Improved support for UAV-based computer vision applications in Search and Rescue operations via RAN Intelligent Controllers

Ciro J. A. Macedo, Elton Vivot Dias, Jones José da Silva Júnior, Gustavo Zanatta Bruno, Felipe H. Grings, Antonio M. Alberti, Cristiano Bonato Both e Kleber Vieira Cardoso

***Abstract*— Search and Rescue (SAR) operations can take significant advantage through the use of Unmanned Aerial Vehicles (UAVs) and communication systems. In this article, we explored the functionalities of RAN Intelligent Controllers (RICs) for managing and orchestrating network components aimed at critical mission operations assisted by UAVs. For example, UAVs are used to fly over an area where the victim is believed to be located, collect high-resolution video information and transmit it back to a ground base station. The proposed architecture exploits the components of the Open Radio Access Network (O-RAN) standard specification. Another contribution of this article is an assessment of a highly complex use case that explores new market trends, such as SAR operations assisted by UAV-based computer vision. The experimental results indicate, for instance, the proposed architecture can substantially improve the performance of applications with sensible Key Performance Indicators (KPIs), through a cognitive loop, able to act on the elements of the communication infrastructure to improve support for critical missions operations assisted by UAV.**

***Keywords*— Communication networks, Artificial Intelligence (AI), UAVs, RICs, Search and Rescue operations.**

I. Introduction

SAR operations are considered special, and are characterized by a similar set of constraints: the operational environments are unfriendly, e.g., caves, underwater, mountains, or disaster scenes — time is critical, and any delay can result in severe consequences, e.g., lost human lives and impact on wildlife [1]. UAVs can perform the SAR mission in these scenarios, reducing human intervention. UAVs are agile, fast, have low operating costs, and can exhibit autonomous behavior by organizing themselves to exchange information. UAVs can be deployed, performing sensory operations to collect evidence of a victim's presence, and reporting the collected information to a remote ground station, as illustrated in Fig. 1.

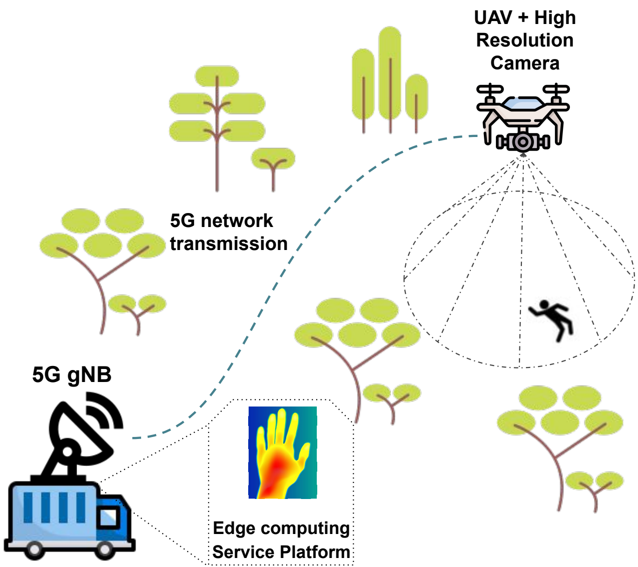

Fig. 1. SARs assisted by UAVs: UAV fly over an area where the victim is believed to be located, collect video information and transmit it back.

Other key elements in SARs operations is the communication systems. These systems provide the means for proper coordination among the several teams involved and provide the communication infrastructure so that UAVs can be integrated to improve the operation. However, some technological challenges need to be overcomed. Current communication networks do not support the demands of highly dynamic scenarios of SARs operations integrated with UAVs. This situation worsens in disaster scenarios, where the network infrastructure is affected and often does not allow continuous communication between several teams involved, including UAVs. In this way, the adaptability of the network has been constantly investigated, considering 5th Generation (5G)/Beyond 5G (B5G) scenarios, especially through the use of AI as a managing and orchestrating tool built into the elements that make up the communication infrastructure [2]. Moreover, standardization

Ciro J. A. Macedo, Instituto Federal de Goiás, Cidade de Goiás-GO, e-mail: ciro.macedo@ifg.edu.br; Elton Vivot Dias, Instituto de Informática, Universidade Federal de Goiás, Goiânia-GO, e-mail: eltondias@inf.ufg.br; Jones José da Silva Júnior, Escola Superior da Polícia Civíl, Goiânia-GO, e-mail: jjsjunior@gmail.com; Gustavo Zanatta Bruno, PPG em Comp. Aplicada, Universidade do Vale do Rio dos Sinos, Porto Alegre-RS, e-mail: zanattabruno@edu.unisinos.br; Felipe H. Grings, PPG em Comp. Aplicada, Universidade do Vale do Rio dos Sinos, Porto Alegre-RS, e-mail: felipehg@edu.unisinos.br; A. M. Alberti, Information and Communications Technology Laboratory, Instituto Nacional de Telecomunicações, Santa Rita do Sapucaí - MG, e-mail: alberti@inatel.br; Cristiano Bonato Both, PPG em Comp. Aplicada, Universidade do Vale do Rio dos Sinos, Porto Alegre-RS, e-mail: cbboth@unisinos.br; Kleber Vieira Cardoso, Instituto de Informática, Universidade Federal de Goiás, Goiânia-GO, e-mail: kleber@inf.ufg.br. This work was conducted with partial financial support from the National Council for Scientific and Technological Development (CNPq) under grant number 405111/2021-5. Moreover, RNP partially supported the work, with resources from MCTIC, Grant No. 01245.010604/2020-14 under the 6G Mobile Communications Systems project from the Radiocommunication Reference Center (CRR) of the National Telecommunications Institute – Inatel, and by MCTIC/CGI.br/São Paulo Research Foundation (FAPESP) through the Project SAMURAI - Smart 5G Core And MUltiRAn Integration under Grant 2020/05127-2.

efforts toward AI-powered communication networks are being disseminated by the 3rd Generation Partnership Project (3GPP), European Telecommunications Standards Institute (ETSI) Experiential Networked Intelligence (ENI) and O-RAN [3], [4], [5].

This article explores the features of the Artificial Intelligence/Machine Learning (AI/ML) pipeline, which integrates the RICs [6], specified in the O-RAN reference architecture [7]. Considering a scenario involving User Equipment (UE)/UAV, connected to an O-RAN ground vehicle Base Station (BS), and whereas this UE/UAV collects high-resolution video in real-time and transmits it to a Deep Neural Network (DNN) object detection model. In this context, we propose an approach aimed at improved support for critical missions operations assisted by UAV. The main contributions of this work are: (i) demonstrates the feasibility of using O-RAN AI/ML pipeline, aimed to improved support for critical mission operations assisted by UAV; (ii) explores, in the same use case, AI applied to the enhancement of O-RAN components, and O-RAN components providing enhanced support to an AI application, in the context of SAR operation.

The remaining of this article is organized as follows. Section II, presents 5G/B5G related works, considering the main initiatives towards AI/ML solutions, UAV, and SARs operations. In Section III, we present O-RAN reference architecture and describe how our AI/ML pipeline version can act to the enhancement of O-RAN components, and, in Section IV, we show the experiments with AI for UAV-based SARs. Finally, conclusions and open issues are discussed in Section V.

II. Related works

Over generations of wireless cellular networks, telecommunication standardization bodies, e.g., 3GPP, ETSI, and O-RAN, have been working on specifications to make the 5G ecosystem more efficient and optimized. Much of the effort is towards AI/ML usage to deal with the complexity of new applications and use cases presented by recent market trends.

Critical mission applications have been constantly investigated in the scientific literature [8] and must be supported by telecommunication infrastructures [3], [1]. However, many applications have requirements beyond the capabilities of the elements that make up the infrastructure [9]. Considering computer vision based on UAVs, recent research proposes the inclusion of DNN models, generating, as a consequence, a high throughput between edge and cloud networks [10]. In work [11], the authors present a set of use cases, providing an overview of deployable 5G network concepts, including architecture options, system performance analysis, and coexistence aspects. Focused on the context of mission-critical operations, the authors in [12] provide an overview of service requirements for public safety mission-critical communications, identifying key technical challenges and explaining how 5G New Radio (NR) features are being evolved to meet the emerging safety-critical requirements. Prospects of using AI/ML at the edge of the network have become stronger with the O-RAN Alliance, which arises with the general objective of standardizing an architecture and a set of interfaces to perform an open Radio Access Network (RAN). The aim is to produce an open RAN based on disaggregated, virtualized software components, which exchange information through open, standardized interfaces being interoperable among different vendors, following prerogatives similar to those specified in the 5G Core (5GC) [13]. Section III presents more details of the O-RAN architecture, as well as the possibility of integration with AI elements, focused on the context of improved support for critical mission operations assisted by UAVs.

III. O-RAN and AI for UAV critical missions

We adopted the following considerations to emulate a concrete scenario: (i) SAR operation as a critical mission; (ii) application requirements associated with KPIs of a DNN object detection model based on high resolution real-time video; (iii) UAVs-based high-resolution video stream; and (iv) 5G mobile networks, specifically O-RAN, due to its flexibility and characteristics of components' dissociation.

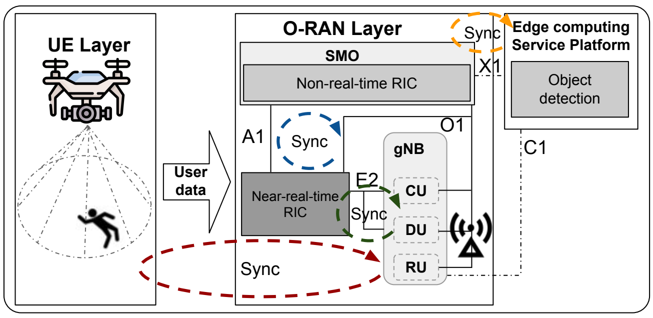

Fig. 2. O-RAN reference architecture, providing improved support for critical mission operations assisted by UAV.

The O-RAN reference architecture was structured based on four guidelines:

- **Virtualization.** Introduction of new components to manage and optimize the network infrastructure and its operations.
- **Disaggregation.** Divides the BS into Central Unit (CU), Distributed Unit (DU), and Radio Unit (RU), following the proposal by 3GPP for the RANs segmentation.
- **Open interfaces.** The inclusion of open interfaces connecting different O-RAN architecture components allows interoperability between the CU, DU, and RU.
- **Intelligent data-driven control through RICs.** RICs are programmable components that can execute optimization routines that orchestrate the RAN.

Fig. 2 shows O-RAN reference architecture [7] (*O-RAN Layer* boundary box), acting in a critical mission scenario, providing improved support for SAR operation assisted by UAVs. In the illustration, *UE Layer* shows a UAV carrying a high-resolution camera, collecting information and transmitting it back to the UAV control vehicle (ground BS illustrated in Fig. 1), to be used to identify possible victims through DNN object detection model [14].

The Fig. 2 also illustrates our view of the interaction between the elements: (i) the *sync* node labeled with red color is responsible to update UAV control with insights produced by third-party applications, i.e., *xApps* and *rApps* designed to provide value-added services to support the RAN optimization process, (ii) the *sync* node labeled with blue color acts in the near-real-time RAN Intelligent Controller (near-RT RIC) with insights produced by non-real-time RAN Intelligent Controller (non-RT RIC), (iii) the *sync* node labeled with green color performs Next Generation Node Base (gNB) optimization with insights produced by near-RT RIC, and (iv) the *sync* node labeled with orange color reports information about application performance to *rApps*.

We consider all network equipment deployed in vehicular ground BS. Moreover, we consider the transmission of control data and application data. Control data has KPIs that indicate low latency and low bandwidth demand. e.g., configuration changes, UAV flight status, UAV data reporting, and UAV navigation commands. However, application data includes 4K high-definition video data, demanding significant uplink network bandwidth requirements.

Knowing the network requirements associated with computer vision application and aiming to guarantee the communication during the SARs operation, we consider that the data produced by the network elements is collected and used by RICs components. An O-RAN internal Machine Learning (ML) pipeline processes the data, outputting insights used as input to four control loops, involving near-RT RIC, non-RT RIC, O-RAN disaggregation components, and Edge computing Service Platform, as illustrated in Fig. 2. The exchange of information between O-RAN internal components occurs through open interfaces A1, O1, and E2. In addition, we propose the X1 and C1 interfaces to reporting pertinent information about application performance to *rApps* and maintain connectivity between UEs and the edge computing service platform, respectively. The following paragraphs discuss details of RICs, *rApps*, *xApps*, Service Management and Orchestration (SMO), O-RAN open interfaces, and how AI can contribute to improving enhanced support for critical mission operations assisted by UAVs.

A. Service Management and Orchestration Framework

SMO handles all management, orchestration, and automation procedures to control RAN components. As illustrated in Fig. 2, non-RT RIC is part of SMO. Through the A1 and O1 interfaces, the SMO components interact with the other components, enabling the data collection to serve as input to AI/ML models. The outputs of the AI/ML models are intended to facilitate network monitoring and control [5].

B. E2, O1 and A1 Interfaces

One of the O-RAN structuring guidelines was the inclusion of the A1, E2, and O1 interfaces, connecting different architecture components and enabling interoperability among CU, DU, and RU. The E2 interface interconnects the CU and DU elements to the near-RT RIC. Through this interface, the near-RT RIC can collect RAN metrics and act in control procedures of the CU and DU elements [15]. The O1 interface enables SMO to manage the lifecycle of O-RAN components. It is an interface focused on operation and maintenance activities, allowing perform initialization/configuration activities of components and performance assurance control. The A1 interface connects non-RT RIC and near-RT RIC. Through this interface, non-RT RIC can forward high-level optimization goals and act in the management of ML models (e.g., deploy, update or undeploy ML trained models used in *xApps*) [5].

C. non-real-time RAN Intelligent Controller (non-RT RIC)

non-RT RIC is contained inside SMO and is one of the core components of the O-RAN reference architecture. It was designed to complement near-RT RIC, to support the execution of third-party applications, acting on control actions over RAN, with timescales larger than 1 second. Furthermore, because it is located inside SMO, non-RT RIC can influence the SMO operations and indirectly control all RAN components connected to SMO through the A1 and O1 open interfaces. non-RT RIC is composed of three main elements: (i) Data management and exposure, which are responsible for managing data and exposing services in the context of SMO; (ii) *rApps*, designed to provide value-added services to support the RAN optimization process; and (iii) AI/ML workflow, responsible for data collection/processing; training; validation/publishing; deployment; AI/ML execution/inference, and continuous operations of AI/ML models.

Considering improved support for critical mission operations assisted by UAV, we distributed the elements that make up the AI/ML workflow between non-RT RIC and near-RT RIC. On non-real time scale, data is collected and processed, and AI/ML models are trained and validated by the training host located in non-RT RIC. Training host is used for near-RT RIC continuous online learning. Inside non-RT RIC, there is the ML model repository that is used to save backup ML training models.

D. near-real-time RAN Intelligent Controller (near-RT RIC)

near-RT RIC connects the O1, A1, and E2 interfaces, hosting the *xApps* and the components required to operate and manage the *xApps*. Considering improved support for critical mission operations assisted by UAV and control KPIs that have low latency demand and uplink/downlink service asymmetry, we placed the ML inference host as part of near-RT RIC. After training and validation steps in non-RT RIC, the ML models are deployed to the near-RT RIC inference host.

We consider the existence of a Data Analytics component inside near-RT RIC, which can contain several ML data analytics models. Each of them is trained and acting in a specific context, such as working in a UAV context decisions, e.g., navigation commands or flight status data reporting, or acting in context decisions associated with RAN control action and guidance, e.g., configuration changes. The outputs produced by ML Data Analytics components may correlate with actions: (i) Load Balancing; (ii) Anomaly Detection; (iii) Mobility Prediction; (iv) Resource Forecast; and (v) Quality of Service (QoS) Assurance. Considering improved support for SARs operations, these correlations are explored by the decision engine to generate actions aimed at RAN resource scaling, QoS targeted at SARs operation, and UAV mobility management. Finally, through a constant synchronization process, the insights generated by the near-RT RIC inference host stimulate *xApps* to act on the RAN elements, generating a cognitive control loop, able to act on the elements of the communication infrastructure to improve support for critical mission operations assisted by UAV.

We describe the sets of experiments and their results acting on the communication infrastructure elements in the following.

IV. Results

We demonstrated the flexibility and usefulness of the O-RAN elements, considering a SAR operation, in which an UAV flies over an area where the victim is believed to be located, collects high-resolution video information, and transmits it back to a vehicular ground BS. Edge computing services are deployed at the ground BS, processing local video and providing control information. In addition, real-time data services collect video server and application health information. We use a DNN YOLOv3 [14] to perform object detection.

Fig. 3 illustrates the network topology, we used to represent the scenario, in which O-RAN components improve connectivity between UAV/UE and Edge computing services, to provide better performance in YOLOv3 object detection. This topology comprises a Cloud node, O-RAN node, vehicular ground BS, and UE[1].

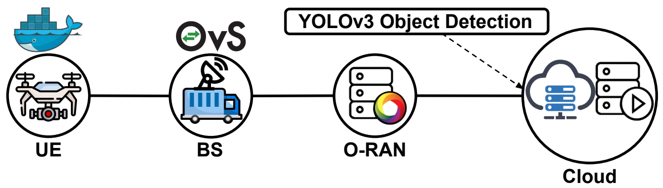

Fig. 3. Network topology representing UAV/UE, vehicular ground BS, O-RAN and YOLOv3 Object Detection.

UE represents UAV, emulated in a containerized Hypertext Transfer Protocol (HTTP) server that streams a high-resolution video through the network. Our vehicular ground BS node provides the network, representing RAN. The BS node is emulated by a virtual switch created by Open vSwitch (OVS) and connects UE to the O-RAN infrastructure. Finally, the cloud node represents a location where all the services and applications external to O-RAN are stored. In our case, the cloud node is a Darknet algorithm with YOLOv3 [14] pre-trained model, running on a Compute Node of Open-Access Research Testbed for Next-Generation Wireless Networks (ORBIT) testbed outfitted with NVIDIA Compute Unified Device Architecture (CUDA) capable Graphics Processing Units (GPUs).

Considering high-resolution images are more suitable for the smaller object detection [16], and considering the availability in the market of models of UAVs properly designed to capture 4K images, we stream a 4K video filmed by a UAV, showing a group of people on the highway. We limited the video stream time to have a more controlled experiment and used a 30 Frames per Second (fps) evaluation rate on YOLOv3. Table I summarizes all the parameters used and their respective values.

[1]https://github.com/LABORA-INF-UFG/paper-CEJGACK-2022

TABLE I
Parameters and their respective values, used in the experiments.

Param Description	Value
Object Detection Algorithm	YOLOv3 [14]
Object Detection Dataset	Pascal VOC
Video Strem Duration	3 seconds
Video Stream Frame Rate	30 fps
Video Strem Resolution	4K (3840 x 2160 pixels)

We used a Real Time Streaming Protocol (RTSP) server containerized together with FFmpeg, positioned at the UE and cloud nodes as illustrated in the topology of Fig 3 to simulate a UAV-based high-resolution video stream. The FFMpeg UE node stream the video to the cloud node that contains FFmpeg Tools container. The cloud node receives the streaming and extracts the frames used in the object detection process. Moreover, we label each of the frames using a Computer Vision Annotation Tool to check the precision, marking the position of the objects to be detected in each frame.

In experiments, we employed a Traffic Control Tool in the Linux kernel to emulate the following situations: (i) dropping bandwidth and (ii) increasing packet loss. First, as shown in Fig 4, we started the process of streaming with the normal network infrastructure conditions. Therefore, the infrastructure starts to behave abnormally, affecting application performance. At this time, RICs detects this abnormal behavior, acts in the network infrastructure, return to normal conditions and improving the application performance.

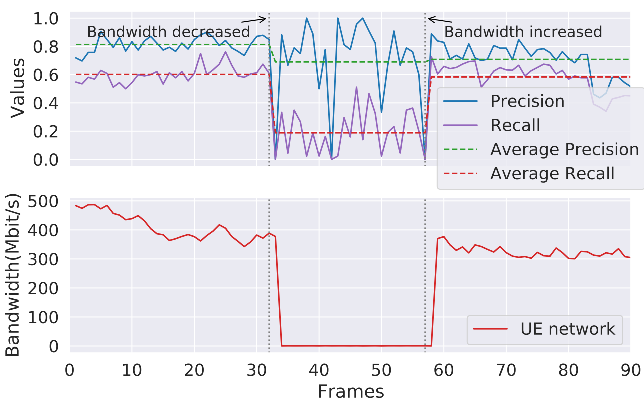

Fig. 4. AI application behavior in a scenario of bandwidth decrease.

Fig 4 shows the behavior of the object detection application in a scenario of decreased bandwidth. The x-axis represents a video stream timeline (fps). At a given moment, we note a disturbance in the average of the recall, caused by a decrease in the bandwidth, and then the values returned to the previous

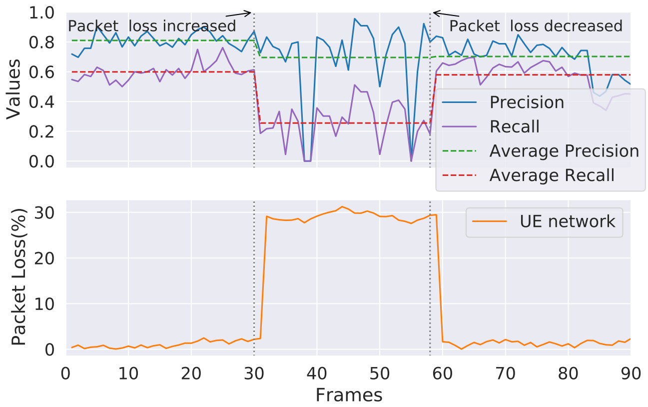

Fig. 5. AI application behavior in a scenario of packet loss.

averages. Fig 5 illustrates the object detection application's behavior in a packet loss scenario. In this case, we note a disturbance in the recall average caused by increasing packet loss.

In both scenarios, SMO has access to the data that indicate this momentary disturbance through the X1 interface, shown in Fig 2. By correlating the application data with the other data produced by the elements that make up the communication infrastructure, the ML pipeline can act to reestablish the communication infrastructure.

The results demonstrate the importance of a communication infrastructure to remain stable and functional in critical mission scenarios, in which time is relevant and any delay can result in serious consequences. When considering these operations assisted by UAV, the scenario becomes more complex. Therefore, it is necessary to make networks more intelligent and integrated with AI/ML elements to act in communication infrastructures' management and control process, such as RICs.

V. Open Issues and Conclusion

This work discussed, the benefits and challenges of AI in critical mission operations assisted by UAVs integrated with RAN intelligent controllers. As shown above, UAVs significantly contribute as an auxiliary tool to improving critical mission operations. However, the inherent features of UAVs summed up with the complexity of new applications and use cases presented by new market trends, demand standards, and efficient AI/ML for the optimized performance of communication infrastructure. Standardization and wide adoption of AI/ML built into the elements that make up the communication infrastructure are key strategies to reduce cost and achieve SAR operations efficiency that heavily depends on communication and computing systems. We explored the functionalities of RAN intelligent controllers, described in the O-RAN standard specification [6], presenting an architecture proposal for providing full AI/ML capabilities for standardized systems aimed at critical mission operations assisted by UAVs. In addition, we also presented experiments that illustrate the benefits for describing the whole architecture elements and their interaction. The initial results show the challenges imposed in some SAR operations, especially when assisted by UAVs, and at the same time, show the potential gains obtained with the use of intelligent strategies closer to the edge, according to the proposal of O-RAN architecture.

While the experiments in this work explored some of the essential concepts of RAN intelligent controllers, the implementation of non-RT RIC and near-RT RIC that covers all components presented in Fig 2 is still lacking. Therefore, in future work, we intend to develop and make publicly available a minimally functional implementation of non-RT RIC and near-RT RIC that illustrates all components described in Section III. Moreover, we are interested in investigating the interaction between the different intelligent elements, which act at different points of the new communication infrastructures, for example, RICs positioned closer to Edge, and Network Data Analytics Function (NWDAF) placed at the core of the network.

References

[1] Roldan *et al.*, "ITU Guidelines for national emergency telecommunication plans; ITU-D Emergency Telecommunications Disaster Response," tech. rep., International Telecommunication Union (ITU), 2019.
[2] Both *et al.*, "System Intelligence for UAV-Based Mission Critical with Challenging 5G/B5G Connectivity," 2021.
[3] 3GPP-TR22.125, "Unmanned Aerial System (UAS) support in 3GPP," technical report, 3rd Generation Partnership Project (3GPP), 2021-09. Version 17.0.4.
[4] Y. Wang *et al.*, "From Design to Practice: ETSI ENI Reference Architecture and Instantiation for Network Management and Orchestration Using Artificial Intelligence," *IEEE Communications Standards Magazine*, vol. 4, no. 3, pp. 38–45, 2020.
[5] O-RAN Working Group 2, "O-RAN AI/ML workflow description and requirements 1.03," technical specification, O-RAN.WG2.AIML-v01.03, 2021-10.
[6] O-RAN Working Group 1, "O-RAN Architecture-Description 6.0," technical specification, O-RAN.WG1.O-RAN-Architecture-Description-v06.00, 2022-03.
[7] Polese *et al.*, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," 2022.
[8] W. Shule *et al.*, "UWB-Based Localization for Multi-UAV Systems and Collaborative Heterogeneous Multi-Robot Systems," *Procedia Computer Science*, vol. 175, pp. 357–364, 2020.
[9] X. Wu, W. Li, D. Hong, R. Tao, and Q. Du, "Deep Learning for Unmanned Aerial Vehicle-Based Object Detection and Tracking: A Survey," *IEEE Geoscience and Remote Sensing Magazine*, pp. 2–35, 2021.
[10] Zhou *et al.*, "Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1738–1762, 2019.
[11] Li *et al.*, "Towards Providing Connectivity When and Where It Counts: An Overview of Deployable 5G Networks," 2021.
[12] Li *et al.*, "5G New Radio for Public Safety Mission Critical Communications," 2021.
[13] Cardoso *et al.*, "A softwarized perspective of the 5G networks," 2020.
[14] Redmon *et al.*, "YOLOv3: An Incremental Improvement," 2018.
[15] O-RAN Working Group 3, "O-RAN Near-Real-time RAN Intelligent Controller Architecture & E2 General Aspects and Principles 2.01," technical specification, O-RAN.WG3.E2GAP-v02.01, 2022-03.
[16] H. Zhang *et al.*, "An empirical study of multi-scale object detection in high resolution UAV images," *Neurocomputing*, vol. 421, pp. 173–182, 2021.