Fast Sort YoloV11 Object Detection approach towards detection and classification of Foreign Objects in Railway lines and Tracks with Multiattention mechanism

[1] Malle Srihari, [2] Mathina Prudhvi Sankar, [3] Nakkala Venkata Mahesh, [4] Pyneni Sathwik, [5] Ms. K. Rajeswari

[5] Assistant professor, Mail: rajeswarik@mahendra.info

[1] mallesrihari69@gmail.com, [2] mathinaprudhvisankar@gmail.com, [3] maheshnakkala710@gmail.com, [4] pynanesathwik527@gmail.com

ABSTRACT:

Due to rapid increase in usage of railway transformation across the world due to its multiple benefits such as cost effectiveness, reliability and environment friendliness for long distance transportation of passengers and goods. Thus it becomes extremely significant to detect the foreign objects intrusion on the high speed railway lines and its track beds to safeguarding train operations. Traditionally many Object detection methods have been employed to detect the foreign objects in railway lines to enhance safety, operational efficiency, and the overall reliability of rail transportation. Despite of several advantages, those architectures fails to address the following challenges such as feature extraction and aggregation of the object detection and recognition approaches increase the complexity of the model and it leads to poor efficiency in detecting the variation in the object appearance due to viewpoint, deformation , occlusion and lightening conditions. In this paper, new fast sort Yolov11 object detection architecture is designed to enhance the detection accuracy and reduce complexity of the model on incorporation of attention mechanism to detect the foreign objects in railway track and railway lines particularly small and occluded object effectively. In this work, RailFOD23 dataset is extracted from Figshare repository for training and testing the model which gathered data using LiDAR, radar, thermal imaging, and sensor networks. Proposed architecture is composed of multiple components to perform object detection. Initially extracted data is applied to backbone component which contains DenseNet121 architecture is to segment and extract the features of the image through convolution layer and fully connected layer of the model. Extracted features are employed to Neck component which performs feature aggregation for better feature presentations. Finally aggregated features is processed in the head component which detect and classify the foreign object in both railway line and railway track with high detection accurately and efficiency. Experimental analysis of the proposed model is performed and it is identified to efficient to process diverse objects such as trains, maintenance equipment, trespassers, or obstructions. Further performance analysis of the model is performed on basis of detection accuracy and detection efficiency, proposed model produces 97.2% accuracy and it is proved to be highly efficient while compared to existing object detection architectures.

KEYWORDS: Artificial intelligence, Object detection, Object Recognition, Deep learning, RailFOD23 dataset, YoloV11 architecture, Attention Mechanism

1. INTRODUCTION

Modern transit is based on the railway system, which makes it easier to carry people and products over long distances. The sustained success of railway networks depends critically on maintaining their safety, dependability, and effectiveness. These objectives can be greatly advanced by object detection along railway lines. In order to enable prompt intervention and preventative measures, it entails the identification and classification of objects or abnormalities that could have an influence on railway operations [1]. Railway networks connect urban centres, commercial hubs, and isolated areas across nations and continents. They support economic growth and development by acting as the foundation of trade and transportation. These networks require sophisticated technologies for preserving safety, improving operations, and reducing disruptions because to their sheer size and complexity [2].

Object detection in railway lines handles a wide range of issues, including preventing train crashes with obstacles, keeping an eye on track conditions, preventing unauthorised access, and even detecting wildlife crossings [3]. The capacity to recognise and react to objects or occurrences along the railway lines has undergone a revolutionary change because to technological advancements, particularly in the areas of computer vision, sensors, and machine learning. In addition to enhancing safety and operational effectiveness, this field has helped the train industry continue its continuing digital transformation[4]. The techniques used to detect objects range greatly, from conventional sensor-based systems to state-of-the-art machine learning algorithms. A few examples of the technology used to identify objects and abnormalities are LiDAR, radar, cameras, and thermal imaging. These systems are essential for asset protection, traffic management, and predictive maintenance in addition to immediate safety concerns. Additionally, this system sets the stage for comprehending the relevance of object detection on railway lines and gives an overview of its significance in contemporary rail transportation. Fig 1 shows the vision-based obstacles detection. [5]

Traditionally many Object detection methods have been employed to detect the foreign objects in railway lines to enhance safety, operational efficiency, and the overall reliability of rail transportation. Despite of several advantages, those architectures fails to address the following challenges such as feature extraction and aggregation of the object detection and recognition approaches increase the complexity of the model and it leads to poor efficiency in detecting the variation in the object appearance due to viewpoint, deformation , occlusion and lightening conditions. In this paper, new fast sort Yolov11 object detection architecture[6] is designed to enhance the detection accuracy and reduce complexity of the model on incorporation of attention mechanism to detect the foreign objects in railway track and railway lines particularly small and occluded object effectively.

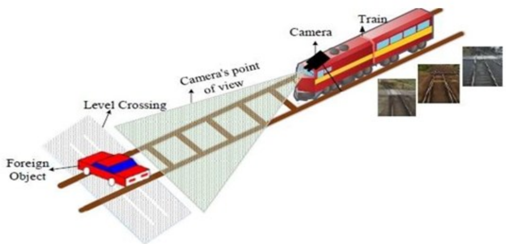

Figure 1: Object Detection Prototype in the Railway lines

RailFOD23 dataset[7] is extracted from Figshare repository for training and testing the model which gathered data using LiDAR, radar, thermal imaging, and sensor networks[8]. Proposed architecture is composed of multiple components to perform object detection. Initially extracted data is applied to backbone component which contains DenseNet121 architecture is to segment and extract the features of the image through convolution layer and fully connected layer of the model. Extracted features are employed to Neck component which performs feature aggregation for better feature presentations. Finally aggregated features is processed in the head component which detect and classify the foreign object in both railway line and railway track with high detection accurately and efficiency.

The remaining part of the article is sectioned as follows, section 2 mentions the review of literature related to object detection technique for foreign object detection and classification in railways. Section 3 defines the proposed methodology for enhancing object detection and classification accuracy of the foreign objects occurrence in railway lines and tracks using a new fast sort yolov11 architecture. Section 4 discusses the experimental results along performance analysis of proposed model against the existing object detection architecture with respect to the detection accuracy. Finally section 5 concludes the article

2. RELATED WORK

In this section, various literature on basis of object detection and classification architectures towards foreign object detection in the railway line and railway tracks has been analysed on multiple aspects are follows

Cristian wisultschew, et.al,[1] implemented an embedded implementation for real-time object detection and tracking algorithm which is used within level crossing scenarios. The literature presents mechanism to reduce the number of accidents in high-risk areas by monitoring the railway level crossing to inform the train driver about the existence of possible obstacles, warrantying a fast response to avoid potential accidents..

Maoli wang, et.al[2] proposed a MRCNN to edge position defects which occurs in rail surface through processing information about small size defects using feature extraction, and semantic segmentation.

3. PROPOSED MODEL

In this section, design of new fast sort YoloV11 architecture composed of multiple processing components to perform foreign object detection in railway lines and railway tracks are as follows

3.1. Backbone Network –DenseNet121

Backbone Network acts as object segmentation and feature extraction module or component. It incorporates DenseNet121 architecture[11] from convolution Neural Network. DenseNet121 composed of multiple convolution layers with different convolution and kernel function to segment the small and occluded object effectively. Furthermore, it supports multi-object segmentation, identifying multiple instances of the same object class within a single image, a valuable feature in scenarios with numerous objects of the same category.

Segmentation model can generate region proposals or superpixels, grouping contiguous pixels with the same semantic label, creating coherent regions that serve as candidate regions for subsequent object detection. Additionally, it allows the generation of regions of interest (ROIs), pinpointing areas in the image where objects are present[12]. The precise delineation of object boundaries by semantic segmentation guides object detection algorithms to localize objects more accurately.Those extracted objects are further extracted using different convolutions to extract the features of multiple scales. Processing step of the backbone network towards object segmentation and classification is as follows

- **Convolution Layer**

Convolution layer perform segmentation and feature extraction process simultaneously on training data. Initially convolution layer uses kernel function[13] to extract the region of interest of the object on basis of the variation in pixel intensity and it is represented as image segments. Segmentation process is as follows

Represents the image as g(x,y)

Compute mean and standard deviation for each pixel of the image

If (g(x1,y1)>(g(x2,y2)

Segment the g(x1,y1) separately from g(x2,y2) and represents g(x1,y1) as segment 1 and g(x2,y2) as segment 2

Each segment is further processed different convolution layer with kernel function and activation function to extract the features. Those features are considered as low level feature and it is organized in form feature map.

Convolution Layer C = Kernel (Image Segment)

Extract low level features (Color, Intensity and texture)

Feature Map $F_m = \sum_{i=0}^{n} f(x_i, y_i)$w

Where W is the segmented window and $f(x_i,y_i)$ is the feature of the segment.

- **Max Pooling Layer**

Max Pooling layer is considered as down sampling layer which down samples the image features of the segment. Down sampling operation of the max pooling layer obtain the high level features on incorporating spatial and temporal attention mechanism to extract the spatial features and temporal features separately. It produces the spatial feature and temporal features of the segment. High level feature map of the segment of spatial and temporal feature are represented as

Spatial Feature map $HSF_m = \sum_{i=0}^{n} sf(x_i, y_i)$

Temporal Feature map $HTF_m = \sum_{i=0}^{n} Tf(x_i, y_i)$w

Where SF is spatial feature and TF is temporal features.

3.1.2. Neck Component

Neck Component performs the aggregation of the spatially and temporal features to enhance the feature presentation. Further Neck components aggregate the multilevel spatial features[14] and multilevel temporal features on basis of the gradient association. Gradient association determines the correlation of the feature on specified gradient of the image. Associated spatial and temporal feature on the specified gradient is transformed into unique spatial representation.

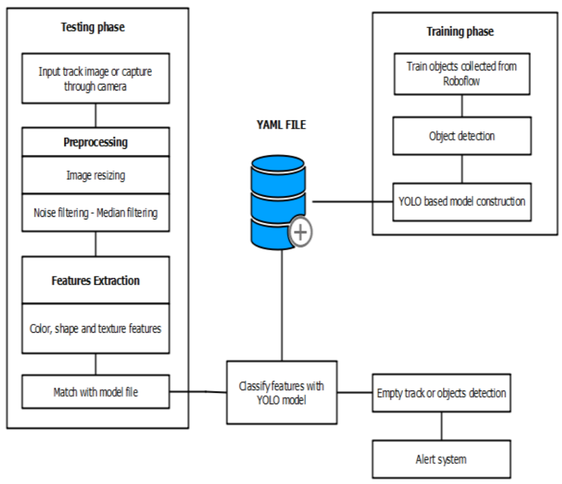

Figure 2: Proposed architecture

Figure 2 represent the architecture of the proposed model. Especially Neck component uses the convolution layer of the backbone to process spatial feature maps and temporal features map towards aggregation and transition of features with kernel. Further resultant feature is considered as essential feature for object detection[15].

$$\text{Feature Aggregation } AF_m = G_f(HSF_m + HTF_m)$$

$$\text{Feature Transition } T_s = T(HTF_m)_s$$

Where G_f is image gradient and $T(HTF_m)_s$ is transistion of feature into spatial feature

3.1.3. Head Component

Head Component acts as detection mechanism towards object localization and classification on processing feature map as fully connected layer. Fully connected layer of neck component composed of activation function, softmax function and loss function[16]. Activation function linearize the feature map and softmax function uses the classifier to generates the object classes with bounding box coordinates[17] and classes labels to image segments with different occlusions.

$$C_f = \frac{1}{N}(Y_i\ (AF_m\ )$$

Where N is number of feature , Y_i is a Classifier function

Further softmax function computes the object score to differentiate the object between the classes effectively. In this work, random forest classifier is used as softmax function. Finally loss function uses cross entropy function reduces the over fitting and underfitting issues.

Algorithm 1: Fast Sort YoloV11
Input: RailFOD23 Dataset
Output: Obstacles Classes
Process ()
YoloV11 (DenseNet121 ()))
Backbone_Segmentation and Feature Extraction ()
Convolution Layer (Kernel Function, Image)
Feature Map = Low level Feature {Color, Texture, Intensity}
Max Pooling Layer –Spatial Attention & Temporal Attention
Spatial Feature Map = Spatial feature (FeatureMap)
Temporal Feature Map = Temporal feature (FeatureMap)
Neck Component_ Feature Aggregation ()
Transition of Temporal feature into spatial features
Aggregated Features of Spatial features
Head Component _Object Detection and Classification()
Softmax_Random forest Classifier Function ()
Object Score(aggregated Feature Map)
Class = {trains, maintenance equipment, trespassers, or obstructions}

4. Experimental Analysis

Experimental analysis and performance analysis of the foreign object detection and classification model in railway line and railway track were as follows

4.1. Experimental Analysis

Experimental analysis is performed using extracted RailFOD23 dataset in Google colab python environment. Extracted dataset is partitioned into training and testing data for model training and model testing. Model training is performed using hyperparameter setting of the YoloV11 architecture to segment and detect the foreign object occurrences in railway lines and tracks. Table 1 represents the hyperparameter values of the yolov11 model.

Table 1: Hyperparameter of YOLOV11 Model

Parameter	Value
Learning Rate	10^{-8}
Batch Size	20
Epoch	50
Activation Function	ReLu
Softmax Function	Random Forest

4.2. Performance analysis

Performance analysis of the model is performed using test data. Test data of the model is processed using Confusion matrix to obtain following parameter to compute the efficiency of the model is as follows

4.2.1. Precision Analysis

Precision is to measure correctly selected features to the foreign object class among the total extracted features [18]. It is also represented as ratio of correctly selected features among the total extracted features to the particular object class. It is represented as

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{...Eq.4}$$

Where TP is considered as true positive and FP is considered as false positive parameters of the confusion matrix. Figure 3 represents the precision analysis of the fast sort Yolov11 model against traditional yolo models

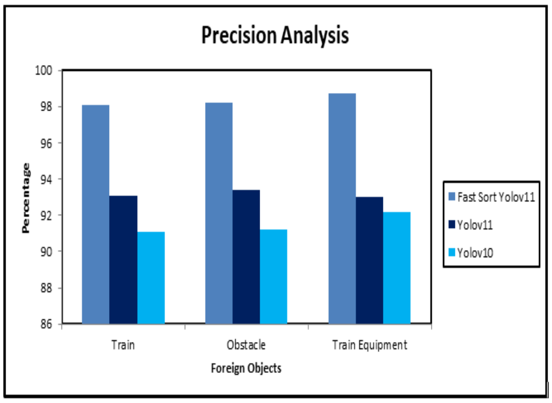

Figure 3: Precision Analysis

4.2. Recall

Precision is to measure incorrectly selected features to the foreign object class among the total extracted features [19]. It is also represented as ratio of incorrectly selected features among the total extracted features to the particular object class. It is represented as

$$\text{Recall} = \frac{TN}{TP+FP} \text{ ...Eq.4}$$

Where TN is considered as true negative, TP is considered as true positive and FP is considered as false positive parameters of the confusion matrix. Figure 4 represents the Recall analysis of the fast sort Yolov11 model against traditional yolo models.

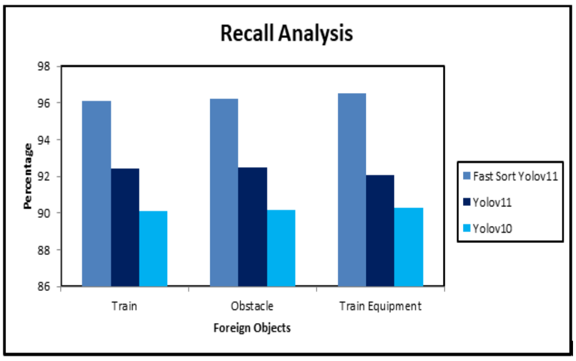

Figure 4: Recall Analysis

4.3. Accuracy

Accuracy is defined as aggregation recall and precision on detecting the features among the total extracted features into different foreign object classes[20]. It is also mentioned as correctly selected features to foreign object class among extracted features from the segment of the images. It is represented as

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} *100 \text{ ... Eq.6}$$

Where TN is considered as true negative, TP is considered as true positive and FP is considered as false positive parameters of the confusion matrix. Figure 4 represents the Accuracy analysis of the fast sort Yolov11 model against traditional yolo models.

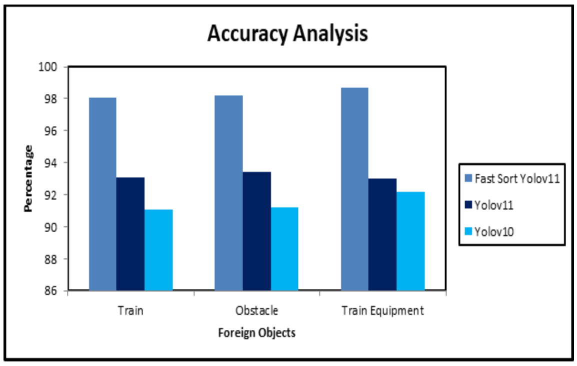

Figure 5: Accuracy analysis

Finally performance of fast sort Yolov11 architecture performs better with detection accuracy 97.2% on compared to conventional approaches. Table 2 mentions the performance evaluation of the yolo architectures in foreign object detection.

Table 2: Performance Evaluation

Technique	Classes	Precision	Recall	Accuracy
Fast Sort YoloV11	Trains	98.1	96.2	97.4
	Train Equipments	98.4	96.5	97.2
	Obstacles	98.8	96.4	97.6
YoloV11	Trains	93.4	92.7	94.7
	Train Equipments	93.6	92.8	94.3
	Obstacles	93.3	92.4	94.1
YoloV10	Trains	91.6	90.1	91.6
	Train Equipments	92.3	90.5	91.3
	Obstacles	93.7	90.2	91.7

6. CONCLUSION

In this work, fast sort Yolov11 object detection architecture is designed and implemented to enhance the detection accuracy and reduce complexity of the model on incorporation of attention mechanism. Proposed model is highly capable in detecting the foreign objects in railway track and railway lines particularly small and occluded object effectively on processing the extracted RailFOD23 dataset from Figshare repository. Proposed architecture process the data towards segmentation and detection on multiple components. In particular, backbone component employs DenseNet121 architecture is to segment and extract the features of the image through convolution layer and fully connected layer. Those Extracted features were processed in Neck component to obtain feature presentations. Finally aggregated features are processed in the head component to detect and classify the foreign object in both railway line and railway track against and evolving situations with high detection accurately and efficiency. Experimental analysis and performance analysis of the proposed model is performed to detect and classify diverse objects such as trains, maintenance equipment, trespassers, or obstructions with accuracy of 97.2% which is high compared to existing object detection architectures. Further it tracks the movement of trains, identifying the presence of trespassers or unauthorized personnel on rapidly changing conditions

REFERENCES

[1] Wisultschew, Cristian, et al. "3D-LIDAR based object detection and tracking on the edge of IoT for railway level crossing." IEEE Access 9 (2021): 35718-35729.

[2] De Donato, Lorenzo, et al. "A survey on audio-video based defect detection through deep learning in railway maintenance." IEEE Access 10 (2022): 65376-65400.

[3] Wang, Maoli, et al. "Detection of surface defects on railway tracks based on deep learning." IEEE Access 10 (2022): 126451-126465.

[4] Zhang, Ziwen, Mangui Liang, and Zhe Wang. "A deep extractor for visual rail surface inspection." IEEE Access 9 (2021): 21798-21809.

[5] Zhang, Xin, et al. "A GANs-based deep learning framework for automatic subsurface object recognition from ground penetrating radar data." IEEE Access 9 (2021): 39009-39018.

[6] Singh, Prashant, et al. "Deployment of autonomous trains in rail transportation: Current trends and existing challenges." IEEE access 9 (2021): 91427-91461.

[7] Sresakoolchai, Jessada, and Sakdirat Kaewunruen. "Integration of building information modeling and machine learning for railway defect localization." IEEE Access 9 (2021): 166039-166047.

[8] Li, Shao Jia, et al. "DF-YOLO: Highly accurate transmission line foreign object detection algorithm." IEEE Access 11 (2023): 108398-108406.

[9] Li, Hui, et al. "An improved YOLOv3 for foreign objects detection of transmission lines." IEEE Access 10 (2022): 45620-45628.

[10] Lainsa, Mikel, and Daeyoung Kim. "Zero-Shot Texture Analysis and Regression-Based Deformation Recognition for Rail Anomaly Detection." IEEE Access (2024).

[11] Klammsteiner, Mirjam, et al. "Vision Based Stationary Railway Track Monitoring System." *2023 33rd Conference of Open Innovations Association (FRUCT)*. IEEE, 2023.

[12] S. Zhai, D. Shang, S. Wang, and S. Dong, ''DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion,'' IEEE Access, vol. 8, pp. 24344–24357, 2020

[13] H. Tan, J. Ding, and P. Tan, ''Insulator fault detection based on edge features,'' J. Zhejiang Univ. Sci. Technol., vol. 34, no. 6, pp. 521–527, Dec. 2022.

[14] L. Zhang, S. Hu, and J. Zhang, ''Classification detection of insulator defects in distribution network based on YOLOv5,'' Electr. Power Sci. Eng., vol. 38, no. 11, pp. 41–48, Nov. 2022.

[15] Z. Wang, Y. Wang, Q. Wang, S. Kang, and V. Mikulovich, ''Two stage insulator fault detection method basedon collaborative deep learning,'' Diangong Jishu Xuebao, vol. 36, no. 17, pp. 3594–3604, 2021.

[16] Z. Qiu, X. Zhu, C. Liao, D. Shi, and W. Qu, ''Detection of transmission line insulator defects based on an improved lightweight YOLOv4 model,'' Appl. Sci., vol. 12, no. 3, p. 1207, Jan. 2022.

[17] L. Li, Y. Zhang, P. Chen, K. Zhang, W. Xiong, and P. Gong, ''Complex scene insulator defect detection algorithm based on lightweight YOLOv4,'' Photoelectron Laser, vol. 33, no. 6, pp. 598–606, Jun. 2022.

[18] L. Tang, M. Yu, M. Wu, and C. Yang, ''Insulator defect detection algorithm based on improved YOLOv5,'' J. Central China Normal Univ., vol. 56, no. 5, pp. 771–780, Oct. 2022.

[19] Z. Feng, L. Guo, D. Huang, and R. Li, ''Electrical insulator defects detection method based on YOLOv5,'' in Proc. IEEE 10th Data Driven Control Learn. Syst. Conf. (DDCLS), May 2021, pp. 979–984.

[20] T. Zhang, Y. Zhang, M. Xin, J. Liao, and Q. Xie, ''A light-weight network for small insulator and defect detection using UAV imaging based on improved YOLOv5,'' Sensors, vol. 23, no. 11, p. 5249, Jun. 2023