Joint Optimization of Digital Semantic Communication and Radar Sensing for Enhanced ISAC

Ouwen Huan, *Graduate Student Member, IEEE*, Chuanhong Liu, Nuocheng Yang, Zhilong Zhang, *Member, IEEE*, Tao Luo, and Mingzhe Chen

***Abstract*—In this work, we propose a novel integrated sensing and communication (ISAC) framework for connected and autonomous vehicles (CAVs), which incorporates digital semantic communication (SemCom) to achieve both reliable communication and accurate sensing. Within this framework, the transmitting vehicle extracts semantic symbols from the source data and transmits them over orthogonal frequency division multiplexing (OFDM) sub-carriers, while simultaneously utilizing echo signals for radar-based environmental sensing. To achieve reliable SemCom, the transmitter must jointly optimize the quantization bitwidth for semantic symbols, the modulation order, the power allocation across semantic symbol dimensions, and the transmit beamforming strategy. These optimizations must also consider sensing performance, leading to a tradeoff between radar sensing and task-oriented SemCom. To address this joint optimization problem, we decompose it into three subproblems and develop corresponding solutions: 1) a hierarchical constrained proximal policy optimization (H-CPPO) algorithm to determine the quantization bitwidth, modulation order, and power allocation under frequency-flat channels, 2) a joint beamforming strategy to optimize the dual-function radar-SemCom transmit beamforming vector, and 3) a semantic importance-based signal-to-noise ratio (SNR) matching strategy that effectively adapts the optimal power allocation obtained under frequency-flat conditions to fading channels with random gains. Simulation results on a road image segmentation task show that, the proposed SemCom scheme achieves near-optimal segmentation accuracy while reducing radar beamforming error by up to 82% compared to the conventional digital system using the same quadrature phase shift keying (QPSK) modulation.**

***Index Terms*—Integrated sensing and communication, radar sensing, vehicular communication, machine learning.**

Received 3 December 2025; revised 20 March 2026; accepted 25 April 2026. Date of current version 11 May 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 62271065 and Grant U22B2001. The associate editor coordinating the review of this article and approving it for publication was H. Guo. *(Corresponding author: Tao Luo.)*

Ouwen Huan, Nuocheng Yang, Zhilong Zhang, and Tao Luo are with Beijing Laboratory of Advanced Information Network, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: ouwenh@bupt.edu.cn; yangnuocheng@bupt.edu.cn; zhangzhilong@bupt.edu.cn; tluo@bupt.edu.cn).

Chuanhong Liu is with Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: 2016_liuchuanhong@bupt.edu.cn).

Mingzhe Chen is with the Department of Electrical and Computer Engineering and the Frost Institute for Data Science and Computing, University of Miami, Coral Gables, FL 33146 USA (e-mail: mingzhe.chen@miami.edu).

Digital Object Identifier 10.1109/TWC.2026.3690165

I. Introduction

FUTURE sixth-generation (6G) communication systems are expected to support a variety of intelligent applications such as vehicle-to-everything (V2X) communication, autonomous driving, and smart cities, that require not only massive connectivity but also high-precision sensing capabilities [1], [2], [3], [4]. To address these demands, integrated sensing and communication (ISAC) has emerged as a key enabling technology for 6G. ISAC allows the dual use of radio signals and infrastructure for both communication and sensing functions. However, the inherent tradeoff between communication and sensing performance imposes fundamental limitations on both functionalities, presenting significant challenges for the design of efficient ISAC systems.

Recent research has extensively explored performance tradeoffs in ISAC system optimization. Studies such as [5] and [6] have investigated transmit beamforming designs, with [5] examining both separated and shared antenna deployments and formulating weighted beamforming optimizations under various constraints, while [6] proposed a joint beamforming model for multiple-input multiple-output (MIMO) radar-communication systems that optimizes radar performance while ensuring user signal-to-interference-plus-noise ratio (SINR) thresholds. However, these works primarily focus on beamforming strategies and overlook the impact of waveform randomness on overall performance. Another line of research, including [1], [7], [8], [9], [10], [11], and [12], has addressed ISAC waveform optimization to characterize the communication-sensing tradeoff. Reference [7] introduced the Cramér-Rao Bound (CRB)-rate region, revealing both subspace and deterministic-random tradeoffs. The authors in [1] extended this to multicast communication and multi-target sensing scenarios. In [8], the authors formulated a power allocation problem to minimize transmit power under detection probability and quality of service constraints, revealing the tradeoff between detection probability and achievable rate. Reference [9] established a capacity-Bayesian CRB framework for orthogonal frequency division multiplex (OFDM) systems, demonstrating the asymptotic optimality of Gaussian inputs. Reference [10] employed mutual information as a unified metric for multi-antenna multi-carrier systems, and [11] developed a deep learning-based approach for joint waveform and beamforming optimization. Reference [12] further investigated

joint waveform design and sub-carrier allocation in multi-user MIMO ISAC. Despite these contributions, existing waveform optimization methods remain incompatible with practical digital communication systems, as they assume arbitrary symbol distributions to approach theoretical bounds, whereas practical systems employ structured constellations such as phase shift keying (PSK) or quadrature amplitude modulation (QAM) constellations.

In contrast, [13], [14], [15], [16], [17], [18], [19], [20], [21] have incorporated practical system constraints into ISAC joint optimization. For instance, [13] proposed and experimentally validated waveform designs suitable for dual-functional data transmission and radar sensing. In [14], the authors introduced a sensing-integrated discrete Fourier transform spread OFDM (SI-DFT-s-OFDM) system for Terahertz ISAC, achieving a 5 dB bit error rate (BER) gain over conventional OFDM while maintaining millimeter-level range and decimeter-per-second-level velocity estimation accuracy. Reference [15] developed a minimum-BER precoder for orthogonal time frequency space (OTFS)-based ISAC under transmit power and sensing constraints. Theoretical analyses in [16] established OFDM as the globally optimal cyclic prefix (CP)-based waveform for minimizing ranging sidelobes with QAM/PSK constellations. Reference [17] demonstrated that symbol phase randomness has negligible impact on OFDM ambiguity functions, favoring PSK for sensing. This results is reinforced by [18], where PSK was shown to be mutual-information-optimal for radar. Reference [19] further validated that uniform PSK power allocation minimizes outlier probability in delay-Doppler estimation. However, while PSK is favorable for sensing, it is detrimental to reliable communication due to its higher BER compared to QAM. To address this, [20], [21] explored probabilistic constellation shaping (PCS) for dynamic tradeoffs. Reference [20] compared geometric, probabilistic and joint constellation shaping for OFDM-ISAC systems using an autoencoder framework. Reference [21] proposed a QAM-based PCS to enhance sensing performance. Nevertheless, PCS sensing performance inevitably degrades with increasing modulation order.

Driven by advances in deep learning, semantic communication (SemCom) has emerged as a game-changing technology for scenarios involving large-volume data transmission under constrained communication resources [22], [23]. As a promising enabler for 6G, SemCom focuses on conveying semantic features extracted from the source data, substantially reducing transmission overhead. By learning the properties of the wireless channel, it enhances transmission robustness and, supported by knowledge bases, ensures reliable task execution [24]. Due to its high efficiency and robustness, SemCom mitigates the tradeoff between sensing and communication in ISAC systems, achieving superior sensing performance while preserving downstream task accuracy comparable to conventional schemes. Despite this potential, existing research on integrating SemCom into ISAC systems has yet to fully explore how its capabilities can be harnessed to resolve the inherent tradeoff and improve overall system performance. For example, [25], [26], [27], [28] focus exclusively on enhancing security within ISAC frameworks. In [29], an efficient non-orthogonal multiple access (NOMA)-based integrated sensing and SemCom system was designed. The work in [30] introduces task-oriented semantic signal processing to reduce packet sizes and boost radar-centric ISAC performance. However, although both [29] and [30] address performance optimization in the context of SemCom integration, the optimization variables and system designs adopted in these works are not closely aligned with practical digital communication architectures, limiting their real-world applicability.

The main goal of this work is to design a novel ISAC framework incorporating digital SemCom technique. The key contributions include:

- We propose a novel ISAC framework for connected and autonomous vehicles (CAVs) that integrates digital SemCom. In this framework, the transmitting vehicle extracts semantic symbols from source data and transmits each dimension in parallel over OFDM subcarriers, allowing receivers to perform downstream tasks. The same OFDM signals are concurrently used for radar sensing via echo processing. To support reliable SemCom, the transmitter jointly optimizes quantization bits, beamforming strategy, and power allocation across semantic dimensions, accounting for both semantic importance and channel conditions. These decisions also influence sensing performance, creating a fundamental tradeoff between radar sensing and task-oriented SemCom.
- We formulate this challenge as an optimization problem that aims to jointly optimize beamforming and sub-carrier power allocation under constraints that ensure high SemCom performance, total transmit power limits, and transmission efficiency. To address this problem, we decompose it into three subproblems and develop corresponding solutions: a hierarchical constrained proximal policy optimization (H-CPPO) algorithm, a joint beamforming strategy, and a semantic importance-based signal-to-noise ratio (SNR) matching strategy.

Simulation results on a road image segmentation task show that, the proposed SemCom scheme achieves near-optimal segmentation accuracy while reducing radar beamforming error by up to 82% compared to the conventional digital system using the same quadrature phase shift keying (QPSK) modulation.

II. System Model

As shown in Fig. 1, we consider an OFDM-based millimeter-wave (mmWave) vehicular ISAC system, where a vehicle simultaneously performs radar sensing and multicast road information transmission using dual-functional OFDM waveforms. In particular, the vehicle is equipped with a uniform linear array (ULA) of N_{t} antennas for transmitting road information to surrounding vehicles, and an omni-directional antenna for receiving communication signals from other vehicles. For radar sensing, another N_{r}-antenna ULA at the vehicle is employed to capture the echoes of the dual-functional OFDM signals. With knowledge of the transmitted symbols, the vehicle estimates the sensing-relevant channel state information (CSI) from these echoes to extract parameters of surrounding targets. To jointly achieve high sensing accuracy

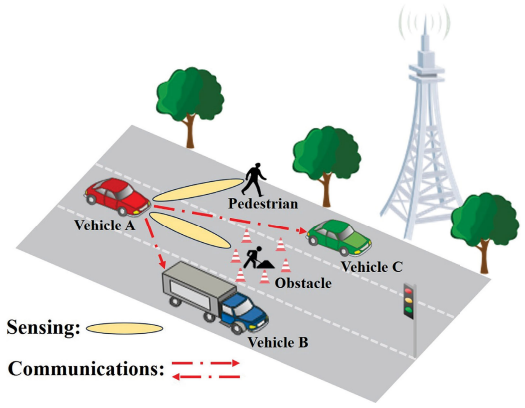


Fig. 1. Considered system.

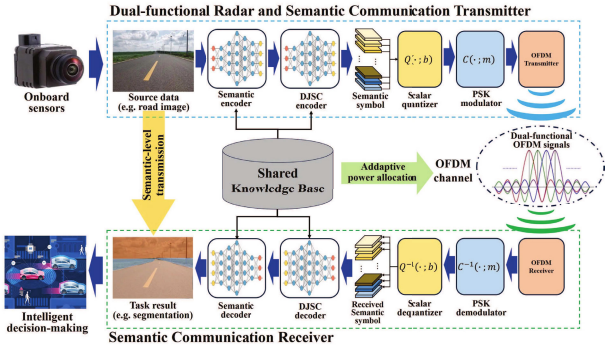


Fig. 2. Designed SemCom system.

and reliable communication, the vehicle adopts a task-oriented digital SemCom scheme for road information transmission. In the following, we first present the designed OFDM-based SemCom framework, and then introduce the OFDM radar sensing model. Finally, we explain the problem formulation.

A. Task-Oriented Digital SemCom

As illustrated in Fig. 2, the vehicle extracts semantic symbols from road information data, and multicasts it to surrounding vehicles. Each vehicle will decode the received semantic information, and executes the downstream intelligent task. Next, we will introduce the semantic transmitter and receiver models, and explain the complete workflow of our designed SemCom scheme.

1) Semantic Transmitter: The semantic transmitter consists of five key components: 1) a semantic encoder, 2) a deep joint source and channel (DJSC) encoder, 3) a scalar quantizer, 4) a digital modulator, and 5) an OFDM transmitter. For a given road information data sample $\boldsymbol{x} \in \mathcal{X}$ (e.g. a road image), a neural network (NN) based semantic encoder is used to extract task-oriented semantic information as $\boldsymbol{\psi} = \Psi(\boldsymbol{x}; \boldsymbol{\alpha})$, where $\boldsymbol{\psi}$ represents the semantic information matrix, $\boldsymbol{\alpha}$ denotes the semantic encoder parameters. To enhance noise robustness, the DJSC encoder transforms $\boldsymbol{\psi}$ into a transmittable semantic symbol, given by $\boldsymbol{\phi} = \Phi(\boldsymbol{\psi}; \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is the DJSC encoder parameters, and $\boldsymbol{\phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_{N_c}]$ is the semantic symbol that will be transmitted over N_c OFDM sub-carriers in parallel. To enhance compatibility with practical digital systems [24], each dimension of the analog semantic symbol $\boldsymbol{\phi}$ is quantized by a scalar quantizer as $\boldsymbol{q}_n = Q(\boldsymbol{\phi}_n; b)$, $n \in \{0, 1, \cdots, N_c - 1\}$, where $\boldsymbol{q}_n$ denotes the quantized bitstream of dimension n, and b is the number of quantized bits for each number in $\boldsymbol{\phi}_n$. Then, we perform PSK/QAM modulation on each $\boldsymbol{q}_n$ with a digital modulator. The obtained constellation symbol vector is $\boldsymbol{c}_n = C(\boldsymbol{q}_n; m) = [c_{n,1}, \cdots, c_{n,L}] \in \mathbb{C}^{1\times L}$, where m is the modulation order, and L denotes the number of constellation symbols modulated from $\boldsymbol{q}_n$. The OFDM transmitter then maps $\boldsymbol{c}_n$ onto sub-carrier n for transmission, and allocates power to each sub-carrier based on factors like channel condition and semantic importance [24]. Note that, in contrast to methods that discard less important semantic dimensions based on importance thresholds (e.g., [31], [32]), we transmit all dimensions in parallel. This design stems from two considerations. First, transmitting complete semantic symbols yields theoretically superior SemCom performance. Second, discarding semantic dimensions is typically adopted when the number of available subchannels is severely limited, trading a marginal loss in reliability for improved efficiency. However, this tradeoff is not the focus of this work, as OFDM-ISAC systems are generally configured with a sufficient number of subcarriers to ensure adequate range resolution for sensing [13]. The complex envelope of OFDM symbol l is thus given by

$$s_l(t) = \frac{1}{\sqrt{N_c}} \sum_{n=0}^{N_c-1} \sqrt{p_n} c_{n,l} e^{j2\pi n \Delta f t} \cdot \text{Rect}(t), \quad (1)$$

where p_n is the power allocated to sub-carrier n, Δf is the sub-carrier spacing, $\text{Rect}(\cdot)$ is a rectangular pulse, $\text{Rect}(t) = 1$ when $t \in [0, T_{\text{sym}}]$, and = 0 otherwise, and $T_{\text{sym}} = T + T_{\text{cp}}$ denotes the total symbol duration (elementary duration $T = \frac{1}{\Delta f}$ plus cyclic prefix T_{cp}). Let $\boldsymbol{w} \in \mathbb{C}^{N_t \times 1}$ be a joint radar-communication beamforming vector, the complete passband OFDM signal becomes

$$\boldsymbol{s}(t) = \text{Re}\left\{ \sum_{l=0}^{L-1} \boldsymbol{w} \cdot s_l(t - lT_{\text{sym}}) e^{j2\pi f_c t} \right\} \in \mathbb{R}^{N_t \times 1}, \quad (2)$$

with f_c being the carrier frequency.

2) Semantic Receiver: We assume that the number of receiving vehicle is V. The semantic receiver at each vehicle includes: 1) an OFDM receiver, 2) a digital demodulator, 3) a scalar de-quantizer, 4) a DJSC decoder, and 5) a semantic decoder. Specifically, the received constellation symbol l on sub-carrier n at receiving vehicle v is

$$\hat{c}_{v,n,l} = \left(\boldsymbol{h}_{v,n,l}\right)^{\text{T}} \boldsymbol{w} \sqrt{p_n} c_{n,l} + n_{v,n,l}, \quad (3)$$

where $\boldsymbol{h}_{v,n,l} \in \mathbb{C}^{N_t \times 1}$ denotes the communication CSI of received constellation symbol l on sub-carrier n, $n_{v,n,l} \sim \mathcal{CN}(0, \sigma_c^2)$ is an additive Gaussian noise (AWGN) with variance σ_c^2. Let $\beta_{v,n,l} = \left(\boldsymbol{h}_{v,n,l}\right)^{\text{T}} \boldsymbol{w}$ be the effective end-to-end channel gain for $c_{n,l}$, which can be estimated at the receiver and used to equalize the received symbols. Hence, given $\hat{c}_{v,n,l}$ and the estimated $\beta_{v,n,l}$, the constellation symbols are demodulated to obtain the received bitstream

$\hat{\boldsymbol{q}}_{v,n} = C^{-1}(\hat{\boldsymbol{c}}_{v,n}; m)$, where $C^{-1}(\cdot)$ denotes the demodulator, and $\hat{\boldsymbol{c}}_{v,n} = \left[\frac{\hat{c}_{v,n,1}}{\beta_{v,n,1}\sqrt{p_n}}, \cdots, \frac{\hat{c}_{v,n,L}}{\beta_{v,n,L}\sqrt{p_n}}\right]$. The bitstream is de-quantized to recover semantic symbols as $\hat{\phi}_{v,n} = Q^{-1}(\hat{\boldsymbol{q}}_{v,n}; b)$, where $\hat{\phi}_{v,n}$ is the semantic feature on dimension n recovered at vehicle v, and $Q^{-1}(\cdot)$ represents the scalar de-quantizer. The whole recovered semantic symbol is defined as $\hat{\boldsymbol{\phi}}_v = \left[\hat{\phi}_{v,1}, \hat{\phi}_{v,2}, \cdots, \hat{\phi}_{v,N_c}\right]^{\mathrm{T}}$. Finally, $\hat{\boldsymbol{\phi}}_v$ will be sequentially processed by the DJSC decoder and the semantic decoder to complete the intelligent task at receiving vehicle v, which can be expressed as $\hat{\boldsymbol{y}}_v = \Psi^{-1}\left[\Phi^{-1}\left(\hat{\boldsymbol{\phi}}_v; \boldsymbol{\mu}\right); \boldsymbol{\eta}\right]$, where $\Phi^{-1}(\cdot; \boldsymbol{\mu})$ and $\Psi^{-1}(\cdot; \boldsymbol{\eta})$ are the DJSC and semantic decoders with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$, respectively, and $\hat{\boldsymbol{y}}_v$ is the final task prediction.

B. OFDM Radar Sensing Model

We consider an OFDM-based radar sensing model that searches for spectral peaks on the delay–Doppler–angle grid based on the estimated echo CSI [13]. Specifically, the transmitted signal is first reflected by the surrounding sensing targets, and subsequently captured by the receiving ULA of the vehicle. The back-scattered symbol at antenna i is given by [33]

$$\boldsymbol{R}_i = \boldsymbol{H}_i \odot (\sqrt{\boldsymbol{p}} \cdot \boldsymbol{1}_L^{\mathrm{T}}) \odot \boldsymbol{C} + \boldsymbol{N}_i \in \mathbb{C}^{N_c \times L}, \tag{4}$$

where $\odot$ represents the matrices' element-wise product, $\boldsymbol{1}_L$ is an L-dimension all-one column vector, $\boldsymbol{H}_i$ denotes the echo CSI at antenna i, $\boldsymbol{C} \in \mathbb{C}^{N_c \times L}$ is the matrix of all transmitted constellation symbols $c_{n,l}$, $\boldsymbol{p} = [p_1, p_2 \cdots, p_{N_c}]^{\mathrm{T}}$ includes the power allocated to each sub-carrier, $\boldsymbol{N}_i \sim \mathcal{CN}(0, \sigma_s^2)$ denotes the i.i.d. circularly symmetric Gaussian matrix. We assume that each sensing target is a point-like scatterer, and thus, $\boldsymbol{H}_i$ is [2]

$$\boldsymbol{H}_i = \sum_{u=0}^{U-1} \alpha_u [\boldsymbol{a}_t(\theta_u)\boldsymbol{w}] \boldsymbol{a}_{\mathrm{del}}(\tau_u) \boldsymbol{a}_{\mathrm{dop}}^{\mathrm{H}}(\nu_u) [\boldsymbol{a}_r(\theta_u)]_i, \tag{5}$$

where U is the number of sensing targets, α_u is the complex channel gain which integrates the effect of path attenuation and radar cross-section (RCS), $\tau_u = 2d_u/C$ and $\nu_u = 2v_u/\lambda_c$ respectively denote the round-trip delay and Doppler shift, with d_u, v_u, C, λ_c being the distance and radial velocity of target u, the speed of light, and the wavelength of carrier. θ_u denotes the angle of arrival (AoA) of target u, $\boldsymbol{a}_t(\theta)$ is the steering vectors of the transmit and receive ULAs [3] with an antenna spacing $d = \lambda_c/2$, and $[\boldsymbol{a}_t(\theta)]_i$ denotes element i of the receive steering vector. $\boldsymbol{a}_{\mathrm{del}}(\tau) \in \mathbb{C}^{N_c \times 1}$ and $\boldsymbol{a}_{\mathrm{dop}}(\nu) \in \mathbb{C}^{K \times 1}$ are respectively defined as

$$\boldsymbol{a}_{\mathrm{del}}(\tau) = \left[1, \cdots, e^{-j2\pi n \Delta f \tau}, \cdots, e^{-j2\pi(N_c-1)\Delta f \tau}\right]^{\mathrm{T}}, \tag{6}$$

$$\boldsymbol{a}_{\mathrm{dop}}(\nu) = \left[1, \cdots, e^{-j2\pi l T_{\mathrm{sym}} \nu}, \cdots, e^{-j2\pi(L-1)T_{\mathrm{sym}}\nu}\right]^{\mathrm{T}}. \tag{7}$$

The goal of radar sensing is to detect each target and estimate their parameters $\{\tau_u, \nu_u, \theta_u\}_{u=1}^{U}$, thereby obtaining precise surrounding obstacle information. The complete procedure to achieve this goal typically consists of two main steps [13], [33], as detailed bellow:

1) Delay-Doppler Image Calculation: To compute the delay-Doppler image, we first estimate the echo CSI $\boldsymbol{H}_i$. Specifically, since $\boldsymbol{C}$ is known to the transmitting vehicle, $\boldsymbol{H}_i$ is estimated using the least square (LS) algorithm as

$$\begin{aligned}\hat{\boldsymbol{H}}_i &= \boldsymbol{R}_i \oslash \left[(\sqrt{\boldsymbol{p}} \cdot \boldsymbol{1}_L^{\mathrm{T}}) \odot \boldsymbol{C}\right] \\ &= \boldsymbol{H}_i + \boldsymbol{N}_i \oslash \left[(\sqrt{\boldsymbol{p}} \cdot \boldsymbol{1}_L^{\mathrm{T}}) \odot \boldsymbol{C}\right].\end{aligned} \tag{8}$$

where $\oslash$ is the element-wise division. Then, the delay-Doppler image at antenna i can be obtained by performing discrete Fourier transformation on $\hat{\boldsymbol{H}}_i$ along delay and Doppler axes, which can be expressed as

$$\Lambda_i(\tau, \nu) = \boldsymbol{a}_{\mathrm{del}}^{\mathrm{H}}(\tau) \hat{\boldsymbol{H}}_i \boldsymbol{a}_{\mathrm{dop}}(\nu). \tag{9}$$

2) Delay-Doppler-Angle Grid Search: Given the delay-Doppler image in (9), we next introduce the pipeline of delay-Doppler-angle grid search. Specifically, we first integrate the delay-Doppler image over all antenna elements by

$$|\Lambda(\tau, \nu)|^2 = \sum_{i=0}^{N_r-1} |\Lambda_i(\tau, \nu)|^2. \tag{10}$$

Based on (10), multi-target detection and delay-Doppler estimation can be performed by searching for peaks in a τ and ν grid [33]. Then, for each detected target u with delay-Doppler bin $(\hat{\tau}_u, \hat{\nu}_u)$, we substitute $(\hat{\tau}_u, \hat{\nu}_u)$ into (9) to obtain its spatial domain observation vector $\boldsymbol{g}_u = [\Lambda_0(\hat{\tau}_u, \hat{\nu}_u), \cdots \Lambda_i(\hat{\tau}_u, \hat{\nu}_u), \cdots, \Lambda_{N_r-1}(\hat{\tau}_u, \hat{\nu}_u)]$. Finally, the AoA of target u can be estimated via the 1-D Multiple Signal Classification (MUSIC) algorithm [34] applied to $\boldsymbol{g}_u$.

C. Problem Formulation

Given the defined system model, our goal is to minimize the sensing performance loss $F_s(b, m, \boldsymbol{w}, \boldsymbol{p})$, while ensuring sufficiently reliable SemCom by constraining the task-specific loss $L_t(\boldsymbol{y}, \hat{\boldsymbol{y}}_v)$ below a predefined threshold. The problem involves the joint optimization of OFDM parameters $b, m, \boldsymbol{p}$ and the dual-functional beamforming vector $\boldsymbol{w}$, subject to constraints on transmission efficiency and resource availability. Accordingly, the optimization problem is formulated as

$$\min_{b,m,\boldsymbol{w},\boldsymbol{p}} \quad F_s(b, m, \boldsymbol{w}, \boldsymbol{p}) \tag{11}$$

$$\text{s.t.} \quad L_t(\boldsymbol{y}, \hat{\boldsymbol{y}}_v) \le D_{\max}, v \in \{0, \cdots, V-1\}, \tag{11a}$$

$$\frac{b}{m} \le T_{\max}, \ m \in \{1, \cdots, b\}, \tag{11b}$$

$$\sum_{n=0}^{N_c-1} p_n \le P_t, \ p_n \in (0, P_t), \tag{11c}$$

$$|[\boldsymbol{w}]_i|^2 = \frac{1}{N_t}, \ i \in \{0, \cdots, N_t - 1\}, \tag{11d}$$

where (11a) ensures that the task losses at all receiving vehicles do not exceed a threshold $D_{\max}$, (11b) constrains the average transmission delay for each number of the semantic symbol not to be greater than $T_{\max}$ symbol periods, which reflects the requirements for efficiency, (11c) limits the total power of each OFDM symbol to be no more than P_t, (11d) guarantees uniform transmission power across each

antenna, which aligns with the general constraints of analog beamforming.

Problem (11) is challenging to solve using conventional methods for several reasons. First, the objectives for both sensing and SemCom involve a combination of discrete and continuous optimization variables. Second, since the predicted label $\hat{\boldsymbol{y}}_v$ depends on neural network-based codecs, the relationship between SemCom performance and the optimization variables lacks an explicit mathematical formulation. To address these challenges, we first derive a concrete sensing objective $F_{\mathrm{s}}(b, m, \boldsymbol{w}, \boldsymbol{p})$ by analyzing radar sensing performance, which clarifies the inherent tradeoff between digital SemCom and radar sensing. We then introduce a decomposition procedure for problem (11) and propose a reinforcement learning (RL)-based solution framework. This framework enables the transmitter to dynamically optimize the variables according to semantic importance and channel state information, thereby simultaneously satisfying the sensing and SemCom performance requirements for multicast receivers.

III. Sensing Performance Evaluation

Next, we introduce the metrics for sensing performance evaluation. From (8) to (10), it can be observed that multi-target detection and delay-Doppler-angle estimation depend on two key factors: 1) the transmit beamforming vector $\boldsymbol{w}$, which determines the power radiated toward the target directions and the power wasted in unintended directions, and 2) the additive noise in the delay-Doppler image, which affects the accuracy of peak detection over the delay-Doppler-angle grids. Since we consider a general scenario where no prior knowledge of target locations is available, we do not adopt the CRB as the sensing metric. Instead, we select sensing performance metrics from two perspectives that do not rely on such prior knowledge: 1) the beam pattern matching error, which characterizes the quality of transmit beamforming, and 2) the noise level in the delay-Doppler image, which reflects the reliability of peak detection. These metrics collectively capture the sensing capability without assuming knowledge of target parameters. Specifically, the matching error between the transmit beam pattern and an ideal reference radar beam pattern $P_{\mathrm{d}}(\theta)$ is defined as [35]

$$f_{\mathrm{s}}(\alpha, \boldsymbol{w}) = \sum_{i=0}^{N_\theta - 1} |\alpha P_{\mathrm{d}}(\tilde{\theta}_i) - \boldsymbol{a}_{\mathrm{t}}^{\mathrm{H}}(\tilde{\theta}_i)\boldsymbol{w}\boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}_{\mathrm{t}}(\tilde{\theta}_i)|^2, \quad (12)$$

where α is a scaling parameter, and $\tilde{\theta}_i$ denotes the sampled direction i from the radar's detection range, with N_θ being the number of direction samples. Meanwhile, the additive noise in delay-Doppler image $\Lambda_i(\tau, \nu)$ is given by

$$\begin{aligned}\widetilde{N}_i &= \boldsymbol{a}_{\mathrm{del}}^{\mathrm{H}}(\tau)\hat{\boldsymbol{H}}_i\boldsymbol{a}_{\mathrm{dop}}(\nu) - \boldsymbol{a}_{\mathrm{del}}^{\mathrm{H}}(\tau)\boldsymbol{H}_i\boldsymbol{a}_{\mathrm{dop}}(\nu) \\ &= \boldsymbol{a}_{\mathrm{del}}^{\mathrm{H}}(\tau)\left[\boldsymbol{N}_i \oslash \left((\sqrt{\boldsymbol{p}} \cdot \mathbf{1}_L^{\mathrm{T}}) \odot \boldsymbol{C}\right)\right]\boldsymbol{a}_{\mathrm{dop}}(\nu). \end{aligned} \quad (13)$$

Due to its stochastic nature, $\widetilde{N}_i$ can degrade the sensing performance. We therefore utilize the power of $\widetilde{N}_i$ to measure its influence. Specifically, for the practically used PSK and QAM modulations, the noise power are presented by the following lemma.

Lemma 1: Assuming each $c_{n,l}$ is independent and identically distributed with normalized power (i.e., $\mathbb{E}(|c_{n,l}|^2) = 1$), the power of $\widetilde{N}_i$ for PSK and QAM constellations with a given modulation order m can be respectively expressed as

$$\begin{aligned}\mathbb{E}_{\mathrm{PSK}}(|\widetilde{N}_i|^2) &= \sigma_{\mathrm{s}}^2 L \left(\sum_n \frac{1}{p_n}\right), \\ \mathbb{E}_{\mathrm{QAM}}(|\widetilde{N}_i|^2) &= \frac{8(M-1)}{3M}\left(\sum_n \frac{1}{p_n}\right) \\ &\quad \cdot \left[\sum_{p=1}^{\frac{\sqrt{M}}{2}}\sum_{q=1}^{\frac{\sqrt{M}}{2}} \frac{\sigma_{\mathrm{s}}^2 L}{(2p-1)^2 + (2q-1)^2}\right], \end{aligned} \quad (14)$$

where $M = 2^m$ denotes the size of constellation.

Proof: See Appendix A.

Lemma 1 shows that $\mathbb{E}(|\widetilde{N}_i|^2)$ depends on the power allocation across sub-carriers and the specific constellation pattern. Moreover, we observe that the noise power of PSK remains independent of the modulation order m, whereas that of QAM increases with larger m. When $m = 2$, $\mathbb{E}_{\mathrm{QAM}}(|\widetilde{N}_i|^2)$ reaches its minimum value and equals $\mathbb{E}_{\mathrm{PSK}}(|\widetilde{N}_i|^2)$. This implies that in OFDM-based radar systems, PSK offers superior sensing performance, despite its higher bit error rate in conventional communication settings. However, thanks to the high efficiency and robustness of SemCom, the BER limitation of high-order PSK can be effectively alleviated. Therefore, with the support of SemCom, we adopt PSK modulation to achieve enhanced sensing performance. Since $\mathbb{E}_{\mathrm{PSK}}(|\widetilde{N}_i|^2)$ depends on the actual value of system parameters, we further normalize it to better serve as an optimization objective, yielding:

$$g_{\mathrm{s}}(\boldsymbol{p}) = \frac{P_{\mathrm{t}}}{N_{\mathrm{c}}} \sum_{n=0}^{N_{\mathrm{c}}-1} \frac{1}{p_n}. \quad (15)$$

Given (12) and (15), the sensing objective $F_{\mathrm{s}}(b, m, \boldsymbol{w}, \boldsymbol{p})$ is defined as the weighted sum of $f_{\mathrm{s}}(\alpha, \boldsymbol{w})$ and $g_{\mathrm{s}}(\boldsymbol{p})$. Thus, problem (11) can be reformulated as

$$\begin{aligned}\min_{\alpha, b, m, \boldsymbol{w}, \boldsymbol{p}} \ & \frac{1}{N_\theta} f_{\mathrm{s}}(\alpha, \boldsymbol{w}) + \rho\left[\frac{1}{N_{\mathrm{c}}} g_{\mathrm{s}}(\boldsymbol{p})\right] \quad (16) \\ \text{s.t.} \ & \alpha > 0, \\ & \text{(11a), (11b), (11c), and (11d),} \quad (16a)\end{aligned}$$

where $\rho > 0$ is a weight parameter, (16a) ensures the scaling parameter α to be a positive number. (16) exhibits a two-fold tradeoff between performance of radar sensing and SemCom. First, an optimal radar beam pattern may degrade the communication SNR. In the considered system, to ensure low transmission errors, a lower modulation order m is required. Due to the transmission efficiency constraint in (11b), a reduction in m further limits the quantization bitwidth b, thereby impairing the quantization accuracy of semantic symbols and ultimately degrading SemCom performance. Second, in SemCom, different semantic symbols or different dimensions of the same symbol often carry varying levels of semantic importance [24]. As a result, optimal resource allocation, such as power distribution, tends to prioritize more important dimensions. This leads to a situation where the

power allocation $\boldsymbol{p}$ that favors SemCom performance may conflict with the optimization of $g_{\mathrm{s}}(\boldsymbol{p})$. Given that (16) is intractable for conventional optimization algorithms, we next introduce the problem decomposition approach to simplify the original formulation, followed by detailed explanation of the complete solution pipeline.

IV. Proposed Algorithms

In Problem (16), the beamforming vector $\boldsymbol{w}$ governs the received SNR at each vehicle. However, even with identical received SNRs and the same $b, m, \boldsymbol{p}$, the SemCom performance may vary significantly across vehicles due to statistically independent sub-channel gains. As a result, directly solving (16) would require characterizing the relationship between all possible sub-channel gain distributions and the non-convex SemCom performance metric $L_{\mathrm{t}}(\boldsymbol{y}, \hat{\boldsymbol{y}}_v)$, which poses substantial challenges. To address this, we decompose (16) into two distinct channel regimes: frequency-flat channels with uniform sub-channel gains, and frequency-selective channels with random sub-channel gains. Under the frequency-flat assumption, the problem is further divided into two subproblems. First, for a given SNR Γ, we optimize the sensing metric g_{s} using a proposed Hierarchical Constrained Proximal Policy Optimization (H-CPPO) algorithm, which establishes a functional mapping $G(\Gamma)$ between Γ and the optimal g_{s}. This decoupling of f_{s} and g_{s} significantly reduces the training complexity of H-CPPO. Moreover, with the aid of $G(\Gamma)$, the original non-convex problem can be reformulated into a conventional joint beamforming form that is tractable via standard optimization methods. For frequency-selective channels, we introduce an efficient SNR matching strategy that adapts the optimal power allocation $\boldsymbol{p}$, originally derived for flat channels, to varying channel realizations. This adaptation is formulated as a convex optimization problem, enabling efficient solution using standard tools. In the following sections, we first present the problem decomposition and the H-CPPO framework for frequency-flat channels, then detail the SNR matching strategy for frequency-selective channels, and finally analyze the computational complexity of the overall approach.

A. Problem Decomposition Under Frequency-Flat Channels

The beamforming design affects communication performance via the receiving SNR. We therefore derive the optimal $g_{\mathrm{s}}(\boldsymbol{p})$ for a given SNR $\Gamma = \frac{|\beta|^2 P_{\mathrm{t}}}{N_{\mathrm{c}}\sigma_{\mathrm{c}}^2}$ with uniform gain $|\beta|^2$ across all sub-carrier channels. This problem can be formulated as

$$\min_{b,m,\boldsymbol{p}} \ \frac{1}{N_{\mathrm{c}}} g_{\mathrm{s}}(\boldsymbol{p}) \tag{17}$$

$$\text{s.t.} \ \ L_{\mathrm{t}}(\boldsymbol{y}, \hat{\boldsymbol{y}}) \le D_{\max}, \quad (11\text{b}), \text{ and } (11\text{c}). \tag{17a}$$

Leveraging the performance mapping function $G(\Gamma)$ between Γ and the optimal $\frac{g_{\mathrm{s}}}{N_{\mathrm{c}}}$ of (17), we can transform (16) into an equivalent joint beamforming optimization problem, as formalized in the following Lemma.

Lemma 2: Given $G(\Gamma)$, the optimal α^*, $\boldsymbol{w}^*$ to problem (16) under frequency-flat channel conditions can be equivalent obtained by solving the following joint beamforming problem:

$$\min_{\alpha,\boldsymbol{w},\Gamma} \ \frac{1}{N_\theta} f_{\mathrm{s}}(\alpha, \boldsymbol{w}) + \rho \cdot G(\Gamma) \tag{18}$$

$$\text{s.t.} \ \ \frac{|(\boldsymbol{h}_{v,n,l})^{\mathrm{T}} \boldsymbol{w}|^2 P_{\mathrm{t}}}{N_{\mathrm{c}}\sigma_{\mathrm{c}}^2} \ge \Gamma, v \in \{0, \cdots, V-1\}, \tag{18a}$$

$$\Gamma \ \ge \ \Gamma_{\min}, \quad (11\text{d}), \text{ and } (16\text{a}), \tag{18b}$$

where (18a) guarantees a SNR threshold Γ for all receiving vehicles, (18b) defines the minimum SNR as $\Gamma_{\min}$. Subsequently, given the optimal solution Γ^* from (18), the optimal b^*, m^*, and $\boldsymbol{p}^*$ for (16) can be determined by solving problem (17) with $\Gamma = \Gamma^*$.

Proof: See Appendix B.

Lemma 2 enables us to decompose (16) into two simpler sub-problems (17) and (18), under frequency-flat channel conditions. Once $G(\Gamma)$ is approximated by convex functions, subproblem (18) becomes tractable to standard convex optimization method via semi-definite relaxation (SDR) [5] techniques. However, (17) remains challenging for conventional optimization approaches, as no analytical expression relates SemCom performance to the optimization variables. Additionally, constraints such as the total power limit complicate the use of heuristic algorithms like PPO [36] in balancing exploration, constraint satisfaction, and convergence stability. To overcome these challenges, we propose a Hierarchical Constrained Proximal Policy Optimization (H-CPPO) algorithm. Recognizing that $\boldsymbol{p}$ must be allocated per dimension, while b, m are determined globally for each transmission, we design a two-level hierarchical RL structure: a low-level policy allocates power $\boldsymbol{p}$ across semantic dimensions based on their semantic importance, and a high-level DQN-based policy efficiently selects the quantization bits b and PSK modulation order m. To ensure reliable constraint satisfaction without compromising training stability, we integrate First-Order Constrained Optimization in Policy Space (FOCOPS) [37] into the PPO framework, resulting in a Constrained PPO (CPPO) formulation for the low-level policy. In the following, we describe the key components and training process of H-CPPO, and then present the joint beamforming strategy derived from it for solving (18).

B. Components of the H-CPPO Algorithm

We now describe the components of H-CPPO. In the low-level CPPO, we utilize dedicated networks to separately estimate 1) the cost associated with constraint violations, and 2) the reward from objective optimization, which reduces advantage function approximation errors and enhances training stability. Moreover, the Lagrange multipliers are dynamically adjusted based on the magnitude of constraint violations, ensuring stricter compliance with constraints throughout the optimization process. For the high-level policy optimization, we adopt a DQN approach [38] to achieve efficient decision making.

1) Low-Level CPPO: **Agent:** The low-level agent is responsible for determining power p_n of each dimension according to its task-oriented semantic importance.

Action: At step n, the low-level action corresponds to allocating transmission power p_n to the feature map at dimension n. To handle the dynamic action space caused by diminishing remaining power, we impose a maximum power limit $p_{\max}$ per dimension, restricting each action to $p_n \in (0, p_{\max}]$.

State: The state is defined as $\boldsymbol{s}_n^{\text{low}} = \left[(i_n)^{\epsilon_s}, b, m, \frac{\bar{P}_n}{P_t}, \frac{n}{N_c}\right]$, where i_n denotes the semantic importance of the feature map at dimension n, ϵ_s is a parameter that scales the values of i_n. When $\epsilon_s > 1$, the gaps between i_n increase, and when $\epsilon_s \in [0, 1]$, the gaps decrease. $\bar{P}_n$ represents the remaining power at step n, b, m are the number of quantization bits and modulation order for the current episode, both being determined by the high-level agent.

Policy: The policy governs the decision of the agent through specifying the conditional probability distribution function (CPDF) over possible actions p_n given the current state $\boldsymbol{s}_n^{\text{low}}$. Specifically, we implement this low-level policy using an actor network with parameters $\boldsymbol{\omega}_{\text{low}}$, such that the CPDF can be expressed by $\pi_{\boldsymbol{\omega}_{\text{low}}}(p_n|\boldsymbol{s}_n)$.

Reward: We define the reward function for solving problem (17) as

$$R_{\text{obj}}\left(p_n, \boldsymbol{s}_n^{\text{low}}\right) = -\left(\frac{P_t}{p_n N_c} - 1\right)^2 + R_0, \tag{19}$$

where R_0 is a constant bias that ensures the reward to be greater than zero.

Cost: The CPPO algorithm decouples constraints from the optimization objective by modeling them as separate cost functions. Specifically, the performance cost for SemCom is designed as

$$C_{\text{sem}}(p_n, \boldsymbol{s}_n^{\text{low}}) = \begin{cases} (i_n)^{\epsilon_s}\|\boldsymbol{\phi}_n - \hat{\boldsymbol{\phi}}_n\|_F, \text{ if } n < N_c - 1, \\ (i_n)^{\epsilon_s}\|\boldsymbol{\phi}_n - \hat{\boldsymbol{\phi}}_n\|_F + \lambda_{\text{sem}}\left[L_t(\boldsymbol{y}, \hat{\boldsymbol{y}}) \right. \\ \qquad \left. - D_{\max}\right], \text{if } n = N_c - 1, \end{cases} \tag{20}$$

where $\lambda_{\text{sem}} > 0$ is a weight parameter that penalizes large task losses. The term $\|\boldsymbol{\phi}_n - \hat{\boldsymbol{\phi}}_n\|_F$ captures the combined effect of quantization and transmission distortions at step n. By serving as an immediate cost signal, it effectively mitigates the sparse reward problem arising from the delayed computation of $\hat{\boldsymbol{y}}$, which is only available after all semantic dimensions have been transmitted. Meanwhile, for the power constraint (11b), the cost function is defined as

$$C_{\text{pow}}\left(p_n, \boldsymbol{s}_n^{\text{low}}\right) = \begin{cases} \lambda_{\text{pow}} \cdot (-\bar{P}_n), & \text{if } \bar{P}_n < 0, \\ 0, & \text{else}, \end{cases} \tag{21}$$

where $\lambda_{\text{pow}} > 0$ is the weight parameter that controls the penalty for excessive power allocation.

State-value function: The state-value function estimates the expected cumulative rewards or costs starting from a given low-level state $\boldsymbol{s}_n^{\text{low}}$. In our CPPO framework, we employ three distinct critic networks to approximate the value functions for the reward R_{obj} and constraint costs $C_{\text{sem}}, C_{\text{pow}}$, denoted as $V_{\boldsymbol{\omega}_i}(\boldsymbol{s}_n^{\text{low}}), i \in \{\text{obj}, \text{sem}, \text{pow}\}$, where $\boldsymbol{\omega}_{\text{obj}}$,$\boldsymbol{\omega}_{\text{sem}}$, and $\boldsymbol{\omega}_{\text{pow}}$ represent trainable parameters of the respective critic networks.

2) High-Level DQN: **Agent:** The high-level agent determines b and m for the transmission of one road information data.

Action: The high-level action is defined as (b, m), where both parameters remain fixed during each low-level RL episode. In particular, for a given b, the modulation order m can be optimally determined through constraint (11b) as $m = \lceil \frac{b}{T_{\max}} \rceil$, since lower modulation orders inherently enhance transmission reliability by reducing bit error rates. This analytical relationship allows the high-level neural network to focus exclusively on optimizing b, while m is derived as a deterministic function of b. To reduce the decision space, we also define a maximum bit number $b_{\max}$ for each source data, such that $b \in \{1, 2, \cdots, b_{\max}\}$.

State: The high-level state is defined as $\boldsymbol{s}^{\text{high}} = [H(\boldsymbol{\phi}), \tilde{L}_t, \tilde{g}_s, \tilde{L}_c]$, where $H(\boldsymbol{\phi})$ represents the binned entropy of semantic symbol, $\tilde{L}_t, \tilde{g}_s$ respectively quantify the SemCom task loss L_t and sensing objective g_s in the last RL episode, and $\tilde{L}_c = \frac{1}{N_c}\sum_n (i_n)^{\epsilon_s}\|\boldsymbol{\phi}_n - \hat{\boldsymbol{\phi}}_n\|_F$ measures the cumulative quantization and transmission distortions.

Policy: The high-level policy $\pi_{\text{high}}\left(b|\boldsymbol{s}^{\text{high}}\right)$ characterizes the CPDF of quantization bit number b given the high-level state $\boldsymbol{s}^{\text{high}}$.

Reward: The high-level reward function is designed to guide the agent toward optimal sensing and SemCom performance, which is defined as

$$R_{\text{high}}\left(b, \boldsymbol{s}^{\text{high}}\right) = -\sum_{n=0}^{N_c-1}\left(\frac{P_t}{p_n N_c} - 1\right)^2 - \lambda_t[L_t(\boldsymbol{y}, \hat{\boldsymbol{y}}) - D_{\max}] + R_1 \tag{22}$$

where λ_t is a weight parameter, and R_1 represents the constant reward bias.

Action-value function: The action-value function estimates the expected cumulative reward when executing a particular action in a given state. In our framework, we parameterize the high-level action-value function using a Q-network denoted as $Q_{\boldsymbol{\omega}_{\text{high}}}(\boldsymbol{s}^{\text{high}}, b)$, where $\boldsymbol{\omega}_{\text{high}}$ represents the trainable network parameters. This Q-function approximation enables the derivation of the high-level DQN policy, which selects actions through a greedy operation given by

$$\pi_{\text{high}}\left(b|\boldsymbol{s}^{\text{high}}\right) = \delta\left[b - \arg\max_{b'} Q_{\boldsymbol{\omega}_{\text{high}}}(\boldsymbol{s}^{\text{high}}, b')\right]. \tag{23}$$

C. Training Process of H-CPPO

We now introduce the H-CPPO training procedure for solving (17). Our framework simultaneously maximizes cumulative rewards while enforcing cumulative cost constraints through three key stages: 1) trajectory collection, 2) low-level policy updates, and 3) high-level policy updates, which are specified as follows.

1) Trajectory Collection: In the trajectory collection phase, the high-level agent initially samples a source data sample $\boldsymbol{x} \in \mathcal{X}$ and determines (b, m) through policy $\pi_{\boldsymbol{\omega}_{\text{high}}}$. Subsequently, the low-level agent generates a sequence of transitions $\mathcal{T} = \{\boldsymbol{\tau}_0, \cdots, \boldsymbol{\tau}_n, \cdots, \boldsymbol{\tau}_{N_c-1}\}$ according to policy $\pi_{\boldsymbol{\omega}_{\text{low}}}$, where each transition $\boldsymbol{\tau}_n = \left[\boldsymbol{s}_n^{\text{low}}, b, m, p_n, R_{\text{low}}, C_{\text{sem}}, C_{\text{pow}}\right]$

captures the complete state, action, reward, and cost information at step n.

2) Low-Level Policy Update: During the low-level update stage, we optimize both actor and critic networks through constrained policy gradient methods. We first calculate the temporal difference (TD) errors $\Delta_n^i, i \in \{\text{obj}, \text{sem}, \text{pow}\}$ for the reward and cost functions [37], with a discount factor γ. The advantage functions of reward and costs are then estimated as

$$\hat{A}_n^i = \Delta_n + (\gamma\lambda)\,\Delta_{n+1}^i + \cdots + (\gamma\lambda)^{N_\text{c}-n+1}\,\Delta_{N_\text{c}-1}^i, \quad i \in \{\text{obj}, \text{sem}, \text{pow}\}, \tag{24}$$

with λ being the variance discount factor. Based on (24), the composite advantage function that combines these estimates is given as

$$\hat{A}_n = \hat{A}_n^{\text{obj}} - \nu_{\text{sem}}\hat{A}_n^{\text{sem}} - \nu_{\text{pow}}\hat{A}_n^{\text{pow}}, \tag{25}$$

where $\nu_{\text{sem}}, \nu_{\text{pow}}$ are adaptive Lagrangian multipliers. In contrast to FOCOPS [37] that uses KL divergence to explicitly enforce trust region constraints, our CPPO employs PPO's clipping mechanism for the trust region constraint approximation, achieving comparable policy stability while significantly simplifying the algorithm implementation. Specifically, the clipping loss of low-level policy is

$$L^{\text{low}}(\boldsymbol{\omega}_{\text{low}}) = -\hat{E}_n\big[\min\big(\rho(\boldsymbol{\omega}_{\text{low}})\,\hat{A}_n, \text{clip}\big(\rho(\boldsymbol{\omega}_{\text{low}}), 1-\epsilon, 1+\epsilon\big)\hat{A}_n\big)\big], \tag{26}$$

where $\rho(\boldsymbol{\omega}_{\text{low}}) = \frac{\pi_{\boldsymbol{\omega}_{\text{low}}}(p_n|\boldsymbol{s}_n^{\text{low}})}{\pi_{\tilde{\boldsymbol{\omega}}_{\text{low}}}(p_n|\boldsymbol{s}_n^{\text{low}})}$ denotes the probability ratio between new and old policies, with $\tilde{\boldsymbol{\omega}}_{\text{low}}$ being the old policy parameters, and clip $(\cdot, 1-\epsilon, 1+\epsilon)$ is an ϵ-clipping function that limits the value of $\rho(\boldsymbol{\omega}_{\text{low}})$ within range $[1-\epsilon, 1+\epsilon]$. The critic losses are given by

$$L^i(\boldsymbol{\omega}_i) = \frac{1}{N_\text{C}}\sum_{n=0}^{N_\text{C}-1}\left(\Delta_n^i\right)^2, \; i \in \{\text{obj}, \text{sem}, \text{pow}\}. \tag{27}$$

Based on (26) and (27), the network parameters are updated via gradient descent method [39] with learning rate ξ. To ensure robust constraint compliant, the Lagrangian multipliers are also adaptively tuned using sub-gradient scheme, which are expressed as

$$\begin{aligned}\nu_{\text{sem}} &\leftarrow \text{clip}\left(\nu_{\text{sem}}^{\min}, \nu_{\text{sem}}^{\max}, \nu_{\text{sem}} + \eta_{\text{sem}}\left[L_\text{t}(\boldsymbol{y}, \hat{\boldsymbol{y}}) - D_{\max}\right]\right),\\ \nu_{\text{pow}} &\leftarrow \text{clip}\left(\nu_{\text{pow}}^{\min}, \nu_{\text{pow}}^{\max}, \nu_{\text{pow}} - \eta_{\text{pow}}\bar{P}_{N_\text{c}-1}\right),\end{aligned} \tag{28}$$

where η_{sem} and η_{pow} are adaptive step sizes, $[\nu_{\text{sem}}^{\min}, \nu_{\text{sem}}^{\max}]$ and $[\nu_{\text{pow}}^{\min}, \nu_{\text{pow}}^{\max}]$ are the ranges of ν_{sem} and ν_{pow}, respectively. This automatic adjustment prioritizes constraint satisfaction by increasing weights for violated constraints during policy updates.

3) High-Level Policy Update: The high-level policy update stage optimizes the selection of b and m to simultaneously satisfy SemCom constraints and maximize sensing performance. For computational efficiency, we implement this using a DQN framework. The loss function for the high-level Q-network is derived from the reward function in (22), given by

$$L^{\text{high}}(\boldsymbol{\omega}_{\text{high}}) = \Big[R_{\text{high}}\left(\tilde{b}, \tilde{\boldsymbol{s}}^{\text{high}}\right) + \gamma \cdot \max_b Q_{\boldsymbol{\omega}_{\text{high}}}(\boldsymbol{s}^{\text{high}}, b) - Q_{\boldsymbol{\omega}_{\text{high}}}(\tilde{\boldsymbol{s}}^{\text{high}}, \tilde{b})\Big]^2, \tag{29}$$

where $\tilde{\boldsymbol{s}}^{\text{high}}$ represents the high-level state from the previous episode. The Q-network parameters are then updated via gradient descent [40], [41].

Algorithm 1 Training Process of H-CPPO

Initialize: Model parameters $\boldsymbol{\omega}_{\text{low}}$, $\boldsymbol{\omega}_{\text{obj}}$, $\boldsymbol{\omega}_{\text{sem}}$, $\boldsymbol{\omega}_{\text{pow}}$, $\boldsymbol{\omega}_{\text{high}}$, training iteration N_{iter}, maximum task loss $D_{\max}$, total transmission power P_t, symbol period upper bound $T_{\max}$, training SNR Γ, pretrained semantic and DSJC codecs, other hyper-parameters for training.
Input: Task-oriented dataset $(\mathcal{X}, \mathcal{Y})$.
for $k = 0 \to N_{\text{iter}} - 1$ **do**
  Sample source data $\boldsymbol{x} \in \mathcal{X}$, ground truth label $\boldsymbol{y} \in \mathcal{Y}$.
  Determine b using π_{high}.
  Calculate m by $m = \lceil \frac{b}{T_{\max}} \rceil$.
  for $n = 0 \to N_\text{c} - 1$ **do**
    Collect $\boldsymbol{\tau}_n$ based on $\pi_{\boldsymbol{\omega}_{\text{low}}}$, and (19) to (21).
  end for
  Calculate R_{high} according to (22).
  for $n = 0 \to N_\text{c} - 1$ **do**
    Calculate L^{low} according to (26).
    Compute L^{obj}, L^{sem}, and L^{pow} based on (27).
    Update $\boldsymbol{\omega}_{\text{low}}, \boldsymbol{\omega}_{\text{obj}}, \boldsymbol{\omega}_{\text{sem}}, \boldsymbol{\omega}_{\text{pow}}$ using gradient descent method.
  end for
  Update $\nu_{\text{sem}}, \nu_{\text{pow}}$ by (28).
  Calculate L^{high} according to (29).
  Update $\boldsymbol{\omega}_{\text{high}}$ using gradient descent method.
end for

Through iterative updates of both hierarchical policies and adaptive cost weights, the algorithm converges to a locally optimal solution for joint quantization, modulation, and power allocation that maximizes sensing performance while guaranteeing SemCom requirements. The complete H-CPPO training procedure is summarized in Algorithm 1.

D. H-CPPO Based Joint Beamforming

Next, we present the solution to the joint beamforming problem (18) using the proposed H-CPPO framework. Assuming the transmitting vehicle has access to the complete task-oriented dataset $\mathcal{D}_\text{t} = (\mathcal{X}, \mathcal{Y})$, we approximate the performance mapping function $G(\Gamma)$ as follows. First, we uniformly sample N_{snr} SNR values from the interval $[\Gamma_{\min}, +\infty)$, denoted as $\{\Gamma_k | k = 0, \cdots, N_{\text{snr}} - 1\}$. For each Γ_k, we train a dedicated H-CPPO model and compute the corresponding worst-case sensing performance $\hat{g}_\text{s}(\boldsymbol{p}) = \max_{x \in \mathcal{X}} \hat{g}_\text{s}(\boldsymbol{p})$ over the entire dataset. The resulting pairs $(\Gamma_k, \hat{g}_\text{s}(\boldsymbol{p}))$ are then used to fit $G(\Gamma)$ via convex regression. Leveraging the convexity of $G(\Gamma)$, (18) can be approximately solved through SDR techniques [5]. The key reformulation involves

substituting the $\boldsymbol{w}$ with its covariance matrix $\boldsymbol{W} = \boldsymbol{w}\boldsymbol{w}^{\mathrm{H}}$ as the new optimization variable, and relaxing the rank-one constraint on $\boldsymbol{W}$. This transformation enables the problem to be efficiently solved using standard convex optimization tools (e.g. CVXPY toolbox [42]). The beamforming vector $\boldsymbol{w}^*$ is subsequently recovered from the relaxed covariance matrix $\boldsymbol{W}^*$ using the Gaussian randomization method [5]. Given the optimal SNR threshold Γ^*, the power allocation vector $\boldsymbol{p}^*$ is determined by selecting the low-level policy trained at the largest SNR value Γ_k that does not exceed Γ^*, i.e., $k = \arg\max_{k'} \Gamma_{k'}$, s.t. $\Gamma_{k'} \leq \Gamma^*$.

E. SNR Matching Under Frequency-Selective Channels

In practical scenarios, sub-channels corresponding to different multicast receivers often experience random and statistically independent gains due to frequency-selective fading. As a result, directly applying the power allocation vector $\boldsymbol{p}^*$ optimized under frequency-flat channel assumptions may lead to significant disparities in per-sub-channel SNRs across receivers, thereby causing substantial performance degradation. To enhance robustness under realistic channel conditions, we propose an efficient SNR matching strategy. This approach retains the optimal parameters $\alpha^*, \boldsymbol{w}^*, b^*, m^*$ obtained under the flat-channel assumption, while adaptively adjusting only the power allocation vector $\boldsymbol{p}^*$ to accommodate varying channel realizations. The objective is to align the actual worst-case SNR on each sub-channel with its reference value from the flat-channel scenario, weighted by semantic importance, thereby enhancing the worst-case semantic communication performance across all receivers. The corresponding optimization is formulated as follows:

$$\min_{\tilde{\boldsymbol{p}}} \sum_{n=0}^{N_{\mathrm{c}}-1} (i_n)^{\epsilon_{\mathrm{s}}} \left(\frac{\min_{v} |\beta_{v,n,l}|^2 \tilde{p}_n}{|\bar{\beta}_{v,l}|^2 p_n^*} - 1 \right)^2 \tag{30}$$

$$\text{s.t.} \quad \sum_{n=0}^{N_{\mathrm{c}}-1} \tilde{p}_n \leq P_{\mathrm{t}},\ \tilde{p}_n \in (0, P_{\mathrm{t}}), \tag{30a}$$

$$g_{\mathrm{s}}(\tilde{\boldsymbol{p}}) \leq (1+\epsilon_{\mathrm{p}}) g_{\mathrm{s}}(\boldsymbol{p}^*), \tag{30b}$$

where $|\bar{\beta}_{v,l}|^2 = \frac{1}{N_{\mathrm{c}}} \sum_{n=0}^{N_{\mathrm{c}}-1} |\beta_{v,n,l}|^2$ denotes the average channel gain, $\epsilon_{\mathrm{p}}>0$ is a tolerance parameter controlling the allowable margin for $g_{\mathrm{s}}(\tilde{\boldsymbol{p}})$. (30a) enforces the total transmit power per OFDM symbol to P_{t}, (30b) imposes an upper bound on $g_{\mathrm{s}}(\tilde{\boldsymbol{p}})$ post power adaptation. Given the convexity of both (30a) and (30b), we can efficiently solve (30) using standard convex optimization methods.

In summary, an approximate solution to the original optimization problem (16) can be obtained by sequentially solving subproblems (17), (18), and (30). The complete procedure for solving (16) is outlined in Algorithm 2.

Algorithm 2 Complete Process of Solving (16)

1: **Input:** SNR sample number N_{snr}, SNR range $[\Gamma_{\min}, +\infty)$.
2: **for** $k = 0 \to N_{\mathrm{snr}} - 1$ **do**
3: Sample Γ from $[\Gamma_{\min}, +\infty)$.
4: Train H-CPPO using **Algorithm 1**, and obtain the minimum $\frac{P_{\mathrm{t}}}{N_{\mathrm{c}}} g_{\mathrm{s}}(\boldsymbol{p})$
5: **end for**
6: Fit $G(\Gamma)$ with some convex functions.
7: Solve α^*, $\boldsymbol{W}^*$, and Γ^*.
8: Approximate $\boldsymbol{w}^*$ using Gaussian randomization based on $\boldsymbol{W}^*$.
9: Solve (17) using the trained H-CPPO model with $\Gamma = \Gamma^*$, to obtain $b^*, m^*, \boldsymbol{p}^*$.
10: Adjust the power allocation $\boldsymbol{p}^*$ by solving (30).

F. Complexity Analysis

Next, we analyze the computational complexity of the proposed algorithms for solving the joint radar-SemCom optimization problem (16). Specifically, the complexity of the H-CPPO algorithm can be decomposed into the low-level CPPO complexity and the high-level DQN complexity, where the CPPO complexity is primarily determined by the dimensions of low-level state and action spaces, the semantic symbol dimension, and the architectures of low-level actor and critic networks. Let L_{act} and L_{crt} denote the number of layers in the low-level actor and critic networks respectively, the CPPO complexity can be expressed as $\mathcal{O}\left(\sum_{n=0}^{N_{\mathrm{c}}-1} \mathrm{Dim}(\boldsymbol{s}_n^{\mathrm{low}}) \left(\prod_{l=1}^{L_{\mathrm{act}}} H_l^{\mathrm{act}} + 3\prod_{l=1}^{L_{\mathrm{crt}}} H_l^{\mathrm{crt}}\right)\right)$, where H_l^{act} and H_l^{crt} represent the number of neurons in layer l of the actor and critic networks. The DQN complexity depends on the dimensions of high-level state and action spaces and the Q-network structure. Since the high-level policy is executed only once per training episode while the low-level policy runs N_{c} times (with $N_{\mathrm{c}} \gg 1$), the computational complexity of the high-level DQN becomes negligible compared to that of CPPO, making the overall complexity of H-CPPO predominantly determined by the low-level CPPO. For the standard convex optimization tool used to solve (18) and (30), the computational complexities are given by $\mathcal{O}\left((N_{\mathrm{c}}^3 + V N_{\mathrm{c}}^2 + N_\theta N_{\mathrm{c}}^2)\sqrt{N_{\mathrm{c}}}\log(1/\epsilon_1)\right)$, and $\mathcal{O}\left(N_{\mathrm{c}}^{3.5}\log(1/\epsilon_2)\right)$, respectively, where ϵ_1 and ϵ_2 are the solution accuracy thresholds for the two subproblems.

V. Simulation Results and Analysis

This section evaluates the performance of the proposed integrated radar and SemCom scheme for road image segmentation. We begin by introducing the task-oriented dataset and detailing the architectures of the SemCom and DJSC codecs. The experimental setup is then specified, followed by a comprehensive performance analysis based on numerical results.

A. Experimental Setups

We employ the CamVid road image segmentation dataset [43] for task-oriented SemCom, which consists of images with a dimension of $3 \times 360 \times 480$ and provides segmentation labels for 11 driving-related categories. The SemCom performance is measured by pixel-level accuracy, with the codec model achieving a mean accuracy of 0.83 on the dataset. Accordingly, the performance threshold is set to 0.83 to represent a sufficiently high performance requirement for

TABLE I
SIMULATION PARAMETERS

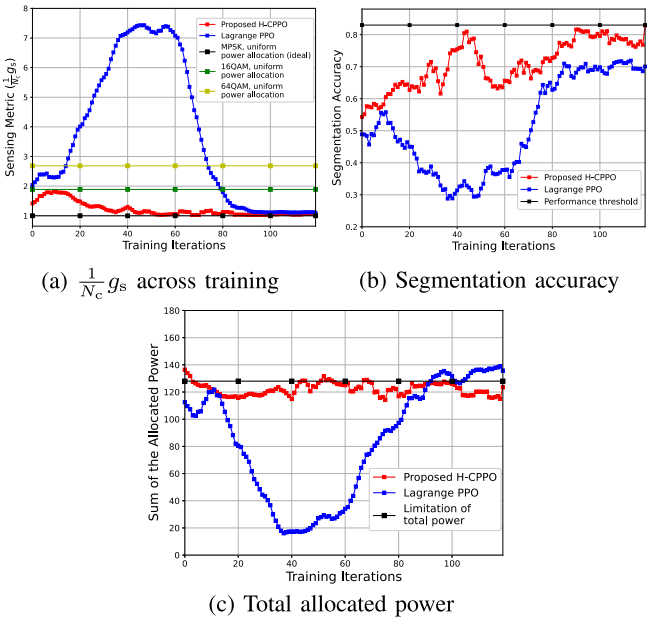
Parameter	value	Parameter	Value
N_c	128	f_c	28 GHz
Δf	167 kHz	N_t, N_r	16
T, T_{cp}	6 μs, 1.5 μs	D_{max}	0.83
T_{max}	2	P_t	20 dBm
$\sigma_c^2/\Delta f, \sigma_s^2/\Delta f$	-174 dBm/Hz	N_θ	181
ρ	2	ϵ_s	0.5
ϵ_p	$0.05, 0.1, 0.15$	p_{max}	2
b_{max}	5	γ, λ	0.95
ϵ	0.1	ξ	1e-4

SemCom. Road image segmentation is selected as the semantic task due to its direct relevance to cooperative perception in vehicular networks. In particular, onboard cameras are essential for CAV decision-making, and sharing image data via vehicle-to-vehicle (V2V) links extends individual perception to enable network-level awareness [44]. However, transmitting raw images with high fidelity incurs substantial overhead. Since segmentation outputs already provide critical information such as object categories and locations, it serves as an efficient and suitable downstream task for V2V SemCom.

The H-CPPO framework employs fully connected networks for the actor, critics, and Q-network. To approximate $G(\Gamma)$, RL models are trained at 3 dB intervals from–6 dB to 12 dB. The semantic importance i_n is computed based on its gradient contribution to the decoder output, following [24]:

$$i_n = \frac{1}{N_L \times N_H \times N_W} \sum_{l=0}^{N_L-1} \sum_{p=0}^{N_H-1} \sum_{q=0}^{N_W-1} \left| \frac{\partial \hat{y}_l}{\partial \phi_{n,p,q}} \right|, \quad (31)$$

where $\hat{y}_l$ is element l of a flatten segmentation label $\text{Vec}(\hat{\boldsymbol{y}})$, with N_L being element number of each segmentation label, N_H and N_W denote the height and width of $\boldsymbol{\phi}_n$, and $\phi_{n,p,q}$ is the value at row p, column q of $\boldsymbol{\phi}_n$. Intuitively, this metric captures how variations in each semantic symbol dimension affect downstream task performance. To better differentiate feature dimensions and enhance their distinctiveness, we statistically analyzed the semantic symbols across the entire dataset and computed the mean feature map per dimension. Singular value decomposition (SVD) was then applied to these averaged semantic symbols. The forward SVD matrix is used at the transmitter, and the inverse matrix is placed at the receiver. Accordingly, when computing semantic feature gradients, the inverse SVD matrix must also be applied to maintain consistency. Note that due to differences in performance metrics and the distribution of semantic importance across tasks, policies trained on one task may not be directly transferable to others. They typically require task-specific training or fine-tuning via transfer learning. Table I summarizes the training hyperparameters and the key parameters of the ISAC system used in the simulations. The system parameters are chosen to be in a similar range as those defined in the NR-V2X standard, as NR-V2X offers flexible sub-channel configurations and sidelink control information (SCI), which can effectively support the practical implementation of the proposed digital SemCom scheme.



(a) $\frac{1}{N_c}g_s$ across training (b) Segmentation accuracy

(c) Total allocated power

Fig. 3. Optimization metrics across training.

B. Performance Evaluation

Fig. 3 illustrates the variation of optimization metrics over training iterations at 0 dB SNR. For comparison, a Lagrange PPO algorithm serves as the baseline, using a reward function that weights sensing and SemCom objectives as well as power constraints with fixed Lagrangian multipliers, and a single critic network to approximate the state-value function. As shown in Fig. 3 (a), the proposed method reduces g_s by 40% and 58% compared to 16QAM and 64QAM, respectively, demonstrating its ability to jointly optimize sensing and SemCom performance. Furthermore, Figs. 3 (a), (b), and (c) collectively reveal that the proposed method converges faster and more stably toward an effective policy, whereas the baseline Lagrange PPO undergoes a prolonged exploration phase. This improvement stems from two key factors: 1) the CPPO framework uses dedicated critic networks to separately estimate reward and cost, leading to a more accurate state-value approximation, and 2) the Lagrangian multipliers are adaptively updated for more consistent constraint satisfaction. Consequently, with the same number of training iterations, the proposed CPPO achieves approximately 9% higher SemCom performance than the Lagrange PPO baseline, while the total allocated power remains consistently near the predefined limit.

Fig. 4 further compares the SemCom performance and sensing metric g_s achieved by the proposed H-CPPO and the Lagrange PPO under various SNR conditions. A baseline with uniform power allocation across all sub-carriers is also included. While this uniform scheme does not consider semantic importance, it can attain the optimal g_s values. The results from this figure show that both H-CPPO and Lagrange PPO effectively improve SemCom performance through optimized power allocation, while maintaining g_s close to that of the optimal uniform MPSK scheme, which demonstrates the effectiveness of the RL-based approaches and the advantage of importance-aware power allocation. Moreover, we see that

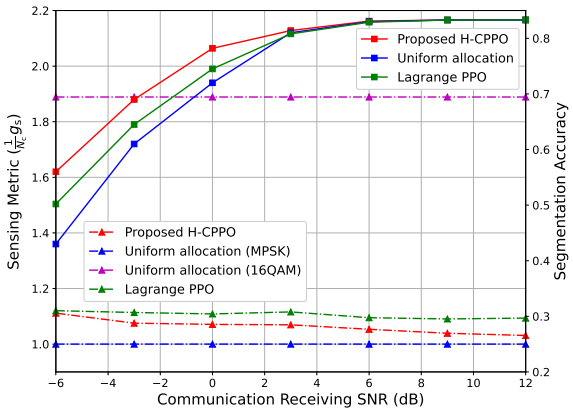


Fig. 4. Performance comparison with Lagrange PPO.

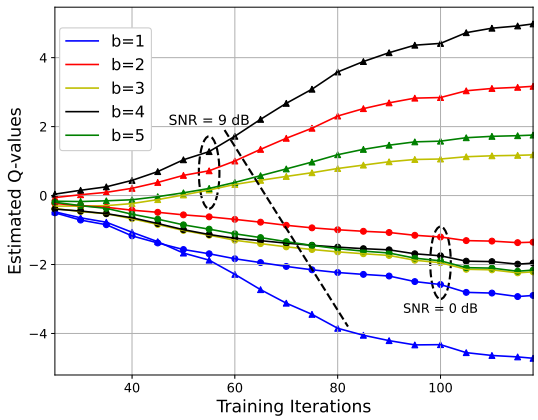


Fig. 5. Estimated Q-values.

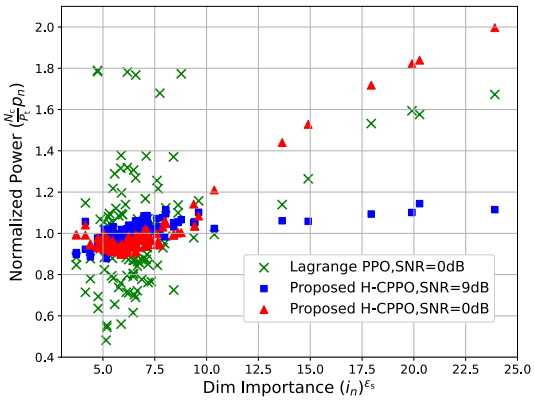


Fig. 6. Relationship between semantic importance and allocated power.

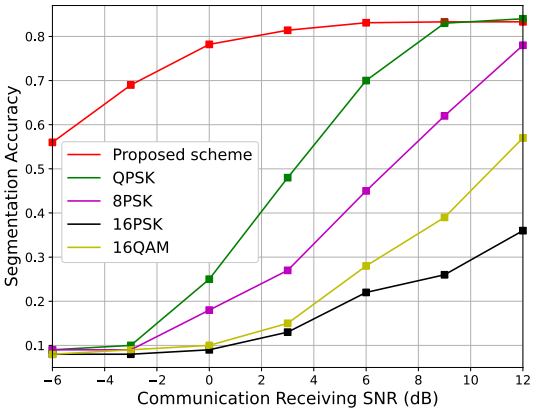


Fig. 7. Performance compared to conventional scheme.

H-CPPO consistently outperforms Lagrange PPO in both communication and sensing metrics. This improvement also stems from H-CPPO's use of separate critic networks to independently model each reward or cost component, along with dynamically adjusted Lagrange multipliers. In contrast to Lagrange PPO, which is more susceptible to training collapse due to severe penalties from constraint violations, H-CPPO strikes a better balance between exploration and constraint satisfaction, enabling faster and more stable convergence to superior policies within the same number of training episodes.

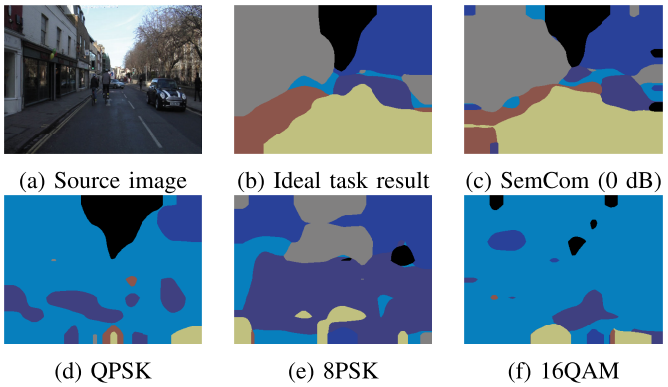
In Fig. 5, we present the estimated Q-function for the high-level policy under 0 dB and 9 dB SNR. we see from this figure that 2-bit quantization is optimal at 0 dB, while 4-bit is preferred at 9 dB. This is because higher quantization bits reduce quantization error but require higher-order modulation, which increases transmission distortion under a fixed delay constraint. The proposed H-CPPO framework effectively balances this tradeoff via the high-level DQN, efficiently identifying the optimal quantization bits.

Fig. 6 illustrates the relationship between transmission power allocation and semantic importance across different dimensions of the semantic symbols. After applying SVD to the semantic symbol set, inter-dimensional correlations are effectively removed, concentrating the most significant semantic information into a reduced number of dimensions. The results indicate that at 0 dB SNR, the proposed method achieves a stronger positive correlation between power and semantic importance compared to the Lagrange PPO baseline. This improvement stems from its superior convergence, which enables more accurate identification of critical dimensions, allocating higher power to them while assigning near-average power to less important ones. As a result, the additive noise variance in the delay–Doppler images is effectively reduced. In contrast, at a higher SNR of 9 dB, the proposed method distributes power more uniformly across all dimensions. This behavior arises because, under more favorable channel conditions, uniform power allocation is sufficient to maintain satisfactory semantic communication performance.

Fig. 7 compares the segmentation performance of the proposed SemCom scheme with that of conventional digital communication systems, which employ 8-bit UINT-8 pixel quantization with PSK or QAM modulation. For a fair comparison, none of the schemes use additional digital channel coding. The results show that the proposed method outperforms conventional digital approaches, particularly in low-SNR regimes, demonstrating the inherent reliability of SemCom. For instance, at 3 dB SNR, the proposed SemCom scheme achieves approximately 70% higher segmentation accuracy than the conventional digital scheme using the same QPSK modulation. This gain is attributed to the ability of pretrained semantic codecs and joint source-channel coding to effectively mitigate transmission distortions. Furthermore,

TABLE II
REQUIRED OFDM SYMBOL PERIODS

Transmission scheme	Requred OFDM periods	Transmission scheme	Requred OFDM periods
SemCom	160	16QAM/PSK	8100
QPSK	16200	64QAM	5400
8PSK	10800	256QAM	4050



(a) Source image (b) Ideal task result (c) SemCom (0 dB)

(d) QPSK (e) 8PSK (f) 16QAM

Fig. 8. Visualization of road image segmentation.

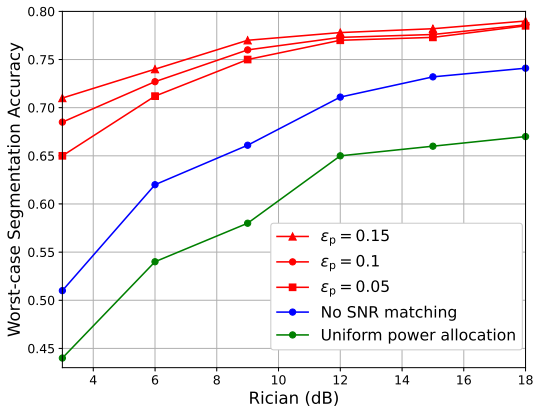


Fig. 9. Worst-case task performance under different multi-path condition.

Table II indicates that the proposed scheme requires fewer OFDM symbols per image, confirming its enhanced efficiency and reduced transmission latency.

Fig. 8 visualizes the segmentation results at the receiver under 0 dB SNR for both the proposed SemCom scheme and conventional digital communication. It can be observed that even at a low SNR, the proposed scheme maintains high segmentation accuracy, while the conventional approach introduces significant distortion in the task results.

In Fig. 9, we evaluate the proposed SNR matching strategy in mitigating the randomness of sub-channel gains, which clearly demonstrates the generalization capability of the proposed scheme under varying channel conditions. We simulate a multicast scenario with six receivers, in which the worst-case SNR is 0 dB. The SemCom performance is assessed based on the worst task performance among all receivers. Each sub-channel gain is modeled using a Rician distribution, with the K-factor varied to emulate different multipath intensities. The results indicate that at a Rician factor of 3 dB, the proposed

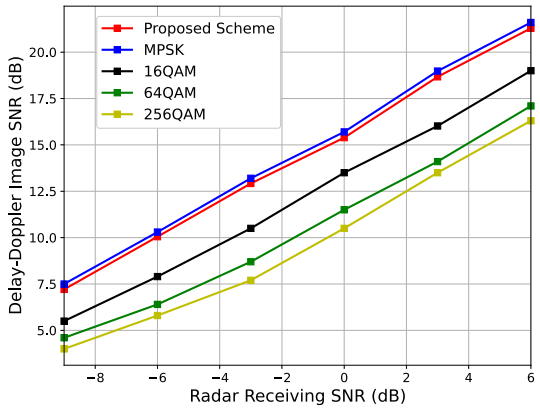


Fig. 10. Delay-Doppler image SNR vs. Radar receiving SNR.

scheme enhances worst-case SemCom performance by up to 39% and 61%, respectively, compared to systems without SNR matching and with equal power allocation. This gain is achieved by ensuring the worst-case SNR per sub-channel approximates the AWGN case, thereby enabling multicast SemCom performance close to that in AWGN. Furthermore, system performance improves with larger values of ϵ_{p}, since a higher ϵ_{p} allows greater flexibility in power allocation.

Fig. 10 shows the relationship between the radar received SNR and the SNR of the delay-Doppler image for considered schemes. Specifically, the power allocation across sub-carriers is configured at a communication SNR of 0 dB, and tolerant factor for SNR matching is applied with $\epsilon_{\mathrm{p}} = 0.1$. We see from this figure that the proposed method performs close to the optimal case (i.e., MPSK modulation with uniform power allocation across sub-carriers), indicating its effectiveness in controlling additive noise in delay-Doppler images. Moreover, at an extremely low radar received SNR of–9 dB, the proposed method improves the delay-Doppler map SNR by 1.7 dB, 2.6 dB, and 3.2 dB compared to 16QAM, 64QAM, and 256QAM, respectively. This demonstrates that the proposed method can better guarantee detection performance and estimation accuracy for target parameters when the target echo is weak.

Figure 11 presents the joint radar-communication beamforming patterns of various schemes. The ideal beam pattern is defined as five narrow beams, each with a width of 6°, covering $-30°$ to 30° for sector detection. Six receiving vehicles are positioned at randomly generated azimuths $\{-82°, -68°, -39°, 6°, 23°, 77°\}$ and distances $\{150\mathrm{m}, 120\mathrm{m}, 117\mathrm{m}, 90\mathrm{m}, 92\mathrm{m}, 130\mathrm{m}\}$. When solving the joint beamforming problem (20), we set $\Gamma_{\min} = 3$ dB to ensure sufficient SemCom performance (i.e., a segmentation accuracy greater than 0.8). Conventional QPSK requires a minimum of 8 dB SNR for equivalent performance, with higher-order PSK/QAM requiring at least 12 dB. From this figure, we observe that the joint beam pattern generated by the proposed method more closely matches the ideal pattern. This is due to the high reliability of SemCom at low SNR, which permits greater power concentration in sensing directions. Conversely, conventional methods necessitate higher received SNR, which disperses beamforming gain and introduces strong

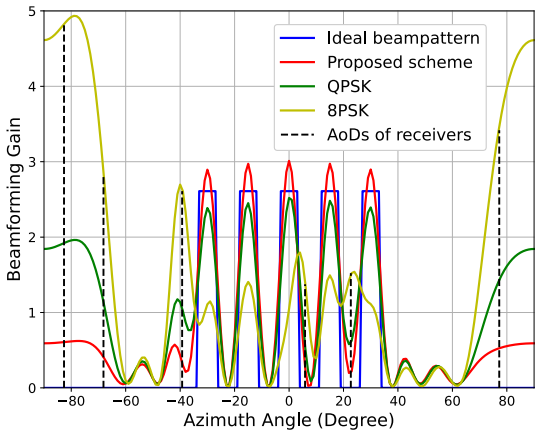


Fig. 11. Example of joint beamforming pattern.

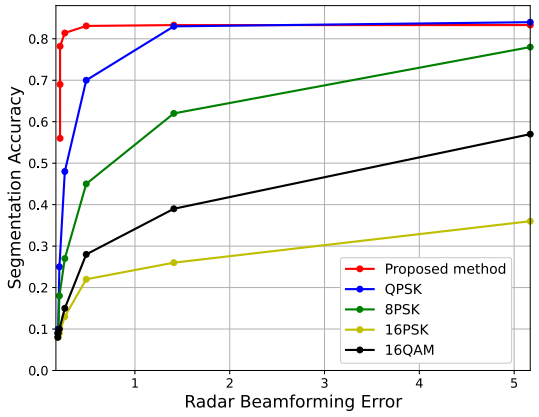


Fig. 12. Task performance vs. Radar beamforming error.

communication-target echoes that may degrade radar detection at expected directions.

In Fig. 12, we illustrates the tradeoff between beamforming error and segmentation performance under the same ideal radar beam pattern and communication target settings as in the previous figure. The results show that the proposed scheme significantly reduces radar beamforming error while maintaining comparable semantic communication task performance. For example, when achieving a segmentation accuracy exceeding 0.8, the proposed SemCom scheme reduces radar beamforming error by approximately 82% compared to the conventional digital transmission scheme using the same QPSK modulation. This improvement is attributed to two key factors: the inherent robustness of the semantic codecs, and the H-CPPO-based optimization of transmission parameters, which enhances SemCom performance at given SNR levels. As a result, the proposed scheme effectively alleviates the inherent sensing–communication conflict in integrated systems and significantly advances the achievable performance frontier. The conventional QPSK-based scheme slightly outperforms the proposed method at the point of maximum beamforming error, as the high communication SNR in that region enables near-error-free transmission and perfect image recovery. In contrast, even with minimal transmission errors, the proposed scheme incurs a small quantization loss, leading to a marginal performance degradation.

VI. Conclusion

In this work, we propose a novel OFDM-ISAC framework for CAV that incorporates digital SemCom to achieve reliable communication and accurate sensing simultaneously. In this framework, the transmitting vehicle extracts semantic symbols from the source data and transmits each dimension in parallel over OFDM sub-carriers to receiving vehicles. At the same time, it utilizes OFDM signal echoes for radar-based environmental sensing. To support reliable SemCom, the transmitter must jointly optimize the quantization bits for semantic symbols, the power allocation across semantic symbol dimensions, and the beamforming strategy. These optimizations must also account for sensing performance, resulting in a tradeoff between radar sensing and task-oriented SemCom. To tackle this challenge, we decompose it into three subproblems and develop corresponding solutions: an H-CPPO algorithm, a joint beamforming strategy, and a semantic importance-based SNR matching strategy. Simulation results demonstrate that the proposed framework outperforms conventional digital transmission schemes in both communication and sensing performance.

Appendix

A. Proof of Lemma 1

According to (13), $\widetilde{N}_i$ can be expanded as

$$\widetilde{N}_i = \sum_{n=0}^{N_c-1}\sum_{l=0}^{L-1} \frac{e^{j2\pi n\Delta f\tau}e^{-j2\pi lT_{\text{sym}}\nu}}{\sqrt{p_n}c_{n,l}} n_{n,l}^{(i)}, \tag{32}$$

where $n_{n,l}^{(i)}$ is the element of $\boldsymbol{N}_i$ at row n column l. Since each $n_{n,l}^{(i)}$ is i.i.d. circularly symmetric Gaussian variable, $\widetilde{N}_i$ is also a circularly symmetric Gaussian variable when $\boldsymbol{C}$ is given. Hence, the conditional mean and variance of $\widetilde{N}_i$ are respectively expressed as

$$\mathbb{E}\left(\widetilde{N}_i|\boldsymbol{C}\right) = \sum_{n=0}^{N_c-1}\sum_{l=0}^{L-1} \frac{e^{j2\pi n\Delta f\tau}e^{-j2\pi lT_{\text{sym}}\nu}}{\sqrt{p_n}c_{n,l}} \mathbb{E}\left(n_{n,l}^{(i)}\right) = 0, \tag{33}$$

$$\mathbb{D}\left(\widetilde{N}_i|\boldsymbol{S}\right) = \sum_{n=0}^{N_c-1}\sum_{l=0}^{L-1} \frac{\sigma_s^2}{p_n|c_{n,l}|^2}. \tag{34}$$

Based on (33) and (34), the power of $\widetilde{N}_i$ is calculated as

$$\begin{aligned}\mathbb{E}(|\widetilde{N}_i|^2) &= \mathbb{E}_{\boldsymbol{C}}\left[\mathbb{D}\left(\widetilde{N}_i|\boldsymbol{C}\right)\right] + \mathbb{D}_{\boldsymbol{C}}\left[\mathbb{E}\left(\widetilde{N}_i|\boldsymbol{C}\right)\right] \\ &= \sigma_s^2 L\left(\sum_n \frac{1}{p_n}\right)\mathbb{E}\left(\frac{1}{|c|^2}\right).\end{aligned} \tag{35}$$

By computing $\mathbb{E}\left(\frac{1}{|c|^2}\right)$ for both PSK and QAM constellations and substituting the resulting values $\mathbb{E}_{\text{PSK}}\left(\frac{1}{|c|^2}\right)$ and $\mathbb{E}_{\text{QAM}}\left(\frac{1}{|c|^2}\right)$ into (35), we arrive at the final expressions given in (14).

This completes the proof. □

B. Proof of Lemma 2

We consider optimizing $b, m, \boldsymbol{p}$ when α and $\boldsymbol{w}$ are given. Since $f_{\mathrm{s}}(\alpha, \boldsymbol{w})$ is independent of $b, m, \boldsymbol{p}$, this subproblem can be formulated as

$$\min_{b,m,\boldsymbol{p}} \left[\frac{1}{N_{\mathrm{c}}} g_{\mathrm{s}}(\boldsymbol{p})\right] \tag{36}$$

$$\text{s.t. (11a), (11b), and (11c).} \tag{36a}$$

Given that the sub-channel gains are constant, the optimal solution to problem (36) depends solely on the received SNR. Denote by $G(\Gamma)$ the optimal value of (36) for a given SNR Γ, with the corresponding optimal variables $b^*, m^*, \boldsymbol{p}^*$. Then, as long as the received SNR of each vehicle exceeds Γ, and the OFDM parameters are configured as $b^*, m^*, \boldsymbol{p}^*$, both the SemCom and sensing performance will attain the level achieved by the solution to problem (36). We accordingly define the worst-case received SNR under a given $\boldsymbol{w}$ as

$$\Gamma(\boldsymbol{w}) = \min_{v} \frac{|(\boldsymbol{h}^{\mathrm{c}}_{v,n,l})^{\mathrm{T}} \boldsymbol{w}|^2 P_{\mathrm{t}}}{N_{\mathrm{c}} \sigma_{\mathrm{c}}^2}. \tag{37}$$

Based on $G(\Gamma)$ and $\Gamma(\boldsymbol{w})$, the original problem (16) can be transformed into

$$\min_{\alpha,\boldsymbol{w}} \frac{1}{N_\theta} f_{\mathrm{s}}(\alpha, \boldsymbol{w}) + \rho \cdot G[\Gamma(\boldsymbol{w})] \tag{38}$$

$$\text{s.t. } \Gamma(\boldsymbol{w}) = \min_{v} \frac{|(\boldsymbol{h}^{\mathrm{c}}_{v,n,l})^{\mathrm{T}} \boldsymbol{w}|^2 P_{\mathrm{t}}}{N_{\mathrm{c}} \sigma_{\mathrm{c}}^2},$$

$$\text{(11d), and (16a),} \tag{38a}$$

By defining $\Gamma(\boldsymbol{w})$ as a slack variable, constraint (38a) can be reformulated as (17a). As a result, problem (38) is equivalent to (17).

This completes the proof. □

REFERENCES

[1] Z. Ren et al., "Fundamental CRB-rate tradeoff in multi-antenna ISAC systems with information multicasting and multi-target sensing," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3870–3885, Apr. 2024.

[2] C. Wen, Y. Huang, and T. N. Davidson, "Efficient transceiver design for MIMO dual-function radar-communication systems," *IEEE Trans. Signal Process.*, vol. 71, pp. 1786–1801, 2023.

[3] J. Wu, Z. Wang, Y.-F. Liu, and F. Liu, "Efficient global algorithms for transmit beamforming design in ISAC systems," *IEEE Trans. Signal Process.*, vol. 72, pp. 4493–4508, 2024.

[4] O. Huan, T. Luo, and M. Chen, "Multi-modal image and radio frequency fusion for optimizing vehicle positioning," *IEEE Trans. Mobile Comput.*, vol. 24, no. 2, pp. 696–708, Feb. 2025.

[5] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO communications with MIMO radar: From co-existence to joint transmission," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2755–2770, Apr. 2018.

[6] X. Liu, T. Huang, N. Shlezinger, Y. Liu, J. Zhou, and Y. C. Eldar, "Joint transmit beamforming for multiuser MIMO communications and MIMO radar," *IEEE Trans. Signal Process.*, vol. 68, pp. 3929–3944, 2020.

[7] Y. Xiong, F. Liu, Y. Cui, W. Yuan, T. X. Han, and G. Caire, "On the fundamental tradeoff of integrated sensing and communications under Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 69, no. 9, pp. 5723–5751, Sep. 2023.

[8] J. An, H. Li, D. W. K. Ng, and C. Yuen, "Fundamental detection probability vs. Achievable rate tradeoff in integrated sensing and communication systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 9835–9853, Dec. 2023.

[9] Z. Huang, A. Liu, R. Du, and T. X. Han, "Capacity-CRB tradeoff in OFDM integrated sensing and communication systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Rome, Italy, May 2023, pp. 2437–2442.

[10] Z. Wei et al., "Waveform design for MIMO-OFDM integrated sensing and communication system: An information theoretical approach," *IEEE Trans. Commun.*, vol. 72, no. 1, pp. 496–509, Jan. 2024.

[11] Q. Qi, X. Chen, C. Zhong, C. Yuen, and Z. Zhang, "Deep learning-based design of uplink integrated sensing and communication," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 10639–10652, Sep. 2024.

[12] M. Hatami, N. T. Nguyen, and M. Juntti, "Joint waveform design and sub-carrier allocation for multiuser MIMO ISAC," in *Proc. IEEE 5th Int. Symp. Joint Commun. Sens. (JC&S)*, Oulu, Finland, Jan. 2025, pp. 1–6.

[13] C. Sturm and W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," *Proc. IEEE*, vol. 99, no. 7, pp. 1236–1259, Jul. 2011.

[14] Y. Wu, F. Lemic, C. Han, and Z. Chen, "Sensing integrated DFT-spread OFDM waveform and deep learning-powered receiver design for terahertz integrated sensing and communication systems," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 595–610, Jan. 2023.

[15] J. Wu, W. Yuan, Z. Wei, K. Zhang, F. Liu, and D. Wing Kwan Ng, "Low-complexity minimum BER precoder design for ISAC systems: A delay-Doppler perspective," *IEEE Trans. Wireless Commun.*, vol. 24, no. 2, pp. 1526–1540, Feb. 2025.

[16] F. Liu et al., "CP-OFDM achieves the lowest average ranging sidelobe under QAM/PSK constellations," *IEEE Trans. Inf. Theory*, vol. 71, no. 9, pp. 6950–6967, Sep. 2025.

[17] L. Hu, Z. Du, and G. Xue, "Radar-communication integration based on OFDM signal," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Guilin, China, Aug. 2014, pp. 442–445.

[18] N. Bekkali, S. Bidon, D. Roque, and M. Benammar, "The price of information in OFDM-DFRC systems," in *Proc. IEEE Radar Conf. (RadarConf24)*, Denver, CO, USA, May 2024, pp. 1–6.

[19] Y. Zhang, S. Aditya, and B. Clerckx, "Input distribution optimization in OFDM dual-function radar-communication systems," *IEEE Trans. Signal Process.*, vol. 72, pp. 5258–5273, 2024.

[20] B. Geiger, F. Liu, S. Lu, A. Rode, and L. Schmalen, "Joint optimization of geometric and probabilistic constellation shaping for OFDM-ISAC systems," in *Proc. IEEE 5th Int. Symp. Joint Commun. Sens. (JC&S)*, Oulu, Oulu, Finland, Jan. 2025, pp. 1–6.

[21] Z. Du, F. Liu, Y. Xiong, T. X. Han, Y. C. Eldar, and S. Jin, "Reshaping the ISAC tradeoff under OFDM signaling: A probabilistic constellation shaping approach," *IEEE Trans. Signal Process.*, vol. 72, pp. 4782–4797, 2024.

[22] D. Gündüz et al., "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 5–41, Jan. 2023.

[23] Y. Yang, C. Guo, F. Liu, L. Sun, C. Liu, and Q. Sun, "Semantic communications with artificial intelligence tasks: Reducing bandwidth requirements and improving artificial intelligence task performance," *IEEE Ind. Electron. Mag.*, vol. 17, no. 3, pp. 4–13, Sep. 2023.

[24] C. Liu, C. Guo, Y. Yang, W. Ni, and T. Q. S. Quek, "OFDM-based digital semantic communication with importance awareness," *IEEE Trans. Commun.*, vol. 72, no. 10, pp. 6301–6315, Oct. 2024.

[25] Y. Yang, M. Shikh-Bahaei, Z. Yang, C. Huang, W. Xu, and Z. Zhang, "Secure design for integrated sensing and semantic communication system," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Dubai, United Arab Emirates, Apr. 2024, pp. 1–7.

[26] J. Dai, H. Fan, Z. Zhao, Y. Sun, and Z. Yang, "Secure resource allocation for integrated sensing and semantic communication system," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Denver, CO, USA, Jun. 2024, pp. 1225–1230.

[27] H. Amiriara, M. Mirmohseni, A. Elzanaty, Y. Ma, and R. Tafazolli, "A physical layer security framework for IRS-assisted integrated sensing and semantic communication systems," *IEEE Trans. Cognit. Commun. Netw.*, vol. 11, no. 5, pp. 2955–2969, Oct. 2025.

[28] S. Liu, H. Yang, W. Xie, and M. Zheng, "Intelligent semantic communication scheme integrating ISAC for low-altitude intelligent networks," *IEEE Trans. Commun.*, vol. 74, pp. 3018–3033, 2026.

[29] Z. Zhao et al., "Efficient design for NOMA enabled integrated sensing and semantic communication," in *Proc. IEEE 99th Veh. Technol. Conf. (VTC-Spring)*, Singapore, Jun. 2024, pp. 01–05.

[30] B. Uykulu, T. M. Duman, and O. Arikan, "Radar-centric integrated sensing and communications using goal-oriented semantic downlink communications," in *Proc. IEEE 36th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2025, pp. 1–7.

[31] W. Zhang, Y. Wang, M. Chen, T. Luo, and D. Niyato, "Optimization of image transmission in cooperative semantic communication networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 861–873, Feb. 2024.

[32] Y. Wang et al., "Performance optimization for semantic communications: An attention-based reinforcement learning approach," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2598–2613, Sep. 2022.
[33] M. F. Keskin et al., "Fundamental trade-offs in monostatic ISAC: A holistic investigation toward 6G," *IEEE Trans. Wireless Commun.*, vol. 24, no. 9, pp. 7856–7873, Sep. 2025.
[34] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and cramer-rao bound: Further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.
[35] D. R. Fuhrmann and G. San Antonio, "Transmit beamforming for MIMO radar systems using signal cross-correlation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 1, pp. 171–186, Jan. 2008.
[36] H. Yu et al., "Optimizing reinforcement learning training over digital twin enabled multi-fidelity networks," 2026, *arXiv:2603.08931*.
[37] Y. Zhang, Q. Vuong, and K. W. Ross, "First order constrained optimization in policy space," 2020, *arXiv:2002.06506*.
[38] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[39] X. Xu, H. Xu, D. Wei, W. Saad, M. Bennis, and M. Chen, "Transformer-based collaborative reinforcement learning for fluid antenna system (FAS)-enabled 3D UAV positioning," *IEEE J. Sel. Areas Commun.*, vol. 44, pp. 1128–1143, 2026.
[40] M. Chen et al., "Distributed learning in wireless networks: Recent progress and future challenges," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3579–3605, Dec. 2021.
[41] O. Huan, Y. Yang, T. Luo, and M. Chen, "Multi-modal data-based semi-supervised learning for vehicle positioning," *IEEE Trans. Commun.*, vol. 73, no. 3, pp. 1663–1676, Mar. 2025.
[42] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.*, vol. 17, pp. 1–5, Jan. 2016. [Online]. Available: https://stanford.edu/
[43] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognit. Lett.*, vol. 30, no. 2, pp. 88–97, Jan. 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:10759568
[44] Q. Pan et al., "Image segmentation semantic communication over Internet of Vehicles," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Glasgow, U.K., Mar. 2023, pp. 1–6.

Ouwen Huan (Graduate Student Member, IEEE) received the B.S. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2021, where he is currently pursuing the Ph.D. degree with the Information and Communication Engineering Department. His research interests include wireless communications, integrated sensing and communication (ISAC), RF-based positioning, and machine learning.

Chuanhong Liu received the B.Eng. and Ph.D. degrees from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications (BUPT), China, in 2020 and 2025, respectively. From 2023 to 2024, he was a Visiting Ph.D. Student with Singapore University of Technology and Design, Singapore. He is currently a Senior Technical Researcher of China Mobile (Suzhou) Software Technology Company Ltd., China. His research interests include semantic communication, deep learning, and resource allocation. He received the IEEE Wireless Communications and Networking Conference 2021 Best Paper Award.

Nuocheng Yang received the B.S. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2021, where he is currently pursuing the Ph.D. degree with the Information and Communication Engineering Department. He is a Visiting Research Student with Singapore University of Technology and Design (SUTD), working in the Laboratory of Prof. Tony Q. S. Quek. His research interests include reinforcement learning, computing power networks, and machine learning in wireless networks.

Zhilong Zhang (Member, IEEE) received the B.E. degree in communication engineering from the University of Science and Technology, Beijing, China, in 2007, and the M.S. and Ph.D. degrees in communication and information systems from Beijing University of Posts and Telecommunications (BUPT), Beijing, in 2010 and 2016, respectively. From 2010 to 2012, he was a Software Engineer with TD Tech Ltd., Beijing. From 2014 to 2015, he was a Visiting Scholar with Stony Brook University, NY, USA. He is currently a Professor with BUPT. His research interests include optimization theory and machine learning, and their applications in wireless networks and intelligent communications.

Tao Luo received the B.S. degree from Shaanxi Normal University in 1993, the M.S. degree from Shanghai University in 1999, and the Ph.D. degree from Beijing University of Posts and Telecommunications (BUPT), China, in 2002. He is currently a Professor with BUPT. His research interests include mobile communication, semantic communications, the Internet of Vehicles, and machine learning.

Mingzhe Chen is currently an Assistant Professor with the Department of Electrical and Computer Engineering and the Knight Foundation Chair in data science and AI with the Frost Institute for Data Science and Computing, University of Miami. His research interests include federated learning, reinforcement learning, virtual reality, unmanned aerial vehicles, and the Internet of Things. He received the NSF CAREER Award in 2026. He has also received four IEEE Communications Society journal paper awards, including the IEEE Marconi Prize Paper Award in Wireless Communications in 2023, the Young Author Best Paper Award in 2021 and 2023, and the Fred W. Ellersick Prize Award in 2022, and four conference best paper awards at ICCCN in 2023, IEEE WCNC in 2021, IEEE ICC in 2020, and IEEE GLOBECOM in 2020. He has been named as a Highly Cited Researcher by Clarivate since 2025. He serves as an Associate Editor for IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, and IEEE TRANSACTIONS ON MACHINE LEARNING IN COMMUNICATIONS AND NETWORKING.