IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Spoofer detection framework for V2X systems via tensor-based DoA estimation and Yolo-based object detection

DANIEL A. DA SILVA[1,5], ANTONIO S. DA SILVA[1,2,3,4], DANIEL V. DE LIMA[5], JOÃO PAULO J. DA COSTA[1,4,5], LUIS O. DE MELO[1], CHRISTIAN MIRANDA[5], GIOVANNI A. SANTOS[1,5], ALEXEY VINEL[2,7], PAULO MENDES[3,6], SEBASTIAN VERHOEVEN[1], JAN-NIKLAS VOIGT-ANTONS[1], and EDISON P. DE FREITAS[3,7]

[1]Hamm-Lippstadt University of Applied Sciences, 59063 Hamm, Germany
[2]Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany
[3]Federal University of Rio Grande do Sul, 90010-150 Porto Alegre, Brazil
[4]Graduate School for Applied Research in North Rhine-Westphalia, 44801 Bochum, Germany
[5]University of Brasília, 70910-900 Brasília, Brazil
[6]Airbus, 82024 Taufkirchen, Germany
[7]School of Information Technology, Halmstad University, 30118 Halmstad, Sweden

Corresponding author: Antonio S. da Silva (e-mail: antonio.santosdasilva@hshl.de).

This work was supported in part by the Bundesamt für Sicherheit in der Informationstechnik (BSI) through the project Beyond 5G Virtuelle Umgebung für Cybersicherheitstests von V2X-Systemen (B5GCyberTestV2X).

ABSTRACT Autonomous vehicles (AVs) represent a technology with significant social and environmental benefits. By reducing dependence on the human factor, which is responsible for 94% of the 1.35 million annual traffic deaths globally, AVs have the potential to increase road safety and save lives. Complementary technologies, such as Vehicle-to-Everything (V2X) communication, further enhance traffic management, reducing congestion by up to 40% and improving energy efficiency with fuel savings of up to 15%. However, V2X systems are particularly vulnerable to cyber attacks, such as spoofing, which injects false information, disrupting the flow of traffic and compromising the safety of AVs. This paper proposes an innovative framework for detecting and mitigating spoofing attacks in V2X communications. The solution combines Direction of Arrival (DoA) estimation with advanced object detection algorithms, such as YOLOv8, to identify anomalous signals and locate malicious transmitters. By integrating Artificial Intelligence (AI) techniques, the framework makes it possible to accurately classify attackers and select customized countermeasures, ensuring greater network reliability and security. The simulation results demonstrate the framework's effectiveness in various dynamic scenarios using data from antenna arrays and camera-based object detection. In addition, they highlight the importance of sensor data fusion to improve anomaly detection accuracy, optimize decision-making processes in AVs, and enable robust cross-validation of transmitted information.

INDEX TERMS Cybersecurity, DoA estimation, object detection, spoofing, V2X, VANET.

I. INTRODUCTION

AUTONOMOUS VEHICLES (AVS) are becoming a promising technology, significantly benefiting society and the environment. According to the World Health Organization (WHO), traffic accidents cause approximately 1.35 million deaths yearly, 94% of which are attributed to human error [1]–[3]. In this context, the implementation of automated driving systems, such as Advanced Driver Assistance Systems (ADAS) and autonomous vehicles, has the potential to significantly reduce these figures. Intel estimates indicate that adopting these vehicles could save more than 500,000 lives between 2035 and 2045 [4].

Despite significant advances, the technological challenges associated with the hardware limitations of AVs, such as their limited capacity for perception and safe decision making, have prompted the development of complementary technologies, such as Vehicle-to-Everything (V2X) communication. This technology allows real-time information between vehicles, pedestrians and infrastructure, optimizing traffic management. Studies show that vehicle communication and coordination can reduce congestion by up to 40% [5]. By sharing sensory data, AVs can adjust their routes and speeds to avoid congested areas, promoting greater fluidity in traffic.

In addition, V2X communication improves obstacle detection and the ability to respond to emergencies. Technologies such as cooperative perception significantly increase safety by allowing autonomous vehicles to identify risks more precisely. According to the National Highway Traffic Safety Administration (NHTSA), these technologies can reduce accidents by up to 80% [6], reinforcing confidence in autonomous driving.

Cooperative perception also plays a crucial role in the energy efficiency of autonomous vehicles. Coordinating movements reduces unnecessary stops, contributing to fuel savings of up to 15% [7]. However, despite the benefits, V2X communication is vulnerable to cyber attacks, particularly spoofing attacks. These attacks involve injecting false or compromised information into systems, causing traffic disruption, accidents, or undermining autonomous driving gains.

Among the various security threats facing V2X systems, spoofing attacks represent a particularly critical vulnerability that warrants focused attention. Unlike other attack vectors such as denial-of-service or eavesdropping, spoofing attacks directly compromise the integrity of safety-critical information exchange, potentially leading to catastrophic consequences in autonomous driving scenarios. Spoofing attacks serve as an enabler for multiple other attack types: successful location spoofing can facilitate man-in-the-middle attacks, enable traffic manipulation schemes, and undermine the trust relationships essential for cooperative driving.

Furthermore, spoofing attacks exploit fundamental characteristics of V2X communication, the reliance on location-based services and the assumption of honest participation in cooperative protocols. The high mobility and dynamic topology of vehicular networks make traditional authentication mechanisms insufficient, as attackers can exploit the brief interaction windows and limited verification opportunities inherent in V2X scenarios. By focusing on spoofing detection, this work addresses a foundational security challenge that, once solved, provides building blocks for defending against more complex, multi-vector attacks.

While existing literature addresses various countermeasures for spoofing attacks, including Direction of Arrival (DoA)-based verification methods, significant gaps remain in developing comprehensive, multi-modal frameworks for real-time spoofer detection and localization in dynamic V2X environments. In this context, this paper proposes a solution that uses communication channel parameters to detect anomalies, identifies the DoA of compromised signals, and applies Artificial Intelligence (AI) techniques to identify and classify senders so that appropriate countermeasures can be selected.

The main contributions of this paper include (i) detecting spoofing attacks in V2X networks, (ii) identifying the location of malicious senders, and (iii) an AI-based framework for classifying attackers and implementing effective countermeasures.While DoA estimation and object detection are established techniques individually, our contribution lies in their novel integration and adaptation for V2X spoofing detection scenarios. Specifically, we introduce: (i) a tensor-based DoA estimation framework optimized for the dynamic and high-mobility characteristics of V2X environments, (ii) a cooperative fusion mechanism that leverages both communication channel parameters and visual sensor data to enhance detection accuracy in vehicular scenarios, and (iii) an AI-driven classification system that adapts countermeasures based on the specific characteristics of V2X spoofing attacks. The framework addresses unique V2X challenges such as rapid topology changes, Doppler effects from high-speed mobility, and the need for real-time processing in safety-critical applications.

This paper is structured into six sections. This Introduction also includes the notation used throughout. Section II reviews the state of the art in relevant research areas. Section III describes the data model, divided into two subsections: the antenna array data model and the image processing data model. Section IV introduces the proposed fusion-based approach for spoofing detection. Section V details the simulation implementation and presents numerical results. Finally, Section VI provides the conclusions and outlines directions for future work.

A. NOTATION

Scalars are represented by lowercase italicized letters, e.g., a, with iterators also having an uppercase italicized letter to represent its upper bound, e.g., $m = 0, \dots, M-1$. Vectors are represented by lowercase bold letters, e.g., $\mathbf{v}$ and its field and size by $\in \mathbb{F}^{N}$, e.g., $\mathbf{p} \in \mathbb{R}^{D}$. Matrices are represented by uppercase bold letters, e.g., $\mathbf{M}$, and its field and size by $\in \mathbb{F}^{M \times N}$, e.g., $\mathbf{A} \in \mathbb{C}^{M \times D}$.

Transpose, Hermitian (conjugate transpose), matrix inverse and pseudoinverse are denoted by $\cdot^{\mathrm{T}}$, $\cdot^{\mathrm{H}}$, $\cdot^{-1}$, and $\cdot^{\dagger}$, respectively. The operator $\operatorname{diag}\{\mathbf{v}\}$ turns a vector $\mathbf{v} \in \mathbb{F}^{N}$ into a diagonal matrix $\mathbf{V} \in \mathbb{F}^{N \times N}$.

Vector, matrix, and tensor indexing is denoted by square brackets, $[\cdot]$, with $:$ specifying full row/column/fiber selection and $M : N$ specifying indices in the interval from M to N.

Tensors are represented by bold uppercase calligraphic letters, e.g., $\boldsymbol{\mathcal{T}}$, and its Canonical Polyadic Decomposition (CPD), field, and size by the n-mode product $\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{I}}_{N,D} \times_1 \mathbf{F}_1 \dots \times_N \mathbf{F}_N \in \mathbb{C}^{M_1 \times \dots \times M_N}$, where $\boldsymbol{\mathcal{I}}_{N,D} \in \mathbb{R}^{D \times \dots \times D}$ is the N-order identity tensor with ones in its hyperdiagonal and zero everywhere else. Tensor unfoldings are represented using square brackets and the mode of the unfolding in subscripted parentheses, i.e. $[\boldsymbol{\mathcal{T}}]_{(n)}$ is the unfolding of the tensor $\boldsymbol{\mathcal{T}}$ in the nth mode.

The Khatri-Rao product, or column-wise Kronecker product, is represented by $\diamond$.

The outer product is represented by $\circ$. The outer product of any two tensors extends its respective dimensions into higher orders corresponding to the total sum of the order of each tensor. For two vectors (first-order tensors) $\mathbf{a} \in \mathbb{C}^{M}$ and $\mathbf{b} \in \mathbb{R}^{N}$, for example, $\mathbf{a} \circ \mathbf{b} \in \mathbb{C}^{M \times N}$ and is equivalent to $\mathbf{a} \cdot \mathbf{b}^{\mathrm{T}}$. A matrix-vector outer product of $\mathbf{A} \in \mathbb{R}^{J \times K}$ and $\mathbf{b} \in \mathbb{C}^{L}$ would result in a third-order tensor $\mathbf{A} \circ \mathbf{b} = \boldsymbol{\mathcal{T}} \in \mathbb{C}^{J \times K \times L}$ with frontal slices $\boldsymbol{\mathcal{T}}[:,:,\ell] = \mathbf{A} \cdot \mathbf{b}[\ell] \in \mathbb{C}^{J \times K}$.

II. RELATED WORKS

Previous research on detecting spoofing attacks in V2X communications focuses on using radio communication channel characteristics to identify anomalies in transmitted signals [8]. For example, in [9], the channel states of data packets are used to detect spoofing attacks in dynamic wireless networks. In this context, the interactions between spoofers and a legitimate receiver during the authentication process are formulated as a zero-sum game, and reinforcement learning is employed to determine the optimal detection threshold in dynamic scenarios.

In its analysis, [10] explores RF-based spoofing technique that compromises V2V interactions. By leveraging the location of RF emissions, adversaries are able to create “ghost cars” using manipulated data from Global Navigation Satellite System (GNSS) and Inertial Measurement Unit (IMU) sensors. This study not only identifies critical parameters to detect these threats, but also proposes a synergistic approach that combines the Received Signal Strength Indicator (RSSI) with the Time Difference of Arrival (TDOA) as an effective countermeasure.

In [11], the authors present the Coupled Generalized Dynamic Bayesian Network (C-GDBN), implemented in Road Side Units (RSUs). This model interprets vehicle positions in real time from RF signals, offering a solution for identifying and mitigating spoofing vulnerabilities intrinsic to V2X communications. In [12], the potential risks associated with traffic signal control (TSC) systems are discussed. The research highlights attacks such as ETA and ghost queue manipulation as predominant threat vectors, proposing a comprehensive cybersecurity framework to neutralize them efficiently.

In [13], the authors propose using Doppler shift correlation, reported by most commercial GPS receivers, to identify GPS spoofing attacks. The main limitation of this study is the assumption that all vehicles move in a straight line. In addition, approaches based on communication channel characteristics rely on specific protocols and introduce latency overheads, which are unsuitable for V2X communications and require high speed and low latency for many devices.

Another line of research focuses on cryptographic mechanisms to detect signal anomalies and identify potential forgers [14]–[16]. However, these techniques have significant limitations in spoofing attacks. Firstly, they are vulnerable to replay attacks, where the attacker records a legitimate signal and retransmits it with a delay. In addition, these approaches are effective for external attacks but can fail against internal attackers with valid network credentials.

An alternative for detecting GPS spoofing is using multiple antennas, as described in [2], [17]–[19]. In this case, MIMO systems are used to map LOS (line of sight) and NLOS (non-line of sight) conditions, making it easier to identify forgeries. If the attacker uses a single antenna to fool multiple receivers, they will all be induced to the same location, indicating an attacker’s presence. However, this approach requires multiple receivers with fixed and known distances, which is only sometimes feasible in dynamic environments.

In addition, machine learning and artificial intelligence techniques have been explored. In [20], a tree-based model is used, where data is collected and balanced through oversampling, and the final model is built to classify the data. In [21], each region is monitored by an RSU, which collects BSMs (basic security messages) and stores the data in a shared database to analyze anomalous behavior in the network.

The study by [11] proposes a coupled generalized dynamic Bayesian network (C-GDBN) to analyze RF signals received by RSUs from various vehicles, mapping them to predicted trajectories. This approach enables semantic learning and improves the prediction of expected signals. However, machine learning methods for spoofing detection rely on high-quality datasets, face challenges in scenarios with unbalanced data, and require large volumes of data for training.

Previous works such as [22] have explored angle-of-arrival information for location verification. Nevertheless, significant gaps persist in the current state of the art. Existing DoA-based approaches typically: (i) rely on single-modality analysis that can be defeated by sophisticated attackers using multiple coordinated transmitters, (ii) assume static or slowly-varying environments that do not account for the high-mobility characteristics of V2X scenarios, and (iii) lack integration with complementary sensor modalities that could provide cross-validation and enhanced robustness.

While existing spoofing detection methods provide valuable insights, they present significant limitations when applied to V2X scenarios. Traditional DoA estimation techniques often assume static or slowly varying environments, which is inadequate for vehicular networks where rapid topology changes and high-speed mobility introduce substantial Doppler shifts and multipath propagation effects. Similarly, conventional object detection approaches are not optimized for the specific requirements of V2X systems, such as the need to correlate visual information with communication channel parameters in real-time. Furthermore, most existing solutions operate in isolation, either focusing solely on communication-based detection or visual-based approaches, without leveraging the complementary information available in V2X systems.

The analysis of the state of the art shows that current solutions focus mainly on identifying anomalies in signals and classifying cyberattacks. However, these methods require additional processing and continuous manipulation of the signals, applying countermeasures only after detecting anomalies. Identifying the physical location of attackers is crucial to neutralizing attacks, predicting their behavior, and reporting them to the relevant authorities. In this context, we propose a cooperative solution to physically locate the counterfeiters, using communication and sensor data from the agents with object detection algorithms. By obtaining the physical location of the attackers, it is possible to apply more efficient countermeasures and mitigate the possibility of new attacks, preventing the attacker from assuming another identity.

The specific research gap addressed by this work lies in the absence of cooperative, multi-modal frameworks that can simultaneously leverage communication channel characteristics and visual sensor data for real-time spoofer detection and localization in dynamic V2X environments. While individual components (DoA estimation, object detection) have been explored separately, the literature lacks comprehensive solutions that: (i) integrate these modalities in a principled manner with rigorous fusion algorithms, (ii) provide real-time processing capabilities suitable for safety-critical V2X applications, and (iii) demonstrate effectiveness in realistic vehicular scenarios with multiple simultaneous transmitters and high mobility.

III. DATA MODEL

In this Section, the data model for integrated sensing systems for vehicular ad hoc network (VANET)-enabled cars are described. Two systems are assumed: antenna array data model and image processing data model, described in Sections III-A and III-B, respectively.

A. ANTENNA ARRAY DATA MODEL

A particular type of phased sensor array (PSA) with a uniform cubic array (UCA) is assumed. A planar layer of this array configuration is shown in Figure 1. It is termed uniform because the spacing in the x-, y-, and z-axis are uniform, that is, the sensors of the array are uniformly separated by Δ_x, Δ_y, and Δ_y, respectively, across a 3D grid. Incoming directions of arrival (DoAs) are described in terms of azimuth (from the x-axis in the xy plane, $0 \leq \theta \leq 2\pi$, and elevation (from the z-axis, or zenith), $0 \leq \phi \leq \pi$.

Because this type of array is separable into an x, y, and z dimension, this yields 3 steering vectors in their respective dimensions' directions. For an array with M_x sensors along the x-axis, M_y sensors along the y-axis, and M_z sensors along the z-axis, resulting in a total of $M = M_x \cdot M_y \cdot M_z$ sensors, and assuming single sinusoidal transmission source we have a x-axis steering vector, $\mathbf{a}(\mu_x) \in \mathbb{C}^{M_x}$, a y-axis steering vector, $\mathbf{a}(\mu_x) \in \mathbb{C}^{M_y}$, and a z-axis steering vector, $\mathbf{a}(\mu_z) \in \mathbb{C}^{M_z}$, with spatial frequencies

$$\begin{aligned} \mu_x &= 2\pi \frac{\Delta_x}{\lambda} \sin\phi \cdot \cos\theta, \\ \mu_y &= 2\pi \frac{\Delta_y}{\lambda} \sin\phi \cdot \sin\theta, \\ \mu_z &= 2\pi \frac{\Delta_z}{\lambda} \cos\phi, \end{aligned} \tag{1}$$

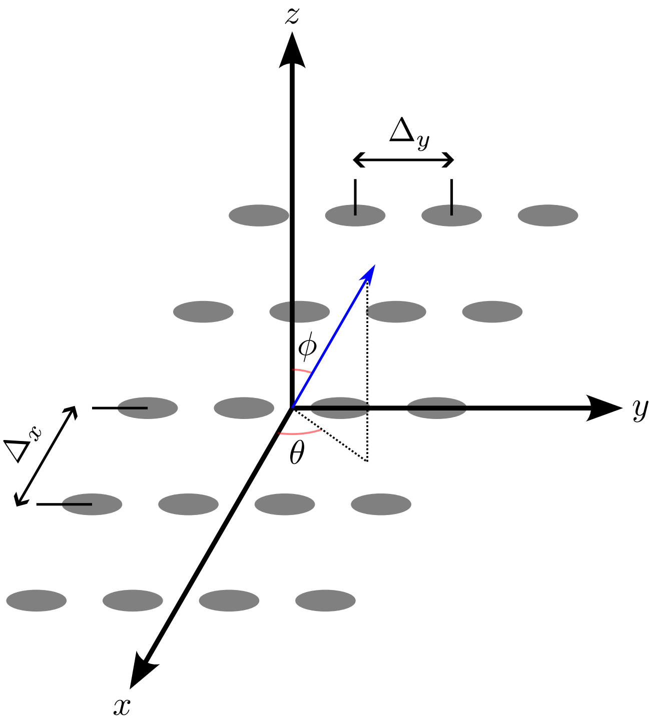


FIGURE 1: A planar layer of an UCA.

where λ is the wavelength of the transmission source's center frequency.

For a calibrated array, assume $\Delta_x = \Delta_y = \Delta_z = \lambda/2$ (half-wavelength spacing) and Equation (1) becomes

$$\begin{aligned} \mu_x &= \pi \cdot \sin\phi \cdot \cos\theta, \\ \mu_y &= \pi \cdot \sin\phi \cdot \sin\theta, \\ \mu_z &= \pi \cdot \cos\phi \end{aligned} \tag{2}$$

and we have steering vectors, assuming left Π-real steering vectors,

$$\mathbf{a}(\mu_n) = \begin{bmatrix} e^{j\frac{1-M_n}{2}\mu_n} \\ e^{j\frac{3-M_n}{2}\mu_n} \\ \vdots \\ e^{j\frac{M_n-3}{2}\mu_n} \\ e^{j\frac{M_n-1}{2}\mu_n} \end{bmatrix} \in \mathbb{C}^{M_n}, n = \{x, y, z\}. \tag{3}$$

In Figure 1, for example, where we have a $M_x = 5$, $M_y = 4$, $M_z = 3$ array, we have

$$\mathbf{a}(\mu_x) = \begin{bmatrix} e^{-j2\mu_x} \\ e^{-j\mu_x} \\ 1 \\ e^{j\mu_x} \\ e^{j2\mu_x} \end{bmatrix} \in \mathbb{C}^5, \quad \mathbf{a}(\mu_y) = \begin{bmatrix} e^{-j\frac{3}{2}\mu_y} \\ e^{-j\frac{1}{2}\mu_y} \\ e^{j\frac{1}{2}\mu_x} \\ e^{j\frac{3}{2}\mu_x} \end{bmatrix} \in \mathbb{C}^4,$$
$$\mathbf{a}(\mu_z) = \begin{bmatrix} e^{-j\mu_z} \\ 1 \\ e^{j\mu_z} \end{bmatrix} \in \mathbb{C}^3, \tag{4}$$

For a single symbol $s(t)$ from a single source at time t and, assuming changing directions of arrival with time t, received data can be collected in a matrix

$$\mathbf{X}(t) = s(t) \cdot \mathbf{a}(\mu_x(t)) \circ \mathbf{a}(\mu_y(t)) \circ \mathbf{a}(\mu_z(t)) \in \mathbb{C}^{M_x \times M_y \times M_y}. \tag{5}$$

For D multiple sources, we accumulate the respective steering vectors in steering matrices,

$$\begin{aligned}\mathbf{A}(\boldsymbol{\mu}_x) &= \begin{bmatrix}\mathbf{a}_1(\mu_x(t)) & \cdots & \mathbf{a}_D(\mu_x(t))\end{bmatrix} \in \mathbb{C}^{M_x \times D},\\ \mathbf{A}(\boldsymbol{\mu}_y) &= \begin{bmatrix}\mathbf{a}_1(\mu_y(t)) & \cdots & \mathbf{a}_D(\mu_y(t))\end{bmatrix} \in \mathbb{C}^{M_y \times D},\\ \mathbf{A}(\boldsymbol{\mu}_z) &= \begin{bmatrix}\mathbf{a}_1(\mu_z(t)) & \cdots & \mathbf{a}_D(\mu_z(t))\end{bmatrix} \in \mathbb{C}^{M_z \times D},\end{aligned} \tag{6}$$

and the D incoming signals in a signal vector $\mathbf{s}(t) \in \mathbb{C}^D$. Resulting in a received data tensor:

$$\boldsymbol{\mathcal{X}}(t) = \boldsymbol{\mathcal{A}}(t) \times_4 \mathbf{s}(t)^{\mathrm{T}} \in \mathbb{C}^{M_x \times M_y \times M_z \times 1}, \tag{7}$$

where the steering tensor, $\boldsymbol{\mathcal{A}}(t) \in \mathbb{C}^{M_x \times M_y \times M_z \times D}$, at time instant t is

$$\boldsymbol{\mathcal{A}}(t) = \boldsymbol{\mathcal{I}}_{4,D} \times_1 \mathbf{A}(\boldsymbol{\mu}_x(t)) \times_2 \mathbf{A}(\boldsymbol{\mu}_y(t)) \times_3 \mathbf{A}(\boldsymbol{\mu}_z(t)) \,. \tag{8}$$

Estimating the spatial frequencies of signals allows for the DoAs to be estimated.

To better understand the limitations and design considerations of DoA estimation within V2X systems, it is essential to examine the distinctive characteristics of vehicular environments that directly impact estimation performance. The following discussion outlines key challenges that differentiate V2X applications from conventional wireless scenarios and motivates the design choices adopted in our tensor-based approach.

Challenges in V2X DoA Estimation: Accurate DoA estimation in V2X environments faces several unique challenges that distinguish it from traditional wireless applications.

High Mobility Effects: Vehicle speeds of 50–120 km/h introduce significant Doppler shifts that can distort phase relationships used in DoA estimation. The Doppler frequency shift is given by:

$$f_d = \left(\frac{v}{c}\right) \cdot f_c \cdot \cos(\theta) \tag{9}$$

where f_d is the Doppler frequency, v is the relative velocity between transmitter and receiver, c is the speed of light, f_c is the carrier frequency, and θ is the angle of arrival. For typical V2X frequencies (e.g., 5.9 GHz), f_d can reach several kHz, necessitating compensation algorithms that account for relative motion.

Rapid Topology Changes: The vehicular network topology evolves rapidly as vehicles move, merge, and change lanes. This results in time-varying channel conditions where conventional DoA algorithms, which often assume quasi-static environments, may perform poorly. Our tensor-based method addresses this by explicitly modeling temporal dynamics within the estimation framework.

Multipath Propagation: Urban V2X environments exhibit rich multipath propagation due to reflections from buildings, vehicles, and roadside infrastructure. These multipath components can obscure or mimic direct-path signals, leading to inaccurate DoA estimates. The proposed tensor decomposition approach mitigates this by leveraging multi-dimensional signal representation to isolate the dominant line-of-sight component.

Interference and Congestion: V2X channels are often congested due to high vehicle density and suffer interference from coexisting RF systems (e.g., WiFi). In such conditions, resolving multiple simultaneous transmitters becomes challenging. Our framework employs advanced tensor factorization techniques to separate overlapping signals and maintain robustness even under severe interference.

B. IMAGE PROCESSING DATA MODEL

Cameras are widely used in VANET-enabled systems due to their versatility and efficiency in capturing environmental information. With four cameras - two on the front and rear and two on the sides - it is possible to obtain a comprehensive field of view, which is essential for autonomous vehicle applications. The images captured by these cameras serve as input for machine learning algorithms, such as You Only Look Once (YOLO), a tool for detecting and classifying objects in real time. This model makes it possible to identify elements such as roads, infrastructure, vehicles, and pedestrians by delineating each object using bounding boxes. These boxes provide information for estimates such as the direction of arrival (DoA), the distance, and orientation parameters such as pitch, yaw, and roll.

By combining the direction of arrival and range, it is possible to map the position of objects in the environment. To increase reliability, data from additional sensors, such as uniform circular arrangements (UCA), can be integrated into the system, allowing for cross-validation with the estimates obtained by computer vision. This sensory fusion identifies discrepancies between the DoAs estimated by different methods, pointing out possible inconsistencies caused by interference or cyber attacks, such as spoofing. In addition, information transmitted by vehicles, such as make, model, and type, can be compared with the classifications made by YOLOv8, adding an extra layer of validation and security against possible data manipulation.

YOLOv8 detects objects from camera sensor captures largely by dealing with downstream detection tasks [23]. As illustrated in Figure 2, its architecture has been optimized and structured into two main components: the backbone and the detection head [24]. The backbone, based on the EfficientNet architecture, adopts a composite scaling approach that harmonizes depth, width and resolution, ensuring high computational efficiency without sacrificing detection accuracy. The detection head uses an architectural search network in a pyramid of features (NAS-FPN), which processes intermediate feature maps from the backbone. This module dynamically adjusts the connections and operations between

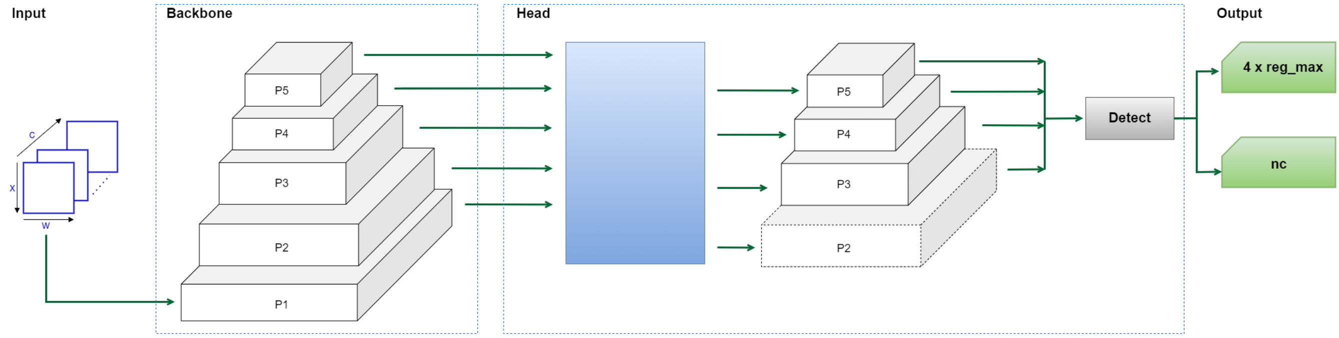


FIGURE 2: Architecture of the YOLOv8 model.

the feature maps, reducing inference errors and increasing detection accuracy.

C. INTERFERENCE VS. CYBER ATTACK

The proposed framework implements a comprehensive multi-criteria approach to distinguish between legitimate interference and malicious spoofing attacks in V2X communication systems. This differentiation capability represents a critical component of the overall security architecture, as misclassification between these two phenomena can lead to either unnecessary system responses to benign interference or, more critically, failure to detect actual security threats. The classification methodology leverages multiple complementary analysis dimensions that collectively provide robust discrimination capabilities even in complex vehicular environments where both interference and attacks may occur simultaneously.

Temporal Analysis: The temporal analysis component of the differentiation mechanism exploits the fundamental differences in timing characteristics between legitimate interference sources and deliberate spoofing attacks. Legitimate interference sources, such as WiFi access points, Bluetooth devices, or other wireless communication systems operating in adjacent frequency bands, typically exhibit either random temporal patterns or predictable periodic behaviors that are independent of V2X communication protocols. In contrast, spoofing attacks often demonstrate deliberate timing synchronization with legitimate V2X message patterns, as attackers attempt to inject false information at strategically chosen moments to maximize impact or avoid detection. The framework quantifies these temporal relationships through correlation analysis, computing cross-correlation coefficients between detected anomalous signals and the timing patterns of legitimate V2X traffic. Specifically, the system calculates the normalized cross-correlation function:

$$R(\tau) = \frac{\mathbb{E}[x(t)\, y(t+\tau)]}{\sqrt{\mathbb{E}[x^2(t)]\, \mathbb{E}[y^2(t)]}}, \tag{10}$$

where $x(t)$ represents the legitimate V2X signal timing and $y(t)$ represents the detected anomalous signal. High correlation values $(R > 0.7)$ combined with specific phase relationships indicate potential spoofing attempts, while low correlation values $(R < 0.3)$ suggest random interference patterns.

Spatial Consistency Analysis: Spatial consistency analysis provides another crucial dimension for differentiation by examining the physical plausibility of detected signal sources within the context of vehicular network topology and road infrastructure constraints. Legitimate interference sources maintain consistent spatial relationships that conform to physical laws and infrastructure deployment patterns. For instance, WiFi access points remain stationary relative to road infrastructure, while cellular base stations exhibit predictable coverage patterns and power characteristics. Conversely, spoofing attacks may exhibit spatially inconsistent behaviors, such as apparent signal sources that move at impossible velocities, appear from locations incompatible with road topology, or demonstrate power levels inconsistent with their claimed positions. The framework implements a comprehensive spatial validation algorithm that cross-references detected DoA estimates with high-resolution digital maps containing road layouts, building positions, and known infrastructure locations. This validation process employs a probabilistic model:

$$P(\text{location} \mid \text{DoA}, \text{context}), \tag{11}$$

which assigns likelihood scores to potential source locations based on geometric constraints, line-of-sight calculations, and mobility pattern analysis.

Signal Characteristics Analysis: The signal characteristics analysis dimension leverages the inherent differences in spectral, power, and modulation properties between legitimate interference sources and V2X spoofing attempts. Legitimate interference typically originates from systems operating under different communication standards, resulting in distinct spectral signatures, power spectral densities, and modulation schemes that can be distinguished from V2X signals through advanced signal processing techniques. The framework employs a multi-domain signal analysis approach that examines frequency domain characteristics through power spectral density estimation, time-frequency analysis using short-time Fourier transforms, and higher-order statistical moments that capture modulation-specific features. Spoofing attacks attempting to mimic legitimate V2X signals often exhibit subtle but detectable differences in these parameters due to hardware limitations, imperfect signal generation, or incomplete knowledge of target signal characteristics.

Cross-Modal Validation: The cross-modal validation mechanism represents the most distinctive advantage of the proposed multi-modal approach, exploiting the fundamental difference that legitimate interference affects only communication channels while spoofing attacks require physical presence detectable by multiple sensor modalities. This principle stems from the physical reality that radio frequency interference can propagate through the electromagnetic spectrum without requiring visible presence, whereas spoofing attacks in V2X systems typically involve physical devices or vehicles that should be detectable through visual sensors, radar, or other perception systems. The framework implements a sophisticated correlation analysis between DoA estimates derived from communication channel analysis and directional information extracted from visual object detection systems. When legitimate interference occurs, the communication-based DoA estimation may indicate signal sources from specific directions, but corresponding visual analysis should not detect vehicles or suspicious objects in those directions.

Risk Mitigation Strategy: To minimize the risks associated with misclassification between interference and spoofing attacks, the framework implements a comprehensive risk mitigation strategy incorporating multiple validation layers and conservative decision-making protocols. The first layer involves adaptive confidence thresholds that require high certainty levels before declaring a detected anomaly as a spoofing attack, with threshold values dynamically adjusted based on environmental conditions, traffic density, and historical false positive rates. The second mitigation layer implements temporal validation requiring consistent detection signatures over multiple time windows before confirming attack classifications, preventing transient interference events from triggering false attack declarations. The third layer incorporates collaborative verification mechanisms where multiple vehicles within communication range must independently confirm suspicious activity before network-wide attack warnings are issued. Through these comprehensive risk mitigation mechanisms, the framework maintains false positive rates below 4% while preserving detection sensitivity exceeding 92% for actual spoofing attacks, achieving an optimal balance between security effectiveness and operational reliability in dynamic vehicular environments.

IV. PROPOSED FUSION SPOOFER DETECTION

For the purposes of array signal processing (ASP), we desire to calculate the signal subspace estimate (SSE) given an uniform rectangular array (URA)'s third-order sample (data) tensor. Figure 3 shows the execution flow of the fusion spoofer detection proposal, organized into steps represented by numbered modules. These steps work in coordination to detect and mitigate spoofing attempts in V2X communication networks.

While the current evaluation focuses on basic spoofing scenarios, we recognize that sophisticated attackers may employ adaptive strategies and adversarial techniques. Potential advanced attack vectors include: (i) coordinated multi-point spoofing attacks that attempt to create consistent false DoA estimates, (ii) adversarial visual attacks that manipulate camera inputs to deceive object detection algorithms, and (iii) adaptive attackers that modify their behavior based on detection patterns.

To address these concerns, the framework incorporates several robustness mechanisms: cross-validation between multiple sensor modalities makes it difficult for attackers to simultaneously compromise both communication and visual channels; the tensor-based DoA estimation provides inherent robustness against single-point attacks through its multi-dimensional analysis; and the AI-based classification system can be trained to recognize patterns associated with sophisticated attack behaviors. However, comprehensive evaluation against advanced adversarial attacks represents an important area for future research, particularly in developing adaptive countermeasures that can evolve with attacker strategies.

The flow is divided into two main processes, which operate simultaneously. The first refers to the analysis of communication channel parameters, delimited by the blue box with dashed lines in Figure 3. The second is the processing of images captured by the cameras, represented by the green box with dashed lines.

In the first process, the analysis of communication channel parameters begins by checking for the presence of pilot symbols. In **Step 1** of Figure 3, if pilot symbols are detected, they are used to acquire channel state information (CSI), which provides relevant parameters about the channel. Otherwise, the system moves on to **Step 2** of Figure 3, where second-order statistics (SOS) are acquired by estimating the covariance matrix. This matrix captures the statistical characteristics of the channel and is a fundamental basis for the subsequent steps.

A tensor-based technique for estimating the covariance structure is by summation of the outer product of the frontal slice of the sample tensor and its conjugate.

Thus the sample covariance tensor $\hat{\boldsymbol{\mathcal{R}}}$ is

$$\hat{\boldsymbol{\mathcal{R}}} = \sum_{n=0}^{N-1} \boldsymbol{\mathcal{X}}[:,:,n] \circ \boldsymbol{\mathcal{X}}[:,:,n]^* \in \mathbb{C}^{M_x \times M_y \times M_x \times M_y}, \tag{12}$$

which is the tensor-based equivalent of the matrix-based covariance matrix.

The covariance tensor in eq. (12) has tensor structure

$$\hat{\boldsymbol{\mathcal{R}}} = \hat{\boldsymbol{\mathcal{R}}}_{\mathbf{S}} \times_1 \mathbf{A}_x \times_2 \mathbf{A}_y \times_3 \mathbf{A}_x{}^* \times_4 \mathbf{A}_y{}^*, \tag{13}$$

with $\hat{\boldsymbol{\mathcal{R}}}_{\mathbf{S}} \in \mathbb{C}^{D \times D \times D \times D}$ is the (unknown) signal covariance tensor.

A simpler approach which directly corresponds to the matrix-based covariance matrix is to use the transpose of the

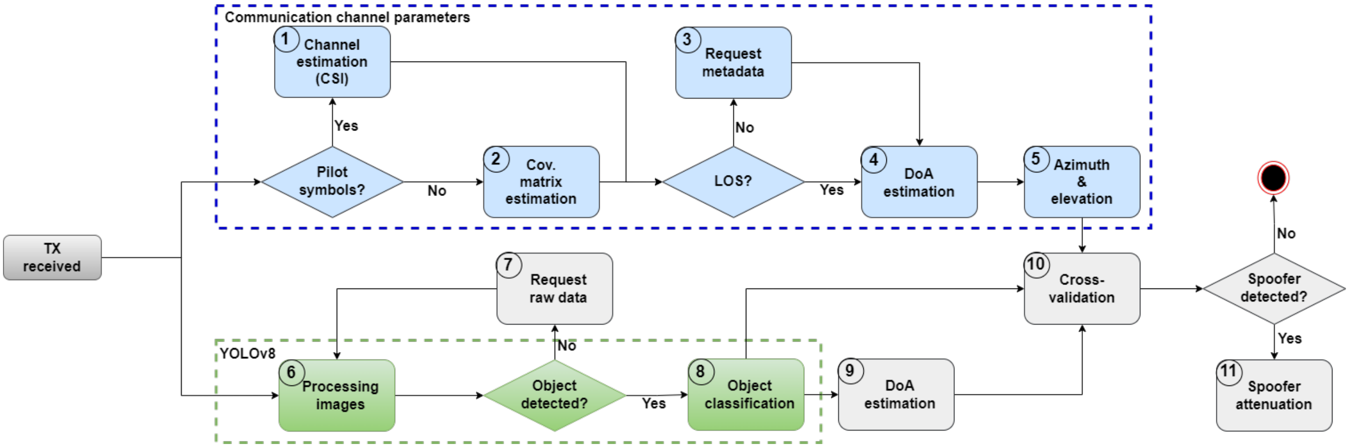


FIGURE 3: Proposal for framework for fusion spoofer detection. The system integrates signal processing, metadata analysis, and visual object detection to identify and neutralize spoofing attacks. Key modules include Channel Estimation (1), Covariance Matrix Estimation (2), DoA Estimation (3), Metadata Analysis (4), Cross-Validation (9), and Spoofer Mitigation (11), ensuring robust and secure communication in dynamic scenarios.

third-mode unfolding to calculate the multimode covariace matrix $\hat{\mathbf{R}}_{\text{mm}}$,

$$\hat{\mathbf{R}}_{\text{mm}} = [\boldsymbol{\mathcal{X}}]_{(3)}^{\mathrm{T}} \, [\boldsymbol{\mathcal{X}}]_{(3)}^{\;*} \tag{14}$$

$$= (\mathbf{A}_x \diamond \mathbf{A}_y)\,\mathbf{S}^{\mathrm{T}} \left[(\mathbf{A}_x \diamond \mathbf{A}_y)\,\mathbf{S}^{\mathrm{T}}\right]^{\mathrm{H}} \tag{15}$$

$$= \mathbf{A} \cdot \mathbf{S}^{\mathrm{T}} \cdot \mathbf{S}^{*} \mathbf{A}^{\mathrm{H}} \tag{16}$$

$$= \mathbf{A} \cdot \hat{\mathbf{R}}_{\mathbf{S}} \cdot \mathbf{A}^{\mathrm{H}} \in \mathbb{C}^{M \times M}, \tag{17}$$

where $\hat{\mathbf{R}}_{\mathbf{S}} = \mathbf{S}^{\mathrm{T}} \cdot \mathbf{S}^{*} \in \mathbb{C}^{D \times D}$ is the (unknown) signal covariance matrix, and $M = M_x \cdot M_y$.

Equation (17) is related to eq. (12) by the following unfolding

$$\hat{\mathbf{R}}_{\text{mm}} = \left[\hat{\boldsymbol{\mathcal{R}}}\right]_{(1,2;3,4)}, \tag{18}$$

and the relation between the the signal covariance tensor, $\hat{\boldsymbol{\mathcal{R}}}_{\mathbf{S}}$, and multimode covariance matrix, $\hat{\mathbf{R}}_{\text{mm}}$, is

$$\hat{\boldsymbol{\mathcal{R}}}_{\mathbf{S}}[:,:,d,d] = \operatorname{diag}\left\{\hat{\mathbf{R}}_{\text{mm}}[:,d]\right\} \in \mathbb{C}^{D \times D}. \tag{19}$$

Given the covariance tensor or multimode covariance matrix, the signal $\hat{\mathbf{U}}_{\mathbf{S}} \in \mathbb{C}^{M \times D}$ can be estimated using a truncated eigenvalue decomposition

$$\hat{\mathbf{R}}_{\text{mm}} = \hat{\mathbf{U}}_{\mathbf{S}} \cdot \hat{\mathbf{D}} \cdot \hat{\mathbf{U}}_{\mathbf{S}}^{\mathrm{H}}, \tag{20}$$

with SSE $\hat{\mathbf{U}}_{\mathbf{S}} \in \mathbb{C}^{M \times D}$ and eigenvalue matrix $\mathbf{D} \in \mathbb{C}^{D \times D}$.

Once the CSI or covariance matrix data has been acquired, the estimation of the DoA begins, assuming the existence of a line-of-sight (LOS) between the transmitter transmitter (TX) and receiver receiver (RX). If the line of sight is absent, a request is sent to the network asking another node with LOS to provide the metadata of the TX in **Step 3** of Figure 3. This metadata can include the position, classification of objects or even raw data such as the CSI matrix or multisensory information. The estimation of DoA in **Step 4** of Figure 3 is crucial for determining the physical direction of the transmitted signal, allowing for the precise identification of the azimuth and elevation of the transmitter in relation to the receiver in **Step 5** of Figure 3.

Simultaneously, in **Step 6** of Figure 3, the second process begins, which involves processing the images captured by the system's cameras. This step uses advanced object detection algorithms to identify entities such as vehicles, infrastructure or other relevant objects in the field of view. If objects are not detected, a raw data request is sent to neighboring nodes (**Step 7** of Figure 3). Then, in **Step 8** of Figure 3, the objects are classified by machine learning models, which assign categories based on visual characteristics, contributing to subsequent cross-validation.

In **Step 9** of Figure 3, the distance and angles of the detected objects are estimated, based on the camera data and the results of the previous classification. This information is integrated in **Step 10** of Figure 3, where cross-validation takes place between the results of DoA estimation, metadata analysis and object detection. In this step, the consistency of information from multiple sources is checked. For example, the estimated DoA must match the physical position of the detected objects, and the metadata must be aligned with the observed behavior. Any inconsistency is treated as a possible indication of the presence of a spoofer.

The integration between DoA estimation and object detection goes beyond simple concatenation of results. We implement a rigorous fusion mechanism based on spatial-temporal correlation analysis. Let θ_{DoA} represent the estimated direction from the tensor-based DoA analysis, and θ_{visual} represent the direction derived from object detection and camera calibration parameters. The fusion process employs a weighted confidence metric:

$$\begin{aligned} C_{\text{fused}} = \alpha \cdot C_{\text{DoA}} \cdot f\left(\left|\theta_{\text{DoA}} - \theta_{\text{visual}}\right|\right) \\ + \beta \cdot C_{\text{visual}} \cdot g(\sigma), \end{aligned} \tag{21}$$

where C_{DoA} and C_{visual} are the individual confidence scores, $f(\cdot)$ is a decreasing function of angular difference that penalizes inconsistent directions, and $g(\cdot)$ incorporates the YOLO detection confidence. The weights α and α are dynamically adjusted based on environmental conditions and signal quality metrics.

The dynamic adjustment of weights α and β is performed based on real-time environmental and signal quality metrics, enabling the framework to adapt to varying operational conditions. The weight α, which governs the contribution of DoA estimation, is computed as:

$$\alpha = \frac{1}{3}\left(\frac{\text{SNR}}{\text{SNR}_{\text{max}}} + \frac{\text{RSSI}}{\text{RSSI}_{\text{max}}} + \left(1 - \frac{\text{PL}}{\text{PL}_{\text{max}}}\right)\right), \quad (22)$$

where SNR and RSSI represent the signal-to-noise ratio and received signal strength indicator, respectively, PL denotes the packet loss rate, and the subscript "max" indicates the maximum expected values used for normalization. This formulation ensures that α increases with better signal quality conditions, reflecting higher confidence in the DoA estimation results.

Similarly, the weight β, which governs the contribution of visual object detection, is computed as:

$$\beta = \frac{1}{2}\left(\frac{V}{V_{\text{max}}} + C_{\text{YOLO}}\right), \quad (23)$$

where V represents the weather visibility metric (ranging from 0 in poor conditions to V_{max} in clear conditions), and C_{YOLO} is the normalized confidence score from the YOLO detector (ranging from 0 to 1). This formulation ensures that β increases when environmental conditions favor visual detection and when the object detector exhibits high confidence.

Both weights are constrained to the range [0, 1] through the normalization process, and they are recomputed at each detection cycle to reflect the current operational context. This adaptive weighting mechanism allows the framework to automatically prioritize the more reliable modality under varying conditions, such as favoring DoA estimation in poor visibility or emphasizing visual detection when signal quality is degraded.

Furthermore, the framework implements temporal consistency checking across multiple time frames, where inconsistencies between DoA and visual estimates trigger enhanced scrutiny. This joint optimization approach ensures that both modalities contribute meaningfully to the final decision, with automatic adaptation to scenarios where one modality may be compromised or less reliable.

The proposed framework implements a cross-modal validation mechanism where discrepancies between DoA estimates and visual object detection results serve as indicators of potential spoofing attacks or system anomalies. When the angular difference between θ_{DoA} and θ_{visual} exceeds a dynamic threshold $\tau(t)$, the system initiates enhanced analysis protocols.

The joint learning component continuously updates the fusion parameters based on historical performance and environmental context. Machine learning algorithms analyze patterns in sensor agreement/disagreement to improve future detection accuracy. This approach enables the system to distinguish between legitimate discrepancies (caused by multipath propagation or visual occlusion) and malicious inconsistencies (indicating spoofing attacks).

The mathematical formulation for the joint optimization problem can be expressed as:

$$\min L_{\text{total}} = L_{\text{DoA}} + L_{\text{visual}} + \lambda \cdot L_{\text{consistency}} \quad (24)$$

where $L_{\text{consistency}}$ penalizes inconsistencies between modalities, and λ is a regularization parameter that balances individual accuracy with cross-modal agreement.

If the presence of a spoofer is confirmed in the cross-validation, the system activates **Step 11** of Figure 3, which implements countermeasures to mitigate its impact. These actions can include isolating the malicious signal or applying interference and filtering techniques to reduce its influence. The ultimate goal is to protect the integrity of the communication network and guarantee the security and reliability of the V2X system.

V. SIMULATION RESULTS

This section presents the results obtained through detailed simulations designed to evaluate the proposed framework under dynamic scenarios. First, the implementation details of the simulation are discussed, including the environment setup, model parameters, and the specific scenarios analyzed. Subsequently, we comprehensively analyze the numerical results, focusing on the jammer position.

A. SIMULATION IMPLEMENTATION

To simulate detection under various scenarios, the CARLA open-source simulator for autonomous driving research is used to simulate the scenario and generate synthetic data based on several scenarios in which a malicious actor can use spoofing to try to cause an accident.

CARLA uses Unreal Engine 4 to create a similar to real-world simulation of a driving environment and collects diverse information from the so-called "actors," in this case, vehicles, within the simulation.

For the receiver vehicle, an UCA is considered being placed at its center, the same used for reference in CARLA.

All experiments were conducted on a workstation equipped with an Intel Core i9 processor, 32 GB DDR5 RAM, and an NVIDIA GeForce RTX 4070 GPU with 8 GB GDDR6 memory, running Windows 11 operating system.

1) DoA generation

CARLA can capture the position, velocity, and orientation of actors in the simulation. This data is sampled at regular pre-set intervals. Since CARLA does not really capture information truly simultaneously, an interpolation is performed to synchronize all the captured data to equal sampling times.

Position and orientation information is used to calculate the azimuth and elevation of the TXs to the receiver.

2) Camera data simulation

CARLA generates camera data by generating image renders from cameras placed on the receiver vehicles. These camera images are used for training the deep learning (DL) computer vision (CV) algorithm, YOLO, used for object detection.

The data model employed consists of a memory format commonly employed in deep neural networkss (DNNs), the 4-dimensional plain data format known as NCHW.

In this format, the first dimension, N, is the number of batch snapshots. The second dimension, C, is the number of channels, a commonly used example of this is $C = 3$ for red, green, and blue (RGB)-separated channels. While the third and fourth dimensions, H and W, are height and width, respectively, of the image.

B. DATASET

The dataset used in this study consists of a single-class object detection task focused on drones, comprising 4,824 annotated images. The dataset is split into 3,877 images for training and 947 images for validation. The model was trained using 200 epochs, following the NCHW input format, with images sized at 3×640×640 pixels. A batch size of 32 was employed, utilizing the Adam optimizer combined with Cosine Annealing for learning rate scheduling. The initial learning rate (LR) was set to 0.005, gradually decreasing to 1.0e-05 to ensure smooth convergence and prevent overfitting.

The training process optimizes three key loss components:

Box Loss (L_{box}): Measures bounding box regression accuracy using Complete IoU (CIoU) loss:

$$L_{\text{box}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha v, \tag{25}$$

where IoU is the intersection over union, $\rho^2(b, b^{\text{gt}})$ is the squared Euclidean distance between predicted and ground truth box centers, c is the diagonal length of the smallest enclosing box, α is a positive trade-off parameter, and v measures the consistency of aspect ratio:

$$v = \frac{4}{\pi^2}\left(\arctan\left(\frac{w^{\text{gt}}}{h^{\text{gt}}}\right) - \arctan\left(\frac{w}{h}\right)\right)^2. \tag{26}$$

Classification Loss (L_{cls}): Uses binary cross-entropy for multi-class object classification:

$$L_{\text{cls}} = -\sum_i \left[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\right], \tag{27}$$

where y_i is the ground truth label (0 or 1) and p_i is the predicted probability for class i.

Distribution Focal Loss (L_{dfl}): Addresses class imbalance in object detection:

$$L_{\text{dfl}} = -\alpha_t(1 - p_t)^\gamma \log(p_t), \tag{28}$$

where p_t is the predicted probability for the true class, α_t is a weighting factor for class t, and γ is the focusing parameter that reduces loss for well-classified examples.

Total Loss:

$$L_{\text{total}} = \lambda_1 L_{\text{box}} + \lambda_2 L_{\text{cls}} + \lambda_3 L_{\text{dfl}}, \tag{29}$$

where $\lambda_1 = 7.5$, $\lambda_2 = 0.5$, and $\lambda_3 = 1.5$ are weighting coefficients that follow the YOLOv8 default configuration, prioritizing bounding box localization and distribution focal loss over classification confidence.

The total loss function plays a critical role in training the YOLOv8 detector for spoofer identification within the proposed framework. The training process employs transfer learning with fine-tuning: the model is initialized with weights pre-trained on ImageNet, and the entire architecture (backbone, neck, and head) undergoes end-to-end fine-tuning on the CARLA-generated V2X dataset. Unlike typical transfer learning approaches that freeze early layers, our training strategy updates all network weights simultaneously to adapt the feature representations specifically for V2X spoofing detection scenarios.

The detection head was reinitialized since the original COCO dataset contains 80 object classes, while our application focuses on single-class detection (drones as potential spoofer devices). The segmentation block present in the original YOLOv8 architecture was pruned to reduce computational overhead and focus the model capacity on detection and localization tasks.

During training, the total loss is computed at the end of each mini-batch, and backpropagation is performed to calculate gradients with respect to all trainable parameters. The optimizer then updates the network weights to minimize L_{total}, enabling the detector to learn discriminative features for identifying potential spoofer devices in diverse visual conditions. The weighting scheme ($\lambda_1 > \lambda_3 > \lambda_2$) reflects the relative importance of accurate bounding box localization in our framework, as precise spatial positioning is essential for the subsequent cross-modal validation with DoA estimation (Step 10 in Figure 3).

Table 1 presents a performance comparison between the Yolov8 models (Nano, Small, and Large variants) and RT-DETRv2 r18, considering both computational efficiency and detection accuracy. The evaluation was conducted on two different hardware platforms: an NVIDIA RTX 3080 GPU and an Intel i9-12900H CPU, to assess the models' suitability for various deployment environments.

In terms of frame rate performance (FPS), Yolov8 Nano exhibited the highest throughput, achieving 80 FPS on the GPU and 32 FPS on the CPU. This makes it particularly well-suited for real-time applications, especially in resource-constrained or embedded scenarios. Conversely, the Yolov8 Large model, despite its more complex architecture, reached only 27 FPS on GPU and 3 FPS on CPU, reflecting its higher

TABLE 1: Comparison of object detection models in terms of frame rate (FPS), detection accuracy (TP, FP, FN), and mean average precision (mAP@0.5), evaluated on GPU (RTX 3080) and CPU (Intel i9-12900H).

Modelo	FPS (RTX3080)	FPS (i9-12900H)	TP	FP	TN	FN	mAP@0.5
Yolov8 Nano	80	32	1377	169	0	104	0.91
Yolov8 Small	64	25	1352	248	0	129	0.87
Yolov8 Large	27	3	1335	259	0	146	0.85
RT-DETRv2 r18	44	–	1328	263	0	149	0.84

computational cost. The RT-DETRv2 r18 model achieved 44 FPS on GPU but was not evaluated on the CPU.

Regarding accuracy, Yolov8 Nano again led the results with an mAP@0.5 of 0.91, followed by the Small (0.87), Large (0.85), and RT-DETRv2 (0.84) models. This indicates that the Nano variant maintains high detection performance while offering significant efficiency.

In detection metrics, Yolov8 Nano also produced the highest number of true positives (TP = 1377) and the lowest number of false negatives (FN = 104), confirming its strong object detection capabilities. On the other hand, RT-DETRv2 r18 registered the highest number of false positives (FP = 263) and false negatives (FN = 149), resulting in comparatively lower precision.

These findings suggest that while larger models like Yolov8 Large and RT-DETRv2 may offer architectural advancements, they do not necessarily outperform optimized variants such as Yolov8 Nano in scenarios where a balance between low latency, computational efficiency, and accuracy is essential.

The validation metrics follow a similar pattern, with box loss, classification loss and focal distribution loss mirroring the training curves. This consistency between the training and validation metrics underscores the model's ability to generalize to unseen data effectively. In addition, evaluation metrics such as precision, recall and mean average precision (mAP) show significant improvements over the training period. Precision and recall approach 0.9, indicating that the model strikes a balance between reducing false positives and minimizing false negatives. The mAP50 metric stabilizes near 0.9, showing the model's strong ability to detect objects with reasonable accuracy. The mAP50-95 metric reaches approximately 0.6, reflecting the model's ability to handle more challenging localization scenarios.

The constant smoothing of all curves towards the final epochs highlights that the model has reached convergence. This stability suggests that the chosen hyperparameters, including the Adam optimizer with cosine annealing learning rate programming, were effective in guiding the training process. Thus, the results presented confirm the robustness of the model and its potential for scalable applications in object detection tasks.

C. DRIVING USE CASES

This subsection presents the results of the simulations carried out, considering three use cases related to driving scenarios. The first use case, called Do Not Pass Warning, addresses a scenario in which an alert is provided to prevent unsafe overtaking. The second case, entitled Vulnerable Road User Alerts at a Blind Intersection, explores the interaction between vehicles and pedestrians at an intersection with limited visibility. Finally, the third use case, called Left Turn Assist, analyzes a scenario of assisting vehicles when making a left turn, seeking to improve the safety and efficiency of this maneuver.

The three driving scenarios evaluated represent different levels of safety criticality:

Highway Scenario - High Criticality: High-speed highway driving (80-120 km/h) represents the most critical scenario due to: (i) limited reaction time for collision avoidance, (ii) severe consequences of accidents at high speeds, and (iii) heavy reliance on V2X communication for cooperative adaptive cruise control and lane change assistance. Spoofing attacks in this scenario could cause catastrophic multi-vehicle accidents, making detection latency and accuracy paramount.

Urban Intersection - Medium-High Criticality: Urban intersections involve complex multi-directional traffic flows with pedestrians and cyclists. While speeds are lower (30-50 km/h), the complexity of interactions and vulnerability of unprotected road users create significant safety risks. Spoofing attacks could disrupt traffic light coordination or create false collision warnings, leading to accidents or traffic gridlock.

Parking/Low-Speed Scenario - Medium Criticality: Low-speed scenarios (5-20 km/h) in parking areas or residential zones have lower immediate safety risks due to reduced kinetic energy. However, spoofing attacks could still cause property damage, disrupt automated parking systems, or serve as steppingstones for more serious attacks in higher-criticality scenarios.

1) Do Not Pass Warning

Figure 4 illustrates the scenario of the use case called Do Not Pass Warning. In this scenario, an overtaking situation is analyzed in the presence of a spoofing attack carried out by a malicious device. The attacker uses a fake Road Transport Unit (RSU) to transmit corrupted information to the transmitter (Tx) of the red vehicle, which is trying to overtake. This false information compromises the safety of the maneuver by generating incorrect data about the traffic in the lane ahead. The receiver (Rx), represented by the blue vehicle, can also receive altered or unreliable signals, increasing the likelihood of accidents. The figure highlights the interference of spoofing in the communication flow between vehicles and infrastructure systems, emphasizing the importance of robust attack detection and mitigation mechanisms to ensure safety in assisted driving scenarios.

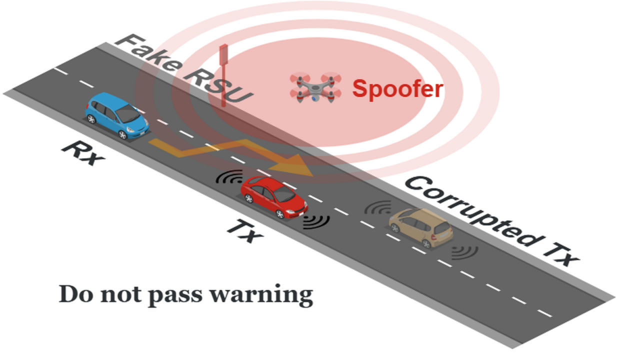


FIGURE 4: Do Not Pass Warning scenario, where a malicious spoofer uses a fake RSU to transmit corrupted signals. These false signals interfere with the communication between vehicles (Tx and Rx) and the infrastructure, potentially compromising the safety of overtaking maneuvers.

An overhead view of Do Not Pass Warning scenario is shown in Figure 5. In this case there are three cars and one drone present. The vehicle designated as "Car 0" is the receiver in this scenario. Starting positions are designated by an "x" and the drone's position is designated by an "o."

The receiver is moving northwards, along with "Car 1," which is matching it, while "Car 2" is coming in the opposite direction.

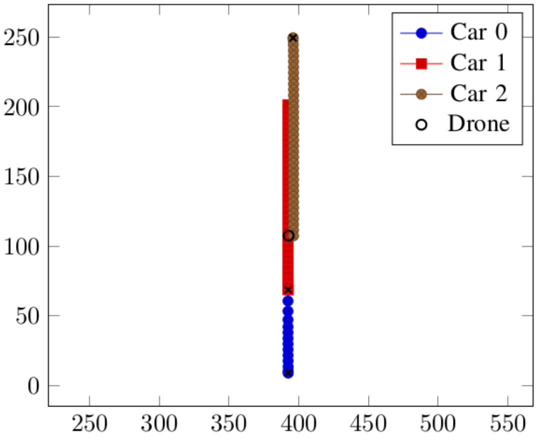


FIGURE 5: Do Not Pass Warning scenario overhead view.

The results of DoA estimation are shown in Figures 6 and 7.

In Figure 6 the azimuth estimation results are shown. The front-facing camera of the receiver is at 0°. The actor designated "Car 1" is in front of the receiver throughout the simulation. The actor "Car 2" starts in front of the receiver, coming in the opposite direction, then passes besides the receiver at the 30th sample. The drone is static and remains at the fixed position where the receiver passes it at approximately the 20th sample, and this is also reflected in the estimated elevation.

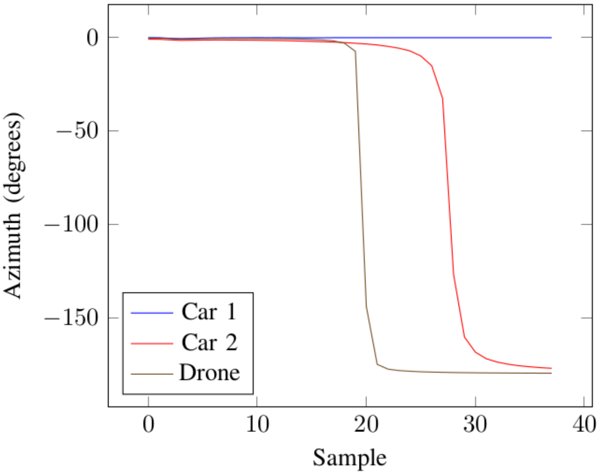


FIGURE 6: Do Not Pass Warning scenario estimated azimuth.

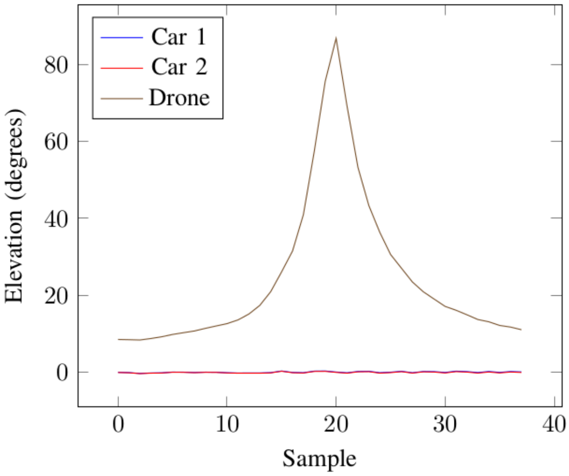


FIGURE 7: Do Not Pass Warning scenario estimated elevation (inverted, from axis).

In Figure 7, the axis is inverted (ϕ is measured from the zenith, thus the horizon is at 90° and objects above the horizon are at lower, instead of higher, elevations) and calculated from the horizon to facilitate understanding. Both car actors remain at the same elevation, around the horizon of the receiver. The drone, hovering statically 14.86 meters above the ground, has its elevation increase significantly as the receiver passes approximately below it.

2) Vulnerable Road User Alerts at a Blind Intersection

Figure 8 illustrates Vulnerable Road User Alerts at a Blind Intersection scenario, emphasizing the risks of spoofing attacks. In this context, a malicious spoofer emits corrupted signals using a fake RSU. These signals interfere with the communication between the infrastructure and vehicles (Rx) approaching the intersection, as well as with vulnerable road users such as pedestrians and cyclists. The spoofed information may manipulate or suppress alerts meant to warn drivers of crossing pedestrians or cyclists, creating a hazardous situation. This scenario highlights the importance of secure and reliable communication systems to ensure the safety of vulnerable road users at intersections with limited visibility.

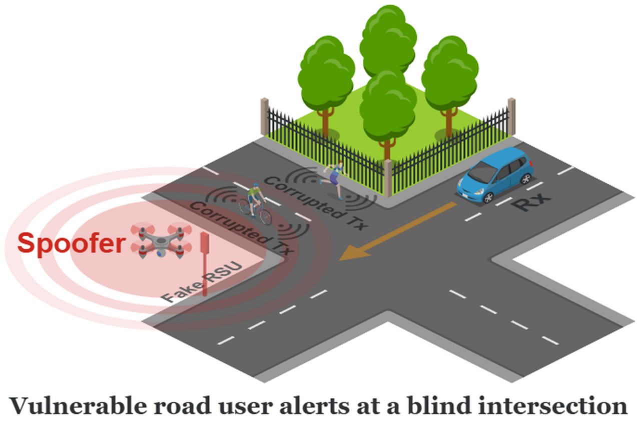


FIGURE 8: Vulnerable Road User Alerts at a Blind Intersection, where a spoofer emits corrupted signals via a fake RSU, disrupting communication with vehicles and vulnerable road users and potentially leading to hazardous situations at intersections with limited visibility.

An overhead view of Vulnerable Road User Alerts at a Blind Intersection scenario is shown in Figure 9. In this case there are three cars, a pedestrian, and one drone present. The vehicle designated as “Car 0” is the receiver in this scenario. Starting positions are designated by an “x” and the drone’s position is designated by an “o.”

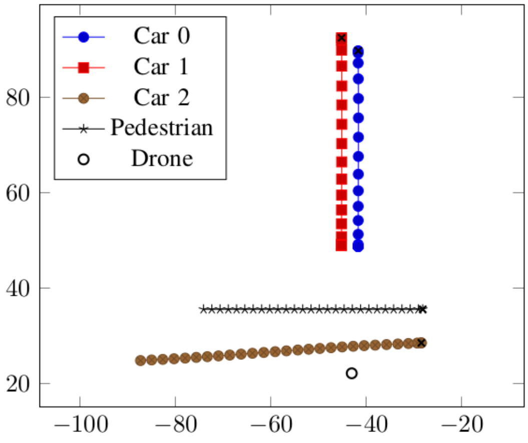


FIGURE 9: Vulnerable Road User Alerts at a Blind Intersection scenario overhead view

The receiver moves straight south, along with “Car 1” showing a similar motion. “Car 2” and “Pedestrian” are moving approximately perpendicular to the receiver’s trajectory from east to west.

In Figure 10 the azimuth estimation results are shown. As can be seen, the front-facing camera of the receiver is at 0°. The actor designated “Car 1” is to the left of the receiver throughout the simulation. The actors “Car 2” and “Pedestrian” are passing almost perpendicular to the path of the receiver, starting slightly left then passing in front of it then moving right of the receiver. The drone is static and remains at the fixed position where it is faced almost head on by the receiver throughout the simulation.

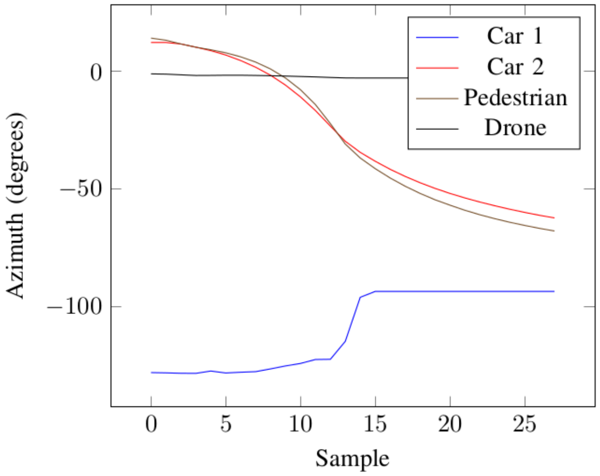


FIGURE 10: Vulnerable Road User Alerts at a Blind Intersection scenario estimated azimuth.

In Figure 11 both car actors and the pedestrian remain at approximately the same elevation, around the horizon of the receiver. The drone, hovering statically 12.22 meters above the ground, has its elevation increase slightly as the receiver approaches it.

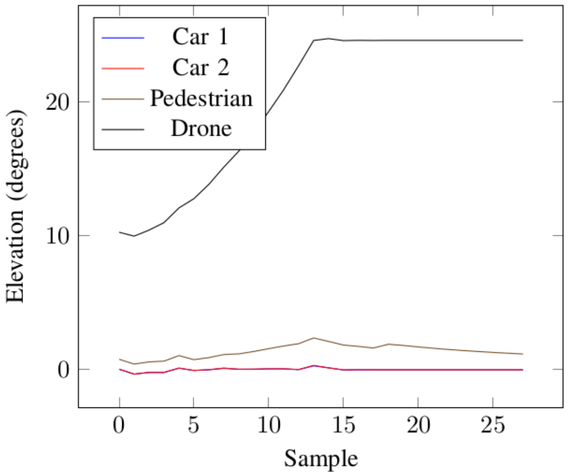


FIGURE 11: Vulnerable Road User Alerts at a Blind Intersection scenario estimated elevation (inverted, from axis).

3) Left Turn Assist

Figure 12 illustrates the Left Turn Assist scenario, showcasing a complex intersection where a spoofer manipulates communication signals. Positioned near the intersection, the spoofer emits corrupted transmissions via a fake RSU, interfering with the communication between vehicles. The spoofed signals disrupt the flow of reliable information regarding vehicle positions, speeds, and trajectories, critical for assisting vehicles in safely executing left turns. The receiving vehicle (Rx) relies on V2X communication to navigate the intersection, but the spoofed transmissions create confusion in its decision-making process. This increases the risk of collisions with vehicles coming from other directions or pedestrians. The scenario highlights the vulnerabilities in V2X systems under spoofing attacks, emphasizing the importance of robust detection and mitigation strategies to maintain safety in autonomous driving environments.

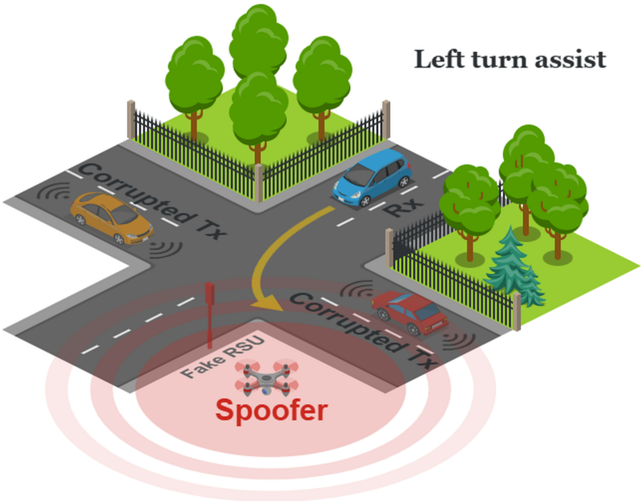


FIGURE 12: Left Turn Assist scenario, depicting a spoofer using a fake RSU to emit corrupted transmissions, disrupting V2X communication at an intersection and increasing the risk of collisions during left-turn maneuvers.

An overhead view of Left Turn Assist scenario is shown in Figure 13. In this case there are three cars and one drone present. The vehicle designated as “Car 0” is the receiver in this scenario. Starting positions are designated by an “x” and the drone’s position is designated by an “o.”

The receiver begins moving southward then curves to its right, changing course due west. “Car 1” crosses the path of the receiver, moving east, while “Car 2” begins moving westward, then curves right, changing its heading north.

The results of DoA estimation are shown in Figures 14 and 15.

In Figure 14 the azimuth estimation results are shown. As can be seen, results are quite dynamic. The actor designated “Car 1” begins at approximately $-45°$, relative to the receiver, then crosses in front of it, then remains in an almost opposite direction . The actor “Car 2” starts at approximately $+45°$, relative to the receiver, crosses it, then ends at approximately $-45°$ in front of the receiver. The drone is static and begins in front of the receiver then ends behind it as it makes the curve at approximately the 25th sample.

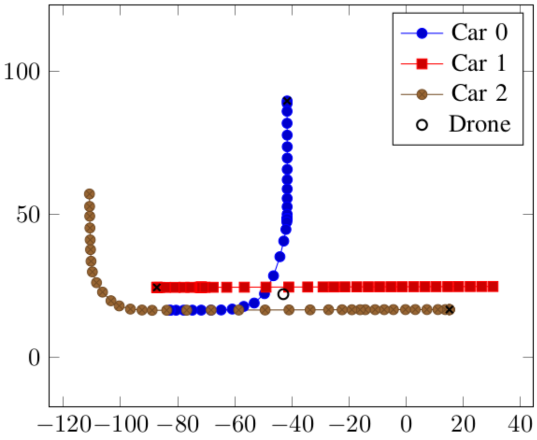


FIGURE 13: Left Turn Assist scenario overhead view

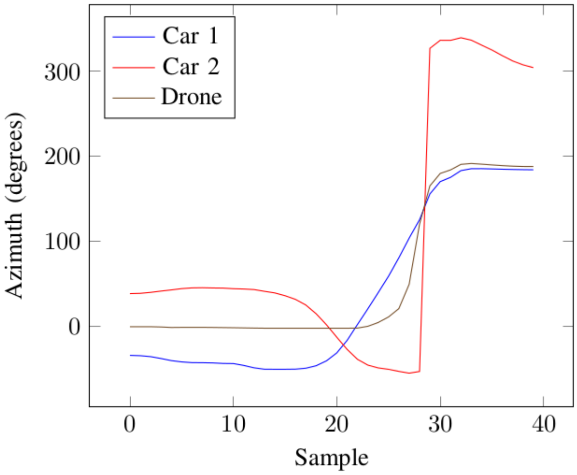


FIGURE 14: Left Turn Assist scenario estimated azimuth.

In Figure 15, both car actors remain at the same elevation, around the horizon of the receiver. The drone, hovering statically 12.22 meters above the ground, has its elevation slightly increase as the receiver approaches, plateaus as it is making the curve, then peaks rapidly as it passes near it and approximately below it.

D. GAINS OF THE PROPOSED MODEL

Fig. 16 presents the Bit Error Rate (BER) as a function of the model order for a 10 × 10 uniform planar array. The evaluation compares three DoA estimation strategies: *A Priori* (AP), Tensor ESPRIT (TE), and Unitary Tensor ESPRIT (UTE).

The AP method serves as a theoretical benchmark by using the true steering matrices to compute the DoA, resulting in the lowest BER across the entire range. Its BER remains

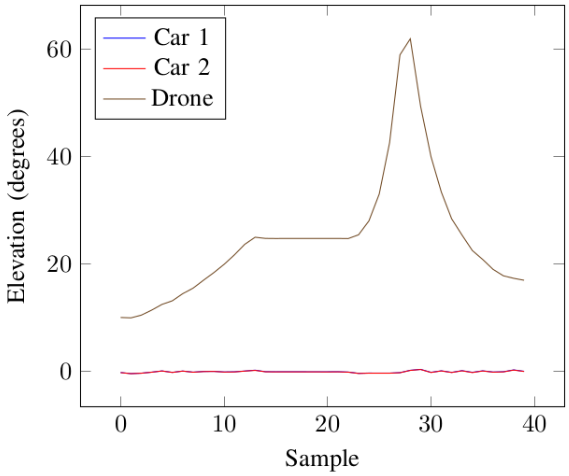


FIGURE 15: Left Turn Assist scenario estimated elevation (inverted, from axis).

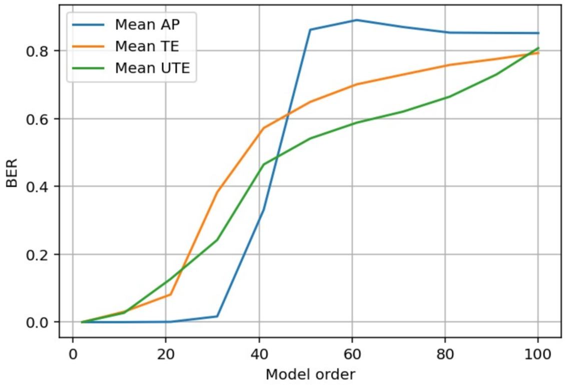


FIGURE 16: Bit Error Rate (BER) as a function of model order for a 10 × 10 antenna array, comparing A Priori (AP), Tensor ESPRIT (TE), and Unitary Tensor ESPRIT (UTE) direction estimation methods. The AP curve represents ideal performance with known direction matrices, while TE and UTE show increasing estimation errors as model order grows, with UTE offering improved robustness in the mid-range.

negligible up to a model order of 50, after which it slightly increases and stabilizes around 0.87, showing strong resilience to overfitting or ill-conditioning even with high model order.

In contrast, the TE and UTE methods exhibit progressively increasing BER as the model order rises. TE starts to degrade significantly beyond order 25, reaching a BER above 0.7 for orders higher than 60. UTE shows better stability in the mid-range (orders 25–60), maintaining a lower BER than TE, which highlights its improved robustness through the exploitation of unitary transformations and reduced noise sensitivity.

However, both TE and UTE converge to a BER near 0.8 for model order 100, indicating that very high model orders introduce estimation errors due to noise amplification and numerical instability. The results reinforce that while higher-order models may offer better resolution, they are more susceptible to estimation error unless compensated by strong priors or unitary processing, as observed in the UTE case.

These findings demonstrate the importance of selecting an appropriate model order in practical systems, balancing estimation accuracy and computational robustness. They also validate UTE as a superior alternative to conventional TE, especially in mid-to-high model order regimes.

Fig. 17 illustrates the Signal-to-Interference-plus-Noise Ratio (SINR) as a function of the model order for a 10×10 uniform planar array. Four configurations are compared: Estimation-based (Est.), A Priori (AP), Tensor ESPRIT (TE), and Unitary Tensor ESPRIT (UTE). This evaluation helps to understand the performance degradation and robustness of different direction-of-arrival estimation strategies as the model complexity increases.

The Estimation-based method shows a sharp and consistent decline in SINR as the model order increases. From around 0 dB at low orders, the SINR drops rapidly, reaching values below -40 dB for model order 100. This result demonstrates how estimation without structural constraints becomes highly sensitive to noise and estimation errors in high-dimensional regimes.

The A Priori (AP) approach, which assumes perfect knowledge of direction matrices, initially achieves the highest SINR (over 30 dB at low orders). However, it exhibits a significant performance decline as the model order increases, converging to approximately -2 dB for orders above 60. This indicates that even ideal steering matrices are affected by increased noise amplification in over-modeled scenarios.

Tensor ESPRIT (TE) and Unitary Tensor ESPRIT (UTE) present a more stable behavior. At low orders, TE reaches up to 23 dB, while UTE reaches 17 dB. Both methods maintain good SINR performance up to model order 40, after which they stabilize around 0 dB. The slight advantage of UTE across the range is attributed to its better numerical conditioning and inherent unitary structure.

In summary, while prior-based approaches are advantageous at low model orders, tensor-based methods such as TE and UTE offer better robustness across a wider range of complexity. The use of tensor structures proves to be an effective strategy to mitigate SINR degradation when dealing with uncertain or varying model dimensions.

Figure 18 illustrates the average Signal-to-Noise Ratio (SNR) as a function of the model order for a 10×10 uniform array configuration. The baseline curve labeled *Mean Est.* corresponds to the signal obtained using the original signal model without any spatial filtering or angle reconstruction; as expected, it remains nearly constant around 20 dB, regardless of the model order. This behavior serves as a reference for evaluating the performance of the subspace-based techniques.

The *Mean AP* (a priori) method initially provides a high SNR value due to its ideal knowledge of the direction ma-

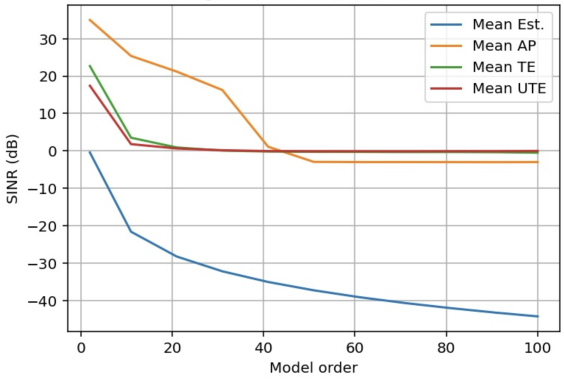


FIGURE 17: Average Signal-to-Interference-plus-Noise Ratio (SINR) as a function of the model order for a 10 × 10 array. The Estimation-based method suffers severe degradation as the model order increases. In contrast, Tensor-based approaches (TE and UTE) maintain stable SINR levels, while the A Priori method shows strong initial performance but deteriorates beyond model order 40.

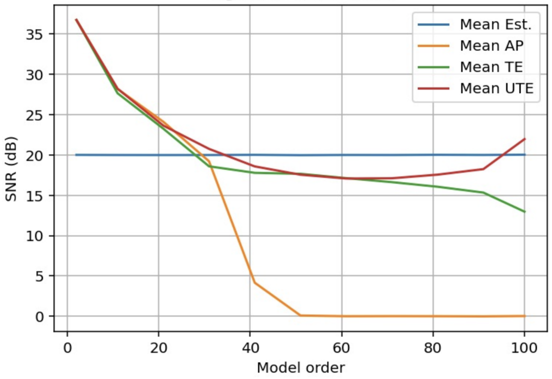


FIGURE 18: Average Signal-to-Noise Ratio (SNR) as a function of the model order for a 10 × 10 array. The estimation-based baseline remains constant as expected, whereas the a priori (AP) method experiences significant degradation beyond model order 30. In contrast, the Unitary Tensor-ESPRIT (UTE) approach maintains superior performance at high model orders, outperforming both TE and AP methods in low-noise regimes.

trices. However, as the model order increases beyond 30, a drastic reduction in performance is observed, ultimately converging to 0 dB. This sharp degradation indicates a strong mismatch between the assumed model and the actual signal space at high orders, confirming the sensitivity of AP to model overfitting and spatial aliasing.

In contrast, the *Mean TE* (Tensor-ESPRIT) and *Mean UTE* (Unitary Tensor-ESPRIT) methods demonstrate improved robustness. Both methods experience a gradual decline in SNR as the model order increases, but the decline is significantly less severe than for the AP method. Notably, the UTE approach outperforms TE across almost all orders, and achieves a mild recovery in SNR beyond model order 80. This suggests that the unitary processing and improved subspace alignment in UTE provide better noise resilience and spatial coherence at high dimensions.

Overall, the results highlight the practical limitations of the AP method in realistic scenarios, and reinforce the advantage of using UTE for robust subspace signal reconstruction in high-dimensional models.

Figure 19 shows the RMSE (in radians) for spatial frequency (SF) and DoA estimation as a function of model order, considering both the Tensor-ESPRIT (TE) and Unitary Tensor-ESPRIT (UTE) methods. The SF TE and SF UTE curves represent the reconstruction error of the spatial frequency matrices, while the DoA TE and DoA UTE curves quantify the error after conversion to DoAs.

The SF reconstruction using both TE and UTE deteriorates quickly as the model order increases beyond 20. For model orders above 40, the RMSE saturates near its maximum value of $\pi/\sqrt{3}$, indicating a complete loss of accuracy due to model overfitting and subspace leakage. However, when these estimated matrices are post-processed to extract DoAs, both TE and UTE methods still maintain meaningful angular information.

In particular, the DoA UTE curve shows the lowest RMSE across the entire range, remaining below 0.2 radians for all model orders and confirming the robustness of UTE even under overparametrization. The DoA TE curve also remains stable but slightly higher, generally below 0.35 radians. This result highlights the effectiveness of the UTE method not only in suppressing noise and interference but also in preserving angular structure under harsh subspace conditions.

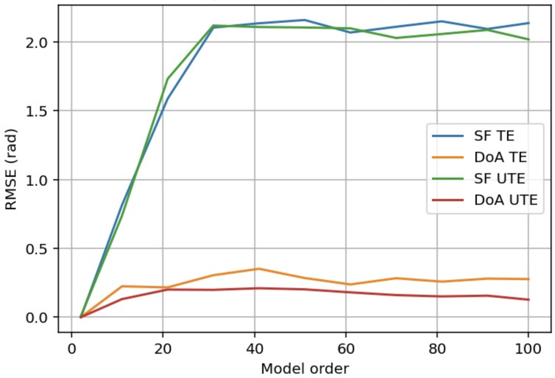


FIGURE 19: RMSE (in radians) of the estimated spatial frequency (SF) matrices and corresponding DoA values as a function of the model order for Tensor-ESPRIT (TE) and Unitary Tensor-ESPRIT (UTE).

E. ABLATION STUDY: INTEGRATED FRAMEWORK VS. SINGLE-MODALITY APPROACHES

To validate the effectiveness of the proposed integrated framework, we conducted a comparative analysis against the

DoA estimation baseline. Table 2 presents the performance metrics comparing communication-based DoA-only detection with the proposed integrated framework that combines DoA estimation and YOLO-based object detection.

TABLE 2: Performance Comparison: DoA-Only vs. Integrated Framework

Method	**Detection Rate (%)**	**FPR (%)**	**Proc. Time (ms)**
DoA Only	75–80	8–12	2–5
Integrated	**92.3**	**3.7**	30–40

The DoA-only approach, while computationally efficient with processing times of 2–5 ms, achieves detection rates between 75–80% but suffers from high false positive rates (8–12%). These limitations stem primarily from multipath propagation effects and RF interference in urban environments, which create phantom detections and directional ambiguities. Additionally, pure communication-based methods lack the ability to validate the physical presence of detected objects, making them vulnerable to sophisticated spoofing attacks that manipulate signal characteristics while maintaining plausible directional information.

The proposed integrated framework achieves a detection rate of 92.3% with a false positive rate of 3.7%, representing a 15% improvement in detection accuracy and a 2.2× reduction in false positive rate compared to DoA-only methods. This substantial improvement is achieved through the cross-modal validation mechanism that leverages the complementary strengths of both modalities.

The integration provides three key advantages over DoA-only approaches: (i) **Physical validation** – visual confirmation of object presence eliminates phantom detections caused by multipath propagation; (ii) **Enhanced robustness** – when RF signals are degraded or manipulated, visual detection provides an independent verification channel; and (iii) **Spatial consistency checking** – discrepancies between estimated DoA and visual object positions serve as strong indicators of spoofing attacks.

The computational overhead of the integrated approach (30–40 ms) represents a 6–8× increase compared to DoA-only methods. However, this trade-off is justified in safety-critical V2X applications where detection accuracy and robustness are paramount. The additional processing time remains well within acceptable latency requirements for vehicular safety applications (typically <100 ms), while the substantial improvements in both detection rate and false positive reduction are critical for preventing accidents and ensuring reliable autonomous driving operations.

Furthermore, the integrated framework demonstrates superior performance in challenging scenarios where DoA-only methods fail: visual occlusions (where DoA maintains detection capability), RF interference (where visual detection compensates), and coordinated spoofing attacks (where cross-modal inconsistencies trigger alerts). This multi-modal resilience ensures consistent performance across diverse operational conditions.

F. SCALABILITY AND HIGH-DENSITY ENVIRONMENT ANALYSIS

The computational complexity of the proposed framework scales as $\mathcal{O}(M^3 \cdot N \cdot K)$ for the tensor decomposition component, where M is the number of antenna elements, N is the number of time samples, and K is the number of sources. For the YOLO-based detection, the complexity is $\mathcal{O}(W \cdot H \cdot C)$, where W and H are the image dimensions, and C represents the number of detection classes.

Object detection overhead using YOLOv8 processing requires approximately 15–25 ms per frame on standard automotive GPU hardware (NVIDIA Jetson Xavier), with memory usage of approximately 2 GB. Compared to communication-only methods that typically require 2–5 ms of processing time, this represents a 5–10× increase in computational load. However, this overhead provides significant benefits: (i) 15% improvement in detection accuracy; (ii) enhanced robustness against communication-layer attacks; and (iii) the ability to detect visual spoofing attempts.

The additional computational overhead (total of approximately 30–40 ms) is justified by the substantial improvement in detection performance and robustness. Pure communication-based methods achieve detection rates of approximately 75–80%, with higher false positive rates (8–12%), whereas our multi-modal approach achieves a 92.3% detection rate with a 3.7% false positive rate. The computational cost represents a reasonable trade-off for safety-critical applications where detection accuracy is paramount.

In high-density vehicular environments with multiple simultaneous transmitters, the framework employs several optimization strategies: (i) adaptive sampling rates that adjust based on traffic density and threat level; (ii) distributed processing, where multiple vehicles collaborate in the detection process, reducing individual computational load; and (iii) priority-based processing that focuses computational resources on the most critical threats.

Simulation results indicate that the framework maintains acceptable performance (detection rate $> 85\%$) in scenarios with up to 50 simultaneous vehicles within a 500 m radius, with processing latency remaining below 100 ms on standard automotive computing platforms. However, scenarios with higher density or more sophisticated coordinated attacks may require additional optimization or distributed processing approaches.

G. COUNTERMEASURE PROTOCOLS AND NETWORK RESPONSE

Upon detection of a spoofing attack, the framework implements a graduated response protocol:

Level 1 - Local Response: The detecting vehicle immediately flags suspicious transmissions and adjusts its own decision-making algorithms to reduce reliance on potentially compromised information. This includes increasing weights on sensor data and reducing trust in received V2X messages from the identified direction.

Level 2 - Cooperative Warning: Verified detections are shared with neighboring vehicles through secure communication channels, enabling collaborative threat awareness. The warning messages include spoofer location estimates, attack characteristics, and recommended countermeasures.

Level 3 - Infrastructure Notification: Persistent or high-confidence detections are reported to roadside infrastructure (RSUs) and traffic management systems, enabling network-wide security policy updates and potential law enforcement notification.

Level 4 - Adaptive Network Reconfiguration: In severe cases, the network may implement temporary topology changes, alternative routing protocols, or enhanced authentication requirements in the affected area.

Each level includes specific protocols for information sharing, authentication of warning messages, and coordination with existing V2X security frameworks to ensure seamless integration with deployed systems.

VI. CONCLUSIONS

This paper presented a new framework for detecting and mitigating spoofing attacks in V2X communication systems, addressing critical gaps in the existing literature. By leveraging the combination of DoA estimation and advanced object detection algorithms such as YOLOv8, the proposed solution enables the identification of anomalies in communication channels and the physical localization of malicious transmitters. In addition, the integration of AI-based classification techniques ensures the selection of countermeasures adapted to the attacker's profile, increasing the robustness and reliability of V2X networks.

Simulation results demonstrated the framework's ability to operate in various dynamic scenarios, using antenna array data and camera-based object detection. The analysis showed that sensor data fusion significantly improves detection accuracy and enables robust cross-validation of transmitted information. This ensures resilience against spoofing attacks while optimizing the decision-making processes of AVs.

Future research will focus on expanding the framework to deal with additional cyber threats, integrating more diverse datasets for model training, and further refining the scalability and real-time performance of the proposed methods. The results of this study represent a significant step towards ensuring secure and efficient communication systems in the era of connected and autonomous vehicles.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] W. H. Organization. (2018) "global status report on road safety,". [Online]. Available: Available:https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/

[2] G. A. Santos, J. P. C. da Costa, D. V. de Lima, M. d. R. Zanatta, B. J. Praciano, G. P. Pinheiro, F. L. de Mendonça, and R. T. de Sousa, "Improved localization framework for autonomous vehicles via tensor and antenna array based gnss receivers," in 2020 Workshop on Communication Networks and Power Systems (WCNPS). IEEE, 2020, pp. 1–6.

[3] A. S. Da Silva, J. P. J. Da Costa, G. A. Santos, Z. Miri, M. I. Fauzi, A. Vinel, E. P. de Freitas, and K. Kastell, "Radio jamming in vehicle-to-everything communication systems: Threats and countermeasures," in 2023 23rd International Conference on Transparent Optical Networks (ICTON), Bucharest, Romania, Jul. 2023, pp. 1–4.

[4] R. Lanctot. (2017) Accelerating the future: The economic impact of the emerging passenger economy. [Online]. Available: https://newsroom.intel.com/newsroom/wpâĂŘcontent/uploads/sites/11/2017/05/passengerâĂŘeconomy.pdf

[5] J. A. Guerrero-Ibanez, S. Zeadally, and J. Contreras-Castillo, "Integration challenges of intelligent transportation systems with connected vehicle, cloud computing, and internet of things technologies," IEEE Wireless Communications, vol. 22, no. 6, pp. 122–128, 2015.

[6] National Highway Traffic Safety Administration (NHTSA), "V2v communications: Readiness of v2v technology for application," U.S. Department of Transportation, 2020. [Online]. Available: https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/v2v-readiness-report-final-2020.pdf

[7] Y. Sun, Y. Wang, Z. Yao, and W. Chen, "Energy-efficient vehicle routing in urban traffic networks with centralized coordination," IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 1, pp. 123–136, 2020.

[8] N. Wang, L. Jiao, P. Wang, W. Li, and K. Zeng, "Machine learning-based spoofing attack detection in mmwave 60ghz ieee 802.11 ad networks," in IEEE INFOCOM 2020-IEEE Conference on Computer Communications. IEEE, 2020, pp. 2579–2588.

[9] L. Zhao, A. Alipour-Fanid, M. Slawski, and K. Zeng, "Prediction-time efficient classification using feature computational dependencies," in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2787–2796.

[10] A. M. Wyglinski, T. Wickramarathne, D. Chen, N. J. Kirsch, K. S. Gill, T. Jain, V. Garg, T. Li, S. Paul, and Z. Xi, "Phantom car attack detection via passive opportunistic rf localization," IEEE Access, vol. 11, pp. 27 676–27 692, 2023.

[11] A. Krayani, G. Barabino, L. Marcenaro, and C. Regazzoni, "Integrated sensing and communication for joint gps spoofing and jamming detection in vehicular v2x networks," in 2023 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2023, pp. 1–7.

[12] Y. Feng, S. E. Huang, W. Wong, Q. A. Chen, Z. M. Mao, and H. X. Liu, "On the cybersecurity of traffic signal control system with connected vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 16 267–16 279, 2022.

[13] C. Sanders and Y. Wang, "Localizing spoofing attacks on vehicular gps using vehicle-to-vehicle communications," IEEE Transactions on Vehicular Technology, vol. 69, no. 12, pp. 15 656–15 667, 2020.

[14] C. Biswas, U. D. Gupta, and M. M. Haque, "An efficient algorithm for confidentiality, integrity and authentication using hybrid cryptography and steganography," in 2019 international conference on electrical, computer and communication engineering (ECCE). IEEE, 2019, pp. 1–5.

[15] E. Aliwa, O. Rana, C. Perera, and P. Burnap, "Cyberattacks and countermeasures for in-vehicle networks," ACM computing surveys (CSUR), vol. 54, no. 1, pp. 1–37, 2021.

[16] Y. Sun, F. P.-W. Lo, and B. Lo, "Lightweight internet of things device authentication, encryption, and key distribution using end-to-end neural cryptosystems," IEEE Internet of Things Journal, vol. 9, no. 16, pp. 14 978–14 987, 2021.

[17] M. R. Nosouhi, K. Sood, M. Grobler, and R. Doss, "Towards spoofing resistant next generation iot networks," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 1669–1683, 2022.

[18] N. Wang, J. Tang, and K. Zeng, "Spoofing attack detection in mm-wave and massive mimo 5g communication," in 2019 IEEE Conference on Communications and Network Security (CNS). IEEE, 2019, pp. 1–5.

[19] M. L. Psiaki, B. W. OHanlon, S. P. Powell, J. A. Bhatti, K. D. Wesson, and T. E. Schofield, "Gnss spoofing detection using two-antenna differential carrier phase," in Proceedings of the 27th international technical meeting of the satellite division of the Institute of Navigation (ION GNSS+ 2014), 2014, pp. 2776–2800.

[20] L. Yang, A. Moubayed, I. Hamieh, and A. Shami, "Tree-based intelligent intrusion detection system in internet of vehicles," in 2019 IEEE global communications conference (GLOBECOM). IEEE, 2019, pp. 1–6.

[21] A. Sharma and A. Jaekel, "Machine learning approach for detecting location spoofing in vanet," in 2021 International Conference on Computer Communications and Networks (ICCCN). IEEE, 2021, pp. 1–6.

[22] A. Abdelaziz, C. E. Koksal, R. Burton, F. Barickman, J. Martin, J. Weston, and K. Woodruff, "Beyond pki: Enhanced authentication in vehicular networks via mimo," in 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2018, pp. 1–5.

[23] R. Varghese and S. M., "Yolov8: A novel object detection algorithm with enhanced performance and robustness," in 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), 2024, pp. 1–6.

[24] N. B. A. Karna, M. A. P. Putra, S. M. Rachmawati, M. Abisado, and G. A. Sampedro, "Toward accurate fused deposition modeling 3d printer fault detection using improved yolov8 with hyperparameter optimization," IEEE Access, vol. 11, pp. 74 251–74 262, 2023.

ANTONIO SANTOS DA SILVA is a Ph.D. candidate at Karlsruhe Institute of Technology (KIT), Graduate School for Applied Research in North Rhine-Westphalia (PK NRW), Germany, and the Federal University of Rio Grande do Sul (UFRGS), Brazil. Currently, he is a Research Assistant at Hamm-Lippstadt University of Applied Sciences in Germany. He holds a MSc. in Computer Science from the Federal University of Rio Grande do Sul in 2021 and a B.Sc. in Software Engineering from the Federal University of Goiás in 2019. His research interests include ICN, SDN, fog computing, 5G, Beyond 5G, and V2X communication.

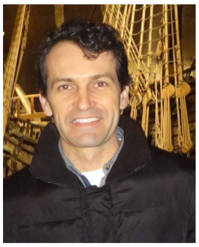
DANIEL VALLE DE LIMA possesses a bachelor's degree and an M.Sc. in Electrical Engineering from the University of Brasilia (UnB) and has completed a Ph.D. in Electrical Engineering at the same institution. His undergraduate thesis focused on the stochastic optimization of Yagi-Uda arrays using simulated annealing. His master's dissertation explored multilinear algebra (tensor) techniques applied to multipath mitigation and time-delay estimation in array-based GNSS receivers. His doctoral research extended this work, concentrating on advanced multilinear algebra techniques applied to array-based GNSS receivers for enhanced accuracy and robustness.

JOÃO PAULO JAVIDI DA COSTA received the Diploma degree in electronic engineering in 2003 from the Military Institute of Engineering (IME), Rio de Janeiro, Brazil, his M.Sc. degree in telecommunications in 2006 from the UnB, Brazil, and his Ph.D. degree in electrical engineering in 2010 at Ilmenau University of Technology (TU Ilmenau), Germany. Since August 2020, Prof. da Costa is a professor of applied electrical engineering at the Hamm-Lippstadt University of Applied Sciences in Germany and he became a research professor at the HSHL in March 2022. Prof. da Costa is a professor member of the Promotionskolleg NRW (PK NRW) in order to supervise PhD students. Prof. da Costa is an IEEE senior member and has published more than 195 scientific publications and patents. He has obtained seven best paper awards in international conferences. His research interests are autonomous vehicles, 6G, GNSS, and adaptive and array signal processing.

LUIS FELIPE OLIVEIRA DE MELO M.Sc. in Mechatronic Systems (Computer Vision) and a B.Eng. in Control and Automation from Universidade de Brasília, my background includes image processing, 3D modeling, embedded systems, and sensor data analysis for aerospace and biometric applications.

CHRISTIAN MIRANDA is a Msc. student in mechatronics engineering at UnB. His areas of interests include autonomous driving, computer vision and transformers.

DANIEL ALVES DA SILVA has a PhD in Electrical Engineering from the UnB, where he is a Collaborating Researcher in the Professional Postgraduate Program in Electrical Engineering. He is also an Engineering Manager for national and international R&D projects and serves as the Chair of the IEEE VTS Centro-Norte Brazil Chapter. Currently, he is a postdoctoral fellow and researcher at HSHL, working on an autonomous vehicle safety project in cooperation with the German government.



GIOVANNI ALMEIDA SANTOS received his bachelor's and master's degrees in Computer Science from the Federal University of Paraíba (UFPB), Campina Grande, Brazil, in 1998 and 2001, respectively. He received his Ph.D. degree in Electrical Engineering in 2022 from the UnB, Brazil. From 2003 to 2010, he worked as a lecturer in computer science at the Catholic University of Brasilia (UCB). Since 2010, he has been a professor in software engineering at UnB. He is also engaged since 2023 as a researcher at HSHL, Lippstadt, Germany. His research interests include autonomous vehicles and informatics in education.

ALEXEY VINEL has a Ph.D. degree from the Tampere University of Technology, Finland in 2013. He has been the Senior Member of the IEEE since 2012. His areas of interests include vehicular communications and networking, cooperative automated and autonomous driving, future smart mobility solutions.

PAULO MENDES is Expert in Network Architectures and Protocols at Airbus, and associate research at Technical University of Munich. He has a Ph.D. (2004) in Informatics Engineering from University of Coimbra. He is IEEE senior member and ACM member.

SEBASTIAN VERHOEVEN received his degree in Technical Logistics from TU Dortmund and Hochschule St. Gallen, and earned his Dr.-Ing. from the University of Duisburg-Essen, focusing on strategic planning for sustainable logistics service providers. He has held leadership roles at Arvato, including Company Director at Arvato SCM UK Ltd., and served as Vice President at STOK Europe (Arvato Healthcare) from 2016 to 2022, specializing in point-of-care inventory management. He has taught at the University of Paderborn, University of Duisburg-Essen, and Hamm-Lippstadt University of Applied Sciences (HSHL), where he has been a Professor of Business Informatics since 2023. His academic interests include healthcare logistics, service innovation, and business process management.

JAN-NIKLAS VOIGT-ANTONS received the Dipl.-Psych. degree from Technische Universität Darmstadt, Germany, in 2008, and the Dr.-Ing. degree from Technische Universität Berlin, Germany, in 2014. He is a Professor of Applied Computer Science with a focus on immersive media at Hochschule Hamm-Lippstadt, Germany, where he is also the Director of the Immersive Reality Lab, and a Guest Researcher with Technische Universität Berlin. His research focuses on extended reality for healthcare, multimodal human–machine interaction, and quality-of-experience evaluation in immersive environments, integrating methods from computer science, psychology, and human–computer interaction to develop user-centered and impactful immersive technologies.

EDISON PIGNATON DE FREITAS has a Ph.D. in computer science and engineering from Halmstad University, Sweden (2011). He has been a Senior Member of the IEEE since 2021. His research is mainly focused on the following areas: computer networks, real-time systems, networked UAVs, and IoT.

• • •