Journal of Engineering and Applied Research (JEAR)

National University of Skills

Summer & Autumn 2024, Vol. 1, No. 2, p. 79-90

Journal Homepage: https://jear.tvu.ac.ir/

doi:10.48301/JEAR.2024.459050.1027

Detecting, Identifying, and Counting Vehicles based on Deep Learning Algorithms in Video Surveillance Systems

Alireza Akoushideh[1*], Seyyed Shafiullah Sadat[2], Asadollah Shahbahrami[3]

[1]Assistant Professor, Electronics Engineering, Technical and Vocational University, Tehran, Iran.
[2]Faculty Member, Computer Engineering, Kabul Polytechnic University, Kabul, Afghanistan.
[3]Professor, Computer Engineering, University of Guilan, Rasht, Iran.

ARTICLE INFO

Article Type:
Original Research

Received: 05.23.2024
Revised: 07.07.2024
Accepted: 07.21.2024

Keyword:
Machine Vision
Deep Learning
Vehicle Detection
Car Classification
Car Counting

***Corresponding Author:**
Alireza Akoushideh
Email: *akushide@tvu.ac.ir*

ABSTRACT

Detection, identification, and automatic counting of vehicles using video surveillance cameras plays an important role in the field of intelligent transportation management. Despite the progress that researchers have made in these cases, its operational implementation still faces challenges such as "various environmental conditions", "unbalanced data sets", "accuracy" and "speed". Therefore, research can be useful in solving these issues. The proposed algorithm for detection, classification, and counting will be based on deep learning. In this research, after applying the proposed initial preprocessing algorithm, the YOLO algorithm was used to detect and classify vehicles. The DeepSORT algorithm was also used to track several vehicles simultaneously. For the accurate counting of vehicles, a developed method was also proposed to increase the processing accuracy. By applying the proposed pre-processing and counting techniques, the practical results showed that the "call" criterion in the video with detection at night challenge was increased to 99.18%.

E-ISSN: 3060-6357

Introduction

Video surveillance cameras are one of the main elements in intelligent transportation systems and are currently installed in all streets, highways, offices, homes, and organizations. One area that uses surveillance cameras is Intelligent Transportation (ITS). Detecting the type of car and other vehicles in the video surveillance system has many applications such as traffic monitoring, traffic congestion estimation, and self-driving car management [1]. Due to the high volume of traffic data and their growing speed, as well as the diversity of traffic data, it is impossible to use human resources in the operations of collecting, storing, analyzing, processing, and extracting data. Therefore, it is necessary to use automatic intelligent systems in this field to make traffic management possible and increase the productivity of resources [2]. Intelligent transportation is a management and service system that pursues three main goals: (a) mobility or changeability which emphasizes the efficiency and performance of transportation systems, (b) sustainable transport which focuses on traffic safety and environmentally friendly development, (c) convenience which provides services to passengers [3]. Surveillance cameras collect image information in different lighting, atmospheric, and environmental conditions. It is noteworthy that the collection of image information is not the same and each one has different environmental, atmospheric[4], and light conditions at different times, each of which is considered a challenge in its place. In other words, extracting information such as statistics of vehicles, classifying them into light and heavy, identifying license plates, and identifying logos, colors, and models [5] from this volume of image information is not an easy task. Therefore, having a robust and flexible detection, identification, and classification algorithm with correct accuracy and real-time speed to overcome all challenges is a serious and important need [6].

Detection and recognition of vehicles are some of the most effective applications in machine vision, and it is divided into two categories: "previous methods" and "methods based on deep learning", which are briefly introduced below [7]. The detection algorithms before the deep learning method (prior methods) can be divided into three categories: (a) using background subtraction [8], (b) using continuous video frame difference [9], and (c) using optical flow [10]. These approaches have good speed but with the change of light and environmental conditions, the presence of movement in the background or the camera, or if the scene is complex in terms of image, they do not provide good results. Researchers gradually turned their attention to Deep Convolutional Neural Networks (DCNN) because they can learn image features from low level to high level with greater resistance to light and environmental changes. With the increase in computing power and image information, there are greater opportunities to develop object recognition based on deep convolutional neural networks [11]. In 2014, a new architecture called Regions with Convolutional Neural Networks RCNN [12] was proposed, which is a milestone in the application of deep convolutional neural networks for object detection. Thus far, deep neural networks have many types, including Faster R-CNN [13], RetinaNet [14], YOLOv3 [15], YOLOv4 [16], and Single Shot MultiBox Detector (SSD) [17]. Methods for detection and detecting objects on images based on deep learning are divided into two categories [11]: (a) two-stage algorithms and (b) one-stage algorithms.

In general, two-stage algorithms have higher accuracy and lower speed than one-stage algorithms due to the volume of calculations. In recent years, one-step algorithms have achieved high accuracy by increasing the number of calculations and changes in the network Backbone and have replaced two-step algorithms. Algorithms for object detection and recognition based on deep learning generally consist of two parts "Backbone", which is usually a network pre-trained with data such as ImageNet, and "prediction" which is responsible for classification and border drawing [16]. In recent years, layers have been added between the backbone and prediction, which collect feature maps from different steps and is called the feature extraction step [16]. The diagnosis and statistical calculations related to vehicles are important applications of the intelligent transportation system and, as explained so far, face many challenges. In general, the vehicle detection process is as follows [18]. First, the image is received from the input of the video surveillance camera, and assuming a fixed angle of view of the camera and the location of the road surface in a certain part of the image, modeling such as the Gaussian mixture can be used in the first few frames of the video to extract the background. After extracting the background, this image is smoothed by tools such as the Gaussian filter, and algorithms such as Mean-Shift are used to make the color of the image uniform. Finally, morphological operations (hole filling) are used to extract the road surface more fully.

In the next section *Statement of the problem, innovation, and mention of goals*, the challenge of real-time detection, recognition, and counting of three types of vehicles (Car, Bus, and Truck) are explained. In research methodology, the proposed method is discussed in detail. This is followed by presenting the experimental results which are the practical results on some benchmarks. The conclusions reached are expounded in the final section.

Statement of the problem, innovation, and mention of goals

This research deals with the problem of detecting, classifying, and counting vehicles in video surveillance systems. In other words, it is used for the analysis and recovery of video data from video surveillance cameras on the roads in real-time to extract knowledge. This knowledge includes information such as the statistics of the passage of vehicles with the separation of its direction (going or returning) and distinguishing the categories of vehicles into three categories: Cars, Buses, and Trucks from each other.

Image data is received from the roads and these data are not the same in raw form and have different optical, atmospheric, and environmental differences. To improve the quality and restore the image in this research, a pre-processing stage is considered the first stage which includes the implementation of Contrast Limited Adaptive Histogram Equalization (CLAHE) [19] used on video frames. Then, at this stage, the Gaussian filter [20] is used to remove the noise of these frames.

In the second stage (detection, classification), the YOLOv4 algorithm is used for detection, detection, and classification. In the third stage (multiple object tracking), the DeepSORT tracker algorithm is used to detect the direction of movement of various types of vehicles. Finally, in the fourth stage (counting), three indicator lines were proposed to be used, and if the vehicle passes all three indicator lines, the counter process will be done. These three indicator lines, with the possibility of setting the position and angle in the area of 0.6, perform the act of counting by separating the categories of vehicles and their direction of movement on the road. Figure 1 depicts the proposed algorithm.

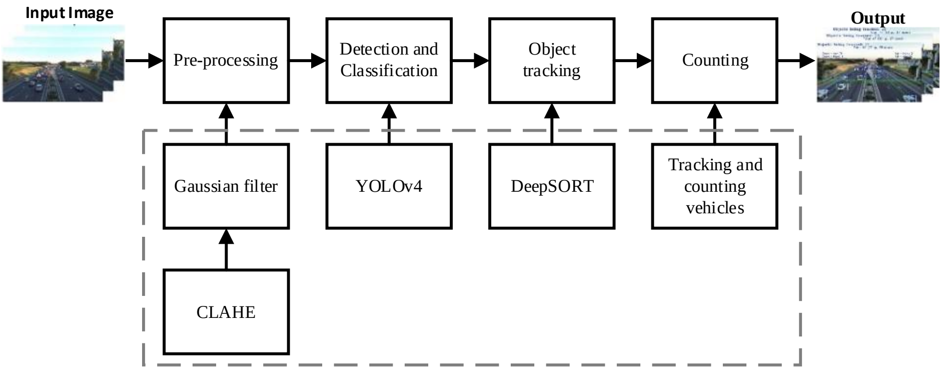

Figure 1. Block diagram of the proposed method.

Research Methodology

The proposed method of this research aimed to create an intelligent monitoring system by obtaining accurate statistics of the traffic flow by separating the category and the direction of movement of vehicles on the roads. This method, whose outline process is shown in Figure 1, includes three steps. In the following section, each step is explained.

Preprocessing: adaptive histogram with contrast restriction

Histogram Equalization is an image enhancement method in which the dynamic range of the image is expanded. The dynamic range of the image is the difference between the lightest gray level of a point in the image and the darkest point in the image. The histogram equalization method creates a transformation function that changes the pixel values of the image in such a way that the histogram of the resulting image is as uniform as possible [21]. Adaptive histogram equalization was created, in which the image is divided into small (tiled) areas and each area is improved separately by histogram equalization. Image segmentation in adaptive histogram enhancement with contrast limitation is similar to the previous method, but with the difference that all areas of the image are not enhanced to the same extent, and adaptive enhancement is performed on different areas of the image [19].

Preprocessing: Gaussian filter

Like any other signal, images can also have different types of noise, particularly when the data is collected from cameras with unfavorable atmospheric and lighting conditions; many methods have been proposed for image smoothing that helps reduce noise. In this research, a Gaussian filter was used to smooth the input image [21]. As can be observed in Figure 2, these two preprocessors can perform real-time performance with better optimization speed and quality.

(a) Before optimization (b) After optimization

Figure 2. Image (a), the original image, and image (b), the improved image.

Detection, identification, and classification

In the current research, the YOLOv4 object detection algorithm was used; this algorithm with deep neural network architecture first performs detection, and then detection and classification and gives its output to the next step by drawing a rectangular border box. YOLOv4 algorithm is a new development on YOLOv3. Using the Darknet-53 technique, adding CSP, SPP, and PAN, and using the mosaic data augmentation method improved the performance of the YOLOv3 algorithm. In the input part, the YOLO algorithm divides the image into c × c cells as shown in Figure 3. Each cell is responsible for detecting a fixed number of objects concerning their center inside the cell. In the layers of this network of the existence of an object, three predictions with different border boxes and group probability are formed. After predicting a large number of objects in the image, YOLO first filters out the bins with a category score lower than the threshold rate [22] (Figure 4). Algorithm variables are explained below.

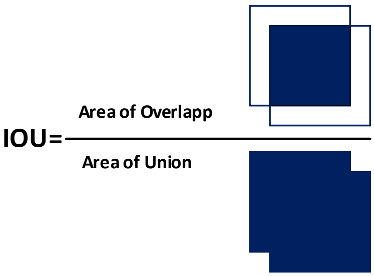

Figure 4. Visual representation of IoU intersection on images [22]

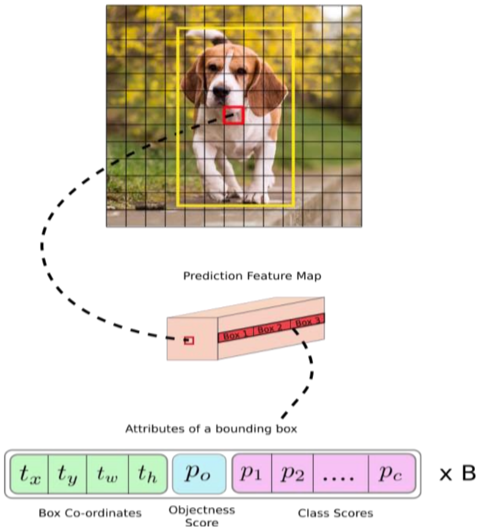

Figure 3. Window formation of cells in multiples of 32 pixels, proposed bounding box coordinates, cells score [22].

Evaluation parameters in research

Average Precision (AP): A popular metric for measuring the accuracy of object detection algorithms such as Faster R-CNN, SSD, and YOLO. In other words, the average recognition accuracy of all classes of objects is still called Mean

Average Precision (mAP). The AP Calculates the average accuracy value for the recall value between 0 and 1.

Precision: This variable measures how accurate the predictions are. That is, what percentage of the prediction is correct?

Recall: This variable measures how well the positives are found. For instance, in a prediction, 80% of possible positive cases are found. Formula (1)-(3) refers to their mathematical definitions:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (3)$$

where *TP* denotes True positive, *TN* is True Negative, *FP* defines False positive, and *FN* is False negative.

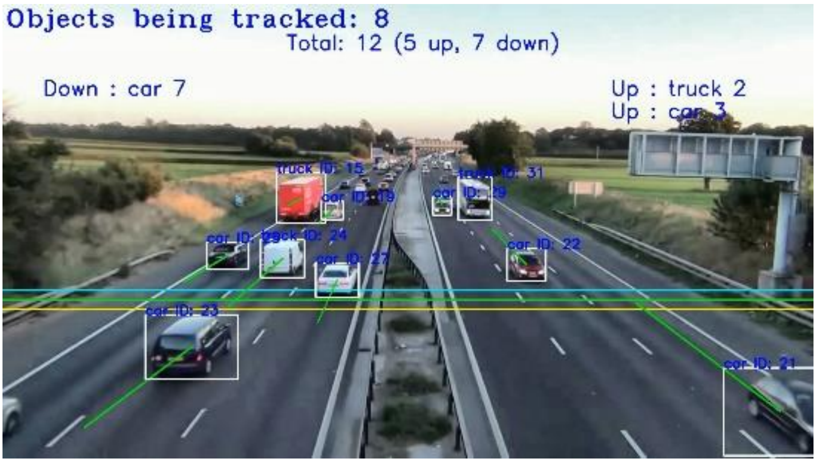

Figure 5. Drawing three indicator lines and identifying the best position for it.

Tracking and counting vehicles

At this stage, the Deep Simple Online and Realtime Tracking (DeepSORT) tracker algorithm was used [23]. A feature of this tracker is that it simultaneously tracks multiple objects in real-time. This algorithm uses the Kalman filter in the frame-to-frame communication image space, and the Hungarian method with the communication metric that measures the overlap of the border boxes; in other words, in the framework of this algorithm, the Kalman filter is used to predict the movements of vehicles and the Hungarian optimization method is used to find the similarity of vehicles and their identification used in different frames. This simple approach has achieved good performance at high frame rates [23]. The act of counting in this research was carried out by drawing three parallel index lines with a distance of 12 pixels from each other in a specific position of the image, which can be shown in Figure 5. In the present research, the accuracy of counting was found to be directly related to considering these lines in different areas of the image. In other words, the placement of these lines completely depends on the distance, height and angle of view of the camera on the road.

Experimental results

The results of the detection evaluation of the proposed method with the YOLOv4 object recognition algorithm on three test videos based on the MS COCO dataset and a dataset from Vietnam are provided in Table 1. As a result of testing these three videos and comparing them, it was discovered that the weights of the MS COCO data set compared to the Vietnam data set have high resistance and accuracy under different lighting and environmental conditions. In this proposed method, vehicles were detected during the day when the vehicle category score reaches more than 75% and at night when it reaches more than 50%. The Vietnam dataset consisted of 15,396 and 33,874 images from two different lighting conditions, day and night, respectively.

Table 1. Identification results of YOLOv3 and YOLOv4 algorithms on different datasets.

Reference	Results (mAP)	Algorithm	Dataset
[15]	57.9%	YOLOv3	MS COCO
[16]	64.95%	YOLOv4	
[24]	99.4% (for 5 classes)	YOLOv4	
[25]	67.7%	YOLOv4	DETRAC
[26]	64.51%	YOLOv4	
[27]	mAP values in the conditions of heavy snow, rain, sand, and fog are 50.33%, 45.50%, 41.65%, and 38.95%, respectively	YOLOv3	DAWN
[18]	88.87% on IoU71.32	YOLOv3	The dataset proposed in [18]
[28]	95.26%	YOLOv4	The dataset proposed in [28]
Proposed	**96.83%**	**YOLOv4**	MS COCO

99.18%	GF+CLAHE+YOLOv4

By observing the performance results of the methods introduced in Table 1, it can be concluded that the YOLOv4 algorithm with the changes in the spine, neck, and head by the researchers demonstrated excellent results compared to YOLOv3 and under adverse weather, optical, and environmental conditions showed highly favorable flexibility compared to other object detection and identification algorithms. Finally, the classification of vehicles with the help of the YOLOv4 algorithm was better than other algorithms. Different environmental conditions and light and unbalanced data sets were important factors that reduced the algorithm's accuracy. However, it is possible to enhance the dataset to a great extent by using various techniques of rotation, cutting, and flipping on the image, as well as fragmented samples of objects. In other words, the vehicle detection, classification, and tracking algorithm seriously need a dataset that includes a large number of images of each category under different lighting and environmental conditions in complex environments.

Conclusion

The current research presents an algorithm for detecting, classifying, and counting vehicles using surveillance cameras based on deep learning algorithms. A method consisting of four steps for achieving accurate statistics of vehicles crossing the road was proposed: (a) Contrast enhancement and image enhancement algorithms using CLAHE and GF which improved the results by 2.43%; (b) Detection and classification of vehicles using the YOLOv4 object recognition algorithm with the spine and neck parts enhancement according to the available hardware equipment which could overcome some critical challenges, especially on low illumination challenges; (c) The DeepSORT tracker algorithm was used to simultaneously track multiple objects in real-time and detect the direction of movement of all three types of vehicles; (d) In the fourth stage, (counting) using three index lines with the possibility of setting the position and angle in the area of *0.6* of the images, the act of counting was carried out by separating the categories of vehicles and their direction of movement (commuting) on the road. Average precision (Ap) of 99.18% was obtained which was better than many state-of-the-art methods.

It is noteworthy that solving some of the challenges discussed in the research, such as complete blockage and challenges created due to excessive increase in vehicles, might not be possible simply by using fixed surveillance cameras. It appears that the machine vision method alone cannot be relied on and it is vital to take advantage of other features that can be extracted from other sensors to reach the final goal of intelligent transportation management. In future research, it is recommended that other methods of improving input images (pre-processing step) and new releases of the YOLO object detection algorithm are used to achieve better accuracy.

Disclosure statement and funding

The authors declare no potential conflicts of interest. The present study received no financial support from any organization or institution.

Acknowledgment

We would like to give special thanks to all the participants in this study.

References

[1] Kaffash, S., Nguyen, A. T., & Zhu, J. (2021). Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis. *International Journal of Production Economics*, *231*, 107868. https://doi.org/10.1016/j.ijpe.2020.107868

[2] Pourjavan, K. (2019). Explanation of Smart City and Urban Smart Transportation Solutions. *Quarterly Scientific Journal of National University of skills*, *16*(1), 15-34. https://karafan.tvu.ac.ir/article_100529.html?lang=en

[3] Akoushideh, A., Tourani, A., Shahbahrami, A., & Masoumnezhad, M. (2021). Design and Implementation of Automatic License Plate Recognition System for Security Gates. *Quarterly Scientific Journal of National University of skills*, *18*(3), 237-252. https://doi.org/10.48301/kssa.2021.130288

[4] Moradi, A., Shahbahrami, A., & Akoushideh, A. (2021). An unsupervised approach for traffic motion patterns extraction. *Institution of Engineering and Technology Image Processing*, *15*(2), 428-442. https://doi.org/10.1049/ipr2.12033

[5] Galea, S., Seychell, D., & Bugeja, M. (2020, December 3-5). *A Survey of Intelligent transportation systems based Modern Object Detectors Under Night-time Conditions* [Conference session]. 3rd International Conference on Intelligent Sustainable Systems ,Thoothukudi, India. https://doi.org/10.1109/ICISS49785.2020.9316076

[6] Tourani, A., Shahbahrami, A., & Akoushideh, A. (2017, November 22-23). *Challenges of video-based vehicle detection and tracking in intelligent transportation systems* [Conference session]. 2nd International Conference on Soft Computing, Rudsar, Guilan, Iran. https://civilica.com/doc/696790/

[7] Shetty, A. K., Saha, I., Sanghvi, R. M., Save, S. A., & Patel, Y. J. (2021, April 2-4). *A Review: Object Detection Models* [Conference session]. 6th International Conference for Convergence in Technology, Maharashtra, India. https://doi.org/10.1109/I2CT51068.2021.9417895

[8] Manikandan, R., & Ramakrishnan, R. (2013). Video object extraction by using background subtraction techniques for sports applications. *Digital Image Processing*, *5*(9), 435-440. https://www.researchgate.net/publication/260288356_Video_Object_Extraction_by_Using_Background_Subtraction_Techniques_for_Sports_Applications

[9] Li, Q-L., & He, J-F. (2011). Vehicles detection based on three-frame-difference method and cross-entropy threshold method. *Jisuanji Gongcheng/ Computer Engineering*, *37*(4), 172-174. https://doi.org/10.3969/j.issn.1000-3428.2011.04.062

[10] Liu, Y., Lu, Y., Shi, Q., & Ding, J. (2013, December 14-15). *Optical Flow Based Urban Road Vehicle Tracking* [Conference session]. 2013 Ninth International Conference on Computational Intelligence and Security, Emeishan, China. https://doi.org/10.1109/CIS.2013.89

[11] Xiao, Y., Tian, Z., Yu, J., Zhang, Y., Liu, S., Du, S., & Lan, X. (2020). A review of object detection based on deep learning. *Multimedia Tools and Applications*, *79*(33), 23729-23791. https://doi.org/10.1007/s11042-020-08976-6

[12] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014, June 23-28). *Rich feature hierarchies for accurate object detection and semantic segmentation* [Conference session]. Proceedings of the Institute of Electrical and Electronics Engineers conference on computer vision and pattern recognition, Ohio, USA. https://doi.org/10.1109/CVPR.2014.81

[13] Ren, S., He, K., Girshick, R. B., & Sun, J. (2015, December 7-12). *Faster r-cnn: Towards real-time object detection with region proposal networks* [Conference session]. Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems, Montreal, Quebec, Canada. http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks

[14] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017, October 22-29). *Focal Loss for Dense Object Detection* [Conference session]. 2017 Institute of Electrical and Electronics Engineers International Conference on Computer Vision Venice, Italy. https://doi.org/10.1109/ICCV.2017.324

[15] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*, 1-6. https://doi.org/10.48550/arXiv.1804.02767

[16] Bochkovskiy, A., Wang, C-Y., & Liao, H-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv*, 1-17. https://doi.org/10.48550/arXiv.2004.10934

[17] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C-Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016* (pp. 21-37). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_2

[18] Song, H., Liang, H., Li, H., Dai, Z., & Yun, X. (2019). Vision-based vehicle detection and counting system using deep learning in highway scenes. *European Transport Research Review*, *11*(1), 51. https://doi.org/10.1186/s12544-019-0390-4

[19] Yadav, G., Maheshwari, S., & Agarwal, A. (2014, September 24-27). *Contrast limited adaptive histogram equalization based enhancement for real time video system* [Conference session]. 2014 International Conference on Advances in Computing, Communications and Informatics, Delhi, India. https://doi.org/10.1109/ICACCI.2014.6968381

[20] Gedraite, E. S., & Hadad, M. (2011, September 14-16). *Investigation on the effect of a Gaussian Blur in image filtering and segmentation* [Conference session]. International Symposium on Electronics in Marine 2011, Zadar, Croatia. https://ieeexplore.ieee.org/abstract/document/6044249

[21] Hasanpour, H., & Asadi Amiri, S. (2015). *Comprehensive concepts of digital image processing along with implementation of algorithms with MATLAB* (S. Gholamian, Ed.). Shahrood University of Technology. https://www.gisoom.com/book/11377594

[22] Sejr, J. H., Schneider-Kamp, P., & Ayoub, N. (2021). Surrogate Object Detection Explainer (SODEx) with YOLOv4 and LIME. *Machine Learning and Knowledge Extraction*, *3*(3), 662-671. https://doi.org/10.3390/make3030033

[23] Wojke, N., Bewley, A., & Paulus, D. (2017, September 17-20). *Simple online and realtime tracking with a deep association metric* [Conference session]. 2017 Institute of Electrical and Electronics Engineers International Conference on Image Processing Beijing, China. https://doi.org/10.1109/ICIP.2017.8296962

[24] Can, V. X., Vu, P. X., Rui-fang, M., Thuat, V. T., Van Duy, V., & Noi, N. D. (2021). Vehicle detection and counting under mixed traffic conditions in vietnam using yolov4. *International Journal of Advanced Research Engineering and Technology*, *12*(2), 722-730. https://doi.org/10.34218/IJARET.12.2.2021.072

[25] Mahto, P., Garg, P., Seth, P., & Panda, J. (2020). Refining yolov4 for vehicle detection. *International Journal of Advanced Research in Engineering and Technology 11*(5), 409-419. https://doi.org/10.34218/IJARET.11.5.2020.043

[26] Ho, N., Pham, M., Vo, N. D., & Nguyen, K. (2020, November 26-27). *Vehicle Detection at Night Time* [Conference session]. 7th National Foundation for Science and Technology Development Conference on Information and Computer Science Ho Chi Minh City, Vietnam. https://doi.org/10.1109/NICS51282.2020.9335870

[27] Hassaballah, M., Kenk, M. A., Muhammad, K., & Minaee, S. (2021). Vehicle Detection and Tracking in Adverse Weather Using a Deep Learning Framework. *Institute of Electrical and Electronics Engineers Transactions on Intelligent Transportation Systems*, *22*(7), 4230-4242. https://doi.org/10.1109/TITS.2020.3014013

[28] Ni, Z., Liu, T., Li, K., Bai, Y., & Zhu, Z. (2021). Real-time vehicle detection and computer intelligent recognition through improved YOLOv4. *Journal of Physics: Conference Series*, *2083*(4), 042006. https://doi.org/10.1088/1742-6596/2083/4/042006