates real-time performance and scalability. While it currently lacks video tracking and accident prediction, future work will address these. Overall, it offers a reliable solution for enhancing railway safety and efficiency. Future improvements could include integrating thermal or infrared sensors for better visibility in low-light conditions, deploying the model on edge devices like Jetson Nano, and using federated learning for private on-device updates. Expanding the dataset to cover varied weather, lighting, and locations would also enhance model robustness.

REFERENCES

[1] Ministry of Railways, *Annual Accident Report 2022*, Government of India, 2022.
[2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 21–37.
[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 779–788.
[4] Ultralytics, *YOLOv8 Documentation*, 2023.
[5] H. Zhang, Y. Li, Q. Wu, and M. Sun, "Improving Object Detection in Challenging Environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1670–1683, 2019.
[6] L. Wang, X. Chen, and J. Liu, "Evaluating Detection Models on Custom Datasets," in *NeurIPS Workshop*, 2020.
[7] Y. Wu, S. Zhao, Z. Xing, Z. Wei, Y. Li, and Y. Li, "Detection of Foreign Objects Intrusion into Transmission Lines Using Diverse Generation Model," 2023.
[8] C. Meng, Z. Wang, L. Shi, Y. Gao, Y. Tao, and L. Wei, "SDRC-YOLO: A Novel Foreign Object Intrusion Detection Algorithm in Railway Scenarios," 2023.
[9] Z. Wu, H. Zhang, P. Wang, and Z. Sun, "RTIDS: A Robust Transformer-Based Approach for Intrusion Detection System," 2022.
[10] J. Fu, J. Liu, H. Tian, Y. Li, and H. Lu, "Dual Attention Network for Scene Segmentation," in *Proc.*

IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 3146–3154.
[11] C. Shorten and T. M. Khoshgoftaar, “A Survey on Image Data Augmentation for Deep Learning,” *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
[12] PyTorch Foundation, *PyTorch Framework Documentation*, 2023.
[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988.
[14] NVIDIA, *Jetson Nano Developer Kit*, 2023.
[15] G. Jocher, J. Chaurasia, A. Qureashi, and K. Stoken, “YOLO by Ultralytics,” *GitHub Repository*, Jan. 2023.
[16] C. Wang, A. Bochkovskiy, and H. M. Liao, “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,” *arXiv preprint arXiv:2207.02696*, 2022.