2021 Final Project Report
for
Academic Consortium for the 21st Century (AC21)
Special Project Fund

Towards Scalable and Resilient Cyber Infrastructure for Mission-Critical UAV-Enabled Wireless Networks

Project Group Leader: Dr. Shih-Chun Lin, Assistant Professor

North Carolina State University

Date of Activities: March 2021 – February 2022

Contents

Project Abstract

Unmanned aerial vehicles (UAVs) for wireless communications have drawn much attention as the mass production of high-performance, low-cost, intelligent UAVs has become more practical and feasible. With the rapid technological advance and wide-range deployment of UAVs, UAV communication networks have been considered a quick solution for post-disaster communications, military defense applications, and diverse commercial scenarios. Led by North Carolina State University, aka NC State, this AC21 SPF 2021 project aims at UN's SDGs #9 (Industry, Innovation, and Infrastructure). This work focused on initiating collaborative research discussion and external grant planning for developing scalable and resilient cyberinfrastructures for UAV-enabled wireless networks in mission-critical services. Dr. Shih-Chun Lin (NC State), Dr. Peng Shi (University of Adelaide), Dr. Masaaki Katayama (Nagoya University), and Dr. Ta-Sung Lee (National Chiao Tung University) joined their efforts in this project with outstanding accomplishments in 2021-2022. These achievements include several seminar talks and teleconferences, technical paper publications, and workshop organization.

Acknowledgment

The project team greatly acknowledges AC21 General Secretariat and the AC21 Special Project Fund (SPF) for their kind help and generous financial support.

Project Description

This project's primary objective is to initiate joint research and workshop on developing scalable and resilient cyberinfrastructures for mission-critical UAV-enabled networks. Team members investigated three-dimensional (3D) fifth-generation (5G) UAV air interface, intersystem control mechanisms, dynamic spectrum sharing, and energy-efficient system orchestration. We discussed working tasks to establish an experimental testbed and collect preliminary data with agricultural and industrial IoT applications to demonstrate the planned research activities. This research will empower the large-scale deployment of communication infrastructure in rural areas or refugee camps with the promised performance of high reliability, low latency, and cost-efficiency. The following planned activities promote creating a community interested in this new opportunity.

As COVID-19 was severe in Asia in 2021 and several places had active restrictions for international travel, most physical meetings and workshops had changed to virtual events. Six virtual talks and seminars have been conducted about these autonomous vehicles and their applications.

- "Towards 6G Smart Fab with Service-Level Agreement Assurance and Reconfigurable Multi-Robot Task Assignment," *TSMC, Ltd.*, Virtual Event, February 22, 2022.
- "Realizing 6G Smart Manufacturing with Service-Level Agreement Assurance and Reconfigurable Multi-Robot Task Assignment," *Department of Computer Science, National Chung Cheng University*, Virtual Event, December 6, 2021.
- "In-Network Computation for Distributed Privacy-Preserving Learning over Wireless Edges," Cisco Systems, Inc., Virtual Event, June 24, 2021.
- "6G Industry Verticals: Connected and Autonomous Vehicle Technologies with Ultra-Low Latency," *Department of Computer Science, National Chung Cheng University*, Virtual Event, May 31, 2021.
- "Beyond 5G Wireless Systems: Industry Trends and Taiwan's Advantages," *ROC National Academy of Civil Service*, Virtual Event, May 12, 2021.
- "6G Industry Verticals: Machine Learning-Enabled and Ultra-Low Latency Connected Transportation," *NC-CAV Seminar Series*, Virtual Event, March 26, 2021.

Figure 1 shows one educational impact of this project, where Dr. Lin delivered a seminar talk on beyond 5G wireless systems in Taiwan's National Academy of Civil Service. Around five hundred government staff attended Dr. Lin's seminar online and learned the industry trends and state-of-the-art development in the information and communication technology (ICT) domain. Figure 2 gives a screenshot of Dr. Lin's virtual talk of "6G industry verticals" in May at National Chung Cheng University. This invited talk was one of the annual seminars for students in the school's computer science department. Dr. Lin presented the research and development insights of connected and autonomous vehicles (CAVs), which were summarized from discussions and meetings among our AC21 project members. This presentation shows our substantive engagement in transportation engineering, radio and antenna designs, system programming, telecommunications, and wireless networking.

Figure 1. Dr. Lin gave a "beyond 5G systems" seminar in Taiwan's National Academy of Civil Service.

Figure 2. Dr. Lin delivered a "6G industry verticals" virtual seminar at National Chung Cheng University.

As the pandemic became less severe by the end of 2021, Dr. Lin conducted one-month research and education exchange in Taiwan. Several intense discussions and meetings were arranged to generate this project's outcomes. Two joint papers about industrial IoT and open-RAN were developed and submitted to top IEEE magazines based on these engagements. For example, another similar talk of "6G smart manufacturing" was delivered at National Chung Cheng University in December 2021. The details are provided in the attached flyer of this report. We plan to expand the initial results and seek external grants to sustain this project and reinforce AC21 members' strategic research partnerships.

In addition, we published two conference papers and two journal papers to prestigious IEEE proceedings for research activities. Three magazine articles were also submitted to the IEEE and are currently under review. The list is provided below.

- S.-C. Lin, K.-C. Chen, and A. Karimoddini, "SDVEC: Software-Defined Vehicular Edge Computing with Ultra-Low Latency," *IEEE Communications Magazine*, vol. 59, no. 12, pp. 66-72, December 2021.
- C.-H. Lin, Y.-H. Fang, H.-Y. Chang, Y.-C. Lin, W.-H. Chung, S.-C. Lin, and T.-S. Lee, "GCN-CNVPS: Novel Method for Cooperative Neighboring Vehicle Positioning System Based on Graph Convolution Network," *IEEE Access*, vol. 9, pp. 153429-153441, November 2021.
- C.-H. Lin, S.-C. Lin, C.-Y. Wang, and T. Chase, "A C-V2X Platform Using Transportation Data and Spectrum Aware Slidelink Access," in *Proc. of IEEE SMC*

Workshop, Virtual Conference, October 2021.

- C.-H. Lin, S.-C. Lin, and E. Blasch, "TULVCAN: Terahertz Ultra-broadband Learning Vehicular Channel-Aware Networking," in *Proc. of IEEE INFOCOM Workshop*, Virtual Conference, May 2021.
- S.-C. Lin, C.-H. Lin, L. C. Chu, and S.-Y. Lien, "Towards Resilient Access Equality for 6G Serverless p-LEO Satellite Networks," under review, 2022.
- S.-Y. Lien, Y.-C. Huang, C.-C. Tseng, and S.-C. Lin, "Universal Vertical Application Adaptation for Open RAN: Sustainable RIC and Intelligent xAPPs," under review, 2022.
- S.-C. Lin, C.-H. Lin, and W.-C. Chen, "Zero-Touch Network on Industrial IoT: An End-to-End Machine Learning Approach," under review, 2022.

Both conference presentations are held virtual. Notably, the TULVCAN work was published by the top-tier communications conference, the IEEE International Conference on Computer Communications (INFOCOM). It introduces novel terahertz (THz) channel-aware spectrum learning solutions that fully disclose the uniqueness of THz channels when performing such ultra-broadband sensing in future vehicular environments. Numerical results show that our solutions outperform the latest generative adversarial network (GAN) realization with higher cosine and structure similarity measures, more minor learning errors, and 56% fewer required training overheads. Besides, we introduce a cellular vehicle-to-everything (C-V2X) verification platform based on an actual traffic simulator and spectrum-aware access. The developed slidelink communication can adopt different operating bands with remarkable spectrum detection performance, validating its practicality in real-world vehicular environments.

Moreover, after a few rounds of discussions and thorough preparation, Dr. Lin and several faculties successfully held a "Special Session on Connected and Autonomous Vehicles" in the IEEE Systems, Man, and Cybernetics Society (SMC) 2021. As shown in Figure 3, this special session focused on state-of-the-art and practice solutions for addressing critical challenges in developing and operationalizing CAVs. The open issues include efficient computer vision algorithms for autonomous vehicle perception systems, communication backbone for CAVs' connectivity, traffic volume prediction techniques, the latest connected autonomous vehicle testbeds, and CAV applications. The session was initially arranged in Melbourne, Australia, but went virtual later due to the COVID-19 situation. Also, spanning this SPF project, Dr. Lin (NC State) and Dr. Lee (National Chiao Tung University) plan to establish a Ph.D. student visiting program between labs at the two institutions. This program will undoubtedly enhance the cross-linkage between the lines of research pursued by the participating members and strengthen the AC21 network in the long term.

Figure 3. IEEE SMC 2021 special session on connected and autonomous vehicles, October 17-20, 2021.

Activities and Reports

Mar. 2021	Had a virtual discussion about a collaborative project roadmap, possible student exchanges, and proposal preparation for potential external grants. • Team members initiated teleconferencing meetings to discuss a detailed timeline and tasks for the research activities.
Apr. 2021	Conducted research on ultra-broadband vehicular learning and software-defined vehicular edge infrastructures. • Prepared manuscripts for IEEE INFOCOM workshop and IEEE Communications Magazine, based on the initial discussions.
May 2021	Delivered "Beyond 5G Wireless Systems" talk to the ROC National Academy of Civil Service.
Jun. 2021	Engaged with Cisco Systems, Inc. to extend this project in wireless edge scenarios.
Jul. 2021	Conducted research on C-V2X verification platform. Prepared a manuscript for IEEE SMC.
Aug. 2021	Conducted research on cooperative neighboring vehicle positioning systems. Prepared a manuscript for IEEE Access.
Sep. 2021	Prepared a special session and invite speakers. • PIs organized the special session on Connected and Autonomous Vehicles in IEEE SMC 2021 http://ieeesmc2021.org/ • It enhanced the cross-linkage among research lines in PIs' labs.
Oct. 2021	Organized the IEEE SMC 2021 special session. • Hosted invited talk(s) and chair presentation and discussion. • Advertised the planning works and sought feedbacks from academic and industrial partners. • The special session improved the international profile of the AC21 consortium and the PIs' institutions.
Nov. 2021	Addressed the review comments and resubmitted the IEEE Access and IEEE Communications Magazine papers.
Dec. 2021 – Jan. 2022	One-month visiting for research and education engagements. • Discussed potential student exchange opportunities. • Visited non-AC21 institutions to deliver forums about the planned collaborative research. • Discussed virtually proposal submissions for external grant for the project sustainability, demonstrating a plan for the project's scope and impact beyond the AC21 grant period. • Discussed virtually the feasibility of prototyping a UAV-enabled network testbed to demonstrate mission-critical services. • Reinforced the complementary strengths of the participated institutions' strategic partnership.

Feb. 2022	Wrapped up the project. • Provided a final report and technical papers. • This collaborative research will empower the large-scale deployment of UAV-enabled applications with the promised performance of high reliability, low latency, and cost-efficiency.

Achievement of Activities

The longstanding relationship and strategic partnership among NC State, UoA, and NU leverage complementary strengths and transdisciplinary scholarship to advance research collaboration and academic exchanges. By jointly planning collaborative research activities, the project reinforced three university linkages through communications between PIs' teams. The PIs used essential conference forums, e.g., IEEE INFOCOM and IEEE SCC, to advance industrial partnerships and achieve funding success. This project stimulated international collaboration with the following contributions and broader impacts on research, education, management, and international exchange.

Research: This project outcomes systematically accomplished scalable and resilient cyberinfrastructures for UAV-enabled networks in mission-critical services. It led researchers' and students' substantive engagement in transportation engineering, radio and antenna designs, system programming, telecommunications, and wireless networking. This discussion involved the multidisciplinary knowledge of 3D UAV channel and energy modeling for UAV hovering and flying, software-defined infrastructure designs for unified control and reliable communication, UAV placement, and self-contained system management.

Education: The project results were incorporated into Optimizations and Algorithms as well as Introduction to Computer Networking courses at undergraduate and graduate levels in the ECE Department at NC State, the school of EEE at AU, and the IMaSS at NU. The PIs also trained their Ph.D. students to become experts in this fast-evolving field and involved M.S. and undergraduate students in the proposed research by assigning them sub-problems to solve.

Management: The funding was used for organizing a special session in IEEE SMC 2021, seminars, and one-month research and education exchange, which bring together researchers in all relevant areas and improves the international profile of NC State, NU, and AU in cutting-edge telecommunication research, enabling the establishment of new research connections and interpretations.

International Exchange: The PIs also plan the visiting/exchange Ph.D. student program at their institutions. Exchanging students will enhance the cross-linkage among team members' research lines and further strengthen the AC21 network.

Conclusion

This project has successfully initiated joint research discussions, seminars, and workshop organization for mission-critical UAV-enabled wireless networks. Several academic and research activities were conducted to place participating AC21 members in a unique leading position. We foresee more engineers and researchers will engage in developing our proposed scalable and resilient cyberinfrastructures and continue to foster our strategic partnerships by seeking external funding based on this project's outcomes.

Achievements Made within SPF Collaboration

The achievements in seminars, publications, and workshops are summarized below. Detailed paper manuscripts are attached for reference.

Seminars:

- "Towards 6G Smart Fab with Service-Level Agreement Assurance and Reconfigurable Multi-Robot Task Assignment," *TSMC, Ltd.*, Virtual Event, February 22, 2022.
- "Realizing 6G Smart Manufacturing with Service-Level Agreement Assurance and Reconfigurable Multi-Robot Task Assignment," *Department of Computer Science, National Chung Cheng University*, Virtual Event, December 6, 2021.
- "In-Network Computation for Distributed Privacy-Preserving Learning over Wireless Edges," Cisco Systems, Inc., Virtual Event, June 24, 2021.
- "6G Industry Verticals: Connected and Autonomous Vehicle Technologies with Ultra-Low Latency," *Department of Computer Science, National Chung Cheng University*, Virtual Event, May 31, 2021.
- "Beyond 5G Wireless Systems: Industry Trends and Taiwan's Advantages," *ROC National Academy of Civil Service*, Virtual Event, May 12, 2021.
- "6G Industry Verticals: Machine Learning-Enabled and Ultra-Low Latency Connected Transportation," *NC-CAV Seminar Series*, Virtual Event, March 26, 2021.

Publications:

- S.-C. Lin, K.-C. Chen, and A. Karimoddini, "SDVEC: Software-Defined Vehicular Edge Computing with Ultra-Low Latency," *IEEE Communications Magazine*, vol. 59, no. 12, pp. 66-72, December 2021.
- C.-H. Lin, Y.-H. Fang, H.-Y. Chang, Y.-C. Lin, W.-H. Chung, S.-C. Lin, and T.-S. Lee, "GCN-CNVPS: Novel Method for Cooperative Neighboring Vehicle Positioning System Based on Graph Convolution Network," *IEEE Access*, vol. 9, pp. 153429-153441, November 2021.
- C.-H. Lin, S.-C. Lin, C.-Y. Wang, and T. Chase, "A C-V2X Platform Using Transportation Data and Spectrum Aware Slidelink Access," in *Proc. of IEEE SMC*

Workshop, Virtual Conference, October 2021.

- C.-H. Lin, S.-C. Lin, and E. Blasch, "TULVCAN: Terahertz Ultra-broadband Learning Vehicular Channel-Aware Networking," in *Proc. of IEEE INFOCOM Workshop*, Virtual Conference, May 2021.
- S.-C. Lin, C.-H. Lin, L. C. Chu, and S.-Y. Lien, "Towards Resilient Access Equality for 6G Serverless p-LEO Satellite Networks," under review, 2022.
- S.-Y. Lien, Y.-C. Huang, C.-C. Tseng, and S.-C. Lin, "Universal Vertical Application Adaptation for Open RAN: Sustainable RIC and Intelligent xAPPs," under review, 2022.
- S.-C. Lin, C.-H. Lin, and W.-C. Chen, "Zero-Touch Network on Industrial IoT: An End-to-End Machine Learning Approach," under review, 2022.

Workshops:

- IEEE Systems, Man, and Cybernetics Society (SMC) 2021 Special Session on Connected and Autonomous Vehicles, Virtual Conference, October 17-20, 2021.

中正大學 National Chung Cheng University

資訊工程學系演講

時間：110年12月6日 14:00 ~ 16:00 視訊演講

題目：Realizing 6G Smart Manufacturing with Service-Level Agreement Assurance and Reconfigurable Multi-Robot Task Assignment

演講者：林士鈞 教授
美國北卡羅萊納州立大學

內容大綱：

Smart manufacturing, also known as Industry 4.0, has emerged as a critical enterprise application of industry IoT. It aims at holistically integrating wireless networking, computing, and automatic control technologies to execute efficient and flexible production. The main objective is to realize networked multi-robot systems, which can rapidly respond to market demands by reconfiguring multi-robot task assignments. State-of-the-art practices utilize software-defined networking paradigms (or softwarization) to build programmable and flexible wireless infrastructures for industrial usages. However, beyond 5G standardizations (i.e., O-RAN adoption) in intelligent factories and their potentials of improving resource utilization and automating network management are still minimal. This seminar will present emerging critical aspects of the 6G smart factory and discuss its technological advance to assure service-level agreements. The talk will cover the current practices of multi-robot coordination, serverless intelligent factory realization, data-driven adaptive control and orchestration, and timely federated multi-task learning for factory workloads over edge infrastructure.

講者簡介：

Dr. Shih-Chun Lin is an assistant professor with the Department of Electrical and Computer Engineering at North Carolina State University, where he leads the Intelligent Wireless Networking (iWN) Laboratory. He received a Ph.D. degree in electrical and computer engineering from Georgia Institute of Technology, Atlanta, USA, in 2017, and an M.S. degree in communication engineering and a B.S. degree in electrical engineering from National Taiwan University, Taiwan. He has published more than 50 peer-reviewed papers, holds 11 U.S. patents, and received the Best Student Paper Award Runner-up in IEEE SCC 2016. As a pioneer of using software-defined networking models for wireless system management, Dr. Lin is researching the low-latency edge computing infrastructure for large-scale connected vehicle deployment and private O-RAN factories. His recent works on intelligent scheduling control and signal interference management have received several international awards, contributing to effective system orchestration for 6G industry verticals. His wireless networking and learning research is supported by many government agencies and industrial companies, including AFRL, NCDOT, CISCO, III, Lockheed Martin, and Academic Consortium 21. His research interests include beyond 5G architectures, mobile edge computing, AIoT, machine learning techniques, mathematical optimization, and performance evaluation.

歡迎蒞臨聽講

ACCEPTED FROM OPEN CALL

SDVEC: Software-Defined Vehicular Edge Computing with Ultra-Low Latency

Shih-Chun Lin, Kwang-Cheng Chen, and Ali Karimoddini

This article introduces novel wireless distributed architectures that embed the edge computing capability inside software-defined vehicular networking infrastructure. Such edge networks consist of open-loop grant-free communications and computing-based control frameworks, which enable dynamic eco-routing with ultra-low latency and mobile data-driven orchestration.

Abstract

New paradigm shifts and 6G technological revolution in vehicular services have emerged toward unmanned driving, automated transportation, and self-driving vehicles. As the technology for autonomous cars becomes mature, real challenges come from reliable, safe, real-time connected transportation operations to achieve ubiquitous and prompt information exchanges with massive connected and autonomous vehicles. This article introduces novel wireless distributed architectures that embed the edge computing capability inside software-defined vehicular networking infrastructure. Such edge networks consist of open-loop grant-free communications and computing-based control frameworks, which enable dynamic eco-routing with ultra-low latency and mobile data-driven orchestration. Thus, this work advances the frontiers of machine learning potential and next-generation mobile systems in vehicular networking applications.

Introduction

Autonomous driving has attracted considerable community interest, which drives the demand for intelligent vehicular systems and connected transportation to ensure ultra-low end-to-end latency. This latency is defined as "the time interval starting from when the vehicle/user sends the connection request to the desired information is received on its side." Such connected and autonomous vehicle (CAV) killer applications rely on fast mobile networking to keep communications latency down to 1 ms for large-scale reliable deployment [1]. However, conventional communications protocol-based vehicular networking can only support the latency in 20–100 ms. First, data transmissions rely on initial control signaling storms, which employ closed-loop physical-layer (PHY) protocols, like three-way handshaking, with thousands of control messages for power control or channel estimation in a single PHY link. Second, cumbersome and reactive protocol stacks and higher-layer optimization (i.e., routing and scheduling in networking) take a significant amount of time in computation and information gathering. Third, the high mobility of CAVs incurs tremendous overhead and complexity of protocols, which in turn causes considerable latency for reliable communications. Fourth, cloud computation using remote data centers for heavyweight analysis can address the rapid growth of mobile applications and their diverse quality of service (QoS) requirements, but at the price of seconds or even longer in end-to-end latency. Finally, cloud-based models require vehicular data to be transmitted to and stored in data centers, increasing chances of information leakage and longer than expected latency, resulting in fatal accidents.

As one 5G and beyond envisioned services, ultra-reliable low-latency communications (URLLC) aims to provide secure data transmissions from one end to another with ultra-high reliability and deadline-based low latency [2]. 3rd Generation Partnership Project (3GPP) standards for URLLC require 1 ms hard latency over the air interface and 99.999 percent system reliability to meet the needs of autonomous vehicles for performing cooperation and safety functions. In August 2018, normative works of New Radio vehicle-to-everything (V2X) and enhanced URLLC were launched in 3GPP Release 16 [3]. In December 2019, the 3GPP RAN Plenary meeting approved 24 new projects for 3GPP Release 17 with one primary focus: bringing sidelink capabilities from automotive to smartphones and public safety. Meanwhile, IEEE 802.11p, the basis for dedicated short-range radio communications (DSRC), established wireless access in vehicular environments (WAVE) [4] and supports communications among vehicles to enable intelligent transportation services. This technology dedicated to connected vehicles helps SAE Level 3 (conditional automation) trail and soon deployment and, targets Level 4/5 (high/full automation) development for self-driving vehicles. However, technological efforts to support mission-critical URLLC requirements are still in an early research stage with no concrete solutions.

Edge Computing and Software-Defined Paradigms

Edge computing provides an alternative to eliminate the reliance on remote data centers for mobile applications. By moving the computing resources close to data sources, edge computation enables near-real-time and low-latency big data processing [5]. This paradigm removes the network bandwidth constraint of sending all the data back and forth to the cloud. Despite its innovative ideas, edge computing in vehicular networking is a challenging and daunting task. First, the underlying vehicular infrastructure consists of highly mobile and unreliable wireless links, hindering network connectivity. Second, efficiently

Digital Object Identifier:
10.1109/MCOM.004.2001124

Shih-Chun Lin is with North Carolina State University; Kwang-Cheng Chen is with the University of South Florida; Ali Karimoddini is with North Carolina A&T State University.

deploying edge compute nodes is difficult as system architects can overestimate distributed processing capability, resulting in unexpected failures. Finally, the overheads of scaling applications into distributed edge nodes, coordinating cross-node state and storage, and delicately handling inconsistent states can also eradicate the benefit of edge computing.

Recently, software-defined networking (SDN) has emerged to facilitate system design and management flexibility, potentially solving the challenges in vehicular edge computing. The main SDN ideas are to separate the data forwarding plane from the network control plane, and to introduce novel network control functionalities based on a network abstraction. Hence, SDN provides the upper-running applications with a centralized view of distributed network states, providing optimal control and system parameters compared to non-softwarized systems. An overview of SDN-enabled Internet of Vehicles is presented in [6], which leverages SDN technologies in communication, computing, and caching to enhance vehicular system performance. However, current SDN realizations mainly emphasize programmable packet forwarding via flow tables and wildcard rules. The literature remains limited regarding technology innovation and practical deployment of distributed computation and learning with the underlying infrastructure.

SDVEC: Software-Defined Vehicular Edge Computing

We introduce a novel distributed computing architecture, called software-defined vehicular edge computing (SDVEC), that is compatible with 5G O-RAN [7] and 3GPP TSG SA selected 5G service-based architectures and employs edge computation for ultra-low-latency vehicular services. SDVEC integrates "intelligence" (i.e., computation capability) seamlessly into anchor nodes (ANs), which connect roadside access points (APs) via fronthaul links with standardized interfaces. The ANs can explore several computation functions at networked vehicular edges while establishing minimum one-time signaling via open-loop transmissions and proactive decisions between vehicles and the infrastructure (i.e., APs). SDVEC extends the network abstraction and provides an interface to allow ANs to control data delivery and processing in harmony with 3GPP Release 16 and the International Telecommunication Union's (ITU's) control plane [3, 8]. Moreover, the network operator can manage SDVEC as a traditional telco or, by industry verticals, solely to act as a private network (e.g., the smart factory in Industry 4.0) due to SDVEC's openness and great flexibility.

As a compact total solution, SDVEC replaces conventional protocol-based vehicular networking with computing-based architectures that support nearby data processing and has the following main contributions. First, SDVEC offsets communications workloads to computation tasks, reducing networking latency while preserving reliability. Second, ANs can respond to vehicles earlier via distributed processing without going through the remote cloud, mitigating total traffic going through the network. Third, with edge computation and machine learning techniques, proactive access/association and anticipatory management can completely resolve high vehicle mobility and subsequent impacts. Finally, ANs can filter sensitive vehicular data, making SDVEC less fragile to security issues with enhanced data privacy. In addition, as SDVEC offloads computation to the vehicular infrastructure, it facilitates on-device computing in autonomous vehicles. More importantly, SDVEC can enable new developments of large-scale distributed machine learning techniques and their applications via its scalability with ultra-low latency and open architecture for computing and networking resources.

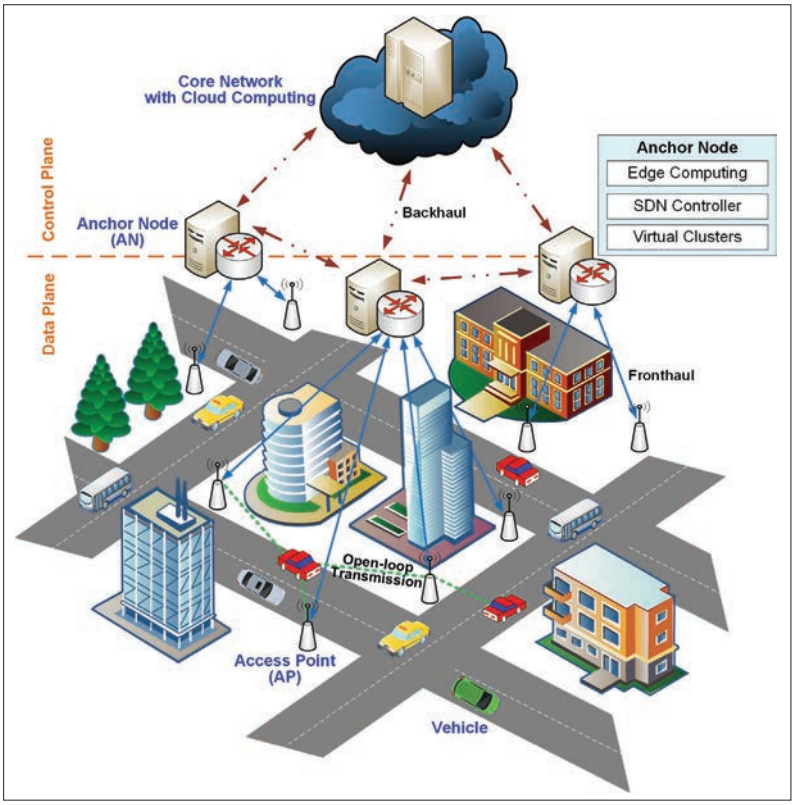

FIGURE 1. The system infrastructure of SDVEC.

The rest of the article is organized as follows. The next section gives SDVEC's software-defined computing infrastructure. Following that, we present ultra-low-latency networking and big-data-enabled orchestration, and then provide SDVEC validation. The final section concludes the article.

Software-Defined Computing Infrastructure

The SDVEC in Fig. 1 systematically integrates SDN and edge computation for delay-sensitive vehicular applications. The data plane is an open, programmable, and virtualizable forwarding infrastructure for endpoint traffic via last-mile open-loop transmissions. As ANs govern APs' functionalities, SDVEC can use distributed transmit points to offer cooperative gain by aggregating massive technology-evolving APs. The control plane, handled by ANs and connecting to the cloud server via backhaul, has two components: network management tools, such as SDN controllers and virtual cluster orchestration for traffic monitoring and real-time management of vehicular operations, and customized applications from service providers, like security and privacy services. This SDN-enhanced computing infrastructure can also realize a suite of innovative computation functions.

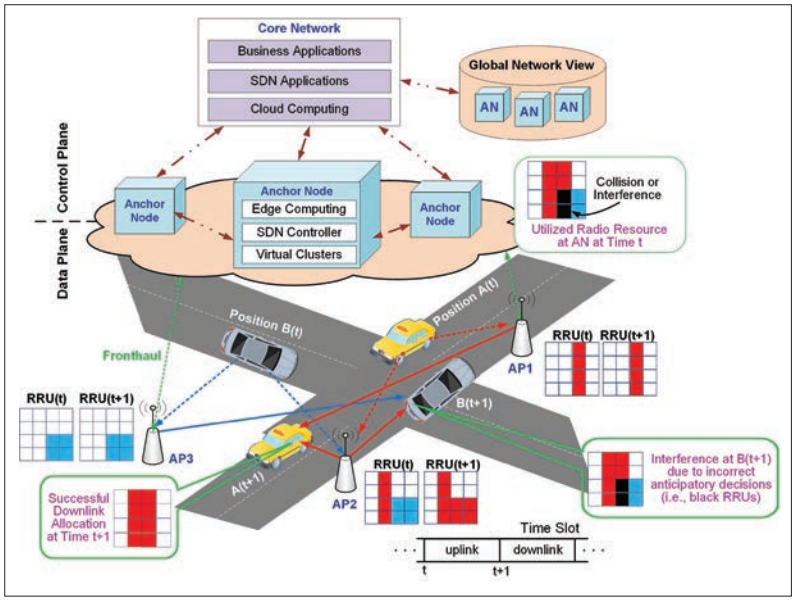

FIGURE 2. The proposed proactive uplink and downlink access.

Wireless Grant-Free Data Plane

Serving as the foundation of SDVEC infrastructure, we developed an ultra-fast vehicular data plane with high reliability. It exploits latency-minimum open-loop transmissions, extends multiuser detection to support wireless grant-free access, provides path-time error control schemes to ensure reliability, and supports dynamic resource slicing among virtual clusters. Our primary goal is to employ grant-free communications between vehicles and APs, which will bring at least 70 percent latency saving due to no re-transmission, being connectionless, and with no schedule needed. The ANs then apply edge computing with mobility management and proactive resource allocation [9] to recover collided open-loop transmissions. In particular, a single vehicle is treated as the center of a virtual AP cluster so that multiple APs can serve this vehicle via cooperative communications. This cell-free (or virtual-cell) concept is different from conventional protocol-based architectures where a cell/cluster uses one base station to serve multiple vehicles. We list the functions of proactive open-loop resource allocation below.

Non-Orthogonal Multiple Access (NOMA) and Fast Beamforming: Grant-free NOMA [10] is adopted for PHY transmissions, consistent with 5G New Radio [8]. A contention transmission unit (CTU), the fundamental multiple access resource, consists of radio resources (e.g., time-frequency blocks), a reference signal (e.g., a preamble), a spreading sequence (e.g., codebook/codeword), and so on. Vehicles perform such contention-based access by randomly selecting CTUs or following pre-configurations. For multi-antenna scenarios, we adopt unsupervised learning [11] to provide real-time beamforming. Our design avoids compute-intensive labeling effort in training, and the overall computational complexity is only 3 percent of conventional WMMSE solutions.

Proactive Resource Allocation: Vehicles proactively allocate radio resources for their uplink transmissions in a distributed manner. For example, each vehicle can select its serving APs by piggybacking signal qualities and randomly allocating CTUs to each associated AP. A more sophisticated design is to apply iterative learning between uplink and downlink access. Specifically, initial uplink resource allocation can be determined via random decisions. Based on uplink results, downlink access is then achieved via machine learning techniques below. Next-round uplink access can further learn from previous downlink allocations.

Multipath Cooperative Communications: Each vehicle packs its data into the selected CTUs and transmits them to proactively associated APs. The size of associated APs indicates the number of upstream cooperative paths. Each AP behaves as a relay node in collaborative multipath networking to forward received signals to an AN. Amplify-and-forward can be selected as a cooperative communications strategy, which allows the AN to decode the vehicular data using selection-combining with proactively set CTUs. Decode-and-forward can be an alternative by which each AP demodulates and decodes before relaying the signals to the AN.

Path-Time Error Control Mechanisms: Since there is no feedback with proactive open-loop communications to shorten the networking latency, multiple networking paths will be employed to ensure reliability and error rate performance. Path-time codes will be developed as the error control coding over open-loop multipath transmissions. Forward error correcting codes can still be applied to open-loop PHY communications. According to our preliminary study, low-density parity check (LDPC) codes serve this purpose well with small-length packets, that is, 128- or 256-bit packet payload.

As proactive resource allocation proceeds in a distributed manner, the multi-armed bandit methodology can adaptively strike the trade-off between reliability and throughput.

Machine-Learning-Enhanced Network Architecture

Downlink resource allocation to support multiple vehicles via different APs can be realized in a centralized manner. However, unlike traditional centralized management, open-loop proactive downlink access relies on uplink access and precise anticipatory mobility management. In particular, proactive uplink access inherits probable collisions due to vehicles' distributed decisions and cannot provide valuable transmissions without further sophisticated designs. Figure 2 provides an example of proactive uplink and downlink access in vehicular networking, where an amplify-and-forward scheme is adopted, and multiple access interference (MAI) happens in the overlapped CTUs (i.e., black blocks) for downlink access. To address this issue, we utilize distributed-learning-enhanced operations in our software-defined computing architecture. As shown in Fig. 3, this architecture refers to 5G URLLC following 3GPP-based specifications and system parameters [8]. Bandwidth reservation and downlink are migrated from IEEE 802.11ax in the air interface, extending DSRC for next-generation vehicular services [9]. We exploit SDVEC's multi-scale computing environment (i.e., on-device, edge, and cloud computations) and design the necessary functions below to combat time-varying MAI for desired last-mile transmissions.

Multiuser Detection for Downlink Receivers: Considering asynchronous proactive communications, we propose two-stage inter-carrier interfer-

ence suppression [9] to mitigate downlink MAI from co-locating virtual cells. While the techniques might incur a high computational complexity, we can develop computation-efficient schemes, given that few APs will possibly be involved for each vehicle due to moderate uplink ranges.

Anticipatory Mobility Management and Intelligent Resource Allocation: Although correct downlink prediction cannot always be assured (e.g., as in Fig. 2), precise anticipatory management is critical for open-loop proactive communications, which determines potential APs in the subsequent downlink access. We introduce a robust data-driven scheme using historical data as shown later. Furthermore, through SDN construction and its centralized control capability, intelligent resource allocation can also be applied optimally assigning downlink associations (i.e., network slices) and CTUs (i.e., radio slices). This smart allocation with details below can significantly improve the throughput and reliability in the proactive downlink.

Based on our designed architecture, SDN controllers at ANs can execute desirable online policies to provide maximum system capacity for proactive multiple access.

Scalable Software-Defined Control Plane

We describe SDVEC's management tools below, which establish a unified control plane among distributed ANs for complete controllability with horizontal controllers and the vertical cloud server in core networks.

Optimal Multi-Controller Placement: The optimal placement problem aims to minimize the required SDN controllers (to reduce infrastructure cost for efficient placement) and, at the same time, support all control traffic from vehicles (to ensure a feasible solution for scalable placement). With vehicle mobility models and ANs' computing capability, we can solve the problem and determine the number of required controllers, their geo-locations, and vehicles' control domain assignments.

Mobility-Aware Control Traffic Regulation: We develop novel control traffic balancing to promise online and adaptive traffic engineering. Dynamic traffic behaviors are modeled, and a nonlinear problem is formulated to find the optimal forwarding paths for vehicles while minimizing the average control traffic latency. The obtained latency values are further fed back to the placement decision, triggering adaptive control for better SDN controller placement and timely control message delivery.

Distributed Controller Consensus: We introduce a timing synchronization protocol that outperforms voting and diffusion protocols and provides linear-time convergence. By adopting a geometric random graph for information exchanges, we derive the closed-form synchronization rate via the graph's Laplacian matrix. The scalability is guaranteed as our protocol preserves the same linear performance regardless of link technologies between machines, machine densities, and machine types. Hence, given the controller's topology, we can establish inter-controller synchronization for the controllers' network and computing-wide consensus, which keeps the network control logic centralized and localizes all decisions to each controller, minimizing control plane response time.

This management framework provides excellent flexibility to control SDVEC's underlying infrastructure components via south-bound APIs for the requested URLLC.

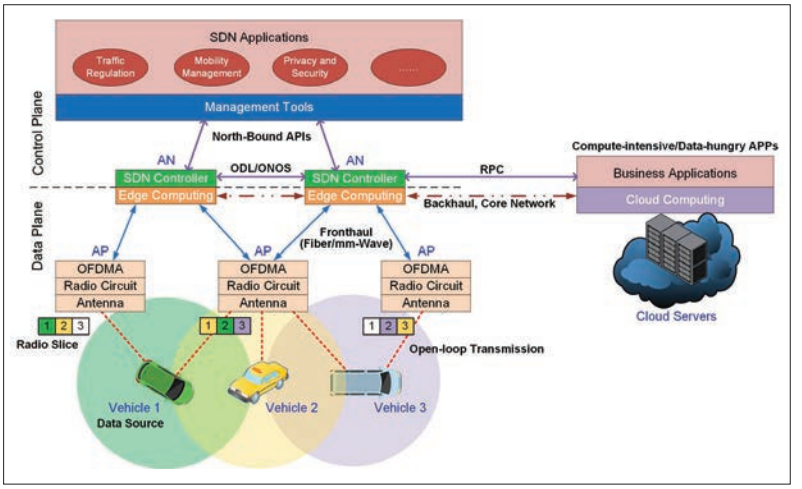

FIGURE 3. Distributed computing architecture with a scalable control plane.

Ultra-Low-Latency Connected Networking

Emerging and growing vehicular applications require highly differentiated networking and computing capabilities to be integrated and deployed over the same vehicular infrastructure. To support infrastructure-as-a-service for applications and latency-minimum associations between vehicles and the infrastructure, we develop virtual cluster orchestration and vehicular network virtualization. In each virtual cluster/network, a vehicular application is provided with the ability to control, optimize, and customize the underlying resources without interfering with other services' performance, leading to cost-efficient operations and enhanced QoS.

Latency-Optimal Vehicle-Centric Clustering

As shown in Fig. 4, by forming vehicle-centric virtual clusters in the SDVEC, each vehicle is served by a selected subset of neighboring APs via open-loop transmissions and coordinated multipoint techniques. This distributed transmit point structure architecture via fronthaul links can significantly boost the coordination gain of APs, eliminate cluster edges from severe inter-cluster interference, and in turn reduce vehicular networking latency. While vehicle-centric clustering is challenging since the clusters are dynamically chosen and may overlap, this clustering method is preferred when dealing with mobile objects.

We implement virtual cluster orchestration via distributed reinforcement learning techniques. The designed orchestration jointly optimizes power allocation and vehicle association, where each vehicle is associated with a dedicated AP cluster for latency-minimum networking. Results show that our design can yield near-optimal performance with high reliability by a nominal number of training episodes compared to the baseline. This preliminary work establishes the feasibility of learning-enabled cluster optimization via edge computation to obtain remarkable efficiency and rapid network operation.

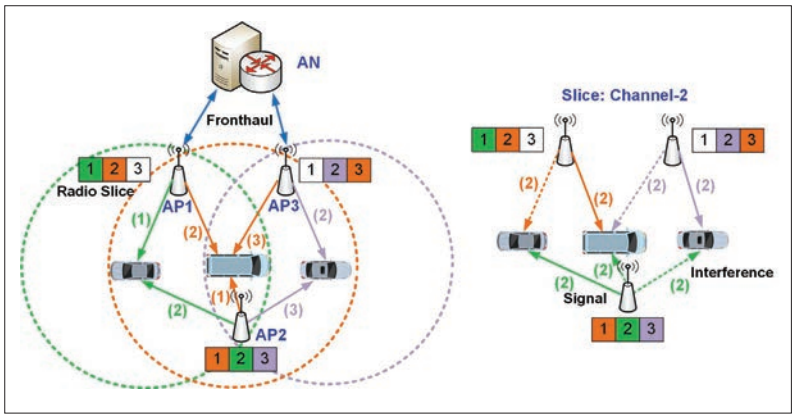

FIGURE 4. The designed virtual clusters with radio resource slices.

Eco-Vehicular Edge Network Slicing

To achieve wireless virtualization in highly mobile environments, a comprehensive set of resources (e.g., antenna elements, APs, AP-to-AN connections, wireless spectrum, transmission power, and ANs' computing capabilities) and the slicing mechanism should be investigated concerning time-varying factors (e.g., vehicle mobility, channel fading). The network slicing technology divides multi-dimensional wireless and computing resources into non-conflicting virtual networks, while the requests of each virtual network are satisfied with the minimum operational expense. Accordingly, we develop eco-vehicular edge networks for connected transportation [12], which use distributed multi-agent reinforcement learning to combat power-hungry edges while ensuring system reliability and data rates. This "dynamic eco-routing" equips ANs with respective edge Q-learners to collaboratively yield best-reward decisions for optimal downlinks. Evaluating by 3GPP C-V2X services [3], our design provides satisfying energy efficiency and system reliability and coverage. To further boost the performance in larger-scale networks and complex transportation environments, we will extend it with deep neural networks and resolve training issues and algorithm complexity.

Big-Vehicular-Data-Enabled Orchestration

Open-loop communications and virtual AP clustering enable SDVEC's core technologies to accomplish ultra-low-latency networking with high reliability. We extend intelligent edge computing at ANs with big data and distributed data processing. Big vehicular data provides unprecedented opportunities for system architects (or SDN controllers) to understand the requirements and behaviors of vehicle mobility and diverse network elements. With both instantaneous and historical data, data-driven optimization allows intelligent real-time decision making in a wide range of services. Useful features (e.g., the correlation between vehicular events and traffic data) can be extracted to make optimal decisions based on long-term strategies. Data-centric mechanisms can consistently achieve minimum latency by leveraging computation capabilities and proactive reinforcement decisions.

Joint Data Caching and Computation Offloading

AN's edge computation can push the frontier of computing applications, data, and services away from centralized cloud infrastructure to local vehicular edges, enabling analytics and knowledge generation to the proximity of data sources. As shown in Fig 5, ANs, endowed with edge servers for computing and storage capacity, can serve vehicular data as a substitute for the cloud. Extra tasks exceeding ANs' computing capacity are offloaded to the cloud, resulting in a hierarchical offloading structure between ANs and the cloud.

We establish new application-defined networks, enabling edge resource orchestration and adaptivity in terms of desired QoS. The proposed dynamic resource orchestration can serve offline training and online prediction distributed among vehicular edges. It enables data and computation parallelism by combining few-shot classification (to reduce trainable weights and keep good performance without model overfitting) and transfer learning (to help the training among edge nodes and possibly with the cloud server). Our design exploits edge infrastructure's real-time and global control and fulfills the computation tasks for distributed learning techniques.

Accordingly, we investigate two edge-cloud operations: caching and offloading for contents and services. First, caching functions refer to caching services and related databases/libraries into edge ANs, enabling local vehicular data processing. Due to ANs' limited computing and storage resources, services cached on ANs can determine tasks to be offloaded to the cloud, significantly affecting the edge computing performance. Second, computation offloading concerns what/when/how to offload vehicular workload from local ANs to the cloud regarding ANs' service availability (i.e., what types of computation tasks/applications). The optimal offloading decisions are complex and coupled both spatially and temporally because of high mobility in vehicular networks. Hence, we will design an efficient solution for jointly optimizing service caching and task offloading for vehicular data and applications. Upon open-loop transmissions and virtual clusters, the objective is to minimize the computation latency under a long-term energy consumption constraint.

Data-Driven Anticipatory Mobility Management

Anticipatory mobility management predicts the proactive network association in the future through machine learning techniques. Rather than using the vehicle's GPS information (which consumes massive communications bandwidth and jeopardizes privacy and security), data analytics at ANs can use the history of AP association vectors to predict the AP association at the next time slot. This data-driven framework avoids losing prediction accuracy due to inaccurate GPS data and probing complicated channel states.

Our recent study provides data-driven mobility management via three designs, including online association learning, a mobility-pattern-assisted mechanism, and transferring learning for reinforcing mobility management. First, with a Markovian model, the posterior belief of vehicle locations can be inferred from historical association vectors and vehicle velocity estimation. Accordingly, asso-

ciation vectors can then be obtained via recursive Bayesian estimation. Second, fixed road structures limit the moving patterns of vehicles with shared similarities. Data-driven frameworks can exploit this mobility pattern to improve prediction accuracy. By modeling vehicle mobility with a Markov jump process, we can mainly acquire a more accurate prediction of current vehicle locations based on the past trajectory, conditioned on the corresponding series of geographic road maps. Finally, in addition to applying empirical probability for predicting mobility patterns, deep learning techniques with big vehicular data can facilitate more robust mobility management. For example, we can use random forest techniques to reinforce anticipatory mobility management. The procedures are: selecting data information (i.e., decision trees), potentially combining data decision trees, and extracting useful features via deep learning.

Virtualization-Enabled Security and Privacy

To uphold the ultra-low-latency design, we establish latency-minimum one-time cryptography functions. SDVEC's infrastructure-as-a-service can enhance system security and privacy. In particular, SDVEC can decouple security functions (e.g., firewalling) from proprietary hardware appliances to run software services, and deliver security enforcement through the separation of the secure control plane and the secure processing and forwarding plane. Specifically, we strengthen the conventional symmetric stream cipher (e.g., WiFi Protected Access from the IEEE 802.11 family) with a lightweight dynamic cipher in Fig. 6. The proposed cypher utilizes the fingerprint concept to create a unique and dynamic keystream. A vehicle's fingerprint is a nonlinear time-dynamic function and can be shared only between the car and edge ANs. A simple fingerprint is a vehicle's historical connections to APs, which is unique to each vehicle and aware by ANs. This dynamic cipher can also provide new keys indecipherable to evicted users. The generated ciphertext is thus suitable for grant-free access, and our solution can offer a secure and low-latency data plane in the SDVEC. Furthermore, the secure control plane can be built through SDN's open architecture. Various security applications can be deployed and run on ANs' SDN controllers. The latest micro-segmentation or sub-slice techniques can be further applied via SDVEC's edge slicing to secure the entire infrastructure and management.

Additionally, this fingerprint-based cipher can enable the following security services.

Authentication: Without requiring real-time authentication, our dynamic cipher can seamlessly integrate with the latest authentication schemes. After the authentication, ANs will send a starting key to start the cipher.

Fingerprint Generation and Verification: While the attempt to associate with APs might be unsuccessful due to a failure/inaccuracy caused by MAI, ANs can still perceive possible connected APs for each vehicle at any time via the scalable SDN control plane. Hence, we develop an integrity checking algorithm to verify the key and be used for downlink transmissions. ANs will maintain a master key for any loss of the vehicle's key.

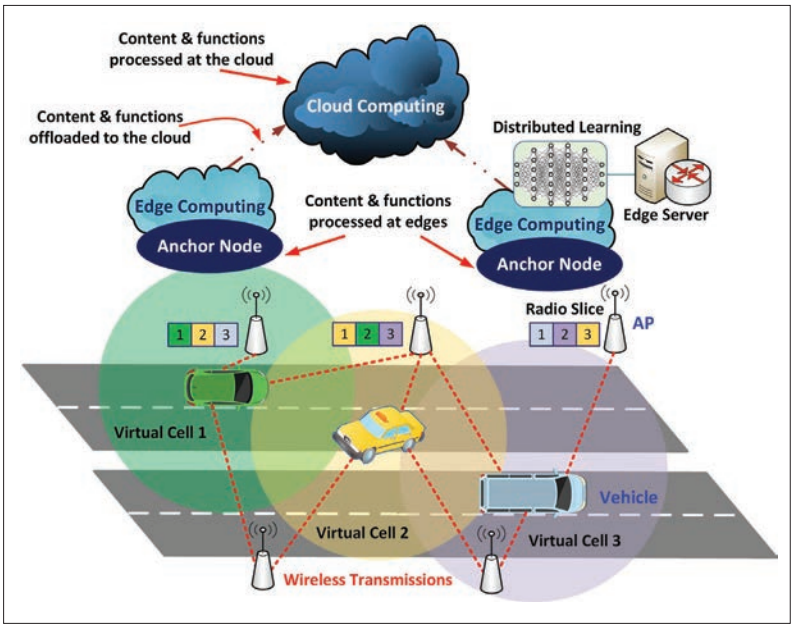

FIGURE 5. ANs can cache and offload both data and services between vehicular edge computation and cloud computation.

Edge Network Dynamic and Fault Diagnosis: Several of our designs (dynamic resource slicing etc.) can bring other degrees of dynamic randomness. By integrating the association information with additional parameters, we can develop new advanced fingerprints. Also, the SDVEC could easily detect whether any fingerprint is being compromised to forge false messages. Such event detection can be fulfilled by our resilient fault diagnosis [13], robust to network, edge, or vehicle operation failures.

For privacy/confidentiality, ANs in the SDVEC can filter or pre-process sensitive vehicular data to avoid information leakage or loss in backhaul and cloud servers. Also, with edge-cloud orchestration, each edge's local training data and neural network structure will not be revealed to other edges or the cloud. Furthermore, our anticipatory mobility management at ANs does not rely on private GPS data of vehicles but employs edge computing to predict plausible APs for vehicles' proactive access. Edge ANs only need to know the past vehicle traces (i.e., virtual clustering patterns that show the past vehicle-AP association), eliminating the privacy concern of using private location information. Moreover, federated multi-task learning can be established on vehicles with their data, and the local edge AN can keep the average learning parameter for a group of cars. Therefore, the decentralized vehicular data has never been passed through the infrastructure, ensuring the utmost privacy level in vehicular environments.

SDVEC Validation and Field Deployment

We set up a full-stack simulation platform that assembles all the proposed designs, and can show each design's results and the end-to-end performance. We also create SDVEC's digital twins with big vehicular data [14] as an infrastructure element that considers connected infrastructure, vehicles, and physical transportation environments. Following O-RAN specifications [7], we

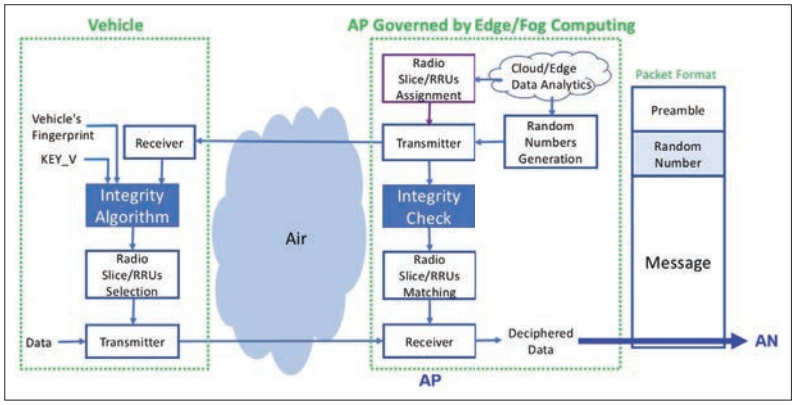

FIGURE 6. Fingerprint-based dynamic cipher for uplink/downlink transmissions.

implement artificial intelligence (AI)-based RAN intelligent controllers dedicated to the SDVEC. We then integrate the O-Cloud (an O-RAN virtualization platform for RAN functions) with simulation of urban mobility (SUMO) [15].

In addition, we built an in-lab experimental testbed for SDVEC architecture. It is a fully functional prototype for an SDN-based edge network and computing platform. This SDVEC testbed consists of: Odroid-C2 and Raspberry Pi 3 Model B for mobile data sources with computation workloads, TL-WDR4300 for programmable APs, Gigabit Ethernet for fronthaul links between APs and ANs, and Intel NUC7i5BNK for high-performance integrated storage servers at ANs and a central cloud. We validate our open-loop communications, grant-free proactive access, edge computation capability with Ryu controllers, and optimal real-time reconfiguration of networking and computing functions. With 128 MB data chunks and the Hadoop application, the testbed can provide a 53 percent reduction in average read and write time. We are expanding this platform with latency-sensitive mission-critical services and for larger-scale validation.

We also deploy SDVEC for intelligent and connected vehicle infrastructure in Greensboro, North Carolina, and the nation. As core investigators in the NC-CAV center funded by the NCDOT in February 2020, our team collaborates with Verizon Wireless for its 5G access. We assess North Carolina's readiness for CAVs in traditional and emerging infrastructure needs and are implementing SDVEC to realize high speed and extensive network coverage. The entire field deployment and testing in North Carolina will be completed in the second quarter of 2022.

Conclusion

Ultra-low-latency connected transportation is envisioned as a critical 6G vertical to reshape autonomous driving and human society in five to ten years. Significant development efforts in vehicular edge computing and wireless infrastructures facilitate the practical deployment of massive CAV technologies. This article presents novel distributed computing architectures, and highlights solutions for system infrastructure, management and orchestration, and data-driven machine learning. It thus realizes a new frontier for vehicular edge networks.

Acknowledgments

This work was supported in part by the North Carolina Department of Transportation (NCDOT) under Award TCE2020-03 and the AC21 Special Project Fund. K.-C. Chen appreciates Cyber Florida for supporting part of this research.

References

[1] U.S. DOT, "Connected Vehicle Pilot Deployment Program"; https: //www.its.dot.gov/pilots/, accessed Oct. 4, 2021.
[2] M. Giordani *et al.*, "Toward 6G Networks: Use Cases and Technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, Mar. 2020, pp. 55–61.
[3] 3GPP TR38.886 v0.5.0, "V2X Services Based on NR; User Equipment (UE) Radio Transmission and Reception," Feb. 2020.
[4] IEEE Std 802.11p-2010, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments," June 2010.
[5] A. Ndikumana *et al.*, "Joint Communication, Computation, Caching, and Control in Big Data Multi-Access Edge Computing," *IEEE Trans. Mobile Comp.*, vol. 19, no. 6, Mar. 2019, pp. 1359–74.
[6] W. Zhuang *et al.*, "SDN/NFV-Empowered Future IoV with Enhanced Communication, Computing, and Caching," *Proc. IEEE*, vol. 108, no. 2, Nov. 2019, pp. 274–91.
[7] O-RAN Alliance, "O-RAN: Operator Defined Open and Intelligent Radio Access Networks"; https://www.o-ran.org/, accessed Oct. 4, 2021, .
[8] S.-Y. Lien *et al.*, "3GPP NR Sidelink Transmissions Toward 5G V2X," *IEEE Access*, vol. 8, Feb. 2020, pp. 35,368–82.
[9] C.-H. Zeng, K.-C. Chen, and D.-Z. Liu, "Two-Stage ICI Suppression in the Downlink of Asynchronous URLLC," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, Jan. 2020, pp. 2785–99.
[10] M. B. Shahab *et al.*, "Grant-Free Non-Orthogonal Multiple Access for IoT: A Survey," *IEEE Commun. Surveys & Tutorials*, vol. 22, no. 3, May 2020, pp. 1805–38.
[11] C.-H. Lin *et al.*, "Unsupervised ResNet-Inspired Beamforming Design Using Deep Unfolding Technique," *IEEE GLOBECOM*, Dec. 2020.
[12] M. F. Pervej and S.-C. Lin, "Eco-Vehicular Edge Networks for Connected Transportation: A Distributed Multi-Agent Reinforcement Learning Approach," *IEEE VTC-Fall*, Oct. 2020.
[13] A. White, A. Karimoddini, and M. Karimadini, "Resilient Fault Diagnosis under Imperfect Observations — A Need for Industry 4.0 Era," *IEEE/CAA J. Autom. Sin*, vol. 7, no. 5, Aug. 2020, pp. 1279–88.
[14] M. Farsi *et al.*, *Digital Twin Technologies and Smart Cities*, Springer, 2020.
[15] C.-H. Lin *et al.*, "A C-V2X Platform Using Transportation Data and Spectrum Aware Slidelink Access," *Proc. IEEE Int'l. Conf. Sys. Man Cybern. Wksp.*, Oct. 2021.

Biographies

Shih-Chun Lin [M'17] (slin23@ncsu.edu) received his Ph.D. degree from the Georgia Institute of Technology in 2017. He is currently an assistant professor with the Department of Electrical and Computer Engineering, North Carolina State University, leading the Intelligent Wireless Networking Laboratory. His research interests include beyond 5G, software-defined infrastructure, machine learning techniques, mathematical optimization, and performance evaluation.

Kwang-Cheng Chen [F'07] (kwangcheng@usf.edu) is a professor of electrical engineering at the University of South Florida. He has widely served in IEEE conference organization and journal editorship. He has contributed essential technology to IEEE 802, Bluetooth, LTE and LTE-A, and 5G-NR wireless standards. He has received several IEEE awards. His research interests include wireless networks, artificial intelligence and machine learning, IoT and CPS, social networks, and cybersecurity.

Ali Karimoddini (akarimod@ncat.edu) received his Ph.D. degree from the National University of Singapore in 2012 and then joined the University of Notre Dame to conduct his postdoctoral studies. He is currently an associate professor with the Department of Electrical and Computer Engineering, North Carolina A&T State University. He is also the director of the NC-CAV Center of Excellence on Advanced Transportation and the ACCESS Laboratory, and the deputy director of the TECHLAV DoD Center of Excellence. His research interests include cyber-physical systems, control and robotics, resilient control systems, flight control systems, multi-agent systems, and human-machine interactions.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 7, 2021, accepted November 5, 2021, date of publication November 12, 2021,
date of current version November 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3127914

GCN-CNVPS: Novel Method for Cooperative Neighboring Vehicle Positioning System Based on Graph Convolution Network

CHIA-HUNG LIN[1], YO-HUI FANG[2], HSIN-YUAN CHANG[2], YU-CHIEN LIN[3,4], (Member, IEEE), WEI-HO CHUNG[2,5], SHIH-CHUN LIN[1], (Member, IEEE), AND TA-SUNG LEE[3,4], (Fellow, IEEE)

[1]Intelligent Wireless Networking Laboratory, Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695, USA
[2]Department of Electrical Engineering, National Tsing Hua University, Hsinchu 300, Taiwan
[3]Institute of Communications Engineering, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan
[4]Center for mmWave Smart Radar Systems and Technologies, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan
[5]Research Center for Information Technology Innovation, Academia Sinica, Taipei 115, Taiwan

Corresponding author: Wei-Ho Chung (whchung@ee.nthu.edu.tw)

This work was supported in part by the "Center for mmWave Smart Radar Systems and Technologies" and the "Center for Open Intelligent Connectivity" under the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) of Taiwan; in part by the Visible Project at the Research Center for Information Technology Innovation, Academia Sinica; in part by the Ministry of Science and Technology (MOST) of Taiwan under Grant MOST 110-2221-E-007-042-MY3, Grant MOST 110-2221-E-A49-025-MY2, Grant MOST 110-2634-F-009-028, Grant MOST 110-2224-E-A49-001, and Grant MOST 110-2622-E-A49-004; in part by the North Carolina Department of Transportation (NCDOT) under Award TCE2020-03; in part by the AC21 Special Project Fund.

ABSTRACT To provide coordinate information for the use of intelligent transportation systems (ITSs) and autonomous vehicles (AVs), the global positioning system (GPS) is commonly used in vehicle localization as a cheap and easily accessible solution for global positioning. However, several factors contribute to GPS errors, decreasing the safety and precision of AV and ITS applications, respectively. Extensive research has been conducted to address this problem. More specifically, several optimization-based cooperative vehicle localization algorithms have been developed to improve the localization results by exchanging information with neighboring vehicles to acquire additional information. Nevertheless, existing optimization-based algorithms still suffer from an unacceptable performance and poor scalability. In this study, we investigated the development of deep learning (DL) based cooperative vehicle localization algorithms to provide GPS refinement solutions with low complexity, high performance, and flexibility. Specifically, we propose three DL models to address the problem of interest by emphasizing the temporal and spatial correlations of the extra given information. The simulation results confirm that the developed algorithms outperform existing optimization-based algorithms in terms of refined error statistics. Moreover, a comparison of the three proposed algorithms also demonstrates that the proposed graph convolution network-based cooperative vehicle localization algorithm can effectively utilize temporal and spatial correlations in the extra information, leading to a better performance and lower training overhead.

INDEX TERMS Cooperative vehicle localization, data fusion, deep neural network (DNN), graph convolution network (GCN), long short-term memory (LSTM), vehicle-to-vehicle (V2V).

I. INTRODUCTION

Intelligent transportation systems (ITS) and autonomous vehicles (AV) enable several convenient applications and are expected to bring about new experiences with enhanced efficiency and safety to road users in the near future [1]. More specifically, ITS applications such as geographic information dissemination, traffic control, and the automatic positioning of accidents can be used to increase the efficiency of road users while traveling. By contrast, automatic collision

The associate editor coordinating the review of this manuscript and approving it for publication was Qi Zhou.

avoidance systems in AVs can be employed to provide safety to road users. However, accurate localization is one of the most vital premises for the implementation of the aforementioned applications [2]. Although several localization methods, such as map matching, fingerprinting, and image/video localization, can be used to provide coordinate information of vehicles, the global positioning system (GPS) is still the most common choice for providing localization results to vehicles. There are two reasons for prohibiting the wide use of these methods. First, to apply these methods, expensive sensors, such as cameras and video recorders, should be installed in the target vehicle to provide the required information for matching or fingerprinting. To achieve accurate localization, acceptable sensors, which can be used to provide a high-resolution or high-quality output, are costly, hindering the widespread use of such methods. Second, to achieve an effective matching or fingerprinting, a database containing sufficient reference samples should be built in advance, prohibiting the wide use of these methods. Moreover, the need for a predefined database also limits the operation area for effective localization. As an alternative, GPS is the most commonly used localization system for vehicle applications, offering a cheap and easily accessible solution for global positioning [3].

Although GPS offers an easy and accessible way to conduct localization, the precision of GPS still has room for further improvement in providing accurate localization. To be more specific, GPS suffers from the influence of several factors (e.g., receiver noise and a multipath effect) such that the received GPS coordinates have large errors with the actual coordinates of the vehicle, thereby posing a threat to the safety of the AV or the precision of ITS applications. To solve this drawback by working on the GPS error refinement, vehicular ad-hoc networks (VANETs) [4] have recently been introduced to the automotive research community where vehicles can communicate with each other to improve their location awareness [2], [5]–[7]. By integrating vehicle-to-vehicle (V2V) communication, an effective ''cooperative driving'' network can be established to share information for GPS refinement usage [1]. To be more specific, several studies have already focused on incorporating GPS with auxiliary information (e.g., ranging measurement and reference points) through optimization-based algorithms to enhance the system performance [2]. In [5], the authors proposed a direction of arrival (DOA)-based cooperative localization method, incorporating GPS with radar to improve the localization. Furthermore, because the information coming from each sensor has its own limitations, the concept of data fusion has been introduced into the GPS refinement to refine the GPS results based on the information acquired from multiple sensors. In addition, in [7], the authors proposed a cooperative neighboring vehicle positioning system (CNVPS), incorporating GPS with various sensors using the weighted average to improve the localization. However, the approach in [7] only employs a linear function for the application of a sensor data fusion for GPS refinement, thereby leaving room for further performance improvements.

Although more powerful algorithms should be developed to improve the performance of existing CNVPS algorithms, the design of advanced optimization-based CNVPS algorithms is not trivial. First, the design is highly dependent on the precise domain knowledge of different information sources (i.e., the error distribution of extra sensors), which may not be easily available under real scenarios. Moreover, the optimization problem should be redesigned if different types of sensors are employed. Second, to achieve accurate results, multiple iterations or complex matrix operations are often employed in the optimization process. Finally, although the problem of interest can be considered time-series data (i.e., multi-time-slot data), existing optimization-based methods only focus on the scenario of a single time-slot data fusion. As a result, the modeling of an optimization problem for extracting the correlation in multiple time-slot data and further improving the performance remains an open problem.

Differing from existing optimization-based algorithms, our idea is to develop CNVPS algorithms based on a deep learning (DL) algorithm owing to such advantages as a low complexity, high performance, and design flexibility. Specifically, given a sufficient training dataset, even without a precise mathematical system model, the DL model can be used to construct a nonlinear function and automatically solve the optimization problem of interest. Moreover, during the online testing stage, only simple matrix operations are executed when generating GPS refinement results, matching the computational limitations of on-board vehicle units. Furthermore, the DL model can extract hidden features (i.e., correlation of a multiple time-slot scenario) to further improve the GPS refinement results, making it almost impossible for optimization-based algorithms to do the same. Based on the aforementioned motivations, we decided to focus on the development of DL-based CNVPS GPS refinement algorithms. As a result, we can not only fully fuse various sensors but also integrate multiple time-slot data by introducing flexible characteristics of DL algorithms. For a single time-slot GPS refinement, we propose a multi-layer perceptron-cooperative neighboring vehicle positioning system (MLP-CNVPS) to achieve DL-based GPS refinement. Considering a multiple time-slot GPS refinement, we propose a long short-term memory-cooperative neighboring vehicle positioning system (LSTM-CNVPS) for obtaining superior results by considering the temporal correlation in multiple time-slot data. Moreover, a graph convolution network-cooperative neighboring vehicle positioning system (GCN-CNVPS) was further developed to better utilize both temporal and spatial correlations and achieve an efficient GPS refinement, leading to an even better performance compared to the aforementioned DL-based CNVPS algorithms. Although some more complex architectures of graph neural networks, such as GraphSAGE [8], [9], can also be used to implement cooperative neighboring vehicle positioning

systems, it is noteworthy that the main idea of this paper is to provide a way for on-board computing units to achieve fast localization enhancement with outstanding performance. As a result, we consider GCN is a more suitable candidate than other graph neural network architectures due to the relatively straightforward and stable graph structure of the considered problem [10] and the real-time demand of vehicle applications.

Simulation results show that the performance of the proposed DL-based localization approach is better than that of existing optimization-based localization algorithms by improving the GPS error mean from $4-20$ m to $2-4$ m. This improvement is comparable to the current V2I localization performance [11], thus verifying the contribution of this study. Moreover, as [12] reported, an error mean of approximately $6-8$ m is sufficient for most vehicular applications using V2V ranging. Furthermore, when it comes to the implementation issues of the proposed approaches, recent studies [9], [13] suggest that the proposed GCN structure can be implemented in on-board computing units to provide real-time GPS enhancement due to the low complexity nature of the proposed GCN structure. Considering both the performance and complexity, our study is a strong candidate to be implemented in V2V applications with the advantages of low-cost hardware, a fast and simple method, and accurate and stable performance.

The remainder of this paper is organized as follows. In Section II, the system model and sensor configurations are as described below. Section III illustrates the proposed method in detail, including the framework of the deep learning-based localization approach and the operations in both the online and offline stages. The simulation results are presented in Section IV, followed by some concluding remarks in Section V.

II. SYSTEM SETUP AND PROBLEM FORMULATION

A. SYSTEM SETUP

As shown in Fig. 1, a cooperative group consisting of N vehicles is considered to enable a CNVPS-aided localization refinement [7]. With the support of vehicle-to-vehicle (V2V) communications, a target vehicle can use extra information from the GPS localization results of neighboring vehicles in the group and sensors in the vehicle to refine the GPS localization result. Specifically, GPS installed in each vehicle is used to estimate the vehicle coordinates. Moreover, commonly used omnidirectional radar is also employed in each vehicle to measure the relative distance and angle of the surrounding vehicles. We further assume that these measurements can be matched to the right surrounding vehicle owing to the matching ability of the sensors and GPS [7]. We also assume that each vehicle in the group can communicate with neighboring vehicles through V2V, which is used as a setting in related studies [2], [5]–[7]. To explain this, the basic safety message (BSM) [14] and the optional part of the BSM, which is supported by both dedicated short-range communication [7], [15] and cellular vehicle-to-everything standards [16], can satisfy our need to exchange information between vehicles in a group. As a whole, in each BSM frame (i.e., time slot), each vehicle can acquire the GPS coordinate results of all surrounding vehicles (from GPSs and V2V exchanges), the relative distance and angle to other vehicles (from on-board omnidirectional radar), and the received signal strength indication (RSSI) of other vehicles (from V2V exchange) for achieving CNVPS-aided localization refinement for a refined GPS localization result.

B. SENSOR CONFIGURATION

In this section, we introduce the sensor configurations used in this study, including GPS, radar, and RSSI.

1) GPS

In this study, GPS provides location information in units of degrees of latitude and longitude. To employ GPS measurements and distance measurements of radars in the developed DL model with the same numerical scale, it is assumed that after the target vehicle receives GPS measurements from the neighboring vehicles, GPS measurements are preprocessed and transformed into coordinates with units of meters. By doing so, the unit scope of input variables is guaranteed to be the same, assisting the problem to be modeled. It is noteworthy that the unit conversion is a one-to-one relationship, a linear calculation can be exploited to acquire GPS refinements of degrees after acquiring a refined localization result in terms of meters. We also use the meter as the unit in the following sections to provide clarity and intuition. Furthermore, it is assumed that the GPS error follows a Gaussian distribution, as there are several independent sources, such as satellite clock bias, atmospheric delay, code acquisition noise, and multipath effect, contributing to the GPS error in practice [17]–[19]. In each case, the relationship between the GPS localization result $\hat{\mathbf{G}} \in \mathbb{R}^2$ and the real position $\mathbf{G} \in \mathbb{R}^2$ of a vehicle can be expressed as follows:

$$\hat{\mathbf{G}} = \mathbf{G} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} = [Re\{|a|e^{j\theta}\}, Im\{|a|e^{j\theta}\}]^T$ denotes the error term with $a \sim N(\mu, \sigma^2)$ and $\theta \sim u(0, 2\pi)$. According to existing literature [20], because the levels of the multipath effect are site-dependent, the statistical properties of the Gaussian distribution should be set differently to reflect the GPS error in different environments. In this study, we simulated three different environments to validate the robustness of different CNVPS-aided refinement algorithms. Specifically, in this study, three environments, including *freespace* (e.g., highway environment), *suburban*, and *urban* scenarios, were considered and set as $N(4.7, 4.84)$, $N(14.8, 49)$, and $N(20.5, 72.25)$, respectively, [2] to reflect the GPS estimation behavior.

2) RADARS

In this study, radar is used to provide the relative distance and angle of the surrounding vehicles as an information source for GPS refinement. It is noteworthy that although the ghost

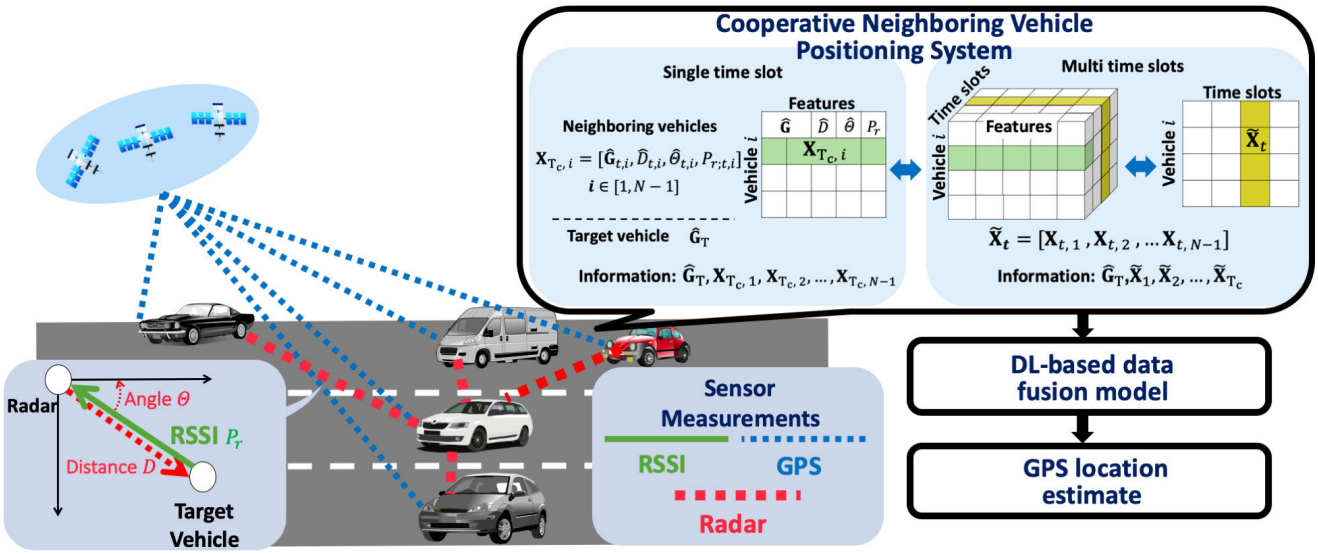

FIGURE 1. The considered scenario of cooperative vehicle localization through V2V. V2V communication is used to transmit additional information acquired from on-board sensors for the use of GPS refinement. The information is arranged in a specific format and fed into the DL-based CNVPS algorithm, generating GPS-refined results.

target effect may appear in the radar sensing stage, we assume that the radar raw data processing step can be finished ideally based on the following reasons. First, the physical characteristics and the detection results from different time-slots can be utilized to aid the ghost target removing. Furthermore, as only short-range radars will be employed in the considered scenario and the number of detecting targets can be regarded as a prioir knowledge according to the pre-defined size of the considered vehicle group, with the aforementioned information, several advanced ghost target removing algorithm [21], [22] can be used to provide correct processed radar data in the considered scenario. As a result, we assume ideally radar raw data is available for the use of proposed algorithms.

To simulate the radar measurements, the relationship between the radar distance measurement $\hat{D} \in \mathbb{R}$ and the real distance $D \in \mathbb{R}$ of a vehicle can be expressed as follows:

$$\hat{D} = D + \check{n}_{\mathrm{D}}, \tag{2}$$

where $\check{n}_{\mathrm{D}} \sim u(-0.025D, 0.025D)$ describes the longitudinal uncertainty of the measurement related to the true distance. Note that the uncertainty increases with the real distance, satisfying the realistic radar measurement behavior. Moreover, the relationship between the angle measurement $\hat{\Theta} \in \mathbb{R}$ and the real angle $\Theta \in \mathbb{R}$ of a vehicle can be expressed as follows:

$$\hat{\Theta} = \Theta + \check{n}_{\Theta}, \tag{3}$$

where $\check{n}_{\Theta} \sim u(-0.5, 0.5)$ describes the lateral uncertainty of the radar measurement with an angular error of 0.5° according to [23].

3) RSSI

Despite the radar measurements, because we adopt BSM to serve V2V communication, the RSSI information can also be acquired in each frame of the BSM to provide another distance measurement for the GPS refinement [24], [25]. To model the RSSI measurements, the practical path loss can be described as a log-normally distributed random variable with a distance-dependent location parameter [12], [26]. That is,

$$P_{\mathrm{r}}(d) = P_0(d_0) - 10n_{\mathrm{p}} \log_{10}(\frac{d}{d_0}) + \tilde{n}, \tag{4}$$

where $P_{\mathrm{r}}(d)$ denotes the received signal strength measured in decibel milliwatts (dBm) at the transmitter-receiver distance d (in meters), $P_0(d_0)$ denotes the reference power (in dBm) at a reference distance d_0 (in meters), n_{p} is the channel path loss exponent, and $\tilde{n}$ is the effect of channel fading. To conduct the simulation in this study, we set the aforementioned parameters as $P_0(d_0) = -34$, $n_{\mathrm{p}} = 2.1$, and $\tilde{n} \sim N(0, 5.5^2)$.

C. PROBLEM STATEMENT

The GPS refinement problem can be considered as a method for improving the GPS localization based on the extra aforementioned information. Hence, the goal of this study is to design a function that takes extra information and the original GPS estimations as input and returns the refined GPS estimation result, minimizing the difference between the predicted result and the real position of the target vehicle by means of data fusion.

Specifically, as shown in Fig. 1, in each vehicle in the group, through a V2V information exchange, the extra information $\mathbf{X_{t,i}}$ is available in each vehicle i at time slot t for the refinement of the GPS localization result. Based on the aforementioned sensor configurations, the extra information can be further expressed as $\mathbf{X}_{t,i} = [\hat{\mathbf{G}_{t,i}}, \hat{D}_{t,i}, \hat{\Theta}_{t,i}, P_{\mathrm{r};t,i}] \in \mathbb{R}^5$, which contains the GPS, relative distance to the target

vehicle, the relative angle to the target vehicle, and the RSSI. Hence, for a target vehicle in the group with GPS measurements $\hat{\mathbf{G}}_{\mathrm{T}}$, the problem we address can be expressed as a function design problem. That is,

$$f^* = arg\min_f \|f(\hat{\mathbf{G}}_{\mathrm{T}}, \mathbf{X}_{\mathrm{T_c},1}, \mathbf{X}_{\mathrm{T_c},2}, \ldots, \mathbf{X}_{\mathrm{T_c},N-1}) - \mathbf{G}_{\mathrm{T}}\|^2, \tag{5}$$

where $\mathbf{X}_{\mathrm{T_c},1}, \mathbf{X}_{\mathrm{T_c},2}, \ldots, \mathbf{X}_{\mathrm{T_c},N-1}$ are the extra information from other $N-1$ vehicles in the group at the current time slot $\mathrm{T_c}$. Note that all algorithms in the previous studies [2], [5], [7], [17], [18] can also be regarded as designing the prediction function through the optimization process. Furthermore, considering that a cooperative group usually exists for several BSM transmission periods (i.e., multiple time slots), for multi-time-slot scenarios, although we can deal with different time slots independently using single time-slot CNVPS solutions, we try to further improve the GPS localization result by using the correlation between multi-time-slot measurements. In light of this, we first consider the GPS refinement problem under the multiple time-slot scenario in this study. In particular, the multiple time-slot problem we considered can be expressed as

$$f^* = arg\min_f \|f(\hat{\mathbf{G}}_{\mathrm{T}}, \widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2, \ldots, \widetilde{\mathbf{X}}_{\mathrm{T_c}}) - \mathbf{G}_{\mathrm{T}}\|^2, \tag{6}$$

where $\widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2, \ldots, \widetilde{\mathbf{X}}_{\mathrm{T_c}}$ are the features of the previous time slots and can be further expressed as $\widetilde{\mathbf{X}}_t = [\mathbf{X}_{t,1}, \mathbf{X}_{t,2}, \ldots, \mathbf{X}_{t,N-1}]$. Note that Eq. (5) can be regarded as a special case of Eq. (6) when only one time-slot information is provided for GPS refinement.

III. DEVELOPMENT OF DL-BASED LOCALIZATION ALGORITHMS

A. OVERVIEW

Our idea is to develop CNVPS algorithms based on DL algorithms owing to their low complexity, high performance, and design flexibility. More specifically, for a single time-slot GPS refinement, we propose a multi-layer perceptron-cooperative neighboring vehicle positioning system (MLP-CNVPS) to achieve a DL-based GPS refinement. For a multiple time-slot GPS refinement, to obtain superior results by considering the temporal correlation in multi-time-slot data, we propose a long short-term memory-cooperative neighboring vehicle positioning system (LSTM-CNVPS). Moreover, a graph convolution network-cooperative neighboring vehicle positioning system (GCN-CNVPS) was further developed to better utilize both temporal and spatial correlations for achieving an efficient GPS refinement, leading to an even better performance compared to the aforementioned DL-based CNVPS algorithms. In the remainder of this section, the motivations and details of the DL-based CNVPS GPS refinement algorithm with different structures are first introduced, and followed by the training specifics at the end of this section.

B. ARCHITECTURE OF THE PROPOSED MLP-CNVPS

As shown in Fig. 2 (a), we propose MLP-CNVPS based on a conventional MLP DL model. Under the multiple time-slots scenario, all extra information and original GPS measurements of the target vehicle will be directly fed into the MLP-CNVPS without preprocessing. As a result, the input layer of MLP-CNVPS can be expressed as a vector of size $1 \times ((5 \times (N-1)+2) \times T)$, where T is the number of time slots, according to the considered system model (i.e., the extra information from N-1 vehicles from the same cooperative group and the GPS measurements of the target vehicle itself). Following the input layer, we constructed four fully connected layers as hidden layers to process the input data. The neurons in each layer were set to 256, 128, 64, and 32, respectively. Behind each layer, the parametric rectified linear unit (PReLU) [27] is employed as the activation function to provide nonlinearity. Then, the output of the last hidden layer is fed into the output layer with two neurons, generating the refined localization of the target vehicle.

As a mathematical expression, the MLP-CNVPS model for multiple time-slot scenarios can be described as follows:

$$\begin{aligned} &f_{\text{MLP-CNVPS}}(\hat{\mathbf{G}}_{\mathrm{T}}, \widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2, \ldots, \widetilde{\mathbf{X}}_{\mathrm{T_c}}; \Theta_{\text{MLP}}) \\ &= \Gamma(W_{\text{out}} \ldots \Gamma(W_2\Gamma(W_1 X_{\text{MLP}} + b_1) + b_2) + \cdots + b_{\text{out}}), \end{aligned} \tag{7}$$

where W_i and b_i represent the weights and bias of the ith layer, respectively, Γ function is the employed PReLU activation function. In addition, $\Theta_{\text{MLP}} = \{\{W_i, b_i\}_{i=1}^4, W_{out}, b_{out}\}$ represents all trainable parameters, and $X_{\text{MLP}} = Vec(\hat{\mathbf{G}}_{\mathrm{T}}, \widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2, \ldots, \widetilde{\mathbf{X}}_{\mathrm{T_c}})$, where $Vec(\cdot)$ is a vectorized operation. Note that the MLP-CNVPS under a single time-slot scenario can be considered as a special case of Eg. (7) when $\mathrm{T_c} = 1$. With the aforementioned DL model architecture, MLP-CNVPS, can be used to construct a nonlinear function and extract hidden information in the given input to refine the GPS estimation results automatically. The simulation results confirm that MLP-CNVPS outperforms the existing CNVPS solutions.

C. ARCHITECTURE OF THE PROPOSED LSTM-CNVPS

Although MLP-CNVPS can extract hidden information in multiple time-slot measurements and provide an improved GPS refinement, the performance achieved can be further improved by further utilizing the correlation of multi-time-slot measurements along the temporal axis. For the considered problem, multiple time-slot measurements actually belong to the time-series data format. However, it should be noted that all features from different time slots are fed into MLP-CNVPS simultaneously, failing to emphasize the correlation of multiple time-slot measurements in the temporal axis. As an alternative, we further propose LSTM-CNVPS to better utilize a temporal correlation in multiple time-slot measurements. More specifically, the proposed LSTM-CNVPS, as shown in Fig. 2(b), is based on LSTM to leverage its ability to extract hidden information in multiple

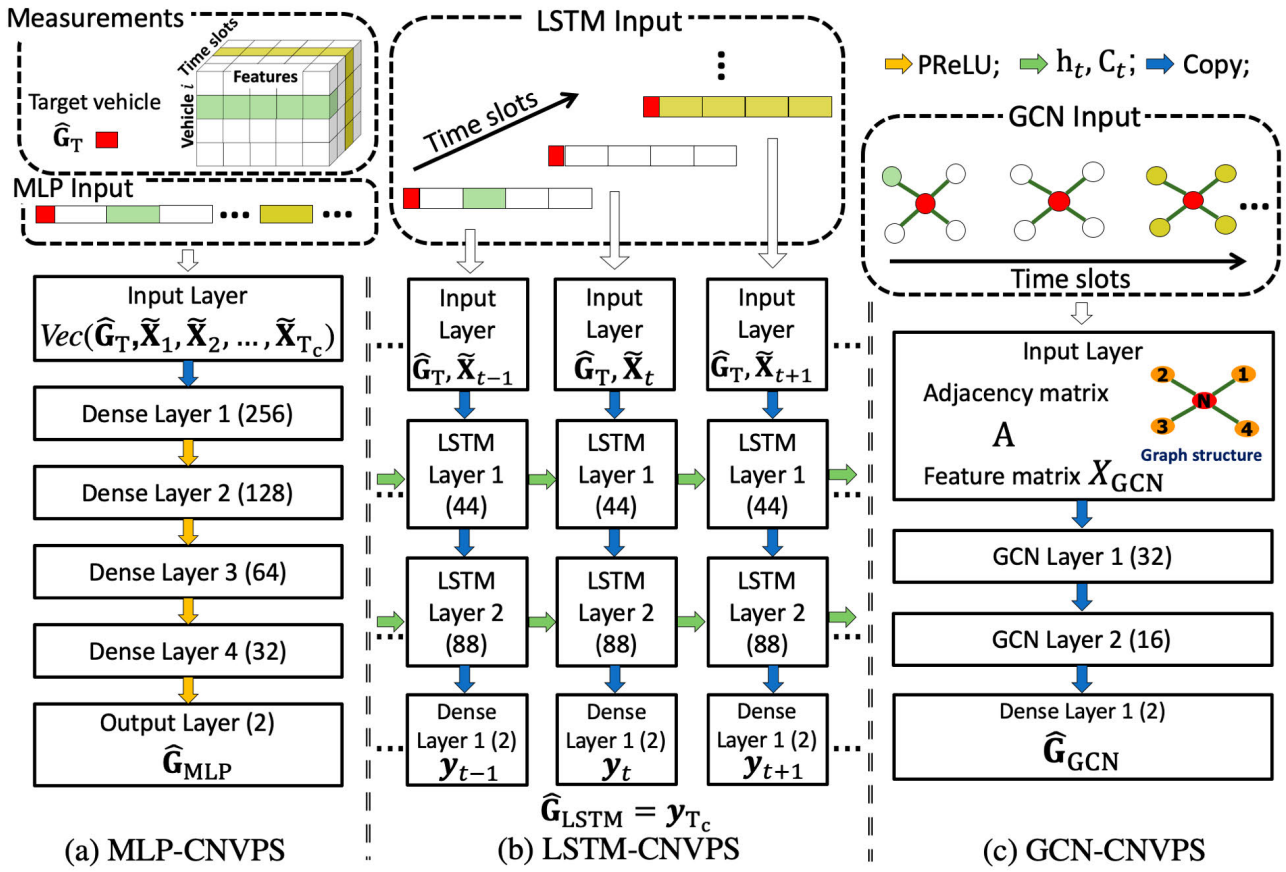

FIGURE 2. Illustration of the proposed model structures under a multiple time-slot condition.(a) MLP-CNVPS: All input data from multiple time slots are fed into the model directly without further pre-processing. (b) LSTM-CNVPS: Input data from multiple time slots are separated and fed into the model in different time slots to emphasize the temporal correlation. (c) GCN-CNVPS: Using the special kernel design of a GCN, the adjacency matrix is used to describe the spatial correlation of input data under the graph structure, and to achieve a superior performance, the temporal and spatial correlation in the input feature matrix will be simultaneously emphasized.

time-slot measurements. In time slot t, the operation of the LSTM-CNVPS can be expressed as follows:

$$\begin{aligned} f_{\mathrm{t}} &= \sigma(W_{\mathrm{f}}^{l}[h_{\mathrm{t-1}}, X_{\mathrm{LSTM,t}}] + b_{\mathrm{f}}^{l}) \\ i_{\mathrm{t}} &= \sigma(W_{\mathrm{i}}^{l}[h_{\mathrm{t-1}}, X_{\mathrm{LSTM,t}}] + b_{\mathrm{i}}^{l}) \\ o_{\mathrm{t}} &= \sigma(W_{\mathrm{o}}^{l}[h_{\mathrm{t-1}}, X_{\mathrm{LSTM,t}}] + b_{\mathrm{o}}^{l}) \\ \widetilde{C}_{\mathrm{t}} &= \tanh(W_{\mathrm{c}}^{l}[h_{\mathrm{t-1}}, X_{\mathrm{LSTM,t}}] + b_{\mathrm{c}}^{l}) \\ C_{\mathrm{t}} &= f_{\mathrm{t}} \odot C_{\mathrm{t-1}} + i_{\mathrm{t}} \odot \widetilde{C}_{\mathrm{t}} \\ h_{\mathrm{t}} &= o_{\mathrm{t}} \odot \tanh(C_{\mathrm{t}}) \end{aligned} \tag{8}$$

where h_t is the state vector, and $X_{\mathrm{LSTM,t}} = Vec(\hat{\mathbf{G}}_{\mathrm{T}}, \widetilde{\mathbf{X}}_t)$ of time slot t. In addition, W_{f}^{l}, W_{i}^{l}, W_{o}^{l}, W_{c}^{l}, and b_{f}^{l}, b_{i}^{l}, b_{o}^{l}, and b_{c}^{l} are the weights and biases of the lth layer. Moreover, $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function, and $\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$ is the hyperbolic tangent function. The $\odot$ operation represents an element-wise multiplication. It is noteworthy that in every time slot, a state vector h will be generated and considered as an input in the next time slot. By doing so, important information from the measurements of the previous time slot will be extracted and kept in this state vector, thereby affecting the operation of LSTM-CNVPS in the next time slot and emphasizing and better utilizing the correlation in the temporal axis compared to MLP-CNVPS.

For the model structure, we constructed an LSTM-CNVPS with two LSTM layers, and the LSTM units of each layer were 44 and 88, respectively. The output is fed into a fully connected layer, which contains two neurons and is regarded as a refined location estimation. The full LSTM-CNVPS model can be described as follows:

$$f_{\text{LSTM-CNVPS}}(\hat{\mathbf{G}}_{\mathrm{T}}, \widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2, \ldots, \widetilde{\mathbf{X}}_{\mathrm{T_c}}; \Theta_{\mathrm{LSTM}}). \tag{9}$$

where Θ_{LSTM} is the set of all trainable weights and biases, and is further expressed as

$$\Theta_{\mathrm{LSTM}} = \{\{W_{\mathrm{f}}^{l}, W_{\mathrm{i}}^{l}, W_{\mathrm{o}}^{l}, W_{\mathrm{c}}^{l}, b_{\mathrm{f}}^{l}, b_{\mathrm{i}}^{l}, b_{\mathrm{o}}^{l}, b_{\mathrm{c}}^{l}\}_{l=1}^{2}, W_{out}, b_{out}\}. \tag{10}$$

Finally, it is worth noting that when we employ LSTM-CNVPS to solve the GPS refinement problem in a single time-slot scenario, because there is only one time slot in (8), the special mechanism of LSTM used to extract the temporal correlation will no longer exist, and the behavior of LSTM-CNVPS is retrograde to that of MLP-CNVPS.

D. ARCHITECTURE OF THE PROPOSED GCN-CNVPS

Although LSTM-CNVPS can better utilize a temporal correlation to improve the performance of MLP-CNVPS, neither MLP-CNVPS nor LSTM-CNVPS consider the spatial correlation of neighboring vehicles. To explain, the extra information provided by different neighboring vehicles should have different weights or levels of confidence considering the relative distance in an adaptive manner, MLP-CNVPS and LSTM-CNVPS cannot support a delicate design for handling this aspect. To further improve the performance of vehicle localization, we propose GCN-CNVPS to simultaneously consider both temporal and spatial correlations. To do so, if the input data belongs to Euclidean space (i.e., image data), a convolutional neural network (CNN) can satisfy the need to consider both temporal and spatial correlations simultaneously based on its special kernel design and consequent convolution operations. However, the input data of the problem considered belongs to a graph representation, limiting the usage of the CNN model. As an alternative, Fig. 2(c) presents the architecture of the proposed GCN-CNVPS. The input for the GCN-CNVPS is represented by an adjacency matrix A and a feature matrix X_{GCN}. Specifically, the adjacency matrix is used to describe the graph structure of interest, allowing the GCN to utilize the spatial correlation in the considered graph structure. We describe this structure to reflect the fact that the target vehicle is able to communicate with the neighboring $N-1$ vehicles, despite these $N-1$ vehicles having no connections with each other. The resulting adjacency matrix of size $N \times N$ can be defined as follows:

$$A=\begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}. \tag{11}$$

However, notice that directly employing adjacency matrix A into GCN-CNVPS causes numerical problems (a gradient explosion and vanishing problem) [28] during the training process, failing to lead to the convergence of optimal weightings. As a result, a normalized adjacency matrix is adopted to prevent the aforementioned issue. In particular, $\hat{A} = \tilde{D}^{-\frac{1}{2}}(A + I_N)\tilde{D}^{-\frac{1}{2}}$ is a normalized adjacency matrix with added-self connections, where I_N is the identity matrix, $\tilde{D} = D + I_N$ is the degree matrix, and $D = diag(\sum_j A_{ij}) \in \mathbb{R}^{N\times N}$. The first $\tilde{D}^{-\frac{1}{2}}$ represents the normalization for each row, and the second represents that for each column. By using the normalized adjacency matrix, the numerical problem during the GCN model training process can be solved. For the model input, the feature matrix X_{GCN} can be represented as a matrix with N rows and $((3 \times (N-1)+2) \times T)$ columns, representing N vehicles and $(3 \times (N-1)+2) \times T$ measurements for T time slots, respectively. To utilize the spatial correlation effectively, the information of each surrounding vehicle should be assigned and placed carefully to be processed separately. Specifically, for the ith row ($i \in [1, N-1]$ for the neighboring vehicles), the measurements are $[\hat{\mathbf{G}}_{t,i}, 0, \ldots, \hat{D}_{t,i}, 0, .., \hat{\Theta}_{t,i}, 0, \ldots, P_{\mathrm{r};t,i}] \in \mathbb{R}^{(3\times(N-1)+2)}$. That is, each neighboring vehicle only acquires the observations of itself and has no information about other vehicles, letting the corresponding values to be set to 0. By contrast, for the Nth row (the target vehicle), the measurements of time slot t can be expressed as $[\hat{\mathbf{G}}_{t,N}, \hat{D}_{t,1}, \ldots, \hat{D}_{t,N-1}, \hat{\Theta}_{t,1}, \ldots, \hat{\Theta}_{t,N-1}.P_{\mathrm{r};t,1}, \ldots, P_{\mathrm{r};t,N-1}] \in \mathbb{R}^{(3\times(N-1)+2)}$. To explain this, the target vehicle possesses its own GPS measurement as well as radar observations and RSSI obtained from neighboring $N-1$ vehicles through V2V communications.

Note that the dimensions of the measurements are slightly different from the previous to separate the acquired information of each vehicle by arranging the measurements in different rows. However, the overall amount of information remains the same.

Finally, for the structure of GCN-CNVPS, to exploit the spatial dependence in the input features, we employed two GCN layers [29], [30], with the number of neurons in each layer being 32 and 16, respectively, in GCN-CNVPS. Specifically, the convolution function of the GCN layer can be expressed as follows:

$$f_{conv}(X_{\mathrm{GCN}}; \hat{A}) = \Gamma(\hat{A}X_{\mathrm{GCN}}W + b). \tag{12}$$

The Γ function is the employed PReLU activation function. Here, W and b represent the trainable weight matrix and bias matrix, respectively. In (12), $\hat{A}X_{\mathrm{GCN}}W$ aggregates all features of neighboring nodes with trainable weights for each node. According to [31]–[33], the operation is analogous to the function of the convolutional kernels in convolutional neural networks (CNNs) and is therefore capable of extracting spatial characteristics in a graph. The output of the last GCN layer was flattened and fed into a fully connected layer. The number of neurons in each layer was 2. The complete operations of GCN-CNVPS can be formulated as follows:

$$\begin{aligned} &f_{\text{GCN-CNVPS}}(X_{\mathrm{GCN}}; \hat{A}; \Theta_{\mathrm{GCN}}) \\ &\quad = f_{out}(f_{conv2}(f_{conv1}(X_{\mathrm{GCN}}; \hat{A}), \hat{A})) \\ &\quad = W_{out}(\Gamma(\hat{A}\Gamma(\hat{A}X_{\mathrm{GCN}}W_1 + b_1)W_2 + b_2)) + b_{out} \end{aligned} \tag{13}$$

Θ_{GCN} is the set of all trainable weights and biases, which can be represented as

$$\Theta_{\mathrm{GCN}} = \{W_1, W_2, b_1, b_2, W_{out}, b_{out}\}. \tag{14}$$

E. TRAINING METHOD

To train the aforementioned DL-based models, supervised learning algorithms were adopted, and the mean square error was employed as the loss function as follows:

$$L(\Theta) = \frac{1}{D}\sum_{i=1}^{D}(\varphi_i - f(X_i; \Theta))^2, \tag{15}$$

where $f(X_i; \Theta)$ is the DL-based model that estimates the result corresponding to a sample X_i with trainable weightings

Θ, φ_i is the true localization, and D is the total number of samples in the training dataset.

Adam [34], a popular gradient descent-based optimizer, was employed to iteratively reduce the loss of each epoch through a backpropagation algorithm during the training process. For MLP-CNVPS, the initial learning rate was set to 0.00001, and the batch size was set to 128. After 1000 epochs, the trained weightings of MLP-CNVPS were recorded, and the offline training process was completed. For LSTM-CNVPS, we set the initial learning rate to 0.00005 and the batch size to 128. After 1000 epochs, the trained weights of the LSTM-CNVPS were recorded, and the LSTM training was completed. For GCN-CNVPS, we set the initial learning rate to 0.0001 and the batch size to 128. After 750 epochs, the trained weights and bias of the GCN-CNVPS were recorded, and the GCN training was completed. Once the offline training process is completed, during the online testing process, the trained DL model can be used to provide vehicle localization estimation results without any further operations.

IV. SIMULATION RESULTS AND DISCUSSION

In this section, three proposed DL-based CNVPS algorithms, MLP-CNVPS, LSTM-CNVPS, and GCN-CNVPS, are evaluated and compared to three existing optimization-based CNVPS algorithms. More specifically, the existing optimization-based CNVPS algorithms, centroid location (CL) algorithm [35], DOA-based location algorithm [5], and optimization-based CNVPS algorithm [7] are implemented in this study as benchmarks. Without the assistance of extra sensors, the CL algorithm simply averages the GPS coordinates of neighboring vehicles to estimate the location of the target vehicle. Thus, the variance of the GPS estimations can be reduced. With the assistance of radar, the DOA-based locating algorithm employs the direction of arrival information of neighboring vehicles to estimate the position of the target vehicle. However, this algorithm cannot exploit additional sensors to further improve the performance. As in the previous study, CNVPS successfully utilizes various sensors to estimate the coordinates of the target vehicle and conducts weighted average localization considering prior knowledge in terms of the standard deviation of each extra sensor. However, because CNVPS only employs a linear function to exploit the information from extra sensors, the achieved performance is limited. Moreover, because the weightings of different sensors are pre-defined according to the statistics of the sensors and remain fixed, the CNVPS cannot adjust the weightings adaptively according to different inputs to achieve better performance. Furthermore, CNVPS also fails to be employed in multiple time-slot scenarios to further improve the performance. In contrast to existing algorithms, DL-based algorithms provide a way to design an adaptive nonlinear function to better utilize the information from extra sensors by extracting temporal and spatial correlations. Moreover, multiple time-slot scenarios can be considered and supported to provide a superior performance compared to a single time-slot scenario. In this section, we first introduce the process of data generation and then compare different algorithms in different scenarios to validate the superiority of DL-based CNVPS algorithms.

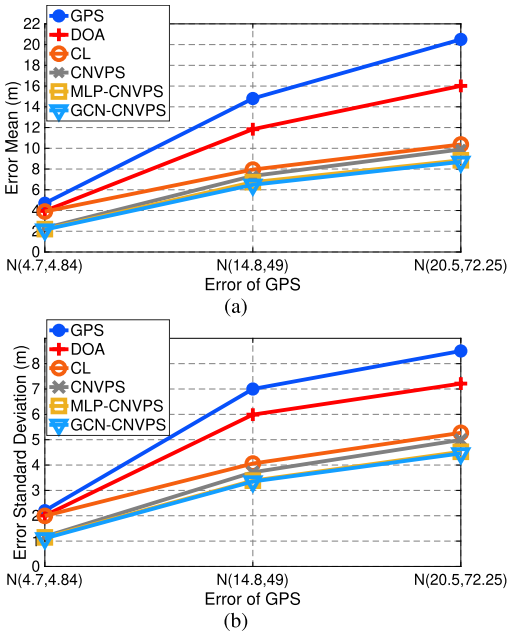

FIGURE 3. (a) Error mean versus scenarios and (b) error standard deviation versus scenario performance with $N = 5$ and $T_c = 1$.

A. DATA GENERATION

To obtain the dataset for model training and testing, we first generate the coordinates of the target vehicle $\mathbf{G}_\mathrm{T}$ and then generate neighboring $N-1$ vehicle coordinates $\{\mathbf{G}_i\}_{i=1}^{N-1}$ with a distance constraint $\|\mathbf{G}_\mathrm{T} - \mathbf{G}_i\|^2 < 10$ (unit: meters). Subsequently, we can obtain measurements according to the sensor configurations mentioned in Sec. II. Specifically, MATLAB software is used to generate virtual measurements for our simulations. We have followed the aforementioned sensor settings and created a scenario as depicted in Fig. 1. In particular, we collect some real data on campus to validate the sensor configurations settings employed in this paper and the results show the same tendency to the generated data based on the system model of this work. As for the generation of multiple time-slot measurements, we specify the vehicle mobility by setting the horizontal velocity V_h and vertical velocity V_v for each vehicle. Moreover, we defined two driving modes, the **straight-through mode** and the **lane change mode**, to set the driver behavior. The vehicle velocities of the former are set as $V_h = 0$ and $V_v \sim u(10, 15)$ m/s, the latter of which are set as $V_h \sim N(0, 1.5^2)$ and $V_v \sim u(10, 15)$ m/s. Based on the aforementioned settings, we set the number of samples for training, validation, and testing datasets to 100000 for each of these three scenarios under different driving modes. We then compute the resulting average mean and standard deviation of the different algorithms to report the error statistics.

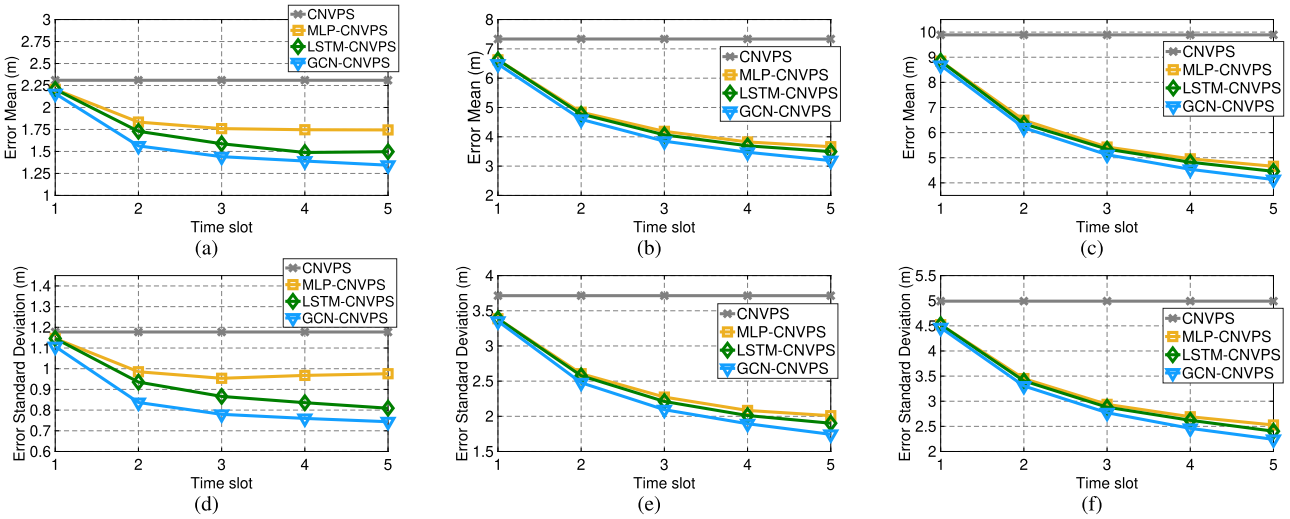

FIGURE 4. Error mean versus time slot performance under (a) free space, (b) suburban, and (c) urban scenarios in straight-through mode with $N = 5$. (d)–(f) Plots of the same for the standard deviation of the error.

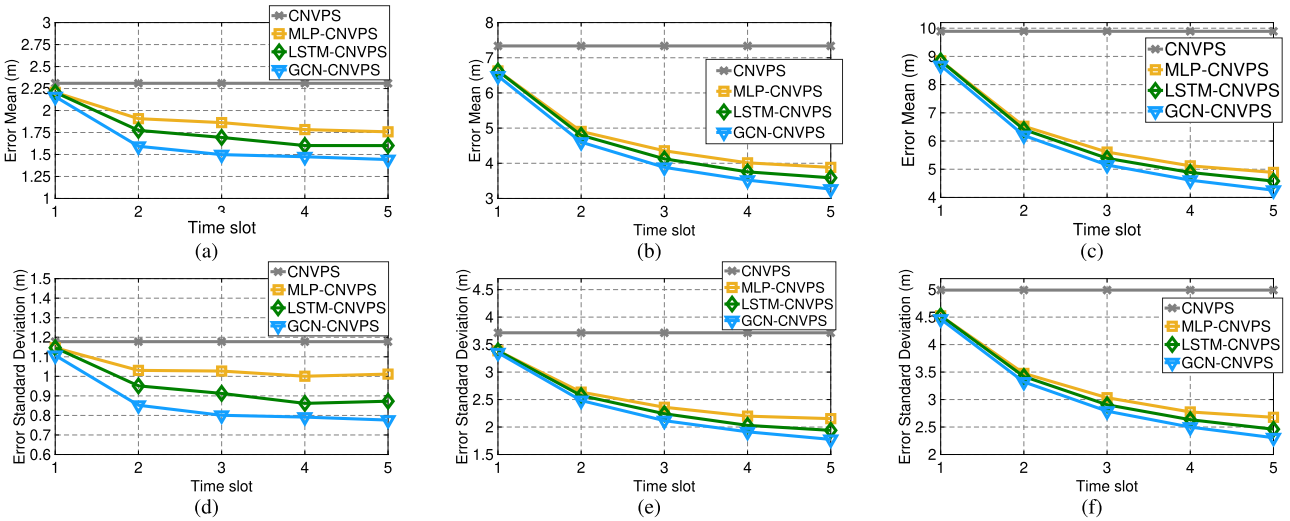

FIGURE 5. Error mean versus time slot performance under (a) free space, (b) suburban, and (c) urban scenarios in lane change mode with $N = 5$. (d)–(f) Plots of the same for the standard deviation of the error.

B. PERFORMANCE ANALYSIS WITH SINGLE TIME-SLOT MEASUREMENTS

In this section, we simulate and discuss the behavior of different CNVPS algorithms in three practical scenarios: *freespace*, *suburban*, and *urban*, under a single time-slot measurement condition. Fig. 3 shows the achieved mean and standard deviation of the estimation error for different algorithms under three scenarios with a cooperative group size equal to 5. Although optimization-based algorithms can improve the GPS estimation error, DL-based algorithms can further improve the GPS measurements by showing a lower achieved mean and standard deviation for all scenarios. Specifically, regardless of the severity of the original GPS estimation error, DL-based CNVPS algorithms can refine the GPS estimations and provide a certain level of improvement. It is also worth noting that GCN-CNVPS slightly outperforms MLP-CNVPS because the spatial correlation is emphasized and better utilized through the special mechanism of GCN operations.

C. PERFORMANCE ANALYSIS WITH MULTIPLE TIME-SLOT MEASUREMENTS

In this section, we discuss the performance of different algorithms under three scenarios with multiple time-slot measurement conditions. Because none of the existing optimization-based algorithms can be extended to multi-time-slot conditions, in this section, we describe the application a single time-slot CNVPS algorithm, which showed the best results among the optimization-based algorithms in the previous simulation, in each time slot instead as a benchmark. Fig. 4 illustrates the error statistics for different algorithms

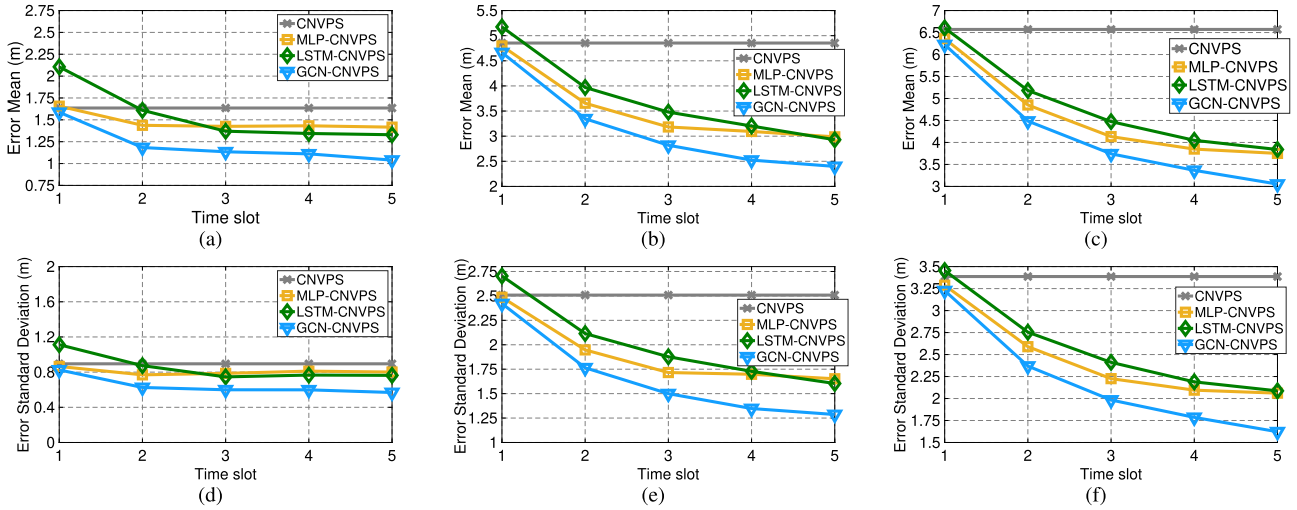

FIGURE 6. Error mean versus time slot performance under (a) free space, (b) suburban, and (c) urban scenarios in straight-through mode with $N = 10$. (d)–(f) Plots of the same for the standard deviation of the error.

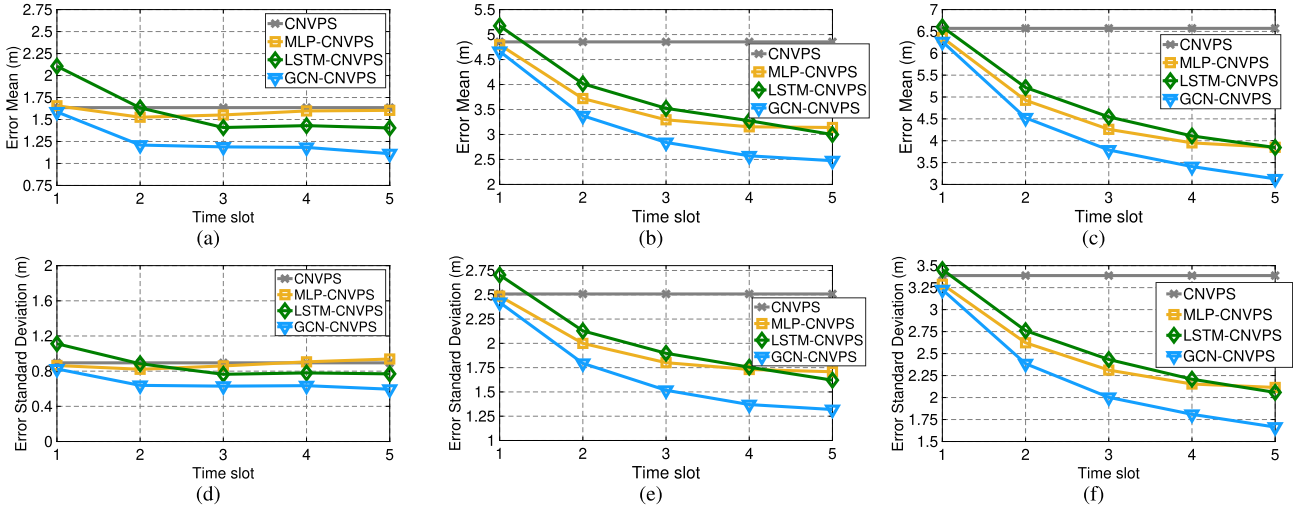

FIGURE 7. Error mean versus time slot performance under (a) free space, (b) suburban, and (c) urban scenarios in lane change mode with $N = 10$. (d)–(f) Plots of the same for the standard deviation of the error.

under the three scenarios when the driving mode is set as the straight-through mode in a cooperative group containing five vehicles. Compared to CNVPS, which fails to utilize information from multiple time-slot measurements to further improve the performance, as the number of time slots increases, all three DL-based CNVPS algorithms can gain from the extra information and achieve a better performance compared to single time-slot measurement conditions. More specifically, LSTM-CNVPS outperforms MLP-CNVPS because of the special mechanism for emphasizing a temporal correlation. Furthermore, GCN-CNVPS can offer a better performance than LSTM-CNVPS because temporal and spatial correlations are both considered through the convolution operations of the GCN model. However, we can also note that the improvement of the urban case is more compelling because the GPS error in this scenario has more room for improvement. However, even under the free-space scenario, DL-based CNVPS algorithms can still be used to improve the original GPS estimation results. Fig. 5 shows the error statistics for different algorithms under the same three scenarios when the driving mode is set as the lane-change mode in a cooperative group containing five vehicles. Nevertheless, we can observe the same behaviors of the three algorithms by showing impressive improvements over the results of the CNVPS algorithm. Note that straight-through mode is easier than lane-change mode because of the relatively fewer variations in directions and relatively higher correlation of locations at different time slots. Hence, we found that all methods perform worse than the straight-through mode. However, among them, GCN-CNVPS always achieved the best performance in

both modes. These results suggest that GCN-CNVPS enhances the performance of GPS by extracting temporal and spatial relationships from historical measurements, confirming our motivation toward the design of GCN-CNVPS.

D. SCALABILITY OF PROPOSED ALGORITHMS

In this section, we further verify the scalability of the proposed algorithms by extending our algorithms to a vehicular scenario that consisting of ten cars. Figs. 6 and 7 present the error statistics for different algorithms in the three scenarios under different driving modes in a cooperative group containing ten vehicles. With additional information provided by increasing the number of surrounding vehicles, the performance of all DL-based CNVPS algorithms improved compared to the previous simulations. However, we also noted that because MLP-CNVPS and LSTM-CNVPS cannot utilize a spatial correlation well, the performance of these two algorithms saturates quickly under this scenario. As an alternative, GCN-CNVPS can handle and utilize the complex spatial correlation between ten vehicles and offer an even better performance than a smaller group of cooperative vehicles, proving the scalability of GCN-CNVPS.

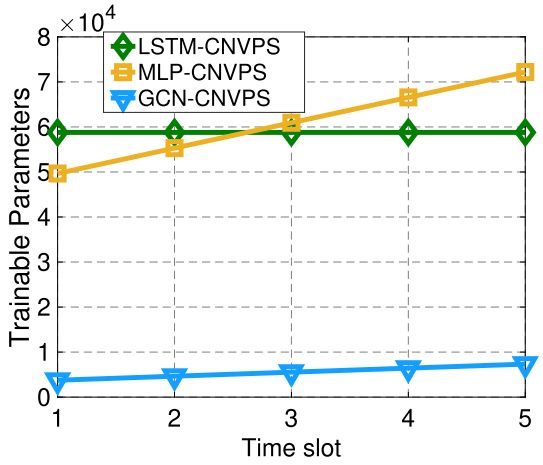

FIGURE 8. comparison of training overhead of different DL-based CNVPS algorithms.

E. TRAINING OVERHEAD OF DIFFERENT DL-BASED CNVPS ALGORITHMS

Fig. 8 shows the number of trainable parameters for different DL-based CNVPS algorithms. Because the dimensions of the input layer of MLP-CNVPS increase with the number of time slots, the number of trainable parameters also increases with this number. For the LSTM-CNVPS, the number of trainable parameters remains the same because LSTM-CNVPS can use the same trainable parameters to process the data from different time slots. Moreover, we can observe that the numbers of trainable parameters of MLP-CNVPS and LSTM-CNVPS are comparable. For GCN-CNVPS, although the number of trainable parameters will also increase with the number of time slots because the number of dimensions of the input layer of GCN-CNVPS also increases with the number of time slots, the rate of increase is fairly flat compared to that of MLP-CNVPS. Finally, we can observe that the training overhead of GCN-CNVPS is far less than that of MLP-CNVPS and LSTM-CNVPS. Note that GCN-CNVPS can also significantly outperform MLP-CNVPS and LSTM-CNVPS. Based on the aforementioned observations, we conclude that GCN-CNVPS is an efficient CNVPS solution with a high performance and low training overhead, because both temporal and spatial correlations are well utilized for aiding the GPS refinement during GCN operations, making GCN-CNVPS a potential solution to assisting the GPS refinement in practice.

V. CONCLUSION

In this study, we proposed several cooperative vehicle localization approaches based on the DL technique to provide precise location estimation results. Specifically, MLP-CNVPS can be used to apply an effective data fusion and aid in the GPS refinement. LSTM-CNVPS was developed by further considering the temporal correlation hidden in the multiple time-slot data. Finally, GCN-CNVPS was developed to consider temporal and spatial correlations simultaneously, offering a higher performance and lower training overhead compared to the existing aforementioned algorithms. Moreover, extensive simulations also confirmed the scalability and robustness of the proposed algorithms, making the developed algorithms potential candidates for use in GPS refinement in practice. Inspired by outstanding performance in this work, we will look for industry partners to test our algorithm in over-the-air scenarios in the future. We also hope that this study will encourage researchers to introduce GCN-based algorithms for efficient vehicular applications.

ACKNOWLEDGMENT

The authors are grateful to all anonymous reviewers for key comments. The contents of this document reflect the views of the author(s) and are not necessarily the views of the university. The author(s) are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of either the North Carolina Department of Transportation or the Federal Highway Administration at the time of publication. This report does not constitute any standards, specifications, or regulations.

REFERENCES

[1] J. Van Brummelen, M. O'Brien, D. Gruyer, and H. Najjaran, "Autonomous vehicle perception: The technology of today and tomorrow," *Transp. Res. C, Emerg. Technol.*, vol. 89, pp. 384–406, Apr. 2018.

[2] K. Liu, H. B. Lim, E. Frazzoli, H. Ji, and V. C. S. Lee, "Improving positioning accuracy using GPS pseudorange measurements for cooperative vehicular localization," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2544–2556, Jul. 2014.

[3] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 829–846, Apr. 2018.

[4] E. C. Eze, S. Zhang, and E. Liu, "Vehicular ad hoc networks (VANETs): Current state, challenges, potentials and way forward," in *Proc. Int. Conf. Autom. Comput.*, Sep. 2014, pp. 176–181.

[5] M. A. Hossain, I. Elshafiey, and A. Al-Sanie, "High accuracy GPS-free vehicular positioning based on V2 V communications and RSU-assisted DOA estimation," in *Proc. 9th IEEE-GCC Conf. Exhib. (GCCCE)*, May 2017, pp. 1–5.

[6] M. Rohani, D. Gingras, and D. Gruyer, "A novel approach for improved vehicular positioning using cooperative map matching and dynamic base station DGPS concept," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 230–239, Jan. 2016.

[7] S. Nam, D. Lee, J. Lee, and S. Park, "CNVPS: Cooperative neighboring vehicle positioning system based on vehicle-to-vehicle communication," *IEEE Access*, vol. 7, pp. 16847–16857, 2019.

[8] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.

[9] J. Liu, G. P. Ong, and X. Chen, "GraphSAGE-based traffic speed forecasting for segment network with sparse data," *IEEE Trans. Intell. Transp. Syst.*, early access, Oct. 6, 2020, doi: 10.1109/TITS.2020.3026025.

[10] S. Banani, S. Kittipiyakul, S. Thiemjarus, and S. Gordon, "Safety message verification using history-based relative-time zone priority scheme," *J. Comput. Netw. Commun.*, vol. 2019, pp. 1–14, Mar. 2019.

[11] R. Halili, M. Weyn, and R. Berkvens, "Comparing localization performance of IEEE 802.11p and LTE-V V2I communications," *Sensors*, vol. 21, no. 6, p. 2031, Mar. 2021.

[12] S. B. Cruz, T. E. Abrudan, Z. Xiao, N. Trigoni, and J. Barros, "Neighbor-aided localization in vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2693–2702, Oct. 2017.

[13] J. Tang, L. Ericson, J. Folkesson, and P. Jensfelt, "GCNV2: Efficient correspondence prediction for real-time SLAM," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3505–3512, Jul. 2019.

[14] *Dedicated Short Range Communications (DSRC) Message Set Dictionary*, Standard J2735_200911, SAE, Sep. 2009.

[15] J. B. Kenney, "Dedicated short-range communications (DSRC) standards in the United States," *Proc. IEEE*, vol. 99, no. 7, pp. 1162–1182, Jul. 2011.

[16] X. Wang, S. Mao, and M. X. Gong, "An overview of 3GPP cellular vehicle-to-everything standards," *GetMobile, Mobile Comput. Commun.*, vol. 21, no. 3, pp. 19–25, Nov. 2017.

[17] R. Parker and S. Valaee, "Vehicle localization in vehicular networks," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2006, pp. 1–5.

[18] N. Drawil and O. Basir, "Toward increasing the localization accuracy of vehicles in VANET," in *Proc. IEEE Int. Conf. Veh. Electron. Saf. (ICVES)*, Jul. 2009, pp. 13–18.

[19] J. Yao, A. T. Balaei, M. Hassan, N. Alam, and A. G. Dempster, "Improving cooperative positioning for vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2810–2823, Jul. 2011.

[20] T. Kos, I. Markezic, and J. Pokrajcic, "Effects of multipath reception on GPS positioning performance," in *Proc. ELMAR*, 2010, pp. 399–402.

[21] M. Goppelt, H.-L. Blöcher, and W. Menzel, "Automotive radar–investigation of mutual interference mechanisms," *Adv. Radio Sci.*, vol. 8, pp. 55–60, Sep. 2010.

[22] F. Folster and H. Rohling, "Data association and tracking for automotive radar networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 4, pp. 370–377, Dec. 2005.

[23] F. de Ponte Müller, E. M. Diaz, and I. Rashdan, "Cooperative positioning and radar sensor fusion for relative localization of vehicles," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jul. 2016, pp. 1060–1065.

[24] N. S. Rajput, "Measurement of IEEE 802.11p performance for basic safety messages in vehicular communications," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS)*, Dec. 2018, pp. 1–4.

[25] S. Al-Stouhi and R. Miucic, "Absolute localization via DSRC signal strength," in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–5.

[26] G. Mao, B. Fidan, and B. D. O. Anderson, "Wireless sensor network localization techniques," *Comput. Netw.*, vol. 51, no. 10, pp. 2529–2553, 2007.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[28] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jul. 2020.

[29] K. Guo, Y. Hu, Z. Qian, H. Liu, K. Zhang, Y. Sun, J. Gao, and B. Yin, "Optimized graph convolution recurrent neural network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1138–1149, Feb. 2021.

[30] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," 2013, *arXiv:1312.6203*.

[31] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[32] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.

[33] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *CoRR*, vol. abs/1606.09375, pp. 1–9, Jun. 2016.

[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–15.

[35] L. Altoaimy and I. Mahgoub, "OWL: Optimized weighted localization for vehicular ad hoc networks," in *Proc. Int. Conf. Connected Vehicles Expo. (ICCVE)*, Nov. 2014, pp. 699–704.

CHIA-HUNG LIN received the B.S. degree in electrical engineering from Chang Gung University, Taoyuan, Taiwan, in 2016, and the M.S. degree from the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, in 2018. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, North Carolina State University.

YO-HUI FANG received the B.S. degree in electronic engineering from the National Changhua University of Education, Changhua, Taiwan, and the M.S. degree from the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan.

HSIN-YUAN CHANG received the B.S. degree in electrical engineering from the National Tsing Hua University, Hsinchu, Taiwan, in 2020, where she is currently pursuing the Ph.D. degree with the Institute of Communications Engineering. Her research interests include communications, neural networks, and signal processing.

YU-CHIEN LIN (Member, IEEE) received the B.S. degree in engineering and system science from the National Tsing Hua University, Hsinchu, Taiwan, in 2016, and the M.S. degree from the Institute of Communications Engineering, National Chiao Tung University, Hsinchu, in 2018. He is currently pursuing the Ph.D. degree with the Department of Electrical Computer Engineering, University of California at Davis, Davis, CA, USA. He is also a Research Assistant with the Institute of Communications Engineering, National Yang Ming Chiao Tung University.

WEI-HO CHUNG received the B.Sc. and M.Sc. degrees in electrical engineering from the National Taiwan University, Taipei, Taiwan, and the Ph.D. degree in electrical engineering from the University of California at Los Angeles, Los Angeles, CA, USA, in 2009. From 2002 to 2005, he was with Chunghwa Telecommunications Company. In 2008, he worked on CDMA systems at Qualcomm, Inc., San Diego, CA. Since January 2010, he has been an Assistant Research Fellow, and promoted to the rank of an Associate Research Fellow with Academia Sinica, in January 2014. Since 2018, he has been a Full Professor and leads the Wireless Communications Laboratory, Department of Electrical Engineering, National Tsing Hua University, Taiwan. He has published over 50 journal articles and over 50 conference papers. His research interests include communications, signal processing, and networks. He received the Ta-You Wu Memorial Award from Ministry of Science and Technology in 2016, the Best Paper Award in IEEE WCNC 2012, and the Taiwan Merit Scholarship from 2005 to 2009.

SHIH-CHUN LIN (Member, IEEE) received the B.S. degree in electrical engineering and the M.S. degree in communication engineering from the National Taiwan University, Taipei, Taiwan, in 2008 and 2010, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2017. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA, where he leads the Intelligent Wireless Networking Laboratory. His research interests include 6G radio and intelligent networking, wireless software-defined architecture, vehicular edge computing, machine learning and mathematical optimization, statistical scheduling, and performance evaluation. He received the Best Student Paper Award Runner-Up at the IEEE International Conference on Services Computing (SCC) 2016 and the Distinguished TPC Member Award at the IEEE INFOCOM 2020.

TA-SUNG LEE (Fellow, IEEE) received the B.S. degree in electrical engineering from the National Taiwan University, in 1983, the M.S. degree in electrical engineering from the University of Wisconsin–Madison, Madison, WI, USA, in 1987, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989. In 1990, he joined the Faculty of National Chiao Tung University (NCTU), Hsinchu, Taiwan, where he is currently a Professor with the Department of Electrical and Computer Engineering. From 2005 to 2007, he was the Chairperson of the Department of Communication Engineering, and from 2007 to 2008 and from 2012 to 2014, he was the Vice President for Student Affairs of NCTU. From 2016 to 2021, he was the Vice President for Research and Development of NCTU. His other past positions include Technical Advisor at Information and Communications Research Labs, Industrial Technology Research Institute (ITRI), and the Managing Director of the Communications and Computer Continuing Education Program, NCTU. He has been the Director of the IoT and Intelligent Systems Research Center, NCTU, since 2017, and a Senior Vice President of the National Yang Ming Chiao Tung University (NYCU, a merger between National Yang-Ming University and NCTU), since 2021. He is active in research and development in advanced techniques for wireless communications, such as smart antenna and MIMO systems, the IoT sensor networks, mobile network resource management, and advanced radar systems for autonomous vehicles. He has led many collaborative projects in several national research programs, such as ''Program for Promoting Academic Excellence of Universities–Phases I and II,'' ''National Science and Technology Program for Telecommunications'' and ''4G Mobile Communications Research Program,'' ''B5G/6G Wireless Communications and Networking Technologies Program,'' and ''Featured Areas Research Center Program of the Higher Education Sprout Project.'' He has published more than 210 peer-reviewed articles and patents. He was the Vice Chairman and the Chairman of IEEE Communications Society Taipei Chapter, from 2005 to 2008, a Board Member of IEEE Taipei Section, from 2007 to 2010, an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, from 2009 to 2013, and IEEE Signal Processing Society Regional Director-at-Large for Region 10, from 2009 to 2013. He is also an Area Editor of *Journal of Signal Processing Systems*. He was appointed as the Commissioner of National Communications Commission (NCC) by the Premier of Taiwan, for the term 2008–2010. He was the Chairman of Telecom Technology Center, a government funded agency for telecommunications RD, from 2013 to 2016. He has won several awards for his research, engineering and education contributions; these include two times National Science Council (NSC) Excellent Research Award, the Young Electrical Engineer Award of the Chinese Institute of Electrical Engineering (CIEE), the Distinguished Electrical Engineering Professor Award of CIEE, two times NCTU Distinguished Scholar Award, and the NCTU Teaching Award. He is an IET fellow. The citation for his elevation to IEEE fellow is ''for leadership and contributions in communication systems and signal processing.''

2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
17-20 October, 2021. Melbourne, Australia

2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC) | 978-1-6654-4207-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/SMC52423.2021.9659109

A C-V2X Platform Using Transportation Data and Spectrum-Aware Sidelink Access

Chia-Hung Lin*, **Shih-Chun Lin***, **Chien-Yuan Wang*** and **Thomas Chase**†

*Intelligent Wireless Networking Laboratory, Department of Electrical and Computer Engineering,
North Carolina State University, Raleigh, NC 27695
†Institute for Transportation Research and Education, North Carolina State University, Raleigh, NC 27695
Email: clin25@ncsu.edu; slin23@ncsu.edu; cwang64@ncsu.edu; rtchase@ncsu.edu

***Abstract*—Intelligent transportation systems and autonomous vehicles are expected to bring new experiences with enhanced efficiency and safety to road users in the near future. However, an efficient and robust vehicular communication system should act as a strong backbone to offer the needed infrastructure connectivity. Deep learning (DL)-based algorithms are widely adopted recently in various vehicular communication applications due to their achieved low latency and fast reconfiguration properties. Yet, collecting actual and sufficient transportation data to train DL-based vehicular communication models is costly and complex. This paper introduces a cellular vehicle-to-everything (C-V2X) verification platform based on an actual traffic simulator and spectrum-aware access. This integrated platform can generate realistic transportation and communication data, benefiting the development and adaptivity of DL-based solutions. Accordingly, vehicular spectrum recognition and management are further investigated to demonstrate the potentials of dynamic slidelink access. Numerical results show that our platform can effectively train and realize DL-based C-V2X algorithms. The developed slidelink communication scheme can adopt different operating bands with remarkable spectrum detection performance, validating its practicality in real-world vehicular environments.**

I. Introduction

With the promising connectivity provided by next-generation communication systems, such as 5G and 6G, several novel applications are developing in full flourish. Among those applications, intelligent transportation system (ITS) and autonomous vehicles are anticipated to bring new experiences with enhanced efficiency and safety to road users in the near future [1]. In macroscopic scale, with the support of ITS algorithms, the data collected from road users can be used to perform traffic flow prediction, optimal path planning, traffic light control, and so on, leading to the improved transportation efficiency. In microscopic scale, via exchanging information with surrounding vehicles and infrastructures, autonomous vehicles can be employed to assist drivers to mitigate fatigue and increase safety during driving. However, an efficient and robust vehicular communication system should act as a strong backbone to offer the needed infrastructure connectivity of ITS and autonomous vehicles [2]. In light of this situation, vehicular communications is a important but challenging research topic to communication society, bringing several special challenges to transitional communication algorithms [2]. Specifically, first, ultra-reliable low latency communications (URLLC) is an important consideration to vehicular communications due to its short reaction-time nature. Secondly, as the high mobility feature of vehicular communications, algorithms should be able to adopt different environments via fast reconfiguration to better serve vehicular communications. As a result, developing suitable algorithms for vehicular communications, which satisfies the aforementioned requirements, is not trivial and challenging.

In the past few years, researchers focus on the development of deep learning (DL)-based solutions for vehicular communications. Compared to traditional optimization-based algorithms, which often involves complex matrix operations and iterations, DL models can be used to construct efficient algorithms with URLLC and adaptability since exhausted computation tasks can be finished in offline training phase. Hence, DL-based algorithms are widely adopted in various vehicular communication applications [1]. However, to train DL-based algorithms, abundant amount of data with high quality should be provided. When it comes to vehicular communications, collecting sufficient amount of real transportation data is difficult and costly. Many works are forced to use simulation data to finish experiments, limiting the power of DL-based algorithms due to the absence of critical features in practical and also raising doubts about its practicality [3]. To tackle this dilemma, in this paper, we propose a vehicular communication platform based on real traffic simulator. As shown in Fig. 1, the proposed platform integrates both transportation and communication modules for the verification of various wireless communication and networking algorithms. Specifically, several open source software, including Simulation of Urban MObility (SUMO), openStreetMap (OSM), and OSM Web Wizard [4], are combined in the transportation module. By doing so, in our transportation module, customized geographic information can be imported as a region of interest first,

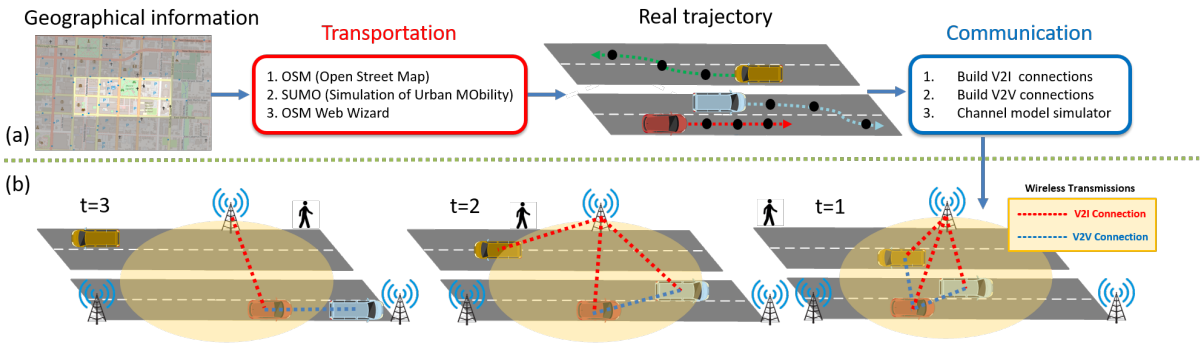

Fig. 1. The Illustration of the proposed vehicular communication platform. In the proposed platform, transportation module and communication module are integrated to generate realistic data for verification of vehicular communication algorithms. To be more specific, when a geographical information is imported, transportation module will create realistic traffic on it, then wireless communications will be simulated. As a result, the proposed platform can be used to assess the achieved performance of various vehicular communication algorithms in real scenarios.

then realistic trajectories of each vehicle will be created on the region of interest based on adjustable parameters. Given those realistic trajectories, in our communication module, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) connections, which is simulated based on latest specifications and papers, are created occasionally to simulate vehicular communication environment. Moreover, we offer sub6GHz and Terahertz (THz) wireless communications for users to choose from now. Different from existing simulation platforms, the trajectories generated from our transportation module will be affected by traffic light and the actions of surrounding vehicles, offering realistic and critical patterns. At the same time, given a interested scenario, the proposed platform can be used to generate huge amount of data quickly, stimulating the development of DL-based vehicular communication algorithms. To validate the proposed platform, we develop a DL-based spectrum management algorithm for sub6GHz vehicular communications. Moreover, the ability of the proposed algorithm to adopt different communication scenarios is also investigated in detail. To explain, spectrum management is an indispensable component to realize efficient vehicular communications. By employing spectrum sensing algorithms in vehicular communications, underutilized spectrum can be reused and consequently leads to a better overall spectrum efficiency [5]. Also, an effective spectrum sensing algorithm can be used as a preceding algorithm to perform critical radio resources allocations in vehicular communications. When it comes to heterogeneous communication systems or military usage [6], spectrum management also shows its importance in terms of provided efficiency improvement. Therefore, we implement critical spectrum sensing algorithms in our platform for further verification. We summarize the major contributions of this work as follows:

- A vehicular communication verification platform is proposed to demonstrate vehicular communication algorithms and assess their performance in real scenarios. To the best of our knowledge, this work is the first to provide an integrated vehicular platform with real traffic simulator and next generation communication systems following specifications and standards. Specifically, realistic traffic in customized region of interest can be created by the transportation module. Then communication connections will be conducted via either sub6GHz or THz bands by the communication module. The proposed platform can be used to demonstrate existing vehicular communication algorithms, especially DL-based algorithm to satisfy the need of hung amount of data with high quality.
- To demonstrate the developed platform, we develop DL-based spectrum management sub6GHz solution based on our previous work [7]. Note all the communication system settings are following latest standard and specifications to better serve current vehicular communications. The comparison of the proposed DL-based and existing generative adversarial network (GAN)-based algorithms is also conducted in different bands. Simulation results reveal that our spectrum management algorithm can offer excellent performance regardless the operation band, showing the practicality of our spectrum management algorithm.

The rest of paper is organized as follows. Section II introduces the considered spectrum management problem. Section III provides the detail of the developed algorithm. In section III, we first introduce the comprehensive information about the developed platform, and then use the generated data to present and discuss the achieved performance of spectrum management algorithms. Finally, section V concludes the whole paper.

II. System Model and Problem Description

A. Signal and Communication Setups

Considering a vehicular communication scenario, followed by cellular-V2X (C-V2X) structure with operation

frequency band $[W_1, W_2]$ (Hz), a base station (BS) serves several user equipments (UEs) in its cell coverage. To increase spectrum efficiency by performing spectrum management, each transmission time-slot will be divided into two phases: sensing phase and data transmission phase. In sensing phases, an UE, which wishes to create new connection using spectrum holes in the radio environment, will perform spectrum sensing to detect existing connections and seek for idle bands first, then the desired connection can be built in the following data transmission phase. Suppose there are I connections existing in the time-slot, the received signal $\mathbf{x}_c(t)$ at the UE side can be modeled as

$$\mathbf{r}_c(t) = \mathbf{x}_c(t) + \mathbf{w}(t) = \sum_{n=1}^{I} \mathbf{h}_n(t) * \mathbf{s}_n(t) + \mathbf{w}(t), \quad (1)$$

where $\mathbf{s}_n(t)$ and $\mathbf{h}_n(t)$ are the transmitted signal and corresponding channel effect of existing transmissions and $\mathbf{w}(t)$ is the noise in receiver side. In a given sensing phase, by using an ADC with Nyquist sampling rate f_s, a perfect discrete time sequence $\mathbf{r}[n] = \mathbf{x}[n] + \mathbf{w}[n] = \mathbf{x}_c(\frac{n}{f_s}) + \mathbf{w}(\frac{n}{f_s}), n = 0, 1, ..., N_s - 1$ can be obtained as $\mathbf{r} \in C^{N_s \times 1}$. Also, the corresponding discrete Fourier transformation (DFT) spectrum can be obtained by performing $\mathbf{R} = \mathbf{Fr}$, where $\mathbf{F}$ is the DFT basis matrix. However, getting DFT spectrum in this way is costly and impractical. To tackle this drawback, spectrum reconstruction problem will be introduced and formulated.

B. Problem Formulation

Sub-Nyquist sampling and spectrum reconstruction based on undersampled measurements can be introduced to avoid the exhausted sampling process. To explain mathematically, a sub-Nyquist sampling process can be expressed as

$$\mathbf{y} = \Phi\mathbf{Fr}, \quad (2)$$

where $\Phi_{M \times N_s}$ is the sensing matrix, which satisfies $M < N_s$ to perform compression. Our goal in this paper is to design a way to reconstruct the the desired signal $\mathbf{x} \in C^{N_s \times 1}$ or corresponding clean DFT spectrum $\mathbf{X} = \mathbf{Fx}$ from the noisy and undersampled measurements $\mathbf{y} \in C^{M \times 1}$. Traditional compressed sensing algorithms can be used to perform reconstruction by solving the optimization problem:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & ||\mathbf{x}||_1 \\ \text{subject to} \quad & ||\mathbf{y} - \Phi\mathbf{Fx}||_2 \leq \epsilon, \end{aligned} \quad (3)$$

where ϵ is the distortion threshold. However, the computational efficiency and execution time can be further improved. To explain, existing spectrum reconstruction algorithms only focus on the design of reconstruction process and ignore the design of compression process, limiting the achieved efficiency. As an alternative, we aim to design a joint compression and reconstruction algorithm to improve the overall efficiency. Hence, instead of solving Eq. (3), the interested joint optimization problem can be expressed as:

$$\underset{f,\Phi}{\text{minimize}} \quad ||\mathbf{X} - f(\Phi\mathbf{Fr})||_2, \quad (4)$$

where f is the reconstruction function.

III. The Development of DL-based Spectrum Reconstruction

To solve the optimization problem in Eq. (4), we develop a DL-based algorithm to perform efficient compression and reconstruction. To explain, designing and evaluating a sensing matrix in the compression process via optimization-based algorithms is not travail, turning out that existing algorithms only focus on solving Eq. (3) instead of solving Eq. (4). In the proposed algorithm, benefiting from the end-to-end learning nature of DL-based algorithm, backprorogation algorithm is used to improve the compression and reconstruction process simultaneously via gradient decent mechanism in each epoch, achieving effective joint optimization. As a result of the joint optimization, undersampled measurements with more critical information compared to existing algorithms can be obtained, leading to better efficiency and consequent better reconstruction results.

A. Deep Learning Model Architecture

We create three consecutive modules (namely compression, coarse reconstruction, and fine reconstruction modules) in our DL model for efficient spectrum reconstruction. The compression module is a specially-designed one-layer convolutional neural network (CNN) with M filters to produce undersampled measurements by making the trainable weights in the compression module act as the content of sensing matrix, being expressed as:

$$\mathbf{z}_{DL} = \Phi_{DL}\mathbf{Fr}, \quad (5)$$

where $\mathbf{Fr} \in \mathbb{C}^{N_s}$ is the original spectrum, $\Phi_{DL} \in \mathbb{C}^{M \times N_s}$ is the sensing matrix designed by the compression module, and $\mathbf{z}_{DL} \in \mathbb{C}^M$ is the obtained undersampled measurements. To be more specific, the input of the compression module $\mathbf{Fr}$ is presented as a real vector with the size of $N_s \times 2$. After the operation of the *compression module*, the output $\mathbf{z}_{DL}$ is a real vector with the size of $M \times 2$, standing for the real part and the imaginary part of the undersampled measurements. Note that there is no activation function in this CNN layer to ensure the whole compression module as a linear operation. After training, the trainable weights in each filter can be represented as a pseudo-random (PN) sequence [8]. Then received signals will be mixed with M PN sequences and pass through a low-pass filter to get the undersampled measurements $\mathbf{z}_{DL}$ in real scenarios. Once we get the undersampled measurements $\mathbf{z}_{DL}$, coarse reconstruction and fine reconstruction module will be employed to

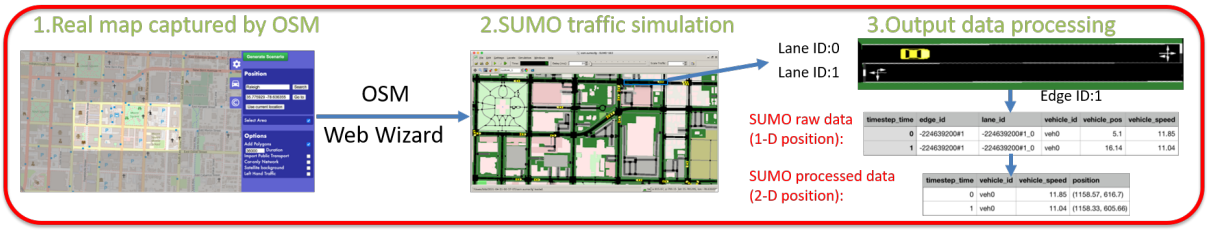

Fig. 2. The detailed process of the transportation module: Using the transportation module, realistic traffic will be created in region of interest and the trajectories of vehicles will be recorded for further use.

perform spectrum reconstruction. The coarse reconstruction is another CNN layer, which has N_s filters with the size of $M \times 2$, to generate initial reconstruction. Different from the compression module, batch normalization (BN) and PRelu are employed in this CNN layer to offer nonlinearty. Then the fine reconstruction module, which is developed based on ResNet-structure and has six residual blocks, will be employed to finish final reconstruction. In each residual block, three CNN layers with number of filters 64, 32, and 2, respectively, are built to refine the initial spectrum reconstruction. Behind each layer, we also employ BN and PRelu in the initial reconstruction module to complete the fine-scale reconstruction.

B. Loss Functions

To leverage the end-to-end training mechanism for solving interested optimization Eq. (4) efficiently, Eq. (4) is employed as as the loss function of the aforementioned model directly. Let Θ_{CR} stands for the trainable weight in the coarse reconstruction module and Θ_{FR} represent the trainable weight in the fine reconstruction module and $f(x; \Theta_{CR}.\Theta_{FR})$ is the nonlinear transformation with Θ_{CR} and Θ_{FR}, the loss function can be expressed as

$$Loss = ||\mathbf{Fs} - f(\Phi_{DL}\mathbf{Fr}; \Theta_{CR}.\Theta_{FR})||^2. \quad (6)$$

During each epoch, Φ_{DL}, Θ_{CR}, and Θ_{FR} will be updated jointly to minimize the training loss until convergence, generating optimal sensing matrix and trainable weights in each module. As for the training specifics, the Adam optimizer with initial learning rate 0.0005 is employed to minimize the loss function. The number of epochs is set as 20 and the mini-batch mechanism is employed with batch size as 128 to facilitate fast convergence.

IV. A New C-V2X Platform and Numerical Results

A. Vehicular Communication Platform Construction

To evaluate the performance of vehicular algorithms in real scenarios, as shown in Fig. 1, we create a platform integrating real transportation module and spec-defined communication module to obtain simulation results.

1) Transportation module: In order to obtain realistic traffic data, we introduce SUMO [4] as the transportation module of the proposed platform. To explain, SUMO is an open source platform that can create microscopic realistic traffic data. Moreover, with the OSM combination, we can import the geographic information from real map to evaluate the achieved performance of vehicular communication algorithms in any region of interest of real locations. As shown in Fig. 2, given a fixed location, in our transportation module, several parameters can be adjusted to simulate the transportation behavior during different time or events. In this paper, a two direction street with four lanes in Raleigh Downtown, North Carolina is selected as region of interest to investigate the behavior of vehicular communication algorithms in Urban scenario. The area we selected for simulation is around 600k m^2 and traffic lights are built and used to control the traffic in this area. The time step is set for 1 second and the average seconds of an new vehicle generated on the map is set as 2.5. Furthermore, we adopt vehicles with 5 m length and 1.8 m width to finish the simulations. The maximum speed of each vehicle is set as 55.56 m per second, and minimum gap of two vehicles is 2.5 m. It is noteworthy that the aforementioned parameters can be adjusted flexibly to reflect the traffic in peak hours or non-peak hours or to match any communication specifications or standards. Moreover, different type of vehicles with different size can be simulated by SUMO platform to generate realistic traffic behavior. Based on the aforementioned settings, SUMO raw dump traffic data will be created. Then the original one-dimensional position data will be converted to two-dimensional position data first and be passed to communication module for further processing.

2) Communication module: After we get the discrete traffic data generated from the transportation module, we follow the latest settings of standards and specifications to construct the communication module. To be more specific, we set one BS on middle of the road. The distance between the V2I connections (i.e., BS and any vehicles) and V2V connections (i.e., any pairs with two vehicles) can be calculated in each time step. In the coverage of the BS or the maximum communication range of V2V connections, we set that for each vehicle there is 0.9 chance

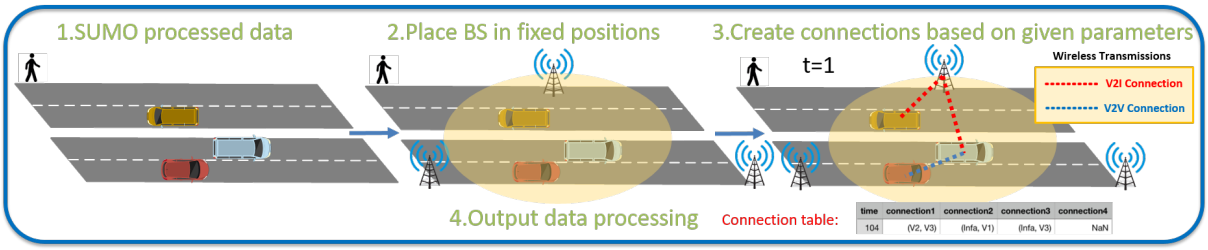

Fig. 3. The detailed process of the communication module: In each time step, based on the realistic trajectories data and communication system settings, wireless connections will be created and the detailed information is also provided in our platform.

that it is communicating with the BS and for each pair of vehicles there is 0.5 chance that they are communicating with each other to create V2V and V2I connections. Next, it is noteworthy that we offer different operation band options with realistic channel behavior to perform simulations in our communication module. Specifically, in Release 17 [9] announced in December 2019, frequency range 1 (sub6GHz) is still considered to support C-V2X communications. We further consider THz band as an option in our platform for next-generation communications. The comparison of achieved performance in those bands can act as a useful reference for system designer to determine the actually band for different applications. We set W_1 and W_2 as $[0, 2GHz]$ according [10] to simulate wideband scenario in sub6GHz band. When it comes to the THz band, we follow our previous paper [7] to set W_1 and W_2 as $[0.1THz, 0.55THz]$. As for the coverage of V2V and V2I connections, we set that as 100m for sub6GHz band and 15m for THz band based on latest papers. In summary, there are 2GHz and 450GHz available bandwidth for the usages of sub6GHz and THz transmission, respectively.

As for the channel effect, we consider frequency and distance-dependent path loss in both bands following the suggestions of 3GPP simulation guideline and latest papers [9]. Specifically, the block slow fading channel with the path loss: $127 + 2 * log_{10}(f) + 30 * log_{10}(d)$, where d (km) is the distance and f is the center frequency of each connections, is set for sub6GHz band. As for THz band, we follow the setting in our previous paper [7] to simulate realistic THz channel behavior with the consideration of non-ideal conditions, such as temperature, atmospheric pressure, air density and non-line of sight effect. The whole simulation process of communication module is shown in Fig. 3, in each time step, different number of connections will be created based on the given trajectories and aforementioned communication parameters. Moreover, the detailed information about the transmitter and receiver of each connection is also provided in the proposed platform.

3) Data set generation: By integrating the above two modules, we generate realistic data set for the training of DL-based algorithms. Specifically, different from existing works, which employs simple transportation model to describe the vehicle mobility, the trajectories of each vehicle are affected by adjacent vehicles and traffic lights on road in our simulations, leading to more realistic simulation results. Based on these trajectories, V2V and V2I connections are built occasionally in our communication module. As compared to latest spectrum reconstruction algorithms, which assumes fixed number of existing connections in their simulations, it is noteworthy that the number of valid connections may change in each sample in our simulations, challenging the ability of DL-based reconstruction algorithms but being more similar to real scenarios. When all the communication connections are generated, we assume $N_s = 256$ subcarriers equally spaced within the available bandwidth. We further assume that each existing users chose a random, non-overlapping group of 5 subcarriers to transmit on, with at least 1 subcarriers of guard on either side. Finally, antenna gains from transmitter G_t and receiver G_r are set as 0 dbi for sub6GHz and 50 dbi for THz bands, respectively for fair comparison. Although we consider the single input single output structure for the communication system design of each band, the achieved results should be the same for multiple input multiple output structure as the only difference is the provided array gain in our interested problem. For each operating band, we generate 50000, 10000, and 10000 samples and corresponding labels for training, validation, and testing sets, respectively. Note that all the inputs and labels are normalized as [1,0] to prevent computational issues of DL-based algorithms. All the reported results are the average values over the testing set.

B. Performance Metrics

We assess different algorithms from both machine learning and communication perspectives in this paper. As for the machine learning performance metrics, mean-square-error (MSE), cosine similarity, and structure similarity (SSIM) between original spectrum and reconstructed results, are reported in this paper [7]. As for the communication performance metrics, we combine the output of different spectrum reconstruction algorithms with simple energy detection mechanism to perform spectrum sensing. In both original

TABLE I
COMPARISON OF SPECTRUM RECONSTRUCTION IN DIFFERENT BANDS WITH SNR = 30 DB AND COMPRESSION RATE = 0.125

	GAN		Proposed	
	sub6GHz	THz	sub6GHz	THz
Machine learning performance metrics				
MSE	0.0685	0.0187	**0.0026**	**9.61e-04**
Cosine Similarity	0.3370	0.4394	**0.9951**	**0.9908**
SSIM	0.3033	0.6289	**0.8523**	**0.9386**
Communication performance metrics				
P_d	2.47e-04	0.0622	**0.9**	**0.9449**
P_f	0	2.68e-04	**2.6e-04**	**0.0014**

and reconstruction spectrum, a subcarrier will be considered as occupied if the power of it exceeds 0.5. Then detection rate P_d and false alarm rate P_f can be calculated to reflect the spectrum sensing performance, Note that the better trade-off between P_d and P_f implies the better spectrum efficiency and better communication quality as new users can always find spectrum holes to conduct interference-free data transmission.

C. Performance Comparison

Table I reflects the achieved performance of different algorithms in different operating bands with signal-to-noise ratio = 30 dB and compression rate = 0.125. In both sub6GHz and THz bands, the proposed algorithm outperforms GAN-based algorithm in various performance metrics significantly. To be more specific, in terms of machine learning performance metrics, the proposed algorithm can offer excellent reconstruction results, which is very close to the original spectrum. Moreover, the proposed algorithm can strike a good trade-off between detection rate and false alarm rate to aid vehicular communications. On the other hand, the reconstruction quality provided by GAN-based algorithm is unacceptable, failing to reflect the utilizing situations of original spectrum and failing to perform effective spectrum management. Specifically, it is noteworthy that the detection rate of GAN-based algorithm is far less than 90%, which is unsatisfactory for any communication systems. It is also noteworthy that the simulation results reveal that the proposed spectrum reconstruction can be employed in different bands as its data-driven nature can adopt different channel characteristics automatically. Finally, although THz bands can provide abundant bandwidth to support data rate-demanding or low latency applications, we report that the extra power consumption or advanced technologies (i.e., beamforming technique, antenna design) should be used to overcome the higher path loss compared to sub6GHz band.

V. CONCLUSION

We propose a vehicular communication verification platform to facilitate the development of DL-based C-V2X communications. A transportation module based on SUMO for realistic transportation data and a communication module based on the latest specifications are integrated to offer practical simulation results. Using the proposed platform, we further develop a DL-based spectrum management algorithm for the sub6GHz band and investigate the adaptability with actual transportation data and different communication settings. Simulation results confirm that the proposed algorithm can be employed in various communication systems to offer impressive results. For future work, we plan to add other 3GPP communication options, such as millimeter wave, and connection types, such as vehicle-to-pedestrian and vehicle-to-network connections. Moreover, we will create computation module to address computation-related issues, including edge computing, on-device learning, and distributed machine learning training in vehicular communications.

ACKNOWLEDGEMENT

This work was supported in part by the North Carolina Department of Transportation (NCDOT) under Award TCE2020-03, the AC21 Special Project Fund, the NC State 2021 Faculty Research and Professional Development (FRPD) Program, and Cisco Systems, Inc.

REFERENCES

[1] C.-H. Lin, Y.-C. Lin, Y.-J. Wu, W.-H. Chung, and T.-S. Lee, "A survey on deep learning-based vehicular communication applications," *Journal of Signal Processing Systems*, pp. 1–20, 2020.

[2] S.-C. Lin, K.-C. Chen, and A. Karimoddini, "Sd-vec: Software-defined vehicular edge computing with ultra-low latency," *arXiv preprint arXiv:2103.14225*, 2021.

[3] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2282–2292, 2019.

[4] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic traffic simulation using sumo," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2575–2582.

[5] M. F. Pervej and S.-C. Lin, "Eco-vehicular edge networks for connected transportation: A distributed multi-agent reinforcement learning approach," in *Proc. IEEE 92nd Vehicular Technology Conference (VTC-Fall)*, 2020.

[6] J. Lu, L. Li, D. Shen, G. Chen, B. Jia, E. Blasch, and K. Pham, "Dynamic multi-arm bandit game based multi-agents spectrum sharing strategy design," in *IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. IEEE, 2017, pp. 1–6.

[7] C.-H. Lin, S.-C. Lin, and E. Blasch, "Tulvcan: Terahertz ultra-broadband learning vehicular channel-aware networking," in *IEEE International Conference on Computer Communications (INFOCOM) Workshop*. IEEE, 2021.

[8] X. Meng, H. Inaltekin, and B. Krongold, "End-to-end deep learning-based compressive spectrum sensing in cognitive radio networks," in *IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.

[9] M. Harounabadi, D. M. Soleymani, S. Bhadauria, M. Leyh, and E. Roth-Mandutz, "V2x in 3gpp standardization: Nr sidelink in release-16 and beyond," *IEEE Communications Standards Magazine*, vol. 5, no. 1, pp. 12–21, 2021.

[10] H. Sun, W.-Y. Chiu, and A. Nallanathan, "Adaptive compressive spectrum sensing for wideband cognitive radios," *IEEE Communications Letters*, vol. 16, no. 11, pp. 1812–1815, 2012.

IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) | 978-1-6654-0443-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/INFOCOMWKSHPS51825.2021.9484613

TULVCAN: Terahertz Ultra-broadband Learning Vehicular Channel-Aware Networking

Chia-Hung Lin*, **Shih-Chun Lin***, and **Erik Blasch**†

*Intelligent Wireless Networking Laboratory, Department of Electrical and Computer Engineering,
North Carolina State University, Raleigh, NC 27695
†Air Force Office of Scientific Research, Arlington, VA 22203
Email: clin25@ncsu.edu; slin23@ncsu.edu; erik.blasch.1@us.af.mil

***Abstract*—Due to spectrum scarcity and increasing wireless capacity demands, terahertz (THz) communications at 0.1-10THz and the corresponding spectrum characterization have emerged to meet diverse service requirements in future 5G and 6G wireless systems. However, conventional compressed sensing techniques to reconstruct the original wideband spectrum with under-sampled measurements become inefficient as local spectral correlation is deliberately omitted. Recent works extend communication methods with deep learning-based algorithms but lack strong ties to THz channel properties. This paper introduces novel THz channel-aware spectrum learning solutions that fully disclose the uniqueness of THz channels when performing such ultra-broadband sensing in vehicular environments. Specifically, a joint design of spectrum compression and reconstruction is proposed through a structured sensing matrix and two-phase reconstruction based on high spreading loss and molecular absorption at THz frequencies. An end-to-end learning framework, namely compression and reconstruction network (CRNet), is further developed with the mean-square-error loss function to improve sensing accuracy while significantly reducing computational complexity. Numerical results show that the CRNet solutions outperform the latest generative adversarial network (GAN) realization with a much higher cosine and structure similarity measures, smaller learning errors, and 56% less required training overheads. This THz Ultra-broadband Learning Vehicular Channel-Aware Networking (TULVCAN) work successfully achieves effective THz spectrum learning and hence allows frequency-agile access.**

I. Introduction

Next-generation communication systems, such as 5G and 6G, are expected to fulfill various requirements in different application scenarios in terms of data rate, latency, and power consumption, among other factors [1]. As a result, next-generation communication systems will likely have a requirement of 100+ Gigabits per second (Gbps) data rate requirement. Thanks to the efforts of researchers [2], [3], a state-of-the-art communications system can already offer up to several Gbps of data rate [4]. However, the Gbps rate is still an order of magnitude below that which is needed to serve popular streaming use-cases. To satisfy the urgent need of multimedia applications, there are two obvious ways of further improving the data rate: increasing the bandwidth used for communication, and improving spectral efficiency at which frequency bands are used [5], [6]. In light of data rate improvements, researchers have started to focus on the utilization of terahertz (THz)-bands for next-generation communications [7]. Although abundant bandwidth is offered, due to the special properties of THz channel such as *ultra-wide bands* and *distance-dependent path loss*, existing physical layer algorithms must be redesigned or enhanced to fully account for the unique properties of THz spectrum communications [7].

Meanwhile, several research papers [8], [9] reveal that the spectrum under-utilization problem, which is caused by the existence of the idle channels, occurs in current communication systems, reducing the overall spectral efficiency. To address this issue, spectrum management [8], [9] has been developed by detecting idle spectrum and temporarily assigning that spectrum to the demanding user, thus improving the overall spectral efficiency. Among all the research topics in spectrum management, spectrum sensing is the most widely discussed in literature [8]–[11] since it is a prerequisite for mitigating error propagation. Recently, a vehicle-to-everything (V2X) data-coordination scenario [12] is a practical usage to employ spectrum sensing algorithms in THz communications. Typically, vehicle-to-infrastructure (V2I) connections will be assigned specific bands for the high bandwidth entertainment applications transmission (e.g., video streaming). On the other hand, vehicle-to-vehicle (V2V) connections may wish to occasionally perform safety message (e.g., vehicle position, speed and heading) transmission using idle bands assigned to the V2I connections. By employing THz Spectrum Sensing (TSS) algorithms in V2X communications, the underutilized spectrum can be reused and consequently leads to a better overall spectrum efficiency [12]. Moreover, TSS also has its potential for frequency resources allocation, heterogeneous communication system, and military usage [11], [13].

Although TSS algorithms in THz communications have potential to better utilize the frequency resources, few existing works contribute to the development of TSS algorithms for THz communications. Most existing SS methods [8], [10] only focus on SS in the narrow band case. While some works [8], [9] concentrate on the development of wide band SS algorithms, there are almost no prior works considering physical channel effect in SS algorithm design, not even saying the unique features of THz communications. When it comes to the ultra-wide band case like THz communications, sub-Nyquist sampling [8], [9] must be introduced to avoid costly

978-1-6654-0443-3/21/$31.00 ©2021 IEEE

and limiting hardware requirements. As a result, compressed sensing (CS) algorithms have been introduced to support spectrum reconstruction from measurements of sub-Nyquist sampling [10] in the past decade. However, they also suffer from high computational complexity to reconstruct the under-sampled signal. Recent research suggests that learning-based compression and reconstruction outperforms traditional CS algorithms since the local correlation is considered to reconstruct the desired output. Although there is a impressive work [14] employing generative adversarial network (GAN) to aid the SS algorithms design, it still employs randomly generated sensing matrix to conduct the compression, failing to perform joint optimization of the compression and reconstruction to get the best reconstruction. Also, the consequently heavy overhead of the training process of GAN-based SS algorithm creates an implementation challenge to be employed in real scenario. In conclusion, the development of a practical SS algorithm, which can be employed in real ultra-wideband communications, is an unsolved problem.

In this paper, we develop a deep learning (DL)-based spectrum reconstruction algorithm, named compression and reconstruction network (CRNet), to offer an efficient SS solution for THz communications. The contributions of this work are listed as:

1) Inspired by CS-based sub-Nyquist SS algorithms [10], we employ DL-based algorithm to perform efficient compression and reconstruction, offering lower computational complexity and decreasing latency for THz communications by considering local correlation of the spectrum in the optimization process.
2) As an addition to existing DL-based algorithms [14], CRNet introduces joint compression and reconstruction mechanism by designing a *structured sensing matrix* and a corresponding reconstruction algorithm in a end-to-end learning manner, further improving reconstruction quality. By doing so, the training overhead can be significantly decreased to achieve when superior performance in terms of reconstruction quality.
3) To the best of our knowledge, no existing works consider channel effect in the development of TSS algorithm. Simulation results demonstrate that the achieved performance of existing DL-based TSS algorithm degenerates severely in THZ communication scenarios. As an alternative, the proposed algorithm can provide robust reconstruction results even in different compression rate scenarios.

II. System Model and Problem Description

A. System Setup and Signal Model

As shown in Fig. 1, consider a small cell including a base station (BS) with N_r receiver antennas and several user equipments (UEs), each with N_t transmitter antennas for vehicular communications. In the downlink phase of the considered THz-band scenario, the BS may occupy frequency bands from f_a to f_b by using some of the N_s subcarriers to perform V2I

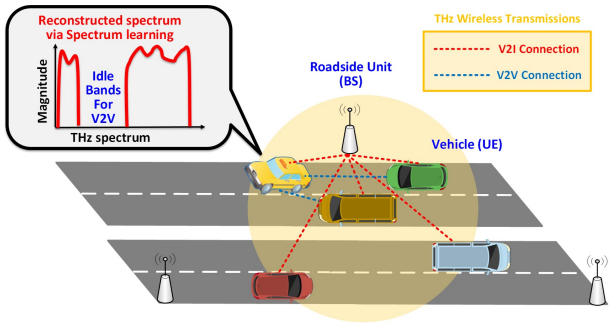

Fig. 1. Considered THz communications for vehicular environment: In downlink phase, the base station (BS) will create several transmission links to different user equirements (UEs) for V2I connections. However, a UE aims to create a V2V connections with surrounding UEs to share safety messages. As a result, the UE should perform SS first to obtain the information of idle spectrum, then use those spectrum to perform V2V connections.

connections. Considering a transmission pair between the BS and i th UE, the complex baseband signal can be expressed as

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i + \mathbf{n}_i, \tag{1}$$

where $\mathbf{y}_i \in \mathbb{C}^{N_t}$ is the received signal, $\mathbf{H}_i \in \mathbb{C}^{N_t \times N_r}$ is the channel matrix, and $\mathbf{x}_i \in \mathbb{C}^{N_r}$ is the transmitted signal, respectively. Assuming a perfect sampling process with a Nyquist sampling rate of $T = 1/2f_b$ discrete-time signals can be obtained from Eq (1). It is assumed that the transmission between the BS-UE pair is with subcarrier f_i and the distance between the BS and UE is d_i in this case, thus THz channel effect $\mathbf{H}_i$ can be further expressed as

$$\begin{aligned} H_i(f_i, d_i) &= H_i^{LOS}(f_i, d_i) + H_i^{NLOS}(f_i, d_i) \\ &= \sqrt{N_t N_r}\alpha_L(f_i, d_i) G_t G_r a_r(\theta_L^r, \phi_L^r) a_t(\theta_L^r, \phi_L^r) \\ &+ \sqrt{N_t N_r} \sum_{i}^{n_{NL}} \alpha_i(f_i, d_i) G_t G_r a_r(\theta_i^r, \phi_i^r) a_t(\theta_i^r, \phi_i^r). \end{aligned} \tag{2}$$

In (2), n_{NL} represents the number of non-line of sight (NLOS) rays. $\alpha_L(f_i, d_i)$ and $\alpha_i(f_i, d_i)$ stands for the complex channel gain of each ray. In each ray, θ/ϕ refers to the azimuth/elevation angles of departure and arrival (AoD/AoA) and $a_t(\theta^t, \phi^t)$ and $a_r(\theta^r, \phi^r)$ are the associated array steering vectors at the transmitter and receiver sides. Finally, G_t and G_r is the transmit and receive antenna gains. Furthermore, the path gain of the LOS ray can be expressed as $|\alpha_L(f_i, d_i)|^2 = L_{spread}(f_i, d_i) L_{abs}(f_i, d_i)$, where $L_{spread}(f_i, d_i)$ represents the spreading loss effect and $L_{abs}(f_i, d_i)$ represents the molecular loss effect, respectively. Given the fact that water vapor molecules cause the majority molecular absorption loss in the THz-band, $L_{abs}(f_i, d_i)$ takes the temperature, atmospheric pressure, and air density into consideration to represent the realistic THz channel behavior. As for the path gain of the NLOS ray, we model it as $|\alpha_i(f_i, d_i)|^2 = \Gamma L_{spread}(f_i, d_i) L_{abs}(f_i, d_i)$, where Γ is the reflection coefficient. Assuming that there are L_1 first-order reflected paths and L_2 second-order reflected paths so that $n_{NL} = L_1 + L_2$. We set the attenuated power of the first-order and second-order reflected paths as 10dB and 20dB, respectively.

B. Problem Description

When a UE aims to create V2V connections with surrounding UEs to share safety messages, a TSS should be performed to detect the idle bands from existing V2V and V2I connections. Yet, in a wideband scenario, to emulate the needed hardware burden, CS must be introduced to aid the reconstruction of the compressed measurements obtained from sub-Nyquist sampling. Considering the aforementioned wideband THz communication scenario, a combining operation can be conducted at the UE to get the time domain measurements r_i, shown as

$$r_i = \mathbf{w}_i^* \mathbf{y}_i = \mathbf{w}_i^* \mathbf{H}_i \mathbf{x}_i + \mathbf{w}_i^* \mathbf{n}_i, \tag{3}$$

where $\mathbf{w}_i^* \in \mathbb{C}^{N_r}$ is the combining weighting. Then the time domain signals of the considered wideband system can be expressed as $\mathbf{r} = \mathbf{s} + \xi = [r_1, ..., r_i, ..., r_{N_s}] \in \mathbb{C}^{N_s}$, where $\mathbf{s} = [\mathbf{w}_1^* \mathbf{H}_1 \mathbf{x}_1, ..., \mathbf{w}_i^* \mathbf{H}_i \mathbf{x}_i, ... \mathbf{w}_{N_s}^* \mathbf{H}_{N_s} \mathbf{x}_{N_s}]$ and $\xi = [\mathbf{w}_1^* \mathbf{n}_1, ..., \mathbf{w}_i^* \mathbf{n}_i, ... \mathbf{w}_{N_s}^* \mathbf{n}_{N_s}]$. Let $\mathbf{F}$ denote a N_s-point discrete Fourier transform (DFT), if the signal is sampled at a sub-Nyquist rate, then the relationship between the clean spectrum $\mathbf{s} \in \mathbb{C}^{N_s}$ and under-sampled measurements $\mathbf{z} \in \mathbb{C}^{N_m}$ can be expressed as

$$\mathbf{z} = \Phi \mathbf{F} \mathbf{r} = \Phi \mathbf{F}(\mathbf{s} + \xi), \tag{4}$$

where $\Phi_{N_m \times N_s}$ is the complex-valued sensing matrix. From Eq. (4), the goal is to design the sensing matrix and the corresponding reconstruction algorithm so that the clean spectrum **Fs** can be recovered from the under-sampled measurements **z** by the reconstructed spectrum $\hat{\mathbf{Fs}}$. It is noteworthy that once a high quality reconstructed spectrum is available, a simple energy detector can be employed to trivially identify the unused frequency bands. Moreover, the reconstructed spectrum with high quality can enable more complex spectrum sharing design in different coexistence models of heterogeneous communication systems [8], [9]. Hence, the motivation is to develop spectrum reconstruction methods.

III. The Development of CRNet

We propose a compression and reconstruction network (CRNet) for efficient spectrum sensing application in THz communications. There are two features of the CRNet. First, conventional SS algorithms, including existing DL-based SS solutions, essentially employ randomly selected (i.e., *unstructured*) sensing matrix to perform compression to get under-sampled measurements, implying there is no special design of the sensing matrix. As an alternative, CRNet firstly introduces the joint design of compression and reconstruction by developing a *structured* sensing matrix and corresponding reconstruction algorithm in a end-to-end learning manner, offering a superior performance compared to existing SS algorithms. Secondly, compared to GAN-based SS algorithms, the training overhead of CRNet is reduced significantly. To be more specific, the under-sampled measurements obtained from the structured sensing matrix are more informative compared to that from unstructured sensing matrix, and the reconstruction can be finished by a low complexity convolutional neural network (CNN) based-model to get a promising reconstruction result. The next section details the CRNet in terms of model architecture, loss function, and training specifics.

A. Model Architecture

As shown in Fig. 2, there are three modules in CRNet, compression, coarse reconstruction, and fine reconstruction modules. The compression module is a specially-designed one-layer CNN to produce under-sampled measurements by making the trainable weights in the compression module act as the content of sensing matrix, being expressed as:

$$\mathbf{z}_{DL} = \Phi_{DL} \mathbf{Fr}, \tag{5}$$

where $\mathbf{Fr} \in \mathbb{C}^{N_s}$ is the original spectrum, $\Phi_{DL} \in \mathbb{C}^{N_m \times N_s}$ is the sensing matrix designed by the compression module, and $\mathbf{z}_{DL} \in \mathbb{C}^{N_m}$ is the under-sampled measurements from the designed sensing matrix. Given an original spectrum **Fr**, in order to feed it into DL-based model, the input of the compression module **Fr** is presented as a real vector with the size of $N_s \times 2$, containing the real part and imaginary part of the original complex vector. After the operation of the *compression module*, the output $\mathbf{z}_{DL}$ is a real vector with the size of $N_m \times 2$, standing for the real part and the imaginary part of the under-sampled measurements. To be more specific, a 1-dimension (1D) CNN layer is constructed with N_m filters in the compression module. In each filter, the trainable weight will be created as a vector with the size of $N_s \times 2$, and then inner product operation between the input and the trainable weight will be conducted on the real part and imaginary part separately to obtain the computed result with the size of 1×2 representing the real part and imaginary part. This operation reflects the matrix operation between each row of the sensing matrix Φ_{DL} and the input **Fr**. As there are N_m filters in this CNN layer, where the size of the output matches Eq. (5) to get the compressed measurements $\mathbf{z}_{DL}$ for the following reconstructions. It is noteworthy that although in the compression module, the computation is performed on the real part and imaginary part separately, the operation is exactly equivalent to the inner product on complex vector as shown in Eq. (5). Moreover, note that there is no activation function in this 1D CNN layer to ensure the whole compression module as a linear operation. Finally, once the training of the compression module is finished, the trainable weights in each filter (i.e., each row of the sensing matrix) can be represented as a pseudo-random (PN) sequence as shown in [14]. By mixing the received signal with N_m PN sequences (as there are N_m rows in the sensing matrix Φ_{DL}) and passing through a low-pass filter, the compressed measurements $\mathbf{z}_{DL}$ can be obtained. For a real scenario, there are no implementation issues to employ the practical CRNet DL-based spectrum reconstruction algorithm.

After CRNet obtains the compressed measurements $\mathbf{z}_{DL}$, the *coarse reconstruction module* aims to provide an initial reconstruction for the following refinements. To do so, another

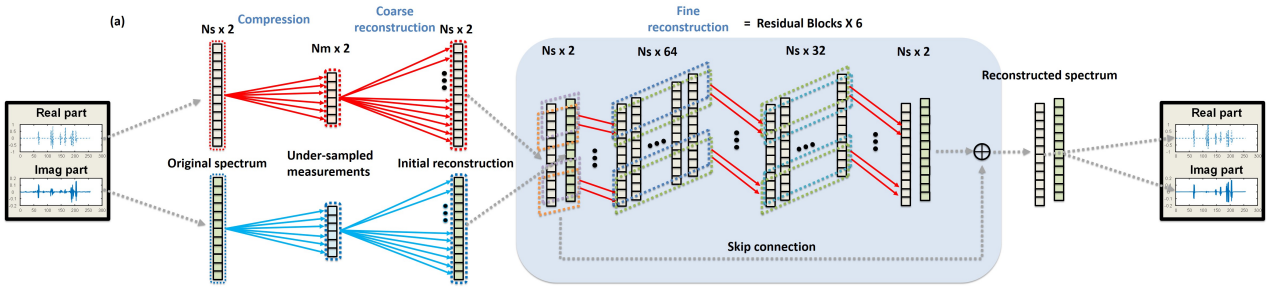

Fig. 2. The model architecture of the CRNet. In CRNet, there are three modules, compression, coarse reconstruction, and fine reconstruction. In the compression and coarse reconstruction modules, the real and imaginary part of original spectrum are compressed and reconstructed, separately. Then, the fine reconstruction module with ResNet-structure is employed to perform fine-scale reconstruction to obtain highly-fidelity reconstructed spectrum. Note that the compression and reconstruction process are performed in a end-to-end training model. As a result, the joint optimization can be performed to obtain optimal weights.

1D CNN layer is employed, which has N_s filters with the size of $N_m \times 2$. After the CNN layer, batch normalization (BN) and parameter-Relu (PRelu) are employed to accelerate convergence and provide non-linearity, respectively. By doing so, an initial reconstruction with the size of $N_s \times 2$, which is the same as the original spectrum, is obtained for the following fine reconstruction module. As for the architecture of the fine reconstruction module, the ResNet-structure [15] gradually refines the initial reconstruction result. To be more specific, the spectrum reconstruction problem is treated as a special image reconstruction problem and employs computer vision techniques to perform meticulous reconstruction. There are two main advantages to introduce ResNet-structure into the design of the proposed spectrum reconstruction algorithm. To explain, a typical DL model with ResNet-structure usually contains several residual blocks, which is built by several CNN layers. Instead of asking a DL model to provide a reconstruction result with high quality from scratch, the ResNet-structure lets a residual block refine the current reconstruction result based on the knowledge from previous residual blocks. As a result, all the residual blocks can coordinate with each other to synergistically produce a final reconstruction result. Another advantage of ResNet-structure is that a DL model with ResNet-structure is more unlikely to suffer from the over-fitting as the special skip-connection mechanism lets the DL model control the number of efficient weightings. For the exact architecture of the fine reconstruction module, in each residual block, three 1D CNN layers with number of filters 64, 32, and 2, respectively, are built to refine the initial spectrum reconstruction. Behind each layer, BN and PRelu are also employed as the setting in the initial reconstruction module. In this paper, the CRNet comprises six residual blocks to perform fine-scale reconstruction as experiments with higher numbers of residual blocks do not improve overall performance but increase computational complexity.

From Fig. 2, a fully-CNN architecture is used in the CRNet DL-based algorithm design. There are two main reasons to use the CNN architecture. First, the needed number of trainable weights can be decreased significantly as the result of the weight sharing mechanism of the CNN. Simulation results confirm that the CRNet model outperforms existing DL models with significantly lower trainable parameters. Secondly, the CRNet fine reconstruction module aims to capture the occupied spectrum to perform fine-scale reconstruction. As the domain knowledge suggests that the occupied spectrum may appear everywhere of the whole spectrum, the convolution operation introduced by CNN can be used to captured the pattern of occupied spectrum regardless the location and number of the occupied spectrum to perform fine-scale reconstruction.

B. Loss function

An end-to-end learning is employed to jointly update all the trainable parameters in CRNet. As a result, the whole compression and reconstruction process can be designed simultaneously to achieve better performance. Let Θ_{CR} stand for the trainable weight in the coarse estimator and Θ_{FR} represent the trainable weight in the fine estimator and $f(x; \Theta_{CR}.\Theta_{FR})$ is the nonlinear transformation with Θ_{CR} and Θ_{FR}, The Mean-square-error (MSE) loss function is set for the model updating, that is

$$Loss = ||\mathbf{Fs} - f(\Phi_{DL}\mathbf{Fr}; \Theta_{CR}.\Theta_{FR})||^2. \quad (6)$$

Note that during each training cycle, Φ_{DL}, Θ_{CR}, and Θ_{FR} will be updated jointly via the back-propagation process to gradually minimize the training loss until convergence, generating optimal structured sensing matrix and trainable weights. Finally, as for the training specifics of the scenario in this paper, the Adam optimizer minimizes the aforementioned loss function. The initial learning rate is set as 0.0005 and the number of epochs is set as 20. The mini-batch mechanism is employed with batch size as 128 to facilitate fast convergence.

IV. Simulation Results

A. Data set preparation

We follow the system model to generate spectrum samples in THz communications for DL-based algorithms training. To be more specific, we set f_a and f_b as 0.1THz and 0.64THz to employ the commonly used transmission window in THz communications. $N_s = 256$ subcarriers are equally spaced within the transmission window. Moreover, we assume that

there are 8 existing users, who all chose a random, non-overlapping group of 5 subcarriers to transmit on, with at least 1 subcarriers of guard on either side. Each user ranges $1-10$ meters from the BS. As for the configuration of the considered communication system, $N_t = N_r = 1$ and $G_t = G_r = 30$ dBi for comparable analysis in [16]. To implement the GAN-based algorithm, we generated the sensing matrix by randomly selecting rows from the inverse DFT matrix to generate under-sampled measurements. As or the CRNet, the original spectrum is considered as input and the under-sampled measurements can be obtained after compression module. Note that all the inputs and labels of the CRNet and GAN-based algorithm are normalized as [1,0] to prevent computational issues. The number of training, validation, and testing sets are 50000, 10000, and 10000 respectively. All the results reported in this paper are the average result over the testing set.

B. Performance metrics

Three performance metrics: mean-square-error (MSE), cosine similarity, and structure similarity (SSIM), are provided to report the achieved performance of the different algorithms. To explain, although MSE can be used to evaluate the reconstruction quality, it cannot reflect the structure similarity between the original spectrum and the reconstructed one. As a result, two additional similarity metrics, cosine similarity and SSIM, evaluate the different algorithms in terms of reconstruction quality. Given a reconstructed spectrum $\hat{\mathbf{Fs}} \in \mathbb{C}^{N_s}$ and the corresponding original spectrum $\mathbf{Fs} \in \mathbb{C}^{N_s}$, the MSE performance metric can be expressed as

$$MSE = ||\mathbf{Fs} - \hat{\mathbf{Fs}}||^2, \tag{7}$$

the cosine similarity ρ can be evaluated as

$$\rho = \frac{< \mathbf{Fs}, \hat{\mathbf{Fs}} >}{||\mathbf{Fs}||||\hat{\mathbf{Fs}}||}, \tag{8}$$

and the SSIM η can be computed as

$$\eta = \frac{(2\mu_{\mathbf{Fs}}\mu_{\hat{\mathbf{Fs}}} + c_1)(2\sigma_{\mathbf{Fs}\hat{\mathbf{Fs}}} + c_2)}{(\mu^2_{\mathbf{Fs}} + \mu^2_{\hat{\mathbf{Fs}}} + c_1)(\sigma^2_{\mathbf{Fs}} + \sigma^2_{\hat{\mathbf{Fs}}} + c_2)}, \tag{9}$$

where μ is the mean value, σ represents the variance, and c is the default constant to stabilize the division. Note that as for the SSIM metric, we compute the SSIM value of real part and imaginary part of spectrum separately and present the average value among those two channels as SSIM only supports the computation of single-channel real vectors.

C. Performance of CRNet in different SNR

Table I presents the achieved performance of the different algorithms for various signal-to-noise ratios (SNR). As shown in Table I, although considering THz channel effect increases the difficulty of performing successful spectrum reconstruction, the CRNet significantly outperforms the GAN-based algorithm. The results demonstrate the advantage of the joint optimization between compression and reconstruction from the CRNet towards the DL model architecture design. The gap between the achieved performance of two algorithms becomes more compelling with higher SNR.

TABLE I
COMPARISON OF SPECTRUM RECONSTRUCTION IN DIFFERENT SNR WITH COMPRESSED RATE = 0.5

SNR (dB)	20		25		30	
Method	GAN	CRNet	GAN	CRNet	GAN	CRNet
MSE	0.0329	**0.0241**	0.0304	**0.0193**	0.0325	**0.0145**
Cosine similarity	0.4659	**0.7828**	0.6588	**0.8315**	0.7511	**0.9107**
SSIM	0.4745	**0.5482**	0.5184	**0.6737**	0.5090	**0.7769**

TABLE II
COMPARISON OF SPECTRUM RECONSTRUCTION IN DIFFERENT COMPRESSED RATE WITH SNR = 30DB

Compressed rate	0.5		0.25		0.125	
Method	GAN	CRNet	GAN	CRNet	GAN	CRNet
MSE	0.0325	**0.0145**	0.0347	**0.0148**	0.0357	**0.0146**
Cosine similarity	0.7511	**0.9107**	0.3743	**0.9346**	0.2136	**0.9161**
SSIM	0.5090	**0.7769**	0.4506	**0.7565**	0.4204	**0.7535**

D. Performance of CRNet in different compression rate

Table II presents the achieved performance of different algorithms in different compression rates. One can notice that the reconstruction quality of GAN-based algorithm degenerates with the smaller compression rate. As an alternative, CRNet shows robustness in different compression rates as the credit of the structured sensing matrix design utilizing fewer under-sampled measurements to perform successful spectrum reconstruction.

E. Illustration of spectrum reconstruction quality of different algorithms

Finally, Fig. 3 illustrates some CRNet and GAN-based reconstruction results from testing samples. Obviously, even in the lower compression rate scenario, CRNet can still offer reconstruction result with high quality due to the joint optimization of compression and reconstruction. On the other hand, the GAN-based algorithm fails to perform valid reconstruction as it struggles to learn the pattern of signal with THz channel effect. It is also noteworthy that the number of trainable weights of GAN-based algorithm is 559,587 (the sum of generator and discriminator) and that is 246,499 in the proposed CRNet DL-based algorithm. In other words, the training overhead is reduced by about 56% when the superior performances in the above simulations are reported.

V. CONCLUSION

This paper proposes the CRNet for enhanced spectrum management for THz communications. From the literature review, this work is the first to consider the channel effect, especially THz channel properties, to truly evaluate the achieved

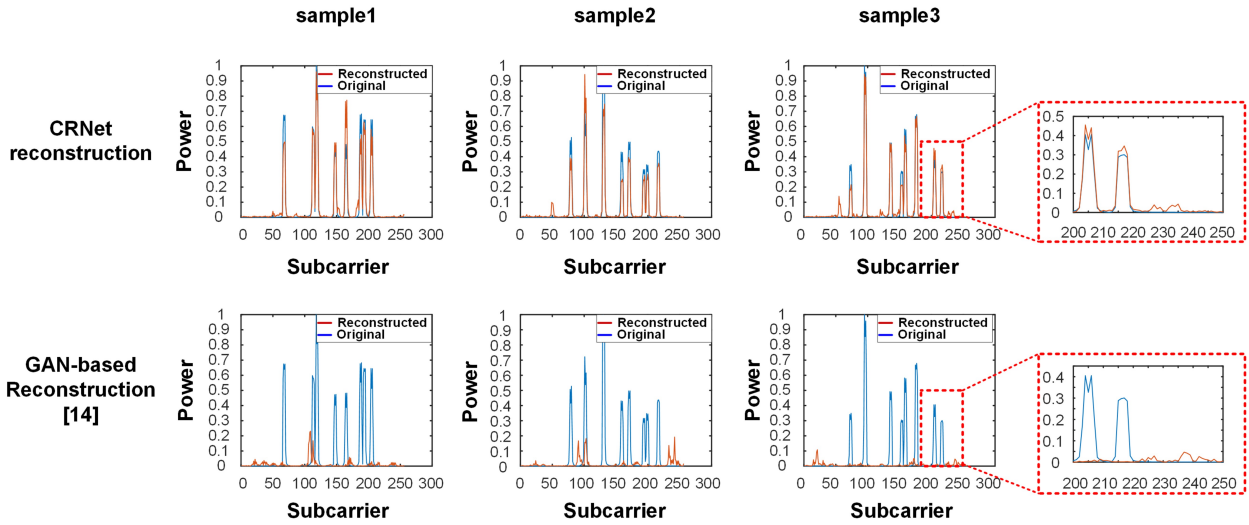

Fig. 3. The magnitude reconstruction results of different algorithms on testing samples with SNR = 30 dB and compression rate = 0.125. The first row: the reconstruction result of CRNet algorithm; the second row: the reconstruction result of GAN-based algorithm. As the THz channel effect, the reconstruction task is nontrivial. However, CRNet can still produce reconstructed spectrum, which is very close to the original one in such a low compression rate.

performance of different spectral sensing algorithms. CRNet efficiently and effectively performs joint design of compression and reconstruction in an end-to-end learning manner. Simulation results reveal that CRNet outperforms existing algorithms and can provide a realistic reconstruction result as the structured sensing matrix design and corresponding reconstruction module design. To be more specific, CRNet can offer superior performance with only 44% training overhead as compared to existing DL-based solutions. It is noteworthy that CRNet assumes that blind spectrum reconstruction is performed in this paper, which means we can only obtain information from the under-sampled measurements. In our future work, we aim to future consider the case when some of the additional information (e.g., user locations, channel statistics) are provided to further improve the spectrum reconstruction.

ACKNOWLEDGEMENT

This work was supported in part by the North Carolina Department of Transportation (NCDOT) under Award TCE2020-03 and in part by the AC21 Special Project Fund.

REFERENCES

[1] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5g evolution: A view on 5g cellular technology beyond 3gpp release 15," *IEEE Access*, vol. 7, pp. 127 639–127 651, 2019.

[2] C.-H. Lin, Y.-T. Lee, W.-H. Chung, S.-C. Lin, and T.-S. Lee, "Unsupervised resnet-inspired beamforming design using deep unfolding technique," in *IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2020.

[3] W. Xiong, J. Lu, X. Tian, G. Chen, K. Pham, and E. Blasch, "Cognitive radio testbed for digital beamforming of satellite communication," in *Cognitive Communications for Aerospace Applications Workshop (CCAA)*. IEEE, 2017, pp. 1–5.

[4] S.-C. Lin and I. F. Akyildiz, "Dynamic base station formation for solving nlos problem in 5g millimeter-wave communication," in *IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2017, pp. 1–9.

[5] G. Wang, K. Pham, E. Blasch, T. M. Nguyen, D. Shen, X. Tian, and G. Chen, "Cognitive radio unified spectral efficiency and energy efficiency trade-off analysis," in *IEEE Military Communications Conference (MILCOM)*. IEEE, 2015, pp. 244–249.

[6] S.-C. Lin and H. Narasimhan, "Towards software-defined massive mimo for 5g&b spectral-efficient networks," in *IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–6.

[7] C. Han and I. F. Akyildiz, "Distance-aware bandwidth-adaptive resource allocation for wireless systems in the terahertz band," *IEEE Transactions on Terahertz Science and Technology*, vol. 6, no. 4, pp. 541–553, 2016.

[8] S.-C. Lin and K.-C. Chen, "Cognitive and opportunistic relay for qos guarantees in machine-to-machine communications," *IEEE Transactions on Mobile Computing*, vol. 15, no. 3, pp. 599–609, 2015.

[9] ——, "Improving spectrum efficiency via in-network computations in cognitive radio sensor networks," *IEEE Transactions on wireless communications*, vol. 13, no. 3, pp. 1222–1234, 2014.

[10] Y. Fang, L. Li, Y. Li, H. Peng, and Y. Yang, "Low energy consumption compressed spectrum sensing based on channel energy reconstruction in cognitive radio network," *Sensors*, vol. 20, no. 5, p. 1264, 2020.

[11] J. Lu, L. Li, D. Shen, G. Chen, B. Jia, E. Blasch, and K. Pham, "Dynamic multi-arm bandit game based multi-agents spectrum sharing strategy design," in *IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. IEEE, 2017, pp. 1–6.

[12] M. F. Pervej and S.-C. Lin, "Eco-vehicular edge networks for connected transportation: A distributed multi-agent reinforcement learning approach," in *Proc. IEEE 92nd Vehicular Technology Conference (VTC-Fall)*, 2020.

[13] G. Wang, G. Chen, D. Shen, X. Tian, K. Pham, and E. Blasch, "Spread spectrum design for aeronautical communication system with radio frequency interference," in *IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*. IEEE, 2015, pp. 2F1–1.

[14] X. Meng, H. Inaltekin, and B. Krongold, "End-to-end deep learning-based compressive spectrum sensing in cognitive radio networks," in *IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.

[15] J. Ma, S.-C. Lin, H. Gao, and T. Qiu, "Automatic modulation classification under non-gaussian noise: A deep residual learning approach," in *IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.

[16] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmwave massive mimo systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, 2018.