CCTV based Enhanced Public Security Management System for Sri Lanka

Sandaruwan B.G.P.
Dept. of Information Technology
Sri Lanka Institute of Information
Technology
Malabe, Sri Lanka

Samaraweera D.H.M.
Dept. of Software Engineering
Sri Lanka Institute of Information
Technology
Malabe, Sri Lanka

Deshan A.S.S.B.
Dept. of Software Engineering
Sri Lanka Institute of Information
Technology
Malabe, Sri Lanka

Amanda Hemantha K.A.D.
Dept. of Information Technology
Sri Lanka Institute of Information
Technology
Malabe, Sri Lanka

Kavinga Yapa Abeywardena
Dept. of Computer Systems
Engineering
Sri Lanka Institute of Information
Technology Malabe, Sri Lanka

Laneesha Ruggahakotuwa
Dept. of Computer Systems
Engineering
Sri Lanka Institute of Information
Technology Malabe, Sri Lanka

ABSTRACT
With the current development of technology, CCTV surveillance is widely used as a method of observing uncivilized public behavior, traffic-related violations, and manmade disasters. With the increase of human misbehavior, traffic on roads, and fire-related accidents in Sri Lanka, automated video tracking of those scenarios is the immediate need. Automation is crucial as the random nature of the above-stated incidents. As the goal of this research, a novel approach, a CCTV-based public security system that is capable of automatically detecting Fires, Road Accidents, Traffic rule violations, and Civil Unrest is proposed to implement using Image processing and Machine Learning. Having an automated security system like this will pave the way to reduce the errors that occur due to manual approaches. This document provides an outline of the proposed system and a summary of outcomes authors have achieved when training appropriate machine learning models to work with real-time data. Systems' functionality includes identifying initial objects through object detection like humans, vehicles, white lines and correctly detect and localizing accidents, white line violation on roads, the occurrence of fires, and fighting related to civil unrest from incoming video frames which belong to CCTV live steam. This system also generates real-time alerts for main controlling admins when the implemented model detects such scenarios.

Keywords
Machine Learning, Automated Security System, Real-Time Alerts.

1. INTRODUCTION
Surveillance is a widespread method used by security personnel as a method of gathering information about the factors that affect security. Surveillance can be done in both human-driven and technological approaches. With the advancement of technology, electronic equipment-based approaches have become more prominent in Sri Lanka. A notable digital approach is the usage of CCTV to gather video and audio data about the environments with security concerns through remote observation instead of physical observation. From 2010 onwards Sri Lanka Police declared open a CCTV unit considering the past historical events like terrorism which highly affected the country's security. According to Sri Lanka police web resources, CCTV surveillance is used to establish the country's security in many aspects. Through the gathered surveillance images/videos Sri Lanka police take action to provide immediate police responses for emergency security situations and direct supervised police actions. Apart from basic systems, there is a demand for automated systems with the capability of automatic detection and alert generation as basic systems are prone to make errors due to their human-driven approaches. The current prevailing system of Sri Lanka completely dwells on manual supervision done by an administrative panel. In this research, the key areas the authors have taken into consideration are occurrences of fires, Traffic rule violation, Accidents, and Civil unrest. The number of accidents that happen in day-to-day life has been approximately getting high. A few-minute delay in reporting an accident will cause the loss of many lives [1]. Heavy traffics, malfunctioning vehicles such as low brakes, uncontrollable vehicle speed, disobey traffic rules by irresponsible drivers, and unconditional roads could be the reasons for accidents in Sri Lanka[2]. Road Safety is an essential factor to create a safe driving environment within a nation. Most countries spend a high amount of money to ensure that the road conditions are always suitable for people to travel more safely, and a higher budget is provided for the police department to ensure that every vehicle is behaving according to the traffic rules and regulations. Energy consumption related to technological equipment is still high today resulting in accidents related to fire. In such kind of situation, the individuals who have been caught in the fire are brutally suffering. So, in that type of condition, the possibility of making them remind to make calls for respective authorities or fire brigades is very less [3]. There is an issue as the time when observers of the situation calling may be delayed and the information provided may false or insufficient to act according to the situation and even many more troubles have occurred due to the inaccurate information and the improper convey of information[3]. Civil unrest scenarios can cause major dangers to human lives as they can cause injuries that can be fatal, also public property destruction, and cause interruptions to the day-to-day activities of people due to fear. The method that the police get notified about sudden occurrences of unrest scenarios is by external sources who

observe the scene and manually inform via emergency hotlines.

The authors have utilized machine learning and image processing techniques to automatically detect incidents related to these key concern areas. The human-driven manual detection system in Sri Lanka has drawbacks. The main intention of this research is to create an effective security system that accurately detects and provides alerts that can enhance security throughout the country.

2. METHODOLOGY

With the vast urbanization and population growth in urban areas of Sri Lanka the rate of incidents that violate security is increasing. Fires, Traffic rule violations, Accidents, and Civil unrest are the most prominent areas where people tend to break rules and disasters happen due to misfocus.

Identification of Fires, Traffic Rule violations mainly based on White lines, Traffic Accident detection and Civil unrest situations involving fights need to be implemented to match the objectives of the proposed system. Since these identification tasks are closely involved with police officials, manually inspecting those situations through CCTV data feeds is time-consuming and unproductive. Considering the above circumstances many research that has been conducted has suggested automatic detection.

To decrease the workload of the administrative officials, the need for an effective notification generation mechanism is also an identified issue. The above-described security incidents must be identified in real-time, and the notifications must be generated right at the moment of detection to conduct successful investigations. Most of the selected security violating incidents are prone to fatal accidents so a need for real-time detecting an alert generating system will reduce further damage.

2.1 Object Detection

The object detection phase is organized by studying the related research papers, identifying the main features implemented and requirements that have been fulfilled by previous researchers. With the broad scope of the CCTV surveillance domain, identifying the relevant areas related to security in Sri Lanka was a challenging step for the authors. When considering the sub-components that each author has, it is considerably challenging to collect requirements and defining the scope for system implementation. After requirement gathering done with Sri Lanka Police CCTV Division, authors have analyzed suitable system requirements for each component that matches the conditions in Sri Lanka. Also, by studying the previous researchers, the authors finalized the features of the system. Below are some multiple steps the authors followed,

• Find related research papers

• Background study of CCTV surveillance domain

• Feasibility study related to Sri Lanka.

• Financial analysis.

• Analyze research paper findings

• Analyzing and completing each components' features.

• Completion of requirements

First, the models must be implemented to identify the initial objects involved in the proposed security areas. CCTV video footage gets processed and separated into frames. For the inspection of the model, frames with identified initial objects are selected to increase efficiency. For instance, frames with vehicles are identified and sent to be identified for possible occurrences of accidents. Similarly, frames with human images get selected to inspect for possible scenarios related to the unrest. To implement the identification and object detection process datasets must be collected which are containing actual CCTV footage which is a very challenging task.

Dataset is the key requirement for implementing the four models. The detection accuracy of the model is dependent on the quality of the selected datasets. Datasets were collected from the Sri Lanka Police CCTV Division which includes Road accidents, Traffic Rule Violation for White lines, Crime incidents related to unrest, and Fires. Online Video datasets also have been used for Fire and civil unrest identification [4][5][6]. Video data related to the fire was also collected by the Fire brigade of Sri Lanka.

Video frames from the collected datasets should be preprocessed and undergo for a labeling process to train the model using machine learning. Each dataset was split into training, testing, and validation sets using the video frames of the same dataset. 80% of the frames were used for testing and 20% of the frames were used for training and validation by referring to the deep learning model training techniques for object detection.

Image augmentation is the process that authors used to artificially increase the dataset size as the available datasets are limited. This is done to increase the generalizing ability of the machine learning model which affects its performance[7].

When selecting an appropriate machine learning architecture authors had to consider the fact that all operations should be done to a sequence of video frames coming from CCTV/RTSP stream. Due to that reason an architecture suitable for processing real-time data needed to be selected which also uses less computational resources. The widely used methodologies by researchers are CNN-based architectures in the past for image and video classification problems. Those methods utilize a massive amount of data and have a very high training overhead [8].Authors have selected YOLO as the object detector after experimenting with CNN as YOLO produced accurate results with less computational time which is very ideal for real-time detection.

Table 1. CNN versus YOLO

Algorithm	Metrics		
	Detection Accuracy	**Computational Time**	**False Positive**
CNN	High	High	Low
YOLO	Comparatively Low	Low	High

YOLO shows some inaccuracies when detecting as stated in Table 1 above, compared to state of art detection algorithms but its detection speed makes it ideal to perform operations to real-time data[9].
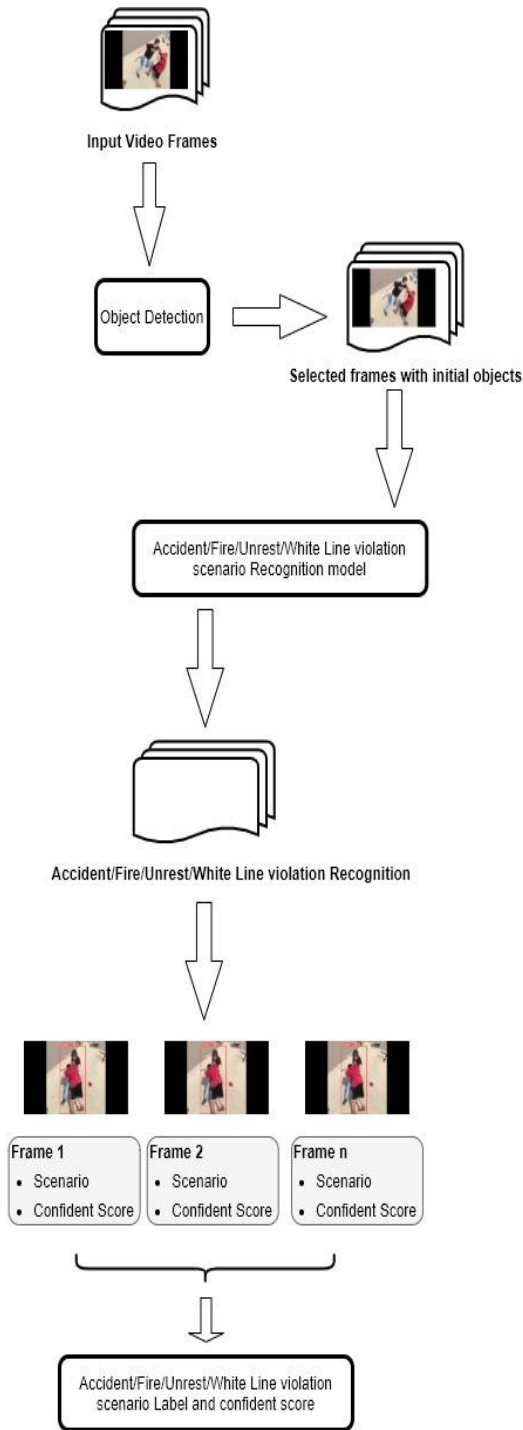
Fig 1: Algorithm architecture

labels needed to be stored and paths for labels and frames needed to be fed to the model. The configuration files which consist of YOLO architecture layers also needed as input. Then the pre-trained convolutional weights are fed to the model. As the result of the training process, a trained weight file is obtained. It is crucial for the development of the proposed system as it determines the output according to the input data. The process of the model starts with the weights file obtained from the training. YOLO v5 uses PyTorch forms to read the weights file.

The creation of the concluded prediction confident score for each class was done by obtaining the average of all detected confident scores of the detected frames which exceed the selected threshold value.

Once the model has identified the object, it temporarily retains the identified features of that frame and checks whether the following frames have a connection to the previously received frame analyzing its data. If the model identifies a connection between the receiving frames, it verifies the detected scenario as true and proceeds to the next steps for alert generation.
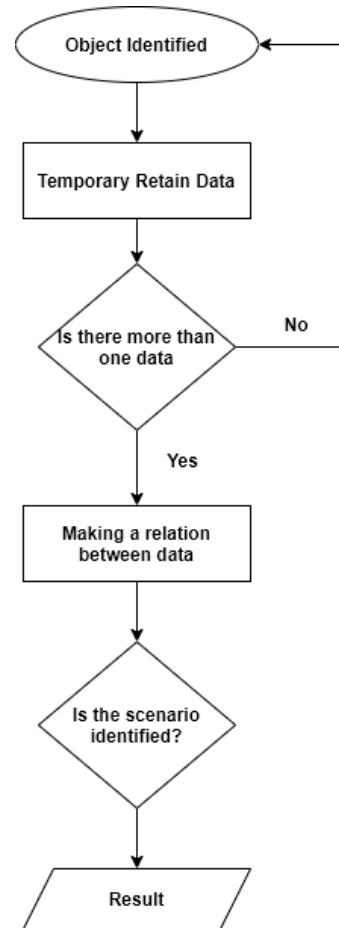

Fig 2: Flow diagram for object detection

Here according to Fig. 1, YOLO is used to detect, localize, and recognize the scenarios in interest which include accidents, fires, civil unrest, and White line violation on roads. It acts as an object detector for identifying initial objects responsible for scenario detection.

2.2 Scenario Detection and Alert Generation

During training, YOLO requires a set of data about each frame. First, it requires coordinates of the objects and scenarios present in the image to locate them. It becomes the label of each class where objects are defined. These generated

As showed in figure 2, the video data captured by the CCTV cameras are transmitted to the system via the RTSP protocol using the OpenCV library and framed into images. The proposed algorithm analyzes the frames. Investigates all aspects of highway accidents, traffic violations, civil unrest scenarios, and fires. Once the model has identified and localized scenarios present in frames, it builds a connection between the frames and confirms the scene. When these factors are verified, external conditions in the environment,

such as wind speed and weather, are obtained from a genuine outside party and given as a percentage of the serious situation that may or may not has occurred.

This is essential for follow-up stages because the algorithm stores the data of the event video segment contains event-verified frames in the database, such as the percentage of the identification, detected camera location, area, and time recorded.

There is research that has come up with various approaches when detecting road accidents, traffic rule violation, violence, and fires. The proposed methodology has to be implemented to increase productivity within the established security system in Sri Lanka. Because of that, a novel methodology that can detect multiple security scenarios is the solution. Most of the research entirely focuses on a single area of concern among all the above factors.

Also apart from the conducted research, this proposed system can generate alerts that are classified according to the severity situation. The proposed system is fulfilling the identification of different scenarios from one platform and operates in real-time based on CCTV surveillance footage. Also, must ensure to consider the various factors that affect Sri Lanka including the current police surveillance system and legal aspects.
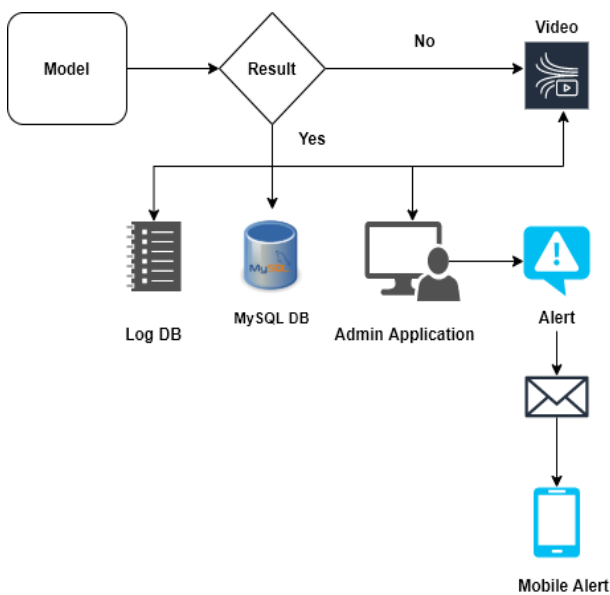

Fig 3: System architecture and alert generation

The notification delivers to the registered users of the system. The users can register to the system through the registration process of administration. After classification, if the results match any of the scenarios, it will trigger the notification system to send necessary details and a preview of the event in the desktop application. The desktop application is handled by the administration unit of the CCTV division of the Sri Lankan police. The incident location, weather details, and the video which contains the details of the incident will be visible in real-time. The severity of the situation is monitored by analyzing background data. As an example, after identifying an accident, if the accident has caused a traffic congestion detail about the number of vehicles and amount of people in the background will be gathered to determine severity. When the detection algorithm identifies objects, for example, humans and vehicles, the tracking process is done through the

number of objects pass through a created virtual line from a tracking algorithm[10]. Using a mobile SMS service researchers were able to inform details about the incident to other relevant authorities and users.

Simultaneously, an identified scenario of the incident from footage will be saved to the MySQL database with obtained details and for recording purposes, a separate log database will be created to maintain the included incident details.

3. RESULTS AND DISCUSSION

After demonstrating and testing different approaches and methods researchers determined that the best and optimal algorithms are suitable for the proposed and desired system. Image processing and object detection were done by YOLO V5, which was the most accurate among all the technologies. Using YOLO V5 researchers were able to obtain above 85% accuracy for all the objects and classify them. A sample image is shown in Fig.4 and 5 below.

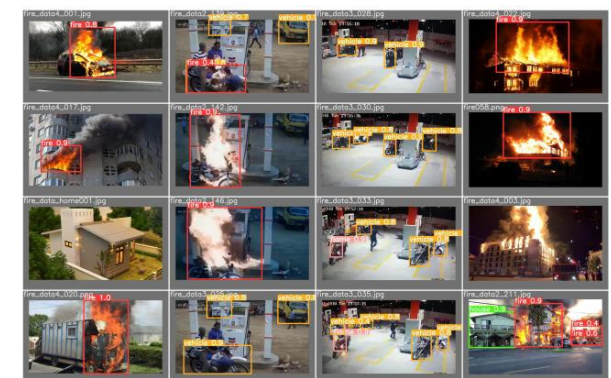**Fig 4: Object detection results for a white line violation**

Fig 5: Object detection results for fire incidents

All the object detection scenarios which are included in the model's performances were evaluated using the test data bundles as mentioned above. Each scenario showed different results according to the data sets and objects although training accuracy for the model was above 92% as well as testing accuracy was above 85% which was obtained by using a split of data that was used to train the model.

To maximize the accuracy for the researchers' desired applications of the model, training was done by using a

minimum of 5000 datasets of positive images in addition to negative images. Using a high number of datasets leads to a high accuracy model.

The scenarios mentioned above will detect objects and additional requirements which are mandatory for a model with ease. Some of the features of the model are below,

- Vehicle detection
- Fire detection
- Road line detection
- Human figure detection
- Fight detection
- Vehicle detection
- Accident detection

Examples for some of the mentioned features are shown in Fig.6 and 7 below,

Fig 6: Accident and vehicle identification

Fig 7: Human figures and fight identification

As seen in Fig. 6 and 7, the model is accurately identified relevant objects with minimum errors. The model reveals the finest accuracy as it trained with datasets that included various scenarios. Researchers used a set of test data to test the trained model which used Yolo v5 and acquired accurate results that are shown as in figure 7 below. As it shows the best accuracy the algorithm provided was when researchers used 60 epochs. With 60 epochs researchers were able to obtain 92% of accuracy.

The reason to use YOLO V5 instead of faster R-CNN was YOLO V5 is mainly because it can detect small objects greater than faster R-CNN furthermore YOLO v5 runs about 2.5 times faster than the competitor which is faster R-CNN. By comparing the accuracy of data which shows in table 2 below, YOLO V5 responds with greater accuracy than the Faster R-CNN. The mean Average Precision (mAP) of YOLO V5 is much higher than the faster R-CNN.

Table 2. Accuracy results for Faster R-CNN and YOLO V5

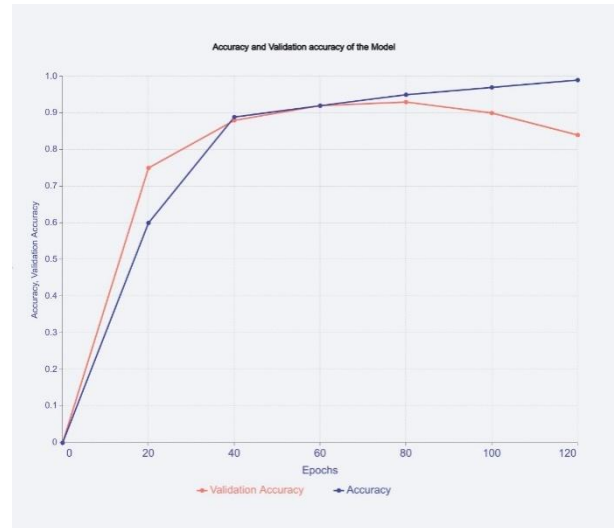|	Faster R-CNN	YOLO V5
AvgmAP	63.0%	82.7%
Accident	55.5%	85.3%
Vehicle	56.8%	91.5%
Fire	67.5%	71.8%
Smoke	65.5%	51.6%
Fight	79.9%	97.5%
Human Figure	79.3%	97.0%
White Line Violation	36.5%	84.6%

Fig 8: Accuracy graph for YOLO V5.

BelowFig. 9,10,11and 12 depict the YOLO V5 model training result for each component respectively.
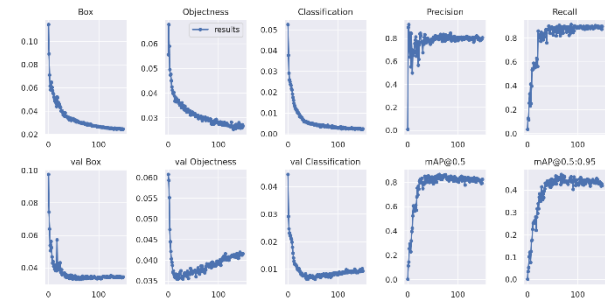

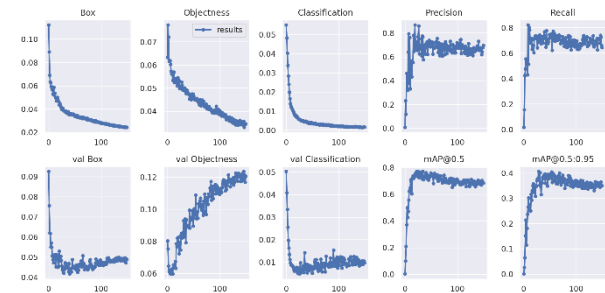**Fig 9: Different evaluation parameters observed for traffic accident detection**

Fig 10: Different evaluation parameters observed for white line violation on roads
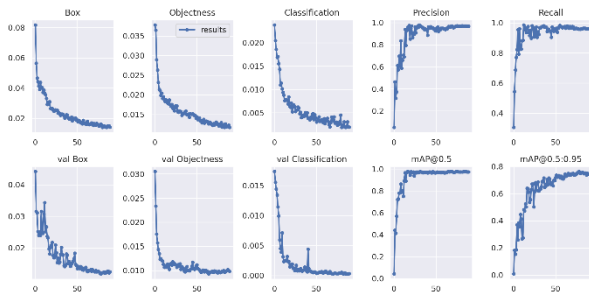
Fig 11: Different evaluation parameters observed for unrest behavior detection
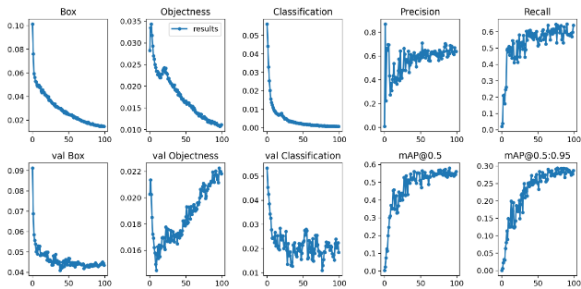

Fig 12: Different evaluation parameters observed for fire detection

Overall, it is shown that the mean average precision is between the range of 0.6 -0.9 for bounding box IoU threshold of 0.5 and the recall is in the range of 0.6-0.9 for the majority of the components.

4. CONCLUSION

This research study presents a solution to enhance the security of Sri Lanka using CCTV surveillance and automatic detection. Here it proposes an intelligent system that can detect security concerning incidents which are fires, road accidents, traffic rule violations related to white lines, and civil unrest. Researchers have discussed the current prevailing methodologies that exist within the country and how this research addresses the drawbacks. Through this research, paper authors have presented the importance of the usage of machine learning-based automatic detection utilizing the results of experimented algorithms to work with real-time data with higher accuracy. Hence, in conclusion, the proposed system will pave the way for efficient disaster management and investigations for the selected security concerning incidents to be conducted productively without delays. Time efficiency is obtained by the real-time alert generating feature for the detected incidents.

For the best performance, CCTVs that are connected to the proposed system can be deployed in places where there is a high need for discipline and security and in places where the above scenarios are frequently occurring. Using the proposed methodology, the implementation of the system will be largely beneficial for the security officials like the police. Also, as a further development amount of CCTV live streams can be increased by allowing public users to register their CCTV streams to the developed network and as a result, a wider land area can be covered for better establishment of security throughout the country.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] E. P. Ijjina, D. Chand, S. Gupta, and K. Goutham, "Computer Vision-based Accident Detection in Traffic Surveillance," 2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019, pp. 1–6, 2019.

[2] "Road safety." [Online]. Available: https://www.who.int/health-topics/road-safety#tab=tab_1. [Accessed: 09-Oct-2021].

[3] "JeevithayaSirasa TV 28th September 2018 - YouTube." [Online]. Available: https://www.youtube.com/watch?v=OooOozD7stk. [Accessed: 09-Oct-2021].

[4] "Fire Detection Dataset | Kaggle." [Online]. Available: https://www.kaggle.com/atulyakumar98/test-dataset. [Accessed: 09-Oct-2021].

[5] M. Perez, A. C. Kot, and A. Rocha, "Detection of Real-world Fights in Surveillance Videos," ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol. 2019-May, no. February, pp. 2662–2666, 2019.

[6] "Accident Detection From CCTV Footage | Kaggle." [Online]. Available: https://www.kaggle.com/ckay16/accident-detection-from-cctv-footage. [Accessed: 09-Oct-2021].

[7] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," Journal of Big Data, vol. 6, no. 1, pp. 1–48, Jul. 2019.

[8] S. Shinde, A. Kothari, and V. Gupta, "YOLO based Human Action Recognition and Localization," Procedia Computer Science, vol. 133, no. 2018, pp. 831–838, 2018.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-Decem, pp. 779–788, 2016.

[10] "People counting with OpenCV – The beginning – Konkludenz." [Online]. Available: https://konkludenz.de/en/people-counting-with-opencv-the-beginning/. [Accessed: 09-Oct-2021].