International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.9756/INTJECSE/V14I5.395 ISSN: 1308-5581 Vol 14, Issue 05 2022

NOVEL METHOD FOR OBJECT DETECTION IN AUTONOMOUS DRIVINGSYSTEM USING CSPResNeXt AND YOLO-V4

[1]S. Geetha Priya Assistant Professor,
Computer Science and Engineering
R.M.D Engineering College, Kavaraipettai
Email Id: sgp.cse@rmd.ac.in,
ORCID : 0000-0002-8740-2103

[2]J. Rajalakshmi
Assistant Professor, Computer Science and Engineering
S.A Engineering College, ChennaiEmail Id: rajalakshmij@saec.ac.in

[3]G.Belshia Jebamalar
Assistant Professor, Computer Science and Engineering,
S.A Engineering College, ChennaiEmail Id: belshia@saec.ac.in

Abstract- *A vehicle's ability to run safely at high speeds requires the detection of objects accurately with real-time detection on the road. To certify a vehicle's safety at high speeds, visible objects on the road must beaccurately detected in real-time. The proposed model is built using the YOLO v4 structure with the alteration in the backbone of the network. The backbone of the YOLO v4 model is CSPDarknet, which is replaced with CSPResNeXt for acquiring optimal speed and accuracy rate for detecting the object. The SPP and PAN together are taken as the neck and YOLO v3 is taken as the head of the network structure. The model has been developed with an alteration in the first part of the network, with CSPResNeXt in the YOLOv4 model, which does feature extraction and classification respectively. The model has been compared with existing models like Faster R-CNN, SSD and Mask R-CNN, and YOLO v2. Compared with these models, the proposed model provides optimal speed with better image resolution, high mAP values with less loss function.*

Keywords*: mAP - Mean Average Precision, YOLO - You Look Only Once, R-CNN - Recurrent - Convolutional Neural Network, SSD - Single-Shot Detector, ResNet - Residual Network, SPP - Spatial Pyramid Pooling, PAN - Path Aggregation Network*

I. INTRODUCTION

Artificial intelligence is a mimic of the human brain which will transfigure every aspect of our lives, including our workplaces, homes, and automobiles. Speaking about automobiles, many companies are currently working towards autonomous vehicles. Autonomous vehicles can identify their surroundings, i.e., their tracks and obstacles, and commute to their destination with the help of a combination of sensors, cameras, and radars. The objective of autonomous vehicles is to both localize themselves in an environment and keep track of objects. The process behind all these objectives starts with detecting objects that reside before the vehicle.

Object detection is the field associated with computer vision that locates the object from a given image and the identified object will be marked using a boundary box. Object detections are used in applications like pedestrian detection, face detection, security purposes, and much more. Many classic algorithms are there for detecting objects, but as technology develops these algorithms are facing issues like accuracy and speed. After the breakthrough of deep learning, many algorithms have come into existence with better accuracy and speed. Some of the algorithms like SSD, Faster R-CNN, and YOLO are much higher in their accuracy rate. The evolution of object detection has so far improved from the traditional methods, which minimizes the number of evaluation steps in the current methodologies and is also more efficient. In the traditional method, a fixed sliding window is usedto slide from left to right and top to bottom in the image to localize the object in the image at different locations. The methodology for detecting an object is likely to be divided into three stages - Region of interest, extracting the features of the object, and classification. An input image is taken and divided into multiple regions. Each region is considered a separate image and the region of interest is set to each image. In the next stage, the Features of the objects are extracted using algorithms like HOG, ResNet, etc. The final stage ofobject detection is the classification where the object is detected using a boundary box and labeled with its class name. Some of the classification algorithms are CNN, R-CNN, YOLO, etc. In our proposed system, feature extraction is done using ResNeXt, and classification is done using YOLO.

A. YOLO-Series :

The YOLO series is a deep learning technology with a regression method. YOLO has four different types in it, each with better improvements. The difference between YOLO and other models is, using other models for object detection, the input image has to be viewed each time for each process, like finding the region of interest, feature extraction, and classification. But, in YOLO the image can be viewed only once, and all the processes are done. Also, in the YOLO model, the region of interest part is done along with the feature extraction and not done separately. YOLO v1 is the basic model of YOLO and is known for its speed at the time of its launch. Though the model is simple and fast, the accuracy tends to degrade with small objects. So, the model is not good enough for dealing with small object datasets. The next model in YOLO is YOLO v2 which is better than YOLO v1 in aspects like speed,

accuracy, and recognition of more objects, around 9000 different objects and so named YOLO9000. Also, this model overcomes the drawback of YOLO v1 in the case of handling smaller objects. Thus, YOLOv2 is uniquely known for its high accuracy rate in handling smaller objects. The next version of YOLO is YOLO v3, which is more complicated than the previous model, but by changing the model structure accuracy and speed can be achieved. The latest version of YOLO is YOLO v4 which has the optimal speed and accuracy rate. The backbone of YOLO v4 is CSPDarknet53 and uses YOLO v3 as the head. As compared to all other YOLO models, this model provides the best results for object detection.

In our proposed system, the YOLO v4 model is taken as the base network structure, with the alteration in the backbone of thenetwork. The backbone of YOLO v4 is altered with CSPResNeXt for achieving better speed and accuracy. The detailed work of the proposed system is explained below.

II. RELATED WORKS

In recent years, many models have been developed for object detection techniques and specifically, many companies are working on autonomous vehicles system for the past few years. Many of the existing systems work as a two-phase algorithm for the detection of objects where the region of proposal is done as a separate module. And in models like YOLO the region of interest module is not done separately but as a part of feature extraction itself. Many models tend to use a single algorithm/ model for both extractions of objects and classification, and some models tend to embed two or more algorithms to perform each process for separate modules. The existing systems related to YOLO [3], [8] model are the wide rangesused for autonomous driving systems and related applications. YOLO performs well in speed but degrades the accuracy when dealing with smaller objects. The other YOLO model like YOLO v2
[9] improves the time computation and speed compared to the previous model.

Many more models with two-stage detections that are used for detecting objects are discussed further here. In that list, HOG isa basic method that provides an accurate result for object detection. And along with HOG, Haar-like [18] algorithm and CSS [12] are used for extracting the features of the object to be detected. The other methods for extracting features are, Selective Search [15] which includes the segmentation and feature of exhaustive search. It groups similar regions using color, shape, and texture. Transfer learning [11] is used for both extracting features and classification. Transfer learning is a method, where a model which is trained for some other relative problem statement is taken for solving the current problem statement. This methodology helps in building the network easily without developing from scratch. Some of the pre- trained models for training the network are ImageNet, ResNet, etc. When building the system with deep learning, the same methods are mostly used for both feature extraction and classification. Fast R- CNN [13] and YOLO [9], [13] are those types, used for both featureextraction and classification. The methods for Classifications are SVM [9], [12], [18], AdaBoost [18], and K-means clustering. These are the most common methods used in a certain period. The idea of the SVM classifier algorithm is simple; it creates a hyperplane that separates two classifiers. AdaBoost or adaptive boosting uses the ensemble learning technique for classification. This combines multiple week classifiers to make it a strong classifier. All the above-discussed methods and algorithms work well with some perspectives, but lag with real-time accuracy and speed. Many algorithms are still been developed for achieving these goals.

Finally, a collaborative system using ResNext-101 and YOLO- v4 is built in our proposed system for optimal speed and accuracy. The methodology of the whole model is discussed in detail in the below sections.

III. PROPOSED WORK

In our proposed work, a neural network is developed using the YOLO v4 structure with an alteration in the backbone of the YOLO v4 model. The actual backbone of YOLO v4 is CSPDarknet, which is replaced with CSPResNeXt in our model to improve the speed and accuracy of the system. The network structure of YOLO v4 has a backbone, a neck, and a head part. In our proposed system, the CSPResNeXt is taken as the backbone of the network, Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PAN) together is the neck and YOLO v3 is the head of the network. The choice of CSPResNeXt is to improve the accuracy of the system by handling the degradation problem in deep networks. The ResNet series is probably used for its better speed and accuracy rate when working with deeper networks. The basic structure of ResNet itself can avoid degradation problems. Degradation problems occur in deep networks where the accuracy of the system saturates at some higher layer of the networks, because of repeating some of the processes at the higher layer more than twice. To avoid this degradation problem, a ResNet series is used, where the model skipsa particular process if it is sequentially repeated twice. Compared with ResNet, ResNeXt offers a parallel stacking layer and adds multi-path into ResNet.

A. SYSTEM DESIGN

The overall architecture of our proposed system is shown in figure
1. The system is built using the neural network structure with abackbone, neck, and head. The images are given as the input to the

system, the first phase of the network is the backbone of the system, which is ResNeXt where the extraction of the feature takes place. The choice of ResNeXt 101 for feature extraction is explained in section 2. Along with this, the ReLU activation function is used by the neural network for converting real-world non-linear data to linear data for further processing. The next phase of the network is the neck, with SPP and PAN network. The pooling process happens in this phase where the exact and useful features of the images are considered or pooled for the next phase. SPP and PAN networks have been explained in section 3. The last phase of the network is the head, where YOLO v3 is used as the head for the network. In this phase the classification and detection of objects take place. The images have been classified and detected using the bounding boxes and give the probability for each label.

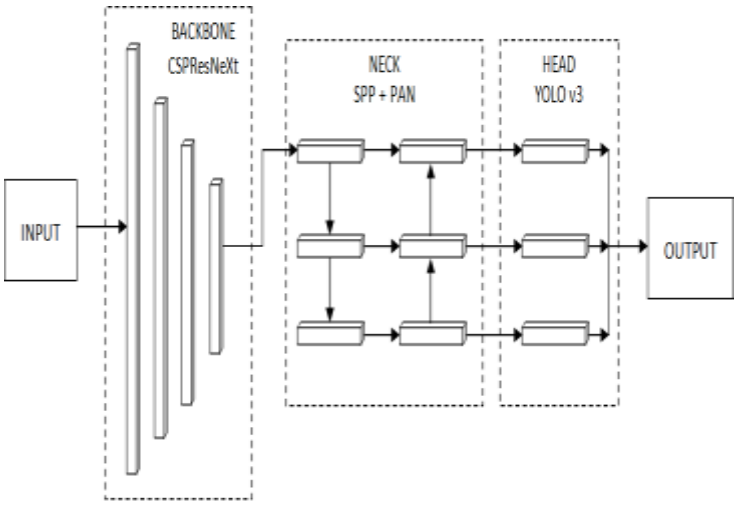

Figure 1 : System Architecture

B. BACKBONE

Cross Stage Partial ResNeXt (CSPResNeXt) is considered the backbone of the YOLO v4 network, which takes care of the feature extraction process. Cross-Stage Partial Networks have been used to avoid duplicate gradient information within network optimization, reducing complexity while maintaining accuracy. CSP can be applied to models like ResNet, ResNeXt, and DenseNet. Applying CSP to these models results in a reduction in the computation effort and improvement in terms of accuracy. In our system, we have taken ResNeXt for working on our system, the working way of CSPResNeXt is shown in Figure 1.

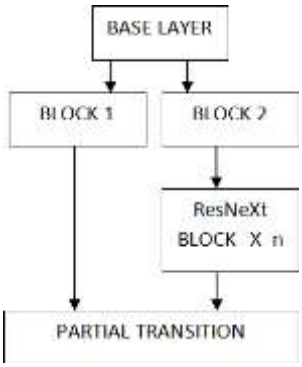

Figure 2 : CSPResNeXt

ResNeXt: The ResNeXt architecture is an enhancement to the deep residual network. Deep Residual Network uses the standard residual block which is altered in ResNeXt as a " split - transform - merge" strategy. Convolutions over input features are not performed over the whole input feature map; instead, the block's input is projected into a series of channels with lower-dimensional representations to which we apply a series of convolution filters before merging them to get the results. Separating the inputs into several channels or separate groups tend to focus on different characteristics feature of the input image. The number of channels or paths in ResNeXt is known as Cardinality. In ResNet, it has high depth and width, whereas ResNeXt has high cardinality which helps in reducing the validation error. The difference between ResNet and ResNeXt architectures is shown in Figure 3. In the example shown in Figure 1 with 128-d, ResNet in Layer -1 has one convolution layer with 64 widths, whereas ResNeXt has 8 different convolution layers with 16widths.

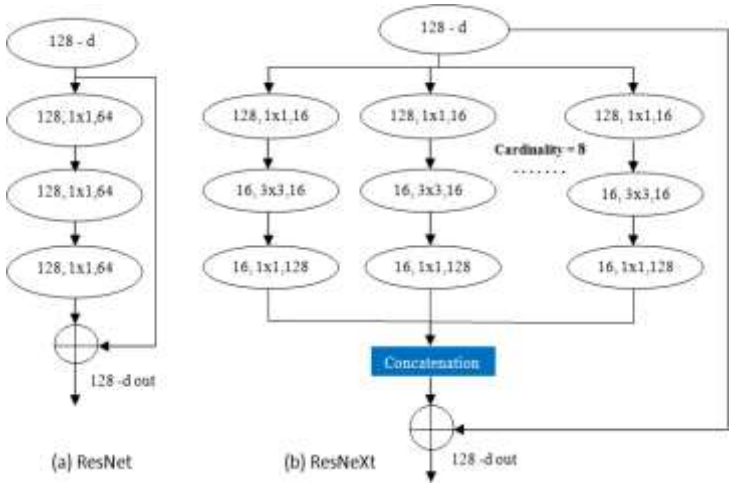

Figure 3 : Work flow of ResNet (a) and ResNeXt (b)

C. NECK

The role of the neck in the network structure is to collect the feature maps of the object in different stages. SPP and PAN are taken along with CSPResNeXt and considered the neck of our network structure. Spatial pyramid pooling is a technique for pooling on multiple kernel sizes instead of a fixed size. Usually, the pooling method resides on the size of the input, but in SPP, irrespective of the input size, the pooling method works. So, SPP can be used in deep networks with different scale sizes which shows better accuracy in classification and detection tasks. Path Aggregation (PAN) is another block present as the neck of the YOLOv4 for the feature aggregation of the network. PAN concentrates on preserving the spatial data and enhancing the instance segmentation process. In models like Mask-RCNN and other related models, the pooling process is done at one single stage, where features travel through multiple layers before the pooling process happens. Due to this, the spatial resolution of the image decreases and increases the complexity of the feature. To overcome this drawback, the PAN network uses the feature from all the layers and lets the network select the useful one.

D. HEAD

YOLO v3 has been considered the head of the YOLO v4 network. YOLO (You Only Look Once) is a deep convolution neural network that performs object detection, as the name suggests; the network uses 1x1 convolution for prediction. This part of the network does the classification and detection part. YOLO uses score regions for predicting objects, the highest score regions are noted as positive detections. YOLO algorithms are faster than any other algorithms, in that it looks at the whole image once while testing and give the predictions and still get its accuracy. The network divides the input image into regions and predicts using the boundingboxes, each detected object is marked using a bounding box. To measure the accuracy rate, bounding boxes will have their score rate. The main advantage of YOLO is, it looks at the image only once and predicts the image, this makes the network extremely fast. Compared to other algorithms, YOLO exceeds in speed, accuracy, and precision. Also, for working with small object detections, the YOLO algorithm provides accurate detection and real-time speed.

IV. RESULT AND DISCUSSION

The proposed system has been experimented with using the Udacity Self Driving Car Dataset as the training sample. The dataset contains 97,942 labels with 15,000 images and 11 classes like vehicles, pedestrians, traffic signals, etc. 1,720 unsupervised examples can be used for both training and testing purposes. From the dataset, 80% of the data is used for training the network and 20% of the data is used for testing the network. The images given to the network are set to the feature extraction process which is done using the backbone of the network,

CSPResNeXt. At the next level, the pooling process happens where the whole image is reduced to the exact data of what is needed for the further process; i.e., the unwanted data from the image is removed in the pooling process. After this, the extracted features are used in the next phase for classification and detection, which is performed by YOLO-v3. The detected object is marked using the bounding boxes. For the classification and detection process, a confidence score has been used to measure how well the object is marked using the bounding boxes. The Confidence score is measured using the following values.

Y = pc, bx, by, bh, bw, c1, c2, c3pc - either object present or not.

bx, by, bh, bw - Coordinates related to the bounding box.

c1, c2, c3 - Numerical values that represent the class of the object.

By experimenting the model with Udacity Self Driving Car Dataset, our model provides an accuracy of 90.6% and less loss rate as compared with some of the existing models. The model has been compared with many other existing models like SSN. Mask- RCNN, Faster R-CNN, and YOLO v2 where each model faces drawbacks like less accuracy rate, loss function, reduction of resolution of the image. Compared with these models, our proposed work shows a better accuracy rate and better spatial resolution of the image with less loss function and also, gives a better accuracy rate in detecting smaller objects too. The compared mAP data has been plotted in the chart and shown in figure 4 and the chart for loss function is shown in figure 5.

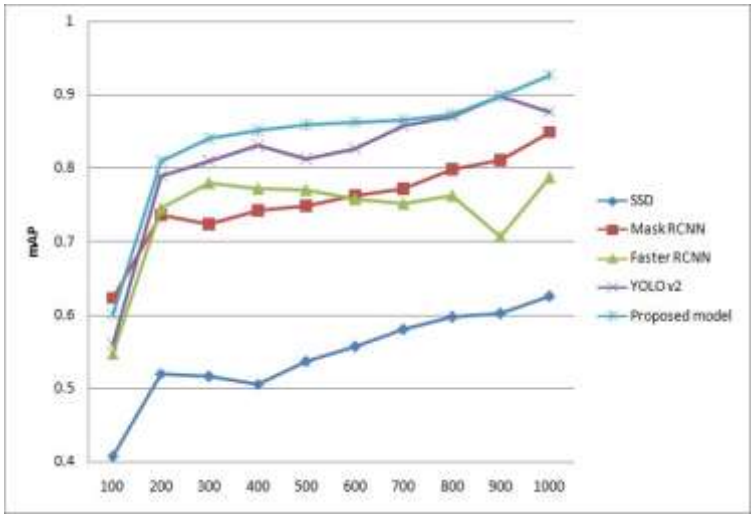

Figure 4 : Comparison of mAP values

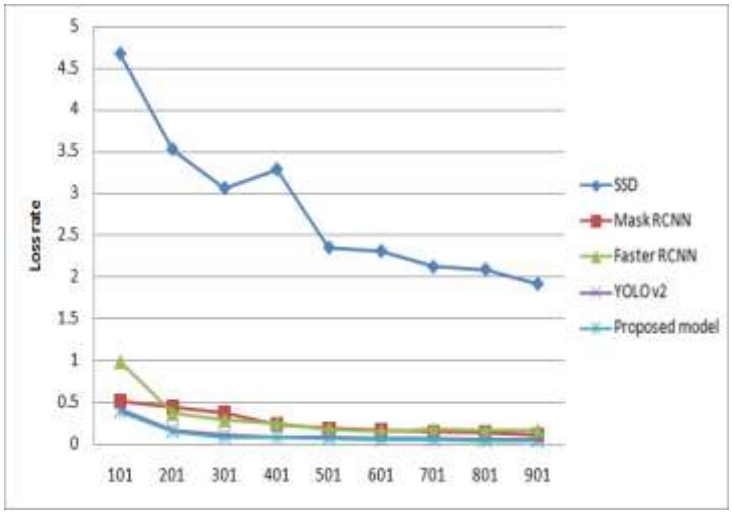

Figure 5 : Comparison of loss values

Thus, the proposed model using CSPResNeXt as the backbone for YOLO v4 and experimented with Udacity Self Driving car Dataset and acquired the accuracy rate of 90.6% with a less loss function. Also, compared with existing models, it proves as a better model for object detection in autonomous driving systems. The output of the proposed system is shown in figure 6, which shows the detected object marked with the bounding boxes with the percentage of class prediction.

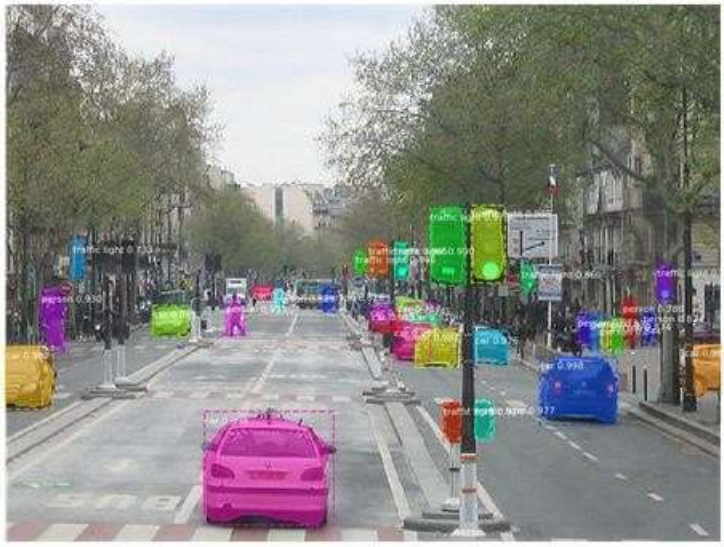

Figure 6 : Result

CONCLUSION

Object detection technology is the most significant task in autonomous driving systems. The success of an autonomous driving system relies on theaccurate detection of an object at the real-time speed of the vehicle. Aimingfor the same, our proposed model has been built using the YOLO v4 model with the alteration in the backbone of the structure, where CSPResNeXt is used instead of DarkNet53. The purpose of the alteration was to avoid degradation problems with deeper networks. By using CSPResNeXt, degradation of accuracy at the higher levels has been avoided. Also, by using the SPP and PAN pooling methods, the reduction in spatial resolution of the image has been avoided. The proposed model has been experimented with Udacity Self Driving Car Dataset and compared with existing models like SNN, Faster R-CNN, Mask R-CNN, and YOLO v2, and obtained the result with a better accuracy rate of 90.6 % with less loss function.

REFERENCE

[1] Rui Wang,1 Ziyue Wang,1 Zhengwei Xu, A Real-Time Object Detector For Autonomous Vehicles Based On Yolov4, Hindawi Computational Intelligence And Neuroscience,Volume 2021, Article ID 9218137.
[2] Chien-Yao Wang, Hong-Yuan Mark Liao," Cspnet: A New Backbone That Can Enhance Learning Capability Of Cnn″ Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition (CVPR) Workshops, 2020, Pp. 390-391
[3] Zaatouri, K., & Ezzedine, T. (2018), "A Self-Adaptive Traffic Light Control System Based On YOLO"; International Conference On Internet Of Things, Embedded Systems And Communications (IINTEC), Pp: 16- 19.
[4] Ahmed, Z., Iniyavan, R., & P, M. M. (2019), "Enhanced Vulnerable Pedestrian Detection Using Deep Learning"; International Conference OnCommunication And Signal Processing (ICCSP), Pp: 0971-0974.
[5] Ash, R., Ofri, D., Brokman, J., Friedman, I., & Moshe, Y. (2018), "Real- Time Pedestrian Traffic Light Detection"; IEEE International Conference On The Science Of Electrical Engineering In Israel.
[6] Brunetti, A., Buongiorno, D., Trotta, G. F., & Bevilacqua, V. (2018), "Computer Vision And Deep Learning Techniques For Pedestrian Detection And Tracking: A Survey"; Neurocomputing, 300, Pp: 17–33.
[7] Chen, E., Tang, X., & Fu, B. (2018), "A Modified Pedestrian Retrieval Method Based On Faster R-CNN With Integration Of Pedestrian Detection And Re-Identification"; International Conference On Audio, Language And Image Processing, Pp: 63-66.
[8] Guanqing Li, Zhiyong Song , Qiang Fu (2018), "A New Method Of Image Detection For Small Datasets Under The Framework Of YOLO Network"; IEEE 3rd Advanced Information Technology, Electronic And Automation Control Conference, Pp: 1031-1035.
[9] Sakshi Gupta, Dr. T. Uma Devi, "Yolov2 Based Real Time Object Detection" , International Journal Of Computer Science Trends And Technology (IJCST) – Volume 8 Issue 3, May-Jun 2020, Pp: 26 - 30.
[10] Lan, W., Dang, J., Wang, Y., & Wang, S. (2018), "Pedestrian Detection Based On YOLO Network Model"; IEEE International Conference On Mechatronics And Automation (ICMA), Pp: 1547-1551.
[11] Rahul Pathak,P. Sivraj, (2018), "Selection Of Algorithms For Pedestrian Detection During Day And Night"; Computational Vision And Bio Inspired Computing, Pp 120-133.
[12] Ghosh, S., Amon, P., Hutter, A., & Kaup, A. (2017), "Reliable Pedestrian Detection Using A Deep Neural Network Trained On Pedestrian Counts";IEEE International Conference On Image Processing.
[13] Hongmeng Song, Wenmin Wang (2017), "Collaborative Deep Networks For Pedestrian Detection"; IEEE Third International Conference On Multimedia Big Data.
[14] Naghavi, S. H., Avaznia, C., & Talebi, H. (2017), "Integrated Real-Time Object Detection For Self-Driving Vehicles"; 10th Iranian Conference OnMachine Vision And Image Processing.
[15] Zhang, H., Du, Y., Ning, S., Zhang, Y., Yang, S., & Du, C. (2017), "Pedestrian Detection Method Based On Faster R-CNN. 2017 13th International Conference On Computational Intelligence And Security.
[16] Hailong Li, Zhendong Wu, & Jianwu Zhang. (2016), "Pedestrian Detection Based On Deep Learning Model"; 9th International Congress On Image And Signal Processing, Biomedical Engineering And Informatics.
[17] Peng, Q., Luo, W., Hong, G., Feng, M., Xia, Y., Yu, Li, M. (2016), "Pedestrian Detection For Transformer Substation Based On Gaussian Mixture Model And YOLO"; 8th International Conference On Intelligent Human-Machine Systems And Cybernetics.
[18] Tomè, D., Monti, F., Baroffio, L., Bondi, L., Tagliasacchi, M., & Tubaro,
S. (2016), "Deep Convolutional Neural Networks For Pedestrian Detection, Signal Processing: Image Communication, Pp: 482–489.