Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1876 - 1881
Received 15 April 2021; Accepted 05 May 2021.

A Real Time Object Detection System Using a Webcam with Yolo Algorithm

V. Neethidevan[1], Dr.S. Anand[2]

[1]Assistant Professor, Department of MCA, Mepco Schlenk Engineering College (Autonomous), Sivakasi, Tamil Nadu, India.

[2]Professor, Department of ECE, Mepco Schlenk Engineering College (Autonomous), Sivakasi, Tamil Nadu, India.

ABSTRACT

The main objective of object detection is to detect various objects in a video stream with more accuracy and with less computation time. Image classification involves assigning a class label to an image, whereas object localization involves drawing a bounding box around one or more objects in an image. The object detection algorithm YOLO to examine the entire image in a single instance and predicts the bounding box coordinates and class probabilities for these boxes. This paper uses Yolo algorithm with Darkent architecture to detect all possible objects in a video stream derived from a Webcam. From the experimental study, the accuracy of object detection varies from 37% to 88%. The existing algorithms take more time to process each frame and accuracy is also less. The Intersection over Union will decide prediction of each object as Good one. The Nonmax suppression technique will get a single prediction per object. The biggest advantage of using YOLOv3 can process 67 **FPS.** The experiment was conducted in a Colab Google environment with GPU. The video taken from various places were used to examine the effectiveness of the proposed approach. The experimental results show that our proposed approach can effectively improve the accuracy rate in detection of various objects in an image.

KEYWORDS

Object Detection, Object Recognition, Computer Vision, Deep Learning, Darknet, Colab.

Introduction

Object detection is a computer vision technique that works to identify and locate objects within an image or video. It identifies and classify various objects present in an image and draws a bounding box around each objects with its detection accuracy. The object detection concepts found many applications in all day to day activities. Image recognition, after identification it assigns a label to each objects in an image. A picture of an apple receives the label “apple”. A picture of two apples, still receives the label “apple”. Object detection, on the other hand, draws a box around each apple and labels the box as “apple”. The model predicts where each object is and what label should be applied. In that way, object detection provides more information about an image than recognition. The various algorithms used for object detection are as follows. Region-based Convolutional Neural Networks (R-CNN), Fast R-CNN, Faster R-CNN, Histogram of Oriented Gradients (HOG), Region-based Fully Convolutional Network (R-FCN), Single Shot Detector (SSD), Spatial Pyramid Pooling (SPP-net), YOLO (You Only Look Once). Each algorithm is used to identify various objects in an image or in a video stream. But each algorithms has its own pros and cons. Either it takes more time to process or accuracy of finding each object is low. Also some challenges like different weather conditions like, fog, night and rainy etc. To improve further in detection accuracy and reduce the computation time, various approaches were suggested by researchers. This paper also suggest an approach to detect various objects present in a video stream by proposing a new architecture.

Literature Survey

In [1], Nguyen et al. proposed a general framework for robust on-road pedestrian and vehicle detection, recognition, and tracking. They combined the advantages of deep learning with those of using multiple local patterns and depth information, and can be operated efficiently. Tracking-based deep learning algorithm was used to track detected obstacles in the next frame. Then it is implemented using GPU for real time use. With different experiments to determine the performance, this framework yields a good detection, recognition, and tracking for real-time driving assistance. Later it could be extended to the detection and recognition of other objects like, traffic signs, traffic lights, and so forth. This paper dealt with detection, recognition, and tracking of preceding obstacles. For future work, the object behavior can be predicted. Next in [2], Deng H. Sun et al, presented region-based CNN method to detect vehicles in aerial images. They combined two CNNs: an AVPN for combining hierarchical feature maps that were more accurate for small object detection, and a VALN for attribute annotation. They conducted experiments with different data sets and they achieved good results for images from UAV or from Google Earth. But still some false detection was there. Accurate vehicle detection is still a challenging task. Future works focus on mining hard negative samples by a bootstrapping strategy and to use a multi-GPU configuration to reduce the computation time. In [3], Y. Cai, et al., proposed a scene-adaptive vehicle detection algorithm and the Bagging (Bootstrap aggregating) mechanism was used to build multiple classifiers and then voting is used to generate confidence scores for the target training samples. A composite deep-structure-based scene-adaptive classifier and its training method have then been designed using the automatic feature extraction ability of DCNN (Deep Convolutional Neural Network) and performing source-target scene feature similarity calculation with a deep auto-encoder. Experiments demonstrate that this method exhibits the advantages of a high degree of automation and a high vehicle detection rate. The limitation of this method is that the confidence assignment method is a simple linear function dependent on the sub classier members which is relatively subjective and lacks a theoretical basis. In [4], Cao et al., developed a novel framework for deep neural networks. A student network is trained to minimize joint loss designed for domain adaptation and knowledge distillation simultaneously. Besides the smaller and faster DCNN model, at the detection side, they have developed a faster method to determine the candidate regions of interest which contain the target objects and established an analytic model to compute the support regions at each convolution layer and integrated this into the existing object detection framework using deep convolution neural networks. Their experimental results on vehicle detection from videos demonstrated that the proposed method is able to speed up the network by up to 16 times while maintaining the object detection performance. In {5], Sommer et al., used Fast R-CNN and Faster R-CNN for vehicle detection in aerial images and evaluated the impact on both detection frameworks. Region Proportional Network (RPN) achieved the best performance of all proposed methods. They proposed their own nets optimized for handling small objects. The best results based on hand-crafted features are achieved for selective search. The best detection performance was achieved for Faster R-CNN that shares the convolutional layers with the RPN. Faster R-CNN provided the best results and in case of future work, examines the generalizability for various GSDs and datasets and how our adaptations can be transferred to further deep learning based detection frameworks. Furthermore, modifications of the detection frameworks are planned to explicitly handle rotated bounding boxes.

In [6], Nguyen et al., presented a real-time and robust region proposal network to detect various vehicles for various driving conditions. It improved the performance of the real-time requirement of Advanced Driver Assistant System (ADAS). For future work, system could be extended to the detection and recognition of various types of obstacles such as traffic signs and traffic lights. They also planned to integrate deep architecture VGG-16 to their proposed system with the help of an advanced pruning-based method to reduce the size of the proposed network, to increase the accuracy and feasibility of the proposed system. In [7], Zhu et al., presented a system with the following features. 1) With the help of UavCT dataset, estimation of real-world city traffic density is done 2) DVCF (deep vehicle counting framework) designed for vehicle detection, classification, tracking, and counting. It could be easily extended to detect and track many other types of objects (e.g., people, bicycles, etc.). 3) A successful attempt to integrate conventional vision-based algorithms and deep learning based approaches. In [8], Geng et al., dealt with an approach for human gesture detection in infrared images, based on the improved YOLO-V3 network. They used three DenseNet blocks to enhance the convolutional feature propagation and gesture detection performance. Many experiments were done that resulted in better detection performance and more accuracy. Under low visibility images such as rainy and foggy and night time conditions, and conditions in which the colors of the targets and the background are similar, it exhibits a high performance. In [9], Avramovic et al., proposed a region-of-interest based approach combined with YOLO architecture to achieve high performance in detecting various objects. In [10], Tourani et al., proposed License Plate Detection (LPD) and Character Recognition (CR) system for Iranian vehicle with real-time performance and high accuracy. They used, two sequential YOLO v.3 deep networks, trained the system with data acquired in various weather, noise, and illumination conditions. The precision and recall measures for the Persian character recognition stage were 0.979 and 0.991, respectively. Average processing time for each image or video frame to extract the license plate character was 119.73 milliseconds. In [11], Li et al., proposed a new approach proposed for road object detection, namely ASPP-CenterNet, which incorporated atrous spatial pyramid pooling (ASPP) with CenterNet. They also, reconstructed the backbone network for better detection performance and lightweight. With respect to CenterNet, this method improved the detection accuracy by 2.0% AP and 1.8% AP on small objects with almost the same inference time (5 FPS), and it is suitable for autonomous driving. The speciality of our system is, it focuses on improving the accuracy of object detection with less computation time.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1876 - 1881
Received 15 April 2021; Accepted 05 May 2021.

Proposed Architecture

The proposed architecture makes use of Yolov3 algorithm with Darkent implementation. It is easy to install and very fast. If the thermal camera based dataset means it could detect objects for different weather conditions. In this, proposed approach consists of Yolo's open source implementation's Darkent, Darknet was written in the C Language and CUDA technology, which makes it really fast and provides for making computations on a GPU, which is essential for real-time predictions. increases accuracy and reduces computation time to detect various objects in a image.

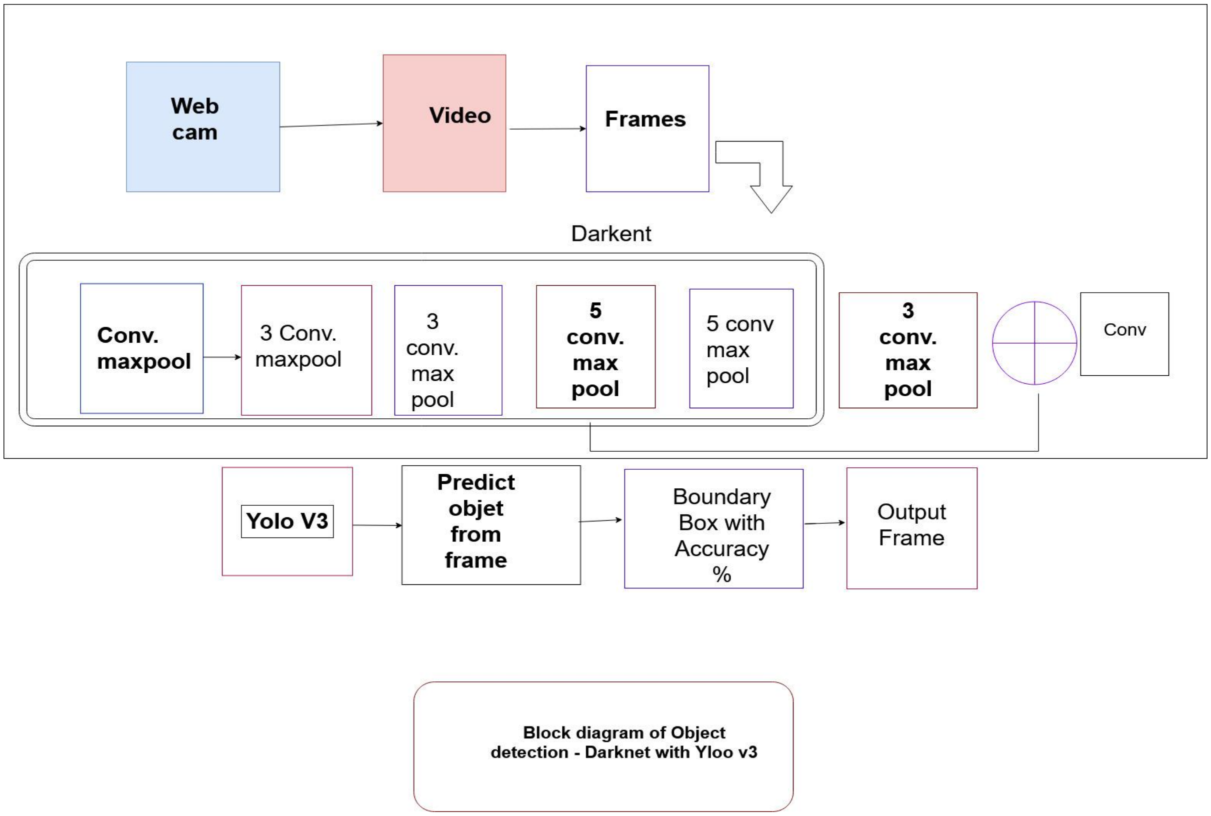


Figure 1. The proposed Architecture to detect various objects in a image

In this architecture, video stream is generated from the webcam of the laptop and then frames are extracted and each frame is given as an input to the system. The system uses **Darknet**, a open source neural network framework, which is written in C language and CUDA technology, that makes it really fast and allows us to make use of computations on a GPU, which is essential for real-time predictions. **Darknet-53** is a convolutional neural network that acts as a backbone for the YOLOv3 object detection approach. I**t** is a convolutional neural network that is 53 layers deep. We can load a pretrained version of the network trained on more than a million images from the ImageNet database [1]. Two common pooling methods are average pooling and max pooling that summarizes the average presence of a feature and the most activated presence of a feature respectively. Yolo is 1000 times faster than R-CNN algorithm and 100 times faster than Fast-CNN Yolo displays detected objects only when the confidence score >.25.

Methodology

Differentiating the cats and the dogs in an image and their locations is one kind of problem in object detection. YOLO is a powerful neural net that giving the bounding box around the detected objects. YOLO, "You Look Only Once", is a neural network capable of detecting what is in an image and where stuff is, in one pass. It gives the

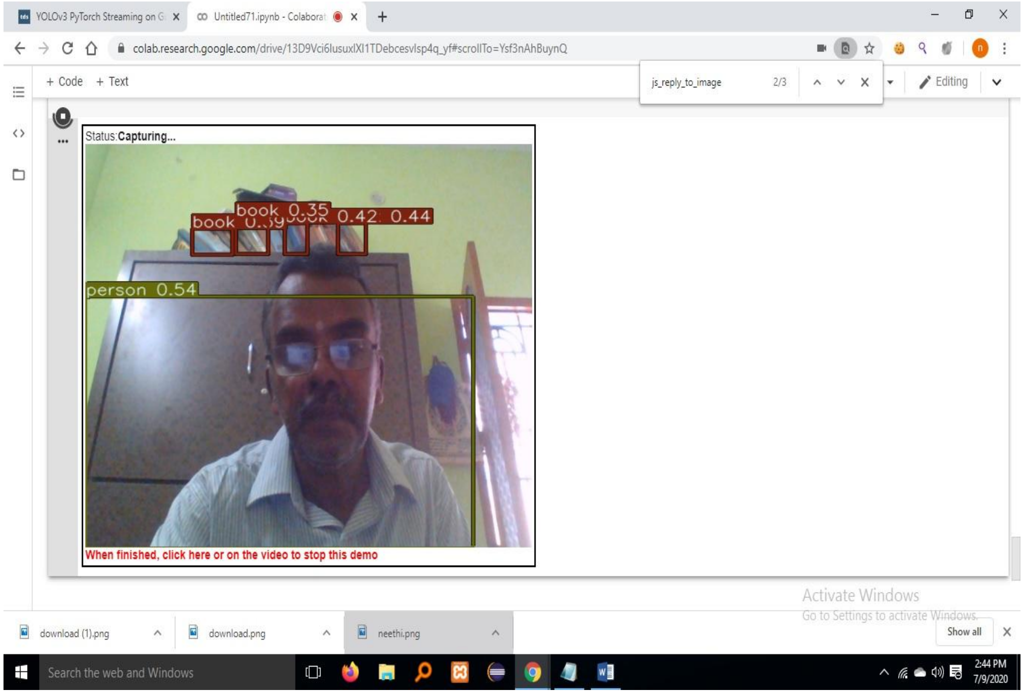
Received 15 April 2021; Accepted 05 May 2021.

bounding boxes around the detected objects, and it can detect multiple objects at a time, as shown in Figure 2. YOLO v3 Permalink adds further small improvements, included the fact that bounding boxes get predicted at different scales. It is trained to detect 1000 different objects like person, bicycle, car, motorbike, aeroplane, bus, train, truck, boat, traffic light, fire hydrant, stop sign, parking meter, bench, bird etc.

In this system, once the application starts, web camera is on and it captures video and various frames are retrieved and given to Yolo algorithm for object detection. It finds the various objects available in the image compared with trained objects earlier and finally assigns a label and draw a bounding box around each object.

Experimental Work

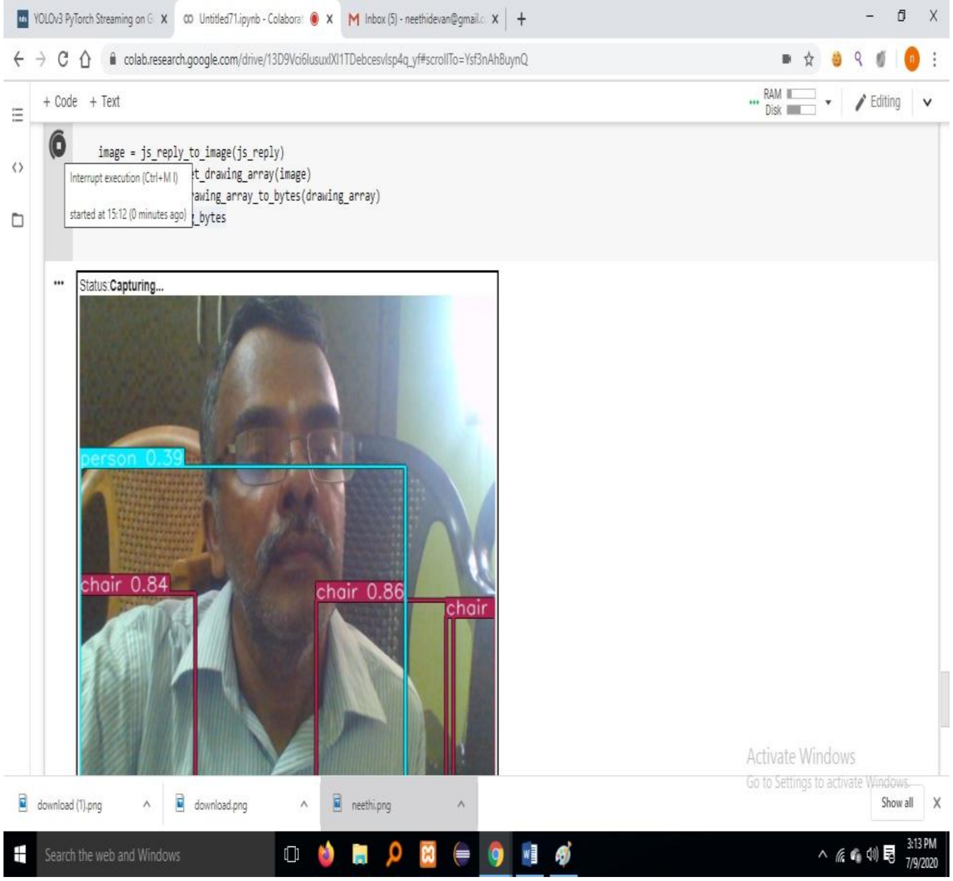
In this system, a laptop with in built webcam is used to generate a video stream. From the video stream, frames are separated and given to the Yolo algorithm. It will detect the various objects present in the image like person, bottle, chair, fan, refrigerator etc with percentage of recognition. A bounding box is drawn on each object present in the image. The implementation is done using colab environment. Video generated using webcam is extracted to get various frame. YOLO V3 is the state of the art existing system. It is an advanced Convolutional Neural Network. It predicts bounding boxes and class probabilities directly from full images in one evolution. Each frame is given as an input to the system and the proposed architecture has Darknet - 53, which is a classifier network and is the first part of the YOLO V3 architecture that helps in building, training and running Neural Network. It is trained to perform classification of image database ImageNet. It has the capability to detect and classify various objects available in the image such as dog, cat, etc. Second part of YOLO V3 analyzes the output of Darknet – 53 and searches for bounding boxes and then classifies the objects inside the bounding box. The system can detect various objects present in the image derived from the web camera, like person, various chairs with more accuracy say 86%, 84% and book objects available at some distance with less accuracy in percentage. It means that yolo algorithm could be able to identify the objects with more accuracy.


YOLOv3 PyTorch Streaming on G
Untitled71.ipynb - Colaborat
colab.research.google.com/drive/13D9Vci6lusuxlXl1TDebcesvIsp4q_yf#scrollTo=Ysf3nAhBuynQ
+ Code
+ Text
js_reply_to_image
2/3
Editing
book 0.35
book 0.
0.42
0.44
person 0.54
Activate Windows
Go to Settings to activate Windows.
download (1).png
download.png
neethi.png
Show all
Search the web and Windows
2:44 PM
7/9/2020

Status:Capturing...

When finished, click here or on the video to stop this demo

Received 15 April 2021; Accepted 05 May 2021.



```
    image = js_reply_to_image(js_reply)
           t_drawing_array(image)
           awing_array_to_bytes(drawing_array)
           _bytes
```

Interrupt execution (Ctrl+M I)

started at 15:12 (0 minutes ago)

Status:Capturing...

YOLOv3 PyTorch Streaming on G
Untitled71.ipynb - Colaborat
Inbox (5) - neethidevan@gmail.c
colab.research.google.com/drive/13D9Vci6lusuxlXI1TDebcesvlsp4q_yf#scrollTo=Ysf3nAhBuynQ
+ Code
+ Text
RAM
Disk
Editing
person 0.39
chair 0.84
chair 0.86
chair
Activate Windows
Go to Settings to activate Windows.
download (1).png
download.png
neethi.png
Show all
Search the web and Windows
3:13 PM
7/9/2020

Figure 2. Objects detected in a room

Conclusion

In this paper, we used one of the most popular implementations of Yolo algorithm with Darknet architecture to detect various objects in a video stream with less computation time and more accuracy. It could process 67 frames per second. Darkent is fast and easy to install and process. The experiment was conducted on a custom dataset captured in a office environment and it could identify the various objects present in the class room, office etc with more detection accuracy in real time. during the winter in different weather conditions (clear weather, rain, fog), during the night, and with different distance from the camera, ranging
from 30 m to 215 m.

References

[1] V. D. Nguyen, H. Van Nguyen, D. T. Tran, S. J. Lee and J. W. Jeon, "Learning Framework for Robust Obstacle Detection, Recognition, and Tracking," *In IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1633-1646, June 2017, doi: 10.1109/TITS.2016.2614818.

[2] Deng, H. Sun, S. Zhou, J. Zhao and H. Zou, "Toward Fast and Accurate Vehicle Detection in Aerial Images Using Coupled Region-Based Convolutional Neural Networks," *In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 8, pp. 3652-3664, Aug. 2017.

doi: 10.1109/ JSTARS. 2017.2694890.

[3] Y. Cai, H. Wang, Z. Zheng and X. Sun, "Scene-Adaptive Vehicle Detection Algorithm Based on a

Composite Deep Structure," *In IEEE Access*, vol. 5, pp. 22804-22811, 2017.

doi: 10.1109/ACCESS.2017.2756081.

[4] W. Cao, J. Yuan, Z. He, Z. Zhang and Z. He, "Fast Deep Neural Networks with Knowledge Guided Training and Predicted Regions of Interests for Real-Time Video Object Detection," *In IEEE Access*, vol. 6, pp. 8990-8999, 2018, doi: 10.1109/ACCESS.2018.2795798.

[5] L. Sommer, T. Schuchert and J. Beyerer, "Comprehensive Analysis of Deep Learning-Based Vehicle Detection in Aerial Images," *In IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 9, pp. 2733-2747, Sept. 2019, doi: 10.1109/TCSVT.2018.2874396.

[6] V. D. Nguyen, D. T. Tran, J. Y. Byun and J. W. Jeon, "Real-Time Vehicle Detection Using an Effective Region Proposal-Based Depth and 3-Channel Pattern," *In IEEE Transactions on Intelligent Transportation Systems,* vol. 20, no. 10, pp. 3634-3646, Oct. 2019, doi: 10.1109/TITS.2018.2877200.

[7] J. Zhu et al., "Urban Traffic Density Estimation Based on Ultrahigh-Resolution UAV Video and Deep Neural Network," *In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 11, no. 12, pp. 4968-4981, Dec. 2018, doi: 10.1109/JSTARS.2018.2879368.

[8] K. Geng and G. Yin, "Using Deep Learning in Infrared Images to Enable Human Gesture Recognition for Autonomous Vehicles," *In IEEE Access,* vol. 8, pp. 88227-88240, 2020.

doi: 10.1109/ACCESS.2020.2990636.

[9] A. Avramović, D. Sluga, D. Tabernik, D. Skočaj, V. Stojnić and N. Ilc, "Neural-Network-Based Traffic Sign Detection and Recognition in High-Definition Images Using Region Focusing and Parallelization," *In IEEE Access,* vol. 8, pp. 189855-189868, 2020, doi: 10.1109/ACCESS.2020.3031191.

[10] A. Tourani, A. Shahbahrami, S. Soroori, S. Khazaee and C. Y. Suen, "A Robust Deep Learning Approach for Automatic Iranian Vehicle License Plate Detection and Recognition for Surveillance Systems," *In IEEE Access,* vol. 8, pp. 201317-201330, 2020, doi: 10.1109/ACCESS.2020.3035992.

[11] G. Li, H. Xie, W. Yan, Y. Chang and X. Qu, "Detection of Road Objects With Small Appearance in Images for Autonomous Driving in Various Traffic Situations Using a Deep Learning Based Approach," *In IEEE Access*, vol. 8, pp. 211164-211172, 2020, doi: 10.1109/ACCESS.2020.3036620.