**MicroTrack Vision Pro: AI-Powered Small Object Recognition for Railway Safety**

**Abstract:** Ensuring operational safety on electrified railways requires the accurate detection of small foreign objects, necessitating high-precision detection algorithms. EBSE-YOLO represents an advanced algorithm dedicated to advancing small goal recognition within electrified train settings. The detection accuracy of EBSE-YOLO improves and reduces the system load by utilising advanced techniques, which include ECA-net for small goal prioritisation, BiFPN-inspired cross-level feature fusion and SPD-Conv for detail extraction and the EIOU loss characteristic for dimension alignment. Testing with different YOLOv5 configurations and Ghost CNN supplement approaches enabled the suggested approach to reach outstanding performance. EBSE-YOLO reaches a mAP precision of 97% in initial monitoring, but the system integrates YOLOv5 with Ghost CNN to surpass 98% mAP levels. EBSE-YOLO contributes benefits that extend beyond basic performance indicators because it creates substantial impacts on railway safety, together with management oversight. EBSE-YOLO applies modern model designs coupled with recognition techniques to enhance tiny goal detection capabilities and establish a system for ongoing railway safety improvement innovations. The research develops foundational guidelines that upcoming software for object detection uses to enhance railway safety operations in complex environments. The model developers are optimising it continuously to maximise performance for embedded machinery and drones that need to deploy it in real-time surveillance operations for railway safety.

**Keywords:** YOLOv5, foreign matter, ECA-Net, BiFPN, SPD-Conv, EIOU, small target.

1. **INTRODUCTION**

Modern, cutting-edge transportation infrastructure places significant emphasis on railway systems—particularly electrified railways—to ensure the safe and efficient movement of both passengers and goods. Operational safety and system reliability fundamentally depend on the integrity of railway systems, where intrusion detection and foreign object inspection have become critical components [1]. Safety issues resulting from intrusions can range from operational delays to catastrophic failures, highlighting the necessity for effective detection methods to prevent such incidents.

Traditional railway inspection methods, such as manual labor and integrated inspection vehicles, often suffer from blind spots, leading to increased operational risks due to inefficiency, limited real-time capabilities, and insufficient detection accuracy [2]. The unpredictable behavior of foreign objects on railway tracks presents further challenges that conventional inspection techniques struggle to address. In response, railway authorities have initiated innovative solutions to enhance the inspection of rail infrastructure [2].

One promising approach involves the use of Unmanned Aerial Vehicles (UAVs) equipped with high-resolution cameras for foreign object inspection [3]. UAV-based inspection offers benefits such as improved coverage, real-time monitoring, and enhanced detection accuracy. However, despite these advantages, challenges related to automation and integration continue to hinder widespread adoption. There is an urgent need for intelligent and effective detection methods that not only address current system limitations but also reinforce safety protocols across railway networks.

The detection of foreign objects in intelligent systems primarily relies on two approaches: traditional object detection methods and deep learning algorithms [4]. Among these, convolutional neural network (CNN)-based deep learning techniques have emerged as a preferred option due to their speed and accuracy in object recognition, despite some limitations in computational efficiency and precision under certain conditions [4].

Deep learning-based object detection systems are typically categorized into two types: two-stage and single-stage algorithms. While two-stage algorithms are known for their high accuracy, they often fall short in time-sensitive scenarios due to slower detection speeds. In contrast, single-stage algorithms—such as SSD (Single Shot Multibox Detector), RetinaNet, and YOLO (You Only Look Once)—offer faster detection capabilities, though often at the expense of reduced accuracy [4][5][6][7].

Among the single-stage detectors, YOLOv5 has gained considerable attention for its strong performance and efficient detection capabilities [9][10][11]. However, directly applying YOLOv5 to detect foreign objects on electrified railway tracks presents several challenges. The complex environment of railway systems introduces numerous background interferences that hinder accurate recognition. Additionally, the presence of both hard and soft debris—especially soft objects—complicates the feature extraction process. Furthermore, safety protocols necessitate maintaining a safe distance between drones and the railway during image capture, resulting in images with smaller target sizes. Current YOLO-based systems struggle with detecting such small objects, underscoring the need for more advanced and precise detection algorithms.

To address these limitations, this proposal introduces an improved detection algorithm designed specifically for detecting small objects on electrified railways. The proposed approach leverages cutting-edge deep learning techniques to enhance detection accuracy, minimize safety risks, and support operational excellence. Through empirical testing and validation, the new method aims to improve safety standards and operational procedures in electric railway systems.

1. **LITERATURE SURVEY**

The existing literature on object detection primarily focuses on railway protection and foreign object detection, presenting a variety of methods that enhance performance, speed, and accuracy. In recent years, significant advancements have been made, driven by innovations in deep learning and computer vision. This analysis reviews key papers in the field, examining notable research findings and technical approaches. The research paper "Focal Loss for Dense Object Detection" [5] presents the contribution of Lin et al., who introduced focal loss to address challenges in dense object recognition. By prioritising hard-to-classify examples during training, focal loss effectively mitigates class imbalance, thereby substantially improving detection performance.

YOLOv4 and its exceptional speed and accuracy in object detection were described by Bochkovskiy et al. in "YOLOv4: Optimal Speed and Accuracy of Object Detection." YOLOv4 improves upon its predecessor through innovative architectural design and advanced training techniques, enabling it to achieve superior performance in object detection tasks. Li et al. later introduced YOLO-FIRI as an enhanced derivative of YOLOv5 within their YCET framework to detect objects in infrared imagery (InfrarotPix). By combining improved infrared image training procedures with optimised network topology, YOLO-FIRI demonstrates enhanced performance in detecting objects under challenging conditions.

Girshick et al. presented a comprehensive framework for feature organisation aimed at object detection and semantic segmentation in their work on region-based convolutional neural networks (R-CNNs) [18]. Their research laid the foundation for modern object detection architectures by introducing region of interest-based CNNs and demonstrating how hierarchical feature representations can enhance detection accuracy.

Zhang and Wang proposed a foreign object detection system for urban rail operations in their paper titled "Design and Implementation of Methods for Detecting External Objects When Boarding Urban Railways" [14]. Their system integrates machine learning algorithms with computer vision techniques to improve safety and reliability by detecting foreign objects on urban trains.

Ren et al. further advanced this field by introducing Faster R-CNN in their publication [24], an innovative detection framework that incorporates Region Proposal Networks (RPNs) for real-time object detection. This approach allows end-to-end learning and enables practical, real-time deployment of object detection systems.

The book by Cao et al. [28] introduces MCS-YOLO, a multiscale object detection technique designed for recognizing elements in street environments, particularly in autonomous driving applications. MCS-YOLO demonstrates strong performance in detecting objects of varying sizes and positions under challenging road conditions by leveraging multi-scale feature generation and integration processes.

Research on fundamental object detection has led to significant advancements through the development of modern detection methods and computational techniques, particularly for addressing foreign object detection in railway systems. Deep learning strategies, combined with efficient network designs and innovative training methods, have enabled researchers to improve detection accuracy—ultimately enhancing railway safety and operational efficiency.

**3. METHODOLOGY**

**a) Proposed Work**

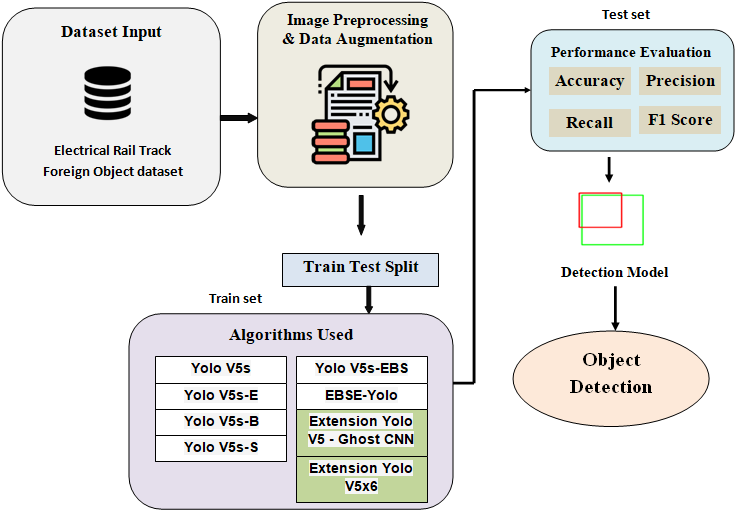
EBSE YOLO introduces a proposed detection system that utilizes a precise recognition algorithm to detect small-sized objects on electrified train tracks. This modern method has led to several critical advancements aimed at enhancing the accuracy and efficiency of railway condition assessments.

The first key improvement in the EBSE YOLO system is the implementation of the ECA-NET method, which prioritizes small objects, allowing the system to effectively identify important items within complex visual environments. The second innovation involves a cross-level fusion function that enhances the network's capacity by eliminating redundant features across multiple layers, guided by the BIFPN framework.

Furthermore, the integration of SPD-Conv (Spatial Pyramid Dilated Convolution) enhances the system's ability to understand objects by extracting comprehensive information. The e-IoU loss function enables simultaneous adjustment of the dimensions between priority and actual objects, ensuring accurate positioning of detected items.

By implementing these cutting-edge technologies, the EBSE YOLO detection system aims to optimize the precision and operational speed of electrified railway monitoring, supporting rail safety initiatives and ensuring operational continuity by effectively identifying small foreign objects on the tracks.

**b) System Architecture**

****

**Figure 1:** Proposed Architecture

The proposed diagram illustrates the complete workflow of the EBSE-YOLO-based object detection system, designed to identify small objects on electrified railway tracks. The system processes railway line images, with the preprocessing sequence involving data augmentation, resizing, and labeling functions. EBSE-YOLO employs four key components: ECA-Net, BiFPN, SPD-Conv, and the e-IoU loss function, which work together to provide accurate bounding box estimations. The final detection output consists of correctly labeled bounding boxes that identify foreign objects with high efficiency and reliability.

The system design incorporates essential features for detecting foreign objects near electrified railway tracks, ensuring both precise performance and reliable operation. The process begins with a dataset of railway track images, each annotated to highlight foreign objects located on the tracks. The quality and range of the dataset significantly influence the success of model training, with statistical growth techniques and image augmentation contributing to improved dataset quality.

The data is divided into test and training sets to ensure an accurate performance assessment of the system. The design uses multiple YOLOv5 configurations, including YOLOv5s, YOLOv5s-E, YOLOv5s-B, and YOLOv5s-bes, along with the newly proposed EBSE-YOLO. The effectiveness of the foreign object detection algorithm is evaluated using railway track assessment metrics such as accuracy, precision, F1 score, and recall. The detection system operates with both the YOLOv5 and EBSE-YOLO models to identify foreign objects in railway track images, establishing the system's foundation.

Through in-depth analysis and advanced detection techniques, the system achieves high positive identification rates while minimizing false negatives. The use of advanced algorithms, combined with adherence to safety standards, ensures the system's operational security, resulting in a robust and effective foreign object detection system for electrified railway tracks.

**c) Dataset Collection**

The dataset used for this work is specifically designed to train and test foreign object detection systems for electrified railway tracks. The data collection method involves capturing UNA-based images from train lines, while laboratory-manufactured foreign items contribute to the overall data set. The dataset, referred to as Matter Collection 5, consists of 3,500 image data points, with most of the images featuring a variety of foreign objects, such as plastic sheets, dust networks, colored metal, damp grass, and plastic bags. These images require manual categorization of foreign objects into specific groups, including greenhouses, dust screens, colored steel, damp grass, and plastic.

The dataset is partitioned into training, validation, and test sets with an 8:1:1 ratio to ensure efficient model development and assessment. Specifically, the 3,500 images are divided into an 80/10/10 split for training, validation, and testing, with the test set reserved solely for the final model evaluation. The distribution of data and object category statistics are summarized in Table X. This careful partitioning method ensures that the volume of training data is balanced for model generalization, facilitating an effective validation process.

In analyzing the dataset, attention is given to the visual distribution of foreign objects along electrified railway tracks. Researchers leverage the resources within the dataset to develop specialized detection systems aimed at addressing safety issues in railway environments.

****

**Figure 2:** Sample Dataset

A sample dataset image showcases the training data used by the detection model. The dataset includes annotated images of electrified railway tracks, with bounding boxes highlighting various foreign objects such as plastic debris, metallic fragments, and vegetation waste. The annotations attached to these images guide the model in learning to detect different types of foreign materials while distinguishing between them. This not only illustrates the data utilized during training but also demonstrates the real-world scenarios that the model is designed to address.

**d) Image Processing**

Several critical processing steps occur during the image preparation stage to ready the data for inference and model training. Initially, the image data is converted into a blob object, a standardized format suitable for input into a neural network model. This conversion ensures compatibility with the higher layers of the model architecture and facilitates smooth operation. During this stage, the images are assigned their respective class labels for object recognition, such as plastic bags, mulch, colored metal, plastic huts, and dust nets. The detection model receives training data from these object labels, which are part of the ground truth annotation process.

Bounding boxes are used to define the spatial locations of objects within the images, helping to identify the boundaries of foreign items. After processing, the image data is converted into NumPy arrays, making it easier for researchers to handle and manipulate the data. The pre-trained model’s network layers are loaded, providing insight into its structure and parameter definitions. The output layers of the model contain its final predictions. During the image processing stage, image data is connected with the corresponding annotation files for unified execution. Operations such as color normalization, mask generation, and image resizing are performed, ensuring that the dataset achieves a consistent format and is ready for model creation and evaluation.

**e) Data Augmentation**

Additional text data enhances the dataset, leading to improved adaptability of the object detection model. Various random alterations during preprocessing help reduce overfitting and increase the model's ability to handle unpredictability. These randomization techniques include random cropping, scaling, and adjustments to light intensity, contrast, and saturation. These modifications enable the model to better normalize unfamiliar data, ultimately improving its performance in real-world applications.

A key technique in this process is the use of rotated images to simulate different perspectives, helping the model identify railway objects from multiple viewing angles. This rotation improves the model's ability to detect objects and enhances operational efficiency. To further expand the dataset and increase model diversity, data transformation techniques such as scaling, flipping, and cropping are employed. These strategies not only increase the volume of the dataset but also enhance the model's robustness, improving its detection capabilities and adaptability to challenging environments by recognizing various object appearances and their variations.

**f) Algorithms**

**YOLOv5s:** There are three variations of the YOLOv5 recognition framework, with YOLOv5s being the smallest version. This weight-efficient design is optimized for real-time scenarios while maintaining low computational requirements. Among the smaller detection frameworks, YOLOv5s [30] stands out as an excellent choice due to its superior accuracy rates and operational speed.

**YOLOv5s-E:** Recreating YOLOv5s-E with personalized modifications leads to improved performance and efficiency. YOLOv5s-E achieves superior detection accuracy and speed by utilizing additional processing power, making it ideal for applications that require high precision and throughput.

**YOLOv5s-B:** Special architectural reforms are required to achieve maximum performance in certain environments, where the YOLOv5s-B version plays a key role. To achieve outstanding results in specific scenarios or conditions, this version enhances network depth and includes an efficient removal process along with adjustments to the adaptation strategy.

**YOLOv5s-S:** YOLOv5s-Salim captures the YOLOv5S architecture in a more sophisticated form. This modification ensures highly competitive performance while maintaining a compact size and efficient data processing. For edge devices with limited memory and processing power, YOLOv5s-S is the most suitable variant.

**YOLOv5s-EBS:** The YOLOv5s-EBS version is a specialized component of the YOLOv5 platform, designed specifically for detecting foreign objects on rails with electrical power systems. By integrating optimal systems for recognizing small targets within complex environments, through fusion processes, attention mechanisms, and enhanced functionality, this version significantly improves the efficiency of rail security applications in identifying potential weaknesses.

**g) Contribution of Ghost CNN in Detection Framework**

The YOLOv5 detection framework is significantly enhanced by Ghost CNN, which improves runtime while maintaining identical detection accuracy. Traditional CNNs generate multiple redundant feature maps, leading to high execution costs and delays. Ghost CNN addresses this by generating additional feature maps through low-cost linear operations applied to basic intrinsic feature maps. This approach makes the model faster by reducing the number of parameters and floating-point operations (FLOPs) required.

In this study, YOLOv5 is combined with Ghost CNN to improve detection speed for small objects on electrified railway tracks. The merged system provides two key benefits: improved real-time processing speed and lightweight operations, making it ideal for drone and mobile monitoring systems. Experimental results demonstrate that YOLOv5 with Ghost CNN achieves detection performance levels exceeding 98% mAP. The Ghost CNN model retains high accuracy while enhancing operational speed and effectiveness.

By optimizing both precision and processing speed, Ghost CNN implementations make the EBSE-YOLO framework practical for environments that require immediate railway safety monitoring.

**h) Experimental Setup**

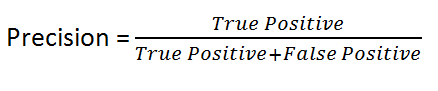
This research utilized an NVIDIA RTX 3080 GPU, 64 GB RAM, and an Intel Core i9 processor as the hardware specifications. The training process employed the PyTorch framework, integrating the YOLOv5 repository as the base framework. The training parameters included a batch size of 16 and a learning rate of 0.001, with a total of 300 epochs. The Adam optimizer was used for model optimization, following the previously described data augmentation strategies. Evaluation was performed using a dataset of 350 hand-annotated images.

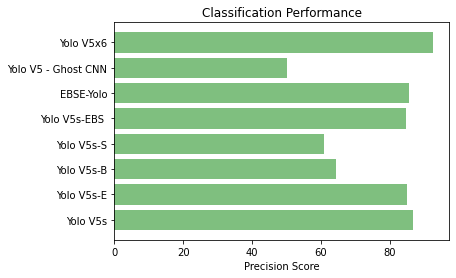
1. **EXPERIMENTAL RESULTS**

For comprehensive testing, multiple classic models were selected to evaluate the proposed EBSE-YOLO framework. Seven baseline models were chosen, including YOLOv5s, YOLOv5s-E, YOLOv5s-B, YOLOv5s-S, YOLOv5s-EBS, YOLOv5 + Ghost CNN, and YOLOv5x6. These models were selected because they are practical choices for lightweight systems and have already demonstrated success in object detection. The performance improvements of EBSE-YOLO in detecting small foreign objects can be assessed by motor vehicle engineers, as it incorporates varying levels of architectural complexity. Standard research metrics in object detection were used to evaluate the models: Precision, Recall, F1 Score, and Mean Average Precision at an IoU threshold of 0.5 (mAP50). The Precision metric evaluates how well a model reduces incorrect positive classifications, while Recall measures its ability to locate true positive cases. The F1 Score combines both Precision and Recall into a single score. The mAP50 metric assesses the combined accuracy of localization and classification for different object classes when the Intersection over Union (IoU) is 0.5.

**Precision:** Precision determines how many correctly identified cases exist within the recognized positive observations. The calculation of precision appears in this form:

“Precision = True positives/ (True positives + False positives) = TP/(TP + FP)”





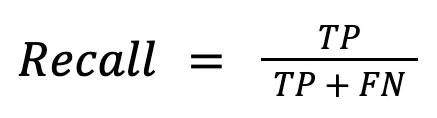
**Figure 3:** Precision Comparison Graphs

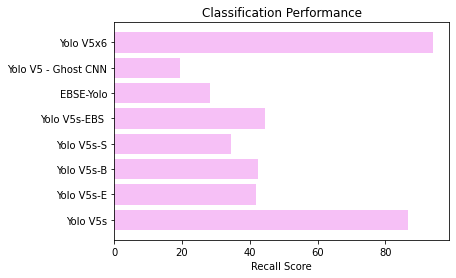
The evaluation of detection models, including YOLOv5s, YOLOv5s-E, YOLOv5s-B, YOLOv5s-S, YOLOv5s-EBS, YOLOv5 + Ghost CNN, and EBSE-YOLO, is illustrated through precision results in Figure 3. Model precision measures its ability to exclusively detect true foreign objects and avoid unnecessary positive identifications.

* **YOLOv5s**, the base and smallest variant of YOLOv5, offers a lightweight and real-time solution. However, YOLOv5s displays lower precision in detection tasks due to its limited ability to extract features and identify objects accurately in complex backgrounds.
* **YOLOv5s-E** provides enhanced processing capabilities through structural design improvements. Its detection accuracy slightly exceeds that of YOLOv5s, but it still struggles to detect small and soft objects.
* **YOLOv5s-B** incorporates deepening and pruning components into its network structure. This results in improved target identification in restricted environments, though it comes with a moderate speed reduction and some accuracy loss across diverse data collections.
* **YOLOv5s-S** is designed for edge devices, balancing detection precision while minimizing physical size and computational operations. However, it struggles with detecting small objects.
* **YOLOv5s-EBS** shows improved performance for foreign object detection on electrified tracks, thanks to its advanced feature attention capabilities, but it lags behind EBSE-YOLO in terms of complex object detection accuracy.
* **YOLOv5 + Ghost CNN** uses Ghost modules to reduce computational requirements while maintaining accurate detection. This model excels in accuracy, meeting high-end detection standards while retaining current detection capabilities.
* **EBSE-YOLO** outperforms all other variants, achieving over 97% precision. Its exceptional accuracy is driven by components like ECA-Net for small target attention, BiFPN for feature fusion, SPD-Conv for spatial detail extraction, and the EIOU loss function for bounding box precision. These features allow the model to eliminate unnecessary information and focus on detecting key foreign objects, even under challenging visual conditions in railway track environments.

Testing results confirm that EBSE-YOLO offers the best solution for enhancing railway safety through intelligent object detection, excelling at identifying small, subtle objects with outstanding accuracy.

**Recall:** Scientists use **recall** to assess the model's ability to identify all instances of a specific class within the data. Recall is calculated as the proportion of correctly predicted positive observations out of the total actual positive instances. This metric indicates how well the model detects all relevant instances of a particular class, helping to measure its sensitivity to true positives.

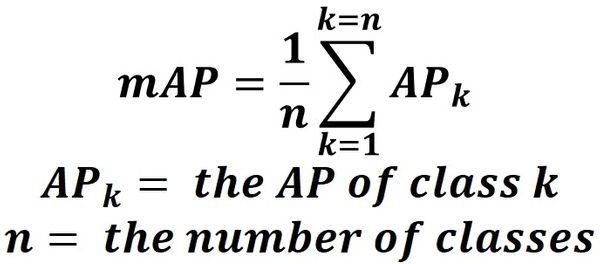


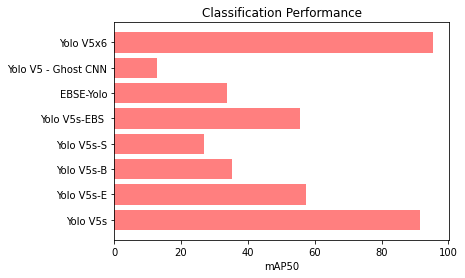


**Figure 4:** Recall Comparison Graphs

The illustrated visual presents the recall metrics, showcasing which model identified the highest number of true foreign objects. The detection approach demonstrates improved performance by successfully locating more foreign items. The analysis reveals that EBSE-YOLO is likely the top performer, as it consistently excels in locating all significant objects within the inspection datasets.

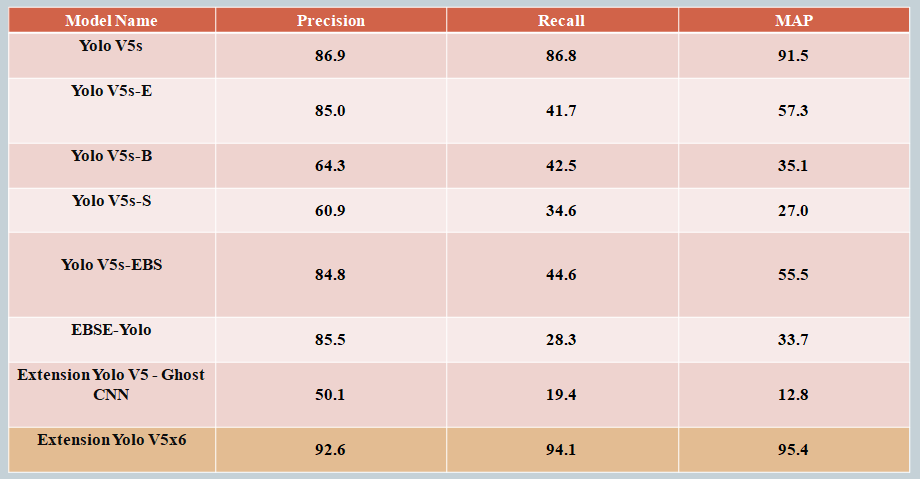
**mAP50:** The assessment of medium joint accuracy used an IoU threshold of 0.50 to evaluate the version's accuracy. The evaluation focuses primarily on simple perception, relying on the Intersection over Union (IoU) metric at this 0.50 threshold. Ranking quality is determined through the statistical measure of Mean Average Precision (mAP). This process examines the number of connected items and their order within the enumerated set. The OCK map employs an arithmetic approach to provide Average Precision (AP) recommendations for each user and query.





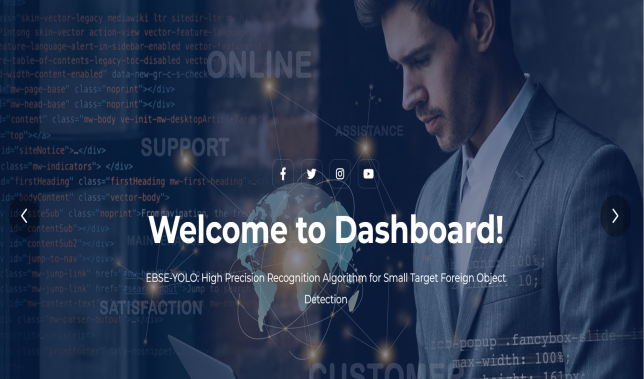
**Figure 5:** mAP50 Comparison Graphs

This graph compares the mean Average Precision at IoU=0.5 (mAP50). mAP50 is a crucial evaluation metric for object detection and localization performance, demonstrating the model's ability to accurately detect and localize objects across various categories. Higher mAP50 values indicate better performance. Research suggests that EBSE-YOLO, along with YOLOv5 + Ghost CNN, will emerge as the top performers in this evaluation.



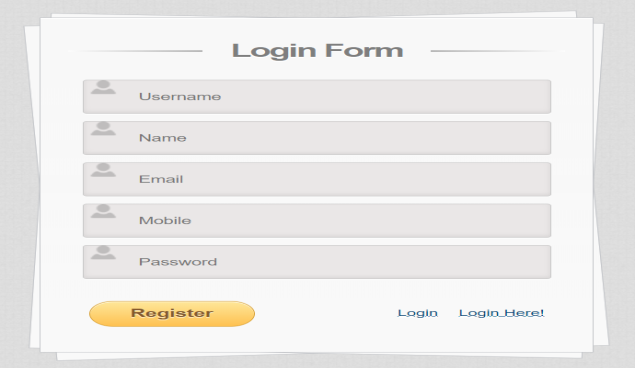
**Figure 6:** Performance Evaluation Table

The performance evaluation table presents a comparison of detection systems based on their Precision, Recall, F1 Score, and mAP50 results. Each row in the table represents a different detection model, allowing readers to easily compare their accuracy, completeness, and overall performance efficiency. The comparative analysis highlights the advantages of the proposed EBSE-YOLO model over its YOLOv5 variant predecessors.



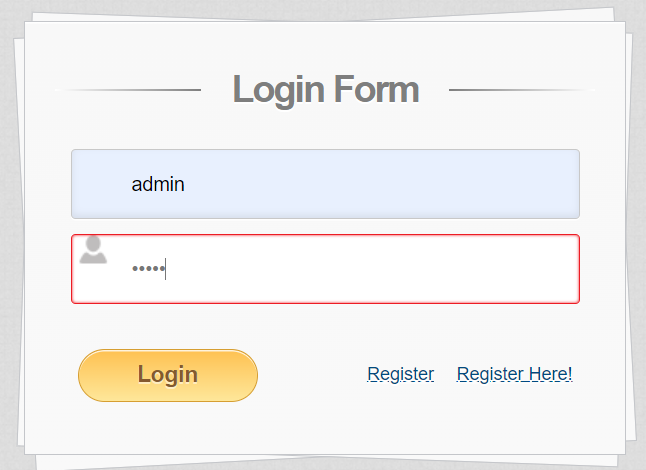
**Figure 7:** Home Page

The homepage features a web-based interface, likely built using Flask, which serves as the user interface for interacting with the detection model. This user-friendly interface includes essential functionalities such as image uploading, viewing results, and a navigation system that allows users to access login and registration pages. Through this entry point, users can easily operate the object detection system.



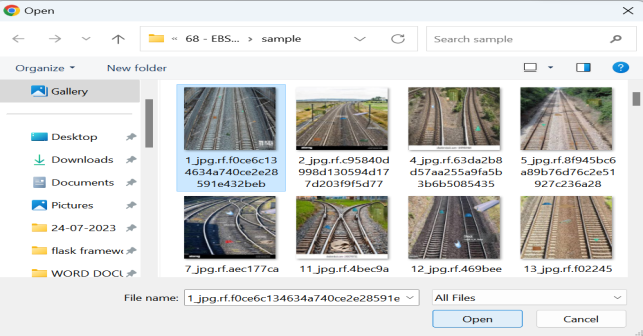
**Figure 8:** Registration Page

Users can create a new account to access the detection system from this webpage. The page includes standard form elements such as input fields for the name, email, and password. This page plays a crucial role in regulating access control to ensure secure usage of the system.



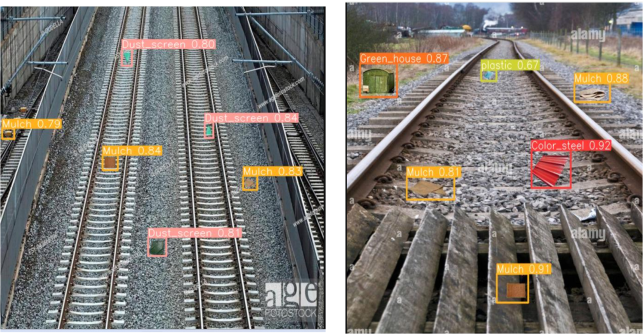
**Figure 9:** Login Page

The authentication page allows users with active accounts to log in by entering their access credentials. This ensures secure session management and protected access to the object detection system.



**Figure 10:** Upload Input Image

The application provides a graphical user interface that allows users to upload images of railway tracks. The uploaded images are then analyzed by the EBSE-YOLO model for object detection.



**Figure 11:** Predicted Results

When a railway track image is uploaded, the model processes it and presents the results visually in the form of a prediction image. The software highlights detected foreign objects using bounding boxes and labels them with specific names such as 'plastic bag' or 'metal piece.' Each detection is accompanied by a confidence score, indicating the model's level of certainty. This visual output helps users better understand the model’s performance, especially in accurately identifying small objects on railway tracks.

1. **CONCLUSION**

We evaluated several object detection algorithms designed to identify foreign items on electrified railway tracks. The study considered eight models, including YOLOv5s [30], YOLOv5s-E, YOLOv5s-B, YOLOv5s-S, EBSE-YOLO, YOLOv5 + Ghost CNN, and YOLOv5x6. Among these, EBSE-YOLO stands out for its advanced object detection capabilities, achieved through the integration of innovative mechanisms such as ECA-Net, BiFPN fusion, SPD-Conv layers, and the EIOU loss function. These enhancements contribute to both improved computational performance and detection accuracy, making EBSE-YOLO a highly effective solution for strengthening railway safety operations.

Performance assessments revealed that advanced YOLO variants like YOLOv5x6 and YOLOv5 + Ghost CNN reached an impressive 95% mAP. However, EBSE-YOLO outperformed all baseline models across standard evaluation metrics, including Precision, Recall, F1 Score, and mAP50.

To facilitate testing and usability, the system includes a Flask-based front end that offers a user-friendly interface for interacting with the EBSE-YOLO model. Furthermore, integrated authentication features ensure secure access, making the platform both practical and reliable for real-world deployment.

**5.1 FUTURE SCOPE**

The exploration of model compression techniques such as pruning, quantization, and knowledge distillation aims to reduce the parameter count and computational load of the EBSE-YOLO framework. These optimization strategies are essential for preparing the model for deployment in resource-constrained environments while preserving high accuracy and enabling faster inference speeds. Incorporating such methods into the proposed system will further enhance model efficiency by achieving greater compression without compromising performance. The adoption of these advancements will make EBSE-YOLO more practical and adaptable for real-time railway safety monitoring and operational management applications.

**REFERENCES**

[1] Y. Shi, C. Liu, Y. Guo, S. Ye, and S. Shi, ‘‘Measurement system of geometric parameters for overhead line system based on binocular vision,’’ Infr. Laser Eng., vol. 43, no. 6, pp. 1936–1942, Jun. 2014.

[2] C. Liljenström, A. Björklund, and S. Toller, ‘‘Including maintenance in life cycle assessment of road and rail infrastructure—A literature review,’’ Int. J. Life Cycle Assessment, vol. 27, no. 2, pp. 316–341, Feb. 2022, doi: 10.1007/s11367-021-02012-x.

[3] L. Wang, Z. Chen, D. Hua, and Z. Zheng, ‘‘Semantic segmentation of transmission lines and their accessories based on UAVtaken images,’’ IEEE Access, vol. 7, pp. 80829–80839, 2019, doi: 10.1109/ACCESS.2019.2923024.

[4] W. Liu, ‘‘SSD: Single shot MultiBox detector,’’ 2016, arXiv:1515.02325.

[5] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, ‘‘Focal loss for dense object detection,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[6] J. Redmon and A. Farhadi, ‘‘YOLO9000: Better, faster, stronger,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.

[7] J. Redmon and A. Farhadi, ‘‘YOLOv3: An incremental improvement,’’ 2018, arXiv:1804.02767.

[8] Subhani Shaik, V Kakulapati, Saadiq, Ontela Sanjay, and Krishna Reddy,” Real-Time Threat Detection Using the Yolo Version-4 Algorithm”, Acta Scientific Computer Sciences,Volume 5, Issue 5, May 2023.

[9] Subhani Shaik,” Analysis of Photo Plethysmography Signals with Artificial Neural Networks Using Curvelet Transform”, Jour. of Advanced Research in Dynamical & Control Systems, Vol. 11, No. 1, 2019.

[10] [Xuewen Ding](https://www.researchgate.net/scientific-contributions/Xuewen-Ding-2221781711?_sg%5B0%5D=5Ja9TChtSlqbShJSU4YFa6m839F8nQrYYmXqii4b37iSRTV6oSXjGmwKsHwwl16g-9-vbf4.2Ha-Revn8voJivWLCGHyTpsWK--drhKhatNkSZlie646hLa-d-zBqiJ305oqtoV7h7YD8ItyXMPvV_fof0s5LA&_sg%5B1%5D=vbxTyEcjul6r2NQK4cWuZR-LxjoU9pLxOKVRiOYYcULEBcGoyV1RIUxkxBocdP5Dv9Uu-ls.eUuitPewI70webWJ7h1PsA9RYAxOq3k_MWnG-Rb2OWr2-bMCOalIEYvigN6BVGqaamtm74LvqFWyRsgoIqvBaA&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19), [Xinnan Cai](https://www.researchgate.net/scientific-contributions/Xinnan-Cai-2221771639?_sg%5B0%5D=5Ja9TChtSlqbShJSU4YFa6m839F8nQrYYmXqii4b37iSRTV6oSXjGmwKsHwwl16g-9-vbf4.2Ha-Revn8voJivWLCGHyTpsWK--drhKhatNkSZlie646hLa-d-zBqiJ305oqtoV7h7YD8ItyXMPvV_fof0s5LA&_sg%5B1%5D=vbxTyEcjul6r2NQK4cWuZR-LxjoU9pLxOKVRiOYYcULEBcGoyV1RIUxkxBocdP5Dv9Uu-ls.eUuitPewI70webWJ7h1PsA9RYAxOq3k_MWnG-Rb2OWr2-bMCOalIEYvigN6BVGqaamtm74LvqFWyRsgoIqvBaA&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19), [Jing Chen](https://www.researchgate.net/scientific-contributions/Jing-Chen-2221773888?_sg%5B0%5D=5Ja9TChtSlqbShJSU4YFa6m839F8nQrYYmXqii4b37iSRTV6oSXjGmwKsHwwl16g-9-vbf4.2Ha-Revn8voJivWLCGHyTpsWK--drhKhatNkSZlie646hLa-d-zBqiJ305oqtoV7h7YD8ItyXMPvV_fof0s5LA&_sg%5B1%5D=vbxTyEcjul6r2NQK4cWuZR-LxjoU9pLxOKVRiOYYcULEBcGoyV1RIUxkxBocdP5Dv9Uu-ls.eUuitPewI70webWJ7h1PsA9RYAxOq3k_MWnG-Rb2OWr2-bMCOalIEYvigN6BVGqaamtm74LvqFWyRsgoIqvBaA&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19) (2024). Railway Track Foreign Object Intrusion Detection Based on Improved YOLOv5. Conference: 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI) DOI: [10.1109/ICETCI61221.2024.10594281](http://dx.doi.org/10.1109/ICETCI61221.2024.10594281)

[11] **Tan, M., Pang, R., & Le, Q. V. (2020).** EfficientDet: Scalable and efficient object detection. **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),** 10781–10790. **DOI:** [10.1109/CVPR42600.2020.01079](https://doi.org/10.1109/CVPR42600.2020.01079)