IJARSCT
ISSN (Online) 2581-9429
Impact Factor: 6.252
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022

Detection and Alert System for Railway Trespassing (DART)

Yogesh Rajgure[1], Karthik Arumugam[2], Harsh Ingle[3], Sagar Dhanake[4]
B.E. Student, Computer Engineering, DYPIET, Ambi, India [1,2,3]
Project Guide and Professor, Computer Engineering, DYPIET, Ambi, India [4]

Abstract: - *Trespassing on Railroads is a dangerous activity and more dangerous when it is committed near the station. Also, is a punishable offense under section 147 of the Railways Act, 1989 where a penalty of imprisonment of up to 6 months and/or fine is leviable in India. Still, a high number of cases of trespassing are reported every day at nearly every station. Even regular patrolling of potential sites is not possible due to increasing crowd and low manpower and high personnel costs. Due to the variability of the crowd, we cannot find a pattern that will cover total trespassing. All these and more reasons raise a need to find an alternate automated approach to deal with this situation. The latest Machine Learning and Computer Vision techniques provide us a direction towards making our Live Detection and Alert System for Railway Trespassing using video surveillance from CCTV cameras all around the station. Unlike various approaches suggested before this paper. DART proposes to deal with the unpredictable trespassing violations on a real-time basis and to allocate the minimum resources at proper places at the proper time to have the highest possible chances of averting the accidents. As a real-time detection system, DART is a single-step detector and hence, like some other two-step detectors, does not leverage much of the sparsity of railroad trespassing activity, and thus has an advantage of being able to detect other activities too. DART system, despite being working on live data handles the trade-off between accuracy and computational time well. DART system demonstrates efficacy on collected videos by achieving a 0.85 f1 score.*

Keyword: - Railway Trespassing detection, Railroad safety, trespassing, Deep learning, Video surveillance, webapp, Background subtraction, Computer vision, YOLO, YOLO tiny, etc.

I. INTRODUCTION

Railroad trespassing is an issue of importance, this term is widely discussed but lacks proper assistance. Reports suggest around 3000 deaths and 10,000 injuries around 2006-15 as a direct result of trespassing and new reports from the Railway Minister of India stated that 30,000 deaths due to trespassing and untoward incidents in 2019-20 alone, and the worst part is that, this is the data of only one country. India accounts for more than 7000 railway stations and around 31,847 railway crossings. Most of them are very rarely patrolled and some of them are too overcrowded that resource allocation is not possible. In most cases, an accident is fatal for the trespasser. Aside from human life, these accidents are exceeding expensive due to property damage and compensation, etc.

Previous research in this area suggested the solutions like Railway police patrolling near tracks and setting up CCTV cameras and a human employee to raise alarms or send personnel to the location. But both the approaches fell short in one situation or another. Another solution suggested is to study the patterns of trespassing from the recordings and be ready at the same time and location to bust the trespasser. For example, College students regularly cross the track at the same time, etc. But this will also mean that only patterns are getting recognized and we are totally ignoring the randomness/unpredictability of the situation.

A. GOALS OF THIS RESEARCH

With the live feed from CCTV cameras, the system has to be provided with the coordinates of our Area-of-Interest for the frames from each CCTV. So, we do not define a person as a trespasser only for being in the camera field of CCTV, but that person should also be inside the ROI for that CCTV. As a result of such an arrangement, the same camera can serve multiple purposes and when decided to change the angle or alignment of the camera, the system can get going by only changing the coordinates of ROI.

We also want to detect trespassing and send an alert in the time as fast as possible. So, we have to embrace the trade-off between detection time and accuracy of the deep learning model that we use for the detection. As we are looking for faster, live actually, predictions, we will have to sacrifice accuracy. Still, the suggested model is worth it and provides good accuracy for good computational time.

B. State-of-the-art / Related work

There has been considerable work done in this field by the research community. Salmane et. al. [6] have proposed a technique to detect hazardous situation crossings, but this technique only detects the movement and is not able to discriminate moving objects between trains, persons, etc., also it does not leverage the advanced deep learning models. Muzammil et. al. [1] have proposed ARTS model to predict trespassing and detections are based on two-step methods that take leverage of sparsity of trespassing and deep learning models. But it suggests predicting the pattern while the trespassing can be totally random and not fit in any pattern, also the use for live detection is not mentioned and it makes the assumptions like all the areas visible in the camera are ROI and any human in that ROI is trespassing.

C. Approach

DART (Detection & Alert System for Railway Trespassing) framework can help greatly in preventing trespassing and related accidents as it adopts a single-step detection process. Whereas, the whole system can be described as a # step process. We already have the videos from CCTV and we have described the coordinates of ROI. In step 1, the detection algorithm is applied over the frames and the humans are detected. In step 2, we check if the detected human lies inside the ROI or outside the region of interest and gather the information for detected bounding boxes that are inside the ROI. In step 3, if we turn on the alert, it may be a sound alert or something else according to the situation and proper assistance can be provided at the place quickly and accidents can be averted.

Also, the data collected can be stored and further utilized for proper analysis, predictions, etc.

II. Problem Definition

Trespassing is not a new issue near railway stations. Especially in a country like India where stations are very crowded that even railway police fail to cover all the area and this results in accidents. The main problem that can be seen is the problem of resource allocation, e.g. If the railway police personnel know the exact place where this violation is being done, they can easily avoid the incidence and punish that person. So, a live trespassing detection is needed to alert the trespassing at right time to avoid an accident.

III. Literature Survey

Sr. No.	Title	Authors	Year	Description/Approach
1	A deep learning approach to trespassing detection using video surveillance data.	Muzammil Bashir, Elke A. Rundensteiner, Ramoza Ahsan	2019	In this paper the researchers have developed their automated framework called ARTS - Automated railroad trespassing detection system. This system uses a 2-stage approach based on CNN to solve the trespassing detection.
2	A deep learning approach towards railway safety risk assessment.	Hamad Alawad , Sakdirat Kaewunruen , Min An	2020	It proposes an effective Realtime risk management solution based on CNN to improve safety throughout the entire railway industry by preventing fatal accidents.
3	YOLO Based Real-Time Human Detection for Smart Video Surveillance at the Edge.	Huy Hoang Nguyen, Thi Nhung Ta, Ngoc Cuong Nguyen, Van Truong Bui, Hung Manh Pham, Duc Minh Nguyen	2021	This paper proposes an approach based on YOLOv2 for human detection. This approach utilizes combined benefits of YOLO residual blocks and multiple spatial pyramid pooling blocks. The proposed model shows high accuracy across various datasets.
4	Analysis of Deep Learning Architectures for Object Detection - A Critical Review	Mohit Pandiya, Sayonee Dassani, Dr. Mangalraj P	2020	This paper proposes a solution having 4 different architectures for comparative analysis and explains the optimality and compatibility.
5	Video Abnormal Event Detection Based on CNN and LSTM.	Guangli Wu*, Zhenzhou Guo, Leiting Li, Chengxiang Wang	2020	In this paper, CNN combined with LSTM of convolutional neural network was used to build the model of abnormal event detection in video. UCSD and UMN dataset to carry out the experiment of abnormal event detection and analysis in video, and achieved certain good results.

IV. Proposed Work

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 6, May 2022

Deep Learning in Object Detection

Continuous development in classification and detection models over the years has paved the way for precise, faster, and computationally efficient models. Right from the success of LeNet, AlexNet, InceptionNet, VGG, ResNet, MobileNet, and EfficientNet to the RCNN family and Yolo family. Of all the above-mentioned, RCNN and YOLO are by far the most used algorithms in today's world.

According to the need of the project, we go with RCNN if we prefer accuracy over speed, or else, we go with YOLO if the speed of detection is more important to us and we can deal with little less accurate detections. If we see the benchmarks for the latest versions of these two, we clearly see that YOLO v5 small gives the speed of 52.8 FPS against 21.7 FPS of Faster RCNN ResNet 50 when tested on NVIDIA GTX 1080 Ti [8]. The main difference between these (Yolo & RCNN) algorithms is that RCNN is in itself a two-step detection algorithm, where regions are proposed in the first step and then the objects are classified in the second step, whereas the Yolo is a single shot detection algorithm where both the steps happen in one shot, hence the name You Only Look Once. This gives us an idea to make use of your models to meet the demand for speed as we are interested in live detections.

V. Methodology

Given the live video as input frames, every third frame is passed to the model and all detected humans are classified into two classes- trespasser (1) and non-trespasser (0) based on their location and defined ROI. And the presence of classified value 1 leads to the activation of alarm in that area.

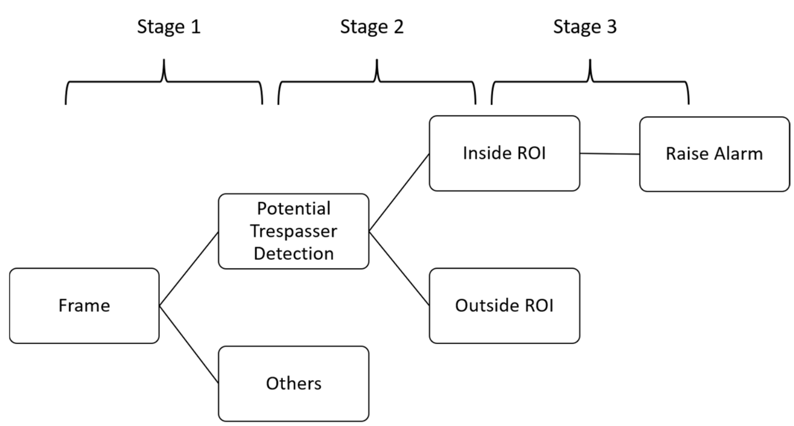


Figure 1: Stages of Application

VI. System Architecture

A. YOLO

When it comes to high-speed object detection, the YOLO algorithm is the best solution till now. Against the 2-step approach, where localization and then detection is done like in RCNN, YOLO adopts a single-step detection approach where it predicts the bounding box coordinates and possible classes of the predictions.

It does so by first dividing the image into N grids, each having the equal dimensional region of SxS. Each of these N grids is responsible for the detection and localization of the object it contains. Now, these grids predict B bounding box coordinates the process is greatly fastened as the computational time is drastically reduced since both the detection and recognition are carried out by cells from different image. This process on other hand brings in a lot of predictions for a single object since multiple cells are predicting the same object with different bounding boxes. This issue is countered by YOLO which uses non-max suppression to cancel out the prediction boxes that are not as relevant or close to the best bounding box which has the image pixels as a part of the local maxima. Even though we got the speed we still need the lighter model so that we can be able to use it with small computation power. This is where YOLO tiny comes into the picture.

B. YOLO Tiny

YOLO tiny is a compressed version of YOLO. It can be considered the simpler form of YOLO. The network structure and the parameters in YOLO tiny are simple and reduced respectively. It is also faster as it is trained on about 29 pre-

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 6, May 2022

trained layers as opposed to 137 pretrained convolutional layers in YOLO v4. We can estimate the speed of YOLO v4 tiny as 8 times faster than YOLO v4, which is a huge difference, also the accuracy is reduced to 2/3 of YOLO v4. Even then, YOLO v4 tiny is selected as the model for this project because we only have to predict the trespasser only once from all the frames in which he/she is present inside the Region of Interest. And that will be enough for raising an alarm.
All the above-mentioned factors contribute to the favor of selecting YOLO tiny as the model for live detections in this project.

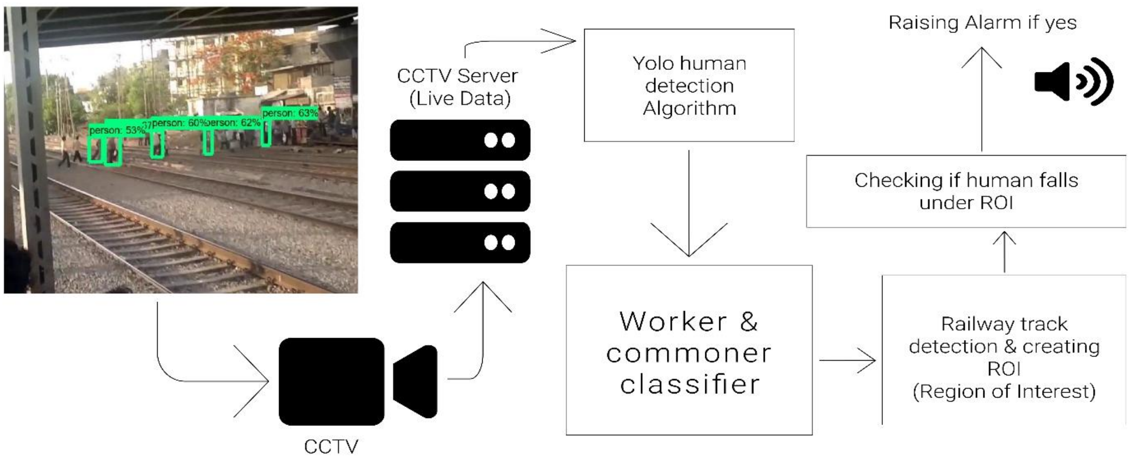


Figure 1: System Architecture

VII. System Implementation

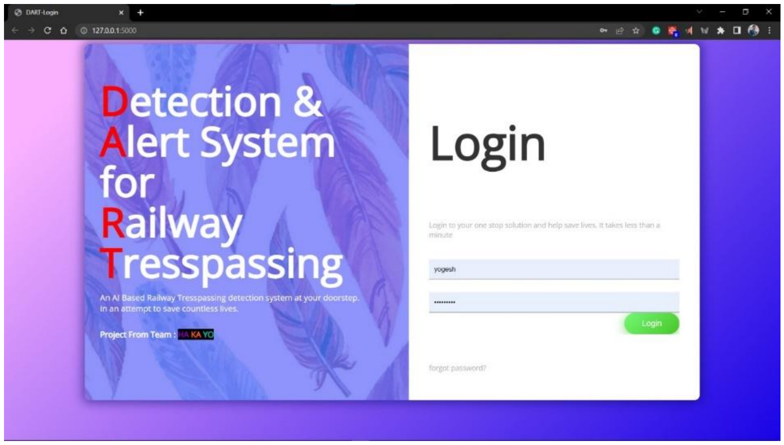
DART Framework

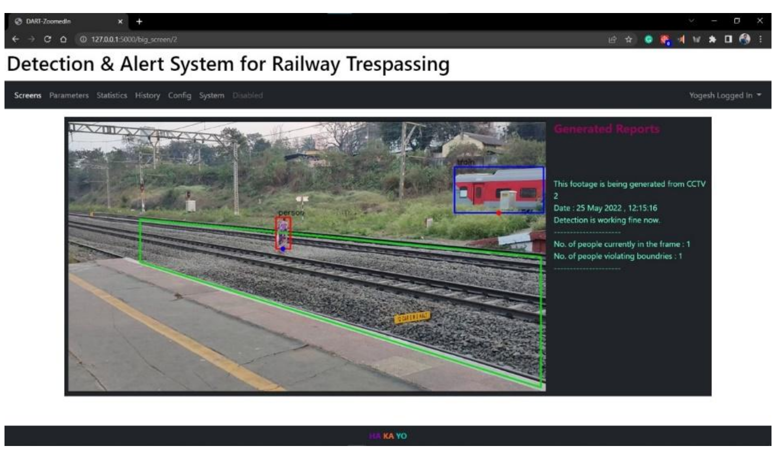
Figure 1 shows the outline of the DART framework. In the first stage, the frames are collected from the live video feed from the CCTV cameras placed at different locations across the station, and every frame is then set to the proper aspect ratio, and different regions of interest are described for each of the CCTV videos. All this information for each video is easily configurable through the config files provided. Then this information is forwarded to the TensorFlow framework to perform detection. Here we have used the YOLO tiny weights for the coco dataset and converted them to the TensorFlow weights for fluent workings of the code. The threshold for detection is set to 0.4, it is kept low in a quest to lower the True Negatives as much as possible so that we get nearly all the trespassers and proper assistance to be provided. The model takes this into consideration and provides the bounding box coordinates and class names for detections.
Figure 2 illustrates the architecture of the DART model. IN step 2, a potential trespasser detected is then allocated a trace-point which is calculated as the middle point of the lower horizontal line of the bounding box. This location of a point is selected as this location is very near to the foot of a human and while walking on railway tracks, feet are on the ground and so is the bottom line of pred box. So, finding a trace-point inside the region of interest (ROI) indicates that a particular human is a trespasser. The system then generates the list of these detected trespassers and sends a signal for the occurrence of even 1 trespasser.
In the last step, an alarm is raised at the required speaker (or in any specified way) based on the output of step 2.
We can say that,

Step 1 of the model takes frames as input and gives the classes and coordinates of detections as the output. (Here, the "potential trespasser" is a human who is not a worker, according to the detector- a person who is not wearing a vest and helmet.)
Step 2 takes classes and coordinates of detections as input and provides a list of detections for every camera and alert (1/0) as output.
Step 3 takes camera wise list of detections and alert key and turns the alert on/off accordingly.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 6, May 2022
Impact Factor: 6.252

All the above steps are attached to the front end using the flask framework. Keeping in mind, the future scope of the project, the front-end is a web app that further opens huge possibilities of data sharing across a network, etc. The flow is easy so that the non-technical person can easily be able to log in and monitor the activities. The sessions and login only with credentials steps are added so that to keep it safe from falling into wrong hands.





VIII. Result

The proposed solution is capable of running on live CCTVs at minimal computation power. The implementation for the same is tested on a computer with i5 9th gen intel quad-core CPU with 8 GB RAM and 8 live input video feeds where the detection accuracy was recorded to be 0.7, but as explained above that even one detection of a trespasser in his/her journey through Region of Interest. The speed of 5 FPS per video was recorded while running 8 videos through the CPU with the above specs. These figures will get better with a better system.

IX. Advantages

- Live detection with better accuracy and computational time.
- The minimum need for new infrastructure, as the cameras are already installed on most of the stations can be utilized by just describing ROI.
- Minimum computational power is required as the model used for the system is very optimized.
- DART system is flexible, so with the availability of a better model, the same system can be used with minor changes.
- Proper assistance and support can be provided to the proper place at the proper time.
- With the decrease in accidents, a lot of money involved can be saved.
- Travel experience of travelers can be improved which attracts more tourists.
- More amount of relevant data can be gathered from the DART system and can be further used to get more insights from it.
- Trespassers can be punished accordingly so that fewer people try to attempt trespassing.

X. Disadvantages

- Detection of many videos requires a better-performing CPU/GPU and a bigger memory size.
- Setup and maintenance can be costly as a skilled person is required.
- Increase in the power consumption due to continuous use of GPU.

XI. Conclusion

In this report, we propose and then comprehensively study the Detection and Alert system for Railway Trespassing, (also called DART) by utilizing the latest techniques of Deep Learning. The proposed system can trade-off between accuracy and computational power and still provide better detections on the live feed. Although this report lays focuses on the use of this system for Railway Trespassing, this system has the potential to be used for many applications in video surveillance or any other application where the focus is on detection in ROI. Also, this system has the capability to obtain and store huge data which can be further utilized for deeper analytics in the area and can even be able to predict the trespassing patterns among other things.

Acknowledgement

This project (research paper along with the proposed solution) would not have been possible if it weren't for a few people. To understand the problem statement better and to come up with a better potential solution was a collection of insights from many colleagues and professors. We would like to express our sincere gratitude, and deepest vote of thanks to our Project Guide Professor Sagar Dhanake, without his constant support and guidance throughout the semester, this project could not have been possible. Finally, we would like to extend our heartfelt gratitude to our friends and family members.

References

[1] Muzammil Bashir, Elke A. Rundensteiner, Ramoza Ahsan, "A deep learning approach to trespassing detection using video surveillance data," 2019 IEEE International Conference on Big Data (Big Data)
[2] Analytics Vidya, "A Practical Guide to Object Detection using the Popular YOLO Framework – Part III (with Python codes)", Available: https://www.analyticsvidhya.com/blog/2018/12/practical-guide-object-detection-yolo-framewor-python/.
[3] Hamad Alawad, Sakdirat Kaewunruen, Min An, "A deep learning approach towards railway safety risk assessment", 2020 IEEE Access.
[4] Huy Hoang Nguyen, Thi Nhung Ta, Ngoc Cuong Nguyen, Van Truong Bui, Hung Manh Pham, Duc Minh Nguyen, "YOLO based Real-Time Human detection for smart video surveillance at the edge", 2021 IEEE Eighth International Conference on Communication & Electronics.

[5] Mohit Pandiya, Sayonee Dassani, Dr. Mangalraj P, "Analysis of Deep learning architectures for object detection – A critical review", 2020, IEEE – HYDCON
[6] Houssam Salmane, Louahdi Khoudour, and Yassine Ruichek. A video analysis-based railway–road safety system for detecting hazard situations at level crossings. IEEE Transactions on Intelligent Transportation Systems, 16(2):596–609, 2015.
[7] YOLO: Real-Time Object Detection Explained, V7 Labshttps://www.v7labs.com/blog/yolo-object-detection
[8] Guangli Wu*, Zhenzhou Guo, Leiting Li, Chengxiang Wang, "Video Abnormal Event detection based on CNN and LSTM", 2020, IEEE 5th International Conference on Signal and Image Processing.