

3D-O-RAN: Dynamic Data Driven Open Radio Access Network Systems

Francesco Restuccia[†], Erik Blasch[‡], Andrew Ashdown[†], Jonathan Ashdown*, Kurt Turck*

[†] Institute for the Wireless Internet of Things, Northeastern University, United States

[‡] Air Force Office of Scientific Research, United States

* Air Force Research Laboratory, United States

Email: {F.Restuccia, Ashdown.A}@northeastern.edu, {Erik.Blasch.1, Jonathan.Ashdown, Kurt.Turck}@us.af.mil

Abstract—This position paper introduces a Dynamic Data Driven Open Radio Access Network System (3D-O-RAN). The key objective of 3D-O-RAN is to support congested, contested and contaminated tactical settings where multimedia sensors, application constraints and operating wireless conditions may frequently change over space, time and frequency. 3D-O-RAN is compliant with the O-RAN specification for beyond 5G cellular systems to reduce costs and guarantee interoperability among vendors. Moreover, 3D-O-RAN integrates computational, sensing, and cellular networking components in a highly-dynamic, feedback-based, data-driven control loop. Specifically, 3D-O-RAN is designed to incorporate heterogeneous data into the network control loop to achieve a system-wide optimal operating point. Moreover, 3D-O-RAN steers the multimedia sensor measurement process in real time according to the required application needs and current physical and/or environmental constraints. 3D-O-RAN uses (i) a semantic slicing engine, which takes into account the semantic of the application to optimally compress the multimedia stream without losing in classification accuracy; (ii) a dynamic data driven neural network certification system that translates mission-level constraints into technical-level constraints on neural network latency/accuracy, and occupation of hardware/software resources. Realistic use-case scenarios of 3D-O-RAN in a tactical context demonstrate system performance.

I. INTRODUCTION

Ensuring U.S. dominance in next-generation wireless technologies is a key priority to ensure national security in the next few years. The Presidential Memorandum on Developing a Sustainable Spectrum Strategy for America's Future [1], released on October 25, 2018, states: “*It is the policy of the United States to use spectrum as efficiently and effectively as possible to help meet our economic, national security, science, safety, and other Federal mission goals now and in the future.*” Moreover, U.S. defense forces are expected to increasingly rely on wireless technologies such as 5G/6G to implement the *Joint All-Domain Command and Control (JADC2)* concept [2], [3]. Critically, 5G/6G networks will necessitate fast, reliable, effective and efficient administration of computation and radio resources to operate at an optimal operating point. In other words, achieving wireless dominance will bring the U.S. closer to multi-domain (space, air, land, sea, and cyberspace) coordination.

While the U.S. remains at the top of the global wireless innovation hierarchy, several nations are fast approaching.

Unmitigated adversarial spectrum usage, coupled with extensive commercial use, will yield an increasingly congested, contested, concealed and contaminated (4C) spectrum environment, where adversarial actions, commercial development, and regulatory constraints will necessarily impede freedom of action in the spectrum. As mentioned in the Memorandum, “*Freedom of action in the spectrum [...] is a required precursor to the successful conduct of operations in all domains and to allow systems to share data seamlessly*”.

Ensuring spectrum freedom will require radically novel wireless networking paradigms, where intelligence, adaptability, flexibility and security are deeply embedded in the wireless communications systems. To this end, the Open Radio Access Network (O-RAN) paradigm can bring together open interfaces, softwarization of network components and multi-vendor interoperability [4]–[6]. Specifically, O-RAN will provide a platform where tactical services can be dynamically deployed on-demand in a matter of seconds via plug-and-play software containers. The O-RAN capability will empower a concrete solution to enable dynamic data driven network control, so as to promptly respond to emergencies, threats and mission requirements. Moreover, O-RAN is by definition *platform-independent*, so (i) data-driven control algorithms can be deployed in vendor-agnostic O-RANs, ultimately reducing costs to develop and operate the network; (ii) network intelligence can be used to operate secure network slices with diverse latency, throughput and reliability requirements and regulate coexistence with other networks. In February 2022, the House passed the America COMPETES Act, which recognizes the importance of O-RAN in establishing supply chain resiliency, by establishing future O-RAN programs and a Public Wireless Supply Chain Innovation Fund [7].

To achieve assured operations through O-RANs, systems will need real-time spectrum and situational awareness through the gathering and processing of heterogeneous sensor data. To this end, future products are expected to extensively rely on artificial intelligence (AI) to automate strategic decision-making in a cost-effective manner [8]. However, as AI becomes such a critical system component, it is imperative to guarantee *certifiability of AI techniques* through explainability of the outputs, interpretability of the processing, and accountability of the decision-making process [9]. Specifically, the AI capabilities are expected to have “*explicit*,

Approved for Public Release; Distribution Unlimited: AFRL-2022-3100.

well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles” [10]. To attest to the criticality of Certifiable AI, on March 5, 2021, the National Security Commission on Artificial Intelligence (NSCAI) has mentioned in its final report [11] the need to “[...] include an evaluation of technical standards and production and transmission pipelines”, while on November 15, 2021, the Defense Innovation Unit (DIU), in its initial “Responsible AI Guidelines” document, indicated the need to “[...] create a task force to study the use of AI and complementary technologies, including the development and deployment of standards and technologies, for certifying content authenticity and provenance” [12].

II. DYNAMIC DATA DRIVEN OPEN RAN SYSTEM

Our vision is simple: **Addressing these issues requires a fundamentally novel approach at the intersection of networking, computing, and optimization.** As such, this paper introduces a new concept in wireless networking, which we refer to as *Dynamic Data Driven Open Radio Access Network System* – in short, 3D-O-RAN. Figure 1 shows a high-level overview of 3D-O-RAN. At its core, 3D-O-RAN is a dynamic data driven application system (DDAS) [13], [14] where the computational, sensing, and networking components are tightly integrated in a highly-dynamic, feedback-based control loop. Specifically, 3D-O-RAN is designed to achieve the following goals: (i) incorporate heterogeneous sensor data (e.g., spectrum, multimedia) into the control loop to dynamically achieve a system-wide optimal operating point; (ii) dynamically steer the multimedia sensor measurement process according to the required application needs and current physical and/or environmental constraints. Ultimately, 3D-O-RAN supports settings where multimedia sensors, application constraints and operating wireless conditions may dynamically change over space, time and frequency [15], [16].

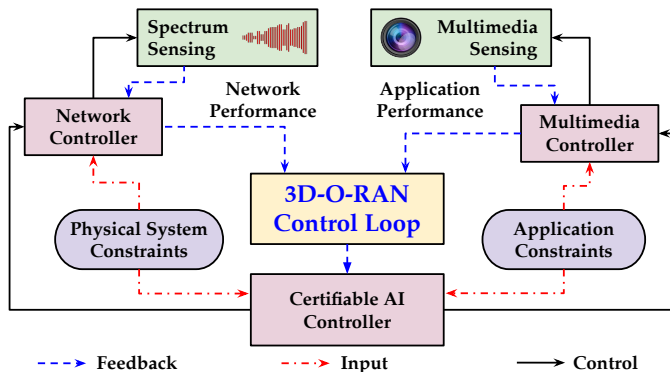


Figure 1: Overview of 3D-O-RAN.

As shown in Figure 1, the 3D-O-RAN system is logically composed by three control engines, namely, the *Network Controller* (NC), the *Multimedia Controller* (MC) and the *Certifiable AI Controller* (CC). Specifically, the NC and MC will (i) acquire real-time information from spectrum and multimedia

sensors for situational awareness; and (ii) influence the physical environment (i.e., spectrum utilization) and the sensing processes (i.e., spectrum sensing and multimedia sensing) with their control decisions. As such, the NC and MC will be highly data driven in nature, and will leverage state-of-the-art AI techniques to perform their functions. Since AI will be deeply embedded into 3D-O-RAN, it becomes fundamental to ensure that its AI algorithms will *certifiably* achieve Key Performance Indicators (KPIs) such as latency, accuracy, and robustness. To this end, it would be highly unrealistic to assume that the AI will *always* deliver the same KPIs as when first deployed, especially when operating in congested, contested, concealed and contaminated (4C) environments. For this reason, the CC will dynamically modify the AI contained in the NC and MC according to the current applications and systems constraints, thus effectively implementing *polymorphic*, *adaptable*, and *certifiable* AI algorithms.

III. DYNAMIC CERTIFIABLE AI IN 3D-O-RAN

Mission-critical requirements on end-to-end deep neural network (DNN) latency and accuracy of operations are expected to continuously change in real-world tactical scenarios. 3D-O-RAN is well poised to be responsive to data availability variations. The key challenge that sets 3D-O-RAN apart from traditional systems is that mission-critical requirements have to be satisfied in highly-dynamic, highly-contested and resource-constrained environments, where (i) system input data (both images/frames and spectrum waveforms), as well as the wireless channel, will be subject to noise/interference, both intentional and unintentional; (ii) the available computational resources, both in terms of hardware and software, are extremely limited and may change dynamically. This implies that dynamic neural networks architectures have to be certified to meet AI-specific constraints (e.g., accuracy), device-specific constraints (e.g., energy/resource consumption) and mission-critical constraints (e.g., end-to-end latency).

Figure 2 shows the block diagram of our proposed Dynamic Data Driven DNN Certification System (in short, 4D-CS). The key innovation of 4D-CS is that mission-driven objectives such as end-to-end latency and accuracy are translated into dynamic technical constraints on network latency, hardware/software (HW/SW) resources, and DNN-level accuracy. Ultimately, the objective of 4D-CS is to establish a dynamic AI certification system, where the AI adapts itself to heterogeneous objectives. Specifically, 4D-CS operates within a data-driven control loop, where technical constraints are continuously fed to a dynamic DNN search engine (DSE), whose objective is to refine the DNN architecture throughout the system lifetime. A core challenge is that in 3D-O-RAN the inference process is distributed and the DNN input data is subject to noise/interference, which further complicates the DNN certification process. In stark contrast with existing work, the DSE will be tasked to dynamically find the right DNN architecture given (i) current 3D-O-RAN computation/network slicing allocation; (ii) current HW/SW resources on the HW platform; and (iii) current noise/interference levels in the DNN input. For this

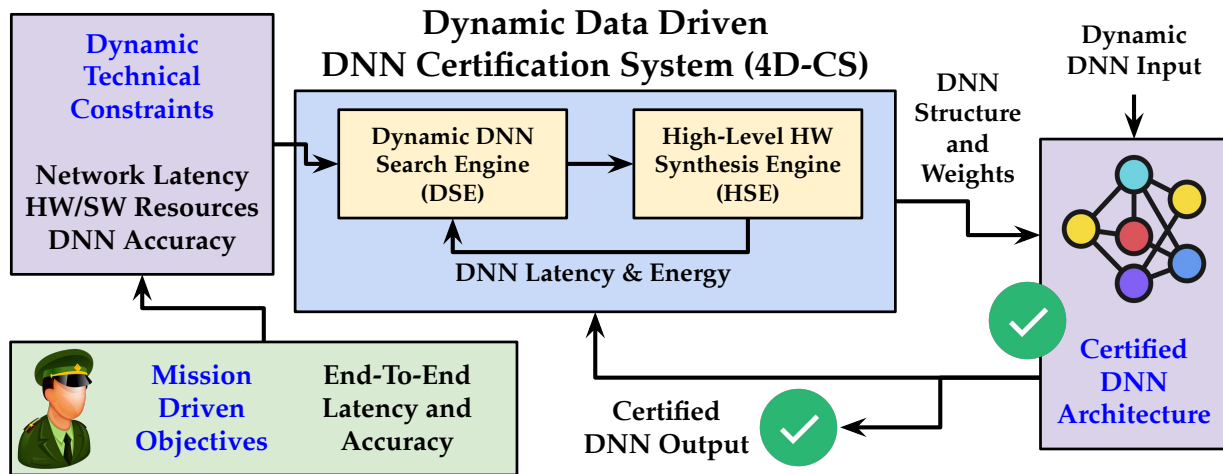


Figure 2: AI Certification System through Dynamic Data Driven DNN Design and Optimization.

reason, the DSE leverages a high-level HW synthesis engine (HSE) to translate a software-defined neural network to an field-programmable gate array (FPGA)-compliant circuit [17], as shown in prior work [18]–[20]. The optimal DNN structure is then incorporated into the DNN for real-time operation in the field. 4D-CS is fed the certified DNN output to determine whether the DNN is still maintaining acceptable performance.

IV. 3D-O-RAN: A TACTICAL APPLICATION SCENARIO

We consider the application scenario depicted in Figure 3, where User Equipments (UEs) and Mobile Sensors (MSs) are deployed in a spectrum sharing zone (SSZ). It is assumed that UEs and MSs are trusted devices – e.g., through traditional cryptography authentication – and may connect to 3D-O-RAN for 5/6G network access. Such trusted devices coexist with Concealed Assets (CAs), whose characteristics such as physical location (e.g., GPS coordinates) and spectrum access technique (e.g., waveform type, modulation, etc) are unknown to the trusted devices. Notice that CAs may be either allied or adversarial devices. The challenge is that trusted devices should identify hostile devices and deny them access to the shared spectrum. 3D-O-RAN is compliant to O-RAN specification, and thus disaggregated into a radio unit (RU), a distributed unit (DU) and a central unit (CU) controlled by a near real-time RAN intelligent controller (RIC).

The UEs contribute to the spectrum sensing process by submitting the sensed spectrum state to the non-real time (NRT) and near-real-time (RT) RICs of 3D-O-RAN. The NRT-RIC collects this information to train data-driven models for classification and control. The RT-RIC leverages the 3D-O-RAN Controller (RANC) to perform beam shaping, network slicing, and MS control. We assume that for sensing, navigation and surveillance purposes, the MSs will generate streams of multimedia sensing data – i.e., video frames and/or audio samples. Frames and samples are then analyzed by state-of-the-art DNNs running at the edge for real-time situational awareness and control [21]. For example, although paths are usually pre-selected, drones are always at risk of significant drift due to adverse weather conditions, loss of power

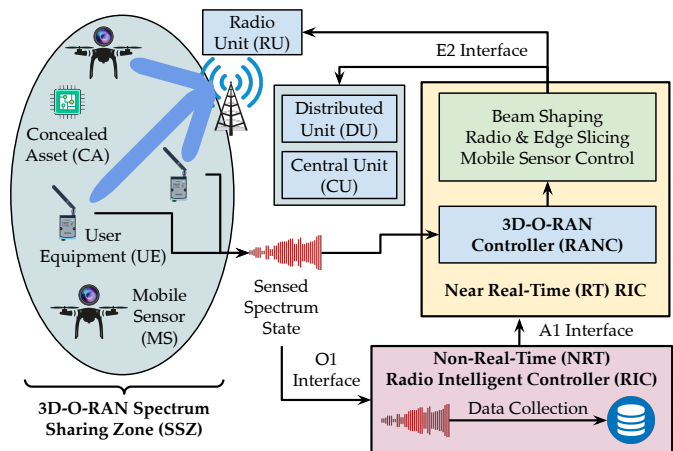


Figure 3: System model and key components of 3D-O-RAN.

and/or GPS connectivity [22], [23]. Therefore, drones require constant monitoring of their surroundings to detect sudden obstacles and correct the trajectory in real time [24], [25]. The fundamental issue is that on one hand, the MSs should submit the highest possible data quality, in order to guarantee the best possible inference through DNNs. On the other hand, continuously streaming high-quality multimedia data could quickly saturate the radio and edge computation capabilities of 3D-O-RAN. Thus, 3D-O-RAN includes the data semantics into the optimization process.

V. 3D-O-RAN: EXPERIMENTAL RESULTS

Mission-critical operations such as obstacle detection and object tracking will depend on the execution of complex computer vision (CV)-based algorithms based on LiDAR or wide-area motion imagery (WAMI) sensor data [26], [27]. However, continuously sending high-resolution multimedia data may eventually saturate the radio resources of 3D-O-RAN. For example, a WAMI image is usually 144 megapixels (216 MB in Lossless JPEG format). **If CV-based inference needs to be executed each second by 100 MSs, the traffic load on 3D-O-RAN would be 21 GB/s.** Moreover, some

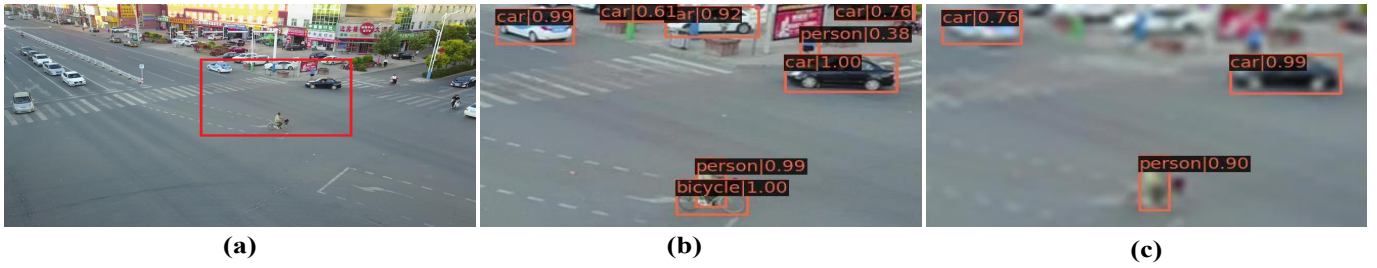


Figure 4: Example of different compression tolerance for different object classes. The original image size is 62.4 KB, and is shown in subfigure “a”. We apply the YOLOX model on the images. Lighter compression rates (0.47x, 29.5 KB output) makes the objects still detectable (subfigure “b”). With stronger compression (0.04x, 2.3 KB output), some objects (e.g., bicycle) are not detected anymore (subfigure “c”).

MSs may require simpler DNN tasks, while others may require significantly more complex tasks, e.g., target recognition [28]. To this end, network/edge slicing allows to virtualize and allocate computational and networking resources according to mission-driven needs [29]–[33]. Existing work in RAN slicing either does not support O-RAN or treats CV tasks equally, which eventually leads to overprovisioning of resources and/or suboptimal performance.

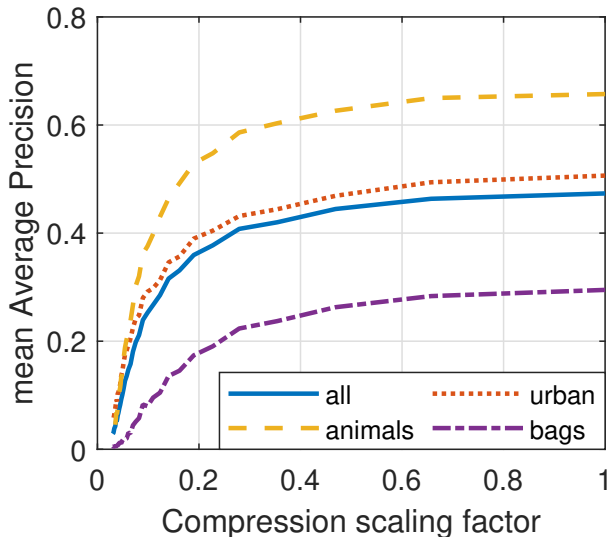


Figure 5: Mean Average Precision (mAP) as a function of the compression scaling factor for different application classes in the COCO dataset.

Our intuition is that object classification tasks have not only different accuracy and delay constraints, but can also be inherently “harder” or “simpler”. For example, as shown in Figure 4, classifying cars at an intersection is semantically less difficult than classifying bicycles, thus car images can be compressed more aggressively than bicycle images. Thus, we can semantically differentiate and compress the images sent to the 3D-O-RAN edge accordingly. To this point, Figure 5 shows the mean Average Precision (mAP) obtained on the widely-known Common Objects in Context (COCO) dataset [34] with the state-of-the-art object detection model YOLOX [35]. For the “urban” applications, we selected the classes *bicycle*, *car*, *motorcycle*, *bus*, *truck*, *traffic light*, *stop sign*, *person*; for “bags”: *handbag*, *backpack*, *suitcase*; for “animals”: *bird*,

cat, *dog*, *horse*, *sheep*, *cow*, *elephant*, *bear*, *zebra*, *giraffe*. We notice that the “animals” class is the “easiest” to be classified, while “bags” is the hardest. Therefore, different compression values may be used to reach the same average mAP. Moreover, tactical applications are expected to be extremely heterogeneous. An autonomous navigation mission could require image segmentation tasks with a minimum expected accuracy of 0.90 mAP and maximum expected delay of 10 ms, while object surveillance might have less stringent accuracy and delay requirements (0.75 and 100 ms). **Thus, semantic slicing is fundamental for proper optimization.** In other words, 3D-O-RAN semantically differentiates tasks and compresses the data accordingly, thus saving networking resources. For example, by assuming an average image size of 100KB and task rate of 50 frames/sec, a **semantic-agnostic slicing algorithm would generate 40 Mbit/s of traffic**. Conversely, with our semantic-based approach, if the class “animals” is selected with 0.4 mAP, then the images can get compressed by a factor of 0.1, **thus reducing by 10 times the traffic**.

VI. RELATED WORK

Thanks to its promise of creating new markets in the telecommunication landscape while drastically improving network performance, Radio Access Network (RAN) slicing has attracted a significant amount of attention over the last years [29], [30], [32]. Moreover, as the RAN gets softwarized, Mobile Edge Cloud (MEC) becomes fundamental to address the ever-stringent latency demands of mobile applications [36], [37]. We refer the interested reader to the following surveys [38], [39]. Specific to slicing of edge resources, Castellano *et al.* [40] proposed a framework to allow the instantiation of multiple virtual network functions on the same infrastructure. However, the framework in [40] was proposed for wired networks only. Van Huynh *et al.* [41] presented a mechanism for slicing of computation, networking and storage through a deep dueling neural network that provides slices admission while avoiding over-provisioning and maximizing the network operator’s reward. Ndikumana *et al.* [42] consider the allocation of heterogeneous resources for MEC task offloading, while in [43] Liu *et al.* propose a framework for MEC-enabled wireless networks called Distributed Cross-Domain Resource Orchestration in Cellular Edge Computing (DIRECT), which however does not consider the case when

MEC and networking resources are on the same edge node. The latter has been investigated by Blasch et al. [44], [45], who modeled the compression-induced information loss in terms of the National Imagery Interpretability Rating Scale (NIIRS), and presented the Compression Degradation Image Function Index (CoDIFI) to predict the NIIRS degradation. However, these approaches do not consider the effect of compression on CV/DL tasks, nor are they O-RAN-compatible, which is one of the primary objectives of this paper.

Recently, Bonati *et al.* [46] have developed deep reinforcement learning (DRL) agents running in O-RAN xApps on the near-real-time RIC to select the best-performing scheduling policy for each RAN slice. In our work, we do not select scheduling policies but instead focus on RAN slicing. D'Oro *et al.* [47] proposed an orchestration mechanism to select the optimal deep learning (DL) models and execution location for each model complying with timescale requirements, resource and data availability. Conversely, we focus on properly slicing end/network resources for timely execution of CV-based DL models under strict accuracy constraints. Existing work of dynamic AI can be broadly classified under the term neural architecture search (NAS) [48], [49], which attempts to find the best neural architecture for a given classification problem. Usually, NAS algorithms operate under the assumption of unlimited computational resources [50]–[52], which is not the case in 3D-O-RAN. Hardware-aware NAS approaches where latency [53], [54] or energy [55] are optimized in resource-constrained devices [56]–[58] have been proposed [59]. However, these approaches assume static datasets and static classification tasks. Moreover, they are not taking into account uncertainty in the wireless communication process, which require lifelong adjustment of the neural network.

VII. CONCLUSIONS

This paper introduces the Dynamic Data Driven Open Radio Access Network System (3D-O-RAN) to integrate computational, sensing, and networking components in a control loop. 3D-O-RAN is compliant with the O-RAN specification for 5G and beyond systems to reduce costs and guarantee interoperability among vendors. 3D-O-RAN incorporates heterogeneous data into the control loop to dynamically achieve a system-wide optimal operating point, and steers the multimedia sensor measurement process in real time according to the required application needs and current physical and/or environmental constraints. 3D-O-RAN includes a semantic slicing engine and a dynamic data driven neural network certification system. Finally, we have provided a realistic use-case scenario of 3D-O-RAN in the context of a tactical setting.

ACKNOWLEDGMENT AND DISCLAIMER

This work is funded in part by the AFRL Visiting Faculty Research Program (VFRP) contract number FA8750-20-3-1003, as well as National Science Foundation (NSF) grant CNS-2134973. Any opinions, findings and conclusions or recommendations expressed in this material are those of the

authors and do not necessarily reflect the views of the NSF or the U.S. Government.

REFERENCES

- [1] Office of the Federal Register, National Archives and Records Administration, "DCPD-201800730 - Memorandum on Developing a Sustainable Spectrum Strategy for America's Future," <https://www.govinfo.gov/app/details/DCPD-201800730>. Full Report available at https://assets.ctfassets.net/3nanhbfr0pc/acoo1Fj5uungnGNPJ3QWY/3a1dafd64f22efcf8f27380aafae9789/2021_RAI_Report-v3.pdf, October 2018.
- [2] J. R. Hoehn, "Joint All Domain Command and Control (JADC2)," <https://crsreports.congress.gov/product/pdf/IF/IF11493>, 2022.
- [3] —, "Joint All Domain Command and Control: Background and Issues for Congress," <https://crsreports.congress.gov/product/pdf/R/R46725>, 2022.
- [4] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," *arXiv:2202.01032 [cs.NI]*, 2022. [Online]. Available: <https://arxiv.org/abs/2202.01032>
- [5] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "Open, Programmable, and Virtualized 5G Networks: State-of-the-Art and the Road Ahead," *Computer Networks*, vol. 182, pp. 1–28, December 2020.
- [6] L. Baldesi, F. Restuccia, and T. Melodia, "ChARM: NextG Spectrum Sharing Through Data-Driven Real-Time O-RAN Dynamic Control," *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, 2022.
- [7] Congress.gov, "H.R.4521 - America COMPETES Act of 2022," February 2022.
- [8] D. C. Tarraf, W. Shelton, E. Parker, B. Alkire, D. G. Carew, J. Grana, A. Levedahl, J. Leveille, J. Mondschein, J. Ryseff *et al.*, "The Department of Defense posture for Artificial Intelligence: Assessment and Recommendations," <https://apps.dtic.mil/sti/citations/AD1088616>, 2019.
- [9] E. Blasch, J. Bin, and Z. Liu, "Certifiable Artificial Intelligence Through Data Fusion," Presented at AAAI FSS-21: Artificial Intelligence in Government and Public Sector, Washington, DC, USA, 2021. [Online]. Available: <https://arxiv.org/abs/2111.02001>
- [10] U.S. Department of Defense, "DOD Adopts Ethical Principles for Artificial Intelligence," <https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/>, February 2020.
- [11] E. Schmidt, B. Work, S. Catz, S. Chien, C. Darby, K. Ford, J.-M. Griffiths, E. Horvitz, A. Jassy, W. Mark *et al.*, "Final Report, National Security Commission on Artificial Intelligence (AI)," <https://apps.dtic.mil/sti/citations/AD1124333>, 2021.
- [12] Department of Defense, "Defense Innovation Unit Publishes 'Responsible AI Guidelines'," <https://www.defense.gov/News/News-Stories/Article/Article/2847598/defense-innovation-unit-publishes-responsible-ai-guidelines/>, November 2021.
- [13] F. Darema, "Dynamic Data Driven Applications Systems: A New Paradigm for Application Simulations and Measurements," in *International Conference on Computational Science*. Springer, 2004, pp. 662–669.
- [14] E. Blasch, G. Seetharaman, and K. Reinhardt, "Dynamic Data Driven Applications System Concept for Information Fusion," *Procedia Computer Science*, vol. 18, pp. 1999–2007, 2013.
- [15] E. Blasch, G. Seetharaman, and F. Darema, "Dynamic Data Driven Applications Systems (DDAS) Modeling for Automatic Target Recognition," in *Automatic Target Recognition XXIII*, vol. 8744. International Society for Optics and Photonics, 2013, p. 87440J.
- [16] E. Blasch, S. Ravela, and A. Aved, *Handbook of Dynamic Data Driven Applications Systems*. Springer, 2018.
- [17] F. Winterstein, S. Bayliss, and G. A. Constantinides, "High-level synthesis of dynamic data structures: A case study using vivado hls," in *Proc. of International Conference on Field-Programmable Technology (FPT)*, Kyoto, Japan, 2013, pp. 362–365.
- [18] F. Restuccia and T. Melodia, "Big Data Goes Small: Real-time Spectrum-driven Embedded Wireless Networking Through Deep Learning in the RF Loop," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2019, pp. 2152–2160.

- [19] —, “DeepWiERL: Bringing Deep Reinforcement Learning to the Internet of Self-Adaptive Things,” *Proceedings of IEEE INFOCOM*, 2020.
- [20] —, “PolymoRF: Polymorphic Wireless Receivers Through Physical-Layer Deep Learning,” *Proceedings of ACM MobiHoc*, 2020.
- [21] A. Munir, J. Kwon, J. H. Lee, J. Kong, E. Blasch, A. J. Aved, and K. Muhammad, “FogSurv: A Fog-Assisted Architecture for Urban Surveillance Using Artificial Intelligence and Data Fusion,” *IEEE Access*, vol. 9, pp. 111 938–111 959, 2021.
- [22] L. Gupta, R. Jain, and G. Vaszkun, “Survey of Important Issues in UAV Communication Networks,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1123–1152, 2015.
- [23] S. Hayat, E. Yanmaz, and R. Muzaffar, “Survey on Unmanned Aerial Vehicle Networks for Civil Applications: A Communications Viewpoint,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2624–2661, 2016.
- [24] D. Wang, W. Li, X. Liu, N. Li, and C. Zhang, “UAV Environmental Perception and Autonomous Obstacle Avoidance: A Deep Learning and Depth Camera Combined Solution,” *Computers and Electronics in Agriculture*, vol. 175, p. 105523, 2020.
- [25] P. Fraga-Lamas, L. Ramos, V. Mondéjar-Guerra, and T. M. Fernández-Caramés, “A Review on IoT Deep Learning UAV Systems for Autonomous Obstacle Detection and Collision Avoidance,” *Remote Sensing*, vol. 11, no. 18, p. 2144, 2019.
- [26] R. Ravindran, M. J. Santora, and M. M. Jamali, “Multi-Object Detection and Tracking, based on DNN, for Autonomous Vehicles: A Review,” *IEEE Sensors Journal*, vol. 21, no. 5, pp. 5668–5677, 2020.
- [27] M. K. Vasić, A. Drak, N. Bugarin, S. Kružić, J. Musić, C. Pomrehn, M. Schöbel, M. Johenneken, I. Stančić, V. Papić *et al.*, “Deep Semantic Image Segmentation for UAV-UGV Cooperative Path Planning: A Car Park Use Case,” in *2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, 2020, pp. 1–6.
- [28] U. Majumder, E. Blasch, and D. Garren, “Deep Learning for Radar and Communications Automatic Target Recognition,” 2020.
- [29] S. D’Oro, L. Bonati, F. Restuccia, M. Polese, M. Zorzi, and T. Melodia, “SI-EDGE: Network Slicing at the Edge,” in *Proceedings of the Twenty-First International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, 2020, pp. 1–10.
- [30] S. Mandelli, M. Andrews, S. Borst, and S. Klein, “Satisfying Network Slicing Constraints via 5G MAC Scheduling,” in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2019, pp. 2332–2340.
- [31] V. Mancuso, P. Castagno, M. Sereno, and M. A. Marsan, “Slicing Cell Resources: The Case of HTC and MTC Coexistence,” in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2019, pp. 667–675.
- [32] S. D’Oro, F. Restuccia, A. Talamonti, and T. Melodia, “The Slice is Served: Enforcing Radio Access Network Slicing in Virtualized 5G Systems,” in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2019, pp. 442–450.
- [33] G. Garcia-Aviles, M. Gramaglia, P. Serrano, and A. Banchs, “POSENS: A Practical Open Source Solution for End-to-End Network Slicing,” *IEEE Wireless Communications*, vol. 25, no. 5, pp. 30–37, 2018.
- [34] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft COCO: Common Objects in Context,” in *Proceedings of European Conference on Computer Vision (ECCV)*. Springer, 2014, pp. 740–755.
- [35] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, “YOLOX: Exceeding YOLO Series in 2021,” *arXiv preprint arXiv:2107.08430*, 2021.
- [36] J. Wang, J. Hu, G. Min, W. Zhan, Q. Ni, and N. Georgalas, “Computation Offloading in Multi-Access Edge Computing Using a Deep Sequential Model Based on Reinforcement Learning,” *IEEE Communications Magazine*, vol. 57, no. 5, pp. 64–69, May 2019.
- [37] J. Zhang, H. Guo, J. Liu, and Y. Zhang, “Task Offloading in Vehicular Edge Computing Networks: A Load-Balancing Solution,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2092–2104, 2019.
- [38] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, “Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies, and Solutions,” *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2429–2453, 2018.
- [39] S. Wijethilaka and M. Liyanage, “Survey on Network Slicing for Internet of Things Realization in 5G Networks,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 957–994, 2021.
- [40] G. Castellano, F. Esposito, and F. Risso, “A Distributed Orchestration Algorithm for Edge Computing Resources with Guarantees,” in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2019, pp. 2548–2556.
- [41] N. Van Huynh, D. T. Hoang, D. N. Nguyen, and E. Dutkiewicz, “Optimal and Fast Real-Time Resource Slicing with Deep Dueling Neural Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1455–1470, 2019.
- [42] A. Ndikumana, N. H. Tran, T. M. Ho, Z. Han, W. Saad, D. Niyato, and C. S. Hong, “Joint Communication, Computation, Caching, and Control in Big Data Multi-Access Edge Computing,” *IEEE Transactions on Mobile Computing*, vol. 19, no. 6, pp. 1359–1374, 2019.
- [43] Q. Liu and T. Han, “DIRECT: Distributed Cross-Domain Resource Orchestration in Cellular Edge Computing,” in *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*. ACM, 2019, pp. 181–190.
- [44] E. Blasch, H.-M. Chen, J. M. Irvine, Z. Wang, G. Chen, J. Nagy, and S. Scott, “Prediction of Compression-Induced Image Interpretability Degradation,” *Optical Engineering*, vol. 57, no. 4, p. 043108, 2018.
- [45] E. P. Blasch, H.-M. H. Chen, W. Zhonghai, G. Chen, K. Liu, and D. Shen, “Image Compression Selection based on Interpretability Loss Estimation,” Nov. 3 2020, uS Patent 10,827,190.
- [46] L. Bonati, S. D’Oro, M. Polese, S. Basagni, and T. Melodia, “Intelligence and Learning in O-RAN for Data-driven NextG Cellular Networks,” *IEEE Communications Magazine*, vol. 59, no. 10, pp. 21–27, October 2021.
- [47] S. D’Oro, L. Bonati, M. Polese, and T. Melodia, “OrchestRAN: Network Automation through Orchestrated Intelligence in the Open RAN,” in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, May 2022.
- [48] T. Elsken, J. H. Metzen, and F. Hutter, “Neural Architecture Search: A Survey,” *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 1997–2017, 2019.
- [49] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated Machine Learning: Methods, Systems, Challenges*. Springer Nature, 2019.
- [50] B. Zoph and Q. V. Le, “Neural Architecture Search with Reinforcement Learning,” *Proceedings of International Conference on Learning Representations (ICLR)*, 2017.
- [51] C. Liu, B. Zoph, M. Neumann, J. Shlens, W. Hua, L.-J. Li, L. Fei-Fei, A. Yuille, J. Huang, and K. Murphy, “Progressive Neural Architecture Search,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 19–34.
- [52] J. Mellor, J. Turner, A. Storkey, and E. J. Crowley, “Neural architecture search without training,” in *International Conference on Machine Learning*. PMLR, 2021, pp. 7588–7598.
- [53] H. Cai, L. Zhu, and S. Han, “ProxylessNAS: Direct Neural Architecture Search on Target Task and Hardware,” *arXiv preprint arXiv:1812.00332*, 2018.
- [54] J. Fernandez-Marques, P. Whatmough, A. Mundy, and M. Mattina, “Searching for Winograd-aware Quantized Networks,” *Proceedings of Machine Learning and Systems*, vol. 2, pp. 14–29, 2020.
- [55] C.-H. Hsu, S.-H. Chang, J.-H. Liang, H.-P. Chou, C.-H. Liu, S.-C. Chang, J.-Y. Pan, Y.-T. Chen, W. Wei, and D.-C. Juan, “MONAS: Multi-objective Neural Architecture Search Using Reinforcement Learning,” *arXiv preprint arXiv:1806.10332*, 2018.
- [56] I. Fedorov, R. P. Adams, M. Mattina, and P. Whatmough, “SpArSe: Sparse Architecture Search for CNNs on Resource-Constrained Microcontrollers,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [57] J. Lin, W.-M. Chen, Y. Lin, C. Gan, S. Han *et al.*, “MCUNet: Tiny Deep Learning on IoT Devices,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 11 711–11 722, 2020.
- [58] E. Liberis, Ł. Dudziak, and N. D. Lane, “μNAS: Constrained Neural Architecture Search for Microcontrollers,” in *Proceedings of the 1st Workshop on Machine Learning and Systems*, 2021, pp. 70–79.
- [59] K. T. Chitty-Venkata and A. K. Somani, “Neural Architecture Search Survey: A Hardware Perspective,” *ACM Computing Surveys (CSUR)*, 2022.