Quantum Deep Deterministic Policy Gradient for Digital Twin-enabled Semantic IoV Networks

James Adu Ansere, *Member, IEEE,* Sasinda C. Prabhashana, *Student Member, IEEE,* Nidhi Simmons, *Senior Member, IEEE,* Octavia A. Dobre, *Fellow, IEEE,* Hyundong Shin, *Fellow, IEEE,* and Trung Q. Duong, *Fellow, IEEE*

***Abstract*—Internet of Vehicles (IoV) networks are becoming more complex as they require real-time decision-making and efficient resource management. These demands make it difficult to maintain stable and reliable operations. The challenges are especially severe in dynamic and time-varying environments. To address these limitations, we propose a framework that integrates the quantum-based deep deterministic policy gradient (Q-DDPG) with digital twin networks (DTN) for distributed semantic optimization in dynamic IoV environments. The framework leverages quantum computing, such as superposition and entanglement, to enhance distributed semantic decisions. DTNs provide real-time modeling for efficient task offloading and adaptive resource allocation in decentralized IoV environments under varying conditions and uncertainties. The numerical results validate the robustness of the proposed approach, significantly reduce latency, and improve energy efficiency.**

***Index Terms*—Internet of Vehicles, Deep Deterministic Policy Gradient, Digital Twin Networks, Quantum Machine Learning.**

I. Introduction

INTERNET of Vehicles (IoV) networks have transformed transportation and mobility through the integration of vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-cloud (V2C) communications [1], [2]. These networks underpin advances such as autonomous vehicles, real-time traffic management, and intelligent transportation systems. However, their expanding complexity presents challenges in resource optimization and task offloading. Efficient resource management and low-latency communication are essential for the seamless operation of decentralized IoV networks, where unpredictability arises from fluctuating traffic patterns, inconsistent user behavior, and dynamic wireless channel conditions [3]. Traditional optimization methods, often based on static assumptions, falter under such uncertainties, and scaling them to the demands of dynamic IoV networks introduces significant computational bottlenecks [4]. In distributed environments, this creates added pressure as scalable solutions rely heavily on decentralized decision-making.

J. A. Ansere, Sasinda C. Prabhashana, O. A. Dobre are with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1B 3X5, Canada (e-mail: {jaansere, cwelhengodag, odobre}@mun.ca).

N. Simmons is with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT7 1NN Belfast, U.K. (e-mail: nidhi.simmons@qub.ac.uk).

H. Shin is with the Department of Electronics and Information Convergence Engineering, Kyung Hee University, 1732 Deogyeong-daero, Giheung-gu, Yongin-si, Gyeonggi-do 17104, Republic of Korea (e-mail: hshin@khu.ac.kr).

T. Q. Duong is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1C 5S7, Canada, and with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT7 1NN Belfast, U.K., and also with the Department of Electronic Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, South Korea (e-mail: tduong@mun.ca).

The work of T. Q. Duong was supported in part by the Canada Excellence Research Chair (CERC) Program CERC-2022-00109 and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant Program RGPIN-2025-04941. The work of O. A. Dobre was supported in part by the Canada Research Chair Program CRC-2022-00187. The work of H. Shin was supported in part by National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2025 00556064).

A. Digital Twin with Semantic Resource in IoV Networks

Digital twin (DT) technology with semantic optimization strategies is crucial to improving IoV network efficiency [5]. This integration addresses the limitations of the IoV network by enabling real-time data processing, dynamic communication, and decision making, which are essential to optimize transportation safety and performance. In [6], a blockchain-powered DT edge network framework is introduced using federated learning and a blockchain based on directed acyclic graph (DAG) for secure model updates and efficient use of resources. The approach faces scalability issues with DAG in devices that are resource-constrained and difficulties in maintaining consensus in large-scale networks. A Markov decision process (MDP)-based approach for joint network selection and power level allocation in vehicular networks, incorporating transfer learning to accelerate learning and reduce latency [7]. The method has computational complexity and relies on accurate prior knowledge for effective transfer learning. The semantic communication framework with generative adversarial networks is developed in [8] to efficiently transmit images, aimed at reducing bandwidth and energy usage. The system experiences performance degradation under noisy channels and struggles to maintain semantic accuracy in varying noise levels.

As IoV networks scale and advance, the critical role of DT with semantic optimization to prioritize critical communications and optimize resource allocation becomes increasingly significant [9]. This prioritization of contextually significant information improves decision-making efficiency, equipping IoV networks to adapt quickly and effectively in dynamic scenarios. DT technology is introduced in [10] to enable task offloading in IoV, employing learning algorithms to predict and optimize task assignments based on real-time insights

from DT models. Although effective, the framework relies heavily on the accuracy of DT representations and faces scalability challenges in large IoV networks with high volumes of vehicles and data. In [11], DT technology is integrated with intelligent reflective surfaces to optimize vehicle task absorption and resource allocation in 6G enabled IoV networks.

The integration of DT into IoV opens up new possibilities for the development of intelligent and adaptive vehicle systems. In [12], a comprehensive study is introduced on integrating digital twin technology into 6G networks to enhance resource management and network optimization. The study identifies challenges in real-time synchronization, scalability, and security in large-scale deployments. In [13], a blockchain-enabled federated learning framework is proposed for DT wireless networks to improve reliability, data privacy, and edge resource management using multi-agent reinforcement learning (MARL). The system requires significant resources for real-time synchronization and faces challenges in edge association due to dynamic resource availability.

Furthermore, the dynamic nature of IoV networks, marked by variable traffic conditions, high vehicular mobility, and unstable bandwidth, poses significant challenges to effective communication and resource management. To address these uncertainties and ensure reliable performance across diverse scenarios, adaptive algorithms are essential [10]. Quantum computing has recently emerged as a transformative solution for complex decision-making in dynamic environments. Leveraging superposition, entanglement, and quantum parallelism, it enables efficient exploration of vast state spaces with minimal computational overhead [14]. By processing rapidly changing factors such as vehicle movements and spectrum variability in parallel, quantum computing supports fast, adaptive decision-making. The use of quantum states further enhances exploration in high-dimensional IoV scenarios, offering remarkable efficiency in managing their inherent complexity [15], [16].

B. Quantum Machine Learning for Optimal Resource Allocation in Wireless Networks

Recently, quantum machine learning (QML) has emerged as a promising technique in wireless network resource allocation. It addresses key limitations of classical machine learning (ML) methods. QML leverages quantum properties such as superposition, entanglement, and quantum parallelism. These quantum principles significantly reduce computational complexity and enhance optimization efficiency. Specifically, quantum superposition and entanglement allow QML algorithms to converge faster. They also require fewer training parameters compared to traditional ML approaches [14], [15]. Moreover, the advantages and challenges of deploying quantum computing-assisted ML and pure QML frameworks in next-generation wireless communication systems are discussed in [16]. It is proposed that quantum-enhanced technologies significantly improve network performance, enabling self-configurable networks that rapidly adapt to complex, multi-dimensional, and dynamic environments. In [17], QML is applied to optimize resource allocation, with a particular focus on user grouping in non-orthogonal multiple access (NOMA) wireless communication. The study introduces a quantum neural network integrated with a reinforcement learning model. It demonstrates practical advantages by enhancing the achievable sum rate in wireless systems. Similarly, in [18], a quantum neural network employing parallel training is utilized to optimize multiple input multiple output NOMA resource allocation by leveraging statistical parameters instead of full datasets. This approach achieves a comparable sum rate to conventional methods while significantly reducing complexity.

Furthermore, quantum-enhanced deep reinforcement learning (QDRL) is employed to minimize the total system cost by jointly considering energy consumption and latency in a digital twin-enabled environment [19]. This approach utilizes the Q-DDPG algorithm to improve learning efficiency and decision-making performance in complex IoV communication networks. Additionally, layerwise QDRL is used to optimize UAV trajectory, user grouping, and power allocation in [20]. According to this study, QDRL outperforms classical DRL in UAV-assisted communication networks, especially under energy-constrained and dynamically changing environments. To investigate the performance of multi-agent QDRL, cooperative mobile access in multi-UAV systems was considered in [21]. The proposed method uses quantum computing principles to mitigate the scalability and complexity challenges in traditional multi-agent DRL methods. According to the simulation results, the QDRL-based solution substantially improves both training convergence speed and the robustness of UAV mobility planning under dynamic environmental conditions. In addition, to address the limitations of frequent handovers and energy constraints in satellites, a quantum-enhanced DRL model is proposed in [22]. The developed model effectively adapts to realistic environments by considering factors such as orbital dynamics, atmospheric interference, and varying communication demands across geographic regions.

C. Quantum DDPG for DT-enabled IoV

Quantum-inspired DDPG algorithm delivers more effective approaches to complex optimization problems than classical methods [23], [24], [25]. Quantum-inspired learning improves traditional DDPG by using quantum features, namely superposition and entanglement, which increase convergence speed, reduce dimensionality, and refine the exploration-exploitation balance [16]. In [26], a framework based on a deep deterministic policy gradient (DDPG) of multiple agents is introduced to handle the task in oIV, improving mobile edge computing by improving information exchange and resource coordination to ensure optimal delivery of services in vehicular networks. Scalability challenges emerge in large-scale IoV networks due to increased computational complexity, with convergence issues arising in highly dynamic environments. An adaptive joint resource allocation scheme for the IoV framework is presented, dividing resources into uplink, computing, and downlink sub-strategies [27]. Twin-delayed DDPG is used to dynamically optimize network capacity, reduce delay, and minimize energy consumption. The effectiveness of the algorithm is limited by the complexity of the model in high-dimensional problems, which can impact performance in highly unpredictable scenarios [28]. This work employs an algorithm

based on DDPG to optimize multi-user computation offloading and caching strategies in vehicular edge computing systems, aiming to minimize execution delay while improving caching and resource utilization for variability in task sequence [20]. The framework's reliance on accurate request modeling and coupled optimization raises computational demands, reducing its effectiveness in highly variable scenarios.

In [15], a framework for machine learning supported by quantum computing was proposed for 6G networks. This framework leverages quantum computing for network orchestration and self-reconfiguration, with the aim of improving performance through real-time learning and adaptation to network dynamics. However, it faces challenges in implementing scalable quantum computing solutions and processing large volumes of data in real-time. In [29], the quantum-enhanced support vector machine (QE-SVM) was developed for sentiment classification. By employing quantum feature maps and optimization techniques such as simultaneous perturbation stochastic approximation, it demonstrated superior performance compared to classical SVM methods. Despite these advancements, significant challenges persist in optimizing quantum circuit parameters and managing the high dimensionality of quantum representations for practical deployment. In [30], a quantum reinforcement learning (QRL) algorithm was introduced to optimize resource management in edge intelligence-assisted IoV systems. The algorithm aimed to reduce latency and enhance system performance by balancing computational and communication resources. Despite its innovative use of hybrid technologies, such as WiFi and cellular networks, maintaining consistent low latency and high performance in dynamic environments remains a challenge.

D. Motivation and contributions

Although there have been recent breakthroughs in implementing DDPG approaches for task optimization in complex IoV applications [31], [32], [33], significant challenges remain, as follows:

1) Inadequate Adaptability to Rapid Network Dynamics: Traditional optimization approaches, such as linear programming and convex optimization, are based on static models that assume a stable network environment [34]. IoV networks, characterized by high mobility and frequent changes in topology, vehicle density, and communication conditions, present challenges to these methods [36]. As a result, traditional approaches struggle to adapt, leading to suboptimal performance and computationally expensive re-optimization for real-time applications.

2) Constrained Scalability and Computational Efficiency: In dynamic environments, the growth of the network size increases computational complexity, leading to longer processing times and reduced efficiency. Traditional optimization methods struggle with scalability in large IoV networks due to the vast data generated by vehicles and digital twins, making timely semantic optimization computationally infeasible [37].

3) Lack of Semantic Awareness and Integration with Digital Twins: Traditional optimization focuses on the quantity of data on semantic relevance. In IoV networks, prioritizing significant data is essential for collision avoidance and traffic management [35]. Lack of semantic awareness leads to excessive data transmission, increasing bandwidth use and latency, thus reducing communication efficiency. DT technology provides virtual replicas of physical assets for real-time monitoring and simulation. Traditional optimization fails to take advantage of DTs for real-time data analysis and feedback [38]. This hinders the integration of physical and virtual layers, decreasing optimization effectiveness in IoV networks.

4) Insufficient Real-Time Processing Capabilities: Dynamic semantic-based communication optimization requires adaptive real-time data processing to facilitate the rapid transmission of information [39]. Traditional optimization methods lack the computational speed and responsiveness to manage the continuous data stream in DT-enabled IoV systems. This delay undermines the effectiveness of semantic-based communication optimization, causing computation overhead and reduced reliability.

Drawing on the insights from the aforementioned studies, we introduce a quantum DDPG (Q-DDPG) framework to reduce latency and energy consumption while maintaining accuracy in DT-enabled IoV networks. By seamlessly integrating Q-DDPG with DT technology, the framework significantly enhances semantic task processing efficiency within high-dimensional hybrid action spaces. This approach overcomes traditional optimization challenges, offering improved scalability and adaptability in complex IoV environments. Table I provides a detailed comparison, illustrating the distinctions between our framework and the existing methodologies. The key technical contributions of this work are summarized as follows:

1) *A Quantum DDPG-Based Framework for IoV Networks*: We propose a quantum DDPG framework for semantic resource optimization in DT-enabled IoV networks. The framework employs quantum-enhanced reinforcement learning to boost task efficiency and tackle scalability and adaptability challenges in dynamic IoV settings.
2) *Integration of Digital Twin Networks for Real-Time Semantic Optimization*: We utilize DT for adaptive semantic intelligence in IoV networks. This integration enables continuous monitoring and real-time semantic decisions for timely adaptation to network uncertainty and changing conditions.
3) *Scalable and Adaptive Distributed Semantic Optimization Solution*: We design a scalable, distributed solution for managing dynamic large-scale IoV ecosystems in high-dimensional continuous action spaces efficiently. The solution provides a reliable framework for efficient resource management and task prioritization in dynamic IoV environments.
4) *Comprehensive Mathematical Modeling and Optimization*: We develop models for semantic networks and stochastic combinatorial optimization, enabling the integration of quantum DDPG and DT technologies in IoV networks.
5) *Performance Metrics in Dynamic IoV Environments*: The simulation results show the effectiveness of the proposed Q-DDPG framework and its superior performance com-

TABLE I: Related works summary to the proposed system

System features	**References**						
	[10]	[13]	[8]	[33]	[34]	[35]	Proposed Q-DDPG
Digital Twin in IoV	✓	✓		✓			✓
Semantic-based communication model		✓	✓	✓	✓	✓	✓
Semantic- communication based optimization		✓	✓	✓	✓	✓	✓
Semantic-aware DT Q-DDPG Framework							✓
Hybrid sequential algorithm							✓

pare to the state-of-the-art methods network uncertainty and variable time conditions.

The remainder of this paper is organized as follows. Section III details the system model and the formulation of the combinatorial optimization problem. Section II provides an overview of QDRL. Section IV introduces the Q-DDPG framework for designed dynamic IoV environment. Extensive experimental results are presented in Section V. Finally, concluding remarks are made in Section VI.

II. Quantum Deep Reinforcement Learning with Semantic IoV Networks

This section provides a comprehensive overview of the QDRL approaches.

A. Overview of Quantum Deep Reinforcement Learning

Deep reinforcement learning (DRL) represents the integration of reinforcement learning (RL) with deep neural networks. This combination enables agents to determine optimal actions in complex, high-dimensional environments by interacting with the environment, observing outcomes, and refining decisions over time. The mathematical formulation of such decision-making problems adopts the MDP framework [40], [41]. An MDP can be defined as a tuple $(\mathcal{S}, \mathcal{A}, P, R, \gamma)$, where $\mathcal{S}$ denotes the complete set of environmental states, $\mathcal{A}$ denotes the action space (discrete or continuous), $P(s'|s,a)$ indicates the transition probability from state s to s' under action a, $R(s,a)$ represents the immediate reward received upon taking action a in state s, and $\gamma \in [0,1]$ is the discount factor applied to future rewards. The learning objective in DRL is the maximization of expected cumulative reward through a policy $\pi(a|s;\theta)$, which is parameterized by a neural network with weights θ. Then, the objective function can be expressed as:

$$J(\pi) = \mathbb{E}_{\pi}\Big[\sum_{t=0}^{\infty} \gamma^t r_t\Big],$$

where r_t denotes the reward received at time step t. Thus, Q-DRL extends DRL by incorporating quantum computing principles to enhance exploration and optimization capabilities. These methods are particularly effective for solving high-dimensional, computationally intensive problems that challenge classical approaches. Moreover, various QDRL algorithms have been developed to address different action space characteristics and decision-making scenarios. Q-DQN augments the classical deep q-network using quantum encoders and quantum neural networks (QNNs) to improve representation of discrete state spaces. Q-SAC targets continuous action problems by applying entropy regularization to maintain a balance between exploration and exploitation. Q-DDPG introduces quantum-enhanced actor and critic networks for continuous control tasks. Q-MADDPG extends the quantum framework to multi-agent systems, facilitating coordinated learning among multiple agents. Q-PPO integrates quantum features into the proximal policy optimization framework to ensure stable policy updates within trust regions. These Q-DRL methods offer practical solutions for handling complex and dynamic environments. The following sections provide detailed explanations of each method.

1) Quantum Deep Q-Network (Q-DQN): Q-DQN enhances the deep Q-Network (DQN) with quantum computing. It optimizes the Q-value function by improving function approximation and enabling faster sampling in high-dimensional environments. Q-DQN has demonstrated computational advantages over classical DQNs in complex settings. In [42], quantum-inspired reinforcement learning techniques were proposed to improve resource allocation, showing better performance in terms of latency and energy efficiency. Classical states s are encoded into quantum states $|\phi(s)\rangle$ using a quantum encoder. The Q-value function is represented by a parameterized quantum circuit $Q_\theta(|\phi(s)\rangle, a)$, where θ denotes the circuit parameters. Then, the loss function can be defined using the Bellman equation which can be expressed as

$$L(\theta) = \mathbb{E}_{\mathcal{D}}\Big[\Big(r + \gamma \max a' Q_\theta(|\phi(s')\rangle, a') - Q_\theta(|\phi(s)\rangle, a)\Big)^2\Big],$$

where $(s, a, r, s') \sim \mathcal{D}$ is sampled from the replay buffer, and γ is the discount factor. Q-DQN improves learning performance in large-scale environments with complex state–action dynamics.

2) Quantum Soft Actor-Critic (Q-SAC): Q-SAC combines soft actor critic algorithm with quantum computing to enhance policy and value optimization by leveraging entropy regularization for effective exploration in dynamic environments [43]. The policy optimization objective in Q-SAC can be defined as

$$J_\pi(\theta) = \mathbb{E}_{(s,a)\sim\mathcal{D}}\Big[Q_w\left(|\phi(s)\rangle, a\right) - \alpha \log \pi_\theta(a \mid |\phi(s)\rangle)\Big],$$

where $|\phi(s)\rangle$ denotes the quantum-encoded state, Q_w represents the quantum-enhanced critic network, and α is a temperature parameter that balances exploration and exploitation. The soft Q-value can be updated by minimizing the loss which can be expressed as

$$L(w) = \mathbb{E}_{(s,a,r,s')}\Big[\Big(Q_w(|\phi(s)\rangle, a) - \big(r + \gamma V_w(|\phi(s')\rangle)\big)\Big)^2\Big],$$

where the entropy-regularized value function can be expressed as

$$V_w(|\phi(s)\rangle) = \mathbb{E}_{a\sim\pi_\theta}\Big[Q_w(|\phi(s)\rangle, a) - \alpha \log \pi_\theta(a \mid |\phi(s)\rangle)\Big].$$

Q-SAC utilizes quantum sampling and expressive quantum state representations to improve exploration and learning efficiency in high-dimensional and complex decision-making tasks.

3) Quantum Deep Deterministic Policy Gradient (Q-DDPG): Q-DDPG integrates deep deterministic policy gradient algorithm with quantum computing to operate in continuous action spaces, leveraging quantum properties to enable efficient learning and improved decision-making in dynamic environments [44]. Q-DDPG can be applied to optimize resource allocation and task scheduling in complex systems by harnessing quantum-enhanced policy learning [45]. The actor network μ_θ outputs the optimal action; the objective function can be expressed as

$$J_\pi(\theta) = \mathbb{E}_{s\sim\mathcal{D}}\Big[Q_w\big(|\phi(s)\rangle,\ \mu_\theta(|\phi(s)\rangle)\big)\Big],$$

where $|\phi(s)\rangle$ denotes the quantum-encoded state representation. The critic network Q_w can be trained by minimizing the loss function, which can be expressed as:

$$L(w) = \mathbb{E}\Big[\big(r + \gamma Q_w\big(|\phi(s')\rangle, \mu_\theta(|\phi(s')\rangle)\big) - Q_w\big(|\phi(s)\rangle, a\big)\big)^2\Big],$$

where r is the received reward and γ is the discount factor.

4) Quantum Multi-Agent Deep Deterministic Policy Gradient (Q-MADDPG): Q-MADDPG extends Q-DDPG to multi-agent systems by integrating quantum computing, using entanglement to facilitate agent interaction and quantum-encoded state representations [46]. Q-MADDPG can be applied to multi-agent coordination and decision-making tasks in dynamic environments. Each agent maintains a quantum actor, and a centralized quantum critic can be used to evaluate joint state-action pairs. The critic target value can be expressed as

$$Q_w\big(|\phi(\mathbf{s})\rangle, \mathbf{a}\big) \approx r_i + \gamma\, Q_w\Big(|\phi(\mathbf{s}')\rangle, \boldsymbol{\mu}_\theta\big(|\phi(\mathbf{s}')\rangle\big)\Big),$$

where $|\phi(\mathbf{s})\rangle$ denotes the quantum-encoded joint state, $\mathbf{a} = (a_1, \ldots, a_N)$ is the joint action vector, r_i is the reward for agent i, and $\boldsymbol{\mu}_\theta$ collects the actor outputs for all agents. Each agent i can optimize its policy by maximizing the expected joint Q-value, which can be defined as:

$$J_{\pi_i}(\theta_i) = \mathbb{E}_{\mathbf{s}\sim\mathcal{D}}\Big[\, Q_w\big(|\phi(\mathbf{s})\rangle,\ a_1, \ldots, \mu_{\theta_i}(|\phi(s_i)\rangle), \ldots, a_N\big)\Big],$$

where θ_i denotes the parameters of agent i's actor network, s_i is agent i's local observation encoded as $|\phi(s_i)\rangle$, and N is the total number of agents. The critic network Q_w can be trained by minimizing a suitable loss over joint transitions, and each quantum actor μ_{θ_i} can be updated via policy gradient methods adapted to the quantum-encoded representations.

5) Quantum Proximal Policy Optimization (Q-PPO): Q-PPO integrates proximal policy optimization with quantum computing to maintain policy stability by restricting updates to a trust region [47]. In Q-PPO, the clipped surrogate objective can be defined as

$$J_\pi(\theta) = \mathbb{E}_{(s,a)\sim\mathcal{D}}\Big[\min\big(r_\theta(s,a)\, A_\phi(|\phi(s)\rangle, a),\ \mathrm{clip}(r_\theta(s,a),\, 1-\epsilon,\, 1+\epsilon)\, A_\phi(|\phi(s)\rangle, a)\big)\Big],$$

where the probability ratio can be defined as

$$r_\theta(s,a) = \frac{\pi_\theta(a \mid |\phi(s)\rangle)}{\pi_{\theta_{\text{old}}}(a \mid |\phi(s)\rangle)},$$

and the advantage function can be expressed as

$$A_\phi(|\phi(s)\rangle, a) = Q_w(|\phi(s)\rangle, a) - V_w(|\phi(s)\rangle).$$

The value-function update loss can be denoted as

$$L_V(w) = \mathbb{E}_{(s,r,s')\sim\mathcal{D}}\Big[\big(r + \gamma\, V_w(|\phi(s')\rangle) - V_w(|\phi(s)\rangle)\big)^2\Big],$$

where r denotes the received reward, s' denotes the next state, and γ denotes the discount factor. These expressions can be applied in quantum-encoded state representations to enable stable policy updates under high-dimensional decision-making tasks.

III. System Model and Combinatorial Problem Formulation

In this section, we present the system model for semantic-communication based optimization in digital twin IoV and offer a mathematical formulation of the problem.

A. Semantic Modeling

We adopt a semantic-centric approach that focuses on the meaning of the data rather than its raw bits. Our semantic model is defined by: semantic complexity, $s_c(\tau)$, which measures the contextual information contained in a task at time τ; semantic accuracy, represented by σ_i and σ_j, which indicate how effectively vehicle i or edge server j preserves the task's meaning during computation; and semantic efficiency, $\sigma_{i,j}$, which assesses the overall effectiveness of both the transmission and processing of the task from vehicle i to server j. A higher $\sigma_{i,j}$ indicates that more meaningful information is preserved during offloading. Semantic performance $\mathcal{S}_p$ of a task at time τ can be defined as a weighted sum of semantic complexity and semantic efficiency which can be denoted as

$$\mathcal{S}_p(\tau) = \varphi_1 \cdot s_c(\tau) + \varphi_2 \cdot \sigma_{i,j}, \tag{1}$$

where φ_1 and φ_2 are weighting factors reflecting the importance of semantic complexity and semantic efficiency, respectively. Furthermore, tasks with higher $s_c(\tau)$ are prioritized with priority score, $\mathcal{P}_{score}$ as

$$\mathcal{P}_{score} = \varphi \cdot s_c(\tau), \tag{2}$$

where φ is the weighting factor. The semantic efficiency $\sigma_{i,j}$ is maximized to transmit only meaningful data $\sigma_{i,j} \geq \sigma_{\min}$, where $\sigma_{\min}$ is the minimum semantic efficiency threshold.

B. Semantic-based Network Model

This subsection presents the semantic-based network model in the DT–enabled IoV. We consider a vehicle set $\mathcal{V} =$

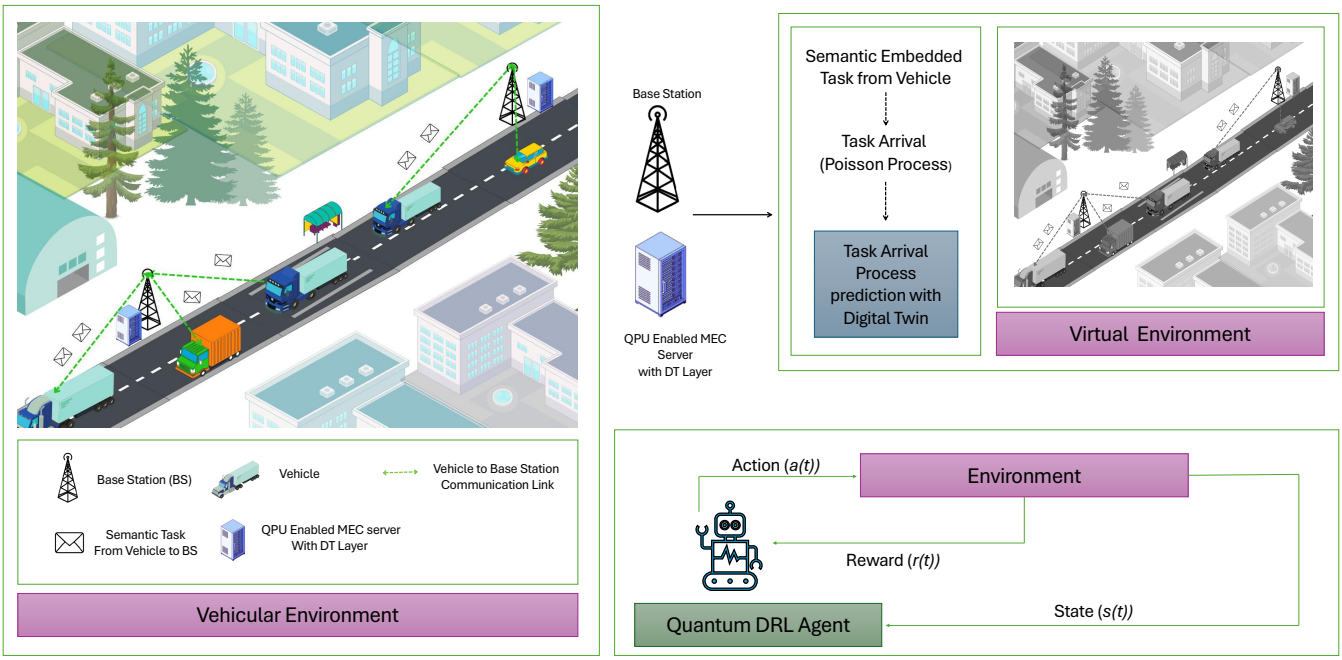

Fig. 1: An illustration of semantic-based communication: task offloading in IoV networks leveraging quantum processing unit and digital twin layer

$\{1, 2, \ldots, N\}$ and an edge-server set $\mathcal{E} = \{1, 2, \ldots, M\}$, each server co-located with or equipped with quantum processing units (QPUs). Fig. 1 illustrates the network topology, including vehicular positions, edge-server locations, and the associated QPUs for task offloading. To capture semantic relevance, we define a semantic efficiency metric $\sigma_{i,j} \in [0, 1]$ between vehicle i and server j, and incorporate it into resource allocation decisions. Edge servers use local digital twin models to mirror real-time vehicular states and semantic task arrival process, allowing predictive and adaptive resource adjustments.

C. Semantic-based Vehicular Mobility Model

We consider the position of the vehicle i in time frame τ, denoted as $\mathbf{p}_i(\tau) = (x_i(\tau), y_i(\tau))$, and the position of the edge server j located at $\mathbf{p}_j = (x_j, y_j)$. The channel gain between vehicle i and edge server j at time τ can be expressed as

$$h_{ij}(\tau) = \frac{g_0}{d_{ij}(\tau)^{\alpha}}, \tag{3}$$

where g_0 is the channel power gain at the reference distance, $d_{ij}(\tau) = \sqrt{(x_i(\tau) - x_j)^2 + (y_i(\tau) - y_j)^2}$ is the distance between vehicle i and edge server$j$, and α is the path loss exponent.

D. Stochastic Task Generation, Task Arrival, and Task Prioritization Models

1) Semantic-based Task Generation Model: Each semantic task S_k is generated by vehicle i at timeframe τ can be described as

$$S_k(\tau) = \{D_i(\tau), s_c(\tau), T_{i,p}(\tau), L_D(\tau), \mathbb{P}(A_i(\tau))\}, \tag{4}$$

where $D_i(\tau)$ denotes the task data size (in bits), $s_c(\tau)$ is the semantic task complexity, indicating its contextual information, $T_{i,p}(\tau)$ is the priority of the task, $L_D(\tau)$ is the task deadline for completion, and $\mathbb{P}(A_i(\tau))$ is the stochastic arrival task process.

2) Semantic-based Stochastic Task Arrival Model: Let $\lambda_i(\tau)$ indicate the arrival rate of semantic tasks in vehicle i, and $\mu_i(\tau)$ represent the service rate of semantic tasks. The number of semantic tasks in the queue at time frame τ follows a queuing model which can be expressed as

$$\eta_i(\tau + 1) = \eta_i(\tau) + \lambda_i(\tau) - \mu_i(\tau), \tag{5}$$

where $\eta_i(\tau)$ is the number of semantic tasks in the queue at time frame τ.

The semantic task generation by vehicle i, modeled as a Poisson process with arrival rate $\lambda_i(\tau)$, can be calculated as

$$\mathbb{P}(A_i(\tau) = k) = \frac{(\lambda_i(\tau))^k \, e^{-\lambda_i(\tau)}}{k!}, \tag{6}$$

where $\mathbb{P}(A_i(\tau) = k)$ is the probability that k semantic tasks arrive at vehicle i in time interval τ.

3) Semantic Task Prioritization Model: Semantic tasks are prioritized based on their importance and urgency, determining their position in the queue, which can be expressed as

$$S_{k,i} = \mathcal{P}_{score} + \beta \cdot (L_D(\tau) - \tau), \tag{7}$$

where β are weighting factor, $L_D(\tau)$ is the task deadline for completion, and τ is the current time. Higher values of $S_{k,i}$ increase the priority of the task queue.

E. Digital Twin for Predict the Task Arrival Process

In our framework, the DT layer functions as a real-time, virtual replica of the physical IoV network and its associated edge servers. The DT continuously aggregates updated state information, historical data, and real-time measurements from both vehicles and edge servers. Then, it constructs an accurate and dynamic model of the network conditions that support the task arrival processes and overall system performance. By forecasting the number of incoming tasks arrivals, the system can adjust resource allocations and offloading strategies before congestion occurs. Then, the DT at time step τ can be expressed as:

$$DT(\tau) = \left(DT_i^{\mathcal{V}}(\tau), DT_j^{\mathcal{E}}(\tau)\right). \tag{8}$$

Considering the uncertain and unpredictable errors between the physical entities and the DT, we model the task arrival rate $\lambda_i(\tau)$ of vehicle i with an uncertain deviation $\delta_\lambda^i(\tau)$ as

$$\lambda_i(\tau) = \hat{\lambda}_i(\tau) + \delta_\lambda^i(\tau), \quad |\delta_\lambda^i(\tau)| \leq \epsilon_\lambda, \tag{9}$$

where $\hat{\lambda}_i(\tau)$ is the estimated arrival rate predicted by the DT, and ϵ_λ is the maximum range of deviations. Importantly. this prediction is key to offloading optimization, as DT's updated estimates are used to adjust resource allocation in real time. This leads to improved offloading decisions based on both current and future network conditions. Moreover, DT's prediction is essential for deciding which tasks to prioritize because tasks with higher semantic complexity are given a higher priority score based on their complexity level. With this information, the DT can alert the system when a large number of tasks is expected, ensuring that resources are allocated to preserve the most important information and maintain strong communication quality.

F. Semantic-based Communication Model

Each vehicle i communicates with the edge server j to execute semantic tasks. The achievable transmission rate $R_{i,j}(\tau)$ between vehicle i and edge server j at time frame τ can be expressed as

$$R_{i,j}(\tau) = B_{i,j}(\tau)\log_2\left(1 + \frac{p_{i,j}(\tau)h_{i,j}(\tau)}{I + N_o}\right), \tag{10}$$

where $B_{i,j}(\tau)$ represents the system bandwidth at time frame τ, $p_{i,j}(\tau)$ is the transmission power. N_o is the noise power, and I denotes the interference.

G. Semantic-based Computation Models

Let $C_i(\tau)$ represents the local computational capacity of vehicle i, and $C_j(\tau)$ be the computational capacity of the edge server j. Let $\alpha_{i,j}(\tau)$ represents the binary offloading decision variable denoting whether semantic task S_k is processed locally or offloaded from vehicle i to edge server j at timeframe τ, defined as

$$\alpha_{i,j}(\tau) = \begin{cases} 1, & \text{if semantic task } S_k \text{ is offloaded to server } j, \\ 0, & \text{if processed locally at vehicle } i. \end{cases} \tag{11}$$

1) Local Computation Model: The computation delay $L_i^{\text{comp}}(\tau)$ for processing the semantic task S_k in the time frame τ can be calculated as

$$L_{\text{comp}}^i(\tau) = \frac{s_c(\tau)}{C_i(\tau)}\sigma_i, \tag{12}$$

where σ_i is the semantic accuracy factor for local computation.

2) Offloading Computation Model: The latency for offloading the semantic task to the edge server and process the task can be expressed as

$$L_{\text{off}}^i(\tau) = \frac{D_i(\tau)\sigma_{i,j}}{R_{i,j}(\tau)} + L_{\text{comp}}^j(\tau), \tag{13}$$

where $L_{\text{comp}}^j(\tau) = \dfrac{s_c(\tau)}{C_j(\tau)}\sigma_j$ is the computation delay at the edge server, and σ_j is the semantic accuracy factor for the edge server. Therefore, the total delay for processing and offloading a semantic task can be calculated as

$$L_{\text{total}}^{i,j}(\tau) = \begin{cases} L_{\text{off}}^i(\tau), & \text{if } \alpha_{i,j}(\tau) = 1, \\ L_{\text{comp}}^i(\tau), & \text{if } \alpha_{i,j}(\tau) = 0. \end{cases} \tag{14}$$

H. Semantic-based Energy Consumption Model

The energy consumption for transmiting the semantic task S_k from vehicle i to server j can be calculated as

$$E_{\text{off}}^{i,j}(\tau) = p_{i,j}(\tau)\cdot\frac{D_i(\tau)\sigma_{i,j}}{R_{i,j}(\tau)}, \tag{15}$$

where $p_{i,j}(\tau)$ is transmission power for vehicle i to server j. Moreover, the local processing energy consumption at vehicle i can be calculated as

$$E_{\text{loc}}^i(\tau) = \kappa_i\left(C_i(\tau)\right)^2 s_c(\tau)\sigma_i, \tag{16}$$

where κ_i is the energy efficiency coefficient of vehicle i processor. Therefore, the total energy consumption for semantic task S_k can be expressed as

$$E_{\text{total}}^{i,j}(\tau) = \begin{cases} E_{\text{off}}^{i,j}(\tau), & \text{if } \alpha_{i,j}(\tau) = 1, \\ E_{\text{loc}}^i(\tau), & \text{if } \alpha_{i,j}(\tau) = 0. \end{cases} \tag{17}$$

I. Stochastic Combinatorial Offloading Problem Formulation

We aim to minimize the total cost while considering energy consumption, latency constraints, semantic accuracy, and arrival rate uncertainties. The total cost Φ at time step τ can be calculated as

$$\Phi(\tau) = \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{E}}\left(w_1 L_{\text{total}}^{(i,j)}(\tau) + (1-w_1)E_{\text{total}}^{(i,j)}(\tau)\right), \tag{18}$$

where w_1 represents the weight factor for latency and energy consumption. Mathematically, the combinatorial optimization

problem can be formulated as follows:

$$\textbf{(P1):} \min_{B_{i,j}(\tau),\alpha_{i,j}(\tau),p_{i,j}(\tau)} \Phi(\tau) \tag{19a}$$

$$\text{s.t.:} \quad L_{\text{total}}^{(i,j)}(\tau) \leq L_D(\tau), \quad \forall i \in \mathcal{V}, \forall j \in \mathcal{E} \tag{19b}$$

$$E_{\text{total}}^{(i,j)}(\tau) \leq E_i^{\max}, \quad \forall i \in \mathcal{V} \tag{19c}$$

$$s_c(\tau)\sigma_i \leq C_i(\tau)\tau, \quad \text{if } \alpha_{i,j}(\tau) = 0, \quad \forall i \in \mathcal{V} \tag{19d}$$

$$\sum_{i \in \mathcal{V}} \alpha_{i,j}(\tau) s_c(\tau)\sigma_j \leq C_j(\tau)\tau, \quad \forall j \in \mathcal{E} \tag{19e}$$

$$\sigma_{i,j} \geq \sigma_{\min}, \quad \sigma_{i,j}(\tau), \forall i \in \mathcal{V}, \forall j \in \mathcal{E} \tag{19f}$$

$$p_{i,j}(\tau) \leq P_i^{\max}, \quad \forall i \in \mathcal{V}, \forall j \in \mathcal{E} \tag{19g}$$

$$\mu_i(\tau) \geq \lambda_i^{\max}(\tau), \quad \forall i \in \mathcal{V} \tag{19h}$$

$$\lambda_i^{\max}(\tau) = \hat{\lambda}_i(\tau) + \epsilon_\lambda, \quad \forall i \in \mathcal{V} \tag{19i}$$

$$\alpha_{i,j}(\tau) \in \{0, 1\}, \quad \forall i \in \mathcal{V}, \forall j \in \mathcal{E} \tag{19j}$$

$$\sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{E}} B_{i,j}(\tau) \leq B_{\max}, \quad \forall i \in \mathcal{V}, \ \forall j \in \mathcal{E} \tag{19k}$$

where constraint (19b) ensures that the total latency must not exceed the task deadline for each semantic task. Constraint (19c) guarantees that the total energy consumption of a vehicle must not exceed its available energy budget. Moreover, constraint (19d) and (19e) ensure that the computational capacities of the vehicles and edge servers are not exceeded. Constraint (19f) ensure that semantic tasks meet a minimum semantic efficiency threshold for local and offloaded processing. Furthermore, constraint (19g) indicates the transmission power limit with the maximum transmit power for vehicle i. Constraint (19h) ensures that the service rates are large enough to accommodate the maximum expected task arrival rate under uncertainty. Constraint (19i) characterizes the upper bound task arrival rate by combining the DT's predicted rate with an uncertainty margin. Moreover, (19j) restricts the offloading decision variable as a binary variable. Constraint (19k) ensures bandwidth does not surpass the system's maximum limit.

IV. Proposed Quantum-based DDPG Design

The combinatorial optimization problem in (19a) for semantic optimization in digital twin IoV networks is computationally intractable and impractical with increasing vehicle numbers. We design a quantum-inspired DDPG framework for efficient exploration in high-dimensional environments, ideal for dynamic and stochastic offloading in IoV. This section elaborates on the designed framework.

A. MDP Problem Formulation

First, we define the problem (19a) as a Markov decision process (MDP), which consists of three key components: states, actions, and rewards. The *state space* ($\mathcal{S}$) represents all possible conditions the system can be in, while the *action space* ($\mathcal{A}$) includes all actions the agent can take. The *reward function* ($\mathcal{R}$) specifies the immediate feedback the agent receives for transitioning from one state to another based on its actions. At each time step t, the agent observes the current state $s(t)$, selects an action $a(t)$, and receives a reward $r(t)$ as a result of its decision.

1) State Space: The state space $\mathcal{S}$ provides essential information about the current environment to support informed decision-making. The state S_i for agent i at each time step τ is defined as

$$s(\tau) = \left[L_{\text{total}}^{i,j}(\tau),\ E_{\text{total}}^{i,j}(\tau),\ S_k(\tau),\ R_{i,j}(\tau) \right]. \tag{20}$$

Here, $L_{\text{total}}(\tau)$ represents the overall delay for processing or offloading a task, $E_{\text{total}}(\tau)$ denotes the total energy consumption, $S_k(\tau)$ includes semantic task features, and $R_{i,j}(\tau)$ indicates the achievable transmission rate between vehicle i and edge server j.

2) Action Space: The action space $\mathcal{A}$ can be defined as the set of allowable decisions at each time step τ. Therefore, the decision taken by the agent at time τ can be denoted as

$$a_i(\tau) = \left[B_{i,j}(\tau),\ p_{i,j}(\tau),\ \alpha_{i,j}(\tau) \right]. \tag{21}$$

In this representation, $B_{i,j}(\tau)$ is the bandwidth allocated between vehicle i and edge server j, $p_{i,j}(\tau)$ is the transmission power used by vehicle i for communication with server j, and $\alpha_{i,j}(\tau)$ is a binary variable indicating whether the task is offloaded ($\alpha_{i,j}(\tau) = 1$) or processed locally ($\alpha_{i,j}(\tau) = 0$).

3) Reward Function: The reward function $\mathcal{R}$ can be defined to provide feedback that guides the agent toward optimal performance. In this formulation, the reward can be designed to minimize a weighted combination of task latency and energy consumption, and can be expressed as

$$r(\tau) = -\left(w_1 L_{\text{total}}(\tau) + (1 - w_1) E_{\text{total}}(\tau) \right) - \sum_{i,j} U_{i,j}, \tag{22}$$

where $\sum_{i,j} U_{i,j}$ denotes the penalties associated with constraint violations. The negative sign ensures that lower cost that results in higher rewards. This formulation encourages the agent to make decisions that not only minimize latency and energy consumption but also comply with all system constraints.

B. Quantum-DDPG (Q-DDPG) Framework

Q-DDPG integrates quantum computing with the DDPG framework by employing a parametric quantum circuit (PQC) as the actor network. Moreover, it leverages higher-order encoding to transform the classical state space into a quantum-representable form. The quantum processing pipeline comprises the encoding of classical data, transformation through the PQC, and quantum measurements to extract actionable outputs.

In higher-order encoding, classical input data $\mathbf{x}$ are mapped to an enriched feature space $\Phi(\mathbf{x})$ that captures nonlinear relationships among input features. These transformed features parameterize a quantum circuit designed to embed feature correlations. the unitary matrix of higher order encoding transformation can be represented as $U_{\mathbf{x}} = \exp\left(i \sum_{j<k} \gamma x_j x_k Z_j Z_k \right)$, where $\mathbf{x}$ denotes the classical input vector with elements x_j and x_k, and γ is a tunable parameter controlling interaction strength. The Pauli-Z operators Z_j and Z_k act on the j-th and k-th qubits, respectively. The product $Z_j Z_k$ encodes pairwise feature dependencies, and the exponential form guarantees unitarity. Thus, the encoded

TABLE II: Summary of key notations.

Notation	Definition	Notation	Definition
$s_c(\tau)$	Semantic complexity at time τ	$\sigma_{i,j}$	Semantic efficiency from vehicle i to server j
σ_i	Semantic accuracy of vehicle i	σ_j	Semantic accuracy of server j
$S_k(\tau)$	Semantic task at time τ	$T_{i,p}(\tau)$	Priority of task k generated by vehicle i
$L_D(\tau)$	Deadline of task completion	$\mathbb{P}(A_i(\tau))$	Probability of k tasks arriving at vehicle i
$\lambda_i(\tau)$	Task arrival rate for vehicle i	$\mu_i(\tau)$	Service rate of vehicle i
$\eta_i(\tau)$	Task queue length of vehicle i	$D_i(\tau)$	Data size of the task (in bits)
$C_i(\tau)$	Computational capacity of vehicle i	$C_j(\tau)$	Computational capacity of edge server j
$R_{i,j}(\tau)$	Achievable transmission rate between i and j	$B_{i,j}(\tau)$	Allocated bandwidth between i and j
$p_{i,j}(\tau)$	Transmission power of vehicle i to server j	$h_{i,j}(\tau)$	Channel gain between i and j
$\alpha_{i,j}(\tau)$	Offloading decision variable (binary)	$L^i_{\text{comp}}(\tau)$	Local computation delay for task k
$L^i_{\text{off}}(\tau)$	Offloading and computation delay to server j	$L^{i,j}_{\text{total}}(\tau)$	Total task processing delay
$E^{i,j}_{\text{off}}(\tau)$	Energy for task offloading from i to j	$E^i_{\text{loc}}(\tau)$	Local processing energy at vehicle i
$E^{i,j}_{\text{total}}(\tau)$	Total energy consumption for task k	$\Phi(\tau)$	Total system cost at time τ
γ	Discount factor in RL	$r(\tau)$	Reward function at time τ
κ_i	Energy efficiency coefficient of vehicle i	$DT(\tau)$	Digital Twin state at time τ

quantum state passes through the PQC, where parameterized single-qubit rotations and entangling gates model complex policies. Finally, measurement operations extract expectation values that serve as outputs for the actor network, enabling continuous action selection in the Q-DDPG framework.

C. PQC-based Actor Network

In our model, a hybrid action space in the Q-DDPG framework enables the agent to handle both discrete strategies and continuous control parameters. The actor network, implemented as a PQC, determines the optimal action $a = \mu(s \mid \theta^\mu)$, where s is the state and θ^μ the PQC parameters. The PQC encodes complex policies using parameterized single-qubit rotation gates $R_y(\theta)$, which embed trainable parameters, and controlled-Z (CZ) gates, which introduce entanglement to capture feature correlations. This combination ensures that the PQC can represent expressive quantum states, optimizing the expected return as evaluated by the critic, which can be defined as

$$J(\theta^\mu) = \mathbb{E}_{s\sim p^\pi}[Q(s, \mu(s|\theta^\mu) \mid \theta^Q)]. \tag{23}$$

The policy gradient with respect to the actor parameters can be computed as

$$\nabla_{\theta^\mu} J(\theta^\mu) \approx \mathbb{E}_{s\sim p^\pi}\left[\nabla_a Q(s, a \mid \theta^Q)\nabla_{\theta^\mu}\mu(s \mid \theta^\mu)\right], \tag{24}$$

where $\nabla_{\theta^\mu}\mu(s \mid \theta^\mu)$ is computed based on the PQC's trainable parameters and quantum gate structure.

D. Classical Neural Critic Network

The critic network is modeled as a classical neural network and is responsible for estimating the action-value function $Q(s, a \mid \theta^Q)$, where θ^Q denotes the set of learnable parameters. The training objective of the critic is to minimize the mean-squared error between the predicted Q-values and the target values. Then, the loss function $L(\theta^Q)$ can be expressed as

$$L(\theta^Q) = \mathbb{E}\left[(Q(s_\tau, a_\tau \mid \theta^Q) - y_\tau)^2\right], \tag{25}$$

where the target Q-value can be defined as

$$y_\tau = r + \gamma Q'(s_{\tau+1}, \mu'(s_{\tau+1} \mid \theta^{\mu'}) \mid \theta^{Q'}). \tag{26}$$

Moreover, the gradient descent update for the critic parameters can be computed as

$$\nabla_{\theta^Q} L(\theta^Q) = \mathbb{E}\left[(Q(s_\tau, a_\tau \mid \theta^Q) - y_\tau)\nabla_{\theta^Q} Q(s_\tau, a_\tau \mid \theta^Q)\right]. \tag{27}$$

E. Q-DDPG Target Networks Update

To stabilize training, the use of target networks can be adopted for both actor and critic components. The actor's target network can be implemented using the parameterized quantum circuit which is illustrated in Fig. (2), while the critic's target network can be maintained as a classical neural network. The update of these target networks can be defined as

$$\theta^{\mu'} \leftarrow \rho\theta^\mu + (1-\rho)\theta^{\mu'}, \tag{28}$$

$$\theta^{Q'} \leftarrow \rho\theta^Q + (1-\rho)\theta^{Q'}, \tag{29}$$

where $\rho \in [0, 1]$ is the target update rate. The use of target networks mitigates the instability caused by rapid parameter changes during training.

Moreover, the complexity analysis of a quantum-inspired hybrid actor–critic DDPG model, each training step involves one discrete action and n_c continuous actions. Let n_s represent the state dimension, n_q denote the number of qubits in the quantum circuit, and r indicate the number of ansatz layers. The actor's computation, which evaluates the quantum circuit, has a complexity of $\mathcal{O}(rn_q^2)$, while the additional classical processing needed to produce both discrete and continuous actions is relatively small. On the other hand, the critic processes inputs of size (n_s+n_c+1) and has a forward-pass complexity of $\mathcal{O}(n_s + n_c)$. Combining these, the total computational complexity per training step is $\mathcal{O}(rn_q^2 + n_s + n_c)$.

V. Experimental Results and Discussions

Extensive experiments were carried out to evaluate the performance of the Q-DDPG algorithm using IBM Qiskit. where vehicles were uniformly deployed RSU locations to either compute tasks locally or offload them to a quantum server. Although vehicle positioning influenced network topology and communication dynamics, the simulation also accounted for network conditions, task arrival rates, offloading decisions,

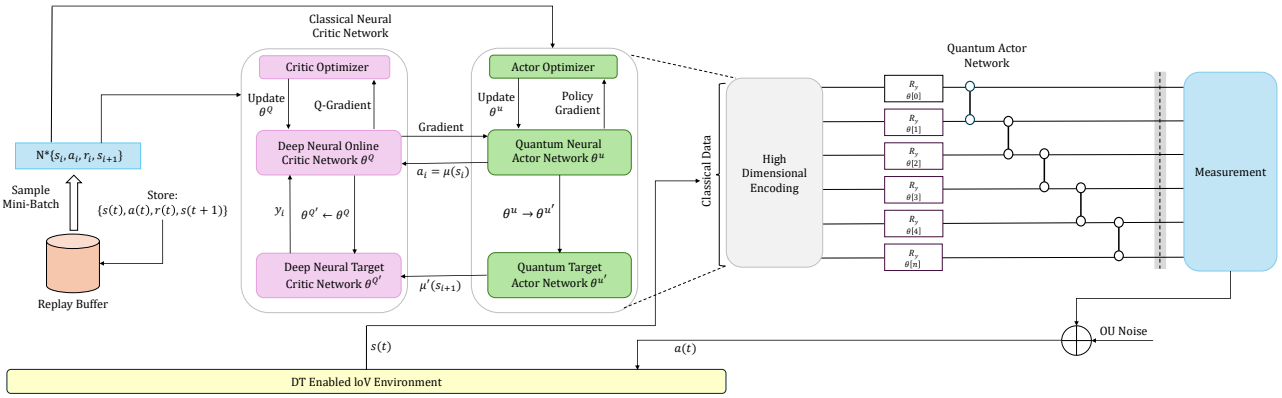

Fig. 2: Q-DDPG network architecture

Algorithm 1: Proposed Q-DDPG Algorithm

Initialize the environment and model parameters.
Initialize the PQC-based actor network $\mu(s \mid \theta^\mu)$ and classical critic network $Q(s, a \mid \theta^Q)$.
Initialize target networks: $\theta^{\mu'} \leftarrow \theta^\mu$, $\theta^{Q'} \leftarrow \theta^Q$.
Initialize the replay buffer $\mathcal{D}$.
Initialize $\tau = 0$.
for $\psi \leftarrow 1$ **to** $\psi_{\max}$ **do**
  Reset the environment and observe initial state s_1.
  for $\tau \leftarrow 1$ **to** $T_{\max}$ *or until terminal state* **do**
    Encode s_τ into a quantum state using higher-order encoding
    Select action $a_\tau = \mu(s_\tau \mid \theta^\mu) + \mathcal{N}(0, \sigma)$, where $\mathcal{N}$ is OU noise.
    Execute a_τ, observe reward r_τ, and next state $s_{\tau+1}$.
    Store $(s_\tau, a_\tau, r_\tau, s_{\tau+1})$ in replay buffer $\mathcal{D}$.
    Sample a mini-batch of transitions from $\mathcal{D}$.
    Compute target Q-value y_τ using (26).
    Update critic network by minimizing loss in (25) via gradient (27).
    Update actor network using policy gradient from (24).
    Soft update target networks using (28):

$$\theta^{Q'} \leftarrow \rho\,\theta^Q + (1-\rho)\,\theta^{Q'}$$
$$\theta^{\mu'} \leftarrow \rho\,\theta^\mu + (1-\rho)\,\theta^{\mu'}$$

energy consumption, and quantum resource allocation. The path loss model was rigorously defined based on previous work [30], and Q-DDPG was tuned with a batch size of 4 and learning rates of 1×10^{-4} and 2×10^{-4} for the actor and critic networks, respectively. Moreover, the number of layers in the PQC is set to 1 and 6 qubits are used with principal component analysis with state space. The buffer size is 1,000,000, and the number of episodes is 500, with each episode consisting of 20 steps. To evaluate the effectiveness of the Q-DDPG framework, we compared it with the state-of-the-art DDPG algorithm [26] under the same simulated conditions. See Table III for other simulation parameters.

TABLE III: Simulation Parameters [10],[28],[29]

Parameter	**Value**
Path loss exponent	2
Number of vehicles	3
Number of edge servers	2
System bandwidth	10 MHz
Noise power density	-70 dBm
Vehicle semantic data size	$1 \times 10^4 - 2 \times 10^4$ bits
Semantic task complexity	$1 \times 10^6 - 2 \times 10^6$ cycles
Weighting factor	0.5
Energy coefficient of processors	1×10^{-27}
Maximum latency	0.05s
Maximum transmit power per vehicle	1 W
Minimum semantic accuracy	0.5

A. Results Discussion

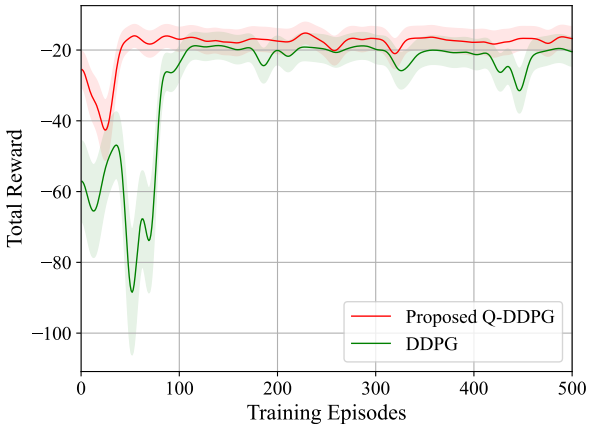

Fig. 3: Total system reward performance

As shown in Fig. 3, the total reward over training episodes for Q-DDPG and DDPG. Q-DDPG converges more quickly and reaches substantially higher final rewards, reflecting its enhanced learning efficiency and optimized offloading strategies.

Its quantum-based approach enables the simultaneous exploration of multiple strategies, significantly accelerating policy optimization. Meanwhile, DTNs provide real-time network feedback, allowing Q-DDPG to dynamically adjust its policy in response to highly volatile IoV conditions. Q-DDPG's reward advantage highlights its robustness and adaptability to network uncertainty.

Moreover, Fig. 4 compares the average cost as a function of the system bandwidth for Q-DDPG and standard DDPG. The proposed Q-DDPG consistently achieves lower costs across all bandwidths (ranging from 3 to 9 MHz). This performance advantage is attributed to the quantum-inspired capabilities of Q-DDPG, which enable the parallel exploration of multiple bandwidth allocation strategies. As a result, Q-DDPG converges more rapidly to optimal offloading policies. The DTN framework further enhances this process by accurately modeling dynamic network states, ensuring adaptive task allocation that minimizes communication costs, even under stringent bandwidth constraints.

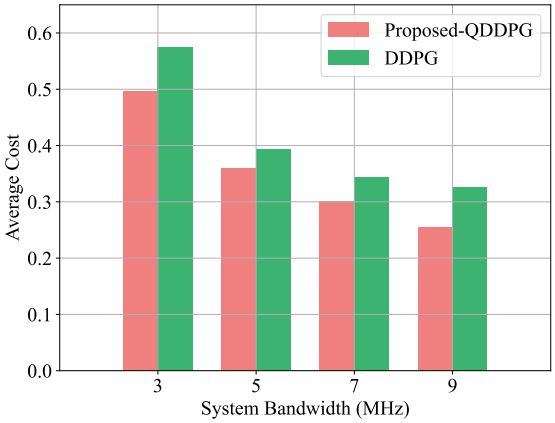

Fig. 4: System bandwidth performance

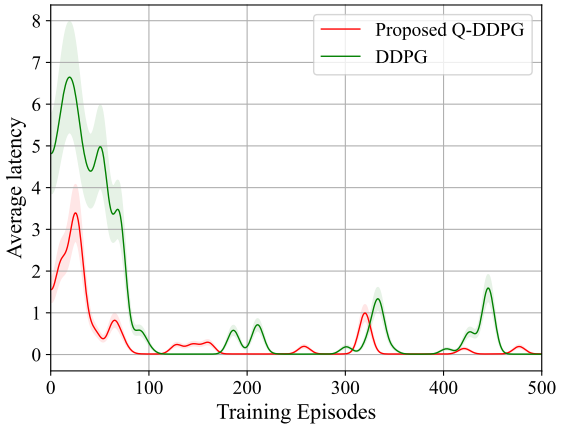

Fig. 5: Average system latency performance

Fig. 5, illustrates the average latency performance of Q-DDPG and DDPG over the course of training episodes. At the beginning of training, both approaches exhibit relatively high latency values. However, Q-DDPG demonstrates a faster reduction in latency, with values decreasing more sharply in the early episodes. As training progresses, Q-DDPG continues to show a consistent decline in latency and reaches a stable state earlier than DDPG. In contrast, DDPG exhibits slower convergence, with higher latency values persisting for a longer duration before eventually stabilizing. Additionally, the latency values of Q-DDPG remain more stable, with fewer fluctuations compared to DDPG, which displays greater variability throughout training.

Moreover, in Fig. 6 shows the average cost trends for Q-DDPG and DDPG over training episodes. Both approaches start with relatively high cost values. Q-DDPG exhibits a sharper decrease in cost during the initial episodes, while DDPG reduces its cost more gradually. As training advances, Q-DDPG stabilizes at a lower cost earlier, whereas DDPG requires more episodes to settle. This highlights that Q-DDPG maintains more consistent performance with fewer fluctuations compared to the varying patterns observed in DDPG.

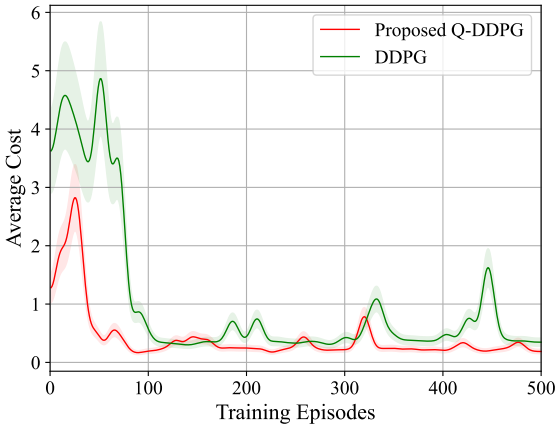

Fig. 6: Average system cost performance

Fig. 7 illustrates the average cost performance of Q-DDPG under two different computational power settings at edge servers, 10 GHz and 5 GHz. The results indicate that as task complexity increases, both configurations experience a gradual rise in cost. However, the 10 GHz configuration consistently achieves lower costs compared to the 5 GHz configuration across all complexity levels. Moreover, this highlights the impact of computational power on system performance when handling tasks of varying complexity. Systems with higher computational power demonstrate better efficiency in managing increased task demands, resulting in reduced costs. The trend further emphasizes the ability of the system to adapt to different resource levels while maintaining stable performance.

Furthermore, Fig. 8 presents the average cost performance of Q-DDPG and DDPG over varying task complexities, measured in mega cycles. As task complexity increases, both methods show a gradual rise in cost. However, Q-DDPG consistently maintains lower costs compared to DDPG across all complexity levels. Furthermore, the results also show that the cost increase for Q-DDPG follows a smooth and steady trend, whereas DDPG exhibits a more linear and higher

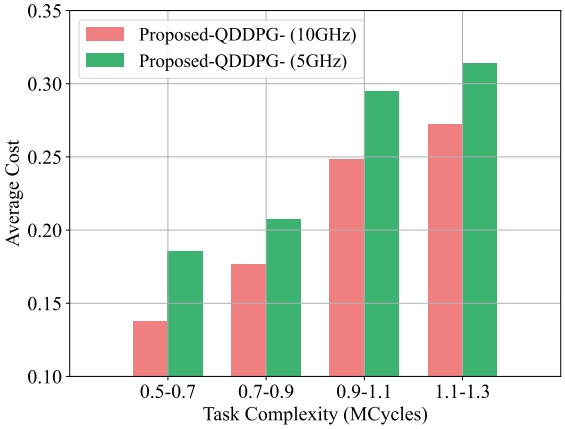

Fig. 7: Task complexity performance for frequency variations.

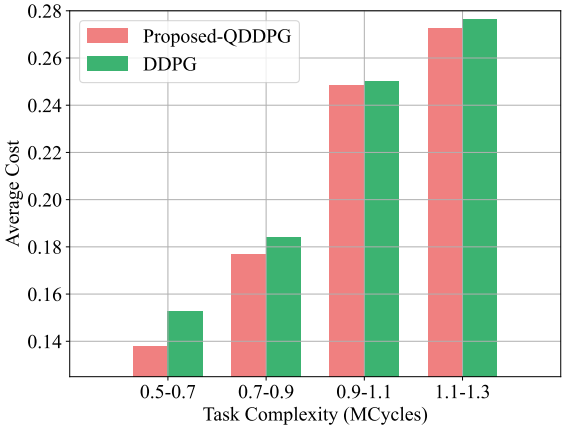

Fig. 8: Task complexity performance for algorithms comparison

cost pattern. The gap between the two methods becomes more evident at higher complexity levels, where Q-DDPG continues to demonstrate lower costs. This suggests that Q-DDPG handles increasing task complexity more effectively than DDPG.

Fig. 9 depits the average cost performance of Q-DDPG with and without DT across different system bandwidth values, measured in MHz. As the system bandwidth increases, both configurations experience a gradual decrease in cost. However, Q-DDPG with DT consistently achieves lower costs compared to Q-DDPG without DT at all bandwidth levels. This trend highlights the effectiveness of DT in optimizing performance, especially as more bandwidth becomes available. The results further suggest that incorporating DT enables the system to better utilize available resources, leading to improved efficiency and lower costs under varying bandwidth conditions. Additionally, the cost gap between the two configurations becomes more noticeable as bandwidth increases, indicating that the benefits of DT become more significant in higher bandwidth environments. This observation emphasizes the

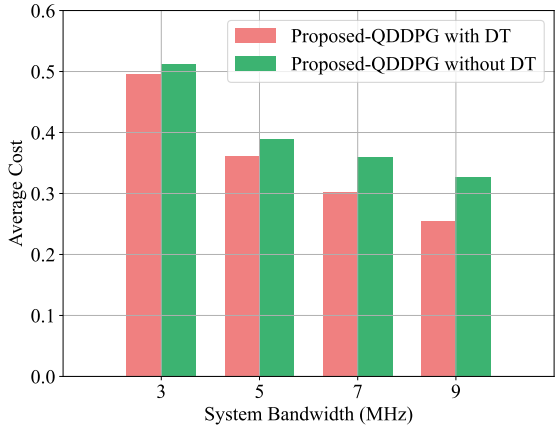

Fig. 9: DTN performance

ability of DT to enhance adaptability and scalability, making it particularly useful in scenarios with dynamic network resources.

VI. Concluding Remarks

This study proposes a Q-DDPG-based framework for semantic optimization in DT-enabled IoV networks, focusing on optimized resource allocation and semantic processing in dynamic vehicular environments. By leveraging a quantum-inspired algorithm and the capabilities of DT, the framework enhances semantic accuracy, reduces energy consumption, and minimizes latency, ensuring both scalability and robustness in highly variable network conditions. The integration of quantum computing with semantic communication enables simultaneous optimization of offloading strategies and communication efficiency, addressing the challenges of real-time adaptability and network uncertainty. Furthermore, this approach lays the foundation for hybrid optimization techniques, seamless security integration, and practical deployment in real-world IoV scenarios, making it a transformative solution for next-generation intelligent transportation systems.

References

[1] K. R. Reddy and A. Muralidhar, "Machine learning-based road safety prediction strategies for Internet of Vehicles (IoV) enabled vehicles: A systematic literature review," *IEEE Access*, vol. 11, pp. 112 108–112 122, Sep. 2023.

[2] J. A. Ansere, G. Han, and H. Wang, "A novel reliable adaptive beacon time synchronization algorithm for large-scale vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 11 565–11 576, Dec. 2019.

[3] C.-C. Chang, Y.-M. Ooi, and B.-H. Sieh, "IoV-based collision avoidance architecture using machine learning prediction," *IEEE Access*, vol. 9, pp. 115 497–115 505, Aug. 2021.

[4] A. Waheed, M. A. Shah, S. M. Mohsin, A. Khan, C. Maple, S. Aslam, and S. Shamshirband, "A comprehensive review of computing paradigms, enabling computation offloading and task execution in vehicular networks," *IEEE Access*, vol. 10, pp. 3580–3600, Jan. 2022.

[5] Z. Liu, H. Sun, G. Marine, and H. Wu, "6G IoV networks driven by RF digital twin modeling," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 3, pp. 2976–2987, Mar 2024.

[6] L. Jiang, H. Zheng, H. Tian, S. Xie, and Y. Zhang, "Cooperative federated learning and model update verification in blockchain-empowered digital twin edge networks," *IEEE Internet of Things J.*, vol. 9, pp. 11 154–11 167, Jul. 2022.

[7] J. Zheng, T. H. Luan, Y. Hui, Z. Yin, N. Cheng, L. Gao, and L. X. Cai, "Digital twin empowered heterogeneous network selection in vehicular networks with knowledge transfer," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 12 154–12 167, Nov. 2022.

[8] M. U. Lokumarambage, V. S. S. Gowrisetty, H. Rezaei, T. Sivalingam, N. Rajatheva, and A. Fernando, "Wireless end-to-end image transmission system using semantic communications," *IEEE Access*, vol. 11, pp. 37 149–37 165, Apr. 2023.

[9] Y. Yigit, L. A. Maglaras, W. J. Buchanan, B. Canberk, H. Shin, and T. Q. Duong, "AI-enhanced digital twin framework for cyber-resilient 6G Internet of Vehicles networks," *IEEE Internet of Things J.*, vol. 11, no. 22, pp. 36 168–36 181, Nov 2024.

[10] J. Zheng, Y. Zhang, T. H. Luan, P. K. Mu, G. Li, M. Dong, and Y. Wu, "Digital twin enabled task offloading for IoVs: A learning-based approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 1, pp. 659–670, Jan 2024.

[11] X. Yuan, J. Chen, N. Zhang, J. Ni, F. R. Yu, and V. C. M. Leung, "Digital twin-driven vehicular task offloading and IRS configuration in the Internet of Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24 290–24 304, Sep. 2022.

[12] N. P. Kuruvatti, M. A. Habibi, S. Partani, B. Han, A. Fellan, and H. D. Schotten, "Empowering 6G communication systems with digital twin technology: A comprehensive survey," *IEEE Access*, vol. 10, pp. 112 158–112 181, Oct. 2022.

[13] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, "Low-latency federated learning and blockchain for edge association in digital twin empowered 6G networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, pp. 5098–5111, Jul. 2021.

[14] N. V. Huynh, B. Zhang, D.-H. Tran, D. T. Hoang, D. N. Nguyen, G. Zheng, D. Niyato, and Q.-V. Pham, "Dynamic spectrum access for ambient backscatter communication-assisted D2D systems with quantum reinforcement learning," *arXiv preprint arXiv:2410.17971*, Oct. 2024.

[15] S. J. Nawaz, S. K. Sharma, S. Wyne, M. N. Patwary, and M. Asaduzzaman, "Quantum machine learning for 6G communication networks: State-of-the-art and vision for the future," *IEEE Access*, vol. 7, pp. 46 317–46 350, Apr. 2019.

[16] T. Q. Duong, L. D. Nguyen, B. Narottama, J. A. Ansere, D. V. Huynh, and H. Shin, "Quantum-inspired real-time optimization for 6G networks: Opportunities, challenges, and the road ahead," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1347–1359, Aug. 2022.

[17] B. Narottama and S. Y. Shin, "Quantum neural networks for resource allocation in wireless communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1103–1116, Feb. 2022.

[18] B. Narottama, T. Jamaluddin, and S. Y. Shin, "Quantum neural network with parallel training for wireless resource optimization," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5835–5847, May 2024.

[19] J. A. Ansere, S. C. Prabhashana, , O. A. Dobre, and T. Q. Duong, "Quantum machine learning DDPG for digital twin semantic vehicular networks," in *Proc. IEEE Int. Conf. Mach. Learn. Commun. Netw. (ICMLN'25)*, Barcelona, Spain, May 2025.

[20] Silvirianti, B. Narottama, and S. Y. Shin, "Layerwise quantum deep reinforcement learning for joint optimization of UAV trajectory and resource allocation," *IEEE Internet of Things J.*, vol. 11, no. 1, pp. 430–443, Jan. 2024.

[21] C. Park, W. J. Yun, J. P. Kim, T. K. Rodrigues, S. Park, S. Jung, and J. Kim, "Quantum multi-agent actor-critic networks for cooperative mobile access in multi-UAV systems," *IEEE Internet of Things J.*, vol. 10, no. 22, pp. 20 033–20 048, Nov. 2023.

[22] G. S. Kim, Y. Cho, J. Chung, S. Park, S. Jung, Z. Han, and J. Kim, "Quantum multi-agent reinforcement learning for cooperative mobile access in space-air-ground integrated networks," *arXiv preprint arXiv:2406.16994*, Jun. 2024.

[23] J. A. Ansere, T. Q. Duong, S. R. Khosravirad, V. Sharma, A. Masaracchia, and O. A. Dobre, "Quantum deep reinforcement learning for 6G mobile edge computing-based IoT systems," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Marrakesh, Morocco, Jun. 2023, pp. 406–411.

[24] S. Y.-C. Chen, "Quantum deep q-learning with distributed prioritized experience replay," in *Proc. IEEE Int. Conf. Quantum Comput. Eng. (QCE)*, Bellevue, WA, Sep. 2023, pp. 31–35.

[25] J. Zhang, G. Zheng, T. Koike-Akino, K.-K. Wong, and F. A. Burton, "Hybrid quantum-classical neural networks for downlink beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 16 498–16 512, Nov. 2024.

[26] Y. Guo, D. Ma, H. She, G. Gui, C. Yuen, H. Sari, and F. Adachi, "Deep deterministic policy gradient-based intelligent task offloading for vehicular computing with priority experience playback," *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 10 655–10 667, Jul 2024.

[27] J. Zhao, H. Quan, M. Xia, and D. Wang, "Adaptive resource allocation for mobile edge computing in Internet of Vehicles: A deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 73, no. 4, pp. 5834–5847, Apr 2024.

[28] L. Liu and Z. Chen, "Joint optimization of multiuser computation offloading and wireless-caching resource allocation with linearly related requests in vehicular edge computing system," *IEEE Internet of Things J.*, vol. 11, no. 1, pp. 1534–1547, Jan 2024.

[29] F. Z. Ruskanda, M. R. Abiwardani, R. Mulyawan, I. Syafalni, and H. T. Larasati, "Quantum-enhanced support vector machine for sentiment classification," *IEEE Access*, vol. 11, pp. 87 520–87 535, Aug. 2023.

[30] D. Wang, B. Song, P. Lin, F. R. Yu, X. Du, and M. Guizani, "Resource management for edge intelligence (EI)-assisted IoV using quantum-inspired reinforcement learning," *IEEE Internet of Things J.*, vol. 9, no. 14, pp. 12 588–12 600, Jul. 2022.

[31] D. Liu, W. Wang, L. Wang, H. Jia, and M. Shi, "Dynamic pricing strategy of electric vehicle aggregators based on DDPG reinforcement learning algorithm," *IEEE Access*, vol. 9, pp. 21 556–21 566, Jan. 2021.

[32] S. Siboo, A. Bhattacharyya, R. Naveen Raj, and S. H. Ashwin, "An empirical study of DDPG and PPO-based reinforcement learning algorithms for autonomous driving," *IEEE Access*, vol. 11, pp. 125 094–125 108, Nov. 2023.

[33] U. Khalid, M. S. Ulum, A. Farooq, T. Q. Duong, O. A. Dobre, and H. Shin, "Quantum semantic communications for metaverse: Principles and challenges," *IEEE Trans. Wireless Commun.*, vol. 30, no. 4, pp. 26–36, Aug. 2023.

[34] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, "Toward semantic communications: Deep learning-based image semantic coding," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 55–71, Nov. 2023.

[35] B. Tang, L. Huang, Q. Li, A. Pandharipande, and X. Ge, "Cooperative semantic communication with on-demand semantic forwarding," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 349–363, Dec. 2024.

[36] Y. Li and X. Tong, "Trust recommendation based on deep deterministic strategy gradient algorithm," *IEEE Access*, vol. 10, pp. 48 274–48 282, Apr. 2022.

[37] M. Jafari, A. Kavousi-Fard, T. Chen, and M. Karimi, "A review on digital twin technology in smart grid, transportation system and smart city: Challenges and future," *IEEE Access*, vol. 11, pp. 17 471–17 484, Feb. 2023.

[38] J. Zhao, Y. Chen, and Y. Huang, "QoE-driven wireless communication resource allocation based on digital twin edge network," *IEEE Trans. Radio Freq. Interference*, vol. 8, pp. 277–281, Sep. 2024.

[39] D. Wheeler and B. Natarajan, "Engineering semantic communication: A survey," *IEEE Access*, vol. 11, pp. 13 965–13 995, Feb. 2023.

[40] T. Q. Duong, L. D. Nguyen, B. Narottama, J. A. Ansere, D. V. Huynh, and H. Shin, "Quantum-inspired real-time optimization for 6G networks: Opportunities, challenges, and the road ahead," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1347–1359, Aug. 2022.

[41] J. A. Ansere, E. Gyamfi, Y. Li, H. Shin, O. A. Dobre, T. Hoang, and T. Q. Duong, "Optimal computation resource allocation in energy-efficient edge IoT systems with deep reinforcement learning," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 4, pp. 2130–2142, Jun. 2023.

[42] A. Paul, K. Singh, C.-P. Li, O. A. Dobre, and T. Q. Duong, "Digital twin-aided vehicular edge network: A large-scale model optimization by quantum-DRL," *IEEE Trans. Veh. Technol.*, pp. 1–17, Jun. 2024.

[43] X. Zhou, X. Zhang, H. Zhao, J. Xiong, and J. Wei, "Constrained soft actor-critic for energy-aware trajectory design in UAV-aided IoT networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1414–1418, May 2022.

[44] R. Wang, M. Shen, Y. He, and X. Liu, "Joint access points-user association and caching placement strategy for cell-free massive MIMO systems based on soft actor-critic algorithm," *IEEE Commun. Lett.*, vol. 28, no. 2, pp. 347–351, Dec. 2024.

[45] A. Bayuwindra, L. Wonohito, and B. R. Trilaksono, "Design of DDPG-based extended look-ahead for longitudinal and lateral control of vehicle platoon," *IEEE Access*, vol. 11, pp. 96 648–96 660, Sep. 2023.

[46] T. Yu, X. Wang, J. Hu, and J. Yang, "Multi-agent proximal policy optimization-based dynamic client selection for federated ai in 6G-oriented Internet of Vehicles," *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, pp. 13 611–13 624, Apr. 2024.

[47] H. Zhou, Z. Zhang, Y. Wu, M. Dong, and V. C. M. Leung, "Energy efficient joint computation offloading and service caching for mobile edge computing: A deep reinforcement learning approach," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 950–961, Jul. 2023.