Unmanned aerial vehicles and machine learning for detecting objects in real time

Mustafa Fahem Albaghdadi, Mehdi Ebady Manaa
Department of Information Networks, College of Information Technology, University of Babylon, Babylon, Iraq

Article Info

Article history:

Received Jun 1, 2022
Revised Jul 15, 2022
Accepted Aug 16, 2022

Keywords:

Object detection
Real-time system
UAV
YOLO

ABSTRACT

An unmanned aerial vehicle (UAV) image recognition system in real-time is proposed in this study. To begin, the you only look once (YOLO) detector has been retrained to better recognize objects in UAV photographs. The trained YOLO detector makes a trade-off between speed and precision in object recognition and localization to account for four typical moving entities caught by UAVs (cars, buses, trucks, and people). An additional 1500 UAV photographs captured by the embedded UAV camera are fed into the YOLO, which uses those probabilities to estimate the bounding box for the entire image. When it comes to object detection, the YOLO competes with other deep-learning frameworks such as the faster region convolutional neural network. The proposed system is tested on a wild test set of 1500 UAV photographs with graphics processing unit GPU acceleration, proving that it can distinguish objects in UAV images effectively and consistently in real-time at a detection speed of 60 frames per second.

Corresponding Author:

Mustafa Fahem Albaghdadi
Department of Information Network, College of Information Technology, University of Babylon
20th Street, Al Hillah, Babylon, Iraq
Email: mustafa.net.msc@student.uobabylon.edu.iq

1. INTRODUCTION

Unmanned aerial vehicles (UAVs) with the ability to operate autonomously have grown in popularity in recent years for a variety of reasons. These include reconnaissance and surveillance, search and rescue, and infrastructure assessment. Visual object identification is a vital component in the development of completely autonomous systems for UAVs of this kind [1]. It is difficult to identify objects on low-cost consumer UAVs with their onboard cameras because of the poor resolution and noise, as well as the tiny size of the things they are trying to capture [2]. This makes the process of object recognition even more difficult. Due to the necessity for near real-time performance in many UAV applications, such as when objects are required for navigation, the task becomes much more complex [3]. The problem to be solved is the difficulty of identifying and locating objects using cheap and lightweight drones. Where the work aims to develop an object detection system using you only look once (YOLO) and determine the location accurately and in real time, while maintaining the system in terms of weight and cost.

Real-time tracking of cars, pedestrians, and landmarks for autonomous navigation and landing has been a common goal of many UAV investigations. Therefore, there are just a few systems that can identify several objects, despite the fact that many UAV applications need the ability to identify numerous targets. It is therefore suggested that there are two practical but important restrictions to blame for this gap between application demands and technology capabilities [4]. It is difficult to build and store a variety of target object models, especially when the objects have a variety of appearances, and real-time object detection requires high computing power even to detect single objects, much less when many target objects are involved, in addition to object recognition algorithms that are tailored to specific object and context types [5].

There are a number of different related works that are associated with UAV object detection, as follows. Al-Sheary and Almagbile [6] examined the risks associated with huge gatherings and developed a variety of safe crowd management strategies. Another option is real-time drone crowd monitoring, which is becoming more popular in order to save lives, preserve the environment, protect property, and maintain peace and authority. According to the findings of this research, pedestrian crowd monitoring systems may be a viable alternative [7]. Crowd density was computed using image segmentation algorithms based on real-time images taken by UAVs; after which the data was evaluated and the results were presented. The provided strategy may be used to make rapid decisions using high-quality data. An 80 percent accuracy rate was found for the photo segmentation method used in this study [8].

Hsieh *et al.* [9] created the car parking lot dataset (CARPK), the world's largest drone observation dataset. It is a difficult dataset for large-scale car counting jobs in parking lots. The research also created a unique strategy for generating viable area suggestions for an item counting task with regularized structures. The learned deep model can count things better if it knows how items are arranged. Counting automobiles from drone view scenes is the purpose of the proposed technique, and they compared it to four other methods: the one-look regression-based counting approach, two popular object identification systems, and a density object counting metric. Based on the methodologies used, region-based convolutional neural network (R-CNN) Faster is comparable to YOLO in terms of object detection success in recent years [10].

Lu *et al.* [11] investigated the difficulties of using drones to detect targets. They built a testbed to examine real-world events. The researchers identified these difficulties after testing perception modules with recent computer vision techniques. Our extensive simulations show that these characteristics have a big influence on the search algorithm design. More robust computer vision algorithms for target search and other drone-related applications are needed, as well as improved techniques to describe the effect of persistent characteristics.

A novel framework for three dimensions (3D) object localization and tracking using drones [12]. It involves object detection, multi-object tracking, ground plane estimation, and 3D target localization. The tracing and 3D localization performances are benchmarked against industry standards and ground truth. To address occlusions and camera rapid movements, their system is resilient. Their work is, nevertheless, bound by several constraints. In spite of this, they found that rapid camera motions do impact group plane estimations. Epipolar searches cannot be performed using a camera that simply spins one way, as is the case with typical drones [13]. Making use of CNN's monocular depth map might therefore be useful to address this aspect. Using the suggested approach, 3D positions may be acquired for each object, allowing for a smoother trajectory than two dimensions (2D). They believe that the addition of constraints to 3D trajectories will make the system more durable and successful.

Singhal *et al.* [14] suggested that a drone might be used to identify items in real-time. The neural network and machine learning algorithms successfully recognized all sorts of things. With so many uses in both autonomous and non-autonomous sectors, merging object detection with drone technology will help mankind. The detection module would identify all target objects and deliver the recognized object data. Object detection will be employed in surveillance, delivery, population analysis, and traffic monitoring, among other applications. Their work also involves a section on the system's future development. UAVs, commonly known as drones, have an important role in disaster response and humanitarian aid [15]. The main purpose of their study is to investigate how unmanned aerial vehicles (or drones) might help survivors in the case of a tsunami, earthquake, flood, or another natural disaster. Initially, it is anticipated that any natural disaster would cause quick damage to infrastructure, transportation, and key services [16]. The goal of this work is achieved by building a Yolo model for the purpose of identifying specific objects and then using this information to determine the exact location by repeating the detection process with more than one projection through a proposed algorithm.

2. THE PROPOSED APPROACH

2.1. System hardware

This section presents a discussion of the hardware used throughout the study, as presented in Table 1. A drone with four propellers was developed with the goal of identifying objects in the sky. The drone was constructed with the help of several electrical components [17].

In this case, the Drone Frame was employed, which is a representation of the aircraft's construction with a size of 450 mm. There were four brushless motors that met the criteria (the rest of the details are stated in Table 1). The fan speed was controlled by the use of an electronic speed controller. The propellers of the F11 aircraft were represented by drone blades, which were employed in this project, as shown in Figure 1 and Figure 2. A lithium-ion battery with four cells was utilized [4].

Table 1. Hardware specifications

Device type	Device model	Supply voltage	Product size	Output interface
Drone frame	F450	-	450 mm	-
Brushless motors	FPV drone	7.4 -11v	4.49 x 3.39 x 0.98 inches	7500 RPM
Electronic speed controller	BEC 2A	7.4 -11v	55x26x8 mm	30A for motor
Drone blades	F11	-	11 x 3 x 0.3 inches	-
Battery	Ovonic 4S	14.8V	2.7*1.26*1.4inch	XT60 plug
PIXHAWK flight controller	Radiolink PIXHAWK	5V	6.65 x 4.25 x 1.93 inches	multi
Radiolink	FS-i6X	12v	9.13 x 8.31 x 4.29 inches	Antenna
Power module	FPV Drone	5v	4.02 x 2.52 x 0.31 inches	XT60 plug
GPS compass	TS100	5v	3.2 x 2.2 x 0.4 inches	Jumper wires
Raspberry Pi 3	Pi 3 B+ Motherboard	5v	3.54 x 2.36 x 0.79 inches	Jumper pins
Camera handlebar	GoPro 10	5V	2.36 x 1.38 x 7.17 inches	Jumper pins

Figure 1. QUAD-COPTER Drone structure

Figure 2. Drone flies up to the sky

The control of the aircraft was accomplished with the use of a Da-Jiang innovations (DJI) Controller. The aircraft was controlled via the use of a radio connection. The power module was responsible for controlling the aircraft’s electrical power source. The plane’s coordinates were determined using a global positioning system (GPS) compass of type TS100, which was connected to satellites. It was decided to utilize the Raspberry Pi 3 to run an artificial intelligence model in order to determine the position of targets. Moreover, the camera handlebar was used so that the user can regulate the stability of the camera as well as the location.

2.2. System overview

The two phases of the proposed system architecture are described below. Each stage has a series of sub-steps that are necessary to accomplish the research goals and accomplish the research goal. Figure 3 depicts the remote motion control and position detection steps.

The QUAD-COPTER motion control is the initial level, which involves four sub-steps. There is an important role for this stage in the QUAD-COPTER’s navigation and avoidance capabilities. Detecting the position of the QUAD-COPTER is the second step in the process, which includes a number of procedures that transfer data to the base station, which displays the location on a map. At this point, the goal is to use real-time object recognition and transmission to relay data back to the base station as quickly as possible.

There are four major components to the proposed system. An Arduino and Raspberry Pi 3 was used to build the drone and manage its fly direction. A number of software applications were installed and downloaded to help specify the drone's components and its readiness for flight. Servo and Raspberry Pi 3 applications, as well as protocols for video transmission and signal transfer between the drone and the computer, were all included in this section. To identify and categorize the newly found items in real-time, a one-dimensional convolutional neural network technique was used (labeling). In addition to the GPS trackers, which were utilized to calculate the beginning and finishing sites of the drones, the trigonometric functions were used in altering the camera angle and the drone height in order to automatically determine the direction of the observed item.

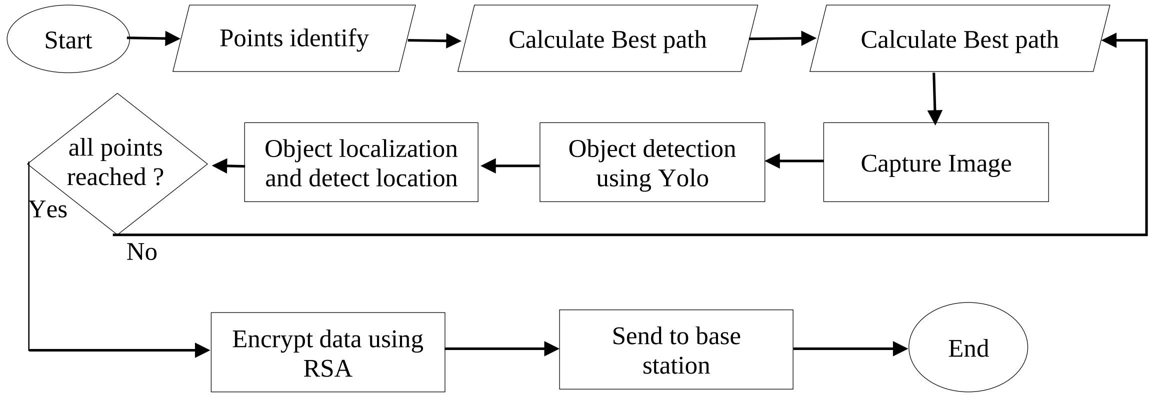

Figure 3. Proposed system

3. METHOD

3.1. Object detection

The recognition of objects in photographs taken by UAVs has been a persistent problem in computer vision. It is difficult to discern items in drone photographs because of the variety of sizes involved, including people, buildings, water bodies, and hills. In this paper, a build-in module in YOLO is used [14].

YOLO was chosen because it works in real time more efficiently than other artificial intelligence (AI) methods. The advantage of YOLO's fast response is that it uses only one stage, which is the CNN without using the reign of interest stage.

3.2. Localization

Locating a UAV's physical position in line with a real or virtual coordinate system is known as localization. When a direct measurement of the UAV's position is unavailable, localization is critical [18]. The accuracy of the estimated location information at a particular point in time is used to assess the performance of a system that uses localization. In this paper, the software is built to calculate X, Y, and Z coordination, as shown in Figure 4 [19].

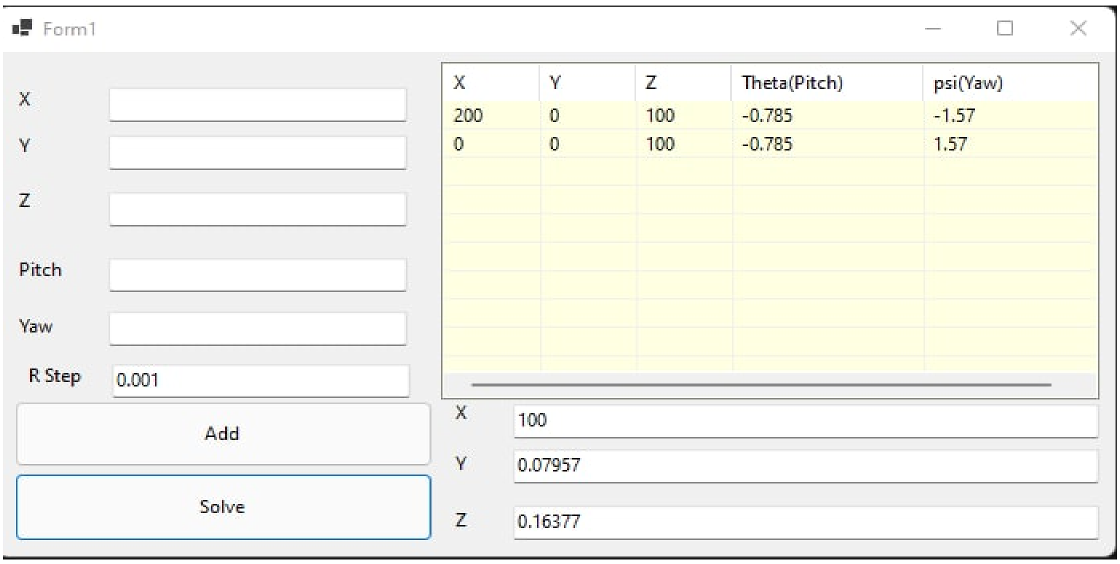

Figure 4. Localization in UAV's systems

Both X, Y, and Z are needed to be position coordinates. Path denotes the value of the lateral axis (Pitch), while Yaw denotes the value of the vertical axis. After the completion of the form and the addition of the values to the right pane [20] these values need to be solved. The results are shown in the x, y, and z coordinates, which correspond to the goal location [21]. Basically, when the drone moves in more than one direction in order to get a precise and steady position to the target, the intersection of these points is found by taking the average of these locations. Consequences for obtaining target coordinates may be found in (1)-(3) [12].

$$\mathrm{Deltax} = \mathrm{r} * \cos(\mathrm{th}) * \sin(\mathrm{psi}) \quad (1)$$

$$\mathrm{Deltay} = \mathrm{r} * \cos(\mathrm{th}) * \cos(\mathrm{psi}) \quad (2)$$

$$\mathrm{Deltaz} = \mathrm{r} * \sin(\mathrm{th}) \quad (3)$$

New x, y, z values are shown using (3)-(5):

$$Xnew = x + Deltax \quad (4)$$

$$Ynew = y + Deltay \quad (5)$$

$$Znew = z + Deltaz \quad (6)$$

The result of the above modules is a straight line that starts from the drone and ends up in ∞, passing through the target position. The straight line is divided into radius r which is utilized to calculate the average of obtained target points. However, the average shows the closest point to the target position that the drone has captured from different trends [22].

4. RESULTS AND DISCUSSION

4.1. Object detection

The preliminary step was investigating the efficiency of YOLOs training on image data in the identification of various items in the lab setting. A variety of things were gathered and placed over the testing area, after which a drone is used to photograph them from a variety of angles and views. The training in Figure 5 shows that the model currency increases slightly and becomes approximately 0.97 whenever the number of training images increased [17]. Figure 6 illustrates the model wrong object detection rates that decreased to 0.05 whenever the number of training images increased.

In order to prove that the proposed method works in a realistic yet controlled context, three separate sets of tests are performed. As part of the initial series of tests, it involves the investigation of how well YOLO can recognize objects from a high distance, as well as how well they can be used in a robot application [23], [24]. To test the performance of YOLO for human recognition, the latency times are compared at different distances, whereby a slight change in the connection latency led to a significant increase whenever the distance became bigger, as shown in Table 2.

Finally, as a basic simulation of a search-and-rescue or surveillance application, the proposed technique is tested by having a drone look for a target item in an interior setting. Figure 7 shows the results of the search-and-rescue simulation.

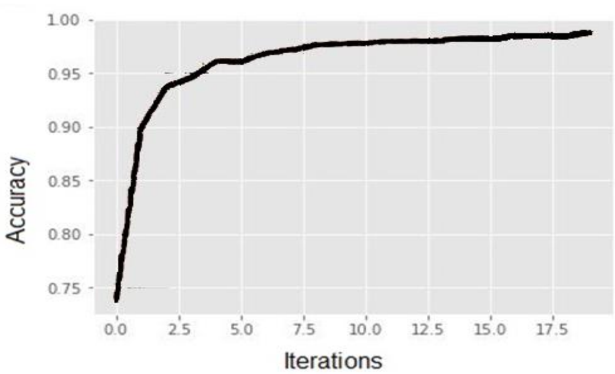

Figure 5. Right object detection rates increase

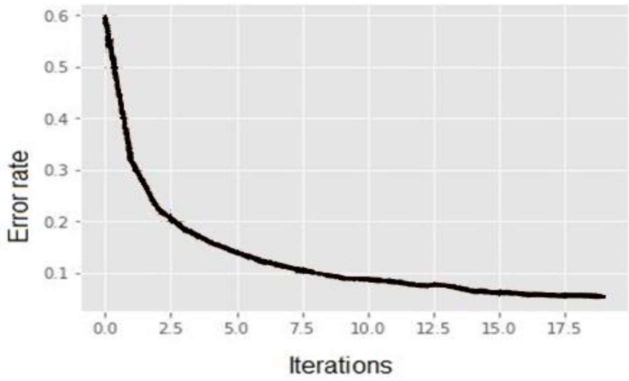

Figure 6. Wrong object detection rates increased

Figure 7. A drone is able to detect person from high distance using YOLO

4.2. Localization testing

As part of the testing requirements, the YOLO model is first tested to check its latency from different distances. The drone has been changed with different angles and views to get as many readings of latency as possible [25]. The results show that when the distance between the base station and drone increased, the latency slightly increases as well, as presented in Table 2. When the distance was 25 m, the latency was only 2 ms, whereas, in comparison to a distance of 200 m, the latency became 35 ms. By analyzing the numbers, it can be stated that the latency is not very high compared to the distance between the drone and the base station [26].

Table 2. Latency between base station and drone

Distance (m)	Latency (ms)
25	2
50	5
100	15
150	23
200	35

5. CONCLUSION

It was hypothesized in this study that UAVs might identify hundreds of object types using CNN. Although YOLO is computationally intensive, a local transmission control protocol (TCP) connection solution to recognition is being considered. It is possible to run object identification algorithms on low-cost consumer UAVs, such as lightweight, low-cost consumer UAVs using the YOLO technique. Even with practically infinite local TCP connection capacity comes a potentially significant and unexpected communication latency, as well as very changeable system loads. As a low-cost hardware platform, the QUAD-COPTER was used to evaluate the proposed method in an actual outdoor setting. In spite of the added communication latency, the findings indicate that the local TCP connection technique might offer speed-ups of almost an order of magnitude, even when identifying hundreds of object types. In a basic target search scenario, it was proven that the proposed technique is effective in terms of identification accuracy and speed.

REFERENCES

[1] A. L. Alfeo, M. G. C. A. Cimino, N. D. Francesco, M. Lega, and G. Vaglini, "Design and simulation of the emergent behavior of small drones swarming for distributed target localization," *Journal of Computational Science*, vol. 29, pp. 19–33, Nov. 2018, doi: 10.1016/j.jocs.2018.09.014.

[2] V. F. Vidal, L. M. Honorio, M. F. Santos, M. F. Silva, A. S. Cerqueira, and E. J. Oliveira, "UAV vision aided positioning system for location and landing," in *2017 18th International Carpathian Control Conference (ICCC)*, May 2017, vol. 168, no. January, pp. 228–233, doi: 10.1109/CarpathianCC.2017.7970402.

[3] K. M. A. Alheeti, M. S. Al-Ani, A. K. N. Al-Aloosy, A. Alzahrani, and D. A. S. Rukan, "Intelligent mobile detection of cracks in concrete utilising an unmanned aerial vehicle," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 176–184, Feb. 2022, doi: 10.11591/eei.v11i1.2987.

[4] M. Wakabayashi, H. G. Okuno, and M. Kumon, "Multiple Sound Source Position Estimation by Drone Audition Based on Data Association Between Sound Source Localization and Identification," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 782–789, Apr. 2020, doi: 10.1109/LRA.2020.2965417.

[5] A. Tullu, B. Endale, A. Wondosen, and H.-Y. Hwang, "Machine Learning Approach to Real-Time 3D Path Planning for Autonomous Navigation of Unmanned Aerial Vehicle," *Applied Sciences*, vol. 11, no. 10, p. 4706, May 2021, doi: 10.3390/app11104706.

[6] A. Al-Sheary and A. Almagbile, "Crowd Monitoring System Using Unmanned Aerial Vehicle (UAV)," *Journal of Civil Engineering and Architecture*, vol. 11, no. 11, Nov. 2017, doi: 10.17265/1934-7359/2017.11.004.

[7] H. Saha *et al.*, "A low cost fully autonomous GPS (Global Positioning System) based quad copter for disaster management," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan. 2018, vol. 2018-Janua, pp. 654–660, doi: 10.1109/CCWC.2018.8301782.

[8] J. Lee, J. Wang, D. Crandall, S. Sabanovic, and G. Fox, "Real-Time, Cloud-Based Object Detection for Unmanned Aerial Vehicles," in *2017 First IEEE International Conference on Robotic Computing (IRC)*, Apr. 2017, pp. 36–43, doi: 10.1109/IRC.2017.77.

[9] M.-R. Hsieh, Y.-L. Lin, and W. H. Hsu, "Drone-Based Object Counting by Spatially Regularized Regional Proposal Network," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, vol. 2017-Octob, pp. 4165–4173, doi: 10.1109/ICCV.2017.446.

[10] P. J. Baeck, N. Lewyckyj, B. Beusen, W. Horsten, and K. Pauly, "Drone based near real-time human detection with geographic localization," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 42, no. 3/W8, pp. 49–53, Aug. 2019, doi: 10.5194/isprs-archives-XLII-3-W8-49-2019.

[11] Y. Lu, Z. Wang, Z. Tang, and T. Javidi, "Target Localization with Drones using Mobile CNNs," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2018, pp. 2566–2573, doi: 10.1109/IROS.2018.8594163.

[12] H. Zhang, G. Wang, Z. Lei, and J.-N. Hwang, "Eye in the Sky," in *Proceedings of the 27th ACM International Conference on Multimedia*, Oct. 2019, no. 1, pp. 899–907, doi: 10.1145/3343031.3350933.

[13] A. Albanese, V. Sciancalepore, and X. Costa-Perez, "SARDO: An Automated Search-and-Rescue Drone-Based Solution for Victims Localization," *IEEE Transactions on Mobile Computing*, vol. 21, no. 9, pp. 3312–3325, Sep. 2022, doi: 10.1109/TMC.2021.3051273.

[14] H. Singhal, S. Rajpure, R. Badgujar, P. Dangle, P. Patil, and P. Guide, "Identifying and Detecting Real-Time Objects Using Drone Camera," vol. 4, no. 12, pp. 39–42, 2019.

[15] M. A. R. Estrada and A. Ndoma, "The uses of unmanned aerial vehicles –UAV's- (or drones) in social logistic: Natural disasters response and humanitarian relief aid," *Procedia Computer Science*, vol. 149, pp. 375–383, 2019, doi: 10.1016/j.procs.2019.01.151.

[16] Z. Li, "Road Aerial Object Detection Based on Improved YOLOv5," *Journal of Physics: Conference Series*, vol. 2171, no. 1, p. 012039, Jan. 2022, doi: 10.1088/1742-6596/2171/1/012039.

[17] Z. Qadir, F. Ullah, H. S. Munawar, and F. Al-Turjman, "Addressing disasters in smart cities through UAVs path planning and 5G communications: A systematic review," *Computer Communications*, vol. 168, pp. 114–135, Feb. 2021, doi: 10.1016/j.comcom.2021.01.003.

[18] J. N. McRae, C. J. Gay, B. M. Nielsen, and A. P. Hunt, "Using an Unmanned Aircraft System (Drone) to Conduct a Complex High Altitude Search and Rescue Operation: A Case Study," *Wilderness & Environmental Medicine*, vol. 30, no. 3, pp. 287–290, Sep. 2019, doi: 10.1016/j.wem.2019.03.004.

[19] H. Nawaz, H. M. Ali, and S. R. Massan, "A study of mobility models for UAV Communication Networks," *3C Tecnología_Glosas de innovación aplicadas a la pyme*, pp. 276–297, May 2019, doi: 10.17993/3ctecno.2019.specialissue2.276-297.

[20] S. Pal and P. P. M. Chawan, "Real-time object Detection using Deep Learning : A survey," *International Research Journal of Engineering and Technology (IRJET)*, no. November, pp. 397–399, 2019.

[21] F. B. Sorbelli, S. K. Das, C. M. Pinotti, and S. Silvestri, "Range based algorithms for precise localization of terrestrial objects using a drone," *Pervasive and Mobile Computing*, vol. 48, pp. 20–42, Aug. 2018, doi: 10.1016/j.pmcj.2018.05.007.

[22] A. R. Kulaib, R. M. Shubair, M. A. Al-Qutayri, and J. W. P. Ng, "An overview of localization techniques for Wireless Sensor Networks," in *2011 International Conference on Innovations in Information Technology*, Apr. 2011, pp. 167–172, doi: 10.1109/INNOVATIONS.2011.5893810.

[23] H. Nawaz, H. M. Ali, and S.-R. Massan, "Applications of unmanned aerial vehicles: a review," *3C Tecnología_Glosas de innovación aplicadas a la pyme*, pp. 85–105, Nov. 2019, doi: 10.17993/3ctecno.2019.specialissue3.85-105.

[24] N. A. Sehree and A. M. Khidhir, "Olive trees cases classification based on deep convolutional neural network from unmanned aerial vehicle imagery," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, p. 92, Jul. 2022, doi: 10.11591/ijeecs.v27.i1.pp92-101.

[25] T. Yang, A. Cabani, and H. Chafouk, "A Survey of Recent Indoor Localization Scenarios and Methodologies," *Sensors*, vol. 21, no. 23, p. 8086, Dec. 2021, doi: 10.3390/s21238086.

[26] G. Golcarenarenji, I. M.-Alpiste, Q. Wang, and J. M. A.-Calero, "Efficient Real-Time Human Detection Using Unmanned Aerial Vehicles Optical Imagery," *International Journal of Remote Sensing*, vol. 42, no. 7, pp. 2440–2462, Apr. 2021, doi: 10.1080/01431161.2020.1862435.

BIOGRAPHIES OF AUTHORS

Mustafa Fahem Albaghdadi obtained the BS degree in Information Networks, College of Information Technology, University of Babylon, Iraq in 2015-2019. He is currently pursuing the M.E. degree in the Information Networks Department, College of Information Technology, Babylon University, Iraq. Research interests include the internet of things, wireless network security and UAV. He can be contacted at email: mustafa.net.msc@student.uobabylon.edu.iq.

Mehdi Ebady Manaa is currently an assistant professor in the department of information network, College of Information Technology, University of Babylon. He received his bachelor's degree from the University of Babylon, college of science in 1999-2000. His master of science from University Utara Malaysia (UUM), Malaysia in 2012. He received his PhD in Computer Science and in the field Network Security and Data Mining using Cloud Computing from the University of Babylon, College of Information Technology in 2016. He is currently focusing on the detection of the attacks. The main interesting fields are data mining techniques (clustering and classification), communication software, network security, cloud computing, internet of things, and unstructured data. He can be contacted at email: it.mehdi.ebady@itnet.uobabylon.edu.iq.