Computer Vision: A Review of Detecting Objects in Videos – Challenges and Techniques

https://doi.org/10.3991/ijoe.v18i01.27577

Mohammad Ali A. Hammoudeh(✉), Mohammad Alsaykhan, Ryan Alsalameh, Nahs Althwaibi
Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia
maah37@qu.edu.sa

Abstract—Traffic safety aims to change the attitude of citizens towards careless traffic on the roads, making this the first step towards changing behavior. Also, teach the rules of safe pedestrian behavior and minimize the risks of road accidents. So many regulations have been set to avoid road accidents and traffic jams, which is the study scope of this paper using IT technology. With the expanding interests in Computer vision use cases such as vehicles self-driving, face recognition, intelligent transportation frameworks and so on individuals are hoping to assemble custom AI models to recognize and distinguish specific objects. Object detection is part of a computer's vision where objects that can be observed externally and are found in videos can be identified and tracked by computers. Therefore, object tracking is an important part of video analysis. There are many proposed methods such as Tracking, Learning, Detection, Mean shift and MIL. In this paper, the computer vision state in object detecting domain along with its challenges are discussed, also we address some requirements and techniques to overcome these challenges. Finally, TensorFlow technology is presented as a recommended solution to support Lane's violation.

Keywords—computer vision, traffic safety applications, object detection, AI models, video analysis, convolutional neural network, tensor-flow, lanes violation tracking

1 Introduction

This In the last 10 decades people didn't know that much about technology as we do now; they did not imagine the new and sophisticated technology yet it was discovered and used properly. Starting with the point that everything is possible this paper will dive into the sea that is object detection. A process of identifying, localizing, and classification of an object in a scene that is object detection which is a branch of computer vision [1].

Computer vision (CV) is a simulation of human vision that is meant to train the computer to understand and interpret vision of the world using several algorithms and applications that support this science [2].

The spread of powerful computers, the availability of high quality, cheap video cameras and the highly need for automated video analysis has created a focused interest in object tracking algorithms. Video analysis has three main steps as detecting a specific moving object, tracking an object between frames and recognizing object behavior by analyzing their tracks. Therefore, object tracking is relevant to the following tasks:

1. Recognition based on motion, such as identifying someone based on gait and automatic object detection, etc.
2. Automated surveillance as Detection of suspect activities or unforeseen from a scene.
3. Video indexing as the videos retrieving automatic annotation in databases of multimedia.
4. Human interaction with the computer as gesture recognition, tracking eye gaze to be stored in computers etc.
5. Traffic monitoring as gathering traffic statistics in real time to direct traffic flow.
6. Vehicle navigation as planning path-based video and capabilities of obstacle avoidance.

Nowadays, object detection has a lot of applications in the real-world, such as video surveillance, robot vision, autonomous driving, etc. Figure 1. shows the increasing growing number of publications that are associated with "object detection" over the last decade.

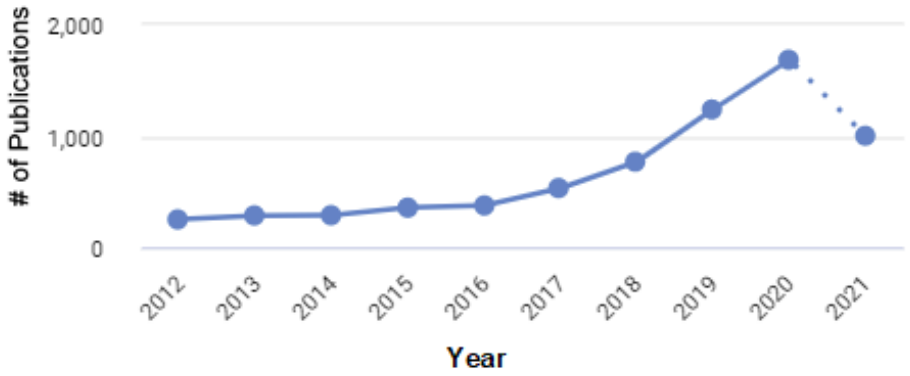

Fig. 1. # of Publications in object detection from 2012 to 2021 (July)

In Saudi Arabia car accidents and traffic jams have been the most concerning things to its members. In addition, these problems must be solved to avoid losing time and reduce car accidents and Jams (Traffic Pollution) Problems.

Seemingly, object detection is easily applied to monitor traffic violations, which lead to reduced car accidents and time lost in pollution and traffic jams. It aims to change the attitude of citizens towards careless traffic on the roads, making this the first step

towards changing behavior. Also, teach the rules of safe pedestrian behavior and minimize the risks of road accidents.

Taking this troubling and challenging situation as a starting point, this paper studies the object detection technology and how it can be used in traffic safety, by highlighting main challenges and proposing a solution. We review some recent research in AI models regarding object detection techniques. The remainder of this article is as follows: Section II presents the related work of traffic detecting objects. Section III reviews some discussions about object detection frameworks; present the feedback of Saudi citizen about traffic violations and solutions. Section IV presents the conclusion and future work.

2 Related work

In this section, a brief review of previous studies were presented as computer vision aspect, focusing on the objects detection concepts in Videos, in order to show some different types of framework in Machine Learning (ML) and deep learning (DL) that used for implementing diverse applications such as object tracking, classification, object detection, analysis and many more, some of which will be presented later.

Object detection is basically defining whether any object is appearing on a video or an image by shape, color and texture. It can be used in video or digital images and it can be static or dynamic in a video. Image object detection is a classification of the image and then localizing the object by pixels and many methods like sliding box as in Figure 2. [3].

Fig. 2. Object detection classification

Object detection in video is being used in many implementations such as video surveillance and autonomous driving. A video has many scenes or frames related to each other. The object is being detected in a video can be dynamic (moving object) or static (stationary object). Sometimes the object can be temporary static or dynamic [4, 5, 6].

Object detection and tracking are very important for different types of applications such as safety, driver assistance, and traffic control and more. The goal in each of these applications is detecting the object at each time period and to know its location. In safety applications, desired objects should be detected and tracked in dangerous environments, to keep them safe from machines, whereas the DAS detects and tracks obstacles, vehicles, and pedestrians to avoid any collision in the movement path. In the traffic control system, the purpose is to ensure smooth traffic, track vehicles and detect lane violations to lessen traffic accidents.

Table 1 summarize some of the studies conducted on improving traffic safety by mathematicians, engineers and city planners with a variety of results. Studies generally rely on different solution methods; every method has advantages and disadvantages [7-12]:

Table 1. Summary of traffic detecting objects system/Model methods in previous studies

Paper	Problem description	Solution method	Advantages	Disadvantages
[7]	The first problem is people violating traffic regulations, which causes another problem, which is the traffic congestion.	-Two methods were used, first method Color-to-Gray scale Algorithm and the second is Cascading Classifiers. -In addition to using Raspberry Pi Camera.	High accuracy object identification and detection	-Requires good internet connection. -The camera resolution is not perfect (5 MP). -Doesn't work properly in low light conditions.
[8]	Traffic density and congestion at traffic junctions.	Background subtraction algorithm and morphological operations, in addition to using Raspberry Pi 3 Model B+ Camera.	High accuracy object identification and detection.	Need some presets manually.
[9]	Detecting small-sized objects, low illumination that effects on object classification.	Using (ASPP)-Center Net method, which contain three-part to solve that: -Down sampling module. -The output layers -Space to depth module.	-Detect a more difficult or invisible object -Detect small size objects.	Requires a large amount of data.
[10]	Difficulty Gathering information about a vehicle.	Using MoG Background subtraction + SVM and Faster RCNN.	-Determine vehicle type and speed. -Count the number of vehicles.	Difficult to detecting object at night.
[11]	The problems that were challenging is unexpected number of moving objects and poorly textured objects Illumination conditions and occlusion.	Combination of genetic dynamic saliency map and background subtraction.	-Practical to handle shadow and occlusion problems efficiently. -Faster object detection method and simple.	-The sizes of the detected objects are bigger than the real sizes. -Generated by the repetitive image resizing may reduce the object detection performance.
[12]	Making object detection take less time and more accurate.	Combining between YOLO and R-FCN Algorithms.	It is faster than normal object detection with other algorithm by 18%.	Needing large set of data.

The most popular types of framework have been used in object detection as the following [13]:

- **TensorFlow:** In Nov. 2015, Google issued TensorFlow; an Open source library software for deep learning to describe, train and deploy ML models.
- **Keras:** François Chollet, a Google engineer, designed Keras and it is an open-source neural network library written in Python, which can run on different machine learning framework like Microsoft Cognitive Toolkit, TensorFlow, PlaidML, R, and Theano. Keras developed to permit fast application with neural networks. It is easy to use for beginners, extensible and modular.
- **PyTorch:** is one of the common machine-learning frameworks and it is created by Facebook's Artificial Intelligence Research Lab and it is released in October 2016. PyTorch is a neural network library basically and it's built on another machine learning library which is Torch.
- **Caffe:** Berkeley AI Research developed Caffe and it was written in Python and C++ programming language and released in 2014. Caffe is a DL framework that supports different types of modules and it is made with modularity, speed, and expression in mind.
- **Theano:** is a deep learning library run on both the central processing unit (CPU) and graphics-processing unit (GPU). It is designed to deal with mathematical expressions to describe, enhance, and assess it, engaging multi-dimensional arrays smoothly. It is the oldest library developed in 2007 by Yoshua Bengio.
- **CNTK (Microsoft Cognitive Toolkit):** In Jan. 2016, Microsoft dispatched a deep learning framework called Microsoft Cognitive Toolkit. It is open-source and backing in a several programming languages, for example, C#, C++, Java and Python. CNTK is utilized in numerous Microsoft applications, for example, Skype, Cortana and Xbox.
- **DeepLearning4J (DL4J):** is an open source (DL) framework based on Java Virtual Machine. DL4J supplies adjustment of instruments for construction production-grade Deep Learning applications. DL4J permits training and inference on CPU or GPU groups to further accelerate machine-learning workloads.
- **MXNET:** is a DL framework made by Apache, which bolsters a plenty of languages, similar to Python, Julia, C++, R, or JavaScript. It has been embraced by Microsoft, Intel and Amazon Web Services.
- **Chainer:** is A Python framework to rapidly execute, prepare and assess profound learning models. It's a ground-breaking, adaptable, and natural neural networks framework that overcomes any barrier among algorithms and executions.

3 Discussion

In this section, we will introduce deep learning methods based on object detection, a comparative analysis of existing AI models, feedback of Saudi citizen about traffic safety then we list some basic challenges in video object tracking and requirements based on their features, after that we define the recommended solution that help to solve and reduce different problems.

3.1 Deep learning methods based object detection

Recently the state-of-the-art results related to object detection; Deep learning has become a buzzword in this field and its methods that give a solution to the computer vision to develop object detection. The deep learning method uses several algorithms and it is based on convolutional neural net-works (CNNs) families. CNN contains three models as the following:

1. R-CNN: The original goal from R-CNN is to have an image as input, generate a set of boxes around an object in this image and display it as output. The architecture of R-CNN is taking input as an image, extract region proposal around 2000 bottom-up, compute features for each proposal using a CNN, then classifies all region that we funded by support-vector-machine (SVM) it is used to decide the type of object [14].
2. Fast R-CNN: This architecture like R-CNN but seek to reduce the time through send whole an input to CNN, CNN determines ROI, then apply ROI Pooling layer on all extract regions of interest (ROI), the benefits of this layer just to make sure all regions in the same size, all regions passed to complete connected network to classifies and returns the bounding boxes using SoftMax and linear regression layers [14].
3. Faster R-CNN: This architecture is the modified version of Fast R-CNN. Faster RCNN uses "Region Proposal Network" (RPN), while Fast RCNN uses the elective search for producing (ROI). RPN uses image properties maps as an input and produces a group of proposed objects, each with an abjectness score as output [15].

The Feature of each of the previous models as shown in Table 2. are:

a) R-CNN: uses eclectic search to produce regions. Expands around 2000 regions from each image.
b) Fast R-CNN: in Fast R-CNN the single image passed to the CNN in or-der to extract feature maps. Predictions are generated by selective search from these maps. Gather all RCNN models.
c) Faster R-CNN: substitutes the local proposal network with a selective search method to make it faster.

Table 2. Comparison between CNN algorithm

Parameters	R-CNN	Fast R-CNN	Faster R-CNN
Test Time Per Image	49 Seconds	22 Seconds	0.2 Seconds
Speed	1x	25x	250x
Mean Average Precision	66.0	66.9	66.9

3.2 Comparative analysis of AI models

The following Figure 3. and Table 3 are showing some comparative cases between AI Models and the machine learning frameworks, where can find that the Tensorflow is dominant in online Job Listing, Google Search Volume, KDnuggets Usage Survey, ArXiv Articles, Amazon Books, Medium Articles and GitHub Activity [16]:

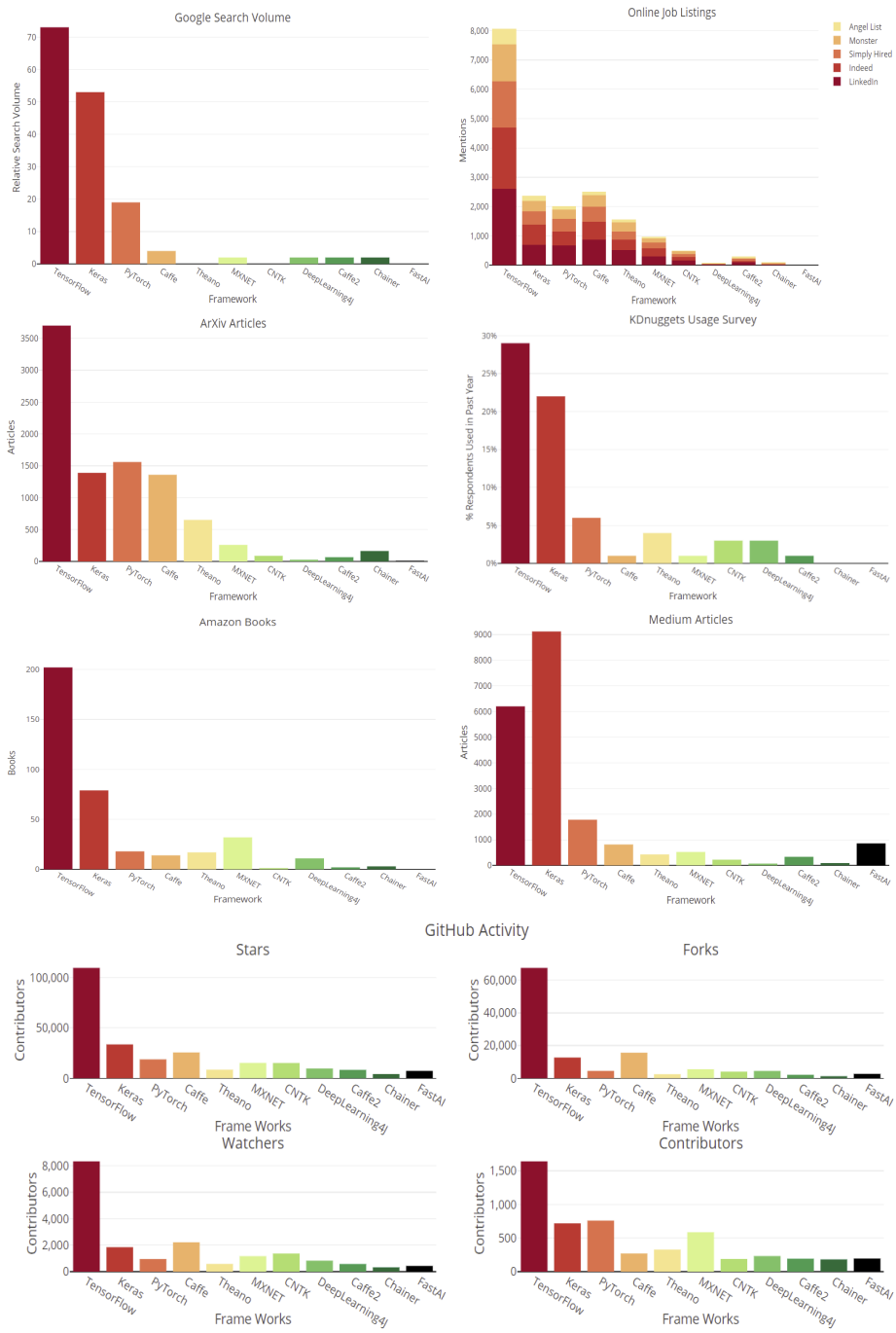

Fig. 3. Comparative analysis of AI models

Table 3. ML framework comparison

	Online Job Listings					**KDnug-gets**	**Google**	**Me-dium**	**Ama-zon**	**ArXiv**	**GitHub Activity**			
Framework	Indeed	Monster	Simply Hired	LinkedIn	Angel List	Usage Survey	Search Volume	Articles	Books	Articles	Stars	Watchers	Forks	Contributors
TensorFlow	2,079	1,253	1,582	2,610	552	29.90%	73	6,200	202	3,700	109,576	8,334	67,551	1,642
Keras	684	364	449	695	177	22.20%	53	9,120	79	1,390	33,558	1,847	12,658	719
PyTorch	486	309	428	665	120	6.40%	19	1,780	18	1,560	18,716	952	4,474	760
Caffe	607	399	515	866	123	1.50%	4	815	14	1,360	25,604	2,218	15,633	270
Theano	356	316	279	508	95	4.90%	0	428	17	652	8,477	585	2,447	328
MXNET	266	154	200	298	29	1.50%	2	524	32	260	15,200	1,170	5,498	587
CNTK	126	96	97	160	12	3.00%	0	223	1	88	15,106	1,368	4,029	189
Deep Learning4J	17	5	9	35	3	3.40%	2	70	11	27	9,615	829	4,441	232
Caffe2	55	51	49	109	12	1.20%	2	335	2	67	8,284	577	2,102	193
Chainer	19	19	19	28	3	0.00%	2	91	3	164	4,128	325	1,095	182
FastAI	0	0	0	0	0	0.00%	0	858	0	11	7,268	432	2,647	195

3.3 Feedback of Saudi citizens about traffic safety

This questionnaire is designed to capture the community's perception about a future program that could help in improving traffic safety in Saudi Arabia.

After publishing it, we obtained many answers about the 22 questions starting with asking the gender of the user. We had 245 males as 81.7% and the females were 55 as 18.3%, 55 citizens were 30 or less years old as 91.7% and all of them were Saudis. Most of the citizens have a car as 86.4%.

Regarding the traffic awareness of the traffic rules there is only 1% Unaware completely and the rest were 55% very aware of them and 41.7% somewhat aware. 91.7% of the citizens know the Violation consequences and only 4% do not. When we asked the citizens about if they had an experience with any violation 59% of them did.

Then we asked them "What is the biggest traffic problem?" that have encountered as in Figure 4.

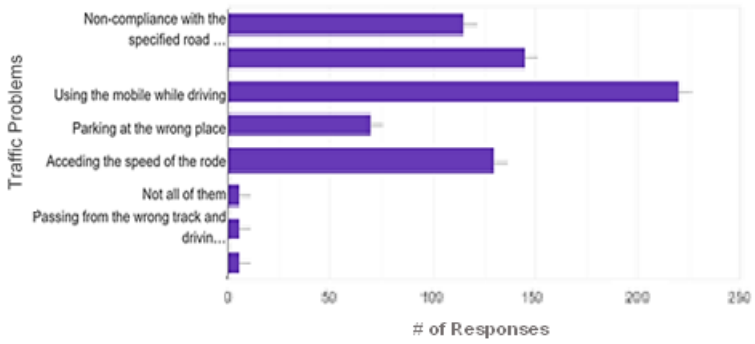

Fig. 4. The biggest traffic problem

The answers of Question "I think AI System (Detecting Objects) has a positive role in enhancing the driver's behavior" as shown in Figure 5.

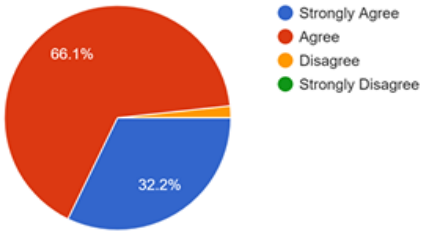

Fig. 5. Importance of AI system

The answers of Question "Please select one of the following that you think has the greatest impact on improving traffic safety" as in Figure 6.

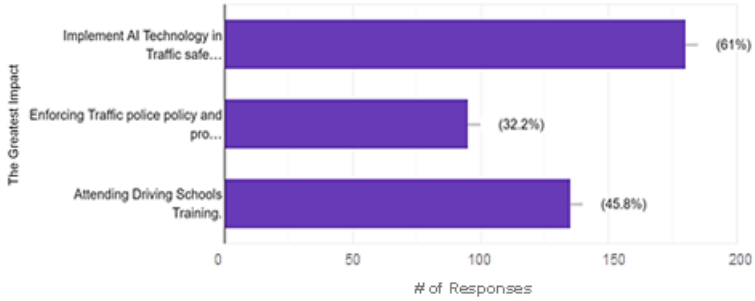

Fig. 6. The greatest impact on improving traffic safety

3.4 Some challenges in video object detection

In video, Object detection has many challenges as the following:

1. Illumination Changes.
2. Occlusion.
3. Speed of the Moving Objects and Intermittent Object Motion.

Illumination challenges, sometimes the Illumination in a scene changes which means that the background color in each frame may be affected and also this could cause some objects to disappear due to changing Illumination [17].

One of the other challenges that you may face video may be captured with a camera that experiences vibrations and instability resulting in inaccurate object detection results and speed of the moving objects and intermittent object motion. It could be one of the obstacles [18, 19, 20]. The significant challenging point in object detection is occlusion and it may be full/partial and can occur anytime an object passes behind another object [17, 21].

3.5 Recommended solution for lanes violation

Whether you are an expert or a beginner, TensorFlow is an end-to-end platform that makes it easy for you to build and deploy ML models. TensorFlow treats every day, real machine learning problems; many companies with various industries apply ML to solve their problems such as ecommerce, social networks and healthcare.

TensorFlow has the following features:

- **Building Models Easily:** TensorFlow has multi-levels of abstraction to train and build models.
- **Robust Machin-Learning Production Anywhere:** TensorFlow permits you to deploy and train your model easily, regardless of what platform or language you use.
- **Powerful Experimentation for Research:** TensorFlow gives you flexibility and control with features like the Keras Functional API and Model Sub classing API for the creation of complex topologies.

Based on research in the previous studies, we recommend using TensorFlow to support traffic safety by applying the following model in Figure 6. It provides powerful methods, flexible and convenient to perform video-object detection that store the entire video analysis in databases and/or real-time visualizations to be used in the future insights. The model is supported by AI model to detect various objects to perform detection for plenty of these items and customize it.

Our model depends on OpenCV's to detect live videos for camera inputs. Live video streams might be loaded from a device camera; cameras connected by cable or IP cameras, and parse it into a specific function.

The proposed work will allow developers to obtain deep insights into any video. This insight can be visualized in real-time, stored in a (NoSQL) database for future review or analysis.

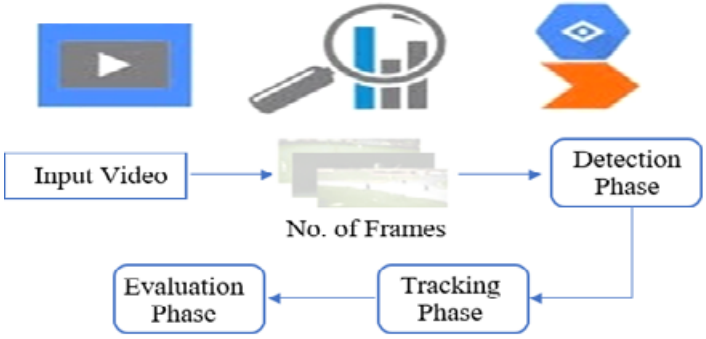

Fig. 7. Video object detection phases

4 Conclusion and future work

In this research paper, we discussed the object detection challenges, techniques and some related statistics, focusing on traffic violation solutions using TensorFlow to detect objects in video.

Finally, TensorFlow was discussed as a recommended solution that supports Artificial Tracking and Analysis in terms of traffic safety. In conclusion, the target of this work is to concentrate on the recent research in this field as a contribution to share liability of knowledge.

As a future work, we plan to examine and evaluate the TensorFlow integration with Muroor authority. Therefore, the proposed work will improve the response time and will have multi-services and feature requirements.

5 References

[1] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., and Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. International journal of computer vision, 128(2): 261-318. https://doi.org/10.1007/s11263-019-01247-4

[2] Crafton, B., Paredes, A., Gebhardt, E., and Raychowdhury, A. Hardware-Algorithm Co-Design Enabling Efficient Event-based Object Detection. 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2021, IEEE, pp. 1-4. https://doi.org/10.1109/aicas51828.2021.9458497

[3] Zhao, Z. Q., Zheng, P., Xu, S. T., and Wu, X. (2019). Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems, 30(11): 3212-3232. https://doi.org/10.1109/TNNLS.2018.2876865

[4] Yahiaoui, M., Rashed, H., Mariotti, L., Sistu, G., Clancy, I., Yahiaoui, L., and Yogamani, S. (2019). Fisheyemodnet: Moving object detection on surround-view cameras for autonomous driving. arXiv preprint arXiv:1908.11789.

[5] Mabrouk, A. B. and Zagrouba, E. (2018). Abnormal behavior recognition for intelligent video surveillance systems: A review. Expert Systems with Applications, 91: 480-491. https://doi.org/10.1016/j.eswa.2017.09.029

[6] Maddalena, L. and Petrosino, A. (2008). A self-organizing approach to background subtraction for visual surveillance applications. IEEE Transactions on Image Processing, 17(7): 1168-1177. https://doi.org/10.1109/tip.2008.924285
[7] Shamrat, F. J. M., Mahmud, I., Rahman, A. S., Majumder, A., Tasnim, Z., and Nobel, N. I. (2020). A smart automated system model for vehicles detection to maintain traffic by image processing. International Journal of Scientific & Technology Research, 9(02): 2921-2928.
[8] Gupta, A., Gandhi, C., Katara, V., and Brar, S. (2020, July). Real-time video monitoring of vehicular traffic and adaptive signal change using Raspberry Pi. In Conference on Engineering & Systems (SCES), 2020, IEEE, pp. 1-5. https://doi.org/10.1109/sces50439.2020.9236731
[9] Li, G., Xie, H., Yan, W., Chang, Y., and Qu, X. (2020). Detection of road objects with small appearance in images for autonomous driving in various traffic situations using a deep learning-based approach. IEEE Access, 8: 211164-211172. https://doi.org/10.1109/access.2020.3036620
[10] Arinaldi, A., Pradana, J. A., and Gurusinga, A. A. (2018). Detection and classification of vehicles for traffic video analytics. Procedia computer science, 144: 259-268. https://doi.org/10.1016/j.procs.2018.10.527
[11] Bouwmans, T. and Garcia-Garcia, B. (2019). Background subtraction in real applications: Challenges, current models and future directions. arXiv preprint arXiv:1901.03577. https://doi.org/10.1016/j.cosrev.2019.100204
[12] Tao, J., Wang, H., Zhang, X., Li, X., and Yang, H. (2017, October). An object detection system based on YOLO in traffic scene. 6th International Conference on Computer Science and Network Technology (ICCSNT), 2017, IEEE, pp. 315-319. https://doi.org/10.1109/iccsnt.2017.8343709
[13] Wang, Z., Liu, K., Li, J., Zhu, Y. and Zhang, Y. (2019). Various frameworks and libraries of machine learning and deep learning: a survey. Archives of computational methods in engineering, 1: 1-24. https://doi.org/10.1007/s11831-018-09312-w
[14] Girshick, R. Fast r-cnn. International conference on computer vision, 2016, IEEE, pp. 1440-1448.
[15] Zhong, Z., Sun, L., and Huo, Q. (2019). Improved localization accuracy by LocNet for Faster R-CNN based text detection in natural scene images. Pattern recognition, 96: 106986. https://doi.org/10.1016/j.patcog.2019.106986
[16] Jeff Hale, "Deep Learning Framework Power Scores", [Online], available on: https://towardsdatascience.com/deep-learning-framework-power-scores-2018-23607ddf297a, (accessed on July 14.2021).
[17] Kim, W. J., Hwang, S., Lee, J., Woo, S., and Lee, S. (2021). AIBM: Accurate and Instant Background Modeling for Moving Object Detection. IEEE Transactions on Intelligent Transportation Systems, 1-16. https://doi.org/10.1109/TITS.2021.3090092
[18] Mantuano, E. S., Garcia-Quilachamin, W., and Santana, J. A. (2021). A Systematic Review of Algorithms in People Images Detection Based on Artificial Vision Techniques for Energy Management in Air Conditioners. International Journal of Online and Biomedical Engineering (iJOE),17(1): 17-33. https://doi.org/10.3991/ijoe.v17i01.17899
[19] Yanni, R., El-Bakry, H., Riad, A., and El-Khamisy, N. (2020). Internet of Things for Surgery Process Using Raspberry Pi. International Journal of Online and Biomedical Engineering (iJOE), 16(10): 96-115. https://doi.org/10.3991/ijoe.v16i10.15553
[20] Lin, H., Cai, K., Chen, H., and Zeng, Z. (2015). The Construction of a Precise Agricultural Information System Based on Internet of Things. International Journal of Online Engineering (IJOE), 11(6): 10-15. https://doi.org/10.3991/ijoe.v11i6.4847

[21] Memon, F. A., Khan, U. A., Shaikh, A., Alghamdi, A., Kumar, P., & Alrizq, M. (2021). Predicting Actions in Videos and Action-Based Segmentation Using Deep Learning. IEEE Access, 9, 106918-106932. https://doi.org/10.1109/ACCESS.2021.3101175

6 Authors

Mohammad Ali A. Hammoudeh is an Assistant professor at Department of Information Technology, College of Computer, Qassim University, Buraydah 51941, Saudi Arabia.

Mohammad Alsaykhan is with department of Information Technology, college of Computer, Qassim University, Buraydah 51941, Saudi Arabia.

Ryan Alsalameh is with department of Information Technology, college of Computer, Qassim University, Buraydah 51941, Saudi Arabia.

Nahs Althwaibi is with department of Information Technology, college of Computer, Qassim University, Buraydah 51941, Saudi Arabia.

Article submitted 2021-10-04. Resubmitted 2021-11-26. Final acceptance 2021-11-27. Final version published as submitted by the authors.