Elephant Detection System for Railway Safety: A Performance Comparison of CNN and R-CNN Models

Dr. David Raj Micheal
Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
davidraj.micheal@vit.ac.in

Gowthami A
Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
gowthami.a2023@vitstudent.ac.in

***Abstract*—The collision of elephants on railway tracks is one of the major threats in wildlife and human safety, especially in the Tamil Nadu, Kerala, and Assam states in India, which hold critical elephant corridors that cross railways. This project proposes a deep learning-based detection system for the prevention of accidents by detecting the exact location of elephants near railway tracks in real time. The system is built using two different deep learning approaches: Convolutional Neural Networks (CNN) and Region-based Convolutional Neural Networks (R-CNN), for the exploration of the best methodology in elephant detection. An image dataset of elephants placed on railway tracks was gathered using web scraping over the internet, which is used to train both models. These images are pre-processed and augmented to make the model robust against changing factors such as pose variations, illumination changes, and long-distance capture. The CNN model acts as a baseline in elephant detection, while the R-CNN model offers a more specialized manner of object detection and localization. In this regard, the performance of these two models is presented with their major metrics concerning detection accuracy, speed, and computational efficiency. Upon detecting the presence of an elephant within a range of the railway track, the system sends real-time alerts to the railway and forest department officials about the necessary subsequent action, such as slowing down or completely stopping trains well in time. Comparison of CNN and R-CNN will bring interesting insights into optimizing deep learning models for wildlife conservation and railway safety that can extend applications to other similar regions.**

***Index Terms*—Comparative Analysis; CNN and R-CNN Models; Elephant Detection; Object Detection; Railway Safety**

I. Introduction

The collision of elephants with trains has become a matter of great concern because of its high impact on biodiversity and human safety whenever railway lines cross crucial pathways of elephant movements, such as in Tamil Nadu, India. Traditional methods for minimizing these events have proven inadequate due to resource intensity and the inability of conventional systems to operate effectively in real time, using manual monitoring and physical barriers. Recent advances in deep learning and artificial intelligence enable the development of automated, real-time detection systems that can accurately detect elephants near railway tracks and send alerts to the respective authorities on time for necessary actions.

The focus of this research is the design of a deep learning-based detection and warning system, realizing real-time elephant detection using CNN and R-CNN. The proposed system utilizes a strong dataset comprising web-scraped images and deploys modern image processing to achieve a better detection rate under various environmental conditions. A further comparative study on CNN-R-CNN has also been conducted in this article to find out the most effective deployment of such systems in realistic scenarios. Such deployment would help in the twin causes of wildlife conservation and railways safety.

II. Objectives

The aim of the project is to develop and compare deep learning models targeted at the detection of elephants near railway tracks and to develop an alert system that can reduce elephanttrain accidents. This work focuses on the design and training of CNN and R-CNN models separately for elephant detection on railway tracks. These models are then to be compared on the basis of detection accuracy, detection time, and computational complexity based on different environmental conditions. Moreover, the project aims to provide an unconstrained and varied dataset of images, which shall eventually include images of elephants on the railway tracks, along with pre-processing the same for training purposes by using both the CNN and R-CNN models. It is expected that this will enhance the performance of the models on various scenarios with changes in lighting conditions, poses, and backgrounds. Comparatively, performance analysis of the CNN and R-CNN models will be conducted in regard to detection accuracy, time efficiency, and resource utilization to find out which one could be most suitable for practical applications. The project further intends to assess the impacts that the proposed detection and alert system could have on wildlife conservation and railway safety by considering its potential to reduce the number of elephant fatalities on railway tracks while enhancing railway safety. Finally, the project probes into the future improvements that could be made in the models by investigating other data types and fine-tuning deep learning models to enhance detection accuracy and reliability, hence their application in wildlife conservation more widely.

III. Related Works

Gunasekara, Jayasuriya et al. (2021) present a CNN-based early warning system to prevent elephant-train collisions, one of the critical issues within the HEC in Sri Lanka. The authors of the study have identified the severe impacts related to the collision, which even results in damage to the endangered elephant species, causing huge economic losses due to disrupted railway services. The authors propose the use of advances in CNN technology to realize a vision-based detection system with an RGB/infrared-capable camera deployed across known hotspots along railway tracks. It is designed to be a low-cost, rugged, self-sustaining system powered by solar panels and batteries, leveraging the YOLO V3 CNN model that is trained on a curated dataset of elephant images collected from several wildlife sites. Preliminary results indicate that the system is indeed able to detect elephants under a variety of lighting conditions, which may enable real-time monitoring and provide early warnings to train drivers. The authors argue that such a detection network could reduce elephant-train collisions and also provide more general surveillance for HEC management, but would need comprehensive training datasets in order to improve the accuracy of object detection in poor visual conditions. [1].

Gupta, Mohan et al. (2021) introduce an advanced surveillance system that employs deep vision technology, designed specifically to mitigate train-elephant collisions, which constitute a major problem in areas where railway lines cross through wildlife habitats, especially in India. The research highlights the necessity for technological solutions in response to the rising incidence of elephant fatalities resulting from these train-related incidents. They propose a lightweight CNN architecture designed specifically for elephant detection near railway tracks. They also propose three transfer learning architectures: ResNet50, MobileNet, and Inception V3. The models are trained on the carefully prepared dataset comprising ELPephant and RailSem19 with the purpose of effective alert systems designed to alert trains on potential collisions. These proposed models showed a great deal of accuracy, with the highest performance recorded at 99.91% from the Inception V3 model. The system is designed to issue real-time alerts to the train operators and audio notifications that will discourage the elephants, hence contributing toward the safety of both the wildlife and the railway itself. This research highlights the application of high-accuracy deep learning models and the concept of transfer learning in enhancing the process of detection and averting human-elephant conflicts on railways. [2].

Parihar, Ghosh et al. (2022) present a novel method for the detection and classification of elephant locomotion using seismic signals in their work "Variational Mode Decomposition of Seismic Signals for Detection of Moving Elephants." The paper addresses one of the main concerns related to human-elephant conflict, particularly near railway tracks, where elephant incidents have become increasingly troublesome. The researchers propose a hybrid approach based on the VMD method for seismic signal processing and multidomain feature extraction, which will later be classified with the support of SVM. This methodology will be compared with other classical signal decomposition methods, including EMD and EWT. The study shows that VMD brings significant improvements in both classification accuracy and detection, thus making it a potentially effective method for developing automated early warning systems with the aim of preventing elephant accidents from occurring because of trains. [3].

In this paper titled "Survey on Smart Siren to Alert Animals About Train Tracks in Image Processing," **Alfored George**

Charles Prem, Rajakumar Singarayan, et al. (2024) discuss a scheme to mitigate incidents of animal-train collisions. The manuscript emphasizes how frequently animals and trains come into conflict with each other, proposing the application of an intelligent siren system that invokes image processing for detecting animals near railway tracks. The system utilizes cameras along the tracks to take pictures and computer vision algorithms to check for the presence of animals within those images. Upon detecting an animal, a siren is activated in order to alert and dissuade it from entering the tracks. The authors argue that this system can significantly reduce animal deaths and can offer humankind a way of peaceful coexistence with wildlife. The manuscript also discusses relevant literature and methodologies regarding animal detection, offering a comparison between current solutions and introducing smart siren as a new and effective approach for reducing animal-train collisions. [4].

Schneider, Taylor et al. (2019), in their work ”Past, Present, and Future Approaches Using Computer Vision for Animal Re-identification from Camera Trap Data,” comprehensively review historical, present, and future applications of computer vision. The study highlights that traditional methods of animal re-identification, such as tagging and DNA analysis, are usually very laborious and invasive, while it emphasizes camera traps as an invaluable tool for ecological data collection that is non-invasive. The authors first discuss the chronological development of feature engineering methods in computer vision with regard to animal re-identification and point out that all the early methods involved the use of special algorithms for extracting features. Deep learning and CNN introduced a significant improvement in the accuracy of re-identification, both for humans and animals, overcoming previous issues such as human bias and reliance on manual feature engineering. Schneider and Taylor et al. predict that the integration of deep learning methods will revolutionize camera trap data analysis, improving the accuracy of population estimates and ecological studies. [5].

Mahmud, Kabir et al. (2023) propose a new wildlife conservation method by designing an automated system for the identification of railway tracks and the detection of foreign objects using the Mask R-CNN algorithm in their manuscript entitled ”Advancing Wildlife Protection: Mask R-CNN for Rail Track Identification and Unwanted Object Detection.”. The authors address the important problem of collisions by trains with wildlife as well as other objects left on the railway tracks, a problem that causes significant ecological and operational impacts. In their approach, a deep learning-based framework is proposed for identifying train tracks and for recognizing extraneous objects specifically situated within the boundaries of rail lines. This work also highlights some of the advantages of using Mask R-CNN for accurate instance segmentation, which ensures high detection accuracy with low false positive rates. Excellent performance was achieved with a mean average precision of 0.9375 and a frame rate of 30 frames per second. The present study therefore proposes a feasible method to help reduce train-wildlife collisions and improve rail safety using advanced computer vision techniques. [6].

Mondal, Mandal et al. (2023) present a fog-assisted framework for monitoring elephants near railway tracks to prevent train-elephant collisions, a significant issue in India where railway networks intersect with elephant habitats. The study emphasizes the drawbacks of cloud computing, such as high latency and network usage, which are critical in latency-sensitive applications like real-time monitoring and collision prevention. In this regard, the authors have proposed a fog-based system that uses fog nodes to process data closer to the source for diminishing latency, network utilization, and hence execution time from cloud-based systems. The framework involves a multi-layer architecture built on top of intelligent cameras and sensors at the edge, fog nodes for processing, and a cloud for storing data and further analytics. The YOLOv5 deep learning architecture is employed for the purpose of detecting elephants, achieving a notable level of accuracy while maintaining low latency. The system under consideration is simulated utilizing the iFogSim toolkit, revealing substantial enhancements in response time, network load, and operational expenses relative to conventional cloud computing methodologies. The study concluded that the use of fog computing for real-time wildlife monitoring and collision prevention is more efficient and effective, hence widening its applications in IoT-based smart systems. [7].

Bhavani, Aathish et al. 2023, proposed a Lora-Based Elephant Detection System Near Railroads to reduce the rate of human-elephant collision near railway tracks. The increasing prevalence of train accidents due to crossing by elephants in railway tracks poses a significant threat to wildlife and human life. Authors propose a system incorporating image processing with IoT, using Lora-Long Range communication for instant detection and alert. Major components of this system involve Node MCU, microcontroller unit, GPS, LCD displays, and speakers, facilitating long-range communications related to elephant sightseeing. The proposed solution has the capability of detecting the presence of elephants near railway tracks and can send alerts to authorities and train operators quickly. This system would lead to much better safety for both humans and elephants. With a detection accuracy of 89%, the system demonstrates its ability in mitigating the risks associated with train-elephant collisions, efficiently, scalably, and economically viable. [8].

Kellenberger, Volpi et al. (2017) propose a method for rapid animal identification in unmanned aerial vehicle imagery by utilizing CNNs in the work Fast Animal Detection in UAV Images Using Convolutional Neural Networks. This work takes into consideration the increasing interest and/or concern in regard to unlawful wildlife poaching and the need to have effective real-time systems that can monitor populations of animals. In a nutshell, the authors fly UAVs over areas to capture images at sub-decimeter resolution and apply a CNN-based model for object detection specific to large animal detection in such images. Their proposed approach illustrates significant improvements in both precision and speed compared with the state-of-the-art Fast R-CNN model. That works out to 72 images per second, compared to only three images per second achieved by the Fast R-CNN baseline, hence a suitable model for real-time monitoring applications. Their system, though, achieved a higher precision through reduced false positives, while in some cases it faced challenges in terms of recall. As indicated, this paper highlights the potential of CNNs to enhance wildlife monitoring and conservation efforts with remote sensing technologies. [9].

Meena Prakash, Vimala, et al. (2023) proposed a methodology for obstacle detection on railway paths using machine learning to avoid all types of railway accidents by utilizing deep learning methods for real-time detection of obstacles in rail infrastructures. Their work focuses on the frequent incidents of collision by trains with large fauna such as elephants and buffalo, causing devastating effects on both fauna populations and railway infrastructure. The system under consideration uses the ESP32 camera to capture the railway tracks imagery, which further gets processed by the pre-trained COCO dataset on the YOLOv3 CNN for obstacle detection and classification. An ultrasonic sensor is also included in the system to measure the distance between the train and the detected obstacles. The data are then sent to the operators of the trains via a messaging alert system. The results of the experiment demonstrate that the system has extremely high accuracy in terms of detection. YOLOv3 outperforms other convolutional neural network architectures like Faster R-CNN and RetinaNet on accuracy and confidence score metrics. This study contributes significantly to the improvement of railway safety and wildlife protection from accidents related to trains. [10].

IV. Methodology

The deep learning-based detection system for real-time elephant detection near railway tracks is proposed. The methodology is divided into four stages: data gathering, preprocessing, and augmentation; model design and training; and performance evaluation. Within this research, the comparison of two implemented deep learning models, CNN and R-CNN, will be carried out on grounds of their accuracy, speed of detection, and computational efficiency.

1) **Data Collection**

   In the project, a comprehensive dataset was built on various sources to ensure diversity in the images around which deep learning models were trained for strength. The dataset's development took place in three broad phases:

   a) Kaggle Elephant Image Dataset- This is the first dataset of elephants collected from Kaggle with images of labelled elephant occurrences in various environments. These images are used while training the basic CNN and R-CNN models.
   b) The Railway Track Dataset, sourced from the Roboflow Universe, is a secondary collection of images containing railway tracks. It constitutes illustrative railway track scenes; these scenes are critical in putting the detection of elephants in context against the background of these tracks.
   c) Web scraping of elephants on railway tracks - To add richness to the dataset and to check if the models were ready for the specific case of elephants on railway tracks, several images have been scraped from Google using Fatkun Batch Download Image extension. Images have then been filtered manually to see whether the images represent true cases where there exist elephants near railway tracks. That enrichment was done for practical suitability

checks, wherein elephants do come in a railway context.

The final dataset consists of images of elephants, railway tracks and elephants on railway tracks, under varied environmental conditions, such as different lighting, angles, and distances. The images were annotated using bounding boxes to train object detection models in both CNN and R-CNN approaches. This heterogeneous dataset will mitigate a wide range of challenges within the railway context, especially those for variations in pose, changes in lighting conditions, and differing distances from the camera, commonly encountered when implementing this detection system.

2) **Data Pre-processing and Augmentation**
These images before training by models, are subject to several pre-processing tasks:
   a) Resizing: All images are resized to a fixed input size (e.g., 224x224 or 256x256 pixels) to ensure consistent input dimensions for the CNN and R-CNN models.
   b) Normalization : Normalized pixel to the range [0, 1] by dividing each pixel value by 255.

$$\text{Normalized Pixel Value} = \frac{\text{Pixel Value}}{255}$$

   c) Augmentation: Various data augmentations involved the use of techniques that utilize random rotations, flipping, zooming, and shifting to enrich the diversity of the dataset; hence, models would be more robust in real-world scenarios. Applied transformations in sequence include the following:
      i) Rotation: Random rotation is performed over a range of ±30 degrees.
      ii) Zoom: Scaling by a factor of 0.8 to 1.2.
      iii) Horizontal and vertical flipping: It creates duplicates of different-viewed elephants.
   d) Generating the bounding boxes (for R-CNN): For R-CNN, bounding box coordinates for each elephant in the images are obtained manually or by other means of annotations. The bounding boxes are very important for training the R-CNN model to localize the elephants.

3) **CNN Model Design**
A simple architecture of a CNN network would be taken as a baseline model for the detection process of elephants. The components involved are
   a) Convolutional Layers: These layers implement detect features associated with edges, textures, and shapes on input images. Among others, the number of filters and filter size can be selected based on the complexity of the dataset. Convolution operation for a specific input image

$$(I * K)(x, y) = \sum_m \sum_n I(m, n) \cdot K(x - m, y - n)$$

   where denotes the convolution operation, and m,n are the indices of the filter and image.
   b) Activation Function: After every convolutional process, the usage of Rectified Linear Unit (ReLU) activation function incorporates nonlinearity in every iteration.

$$ReLU(x) = max(0, x)$$

   c) Pooling layers: After some of the convolutional layers, max pooling layers are used to reduce the spatial dimensions while retaining the most prominent features with commonly used 2x2 windows and a stride of 2.
   d) Fully Connected Layers: Following feature extraction using operations in convolutions and pooling, the output from these operations is flattened and then passed through one or more fully connected layers that produce final class probabilities. Output Layer The final output layer consists of a single neuron with either a sigmoid or softmax activation function to classify whether or not there are elephants in the image.

4) **R-CNN Model Design**
Region-based CNN (R-CNN) provides for selective search, which detects regions in an image where objects may exist. It then applies a CNN to classify and refine the bounding boxes of these regions. The R-CNN approach can be described in the following steps:
   a) Selective Search: This algorithm describes an ag-

gregation of candidate regions (proposals) in an image where objects (such as elephants) tend to be. Selective search unifies segmentation with region merging to create region proposals. The algorithm behind selective search can be encapsulated in the following formula:

$$RegionProposals = SelectiveSearch(I)$$

b) Region Cropping: Apply region proposal from an image cropping to the input size of a CNN.
c) Feature Extraction: The cropped regions are passed through a CNN to extract features as elucidated in the earlier section explaining the CNN architecture. A CNN, in the case of R-CNN, is used as a feature extractor.
d) Bounding Box Regression: To refine the predicted bounding boxes R-CNN implements a bounding box regression model that best minimizes the localization loss:

$$L_{\text{bbox}} = \sum_i (\hat{x}_i - x_i)^2$$

where $\hat{x}_i$ is the predicted bounding box and x_i is the ground truth bounding box coordinates.
e) Classification: The regions are classified after feature extraction to determine either an elephant exists there or not. The classification is done by using a softmax or sigmoid function, applying to it whether it is multi-class or binary classification.

5) **Model Training**
Both CNN and R-CNN models employ SGD with back-propagation as the training procedure. The loss function is defined by the CNN as binary cross-entropy loss:

$$L_{\text{CNN}} = -\frac{1}{N} \sum_{i=1}^{N} \left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right)$$

where N is the number of samples, y_i is the true label, and $\hat{y}_i$ is the predicted probability.
The loss for the R-CNN is the summation of classification loss and bounding box regression loss:

$$L_{\text{R-CNN}} = L_{\text{cls}} + L_{\text{bbox}}$$

where L_{cls} is the classification loss (similar to CNN) and L_{bbox} is the bounding box regression loss

6) Performance Evaluation The performance of both models is evaluated using the following metrics:
Detection Accuracy: Measures the percentage of correct predictions out of the total predictions made. This is computed as:

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

7) Comparison between CNN and R-CNN
This section analyzes the performance characteristics of the CNN and R-CNN models, with regard to detection accuracy, detection speed, and computational efficiency. By referencing this analysis, critical differences between the models will be determined, which highlight which model will be more suitable for real-time detection of elephants near rail tracks. Accuracy in Detection Detection accuracy is one of the most crucial metrics, showing the ability of a model to correctly distinguish elephants from other objects in an image. The predictive accuracy of these models was compared. R-CNN depicts greater accuracy compared with CNN in terms of detection. With an accuracy level of 94.98%, it stands to be contrasted with CNN's accuracy level of 90%. Thus, R-CNN becomes the more reliable detection model, specially on complicated scenarios. However, CNN shows faster detection speed along with computational efficiency. Therefore, it becomes ideal for situations wherein fast performance is required and computational resources are at a minimum. The deployment environment's special requirements regarding the trade-off between the two models should, therefore be put into perspective while seeking a balance toward accuracy, speed, and computational resources.

V. Results

In the evaluation of the proposed elephant detection system, the two models CNN and R-CNN were ranked against each other based on detection accuracy, speed, and computational efficiency. The following table illustrates the results of the accuracy returned by each model:

Model	Accuracy
CNN	0.9000
R-CNN	0.9498

Fig. 1. Model Comparison

The R-CNN model showed higher detection accuracy at 94.98%, which was higher than that of the CNN model at 90.00%. This indeed was close to what was theoretically expected because R-CNN models are usually preferred more for complex object detection when localization matters most. By focusing on the elephants particularly located on railway tracks, the region-based approach of the R-CNN allowed for more precise predictions of bounding box areas covering elephants in various poses and conditions. Furthermore, although CNN had superior speed in the processing of images because of its less complex structure, R-CNN's heightened accuracy emphasized its suitability for critical applications where higher exactness counts than efficiency. The ruggedness of R-CNN, particularly in various light and background conditions, further emphasizes applicability in real-world railway safety systems, which rely on accurate warnings.

VI. Conclusion

The article describes how the models of deep learning, specifically CNN and R-CNN, were utilised in the detection of elephants on railway tracks to prevent wildlife collision. The findings are that even though these models were effective in identifying elephants, R-CNN was more accurate than CNN; hence it is more reliable for scenarios that require the correct localization. Higher accuracy from the current model supports its feasibility for use in real-time railway safety systems that demand detection reliability highly, especially in areas that have high elephant movement across railways. Such an implementation could potentially reduce the numbers of railway accidents caused by elephant attacks, promoting both wildlife conservation and railway safety. Further work could focus on how to further improve this model with real-time optimization techniques, additional environmental data, and fine-tuned increments that have response time without degrading the detection accuracy.

References

[1] Gunasekara, S., Jayasuriya, M., Harischandra, N., Samaranayake, L., & Dissanayake, G. (2021, December). A Convolutional Neural Network Based Early Warning System to Prevent Elephant-Train Collisions. In 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS) (pp. 271-276). IEEE.

[2] Gupta, S., Mohan, N., Nayak, P., Nagaraju, K. C., & Karanam, M. (2022). Deep vision-based surveillance system to prevent train–elephant collisions. Soft Computing, 1-14.

[3] Parihar, D. S., Ghosh, R., Akula, A., Kumar, S., & Sardana, H. K. (2022). Variational mode decomposition of seismic signals for detection of moving elephants. IEEE Transactions on Instrumentation and Measurement, 71, 1-8.

[4] Prem, A. G. C., Singarayan, R., Rajarethnam, S., & Mohan, N. (2023, October). Survey On Smart Siren to Alert Animals About Train Tracks in Image Processing. In 6th International Conference on Intelligent Computing (ICIC-6 2023) (pp. 114-118). Atlantis Press.

[5] Schneider, S., Taylor, G. W., Linquist, S., & Kremer, S. C. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. Methods in Ecology and Evolution, 10(4), 461-470.

[6] Mahmud, I., Kabir, M. M., Shin, J., Mistry, C., Tomioka, Y., & Mridha, M. F. (2023). Advancing Wildlife Protection: Mask R-CNN for Rail Track Identification and Unwanted Object Detection. IEEE Access.

[7] Mondal, M. K., Mandal, R., Banerjee, S., Biswas, U., Lin, J. C. W., Alfarraj, O., & Tolba, A. (2023). Design and development of a fog-assisted elephant corridor over a railway track. Sustainability, 15(7), 5944.

[8] Bhavani, S., Aathish, D., Iswarya, S., Kathiravan, A., & Kavin, P. (2023, February). LoRa Based Elephant Detection System near Railroads. In IEEE Proc.

[9] Kellenberger, B., Volpi, M., & Tuia, D. (2017, July). Fast animal detection in UAV images using convolutional neural networks. In 2017 IEEE international geoscience and remote sensing symposium (IGARSS) (pp. 866-869). IEEE.

[10] Prakash, R. M., Vimala, M., Keerthana, S., Kokila, P., & Sneha, S. (2023, May). Machine Learning based Obstacle Detection for Avoiding Accidents on Railway Tracks. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 236-241). IEEE.