Pedestrian detection based on TensorFlow YOLOv3 embedded in a portable system adaptable to vehicles

Eduard Zadobrischi[1,2], Mihai Negru[2]
[1]Department of Computers, Electronics and Automation, Stefan cel Mare University of Suceava, Suceava, Romania
[2]Technical University of Cluj-Napoca, Cluj-Napoca, Romania
E-mail: eduard.zadobrischi@usm.ro; mihai.negru@cs.utcluj.ro;

Abstract— **With the expansion of accessibility and availability of personal cars towards the population, the environment, road safety, and pedestrian safety faces greater problems. The neediest for the management of information through autonomous systems dedicated to pedestrian detection and improvement of analysis and road prevention algorithms is an important factor addressed in this paper. The purpose of this paper is to demonstrate and propose viable solutions that will help drivers to practice an efficient, safe, and event-free driving style. This paper presents a prototype under development that can avoid various traffic events, the system analyzing, and alerting the driver regarding pedestrian intentions, marking each detection separately according to the degree of danger that it constitutes for both the driver, as well as for pedestrians. This research analyzes and emphasizes that at this point everything is already focused on the paradigm from which it is possible for all these technologies to cooperate in a hybrid platform, offering a real solution to the demands of human users but also of IoT solutions.**

Keywords— pedestrian detection, vehicle safety applications, infrastructure-to-vehicle communications, safety driving, visible light communication system, YOLO V3.

I. INTRODUCTION

Statistically speaking, with over 1.2 million victims and over 50 million people injured annually, road accidents are a major cause for our society. The continuous development of information and communication solutions in this direction encourages the demand for vehicles equipped with road safety systems that can significantly reduce this problem. In recent years, both the academic and industrial environments have provided quality studies and solutions for road safety systems based on different types of communication. One can say that the detection of pedestrians besides the utility in a wide range of applications, also gives rise to a series of challenges that differ quite a lot because the human characteristics are different, the appearance, the movements, the height, the dress styles, characteristics commonly encountered among pedestrians. Despite these challenges, many research groups over the years have analyzed and characterized in various ways these characteristics for complete pedestrian profile identification. The research groups have carried out focused analyzes in this field since the 1970s, the Histogram of Oriented Gradients (HOG) offers an efficient way to extract the characteristics of an image, obtaining a classification model to recognize objects and here we refer to Badler and Smoilar, and others [1]-[3].

For the development in this direction, hardware and software components are used to provide a starting point in the development of systems dedicated to this sector. The detection and predictability of pedestrian movements are made according to the degree of attention or concerns, these details sometimes make the detection process difficult and errors appear automatically in the analysis of objects.

Most vehicles produced after 2014 come with various systems that help the driver to exercise efficient and safe driving through the multitude of sensors and assisted driving modules. The detection and anticipation of pedestrian movement are inherently a rather difficult problem to solve due to the traffic and analysis and classification of movements performed by a pedestrian. The purpose of the paper is to evaluate the different parameters that try to characterize a certain type of behavior, degree of attention and predictability of movements, and based on this information the drivers are alerted to the severity of the danger to which they and the pedestrians are exposed [4]. The information obtained is subsequently processed and analyzed on the navigation interface mounted on the center console of the car, this offers a starting point in the development of road safety solutions [5].

II. DEBATE ON THE EXPERIMENTAL SETUP AND EXISTING SOLUTIONS

This section is presented details about the setup and implementation, also, the brief presentation of the applied methods, and the analysis of the current development stages. The development of applications and systems dedicated to driving safely have also reached points such as [6], thus putting in the foreground the control of the human factor, in this case, pedestrians. The usefulness and potential of these applications in the automotive industry are presented in [7]. The efficiency and results of such a system tested and confirmed, so there are all the premises that pedestrian detection and analysis systems are a real challenge for both the industry and the car manufacturers.

This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI - UEFISCDI, project number PN-III-P1-1.2-PCCDI-2017-0917, contract 21 PCCDI/2018.

978-1-7281-6870-8/20/$31.00 ©2020 IEEE

The communications and the transmission of information are made by visible light, an innovative method that could significantly contribute to reducing the number of road accidents [5]. Autonomous detection and prevention systems are an important topic analyzed and sought by the National Administration for Highway Traffic Safety (NHTSA), especially in the case of new cars. Such systems have been developed since 2010 by Volvo, followed by Mercedes-Benz, Audi, Volkswagen, with the support of Bosch and BMW, which also uses its own communication systems between its own cars based on DSRC technology (Dedicated short-range communications). The analysis and processing of data received from external light components such as headlights, traffic lights, street lighting, vehicles, is another advantage in terms of the entire system increasing the degree of safety offered while driving.

The component developed and presented in this article is part of a complex project based on the development of context-sensitive hybrid communication systems with applicability in the automotive field. A main advantage is that the modularity and flexibility of the system allow the implementation of new functions that can increase the complexity and efficiency. This component is supported by the whole system, but especially with the possibility of using a technological mix in terms of communications, talking about visible light communications (VLC), and RF communications (radio frequency). In addition to the low cost of creating and implementing such a system, there are also the advantages offered by the portability of the device itself. An adaptation of the system either the detection of objects or pedestrians by recognition and image processing or by physical elements consisting of sensor modules makes up a homogeneous component with perspectives in this field. The biggest disadvantage at the moment is the legislation that provides for certain measures to implement systems dedicated to traffic safety. The proposed and presented approach is a topical one in continuous development with all the novelty elements necessary for the current stage.

A. Introduction on the experimental portable setup for detection pedestrian

The development of this experimental setup, commercial Raspberry Pi 3 modules, sensor modules, and cameras dedicated to the development board as well as board cameras were used. The final system has been embedded in a 3D printed case so it can be sized to fit most car consoles. For better exposure of the application, a Raspberry Pi 3 compatible screen with the touchscreen was used to play and create the ideal user environment, see the Figure 1. In addition to detecting and analyzing objects and pedestrians within the application, information received from sensors (speed, distance, weather conditions, humidity, visibility) can be displayed and managed.

Later in the next stage, modules will be developed in the direction of prevention and control of the application through voice commands, a stage in which hardware changes are needed for greater processing power.

Fig. 1: Portable touchscreen system used for the pedestrian detection.

The purpose of these tests is to analyze the behavior of pedestrians in the external environment, to analyze the data received by the sensors, the real-time predictability of pedestrian movements and to try to prevent the loss of human lives. The presented scenarios will have real elements and capture made in the traffic from Suceava in the immediate vicinity of the University Campus of the Stefan Cel Mare University Suceava in December, in favorable weather conditions from all points of view. The current system is in an initial phase in which it is tried to demonstrate the utility of these applications in the direct fight with the main causes for the loss of human lives and of the low level of prevention in the automobile field and of the lack of pedestrians. Thus, putting the human factor in the foreground, and in this case, the pedestrian. Because most do not have proper conduct and carry out various activities that distract them, making them vulnerable.

B. Debate on the detection and predictability pedestrian setup

To manage the problem more efficiently, we used TensorFlow because it offers an exposition of the algorithms of automatic learning and fast implementation that can run on devices with much lower hardware resources, obtaining ideal results. TensorFlow offers flexibility and variety in terms of developed algorithms, processing times and modeling using neural networks, thus improving runtime. The process of detection and analysis in the pedestrian identification stage must be carried out in a short time because the main object is the human factor, completely unpredictable in reactions and movements. For achieving remarkable results the classic Raspberry dedicated camera has been replaced with a dedicated dashboard camera, in the developmental configuration with the mention that we use Raspberry 3, see Figure 1. For a more efficient setup in parallel, a similar system is developed on a configuration. Replacing the Raspberry 4 controller, higher configuration, benefiting from 4GB RAM and a 1.5 GHz processor, which together with a USB Neuronal Stick, will increase system capabilities.

The usefulness of the TensorFlow Detection Object API is that this environment provides detailed training and implementation for creating detection models. Utilization of SSD (Single Shot Detector) and Faster R-CNN (Region Convolutional Neural Network) architectures, as well as extractors such as MobileNet or Inception, supplement the entire detection process and increase processing speed. For better accuracy, COCO SSD should be used, although it offers fast processing, the accuracy of this algorithm does not provide the expected yield [1]-[14]. In the first phase, a mix between the models and algorithms analyzed causes the presented setup to provide satisfactory results according to the hardware architecture and the required requirements. Thus, relying more on MobileNetV2 and YoLo V3 architecture in the analysis of several blocks, for better accuracy [15],[16].

It can be observed in the representation of Fig.2 the data collection from the reference builders and destructors according to the model. For each configuration, a model contains a selection of meta-architecture type, with characteristic extractor (RESNET and INCEPTION RESNET), but also the input resolutions or the number of

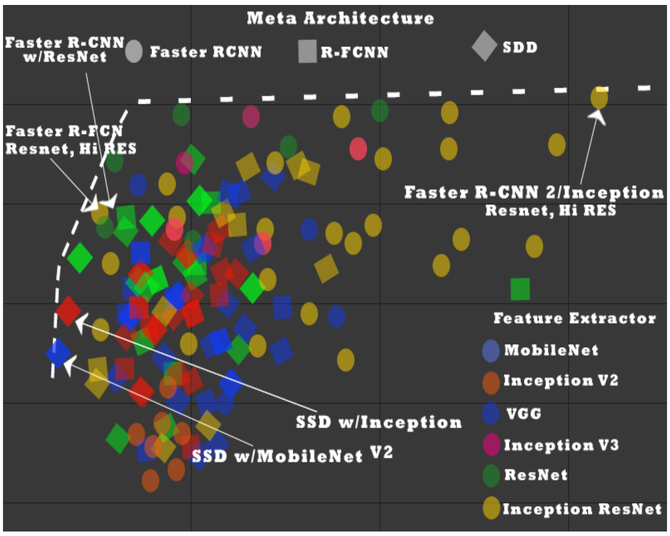

Fig. 2: Graphic display of accuracy vs time, an indication of meta-architecture, and extractor by colored symbols.

propositions for speeding the R-CNN and R-FCN. Toward obtaining results according to the memory request and the path parameters, we resort to cascade operations, so we can provide the time required for a complete journey using a subset of data and an experimental configuration [17],[18].

The diagram shown in Figure 2, represents the dispersion for all the configurations of each model, the representation of the feature extractors and the marker forms that include a meta-architecture. To run an image, it records times of even tens of milliseconds, depending on the number of objects and the complexity of the movements.

We can be said that in general, we see R-FCN, SSD models are most faster as opposed to R-CNN, which tends to get slower models, but much better accuracy, with at least 100 ms being required for each processed image. If we have a single region and a single object to track, then processing times increase, but this is subject to the limitation of the proposed regions. In a first version, the speed of detection is sacrificed more precisely the speed, just to increase the precision, so that the restrictive imaginary threshold of detection can be exceeded. All this process, as well as the architecture, is in a testing phase where we try to develop a solution that will satisfy any requirement in real-time, especially since we can have disruptive factors that can influence the detection model. From the studies performed so far a solution that offers both speed and accuracy comes from the YoLo V3 algorithm. The realized and improved model offers several advantages as opposed to other systems. We can say that it analyzes the whole image during the test, so we can have a prediction through the global contextual analysis of the image. The time response for a prediction is much better than an R-CNN model, which requires thousands of trips for a single image. According to the Darknet, this algorithm is characterized by a speed of approximately 100x compared to R-CNN Fast and approximately 1000x R-CNN [18],[19],[20].

III. Debate Software Architecture and Experimental Results

A. *Debate on the software architecture*

For testing, several detection models have been built strictly for pedestrian analysis and predictability, classifying them according to how they advance in image/video, outlining each type of pedestrian in a transparent red border for those who endanger and they put themselves in danger, and in a green box those who are outside the area of the vehicle and do not expose or endanger the safety of traffic, but also their own lives [19],[20].

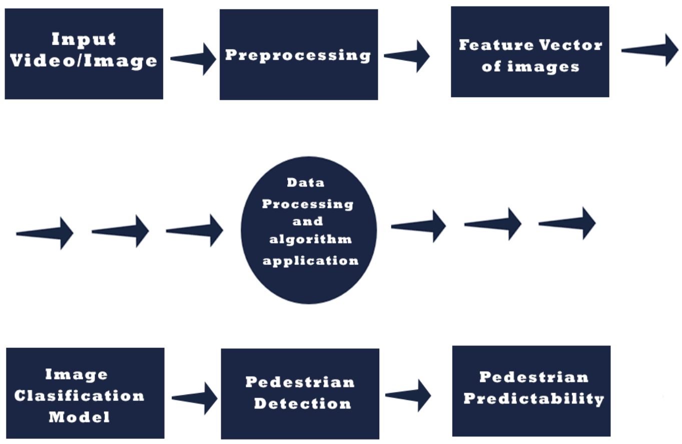

Fig. 3: Architecture of the algorithms applied in image / video processing

The efficiency of the system setup proposed a detailed analysis for all the stages, the number of pedestrians on the respective route, the number of detections, the accuracy, even the problems encountered or elements that do not satisfy the first analysis is also presented. Detection is done by delimiting the image through several boxes using logical regression.

Thus, the points on height and width are analyzed, thus calculating the center of each box has a prediction point. The use of independent logic classifiers offers a high degree of accuracy and a much better efficiency with regard to the prediction class.

The complexity of moving to the parameterized data set is simpler and the overlay analysis and pedestrian detection are performed at the same time. By analyzing the characteristics of two layers overlap/detection we construct a map of the functions that are then combined using the concatenation. Using this method, you can obtain semantic information that samples fine-grained information, see the architecture algorithms for processing images in Figure 3.

B. Experimental evaluation and results

The test plan was made a route for a total distance of about 4,4 km (see the Figure 4) in a timetable established according to the level of traffic, but also of pedestrians, so the simulation took place on a Saturday in the interval 10:00-12:00 Am, for the existence of a flow of pedestrians and traffic participants. The level of congestion in the case of traffic was a normal one without long interruptions that would disturb the test. At this distance, several cases were identified in which not only pedestrian detection was necessary, but also a model for analyzing and predicting cars that could endanger other traffic participants. In the case of the first scenario, the detection stages are performed step by step, taking care to respect the speed of movement, gradually increasing, according to the obtained performances, analyzing the accuracy of the detection according to these characteristics. The first part of the tests benefited of ideal weather conditions, which allowed recording of clear frames, without sunlight or reflections on the asphalt, bonnet, or windshield. Scenario 1 in the first part is satisfactory in relation to the real events, obtaining the

Fig. 5: Scenario 1 – Pedestrians identified under ideal conditions

Fig. 6: Scenario 2 - Unidentified pedestrians

detection of pedestrians who have put in danger and being totally inattentive. The first part of this scenario is carried out on a secondary alley, poorly circulated. The moment when the car travels on the main road a first disturbing factor is a sunlight, falling steadily on the surface of the windscreen, the bonnet, and the road surface. This factor disrupts both the track record and the detection process, on all this time no pedestrians were identified. After returning the car on an ideal route, we observe that while the speed increases the analysis and the detection of pedestrians are delayed, the speed of travel was between 55km / h and 62km / h. In some cases, the detection has been done, but the identification in the image appears to delay, this problem is due to the hardware component.

The problems encountered can be remedied, both by the hardware upgrade and the use of several cameras and filters,

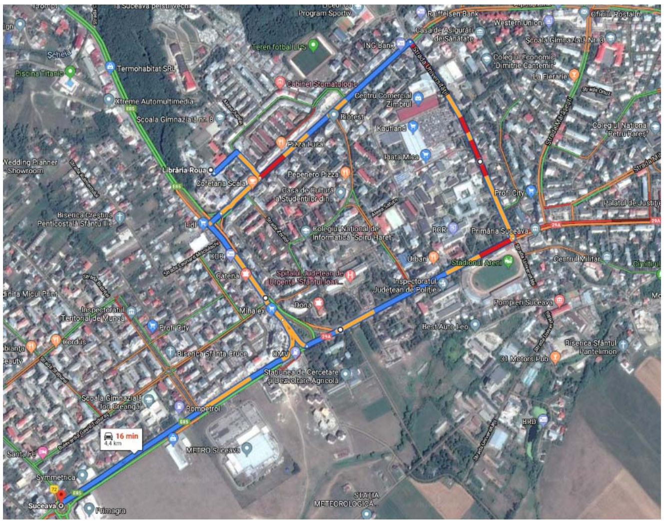

Fig. 4: The route followed in carrying out the tests, the level of traffic congestion and the distance traveled.

positioned in the front area of the car, with the grid. Because between the radiator and the front grille, there is an area away from strong outside light, so one of the rooms can be mounted in that area. The use of more cameras both inside and outside the machine would automatically increase the degree of accuracy, having more sources of analysis and comparison.

In a secondary scenario, traffic jams and congestion zones were taken into account, where most pedestrians force crossing the streets without being assured in both directions, or through unauthorized places, making it difficult for traffic participants to move safely.

We can say that in this scenario the speed of movement was on average about 20km/h, a factor that contributed positively to the detection of vulnerable pedestrians. Analyzing these scenarios we draw some conclusions on pedestrian behavior, but also on the development system. Under ideal conditions the entire process runs according to expectations, medium level traffic, ideal weather factor, without strong light, constant travel speed 5km/h - 40km/h.

If an image processing and filtering algorithm are used to transform the whole frame into a gray image, eliminating colors, then the degree of perception and analysis increases, because the use of HoG of an RGB image gives better performance due to the fact that the detection algorithm no longer exists. It is necessary to analyze and cross the gradient several times, as a linear operator.

The accuracy in the case of pedestrian detection may decrease due to the increase in the total number of cadres per second, this can also be seen in the table below. The total number of detecting but also the incorrectly exposed frames, which are not pedestrian, is analyzed.

Following the analysis of these data obtained from practical simulations, it can be seen that the 2 scenarios have highlighted some of the advantages of the developed system, but also future directions on which we must focus all our attention. Algorithms for image detection and analysis will have to outline and delimit the pedestrian area from the one intended for road traffic, so the recognition process would be much simpler and any pedestrian who leaves the pedestrian grades space is automatically seen as a possible warning.

We can say that the results obtained are due to the detection of single objects according to customized models so the accuracy and response time was one to match. Yolo v3 classification has been improved for browsing, an image and detecting over 20 different classes of objects for the interior, here we are talking about static objects.

TABLE I. DETECTION ACCURACY DETAILS,TOTAL BLOCKS DECODED,FRAMES AND PEDESTRIAS DETECTED

Video duration in seconds	Total number of blocks decoded	Frames detected	Frames lost	Total pedestrians	Pedestrians detected	Detection accuracy %
302	**16954**	**8242**	**0**	**41**	**36**	**88%**
180	**10562**	**5264**	**4**	**38**	**31**	**82%**
130	**7803**	**3876**	**13**	**28**	**18**	**76 %**
130	**10942**	**5461**	**1**	**19**	**18**	**95 %**

According to the literature and simulations, the Yolo v3 version is clearly superior to v2, significantly reducing the errors of localization, reprocessing, increasing the accuracy of the classification, by overlapping the convolutionary layers, results were obtained by approximately 7% more conclusive.

The major advantage was that Yolo v3 uses independent classifiers for the analyzed classes and the losses are of the binary type of cross-entropy in order to predict the class. Using such classifiers that are logistical and independent, an object created in the model can be detected in a wide range just as a man can be a person at the same time being an advantage in such applications but decreasing the accuracy of the detected object. All these elements led to the choice of this detection network capable of analyzing over 80 different categories of built objects. The speed is much higher and promises more using accelerators only as shown above, in the case of accuracy there are points where substantial improvements are needed.

IV. Conclusion

According to the table below, we analyze the system performance and the elements where there is the possibility of improving the system. Each scenario was analyzed and the variations for each of them can be observed, clearly presenting the performances according to the evaluated parameters. The proposed goal of detecting pedestrians using a Yolov3 with markers by overlapping pre-existing images and patterns has been achieved.

Future directions are to increase computing power through the use of a high-performance microcontroller, but also the use of accelerators for processing, whether we are talking about Google Coral Edge or Intel Neural Compute. These elements being automatically improved by installing at least 6 cameras in the front area of the car for a multi-point analysis, thus increasing the accuracy of the results.

The algorithm used is a viable solution at this time with the objection that it can be adjusted as needed, but especially to increase performance and increase the detection area and type of detection (cars, cyclists, animals, imperfect road parts, obstacles). The final obtained system is a prototype within CARSAFE project, which has the goal to obtain a hybrid visible light communication and augmented reality platform for development of smart driver assistance and active vehicle safety systems.

References

[1] Badler N, Smoliar S (1979) Digital representations of human movement. ACM Comput Surv 11(1):19–38

[2] Gavrila DM (1999) The visual analysis of human movement: a survey. CVIU 73(1):82–98

[3] Y. Zhou, G. Wang, G. Xu and G. Fu, "Safety driving assistance system design in intelligent vehicles," *2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014)*, Bali, 2014, pp. 2637-2642. doi: 10.1109/ROBIO.2014.7090740

[4] Gavrila DM, Philomin V (1999) Real-time object detection for smart vehicles. In: CVPR, Fort Collins, Colorado, USA, pp 87–93

[5] S. Jamal, H. Zeid, M. Malli and E. Yaacoub, "Safe driving: A mobile application for detecting traffic accidents," *2018 IEEE Middle East and North Africa Communications Conference (MENACOMM)*, Jounieh, 2018,pp.1-6. doi: 10.1109/MENACOMM.2018.8371000

[6] L. Jöckel, M. Kläs and S. Martínez-Fernández, "Safe Traffic Sign Recognition through Data Augmentation for Autonomous Vehicles Software," *2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, Sofia, Bulgaria, 2019,pp.540-541. doi: 10.1109/QRS-C.2019.00114

[7] C. Bao, C. Chen, H. Kui and X. Wang, "Safe Driving at Traffic Lights: An Image Recognition Based Approach," *2019 20th IEEE International Conference on Mobile Data Management (MDM)*, Hong Kong, Hong Kong,2019,pp.112-117. doi: 10.1109/MDM.2019.00-67

[8] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei and H. Zou, "Fast multiclass object detection in optical remote sensing images using region based convolutional neural networks," *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, TX, 2017, pp. 858-861.doi: 10.1109/IGARSS.2017.8127088

[9] Wu, G. Jin, M. Gao, Z. HE and Y. Yang, "Helmet Detection Based On Improved YOLO V3 Deep Model," *2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC)*, Banff, AB, Canada,2019,pp.363-368.doi:10.1109/ICNSC.2019.8743246

[10] J. Hu, X. Gao, H. Wu and S. Gao, "Detection of Workers Without the Helments in Videos Based on YOLO V3," 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, 2019, pp. 1-4.

[11] H. Yang et al., "Tender Tea Shoots Recognition and Positioning for Picking Robot Using Improved YOLO-V3 Model," in IEEE Access, vol. 7, pp. 180998-181011, 2019.

[12] E. Chen, X. Tang and B. Fu, "A Modified Pedestrian Retrieval Method Based on Faster R-CNN with Integration of Pedestrian Detection and Re-Identification," 2018 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, 2018, pp. 63-66.

[13] Y. Chen, S. Jhong, G. Li and P. Chen, "Thermal-Based Pedestrian Detection Using Faster R-CNN and Region Decomposition Branch," 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Taipei, Taiwan, 2019, pp. 1-2.

[14] P. Vasuki and S. Veluchamy, "Pedestrian detection for driver assistance systems," 2016 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, 2016, pp. 1-4.

[15] H. Shin, "Abnormal Detection based on User Feedback for Abstracted Pedestrian Video," 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea (South), 2019, pp. 1036-1038.

[16] D. Kim, S. Park, D. Kang and J. Paik, "Improved Center and Scale Prediction-Based Pedestrian Detection Using Convolutional Block," 2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin), Berlin, Germany, 2019, pp. 418-419.

[17] J. Wang, H. Li, S. Yin and Y. Sun, "Research on Improved Pedestrian Detection Algorithm Based on Convolutional Neural Network," 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 2019, pp. 254-258.

[18] G. R. Rakate, S. R. Borhade, P. S. Jadhav and M. S. Shah, "Advanced Pedestrian Detection system using combination of Haar-like features, Adaboost algorithm and Edgelet-Shapelet," 2012 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, 2012, pp. 1-5.

[19] X. Wang, H. Hu and Y. Zhang, "Pedestrian Detection Based on Spatial Attention Module for Outdoor Video Surveillance," *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, Singapore, Singapore, 2019, pp. 247-251.

[20] J. K. Park, K. S. Han and S. Y. Yun, "Intensity classification background model based on the tracing scheme for deep learning based CCTV pedestrian detection," *2018 IEEE 9th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT)*, Cape Town, 2018, pp. 224-228.