International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

December 2022 | Issue 53 | Volume 14 | Number 4 | Pages 58-64

REAL-TIME MONITORING FOR SOCIAL DISTANCING VIA YOLOV5 ALGORITHM

H.H. Saeed **A. Alazzawi**

Department of Computer Science, College of Science, University of Diyala, Baqubah, Iraq
scicompms2130@uodiyala.edu.iq, dr.abdulbasit@uodiyala.edu.iq

Abstract- The deadly COVID-19 outbreak has wreaked havoc around the globe. As of 2022, there have been around 42 million cases worldwide, with 1.14 million deaths. A fuller knowledge of the epidemic demonstrates that the irresponsibility of a single person can have far-reaching, irrevocable consequences. COVID-19 must be contained by social isolation. Consequently, a system to monitor and identify the human-endangering distance is required. The proposed technique uses Euclidean distance-derived bounding boxes and distance metrics to leverage the YOLOv5 object identification model to track known individuals. Experimentation demonstrated that the YOLO v5-based Euclidean distance method outperformed other deep learning algorithms such as YOLOv3 and YOLOv4. Our model achieved less inference time and a high process frame score with balanced mAP.

Keywords: Social Distancing, COVID-19, Surveillance System, Object Detection, Deep Learning.

1. INTRODUCTION

By World meter, the report is shown the COVID-19 infections in the world; around 475,377,907 coronavirus cases are confirmed and viewed by a statistic in different countries within the last update in 2022. COVID-19 has turned into a worldwide tragedy. Since the emergence of the COVID-19 pandemic before the new year of 2020, there has been a global health crisis. Several initiatives have been performed worldwide to slow or halt the transmission of the highly infectious viral disease. Many specialists' medicals advise that social distancing is the best approach for preventing the spread of the contagious virus. Social distancing avoids close contact and human interactions that might contribute to the spread of fatal illnesses. International institutions, such as the World Health Organization (WHO), subsequently determined that social isolation was the most efficient method for containing the infectious virus. Multiple researchers conducted a timely inquiry on the social distancing effects on stopping the spread of the COVID-19 pandemic and revealed highly optimistic results, verifying WHO and CDC recommendations [1].

In accordance with the adage "prevention is better than cure", the World Health Organization (WHO) has advised numerous preventive measures to decrease the spread of coronavirus. Social distancing [2]. They have proven to be among the most excellent likely approaches as a transmission deterrent in the current environment, known as "social distancing", which attempts to reduce contact between people who might be affected by COVID-19 and healthy individuals [3]. Figure 1 shows the measurement curve for people who follow the social distance; they have low COVID-19 infection than those who people had not followed the procedure of WHO.

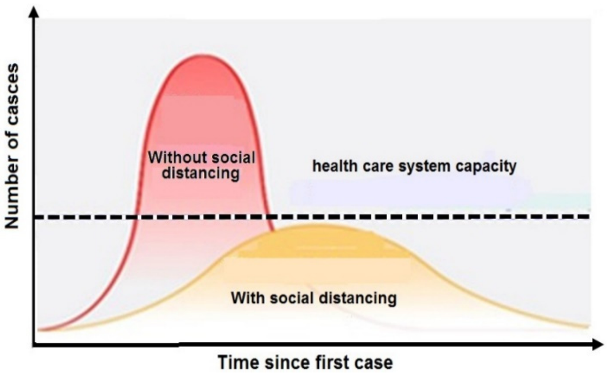

Figure 1. With and without the obligation of social distancing

To maintain social distancing, everyone should keep a space of a minimum (6 feet) between them, as per WHO standard prescriptions. This is a well-known method of breaking the transmission of infection. As a result, social distancing has become the norm in all impacted nations. It is challenging to keep track of social distancing in actual circumstances. There exist two surveillance approaches: manually as well as automatically. The manual approach entails many physical eyes using cameras in surveillance systems to ensure that each person keeps their distance. While an automated surveillance system investigates the space, the system issues a warning when a strange occurrence happens. Security staff may conduct appropriate measures in light of this notice [4, 5].

Consequently, "next-gen" autonomous systems must monitor the vast area of stations to keep safety social distance. Object detection techniques are intended to mimic human behavior and may be applied to various applications to help with the COVID-19 situation [6].

Traditional object detection methods have drawbacks: limited recognition accuracy, calculation, and a sluggish running speed; many good recognition outcomes may occur [7], [8]. This study tries to limit the effects of the COVID-19 outbreak while causing as little disruption as possible to economic artifacts. Furthermore, this research offers an efficient automatic surveillance system that facilitates each individual's location and observes them for the social distancing factor. This program may be employed for both outdoor and indoor monitoring. Malls, streets, megastores, airports, railway stations, and other locations are where it may be deployed. This research is inducted primary contributions, which are as follows:

1.1. Human Tracking and Detection

The unique YOLOv5 algorithm presented for person monitoring and detection in surveillance videos is the first contribution. Human detection refers to a single-step process that begins with the localization and classification of an object.

1.1.1. Efficient System Monitoring of Social Distancing Detection

The Main contribution is to determine distance among persons in public settings utilizing our suggested method. If obeyed, the choice is based on social distancing. Or else, a red rectangle is drawn around the people who do not satisfy the social distancing requirements. Section 2 intends to thoroughly examine both conventional and modern methods to various human detection methods. Section 3 centers on human detection models based on deep learning (DL). Section 4 is devoted to the experiments and their extensive analysis. Finally, in section 5, the conclusion is expressed, accompanied by the upcoming scope.

2. RELATED WORKS

The Sliding Window concept [9] to locate things inside photos. Applying this procedure, an image is broken into particular regions or block sizes. Additionally, these blocks are categorized into numerous classes. This solution leverages the single shot multibox detector (SSD) model to distinguish humans in an image or video. It then puts a red line as an alert on folks whose distances are shorter than the baseline. The technology is applied for real-time social distancing monitoring, giving a mean average precision (mAP) of 88.44 percent [8]. Using YOLOv3 object recognition methodology, the approach recognizes humans in video sequences. Additionally, transfer learning was utilized to improve the accuracy of the model-related while detection procedure, an algorithm that has been previously a trained algorithm was associated with an additional trained layer by utilizing an above human data set. Using the obtained bounding box data and the Euclidean distance algorithm, the detection model calculated the distance between the individuals. Consequently, the detection model obtained 98 percent accuracy with transfer learning and 92 percent accuracy without it [10].

Here, the scholars aided authorities in redesigning public spaces or taking preventative measures to minimize high-risk zones. They employed the YOLOv3 method as a pre-trained model, and the dataset was pre-recorded video as input. It may also be applied in other sectors, for instance, crowd analysis, human action recognition, and autonomous vehicles. Authors [11] utilized machine learning and object detection to spot face masks and social distance in a video feed. A convolutional neural network (CNN) model for detecting facemasks was built utilizing TensorFlow and Keras and trained on a dataset of 3800 images. Using YOLO Object detection, the Euclidean distance between the centroids of the recognized boxes was determined to recognize persons in a frame and test for social distancing [12]. For pedestrian identification, further research used the YOLOv3 algorithm. After that, the video frame was changed to a top-down perspective so that distances could be calculated in the two-dimensional (2D) plane.

A red frame and a red line represent any non-compliant pair of people in the presentation. In addition, a prerecorded film of people walking along the street was used to evaluate the strategy. The approach utilized to establish social distance measures amongst various individuals in the film is demonstrated in conclusion. The traditional system was designed as a real-time detecting instrument [13]. A method of centroid tracking tracked the people throughout the clip. A Mask R-CNN deep neural network was built to distinguish individuals in a video frame to determine if social distance occurred during an encounter between individuals [14]. A framework for effective object classification and localization must identify many objects of differing sizes inside an image. Additionally, it should lower computing costs and false-positive rates. Deep CNN-based object detection has achieved substantial advancements [15]. CNNs are an intensive, feed-forward artificial neural network (ANN) used for categorizing and recognizing pictures in computer vision tasks. CNN can generate robust features by utilizing the convolution technique. The ref. [16] presented a hierarchical model with a large capacity for attribute encoding [17] for identifying human upper body types. This model accommodates local and contextual picture input with an accuracy of up to 86% using a candidate-region (CR)-CNN and several convolutional features.

Recently, a one-stage approach for a social distance detector for monitoring pre-strain, was proposed [9][18]. According to reports, the YOLOv3 approach provided an on-stage solution to identify and track pedestrian distance. The YOLOv4 was used to utilize COCO dataset measurements for real-time social space surveillance. The results revealed that the YOLOv4 algorithm achieved low light power with 97.84 percent leading average precision, the observed primary square error value, and the frame per the second score. The YOLOv4 model was implemented using GPU-based tweaking [19].

Nonetheless, the outcomes reveal limited accuracy and detection. The literature demonstrates that object detection is crucial in computer vision due to the diversity of applications it may be used, including medical imaging, activity recognition, and pedestrian detection, face detection, etc. The significance of object identification in lowering the separation of COVID-19 has been increased by an algorithm that employs the development of an effective object detector to measure the social distance between individuals.

3. MONITORING THE SOCIAL FOR DISTANCING SYSTEM

The proposed the entire context of public monitoring for social distance is depicted in Figure 2. Using image/frame sequences acquired from the COCO dataset, the YOLOv5 object detection module for detecting the presence of people in a setting is characterized by trained using frames. For measuring the level of social separation, metrics such as 'centroid' the object or person position and 'distance' between the evaluation of several centroids. Change the color of the identified human box from green to red to generate a warning. Green is the bounding box's color up until a proper separation between two people is reached. The enclosing boxes turn red to denote a violation of social space as this value declines. Sliding-window-based region proposals are an easy method for developing an efficient object detector.

Change the color of the identified human box from green to red to generate a warning. Until a proper distance separates two people, the bounding box is green. The red tint of the bounding boxes denotes a violation of social space when these lowers. A simple method for creating a trustworthy object detector is to use sliding window-based region proposals. This technique divides the image or frame into blocks or regions of a certain size. Moreover, these blocks have been classified are categorized according to their distinct classes. There are numerous machine learning and deep learning paradigms [20]. May be used to classify blocks. Additionally, regions may contain a piece of the item, resulting in many bounding boxes surrounding it. To address this issue, the Non-Maximum Suppression approach [21] is employed to correctly locate the thing within image by suppressing inferior bounding boxes and keeping just the best.

After monitoring social distancing capture each frame, there is a violation of the distance, and a frame is selected and sent to the respondent's email. this paper demonstrates a deep learning-based technique for detecting human presence. The proposed approach for object recognition and localization, which is detailed in Section 3.1, is highly useful. In addition, these strategies are utilized in the algorithm for social separation to determine if individuals adhere to the distance requirement.

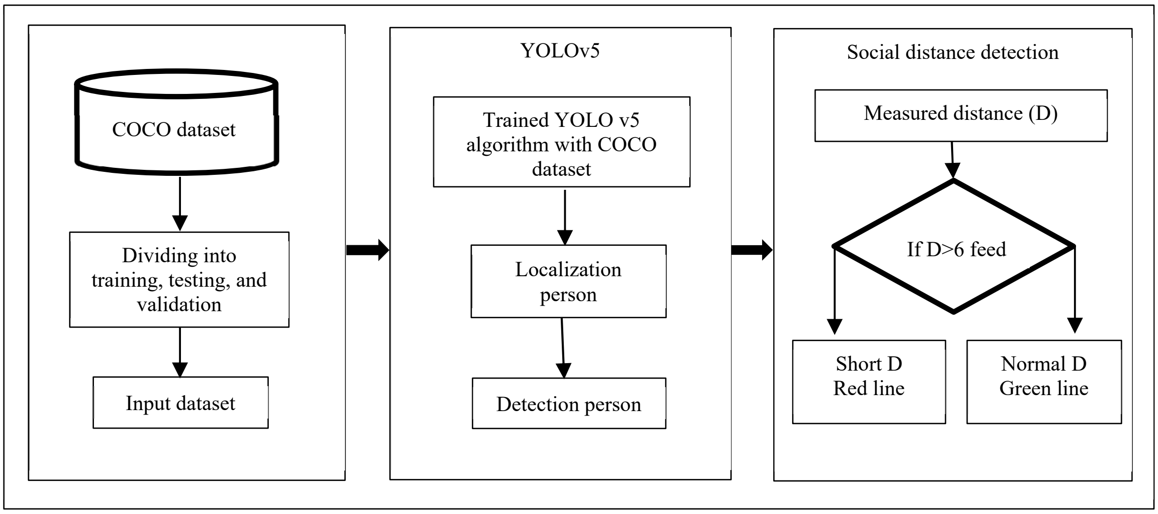

Figure 2. Monitoring the system of social distance

4. PROPOSED YOLOV5 ALGORITHM

YOLOv5 is comprised of YOLOv1-YOLOv4 and is a sophisticated real-time object detector; Figure 3 depicts YOLOv5's network architecture. Three factors led to the selection of YOLOv5 as the first learner. Cross-stage partial network (CSPNet) and Darknet were first combined by YOLOv5, with CSPDarknet serving as the network's core. By including gradient information variations in the feature map, reducing model parameters, and increasing floating-point calculations per second (FLOPS), CSPNet overcomes the challenge of gradient information for large-scale backbones that is repeated. This ensures inference speed and accuracy while reducing model size. Accuracy and speed are essential for identifying forest fires, and the model's height impacts its inference performance on resource-constrained edge devices. As its neck, YOLOv5 employs a (PANet) to improve the transmission of information. By combining an improved bottom-up approach with a revolutionary feature pyramid network (FPN) architecture, PANet is able to boost low-level feature transmission concurrently, an adaptable trait. Additionally, adaptive feature pooling ensures that relevant data from each feature level is sent to the subsequent sub network.

Additionally, PANet enhances precise localization signals at lower levels can considerably enhance the positioning accuracy of an object. The YOLO layer, which serves as the head of YOLOv5, establishes three different sizes of feature maps (18 18, 36 36, and 72 72) for multiscale prediction, allowing the model to manage small, medium, and big objects. A forest fire, for instance, often progresses from a small (ground fire) to a medium (trunk fire) to a huge (forest fire) blaze (canopy fire). The model's multiscale detection capability enables it to monitor variations in size as the fire spreads [22].

The network is segregated into three sections: the backbone, the neck, and the output. The neck portion combines the obtained feature information and produces three scales of feature maps after the backbone section extracts feature information from the input images. Finally, the output portion uses the feature maps developed to identify objects.

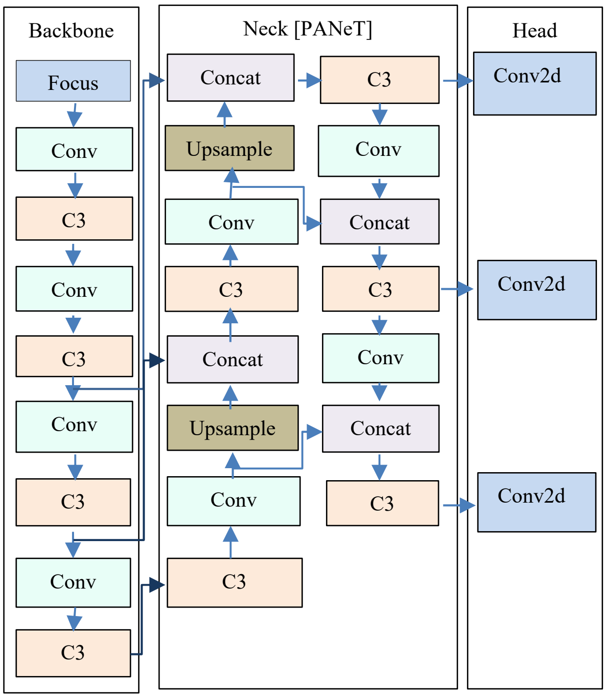

Figure 3. YOLOv5 Method's Architecture [23]

4.1. Distance Calculation

Once the individuals have been located, their bounding box is obtained. After detection, the Euclidean distance is determined. Multiply the observed length by the scaling factor to convert space to conventional units such as feet. If d denotes the Euclidean distance, which is 6 feet for real-time social distancing, then the following Equation holds:

$$ED = 6\,feet(RLM) \quad (1)$$

$$ED = (6 / D)\,feet \quad (2)$$

where, *ED* is Euclidean Distance, *RLM* is Real Life Metric, and *D* is the distance.

5. EXPERIMENTS AND RESULTS

This section shows how YOLOv5 enables an object detection and tracking system for pedestrians with perfect performance. Compare the YOLOv5 object detection algorithm with the last versions of the YOLO algorithm. The simulation shows accurate detection for people. The red color inducted distance between two or more persons closely. Moreover, Python programming is used for object detection training: a TensorFlow-based version of YOLOv5.

5.1. MS COCO Dataset

The Microsoft Common Things in Context (MS COCO) collection [24] for recognizing and segmenting common objects in their natural settings includes 91 common object categories, 82 of which include over 5,000 annotated instances. The twenty categories included in the PASCAL VOC dataset are represented below. The collection consists of 2.5 million labeled samples in 328,000 pictures. Additionally, the MS COCO dataset incorporates multiple perspectives, and all items are in their native settings, providing a wealth of contextual information. COCO contains fewer categories than ImageNet, but more examples per category [25]. Detecting small objects takes stronger contextual reasoning, as is common knowledge. The MS COCO image collection contains a wealth of contextual data. The largest class is "human" with over 800,000 instances, and the smallest is "hair driver" with approximately 600 instances. There are around 800 instances of the class "hairbrush." The difficulty of the remaining 71 categories is equivalent, with the exception of 20 categories having a big or small number of examples.

5.2. Findings and Discussion

This section analyses the findings and compares the suggested framework to current methodologies. For the object detection model, an influential computer vision YOLOv5 is utilized. Three major evaluation factors, inference time, mean average precision (mAP), and framerate, are being used to assess the performance of the recommended version of YOLOv5 while employing a standard view, the application of the YOLOv5 model is continuously observed during the training phase utilizing the mAP with the error and an overall decrease in the detection of the individual [26].

5.2.1. Framerate

This metric is used in computer vision for object detection to track processing performance in frames per second (FPS). In addition, it is another approach for determining the processing speed of the model. Figure 4 compares the performance of the proposed YOLOv5 algorithm to that of two state-of-the-art models, the YOLOv3 and YOLOv4 algorithms. The proposed approach (YOLOv5) performs admirably with (140 fps). The two algorithms (YOLOv3 and YOLOv4) are close behind with (50 and 20 fps), respectively.

5.2.2. Mean Average Precision (mAP)

A commonly employed statistic for evaluating the precision of object detection models. Similarly, it seeks to obtain the average precision for recall values between 0 and 1. Similarly, precision is used to quantify the accuracy of the model's predictions, whereas recall measures the trained model's ability to predict positive samples. Conditions of mAP on a common benchmark dataset with dim and normal lighting. The trial results indicate that it operates admirably with a balanced mAP score (0.57). Figure 5 demonstrates that YOLOv3 and YOLOv4 appeared to have good detection with 0.58 and 0.66 of mAP, respectively.

5.2.3. Training Time

Training time is a crucial component in measuring the complexity of algorithms in machine learning and deep learning. Our model (YOLOv5) outperforms its predecessors (YOLOv3 and YOLOv4) based on the training time required by an algorithm for developing a deep learning model. 14 days and 50k iterations were used to train the YOLOv3 algorithm, whereas 10 days and 1300 iterations were used to introduce the YOLOv4 algorithm. As depicted in Figure 6, the considerable training time with the YOLOv5 method requires 3.5 hours and 1300 iterations due to its low complexity.

5.2.4. Inference Time

Assessment measure aids in determining the predictive performance of a model that has already been trained. The trained model's inference time should be as short as feasible, and the accuracy of the previously trained model. As demonstrated in Figure 7, YOLOv3 and YOLOv4 may be utilized to identify pedestrians on the MS COCO Dataset, with YOLOv5 performing better. Inference time is also evaluated to prevent any mistakes. The inference time outcomes of the suggested YOLOv5 have two YOLOv3 and YOLOv4 models. Again, the inference time must be as short as feasible, and it is observed that the inference time of enhanced YOLOv5 is the fastest, while YOLOv4 is further behind. Upon comparing the upgraded version of YOLOv5 with normal view vision, the suggested model outperforms the YOLOv4 and YOLOv3, indicating that the proposed technique is more effective than the suggested technique to the current state-of-the-art models.

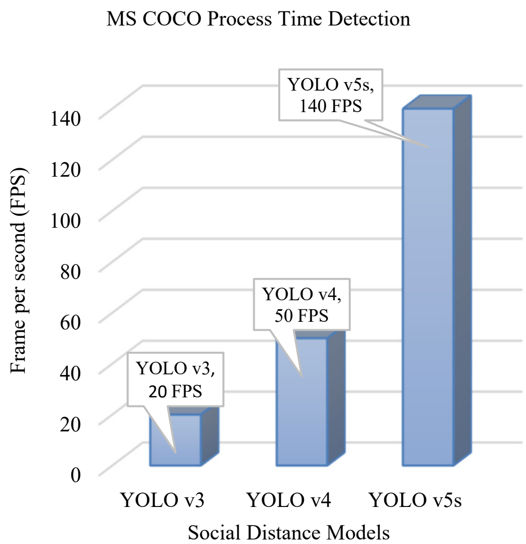

Figure 4. FPS measurement

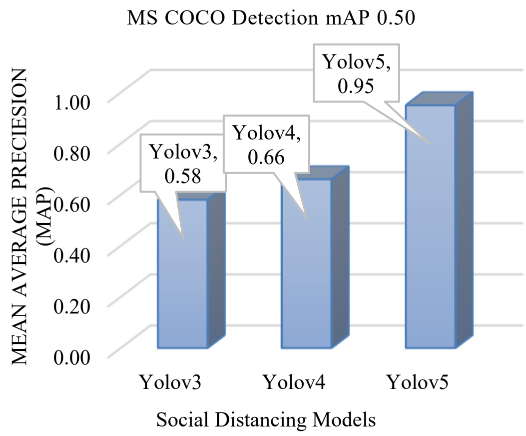

Figure 5. mAP measurement

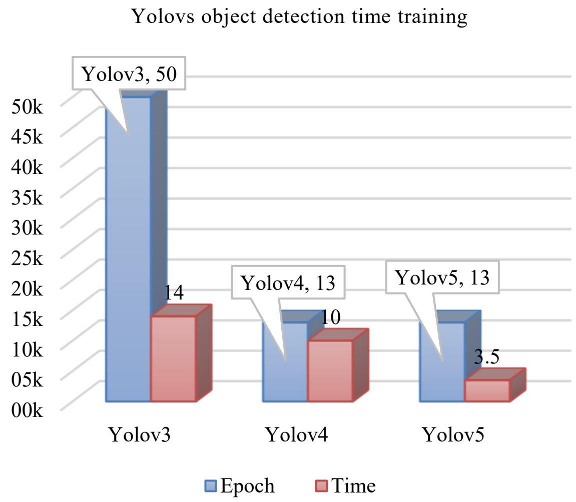

Figure 6. Time training measurement

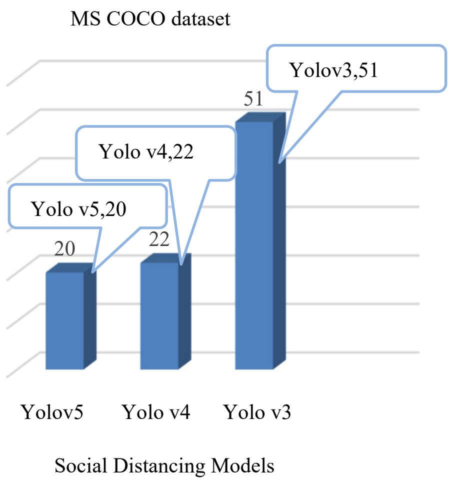

Figure 7. Inference time measurement

6. CONCLUSIONS

The study devised an efficient real-time object detection by automated method for monitoring social distancing, whereby anyone may be identified in real-time utilizing anchor boxes. The produced bounding packages assist the identification of assemblages of people satisfying the proximity property as defined by a pairwise vectorized technique. Calculating the number of groups generated to confirm the number of infractions. By dividing the number of individuals by the number of groups, the violation index is calculated. YOLOv5 outperforms both YOLOv3 and YOLOv4 in terms of process detection and inference time. Yolov5 is comparable to YOLOv3 and YOLOv4 in mean average precision since it simultaneously performs end-to-end training, detection, and classification. Due to the sensitivity of this method to the camera's spatial positioning, the equal procedure can be modified to better adapt to the matched field of view.

REFERENCES

[1] A. Fitwi, Y. Chen, H. Sun, R. Harrod, "Estimating Interpersonal Distance and Crowd Density with a Single-Edge Camera", Computers, Vol. 10, No. 11, pp. 1-23, 2021.
[2] C. Sun, Z. Zhai, "The Efficacy of Social Distance and Ventilation effectiveness in preventing COVID-19 Transmission", Sustain. Cities Soc. J., Vol. 62, pp. 1-10, 2020.
[3] A.P. Balasa, "COVID 19 on Lockdown, Social Distancing and Flattening the Curve - A Review", Eur. J. Bus. Manag. Res., Vol. 5, No. 3, pp. 3-6, 2020.
[4] Y.C. Hou, M.Z. Baharuddin, S. Yussof, S. Dzulkifly, "Social Distancing Detection with Deep Learning Model", The 8th Int. Conf. Inf. Technol. Multimedia (ICIMU), pp. 334-338, 2020.
[5] X. Zou, "A Review of Object Detection Techniques", Proc. - 2019 Int. Conf. Smart Grid Electr. Autom., (ICSGEA), pp. 251-254, 2019.
[6] A. R. Pathak, M. Pandey, and S. Rautaray, "Application of Deep Learning for Object Detection," Procedia Comput. Sci., vol. 132, no. Iccids, pp. 1706–1717, 2018.
[7] S. Saponara, A. Elhanashi, A. Gagliardi, "Implementing a Real-Time, AI-based, People Detection and Social Distancing Measuring System for Covid-19", J. Real-Time Image Process., Vol. 18, No. 6, pp. 1937-1947, 2021.
[8] J. Qin, N. Xu, "Research and Implementation of Social Distancing Monitoring Technology Based on SSD", Procedia Computer Science, Vol. 183, pp. 768-775, 2021.
[9] M.M. Ozbek, M. Syed, I. Oksuz, "Subjective Analysis of Social Distance Monitoring Using YOLOv3 Architecture and Crowd Tracking System", Turkish J. Electr. Eng. Comput. Sci., Vol. 29, No. 2, pp. 1157-1170, 2021.
[10] L. Morawska, et al., "How Can Airborne Transmission of COVID-19 Indoors be Minimised?", Environ. Int., Vol. 142, pp. 1-7, 2020.
[11] G. Kumar, S. Shetty, "Application Development for Mask Detection and Social Distancing Violation Detection using Convolutional Neural Networks", www.scitepress.org, Vol. 1, No. Iceis, pp. 760-767, 2021.
[12] S. Meivel, et al., "Mask Detection and Social Distance Identification Using Internet of Things and Faster R-CNN Algorithm", Comput. Intell. Neurosci., Vol. 2022, pp. 1-13, 2022.
[13] S. Yadav, P. Gulia, N.S. Gill, J.M. Chatterjee, "A Real-Time Crowd Monitoring and Management System for Social Distance Classification and Healthcare Using Deep Learning", J. Healthc. Eng., Vol. 2022, pp. 1-11, 2022.
[14] S. Gupta, R. Kapil, G. Kanahasabai, S.S. Joshi, A.S. Joshi, "SD-Measure: A Social Distancing Detector", The 12th Int. Conf. Comput. Intell. Commun. Networks (CICN), pp. 306-311, 2020.
[15] D. Chahyati, M.I. Fanany, A.M. Arymurthy, "Tracking People by Detection Using CNN Features", Procedia Comput. Sci., Vol. 124, pp. 167-172, 2017.
[16] A. Zhu, T. Wang, T. Qiao, "Multiple Human Upper Bodies Detection via Candidate-Region Convolutional Neural Network", Multimed. Tools Appl., Vol. 78, No. 12, pp. 16077-16096, 2019.
[17] S.A. Agnes, J. Anitha, S. Pandian, J.D. Peter, "Classification of Mammogram Images Using Multiscale all Convolutional Neural Network (MA-CNN)," J. Med. Syst., Vol. 44, No. 1, pp. 1-9, 2020.
[18] K. Bhambani, T. Jain, K.A. Sultanpure, "Real-Time Face Mask and Social Distancing Violation Detection System Using YOLO", IEEE Bangalore Humanitarian Technology Conference (B-HTC), pp. 1-6, 2020.
[19] A. Rahim, A. Maqbool, T. Rana, "Monitoring Social Distancing Under Various Low Light Conditions with Deep Learning and a Single Motionless time of Flight Camera", PLoS One, Vol. 16, No. 2, pp. 1-19, February 2021.
[20] R. Esmaeilzadeh, A. Roshan, "Efficient Electric Price Forecasting Using Neural Networks", Int. J. Tech. Phys. Probl. Eng., Vol. 9, No. 3, pp. 30-34, 2017.
[21] J. Hosang, R. Benenson, B. Schiele, "Learning Non-Maximum Suppression", The 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, Vol. 2017-Janua, pp. 6469-6477, 2017.
[22] R. Xu, H. Lin, K. Lu, L. Cao, Y. Liu, "A Forest Fire Detection System Based on Ensemble Learning", Forests, Vol. 12, No. 2, pp. 1-17, 2021.
[23] U. Nepal, H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs", Sensors, Vol. 22, No. 2, p. 1-15, 2022.
[24] G.T.U.A. Colleges, et al., "Microsoft COCO", Eccv, No. June, pp. 740-755, 2014.
[25] O. Russakovsky, et al., "ImageNet Large Scale Visual Recognition Challenge", Int. J. Comput. Vis., Vol. 115, No. 3, pp. 211-252, 2015.
[26] V. Greu, P. Ciotirnae, I. Sima, F. Popescu, "A New Algorithm for Optimizing IEEE 802.11 Network Based on Interference Activity Evaluation Criteria", Int. J. Tech. Phys. Probl. Eng., Vol. 7, No. 4, pp. 55-61, 2015.

BIOGRAPHIES

Hiyam Hashem Saeed was born in Baghdad, Iraq in March 1984. She received the B.Sc. in computer science from University of Baghdad, Baghdad, Iraq in 2006. Currently, she is pursuing the M.Sc. degree from University of Diyala, Diyala, Iraq. Her research interests include the applications of artificial intelligence, deep learning, and object detection.

Abdulbasit Alazzawi was born in Baghdad, Iraq in July, 1974. He received the B.Sc. degree in computer science from University of Baghdad, Baghdad, Iraq and the M.Sc. degree from University of Technology, Iraq. He completed his Ph.D. from Altinbas University, Istanbul, Turkey. He is still working in the same field as Asst. Prof. His research interests include advanced face recognition system, artificial intelligence, linear regression slope.