Enhancing Railway Safety through Human Activity Recognition

Kushaal S
Department of AIML
RV College of Engineering
Bengaluru, INDIA
kushaal.sathish@gmail.com

Shiva Kumar
Department of AIML
RV College of Engineering
Bengaluru, INDIA
shivakumar.vahani@gmail.com

***Abstract*—Intelligent Transportation System (ITS) is one of the evolving domains where Artificial Intelligence (AI) is creating endless opportunities. In recent days, for transit both passengers and freight prefer railway as the transportation medium. For this reason, the railway should provide the safety and uninterrupted services. These safety measures reduce the fatalities. In this paper, we present an Object Detection (OD) and Activity Recognition (AR) model for railway lines with the goal of reducing accidents, fatalities and improving safety. The proposed work uses the advantage of You Only Look Once Version 8 (YOLO-V8) to detect the objects and recognize the activities. This paper provides a detailed performance analysis of YOLO-V8 on training and validation sets. Also, this paper highlights the other performance indices such as classification accuracy and mean average precision with respect to the epoch number.**

Keywords—YOLO-V8, Object Detection, Activity Recognition, Railway

I. INTRODUCTION

Railway safety is a critical concern in India, where the railway network is among the largest and busiest in the world. The Indian Railways operates over 67,000 kilometres carrying more than 8 billion passengers and over 1.2 billion tons of freight annually [1]. Railway accidents, particularly those involving unauthorized human presence on tracks, profoundly impact transportation. The loss of lives in such incidents is often tragic and widespread, affecting not only the immediate families of the victims but also the larger community. Despite various safety measures, unauthorised human presence on railway tracks remains a significant risk, contributing to numerous accidents and fatalities each year. The National Crime Records Bureau (NCRB) reported that there were an estimated 26,000 fatalities in India in 2021 because of unauthorized entry onto railway lines [2], underscoring the urgent need for improved safety mechanisms. Accidents often prompt government action and policy changes. These responses, while necessary, can sometimes lead to rushed or poorly planned measures that may not effectively address the root causes of accidents. Traditional methods for ensuring track safety, such as manual surveillance, physical barriers, and warning signs, often prove inadequate due to their limited coverage, high operational costs, and inability to provide real-time alerts [3]. These methods struggle to address the complexity and scale of India's extensive railway network, where timely intervention is crucial to prevent accidents. As a result, there is a growing need for automated systems that can continuously monitor railway tracks and detect human presence accurately and swiftly.

Advancements in computer vision and deep learning have made it possible to develop automated systems capable of enhancing railway safety. Object detection algorithms, particularly those based on deep learning, have demonstrated significant potential in various applications, including surveillance, autonomous vehicles, and public safety [4]. The YOLO algorithm, known for its efficiency and real-time detection capabilities, has emerged as a leading choice for tasks requiring rapid and accurate identification of objects [5].

This paper focuses on developing a robust system for detecting human intervention on railway tracks using the YOLO algorithm. The primary goal is to create a solution that can operate effectively in diverse environmental conditions, including varying lighting, weather, and occlusion scenarios. By leveraging the strengths of the YOLO architecture, our paper aims to provide accurate and timely alerts, thereby preventing potential accidents and improving overall railway safety. Our contributions to this field include the creation of a custom dataset for training and testing the model, tailored modifications to the YOLO architecture to enhance detection accuracy, and a comprehensive evaluation of the system's performance in real-world conditions [6]. Additionally, we present an implementation strategy for deploying this system in practical settings, including considerations for real-time processing and alert mechanisms. In the following sections, we will review the current state of railway safety measures and object detection technologies, describe our methodology for developing the human detection system, present our experimental results, and discuss the practical implications and future directions of our work.

II. LITERATURE REVIEW

An extensive literature review is carried out to gain a broad understanding of the latest advancements and research in railway safety and object detection.

Railway safety heavily relies on traditional methods such as manual surveillance, physical barriers, and CCTV cameras. Manual surveillance involves personnel monitoring tracks and crossings, which is effective to some extent but is labour-intensive, prone to human error, and not feasible for continuous monitoring across vast railway networks [3]. While effective in preventing unauthorized access, physical

barriers are costly, require significant maintenance, and cannot be implemented extensively, especially in rural and semi-urban regions [7]. CCTV surveillance provides visual records but demands constant human monitoring, limiting its effectiveness in real-time threat detection [8]. Sensor-based systems, employing motion detectors, infrared sensors, and pressure sensors, can trigger alarms but are often prone to false positives and may not function effectively under adverse weather conditions [9]. Advancements in object detection technologies have introduced more sophisticated approaches, such as background subtraction and 3D localization. Zhang et al. present an Automatic Video Surveillance (AVS) system for railway level crossings using passive stereo vision [6]. This technique uses a selective stereo-matching algorithm for 3D localization and Independent Component Analysis (ICA) to handle noise and distinguish motion from the background. Similar to this, Lee et al. highlighted how conventional motion detection techniques, which look for changes between consecutive frames, struggle with noise and varying environmental conditions [10]. The advent of Convolutional Neural Networks (CNNs) has revolutionized object detection, making significant strides in accuracy and efficiency. R-CNN (Region-based Convolutional Neural Network), introduced by Girshick et al., marked a significant advancement by incorporating region proposals [11]. It uses selective search to generate region proposals and applies a CNN to extract features from each region by incorporating region proposals, Girshick et al.'s introduction of R-CNN (Region-based Convolutional Neural Network) marked a significant advancement to information loss, highlighting the need for more efficient approaches. Fast R-CNN, proposed by Girshick builds on the strengths of R-CNN by eliminating the cropping and warping step, and using Region of Interest (ROI) pooling to enhance training efficiency. It replaces SVM with a SoftmaxL1Loss layer for classification and uses Smooth Loss for bounding-box regression. The integration of classification and regression within a single network improves precision and accelerates training and testing speeds [11]. Faster R-CNN, introduced by Ren et al., incorporates a Region Proposal Network (RPN) that shares the convolutional layers with the detection network to further improve item detection performance [12]. This integration significantly reduces computational overhead, making the detection process more efficient and faster. YOLO (You Only Look Once) algorithms present a paradigm shift in object detection by treating the task as a single regression problem. YOLOv1, introduced by Redmon et al. consider object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. Despite its speed, YOLOv1 struggles with small objects and complex backgrounds. YOLOv2, improves upon YOLOv1 by maintaining high speed while increasing mean Average Precision (mAP). It introduces multi-scale training, allowing the model to operate at various input sizes, improving robustness and adaptability [4]. YOLOv3, introduced by Redmon and Farhadi brings more architectural improvements, such as using a deeper network with residual blocks and multi-scale predictions, enhancing its capability to detect objects at different scales. YOLOv3 offers significant improvements in accuracy, particularly for small objects, without compromising on speed. YOLOv4, developed by Bochkovskiy et al. incorporates numerous enhancements in architecture and training techniques, focusing on improving speed while achieving high accuracy. YOLOv4 uses CSPDarknet53 as the backbone, PANet for path aggregation, and a modified SPP block for better receptive fields, achieving state-of-the-art performance on the COCO dataset with substantial improvements in both accuracy and speed [5].

In precise, the detailed literature survey thoroughly explains the evolution and advancements in object detection technologies, particularly within the CNN family and YOLO algorithms. The progression from R-CNN to Faster R-CNN illustrates efforts to enhance efficiency and accuracy by integrating region proposal generation with object detection [11]. With their novel method of treating object identification as a single regression problem, YOLO algorithms have broken previous records for speed and real-time effectiveness [4]. By leveraging these advancements, our research aims to develop a robust system for detecting human presence on railway tracks, utilizing the strengths of YOLO for real-time, accurate monitoring and alerting. This system has the potential to significantly improve railway safety by providing timely and precise detection of unauthorized human presence, thereby preventing accidents and fatalities.

III. Proposed Work

The proposed work comprises Data Collection, Data Preprocessing, Model Selection and Model Evaluation.

A. Data Collection

In this work, a comprehensive data collection strategy was adopted to ensure the robustness and accuracy of the detection system. In this work, high-resolution pictures of diverse train situations are taken using a Samsung 50 MP GN5 sensor with an f/1.88 aperture. The images were captured with a median resolution of 1080 x 1080 pixels, striking a balance between image quality and computational efficiency. The primary focus was on collecting a diverse set of around 1,000 original images with a person as the primary subject on the tracks. The dataset comprises of three different classes such as sitting (325 images), lying (350 images) and standing (325 images) on the track. These images were

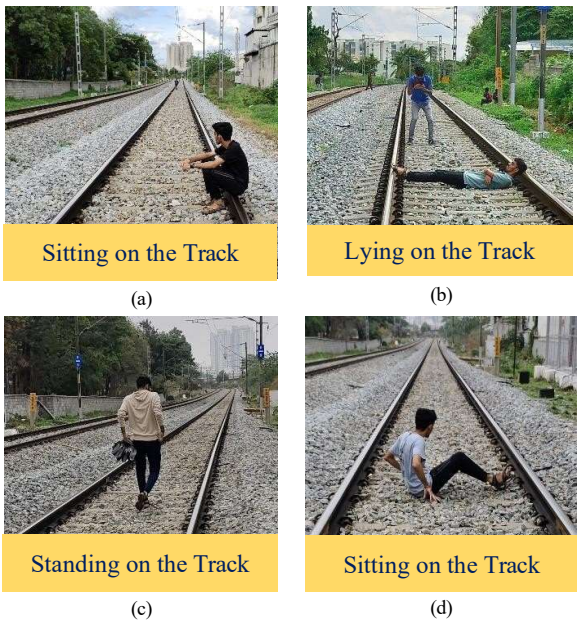

Fig.1 Different Classes used in the proposed work.

meticulously captured under different conditions to simulate real-world variability. To achieve this, data was systematically gathered across various lighting conditions, including daytime, nighttime, dawn, and dusk, ensuring that the model could handle the varied times of the day. Recognizing the impact of weather on visibility and image quality, the data collection also included different weather conditions, such as clear sky, rain and fog were considered to encompass a wide range of environmental factors. To further enhance the dataset's diversity, images included scenarios with partial occlusions where the subject was partially obscured by other objects, such as railway infrastructure or vegetation. Figure 1 depicts the few examples of the data collected for the model training.

B. Data Pre-Processing

Recognizing the importance of having a robust dataset, Roboflow tool is used for annotating the images and labelling done manually. The tool enables us to draw precise bounding box annotations around humans and other relevant objects on the tracks. The dataset comprises of multiple classes such as standing, sitting, sleeping/ lying and track. Annotations were made with precise bounding boxes around the subject and relevant objects. This precision was crucial for training an accurate detection model. Each annotation was carefully reviewed to ensure correct class labeling. Misclassifications were minimized through multiple rounds of review. For images with partial occlusions, the bounding boxes were drawn around the visible parts of the subject. To increase the dataset size and variability, several data augmentation techniques were employed such as horizontal flipping, 90-degree rotations, converting 20% of the images to grayscale, adjusting brightness by ±20%, modifying exposure by ±15%, and rotating images by ±15 degrees. These augmentations expanded the dataset to over 3700 images, providing a rich and diverse training set. This extensive augmentation ensured that the model could learn from a wide array of conditions and scenarios, significantly improving its robustness and accuracy. Considering all the annotations of images into classes, the total annotations present were 1819, 1782, 1406 and 689 for track, standing, sitting, and sleeping respectively. The data was split into 70-20-10, that is, 2635 images, 754 images, and 377 images respectively for training, validation, and testing, respectively.

C. Model Selection

YOLO-V8, a state-of-the-art object detection architecture, was chosen for its efficiency and accuracy. The model architecture allows to process high-resolution images quickly, making it suitable for real-time applications.

The YOLO algorithm has seen numerous iterations, with YOLO-V8 being one of the most advanced versions, offering significant improvements in accuracy and efficiency. For this project, YOLO-V8 was employed due to its superior performance in detecting objects in real time, making it ideal for the critical task of monitoring railway tracks for human presence.

YOLO-V8 introduces several enhancements over previous versions. It begins with an input layer designed to accept images resized to a standardized resolution, typically 640x640 pixels, which ensures consistency and optimal performance. The backbone of YOLO-V8, an evolution from previous DarkNet architectures, utilizes advanced convolutional layers and incorporates innovations like CSP (Cross-Stage Partial) connections and a focus module. These features allow for more efficient feature extraction and better gradient flow, enhancing the model's ability to detect delicate details. Provided the detailed architecture of YOLO-V8.

D. Model Training

Google Colab, a cloud-based platform that offers free access to potent computing resources, was used to train the YOLOv8 model for detecting human interference on railway lines [15]. Google Colab uses a T4 GPU, a service for researchers and developers. It's a powerful inference accelerator with NVIDIA Turing Tensor Cores that can perform multi-precision inference to speed up AI applications [16]. Deep learning applications, image processing, and mid-range machine learning jobs are all excellent fits for the T4 GPU. It features 2,560 CUDA cores and 16 GB of GDDR6 memory [16]. The use of Colab was instrumental in efficiently handling the computational demands of training a deep learning model, particularly on large datasets. The model training pipeline involved several key steps, discussed in the upcoming paragraphs. Data Pre-processing and Augmentation of the dataset, comprising annotated images of railway tracks with humans and other relevant objects, was first pre-processed. This included normalizing pixel values, scaling images to a standard resolution, and translating annotations to the YOLO-V8 required format. [17]. Random cropping, rotation, flipping, and color modifications were among the data augmentation techniques used to increase the diversity of the training set and strengthen the model's resistance to changing circumstances [18].

E. Model Evaluation

Model evaluation is performed using the following metrics:

- **Mean Average Precision (mAP)** averages precision values at different recall levels to assess the model's performance across classes. It displays how well the model can detect items while reducing false positives and false negatives. [5].
- **Precision** is the model's capacity to limit false positives and accurately forecast outcomes and is demonstrated by the ratio of true positive detections to the total number of positive detections. This ratio is essential for avoiding needless warnings in railway monitoring systems [13].
- **Recall** highlights the model's capacity to identify all occurrences of human intervention on the tracks, hence improving system safety. It does this by calculating the ratio of genuine positive detections to the total number of actual positive cases.
- **F1 Score** is a balanced statistic that is particularly helpful in situations, when both false positives and negatives have significant consequences, is the harmonic mean of precision and recall [14].

Together, these metrics offer a detailed evaluation of the YOLO-V8 model's performance, highlighting its strengths and areas for improvement. This comprehensive assessment is crucial for optimizing the model, ensuring high operational

efficiency, and maintaining safety in real-time applications [5].

IV. Results and Discussion

The experimental procedure includes the following steps: organizing the experiment, preparing the code, training the model, evaluating and testing it, and comparing the findings.

A. Experimental Setup

The training environment utilized Google Colab, leveraging its T4 GPU for training. In this work, training set accounted for 70% of the dataset, validation set for 20%, and testing set for 10% with 2,635, 754, and 377 images, respectively. YOLO-V8 models were used for training and produced the best accuracy. The parameters used in the training process for the model are shown in Table I.

TABLE I. Model Training Parameters

Batch Size	16
Epoch	50
Learning Rate	0.00125
Optimizer	AdamW

B. Experimental Results

Two important metrics, mAP50 and mAP50-95 were used to assess the performance of the model.

Figure 2(a) shows the mAP at a 50% Intersection over Union (IoU) threshold, known as mAP50 and mAP averaged over multiple IoU thresholds from 50% to 95%, which is a more stringent and comprehensive measure of model performance.

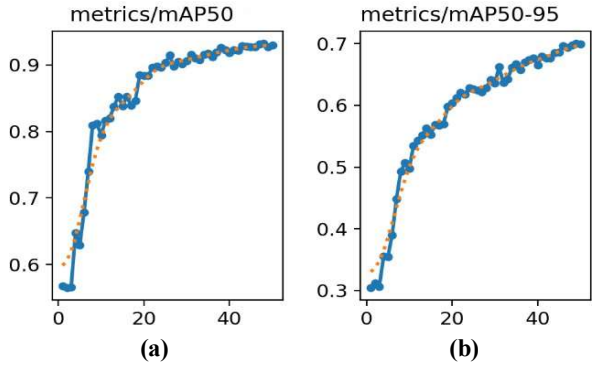

Fig. 2 Mean Average Precision a) mAP 50 and b) mAP50-95

The mAP50 starts around 0.6 and quickly increases as the epochs progress, reaching around 0.93 after 40 epochs. The curve flattens towards the end, indicating that the model is converging and improving less significantly with additional epochs. This shows that the model is highly accurate in detecting and classifying objects with a moderate overlap. This infers the model is performing well in identifying whether a person is standing, sitting, or lying down on the track, as well as distinguishing the track itself.

The mAP50-95 starts around 0.3 and increases steadily, reaching approximately 0.7 after 40 epochs. Like the Fig.2(a), the curve shows signs of flattening, indicating convergence. The lower mAP50-95 (around 0.7) indicates that while the model is effective, it struggles more with precise localization, particularly when distinguishing between classes that may have similar appearances or overlap**.**

The precision starts at around 0.6 and increases rapidly within the first 10 epochs, reaching a plateau around 0.9 as shown in Figure 3(a). The precision remains relatively stable with minor fluctuations after that. The recall starts at around 0.55, dips slightly, and then increases steadily, reaching around 0.85 after 40 epochs as shown in Figure 3(b). The initial dip suggests early challenges in detecting all objects, but these were overcome as the model continued to train. The YOLO-V8 model performance is given the Table II

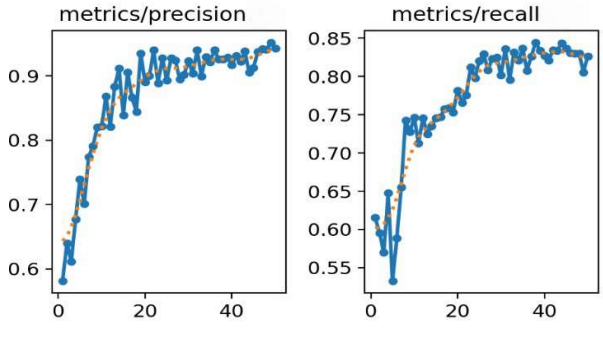

Fig.3 Performance Metrics a) Precision and b) Recall

TABLE II. Model Performance Metrics

Model	YOLO-V8
mAP50	92.3
mAP50-95	69.5
Recall	81.1
Precision	94.4

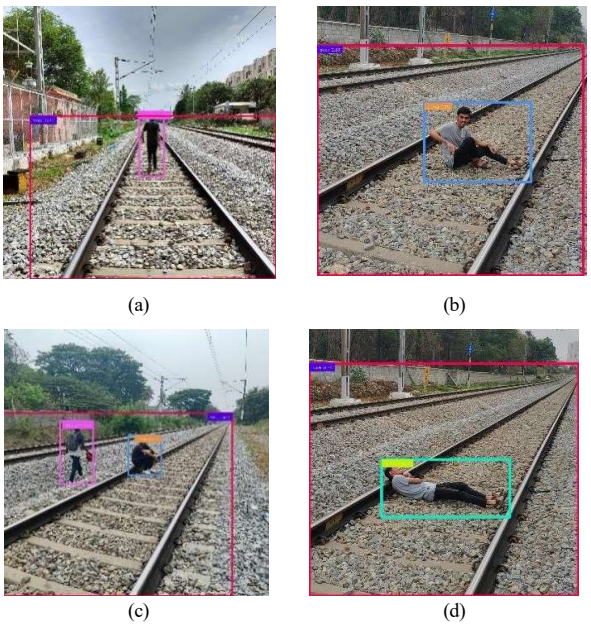

Fig. 4 Various Activities Recognized by YOLO-V8 of Rail Dataset

A few images were randomly selected for testing to assess the predictive capabilities of the trained models. Figure 4 shows the predicted result of YOLO-V8.

V. Conclusion

Railway accidents are more common in recent days and the use of AI technologies helps in preventing it. In this paper, we have created a unique dataset. The dataset is used to identify the different human activities through AI which enhances the railway safety and seamless commuting. The YOLO-V8 model was selected as the optimal choice for this work and shown in-depth analysis. The proposed method achieved 93.00% accuracy. The proposed system can be extended with the sharing of the autonomous notifications with the loco-pilots and nearest railway stations. Additionally, the dataset is now limited to a small number of activities, and it would be beneficial to include a greater variety of activities.

Acknowledgement

The authors would like to thank Dr K N Subramanya, Principal, R V College of Engineering, Bengaluru and Dr B Sathish Babu, Professor and Head, Department of Artificial Intelligence and Machine Learning, R V College of Engineering, Bengaluru for their timely help and constant support to complete this work.

References

[1] Indian Railways, 2021. Indian Railways Statistical Summary. [Online] Available: indianrailways.gov.in/railwayboard/uploads/directorate/stat_econ/pdf/Indian Railways Annual Report %26 Accounts English 2021-22_web_Final.pdf.

[2] National Crime Records Bureau (NCRB), 2021. Accident Deaths & Suicides in India 2021. [Online] Available:https://www.drishtiias.com/pdf/1661934989.pdf.

[3] Cao, Zhiwei and Qin, Yong and Jia, Limin and Xie, Zhengyu and Gao, Yang and Wang, Yaguan and Li, Ping and Yu, Zujun, "Railway Intrusion Detection Based on Machine Vision: A Survey, Challenges, and Perspectives",2024, IEEE Transactions on Intelligent Transportation Systems, vol 25, DOI: 0.1109/TITS.2024.3412170

[4] Joseph Redmon; Santosh Divvala; Ross Girshick; Ali Farhadi,"You Only Look Once: Unified, Real-Time Object Detection." ,2016, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, DOI: 10.1109/CVPR.2016.91.

[5] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection.", 2020, arXiv:2004.10934. [Online] Available: https://arxiv.org/abs/2004.10934v1

[6] Nizar F, Loudahdi K, Jean-Luc B, El-Miloudi E "Intelligent Surveillance System Based on Stereo Vision for Level Crossings Safety Applications" Recent Developments in Video Surveillance, chapter 5, [Online] Available: https://books.google.co.in/books?id=XC-aDwAAQBAJ&lpg=PA75&ots=A6slTPM8Ad&dq=Automatic%20Video%20Surveillance%20for%20Railway%20Level%20Crossings%20Using%20Stereo%20Vision.&lr&pg=PA75#v=onepage&q&f=false.

[7] Trudi Farrington-Darby, Laura Pickup, John. R. Wilson, " Safety culture in railway maintenance." , 2005, MDPI, Safety Science, vol 43, issue 1, DOI: https://doi.org/10.1016/j.ssci.2004.09.003..

[8] H. P. Haryono and F. Hidayat, " Trespassing Detection using CCTV and Video Analytics for Safety and Security in Railway Stations", 2022, International Conference on ICT for Smart Society (ICISS), pp. 01-04, DOI: 10.1109/ICISS55894.2022.9915245.

[9] A. Abduvaytov, R. M. Abdu Kayumbek, H. S. Jeon and R. Oh, "The Real Time Railway Monitoring System suitable for Multi-View Object based on Sensor Stream Data Tracking," 2020, International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2020, pp. 1-4, DOI: 10.1109/ICISCT50599.2020.9351474.

[10] Sergio A. Velastin, Boghos A. Boghossian, Maria Alicia Vicencio-Silva, "A motion-based image processing system for detecting potentially dangerous situations in underground railway stations.",2006, Transportation Research Part C: Emerging Technologies, Vol. 14, Issue. 2, DOI: https://doi.org/10.1016/j.trc.2006.05.006.

[11] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.", 2014, IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[12] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.", 2017, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, DOI: 10.1109/TPAMI.2016.2577031

[13] Vijay Dubey, "Evaluation Metrics for Object Detection Algorithms",2020, Medium, [Online] Available: https://medium.com/@vijayshankerdubey550/evaluation-metrics-for-object-detection-algorithms-0d6489879f3

[14] Timothy C Arlen, "Understanding the mAP Evaluation Metric for Object Detection", 2018, Medium, [Online] Available:https://medium.com/@timothycarlen/understanding-the-map-evaluation-metric-for-object-detection-a07fe6962cf3.

[15] Google Colab: Collaborative Research and Computing. Google Research. (2020). [Online] Available: https://colab.google/.

[16] NVIDIA T4: Performance for Inference, Deep Learning, and Data Analytics. NVIDIA Corporation. (2018). [Online] Available: https://www.nvidia.com/en-in/data-center/tesla-t4/

[17] Shorten, C., Khoshgoftaar, T.M, "Data Augmentation Techniques for Deep Learning Models",2019, Journal of Big Data, [Online] Available: https://doi.org/10.1186/s40537-019-0197-0.

[18] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", 2012, Advances in Neural Information Processing Systems 25 (NIPS 2012), vol:25, [Online] Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf