DeepDist: A Deep-Learning-Based IoV Framework for Real-Time Objects and Distance Violation Detection

Yesin Sahraoui, Chaker Abdelaziz Kerrache, Ahmed Korichi, Boubakr Nour, Asma Adnane, and Rasheed Hussain

Abstract

Crowd management systems play a vital role in today's smart cities and rely on several Internet of Things (IoT) solutions to build prevention mechanisms for widespread viral diseases such as Coronavirus 2019 (COVID-19). In this article, we propose a framework to aid in preventing widespread viral diseases. The proposed framework consists of a physical distancing notification system by leveraging some existing futuristic technologies, including deep learning and the Internet of Vehicles. Each vehicle is equipped with a switching camera system through thermal and vision imaging. Afterward, using the Faster R-CNN algorithm, we measure and detect physical distancing violation between objects of the same class. We evaluate the performance of our proposed architecture with vehicle-to-infrastructure communication. The obtained results show the applicability and efficiency of our proposal in providing timely notification of social distancing violations.

Introduction

With the tremendous growth in the world's population, there is a tendency to gather in communities for common interests such as workplaces, restaurants, music festivals, sport events, religious gatherings, and more. As a consequence, populations are vulnerable to many natural and human-made disasters, especially highly infectious and fast-spreading diseases similar to what we have witnessed in the coronavirus pandemic (COVID-19), which spreads quickly and more efficiently from person to person through respiratory droplets when an infected person coughs, sneezes, or talks. If the droplets land on objects and/or surfaces such as tables or door handles, they also become the cause of infection of other people touching these objects or surfaces, then touching their eyes, noses, or mouths before cleaning their hands.[1] In these situations, crowd management and crowd control became key measures in slowing down the spread of the virus.

Crowd management strategies have become an integral part of smart cities, for managing, planning, and monitoring crowds. To this end, deep learning (DL) approaches such as convolutional neural networks (CNNs) have emerged as a new exciting research area of machine learning for identifying and classifying images' objects, and hence, these solution can be very effective when used for crowd management. In addition, most major crowd disasters can be prevented by crowd management strategies such as physical distancing, which aims to avoid critical crowd densities [1].

The World Health Organization (WHO) has adopted the term *physical distancing* instead of *social distancing* or *isolation* [2], since the latter is damaging economies and causes psychological effects. In contrast, the former maintains a physical distance more than three feet without breaking community contact. Physical distancing is an alternate way to fight against the spread of COVID-19 until a vaccine is available.

In order to slow down the spread of this pandemic, many video surveillance camera systems are deployed in streets to detect physical distancing violations [3]. However, there are a number of challenges in monitoring physical distancing in public areas. Although street cameras might be installed in public places (which is the case in many cities), it has some issues such as:
- Camera installation incurs high expenditure.
- The absence of any pedestrian on some corners makes the camera useless.
- There is a lack of a real-time warning mechanism that allows safe crossing by practicing physical distancing.

The high mobility and cooperation of the Social Internet of Vehicles (SIoV) is one of the most promising technologies to address the challenges mentioned above. SIoV is an improvement of IoV technology advances by applying IoV conceptsand adding the social dimension to it, and makes vehicles, drivers, and passengers able to interact with each other, leveraging different devices such as cellular networks, sensors, and smartphones, and interconnects people within and around vehicles. In SIoV, vehicles can communicate/socialize with each other by sharing information among a cluster of cars, including the traffic situation in an area, accident warning, and so on.

In this article, we propose a framework based on DL using SIoV namely "DeepDist," to help practice physical distancing. We assume that each vehicle is equipped with thermal and vision imaging camras. Using the Faster Region-based CNN (Faster R-CNN) algorithm, these vehicles can detect objects in real time with a red bounding box if the physical distancing among them is violated. Afterward, an automatic alert is sent to pedestrians and vehicles through advertisement boards and smart dashboards, respectively. To the best of our knowledge, this is the first attempt at using SIoV to detect physical distancing and alert about violations through a wide selection of advertising boards.

In summary, the major contributions of this article are as follows:
- Accurately predict whether social and safety distance among pedestrians and inter-vehicle are respected using SIoV.
- Make optimal planning decisions for crowd management using advertising boards and vehicle dashboards.

The rest of this article is organized as follows. We review related works. We present an overview of Faster-RCNN. The proposed approach is described. The related performance evaluation is presented. Finally, we provide some future directions and conclude the article.

Related Work

In this section, we review existing works on object detection and how the distance between similar objects is measured using DL techniques.

Measuring Physical Distance

Kelso *et al.* [4] simulated the effect of four different social distancing intervention measures in the potential control of a future influenza pandemic. The results show that such measures
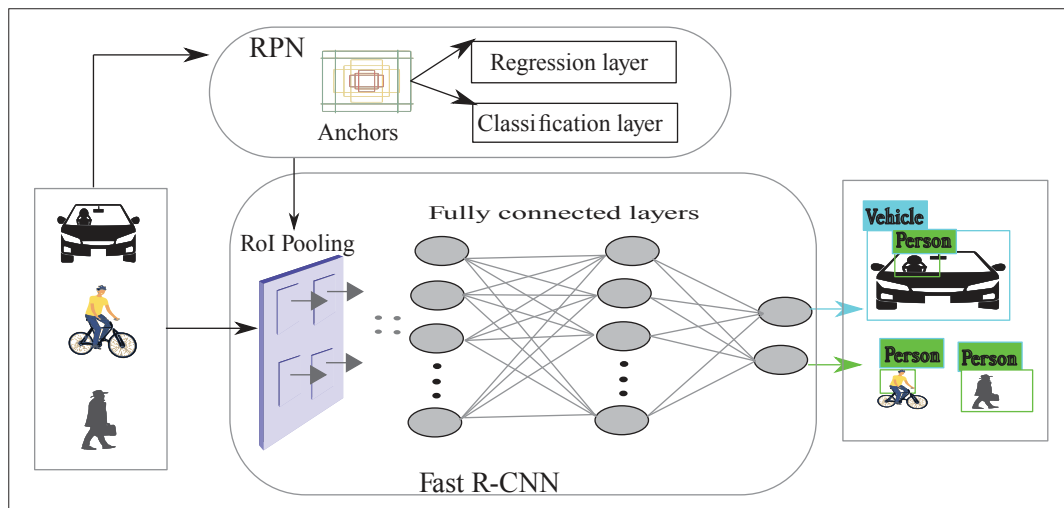
FIGURE 1. The general architecture of Faster R-CNN.

are capable of stopping epidemic development, but only if they are used relatively together for a long time and without delay, while the imposition of such measures would affect personal freedom.

Zhang et al. [5] proposed a method to improve safety on roads, and to assist with blind-spot monitoring during the reversing process, combining DL using the R-CNN algorithm and binocular vision technology for identifying and measuring the distance of traffic obstacles in real time from different angles. In this study, some problems in the detection process need to be solved, including the detection of various types of obstacles as well as the target's details like speed, direction, and angle to get more accuracy in the detection process.

Zaarane et al. [6] proposed an algorithm based on a stereo vision method to detect and measure inter-vehicle distance. The designed method is divided into four steps: The first one starts with capturing images, followed by an object detection step using AdaBoost classifier; then the stereo matching process is applied, and the distance is estimated in the final step. This system showed high accuracy in distance measurement. However, applying a detection algorithm to only one image of a vehicle makes this method very slow. In addition, in the multi-object environment, the problem of the same object selection appearing in both cameras is still open.

Huang et al. [7] proposed a real-time inter-vehicle distance estimation on urban/suburban roads. The authors exploited the vanishing point through a single-lens video camera. The results show the effectiveness of the proposed method. Similarly, Sasaki et al. [8] developed an inter-vehicle distance measurement system through a smartphone camera based on the similarity relationship. The main drawback of this framework is the considerable error rate, which can be addressed by increasing image resolution, but it requires more more processing time.

OBJECT DETECTION AND RECOGNITION

Recently, several papers have extensively treated object detection using DL algorithms and computer vision. For instance, Wagner et al. [10] introduced two architectures for pedestrian detection on the basis of multi-spectral image data, by combining the images of a visible and thermal camera. The first architecture is for early and the second for late fusion, and they are both built upon the R-CNN algorithm. The results of the pre-trained late-fusion-based model show good performance in terms of accuracy. However, the main weakness of this algorithm is related to its slow process.

Punn et al. [10] presented a real-time framework based on DL, the YOLO v3 algorithm, to detect pedestrians and monitor social distancing violations. The experimental analysis of this algorithm gives efficient performance in terms of accuracy and precision, but this algorithm needs a huge dataset for the training phase.

Aqib et al. [11] proposed an emergency evacuation framework, based on DL and big data analytics to predict traffic behavior in emergency and evacuation situations in smart cities using historical traffic data. The correlation between the input and output data gave very high accuracy, but high uniformity in input data may not necessarily give the same performance with other datasets.

FASTER R-CNN OVERVIEW

This section provides an overview of the Faster R-CNN algorthim, its architecture, training, and loss functions.

ARCHITECTURE

Convolutional neural networks are among the most important DL solutions that are commonly used in computer vision applications and visual analysis imagery. There are many available object detection algorithms, but Faster R-CNN is the one offering a good accuracy and cost trade-off [12]. Faster R-CNN aims to find a solution to improve the slow running problem of some algorithms like Region CNN (R-CNN) and Fast R-CNN.

As shown in Fig. 1, Faster R-CNN is one of the most important object detection algorithms that relies on CNN such as You Look Only Once (YOLO) and Single Shot Detector (SSD). Its object detection process is mainly divided into two parts:

- **Part 1:** The region proposal network (RPN) as a region proposal algorithm. It uses anchors as reference boxes for generating proposals. At the regression layer level, the box parameters are predicted for all proposals, while predicting the possibility that the proposals are an object/background using the classification layer, and then define the object with a bounding box.
- **Part 2:** Fast R-CNN as a detector. The proposal regions provided by RPN are passed to the Fast R-CNN algorithm via the region of interest (RoI) pooling layer to reduce the feature maps. Then the output from the RoI pooling layer is passed through fully connected layers, which in turn feed the softmax and regression layer for classification and bounding boxes prediction, respectively.

TRAINING AND LOSS FUNCTIONS

The RPN training process begins by randomly pulling an image from the training dataset and specifying a label to each anchor to form a mini-batch that contains 256 anchors; hence, the loss value is minimized. Afterward, the image is included in RPN as
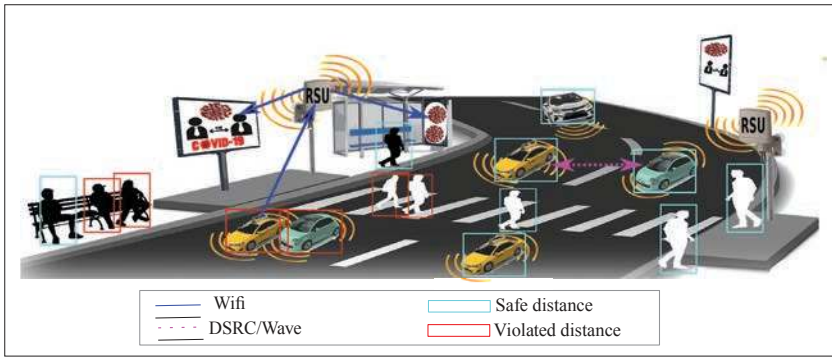
FIGURE 2. General design of our proposed approach.

an input to predict proposals using anchors, then measure and propagate the loss value to update the convolutional kernels rate.

There are two losses in Faster R-CNN: one is classification (category) loss L_{cls}" and other is regression (bounding boxes location) loss L_{reg}, as shown in the equation below [12].

$$L(p_i t_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$$

i	the anchor index.
λ	a balancing parameter.
p_i	the probability of predicting the object.
p_i^*	the ground-truth label, $p_i^* = \begin{cases} 1 \text{ if object} \\ 0 \text{ if background} \end{cases}$
t_i	the predicted bounding box coordinates vector.
t_i^*	the ground-truth box if (+) anchor.

SYSTEM FUNCTIONALITIES

In our proposal, we combine the visible and thermal camera, which is used for much DL-based pedestrian detection and vehicle detection. The use of thermal cameras is motivated by the fact that they are not affected by darkness and bad weather conditions, whereas the reflected temperature influences the accuracy of its measurement, unlike the visible camera, which performs well in daylight, but its performance is affected at night and in low illumination conditions. To overcome the disadvantages of both, it is useful to switch between the two cameras wisely.

Our approach differs from the existing works in the fact that we also involve the highly mobile SIoV, which can cover the whole city rather than only small areas. As mentioned above, we consider vehicles equipped with two fixed cameras, one thermal and the other visible in order to track both pedestrians and vehicles. At the same time, using Faster R-CNN, the system also monitors the physical distance and generates alert messages in the case of distance violations, as shown in Fig. 2.

The three following cases are observed in our model.

Case 1: When vehicles detect physical distancing violations between pedestrians, it generates and sends real-time alerts to the nearby roadside units (RSUs), which in turn incite the closest advertisement boards to exhibit their HD videos/images on the display screen and indicate the necessity of physical distancing for public health and practicing it well.

Case 2: To take advantage of the object identification and distance measurement between similar objects, our proposal also identifies the safety distance violation among vehicles moving in parallel lines. Once the safety distance violation is detected, an event-triggered Decentralized Environmental Notification Message (DENM) is broadcast with a distance violation alert that includes the position of the concerned vehicles [13].

Figure 3 illustrates exactly where we propose to include the safety distance violation alert in the situation container of the DENM

Case 3: We can also leverage a cloud-based system to send alert messages and their localization via vehicle-to-broadband cloud (V2B) communication to the central platform, which counts the total number of these messages in each area and displays it using graphical representation such as a heat map. A localization with intensive alert messages is considered a crowded area. Therefore, the user can have prior knowledge remotely if an area, a road, or a bus stop is crowded or not.

In our design, the data collected in a geographic area are forwarded to an edge computing server where they are temporarily stored, processed, and accessed by interested consumers, (police, health department inspectors, etc.).

According to the 3rd Generation Partnership Project Vehicle-to-Everything (3GPP-V2X) specification in [14], a V2X application server can be implemented according to the multi-access edge computing (MEC) paradigm to support V2X applications. By providing storage and computing resources close to where the data is produced, the V2X application server ensures that data is processed in real time at the periphery of the network, thus also limiting the traffic load in the core network [15].

In addition to traditional road traffic applications, the V2X application server can be expanded to support a variety of IoV applications, including those for epidemic detection.

Of course, several V2X application servers can be deployed in different geographic areas, and their storage and processing resources can be sized according to population density. Depending on their role, interested consumers can access the data from a single edge server or multiple servers. For example, health department inspectors working in a specific area will only access data from that area; on the other hand, if the Ministry of Health is interested in a global map of suspected infections, data from all V2X application servers will be accessible.

PERFORMANCE EVALUATION

Our simulation is carried out using two open source frameworks, Network Simulator (NS-3), which is a discrete-event network simulator, and the Simulation of Urban Mobility (SUMO) to generate realistic mobility behavior to vehicles.
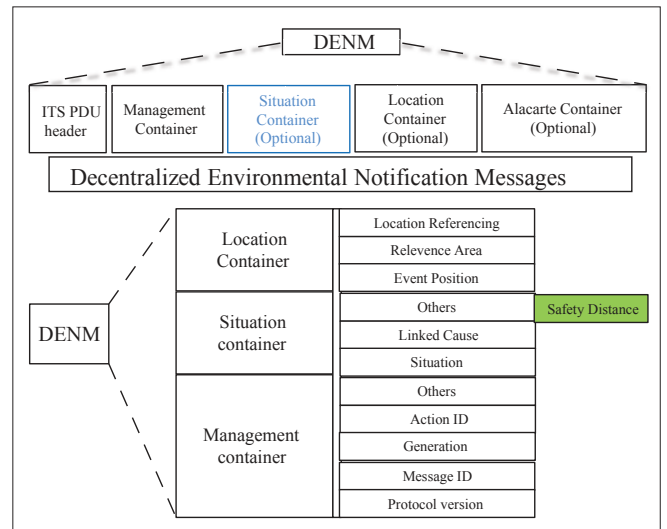

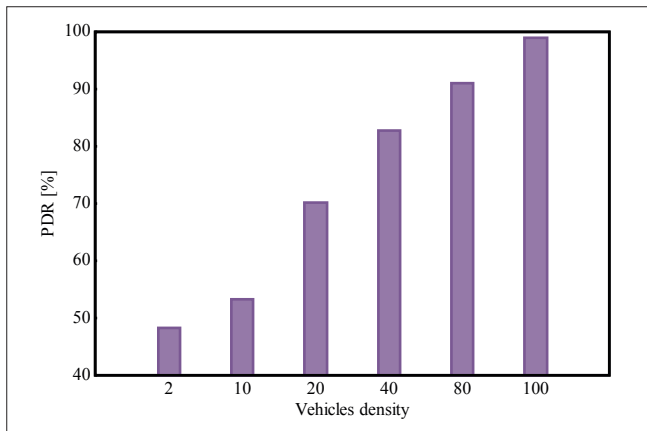FIGURE 3. Extended DENM to support safety distance violation alerts.

FIGURE 4. Average PDR variation according to vehicles density.

FIGURE 5. Mean E2E delay variation according to vehicles density.

FIGURE 6. Precision/Recall (PR) curve for Faster R-CNN.

The vehicles are equipped with 802.11p communication interface, which is considered as an approved amendment to the IEEE 802.11 standard and is the standard of vehicle-to-vehicle and vehicle-to-Infrastructure communications. We assume that each vehicle sends an alert packet to the nearby RSUs, which in turn forward it to the close advertising boards to display HD alert images from its stock once a physical distancing violation is detected.

We evaluate the quality of service (QoS) metrics against a varying vehicular density from 2 to 100 using TCP traffic pattern and two rays ground as a propagation model.

Obtained results of our simulation are presented by measuring different QoS metrics such as end-to-end delay, packet delivery, and precision and recall, which can be defined as follows:

- **End-to-end delay** is the total time required for a packet to reach its destination.
- **Packet delivery ratio** is the rate of packets successfully transmitted over the total transmitted.
- **Precision and recall** are used to measure the accuracy of information retrieval, classification, and identification within a computer program.

Figure 4 shows that the packet delivery ratio result obtained is inversely proportional to the vehicles' density; because of the congestion, packets are dropped. We can observe from Fig. 5 that the IEEE 802.11p standard offers end-to-end delays less than 104 ms despite the change in vehicles' density.

Besides the QoS-related metrics, precision and recall are the two most important metrics to evaluate model accuracy, classification, and identification. Precision is defined as the accuracy of the classifier to find all the instances of the positive class, while recall detects the largest possible number of positive cases.

A large number of existing studies have examined the object detection in different classes using Faster R-CNN and calculate precision/recall (PR) to show up the sensitivity of the detector. Figure 6 shows the PR curve using Stanford Vehicles' Dataset; this algorithm achieves a good mean Average Precision (mAP) value, which approximately reaches 0.76.

CONCLUSION AND FUTURE DIRECTIONS

In order to avoid physical distancing violations, in this article, we design a framework based on deep learning and the SIoV paradigm to track objects using the Faster R-CNN algorithmthrough a switching cameras system and detecting the people or vehicles violating physical distancing. In this architecture, we can alert about violations using vehicle-to-infrastructure communication by leveraging advertisement boards, or directly via vehicle-to-vehicle communication. In future work, we can focus on detecting some details such as wearing a 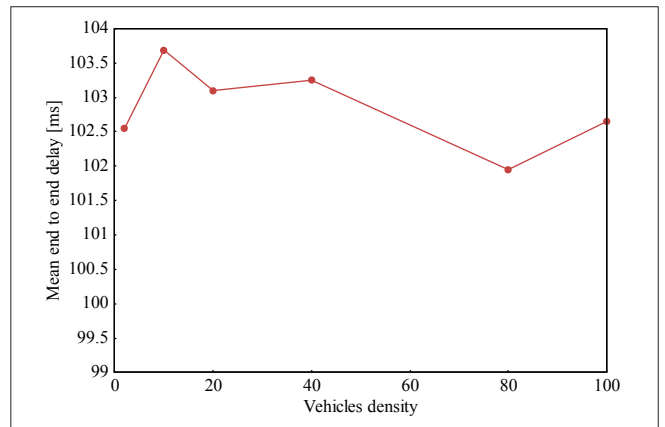face mask, to recognize if pedestrians/passengers are not wearing a face mask or wearing it incorrectly, and check if a vehicle's glass windows are rolled up completely when vehicles are too close to each other to prevent the spread of the virus through air.
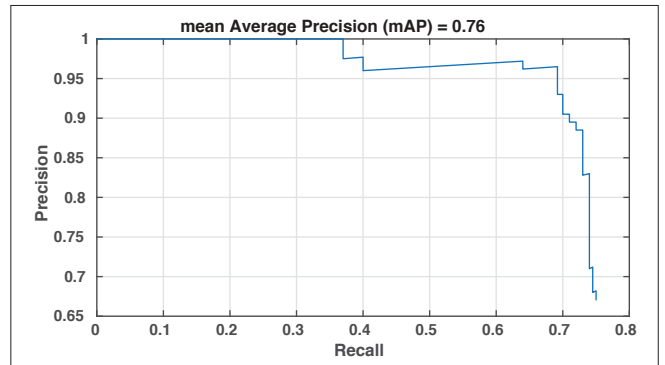
Indeed, most modern cities suffer from traffic congestion in all streets, especially at peak periods, which makes physical distancing between vehicles a challenging task as they are subject to driving regulations, in comparison to pedestrians, who have the choice to move freely. The deep cameras can smoothly track moving objects, and measure physical distancing easily in a low and medium density scene, and give best results; meanwhile, it rests a challenging task in the most crowded places. Some of the pedestrians are hidden together, which makes it not visible in crowded assemblies, even for a human observer; therefore, it seems too hard to put bounding boxes to all in this dense scene. However, image processing and annotation are expensive in both CPU and storage. The use of buses (with a pre-defined path) and taxi/user (with random path but popular) can help in increasing the benefits of edge computation by providing an extra server to enhance the computation capabilities of surrounding cameras.

An alternative method is to benefit from compute-less networks. An edge server can store not only the content but also the computation input/output. Similar computation tasks can use the stored output instead of performing computation from scratch, hence improving the computation deadline and resources. This scenario is widely witnessed in object tracking where various snapshots are overlapping, which increases the similarity of tasks and hence the resulting outputs.

Finally, IoV promises more efficient, cleaner, and safer transportation for all road users. Our article suggests a novel application of IoV data and capabilities at the service of road users and the local community in general. We have been able to prove that the data and services offered by the network of connected vehicles can also be important and a game changer for other road users.

References

[1] L. Boominathan, S. S. Kruthiventi, and R. V. Babu, "Crowdnet: A Deep Convolutional Network for Dense Crowd Counting," *Proc. 24th ACM Int'l. Conf. Multimedia*, 2016, pp. 640–44.

[2] WHO, https://www.who.int/news-room/featurestories/detail/who-expert-mission-to-belarus-recommends-physicaldistancing- measures-as-covid-19-virus-transmits-in-the-community, accessed May 20, 2020.

[3] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Monitoring COVID-19 Social Distancing with Person Detection and Tracking Via Fine-Tuned Yolo v3 and Deepsort Techniques," arXiv preprint arXiv:2005.01385, 2020.

[4] J. K. Kelso, G. J. Milne, and H. Kelly, "Simulation Suggests that Rapid Activation of Social Distancing Can Arrest Epidemic Development Due to A Novel Strain of Influenza," *BMC Public Health*, vol. 9, no. 1, p. 117, 2009.

[5] J. Zhang, S. Hu, and H. Shi, "Deep Learning Based Object Distance Measurement Method for Binocular Stereo Vision Blind Area," *Int'l. J. Advanced Computer Science and Applications*, vol. 9, 2018, p. 1.

[6] A. Zaarane et al., "Distance Measurement System for Autonomous Vehicles Using Stereo Camera," *Array*, vol. 5, 2020, p. 100–16.

[7] D.-Y. Huang et al., "Vehicle Detection and Inter-Vehicle Distance Estimation Using Single-Lens Video Camera on Urban/Suburb Roads," *J. Visual Commun. and Image Representation*, vol. 46, 2017, pp. 250–59.

[8] N. Sasaki, S. Tomaru, and S. Nakamura, "Development of Inter-Vehicle Distance Measurement System Using Camera-Equipped Portable Device," *2017 17th IEEE Int'l. Conf. Control, Automation and Systems*, 2017, pp. 994–97.

[9] J. Wagner et al., "Multispectral Pedestrian Detection Using Deep Fusion Convolutional Neural Networks," *ESANN*, 2016.

[10] N. Singh Punn, S. K. Sonbhadra, and S. Agarwal, "Monitoring COVID-19 Social Distancing with Person Detection and Tracking via Fine-Tuned Yolo v3 and Deepsort Techniques," arXiv, pp. arXiv–2005, 2020.

[11] M. Aqib et al., "A Smart Disaster Management System for Future Cities Using Deep Learning, GPUS, and In-Memory Computing," *Smart Infrastructure and Applications*, Springer, 2020, pp. 159–84.

[12] S. Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[13] ETSI TS 102 869-3 v1.6.1, "Intelligent Transport Systems (ITS); Testing; Conformance Test Specifications for Decentralized Environmental Notification Basic Service (den); Part 3: Abstract Test Suite (ATS) and Protocol Implementation Extra Information for Testing (PIXIT) (2020-04)," 2020.

[14] 3GPP TS 23.285 v15.2.0, "Technical Specification Group Services and System Aspects; Architecture Enhancements for V2X Services," Rel-15, Dec. 2018.

[15] M. Amadeo et al., "Enhancing the 3GPP v2x Architecture with Information-Centric Networking," *Future Internet*, vol. 11, no. 9, 2019, p. 199.

Biographies

Yesin Sahraoui (sahraoui.yesin@univ-ouargla.dz) is currently a Ph.D. student at the University of Ouargla, Algeria. His main research is focused on Social internet of Vehicles, wireless sensor networks, and cloud computing.

Chaker Abdelaziz Kerrache (ch.kerrache@lagh-univ.dz) is an associate professor in the Department of Mathematics and Computer Science, University of Ghardaia, Algeria. He received his M.Sc. degree in computer science from the University of Laghouat, Algeria, in 2012, and his Ph.D. degree in domputer science from the University of Laghouat, Algeria, in 2017. His research activity is related to trust and risk management, secure multi-hop communications, vehicular networks, named data networking, and UAVs.

Ahmed Korichi (ahmed.korichi@gmail.com) received his engineering degree in computer science from the University of Batna, Algeria, in 1987 and his Master's and Ph.D. degrees in computer science from the University of Batna in 2001 and 2009, respectively. He obtained his HDR degree from the University of Biskra, Algeria in 2010. He joined the Department of Computer Science and Information Technology at the University of Ouargla. Since 2001, he has been conducting research in the areas of network design and performance, modeling and simulation, QoS, wireless ad hoc networks, and vehicular networks. He has reviewed contributions in several scientific journals.

Boubakr Nour [GS'17, M'20] (n.boubakr@bit.edu.cn, n.boubakr@ieee.org) received his Ph.D. degree in computer science and technology from Beijing Institute of Technology, China. His research interests include next-generation networking and Internet. He was the recipient of the Best Paper Award at IEEE GLOBECOM 2018, and the Excellent Student Award at Beijing Institute of Technology in 2016, 2017, and 2018.

Asma Adnane (a.adnane@lboro.ac.uk) is a lecturer in the Computer Science Department of Loughborough University. Before joining Loughborough, she was a senior lecturer at the University of Derby (2013–2018) and a KTP associate at the University of Leicester and Crowdlab ,Leicester (2011–2013). Her main research focus is on networks and security. She has been involved in research related to the multipath routing performance in ad hoc networks, and in different applications of trust management for secured routing in ad hoc networks and vehicular networks.

Rasheed Hussain [S'09, M'15, SM'19] (r.hussain@innopolis.ru) is an associate professor and the director of the Institute of Information Security and Cyber-Physical Systems, Innopolis University, Russia. He is also the head of the Networks and Blockchain Lab at Innopolis University. His research interests include information, cyber, and network security, privacy, vehicular networks, vehicular clouds, vehicular social networking, applied cryptography, the Internet of Things, content-centric networking, blockchain, API security, and machine learning in cybersecurity.

FOOTNOTES

[1] https://www.cdc.gov/coronavirus/2019-ncov/