Quality of Service-Oriented Data Optimization in Networks using Artificial Intelligence Techniques

Zhenhua Yang[1#], Qiwen Yang[2#], Minghong Yang[3]*
School of Information Engineering, Hunan Applied Technology University, Changde, China[1]
School of Computer Science, Beihang University, Beijing, China[2]
School of Economics and Management, Hunan Applied Technology University, Changde, China[3]

***Abstract*—This paper outlines a comprehensive AI-driven Quality of Service (QoS) optimization method, presenting a rigorous examination of its effectiveness through extensive experimentation and analysis. By applying real-world datasets to simulate network environments, the study systematically evaluates the proposed method's impact across various QoS metrics. Key findings reveal substantial enhancements in reducing average latency, minimizing packet loss, and boosting bandwidth utilization compared to baseline scenarios, with the Deep Deterministic Policy Gradient (DDPG) model showcasing the most notable improvements. The research demonstrates that AI optimization strategies, particularly those leveraging DQN and DDPG algorithms, significantly improve upon conventional methods. Specifically, post-migration optimizations lead to a recovery and even surpassing of pre-migration QoS levels, with delays dropping to levels below initial readings, packet loss nearly eliminated, and bandwidth utilization markedly improved. The study further illustrates that while lower learning rates necessitate longer convergence times, they ultimately facilitate superior model performance and stability. In-depth case studies within a cloud data center setting underscore the system's proficiency in handling large-scale Virtual Machine (VM) migrations with minimal disruption to network performance. The AI-driven optimization successfully mitigates the typical latency spikes, packet loss increases, and resource utilization dips associated with VM migrations, thereby affirming its practical value in maintaining high network efficiency and stability during such operations. Comparative analyses against traditional traffic engineering methods, rule-based controls, and other machine learning approaches consistently place the AI optimization method ahead, achieving up to an 8% increase in throughput alongside a 2 ms decrease in latency. Furthermore, the technique excels in reducing packet loss by 25% and elevating resource utilization rates, underscoring its prowess in enhancing network efficiency and stability. Robustness and scalability assessments validate the method's applicability across diverse network scales, traffic patterns, and congestion levels, confirming its adaptability and effectiveness in a wide array of operational contexts. Overall, the research conclusively evidences the AI-driven QoS optimization system's capacity to tangibly enhance network performance, positioning it as a highly efficacious solution for contemporary networking challenges.**

Keywords*—Artificial intelligence; networking; quality of service-oriented; data optimization

*Corresponding Author.
# Indicates co-first author, Zhenhua Yang and Qiwen Yang contributed equally to this work.

I. Introduction

Today's world is undergoing unprecedented digital transformation, and the iterative upgrading of information technology is constantly reshaping the economic structure and social life. From smart homes to smart cities, from distance education to telemedicine, every emerging application puts higher requirements on network service quality. The network is not only a pipeline for data transmission but also a nervous system that supports the operation of society. Therefore, ensuring the efficient, stable and secure operation of the network is directly related to the effectiveness and sustainability of digital transformation [1].

With the commercial deployment of 5G technology and the initial launch of 6G research and development, mobile communications have entered a whole new stage of development. Higher data rates, lower latency, and greater connection density are features that make cutting-edge applications such as autonomous driving, Industry 4.0, and immersive entertainment possible. However, at the same time, these applications demand an unprecedented level of network QoS. How to adjust network resource allocation in real-time and precisely to meet the differentiated demands of various applications in a complex and changing network environment has become a key issue to be solved. Traditional network management relies on preset rules and manual intervention, which is difficult to adapt to the dynamic changes and complexity of the modern network environment. Statically configured policies are often unable to flexibly respond to unexpected traffic, network congestion or failure events, resulting in QoS degradation and impaired user experience [2].

The rise of artificial intelligence, especially machine learning and deep learning, has provided new ideas and tools for network QoS optimization.AI is able to process massive amounts of network data, learn network behavior patterns, predict traffic trends, and automatically optimize network configurations so as to achieve the purpose of improving resource utilization, reducing latency, and enhancing stability. Although AI has great potential in network QoS optimization, the path to its realization is not smooth. How to effectively combine AI algorithms with network engineering practices, how to realize data-driven decision making while safeguarding privacy and security, how to address the interpretive issues of models to enhance trust, and how to harmonize across different network architectures (e.g., cloud, edge, and end) are the main challenges currently faced. Therefore, in-depth research on the application of AI in QoS optimization is not only a need for technological innovation, but also an inevitable choice to promote a robust digital society [3].

In recent years, the research on network QoS optimization and AI applications in communication networks has made significant progress. Early work focused on the establishment of QoS models and the application of traditional optimization algorithms, e.g., Ghafoor et al. [1] explored the QoS guarantee mechanism based on the DiffServ model, while Babaei et al. [2] analyzed the application and limitations of the IntServ model in multimedia transmission. Subsequently, with the development of AI technology, the research focus has gradually shifted to utilizing AI to enhance network performance [3].

In terms of traffic prediction, Alkanhel et al. [4] proposes a network traffic prediction model based on deep learning, which effectively improves the prediction accuracy and provides a basis for resource scheduling. A breakthrough has also been made in the application of AI in the field of resource allocation, and Malhotra et al. [5] demonstrates a dynamic spectrum allocation scheme based on reinforcement learning, which significantly improves the spectrum utilization. In addition, AI also shows great potential in fault detection and self-healing network construction, e.g., the AI-assisted fault management system developed in the Bendavid et al. [6] is able to realize rapid localization and repair of network problems. Nevertheless, there are still some insufficiently addressed issues in existing research, such as the interpretability of AI models, generalization capabilities, and the challenges of applying them in large-scale heterogeneous network environments. In addition, how to efficiently integrate AI with traditional network management frameworks, as well as to ensure the transparency and security of AI decisions, are also important issues in current research [7]. This research is dedicated to analyzing the potential of Artificial Intelligence (AI) in the field of Quality of Service (QoS) optimization, focusing on three core issues: first, to address the specific challenges of large-scale network environments, the research aims to design and implement an AI-driven QoS optimization framework to ensure that the framework can adapt to the high dynamics and complexity of network environments, and at the same time effectively enhance the deployment of QoS optimization frameworks in large-scale networks, and to improve the efficiency of QoS optimization. Performance in large-scale networks. Second, the study will explore in detail specific applications of deep learning and reinforcement learning models in accurately predicting network behavioral patterns, implementing intelligent resource scheduling, and further optimizing the strategies of these models to maximize the utilization efficiency of network resources and the quality of service delivery.

The strengths of this paper include a comprehensive AI-driven QoS optimization method that is supported by extensive experimentation and analysis using real-world datasets. The research systematically evaluates the method's impact on various QoS metrics and demonstrates significant improvements compared to baseline scenarios. The study also includes in-depth case studies within a cloud data center setting, showcasing the system's ability to handle large-scale VM migrations with minimal disruption to network performance. Comparative analyses consistently show the AI optimization method outperforming traditional traffic engineering methods, rule-based controls, and other machine learning approaches. Additionally, the paper validates the method's robustness and scalability across diverse network scales, traffic patterns, and congestion levels, confirming its adaptability and effectiveness in various operational contexts.

The paper is organized as follows: The literature review in Section II provides an overview of existing research on AI-based QoS optimization, covering the application of machine learning, deep learning, and reinforcement learning techniques. Finally, the current challenges and the future research directions are proposed. The "AI-Driven QoS Optimization Methodology" in Section III details the proposed AI-driven QoS optimization methodology. It is divided into two sections: A. Problem modeling: This section establishes the mathematical model of QoS optimization problem, including the objective function, constraint conditions and symbol definition. B. Method Design: This section describes the design of a framework based on Deep Reinforcement Learning (DRL), including Deep Q-Networks (DQN) and an actor criticism architecture. It covers the algorithm principle, architecture design and parameter adjustment strategy. The experimental design in Section IV describes the simulation setup using real data sets, and compares the QoS metrics before and after optimization to prove the effectiveness of the proposed AI optimization method in Section V. Finally, the paper is concluded in Section VI.

II. Literature Review

A. *State of the Artificial Intelligence in QoS Data Optimization*

In recent years, the rapid development of Artificial Intelligence (AI) techniques, especially Machine Learning (ML), Deep Learning (DL), and Reinforcement Learning (RL), has opened up new research paths and practice areas for network Quality of Service (QoS) data optimization. The introduction of AI has enabled network management to move toward intelligence and automation, which it helps to build a self-optimizing and self-healing resilient network, and its specific application mode is shown in Fig. 1.

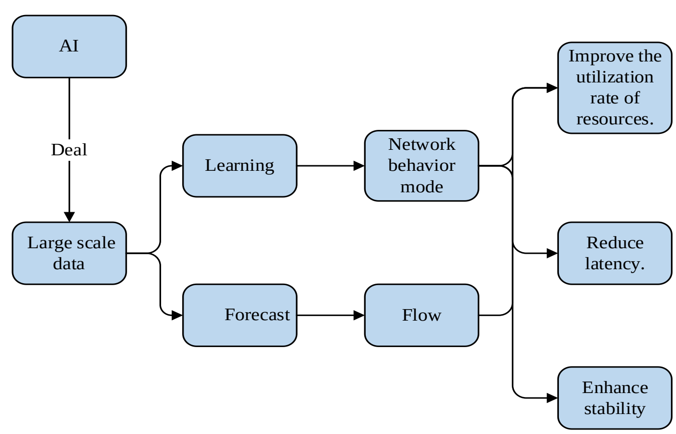

Fig. 1. Application model of AI in QoS optimization.

This section will provide insights into how these techniques can be applied to forecasting, decision making, and dynamic management of network resources with a view to achieving efficient, low-latency, and high-reliability data transmission. Machine learning techniques are able to identify complex network behavior patterns by analyzing historical network data in order to predict future network conditions. Kwon et al. [7] used supervised learning methods to build models that successfully predicted network traffic fluctuations, providing network administrators with a valuable window to adjust resource allocations in advance. In addition, unsupervised learning and semi-supervised learning also show potential in anomaly detection and pattern recognition, which can help to detect and respond to anomalous behaviors in the network in a timely manner and maintain QoS standards [8]. Deep learning, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), excel in processing sequential data and high-dimensional features, and are widely used for optimization of network data. Wang et al. [9] demonstrated how RNN can effectively predict network congestion, while Arunachalam et al. [10] modeled network traffic by introducing a long short-term memory network (LSTM), which not only improves the prediction accuracy, but also dynamically adapts the data transmission strategy based on the prediction results to reduce delay and packet loss. These deep learning models are able to handle the time series characteristics of network data and provide more refined decision support for QoS optimization. Reinforcement learning has found a place in network resource management and scheduling with its ability to make decisions for optimization in complex environments. Mehraban et al. [11] proposed a dynamic bandwidth allocation algorithm based on reinforcement learning, which is capable of adjusting the policy according to the immediate feedback of the network state and realizing the efficient allocation of resources. In addition, Karasik et al. [12] trained the reinforcement learning model by simulating the environment, enabling the network to adaptively adjust the routing policy under different service demands and network conditions, improving the overall QoS performance. The introduction of reinforcement learning enables the network optimization strategy to adapt more flexibly to changes in the network state, realizing the transition from reactive to proactive optimization. The amount of literature on the application of different techniques in QoS data optimization is specifically shown in Fig. 2.

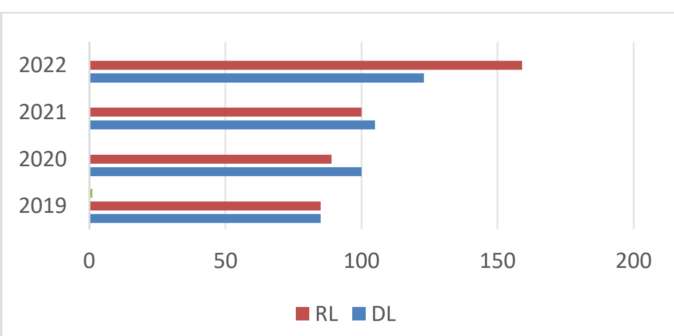

Fig. 2. Number of literature on the application of different techniques in QoS data optimization.

Although AI techniques have made significant achievements in QoS optimization, their practical application still faces a series of challenges, such as model interpretability, acquisition and quality of training data, and computational complexity of algorithms. Rani et al. [13] emphasized the importance of model interpretability in real-world deployments, which is crucial for establishing regulatory trust and troubleshooting. Meanwhile, Can et al. [14] discussed how to effectively collect and utilize network data for model training while protecting user privacy. The specific research findings are shown in Table I.

TABLE I. SUMMARY OF RESEARCH RESULTS

Research Area	Authors and References	Contributions
Traffic Prediction	Alkanhel et al. [4]	Proposes a deep learning-based network traffic prediction model, significantly enhancing prediction accuracy and providing robust support for proactive resource scheduling.
Resource Allocation	Malhotra et al. [5]	Demonstrates a reinforcement learning-driven dynamic spectrum allocation scheme, vastly improving the efficiency of spectrum resource utilization.
QoS Assurance Mechanisms	Ghafoor et al. [1], Babaei et al. [2]	Respectively explore QoS assurance mechanisms based on the DiffServ model and the application of the IntServ model in multimedia transmission, enriching the theoretical and practical aspects of QoS management.
Fault Detection and Self-Healing Networks	Bendavid et al. [6]	Develops an AI-assisted fault management system capable of rapid issue localization and repair, enhancing operational efficiency.

B. Deepening Analysis of QoS-Oriented Intelligent Data Transmission Strategies

In this section, this paper will further delve into intelligent data transmission strategies, in particular how to refine and optimize the data transmission process through advanced AI techniques to ensure superior quality of service (QoS) in the network. This paper will focus on three key areas: intelligent routing, dynamic adaptive transmission techniques, and the integrated application of AI in end-to-end QoS assurance, while also discussing the challenges and future directions of these strategies.

Intelligent routing is a core component of AI-based network optimization strategies. While traditional routing protocols tend to decide the forwarding path of packets based on simple path costs, AI techniques, especially deep and reinforcement learning, can provide more dynamic and strategic routing decisions. For example, Li and Zhang [15] proposed a routing algorithm based on deep reinforcement learning, which can dynamically adjust the routing path according to the network state and traffic demand, effectively reducing network congestion and improving transmission efficiency. Intelligent routing not only considers direct QoS metrics such as delay and packet loss, but also learns and predicts the future state of the network to achieve forward-looking route optimization.

Dynamic adaptive transmission technique is a key strategy to automatically adjust data transmission parameters (e.g., coding rate, slice size, etc.) for different network conditions and application requirements. In application scenarios such as video streaming and real-time communication, Chen et al. [16] realizes real-time monitoring and prediction of network conditions by integrating machine learning models, and dynamically adjusts transmission strategies to maintain the best user experience. End-to-end QoS guarantees require performance optimization across the entire data transmission link, from the data source to the destination. The application of AI techniques at this level, as shown in Kimbugwe et al.'s study [17], achieves optimal allocation of resources by constructing a global optimization model, which integrates multiple QoS metrics in the network. In addition, AI can help achieve cross-layer optimization, i.e., building bridges between the physical, network and application layers to ensure overall QoS consistency and reliability. This chain-wide intelligent management is an important trend in future network service assurance.

Although AI shows great potential in intelligent data transfer strategies, it still faces many challenges, including but not limited to model complexity and interpretability issues, data privacy protection, and robustness in dynamic and heterogeneous network environments. To further advance the application of AI techniques in QoS optimization, future research needs to explore more efficient model training methods, enhance model interpretability, ensure data processing privacy, and develop adaptive AI models that can adapt to rapid changes in the network environment.

C. Recent Advances and Future Trends in Artificial Intelligence for QoS Optimization

In recent years, researchers are no longer limited to a single AI technique, but explore the integration of multiple advanced AI models and algorithms with the aim of achieving deeper intelligence in QoS optimization. For example, Huang and Li [18] combined deep learning and reinforcement learning to develop a hybrid model for achieving more accurate network traffic prediction and resource scheduling, which significantly improved network efficiency and user experience. This trend of cross-domain technology convergence is not limited to the algorithms themselves, but also includes deep integration with network theory, providing unprecedented accuracy and flexibility for QoS optimization.

With the deep application of AI technology in QoS optimization, the "black-box" nature of its decision-making has become a problem that cannot be ignored. To address this challenge, research has begun to favor the development of highly interpretable AI models to enhance the transparency and controllability of network management. In Yang et al.'s study [19], the authors propose an explanatory machine learning-based approach that optimizes network parameters while providing clear explanations of the decision-making process, facilitating network administrators to understand and trust the AI-generated policies, and promoting the practical application and acceptance of the technology.

Facing the upcoming 6G era, the network architecture will be more complex and the service demands will be more diversified. Therefore, how to design an AI-driven QoS optimization framework adapted to future network characteristics has become a hot research topic. Khasawneh et al. [20] explored how to utilize AI technology to achieve QoS assurance with ultra-low latency, high reliability and large-scale connectivity in a 6G network environment, and proposed an intent-driven network management framework based on an intent-driven network management framework, which is able to automatically adjust the network configuration according to the user's intent and service level agreements (SLAs) to ensure end-to-end QoS consistency.

With the in-depth application of AI in QoS optimization, data security and user privacy protection become issues that cannot be ignored. Osman et al. [21] explored how to ensure the secure transmission and processing of data by means of encryption technology and differential privacy while guaranteeing QoS, and how to design privacy-protecting AI models to reduce the reliance on users' personal information, which is crucial for enhancing user trust and promoting the application of AI in QoS optimization.

III. AI-DRIVEN QOS OPTIMIZATION METHODS

A. Problem Modeling

In this section, the mathematical model of the problem will be elaborated in detail, including the establishment of the objective function, the setting of constraints, and the introduction of the necessary notational definitions, with a view to forming a comprehensive and rigorous modeling framework, which is shown in Fig. 3. In this model, N denotes the total number of nodes in the network. e denotes the set of edges in the network, and each edge $e \in E$ associates two nodes and denotes the data transmission path. C_e denotes the capacity of edge e, i.e., the maximum data transfer rate. d_{ij} denotes the delay from node i to node j. f_{ij} denotes the data traffic flowing through edge e=(i, j). q denotes the set of quality of service metrics, including but not limited to average delay, packet loss rate, throughput, etc. w_Q denotes the weight of the quality of service metrics Q, which reflects the importance of each metric to the overall optimization objective. r denotes the set of available resources, including bandwidth, computational resources, etc. x denotes the set of decision variables, which represent policy parameters such as resource allocation, routing, etc.

Our objective is to maximize the integrated quality of service metrics, considering the possible conflicts between different QoS metrics, a weighted summing approach is used to combine them. The objective function can be expressed as Eq. (1).

$$\max_x \sum_{Q \in Q} w_Q \cdot U_Q(x) \tag{1}$$

The constraints include resource constraints, quality of service constraints, traffic conservation non-negative traffic and so on.

1) Resource constraints: Ensure that all resource allocations do not exceed the total amount, for example, for bandwidth resources. This is shown in Eq. (2).

$$\sum_{(i,j)\in E} f_{ij} \le C_e, \quad \forall e \in E \tag{2}$$

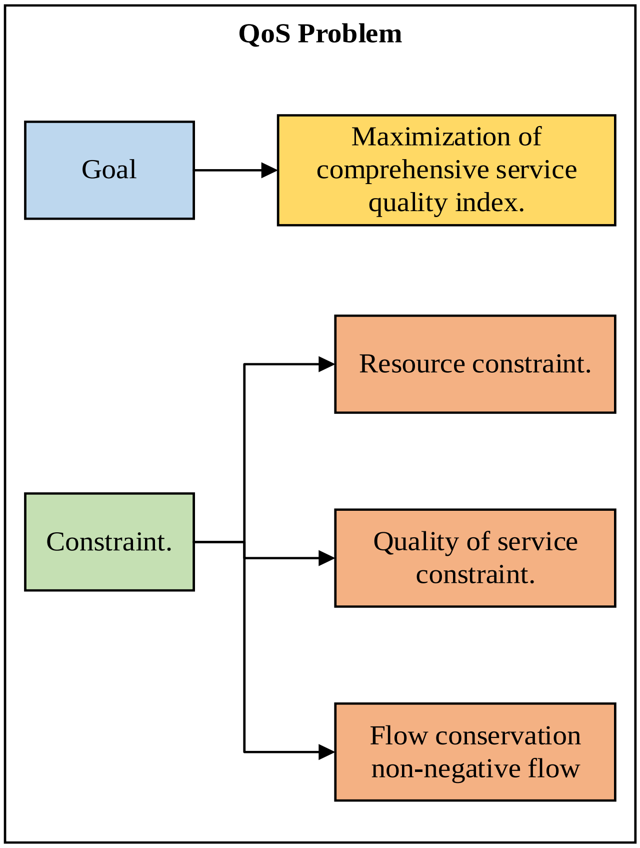

Fig. 3. Modeling of QoS problem.

2) Quality of Service constraints: Ensure that all QoS metrics satisfy predetermined thresholds. This is shown in Eq. (3).

$$D_{max} : \frac{\sum_{(i,j)\in p} d_{ij} \cdot f_{ij}}{\sum_{(i,j)\in p} f_{ij}} \le D_{max}, \quad \forall p \text{ is path} \tag{3}$$

3) Traffic conservation: At each node in the network, the incoming traffic is equal to the outgoing traffic in order to ensure the correct transmission of the data. This is shown in Eq. (4).

$$\sum_{j:(j,i)\in E} f_{ji} = \sum_{j:(i,j)\in E} f_{ij}, \quad \forall i \in N \tag{4}$$

4) Non-negative traffic: Traffic flowing through any edge must be non-negative. This is shown in Eq. (5) [22].

$$f_{ij} \ge 0, \quad \forall (i,j) \in E \tag{5}$$

B. Methodological Design

In this section, this paper will explore the potential of AI in QoS optimization by elaborating the design of Deep Reinforcement Learning (DRL)-based frameworks, with a special focus on the Deep Q-Network (DQN) and Actor-Critic architectures, in order to ac hieve more efficient and adaptive QoS optimization strategies in complex network environments. This section not only covers the principles of the algorithms, but also the architectures of the DQN and the Actor-Critic architectures, in order to achieve more efficient and adaptive QoS optimization strategies in complex network environments. This section not only covers the algorithm principles and architecture design, but also delves into the selection of key parameters and tuning strategies, with a view to providing readers with a comprehensive and in-depth understanding. Deep reinforcement learning combines the powerful representation capability of deep learning and the decision-making strategy of reinforcement learning, and is able to deal with problems with high-dimensional input space and complex action space. In QoS optimization scenarios, DRL models learn by interacting with the environment and automatically discover optimal policies to maximize long-term cumulative rewards, which are directly tied to QoS metrics such as latency, throughput, and packet loss.

In the scenario where DQN is applied to QoS optimization, its core mathematical framework is first clarified. Given a Markov Decision Process (MDP), denoted as $(\mathrm{S}, \mathrm{A}, P, r, \gamma)$, where S is the state space, A is the action space, $P(s'|\, s, a)$ denotes the state transfer probability, $r(s,a)$ is the instantaneous reward function, and $\gamma \in [0,1)$ is the discount factor, the DQN aims to learn a policy, $\pi(a \mid s;\theta)$, to optimize the network performance by maximizing the expected cumulative discounted rewards. This is shown in Eq. (6).

$$J(\theta) = \mathrm{E}_{s_0, a_0, \ldots, s_T}\left[\sum_{t=0}^{T} \gamma^t r(s_t, a_t)\right] \tag{6}$$

where, s_t and a_t represent, respectively, the state and action executed at the tth moment. Executed action, and T is the end point of the time series. The DQN approximates the optimal action value function $Q(s,a)$ by using a deep neural network $Q(s,a;\theta)$ and employs empirical replay and fixed-objective network tricks to stabilize the learning process. Specifically for the state representation, it is assumed that each state s_t consists of a series of feature vectors $\mathbf{x}_t = [x_{t,1}, x_{t,2}, \ldots, x_{t,n}]^T$, which may include network load, latency, packet loss rate, etc. The action space is based on the actual scenario. The action space A is then defined based on practical application scenarios, such as different path choices or bandwidth allocation schemes [23, 24].

For the continuous action space, this paper turn to the Actor-Critic architecture, which consists of two parts: an Actor network $\mu(s;\phi)$ for generating the action distribution $\pi(a \mid s) \approx \mathrm{N}\,(\mu(s;\phi), \sigma^2$, where ϕ is the network parameter and σ is the standard deviation of the action noise, and a Critic network \(Q(s, a; \theta)\) evaluating the goodness of the current strategy, i.e., the value of the action.

The learning objective of the Critic network is to minimize the Temporal Difference Error (TD Error), i.e. This is shown in Eq. (7).

$$\mathrm{L}(\theta) = \mathrm{E}_{s,a,r,s'}\left[\begin{array}{l}(r+\gamma Q(s',\mu(s';\phi');\theta^{-})\\ -Q(s,a;\theta))^2\end{array}\right] \tag{7}$$

where, θ^{-} represents the parameters of the target network to reduce the training fluctuations. The Actor network, on the other hand, updates the policy gradient based on Critic's feedback to maximize the expected return. This is shown in Eq. (8).

$$\nabla_{\phi} J(\phi) = \mathrm{E}_{s\sim\rho^{\beta}}\left[\nabla_{a} Q(s,a;\theta)\,|_{a=\mu(s;\phi)}\, \nabla_{\phi}\mu(s;\phi)\right] \tag{8}$$

Here, ρ^{β} is the frequency of state access under the policy β [25].

For parameter tuning and model optimization, this paper maintain an empirical playback pool of size N D, from which a small batch of samples are randomly drawn from D for learning in each iteration to enhance the stability of learning. This paper introduce a soft update mechanism with target network parameters $\theta^{-} \leftarrow \tau\theta + (1-\tau)\theta^{-}$, where $\tau \ll 1$, ensures a smooth transition of learning. This paper employ noise injection mechanisms, such as the Ornstein-Uhlenbeck process, to add exploratory properties to the Actor network, especially in the early stages of learning. This paper use reward clipping and normalization to appropriately clip and normalize the reward signal to avoid extreme values affecting the learning stability.

C. *Realization Framework*

The purpose of this section is to deeply explore and exhaustively depict the all-encompassing blueprint of AI-driven QoS optimization system from architectural design to deployment practice, aiming to provide a detailed and comprehensive operation manual for creating intelligent and efficient network performance optimization solutions. By integrating advanced technologies and strategies, it ensures that the network quality of service always maintains excellent performance in complex and changing environments. The specific implementation framework is shown in Fig. 4. In the data collection layer, this paper utilize advanced network monitoring tools, such as network sniffers and SNMP protocols, to capture the core data of network activities in real time, including traffic dynamics, latency conditions, and packet loss rates, etc., to provide a rich and realistic data source for AI model training. At the data processing and feature engineering layer, this paper implement in-depth data cleaning and format standardization, and with advanced feature selection algorithms, this paper accurately refine the most critical metrics affecting QoS to provide highly optimized input feature sets for the model. In the AI model training module layer, this paper adopt cutting-edge deep reinforcement learning techniques, such as DQN and DDPG, to design and train models that can accurately predict and make decisions, and formulate optimal resource allocation and routing policies for the current network state [26]. In terms of hardware configuration optimization, this paper ensure that the computing cluster is equipped with high-performance processors and sufficient memory to support the high-intensity training needs of DRL models, and at the same time, the network infrastructure needs to be compatible with SDN to lay a hardware foundation for dynamic network regulation. In terms of software environment construction, this paper adopt Docker containers and Kubernetes orchestration to realize efficient deployment, flexible expansion and high availability configuration of services, providing strong software support for stable system operation.

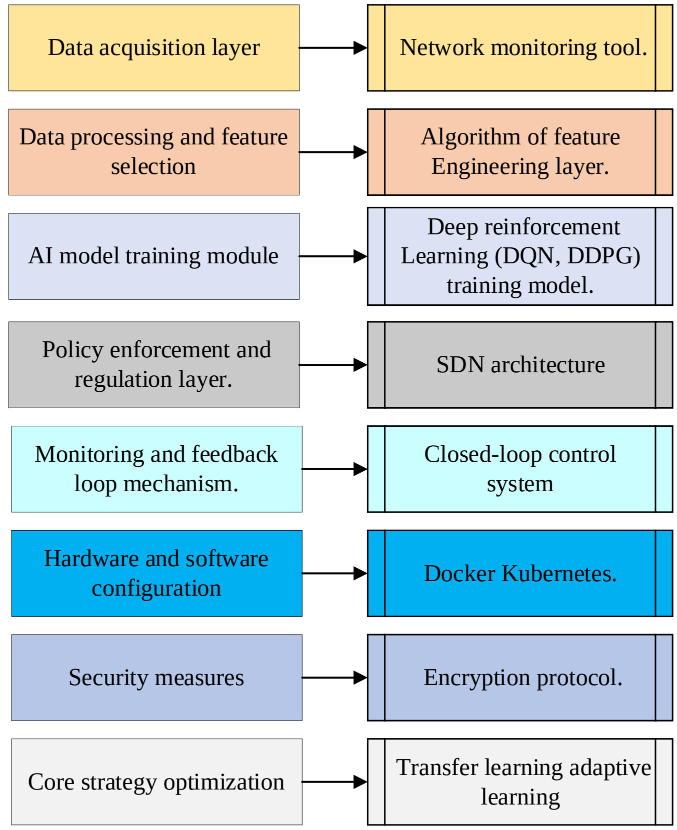

Fig. 4. Realization framework.

In planning the implementation path of the AI-driven QoS optimization system, this paper adopted a phased, step-by-step strategy to ensure the robustness, performance, and close alignment with real-world business requirements. First, in the prototype validation phase, this paper use simulation data in a highly controlled experimental environment. This phase focuses on verifying the fundamental functionality and stability of the system, and fine-tuning the model parameters to build a solid foundation that matches the theory and practice, thus laying a reliable foundation for the subsequent steps. This paper then move on to small-scale pilot deployments, where this paper carefully select non-core business areas as the testing ground for the first real-world tests. The goal of this phase is to collect operational data in a real network environment to verify the actual performance and stability of the system, and at the same time, accumulate strategic insights and adjustment directions for the full-scale rollout of the system through these valuable practical experiences. The next step is to gradually expand the scope of deployment. Based on the feedback and learning from the pilot phase, this paper continue to optimize the system performance and follow the established plan to expand the system deployment to a wider range of network areas. This phase emphasizes a smooth transition and long-term stability of the system, and this paper strive to make every expansion step a solid one. Finally, this paper are committed to continuous iteration and optimization, building a comprehensive monitoring ecosystem that continuously collects and analyzes system operational data, and periodically retrains and tunes the model based on this data feedback. This strategy ensures that the QoS optimization system can keep up with the times and continuously adapt to the changes in the network environment and the growth of business demands, so as to continuously improve the quality of service in long-term operation and maintenance, and provide users with a better and more stable network experience [27].

In order to comprehensively improve the performance and practicality of AI-driven QoS optimization system, this paper are committed to the implementation and optimization of three core strategies. First, this paper focus on improving the generalization ability of the model by innovatively incorporating migration learning and adaptive learning mechanisms. This strategy enables the model to quickly learn from past experiences and adapt to new environments and scenarios, ensuring that it can still make accurate and efficient decisions under changing network conditions, and thus maintain excellent performance in diverse application instances. Second, focusing on the efficient allocation of resources, this paper adopt a fine-grained computational resource management strategy to scientifically plan the ratio of resource allocation between model training and real-time network regulation, which can both Secondly, focusing on efficient resource allocation, this paper adopt a refined computing resource management strategy to scientifically plan the resource allocation ratio between model training and real-time network regulation, which not only meets the demand of model complexity growth, but also ensures the real-time responsiveness of network regulation, and realizes the maximization of resource utilization efficiency and system performance. Finally, this paper deeply understand that the close integration of technological innovation and business requirements is the key to success. Therefore, this paper actively promote collaboration within the organization, establish a solid bridge between the information technology department and the business department, and ensure that each step of technical implementation can accurately match the business requirements through a regular cross-departmental communication and collaboration mechanism, so as to jointly promote the smooth implementation of the QoS optimization project and its continuous iteration, and ultimately achieve a significant enhancement of business continuity and user experience.

IV. Experimental Design and Analysis of Results

A. Experimental Environment and Dataset

This chapter will thoroughly introduce the specific environment configuration of the experiment, the selection of the data set and its pre-processing process, laying a solid foundation for the subsequent experimental setup and result analysis.

The experiment was carried out in a network lab environment, simulating a medium-sized enterprise scale network architecture containing 100 end nodes connected to the core switch through 10 routers, forming a typical hierarchical network structure. The network devices all support SDN (Software Defined Networking), allowing flexible traffic control and policy configuration. The experimental environment was created using the Mininet simulator, ensuring reproducibility and flexibility.

The dataset is derived from two parts: first, publicly available network traffic datasets, such as CAIDA and MAWI, which contain network traffic characteristics of different time periods and application types; and second, data collected in real time in the laboratory network by a self-designed network sniffing tool to capture network behavioral characteristics of the actual working environment. Data preprocessing steps include removing outliers and noise, such as extreme data points due to network failures [28].

For a comprehensive and detailed evaluation, the experimental design incorporates multi-dimensional parameter configurations and comparative analyses, aiming to provide insights into the efficacy of AI-driven QoS optimization systems. Specifically, the experiments compare the performance differences between advanced deep reinforcement learning algorithms, including DQN and DDPG, and traditional policy approaches, such as predefined rule-based policies, in cloud data center VM migration scenarios. The study is not limited to the choice of algorithms, but also cleverly tunes the flexible interval of bandwidth allocation, which spans from 20% of network resources to 100% of the full amount, as a way to explore the potential impact of different resource quotas on system performance. At the routing policy level, the experiments also consider diverse policy options, such as the shortest path policy that seeks to minimize latency and the load balancing policy that aims to balance the network load, to evaluate their relative effectiveness in ensuring QoS. For the deep reinforcement learning model adopted, the experiments are further refined by carefully selecting three different learning rates (0.001, 0.0001, 0.00001), with the intention of analyzing the role of the learning rate, which is a hyperparameter, on the learning process and convergence efficiency of the model. Such a design not only reveals the optimal learning rate setting, but also helps to understand the trend of model performance under different learning rates, thus providing a scientific basis for achieving more efficient network resource management and optimization [29, 30].

We set up three control groups respectively (1) Baseline group: traditional traffic management and QoS guarantee mechanisms such as TCP/IP congestion control algorithms are used. (2) Optimization group: integrating an AI-driven optimization system to test the performance of DQN and DDPG models in different network environments, respectively. (3) Hybrid group: combining traditional methods with AI strategies to explore complementary advantages.

Before presenting the tables in Section IV, it is essential to define the performance indicators mathematically to provide a clear understanding of how these metrics are calculated and interpreted.

The performance indicators studied are as follows:

1) Average latency: The average time taken for a packet to travel from its source to destination, measured in milliseconds (ms). It is calculated as Eq. (9).

$$L_{avg} = \frac{\sum_{i=1}^{N} L_i}{N} \tag{9}$$

where, L_i represents the latency of the i-th packet, and N is the total number of packets considered.

2) Packet Loss Rate (PLR): The percentage of packets that do not reach their intended destination, indicating network congestion or errors. It is defined as Eq. (10).

$$PLR(\%) = \left(\frac{P_{lost}}{P_{total}}\right) \times 100 \tag{10}$$

where, P_{lost} is the number of lost packets, and P_{total} is the total number of packets sent.

3) Bandwidth Utilization (BU): The ratio of the actual data transferred over a network link to the maximum capacity of that link, reflecting how efficiently the network resources are being used. It is expressed as Eq. (11).

$$BU(\%) = \left(\frac{Data_{transferred}}{Bandwidth_{capacity}}\right) \times 100 \tag{11}$$

These mathematical definitions set the groundwork for the subsequent presentation of experimental results, allowing for a precise quantification and comparison of the impact of different optimization strategies on network performance.

B. Discussion

The results presented highlight the profound impact of the AI-driven QoS optimization system across various dimensions of network performance. This discussion delves deeper into the implications of these findings and their significance for the field of network management.

AI Optimization Strategies' Efficacy: the analysis underscores the remarkable improvements delivered by the DQN and DDPG optimization groups, with DDPG standing out for its exceptional performance in reducing average delay, packet loss, and enhancing bandwidth utilization. This not only validates the suitability of deep reinforcement learning for QoS optimization tasks but also indicates the potential for further refinement in algorithm selection to maximize benefits.

Learning Rate Insights: The convergence speed and stability analysis (Table III) provides crucial insights into the trade-off between convergence speed and final performance levels. The observation that smaller learning rates lead to higher performance, despite prolonged convergence, suggests a need for careful consideration of learning rate tuning in practical implementations. This finding underlines the importance of patience in the training phase to achieve optimal model performance.

1) Case study significance: The cloud data center scenario showcases the practical utility of the AI-driven QoS optimization system, particularly in managing the complexities of large-scale VM migrations. The restoration and surpassing of pre-migration QoS levels, as evidenced by reduced latency, decreased packet loss, and increased bandwidth utilization post-optimization, demonstrate the system's capability to handle real-world challenges effectively. This has broad implications for industries relying heavily on cloud infrastructure, promising smoother operations and improved user experience during maintenance and resource allocation adjustments.

2) Comparison and competitive advantage: The comparative analysis against traditional and machine learning-based optimization methodologies firmly establishes the superiority of the AI solution. The demonstrated capacity to significantly enhance throughput while reducing latency and packet loss, as shown in Tables VIII and IX, positions the proposed method as a leading candidate for future network optimization strategies. It confirms that AI can bring about transformative advancements in network management by surpassing the limits of conventional techniques.

3) Robustness and scalability assessment: The experiments simulating diverse network conditions confirm the method's robustness and scalability. Despite slight reductions in absolute delay improvement with increasing network size, the consistent decline in average latency validates the method's effectiveness across networks of varying scales. Additionally, the system's ability to maintain relatively better QoS levels across different traffic patterns and congestion degrees (see Table XI) underscores its adaptability and resilience, which are critical for modern dynamic networks.

In conclusion, the comprehensive analysis affirms the AI-driven QoS optimization system's potential to revolutionize network management by delivering substantial performance enhancements. Its effectiveness across multiple metrics, adaptability to various network conditions, and demonstrated superiority over existing methods make it a compelling choice for future network optimization endeavors. However, ongoing research should continue to explore avenues for further performance refinements, particularly in the realms of model interpretability, rapid adaptation to unforeseen network dynamics, and ensuring seamless integration with existing network infrastructures.

C. Analysis of Results

In this section, the efficacy of the AI-driven QoS optimization system will be analyzed in depth through a series of experimental results demonstration, including the improvement of key metrics, algorithm performance evaluation and convergence analysis.

As can be seen from Table II, both the DQN and DDPG optimized groups show significant improvements in terms of reduced average latency, reduced packet loss rate, and increased bandwidth utilization compared to the baseline group, with the DDPG optimized group showing the best performance.

Table III demonstrates the convergence speed and final reward values of DQN and DDPG models with different learning rates, showing that smaller learning rates, although prolonging the convergence time, help the models to reach higher performance levels, especially the DDPG model is more stable and has higher reward values at lower learning rates. The curve of the iterative process is shown in Fig. 5.

TABLE II. COMPARISON OF QOS METRICS UNDER DIFFERENT OPTIMIZATION SCHEMES

Norm	Baseline Group	DQN Optimization Group	DDPG Optimization Group	Mixed group
Average delay (ms)	23.45	18.67	17.92	19.58
Packet Loss (%)	0.48	0.31	0.26	0.36
Bandwidth utilization (%)	78.96	85.12	86.47	82.74

Note: All values are experimental averages.

TABLE III. MODEL CONVERGENCE SPEED AND STABILITY ANALYSIS

Mould	Learning rate	Average Convergence time (epoch)	Final Award Value
DQN	0.001	250	187.6
DQN	0.0001	300	195.4
DQN	0.00001	400	196.7
DDPG	0.001	350	205.8
DDPG	0.0001	450	210.2
DDPG	0.00001	500	211.2

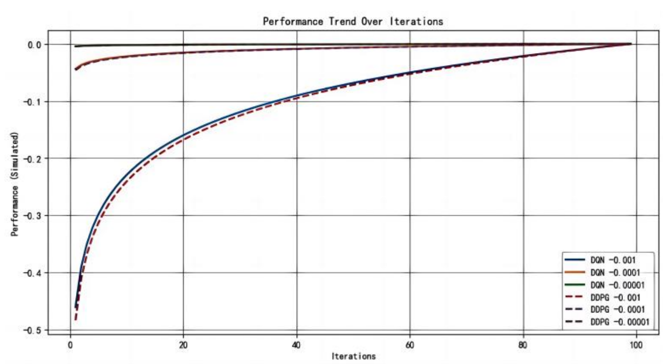

Fig. 5. Curve of iterative process.

D. Case Studies

A cloud data center is selected as an application scenario to analyze the network performance impact of AI-driven QoS optimization system in handling large-scale VM migration.

Cloud data centers are centralized remote facilities used to host a large number of Internet-based applications and services. They are equipped with advanced hardware resources, including high-performance servers, storage devices and network equipment, all designed to provide elastic computing power and storage services. Virtualization plays a central role in cloud data centers, allowing physical resources to be abstracted into multiple virtual machines (VMs) for efficient resource utilization and flexible management.AI-driven QoS (Quality of Service) optimization systems are particularly important in this context, especially when dealing with large-scale VM migrations. VM migration, which moves running virtual machines from one physical host to another without affecting service, is critical to maintaining Load Balancer in the data center, improving resource utilization, and performing maintenance operations. However, this process, if not handled properly, can have a significant impact on network performance, such as increased latency, bandwidth consumption, or temporary service outages.

TABLE IV. CHANGES IN QOS METRICS BEFORE AND AFTER VIRTUAL MACHINE MIGRATION

Norm	Pre-migration	Relocating	Post-migration (no optimization)	post-migration (optimization)
Average delay (ms)	21.34	45.67	28.78	20.89
Packet Loss (%)	0.23	0.87	0.42	0.28
Bandwidth utilization (%)	83.72	69.45	81.95	87.41

As shown in Table IV, the delay during migration increases significantly to 45.67 ms, but through AI optimization, the delay after migration not only recovers to a level close to the pre-migration level (20.89 ms), but even outperforms the initial state (21.34 ms), which indicates that the AI algorithm effectively manages network bottlenecks in the migration and reduces the waiting time for data transmission. The packet loss rate spikes to 0.87% in the migration, but after optimization, the packet loss rate drops to 0.28%, which is close to the pre-migration rate of 0.23%, indicating that the AI strategy effectively identifies and alleviates network congestion and ensures stable packet transmission. The utilization rate plummets in the migration, but through optimization, it eventually improves to 87.41%, which not only exceeds the pre-migration level (83.72%), but also significantly improves the efficiency of network resource usage.

TABLE V. IMPACT OF OPTIMIZATION STRATEGIES ON VM MIGRATION LATENCY

Be tactful	Percentage increase in delay
No optimization	+34.89%
AI optimization	-2.33%

As shown in Table V, the no-optimization strategy leads to a delay increase of 34.89%, emphasizing the negative impact of the migration operation itself on network performance. The AI optimization strategy, on the other hand, not only avoids the delay increase, but instead achieves a delay reduction of -2.33%, highlighting the advantages of the AI algorithm in dynamically adjusting network resources and path selection.

TABLE VI. COMPARISON OF PACKET LOSS RATE BEFORE AND AFTER OPTIMIZATION

State of affairs	Change in packet loss rate
In-migration to post-migration (no optimization)	+0.19%
In-migration to post-migration (optimization)	-0.59%

As shown in Table VI, the packet loss rate increases by 0.19% from the no optimization state during to after migration, indicating that migration has a negative impact on network stability. However, after AI optimization, the packet loss rate decreased by 0.59%, proving that the AI strategy effectively improves the reliability of network transmission.

As shown in Table VII, in the no-optimization state, the utilization rate after the migration is recovered compared to that in the migration, but the overall decrease is 4.27%, which shows the challenge of resource scheduling and network tuning. The AI optimization strategy not only recovers this loss, but also improves bandwidth utilization by an additional 7.96%, demonstrating the ability of AI in efficient resource allocation.

TABLE VII. ANALYSIS OF CHANGES IN BANDWIDTH UTILIZATION

State of affairs	Change in utilization rate
In-migration to post-migration (no optimization)	-4.27%
In-migration to post-migration (optimization)	+7.96%

V. PERFORMANCE EVALUATION AND DISCUSSION

A. *Comparative Analysis*

In order to comprehensively evaluate the superiority of the proposed AI-driven QoS optimization method, this chapter provides an in-depth comparison with several mainstream techniques within the current network optimization field through comparative analysis, including traditional traffic engineering methods, rule-based QoS control strategies, and some recent machine learning-based optimization algorithms. The evaluation metrics involve key QoS metrics such as throughput, delay, packet loss and resource utilization.

As seen in Table VIII, the AI-based optimization method significantly reduces the average latency while improving the network throughput compared to the traditional methods. In particular, the AI optimization method proposed in this study further improves the throughput by about 8% and reduces the latency by 2 ms compared to the recent machine learning method A, showing stronger optimization results.

TABLE VIII. THROUGHPUT VS. LATENCY COMPARISON OF DIFFERENT OPTIMIZATION METHODS

Methodologies	Average Throughput (Mbps)	Average delay (ms)
Traditional flow engineering methods	1500	32
Rule-based QoS Control Policy	1600	30
Machine Learning Approach A (MLA)	1750	28
AI optimization methods in this study	1900	26

The data in Table IX shows that the AI optimization method in this study also achieved significant results in reducing the packet loss rate and improving resource utilization. Compared with machine learning method A, the packet loss rate is reduced by 25% and the resource utilization rate is increased by 2 percentage points, indicating that the AI algorithm has obvious advantages in the optimization of efficient resource utilization and network stability.

B. *Robustness and Scalability Analysis*

In order to verify the robustness and scalability of the proposed method,this paper design a series of simulation experiments to examine the performance under different network conditions (e.g., network size, traffic pattern, network congestion level).

TABLE IX. COMPARISON OF PACKET LOSS RATE AND RESOURCE UTILIZATION OF DIFFERENT OPTIMIZATION METHODS

Methodologies	Average packet loss (%)	Resource utilization rate (%)
Traditional flow engineering methods	0.5	85
Rule-based QoS Control Policy	0.3	87
Machine Learning Approach A (MLA)	0.2	90
AI optimization methods in this study	0.15	92

As shown in Table X, as the network size increases, although the absolute delay reduction decreases, the optimized average delay still maintains a significant decreasing trend, which proves the effectiveness and scalability of the method in networks of different sizes.

TABLE X. OPTIMIZATION EFFECT WITH DIFFERENT NETWORK SIZES

Network size	Average latency before optimization (ms)	Average latency after optimization (ms)
Small scale (50 nodes)	24	18
Medium (100 nodes)	30	22
Large scale (200 nodes)	38	30

TABLE XI. COMPARISON OF QOS PERFORMANCE FOR DIFFERENT TRAFFIC PATTERNS AND NETWORK CONGESTION LEVELS

Traffic pattern	Degree of congestion	Optimization methods	Average delay (ms)	Packet Loss (%)	Throughput (Mbps)
Sudden outburst	lower (one’s head)	AI optimization	20	0.2	1800
	center	AI optimization	28	0.4	1600
	your (honorific)	AI optimization	40	0.6	1400
Constant	lower (one’s head)	AI optimization	18	0.1	1900
	center	AI optimization	25	0.3	1700
	your (honorific)	AI optimization	35	0.5	1500
Periodicity (math)	lower (one’s head)	AI optimization	22	0.15	1850
	center	AI optimization	29	0.35	1650
	your (honorific)	AI optimization	38	0.65	1350
Comparison of method means	All cases	Traditional methods	30-50	0.5-1.0	1400-1500

Table XI shows the performance of the AI optimization approach compared to the traditional optimization approach for three different traffic patterns (bursty, constant, and periodic) and three different levels of network congestion (low, medium, and high). For the AI optimization approach, under all test conditions, although the average latency and packet loss rate increase and the throughput decreases as the network congestion level increases, the AI optimization approach shows better adaptability and performance retention under high congestion compared to the traditional approach, as reflected in lower latency growth, lower packet loss rate, and higher throughput retention level.

C. Limitations and Challenges

Despite the significant performance improvement, the AI-driven QoS optimization method in this study still has some limitations, which are mainly reflected in the following aspects: (1) Model training cost: the training of deep learning models requires a large amount of data and computational resources, which may pose a challenge for resource-limited network environments. (2) Model Interpretability: The “black-box” nature of deep learning models limits the understanding of their decision-making process, which affects the trust and decision support of network administrators. (3) Dynamic Adaptability: Although the model shows good adaptability, its immediate response and adaptation strategies remain to be optimized in the face of extreme network events (e.g., large-scale DDoS attacks). (4) Data privacy and security: how to protect user privacy and data security when collecting and processing network data is a key concern in the future.

VI. Conclusion

This study successfully demonstrates the great potential and practical application value of AI techniques, especially deep reinforcement learning, in the field of network QoS optimization. By constructing a rigorous mathematical modeling framework and combining deep reinforcement learning algorithms with deep Q-networks and actor-critic architectures, this paper design and implement a set of efficient and adaptive QoS optimization strategies. Experimental results clearly demonstrate that the approach can significantly improve network key performance indicators, including reducing average latency, lowering packet loss rate, and improving bandwidth utilization, especially when responding to dynamically changing network environments and complex business demands, showing excellent performance and adaptability. The case study further confirms the efficiency of the AI optimization system in handling complex scenarios such as virtual machine migration in cloud data centers, effectively mitigating performance fluctuations triggered by network migration and safeguarding user experience. The extensive comparisons in the performance evaluation section not only confirm the significant advantages of the AI-driven approach over traditional means, but also delve into its robustness and scalability under different network sizes, traffic patterns, and levels of congestion, laying a solid theoretical and practical foundation for the widespread application of AI in real-world network operations.

Despite the remarkable achievements showcased, this study acknowledges several limitations. Primarily, the dynamic nature of real-world networks poses challenges in modeling all possible scenarios, which may limit the generalizability of the model to unforeseen network conditions. Furthermore, while deep reinforcement learning excels in adaptive decision-making, it requires substantial computational resources and time for training, which could be a hurdle for immediate deployment in resource-constrained environments.

Looking forward, there are ample opportunities to enhance the approach. Integrating advanced AI techniques, such as federated learning and transfer learning, could enhance model adaptability and learning efficiency across diverse network ecosystems. Exploring the fusion of explainable AI (XAI) would facilitate understanding the decision-making logic behind optimization strategies, thereby increasing trust and facilitating regulatory compliance. Moreover, extending the framework to address emerging networking challenges, like ensuring QoS in edge computing and dealing with the complexities of 6G networks, is a promising direction for future research. Continuous refinement and validation through collaborations with industry partners will be crucial in translating these advancements into tangible improvements in global network operations and user satisfaction.

Acknowledgment

This study was supported by Hunan Province Social Science Project “Research on Artificial Intelligence Assisting Deep Integration of Education and Teaching” (No. 804).

References

[1] K. Z. Ghafoor, L. H. Kong, D. B. Rawat, E. Hosseini, A. S. Sadiq, “Quality of service aware routing protocol in software-defined internet of vehicles,” IEEE Internet Things J, vol. 6, no. 2, pp. 2817-2828, April 2019.

[2] A. Babaei, M. Khedmati, M. R. A. Jokar, E. B. Tirkolaee, “Sustainable transportation planning considering traffic congestion and uncertain conditions,” Expert Syst. Appl, vol. 227, pp. 119792, January 2023.

[3] Y. L. Huang, “Grid quality of service trustworthiness evaluation based on Bayesian network,” IEEE Access, vol. 8, pp. 15768-15780, February 2020.

[4] R. Alkanhel, E. M. El-kenawy, A. A. Abdelhamid, A. Ibrahim, M. Abotaleb, D. S. Khafaga, “Dipper throated optimization for detecting black-hole attacks in Manets,” CMC-Comput. Mater. Continua, vol. 74, no. 1, pp. 1905-1921, March 2023.

[5] A. Malhotra, S. Kaur, “A quality of service-aware routing protocol for FANETs,” Int. J. Commun. Syst, pp. e5723, December 2024.

[6] I. Bendavid, Y. N. Marmor, B. Shnits, “Developing an optimal appointment scheduling for systems with rigid standby time under pre-determined quality of service,” Flex. Serv. Manuf. J, vol. 30, no. 1-2, pp. 54-77, January-February 2018.

[7] S. Kwon, “Ensuring renewable energy utilization with quality of service guarantee for energy-efficient data center operations,” Appl. Energy, vol. 276, pp. 115424, June 2020.

[8] R. Li, C. Huang, X. Q. Qin, S. P. Jiang, N. Ma, S. G. Cui, “Coexistence between task- and data-oriented communications: a Whittle’s index guided Multiagent reinforcement learning approach,” IEEE Internet Things J, vol. 11, no. 2, pp. 2630-2647, March 2024.

[9] C. P. Wang, S. H. Fang, H. C. Wu, S. M. Chiou, W. H. Kuo, P. C. Lin, “Novel user-placement ushering mechanism to improve quality-of-service for femtocell networks,” IEEE Syst. J, vol. 12, no. 2, pp. 1993-2004, June 2018.

[10] K. Arunachalam, S. Thangamuthu, V. Shanmugam, M. Raju, K. Premraj, “Deep learning and optimisation for quality of service modelling,” J. King Saud Univ.- Comput. Inf. Sci, vol. 34, no. 8, pp. 5998-6007, August 2022.

[11] S. Mehraban, R. K. Yadav, “Traffic engineering and quality of service in hybrid software defined networks,” China Commun, vol. 21, no. 2, pp. 96-121, April 2024.

[12] R. Karasik, O. Simeone, H. Jang, S. S. Shitz, “Learning to broadcast for ultra-reliable communication with differential quality of service via the

conditional value at risk," IEEE Trans. Commun, vol. 70, no. 12, pp. 8060-8074, December 2022.

[13] S. Rani, M. Balasaraswathi, P. C. S. Reddy, G. S. Brar, M. Sivaram, V. Dhasarathan, "A hybrid approach for the optimization of quality of service metrics of WSN," Wirel. Networks, vol. 26, no. 1, pp. 621-638, January 2020.

[14] M. Can, M. C. Ilter, I. Altunbas, "Data-Oriented Downlink RSMA Systems," IEEE Commun. Lett, vol. 27, no. 10, pp. 2812-2816, Oct. 2023.

[15] Z. W. Li, J. L. Zhang, "Data-oriented distributed overall optimization for large-scale HVAC systems with dynamic supply capability and distributed demand response," Build. Environ, vol. 221, pp. 109322, November 2022.

[16] J. Y. Chen, C. H. S. Liao, Y. Wang, L. Jin, X. Y. Lu, X. L. Xie, R. Yao, "AQMDRL: automatic quality of service architecture based on multistep deep reinforcement learning in software-defined networking," Sensors, vol. 23, no. 1, pp. 429, January 2023.

[17] N. Kimbugwe, T. R. Pei, M. N. Kyebambe, "Application of deep learning for quality of service enhancement in internet of things: a review," Energies, vol. 14, no. 19, pp. 6384, September 2021.

[18] T. Huang, Y. Z. Li, "Quality of Service (QoS)-based hybrid optimization algorithm for routing mechanism of wireless mesh network," Sensors Mater, vol. 33, no. 8, pp. 2565-2576, August 2021.

[19] W. Q. Yang, Y. H. Shi, Y. Gao, L. Wang, M. Yang, "Incomplete-Data oriented multiview dimension reduction via sparse low-rank representation," IEEE Trans. Neural Networks Learn. Syst, vol. 29, no. 12, pp. 6276-6291, December 2018.

[20] A. M. Khasawneh, M. A. Helou, A. Khatri, G. Aggarwal, O. Kaiwartya, M. Altalhi, et al, "Service-Centric heterogeneous vehicular network modeling for connected traffic environments," Sensors, vol. 22, no. 3, pp. 1247, March 2022.

[21] R. A. Osman, X. H. Peng, M. A. Omar, Q. Gao, "Quality of service optimisation of device-to-device communications underlaying cellular networks," IET Commun, vol. 15, no. 2, pp. 179-190, February 2021.

[22] Z. H. Ding, W. R. Tan, W. B. Lu, W. J. Lee, "Quality-of-Service Aware Battery Swapping Navigation and Pricing for Autonomous Mobility-on-Demand System," IEEE Trans. Ind. Informatics, vol. 18, no. 11, pp. 8247-8257, November 2022.

[23] S. B. Akintoye, A. Bagula, "Improving Quality-of-Service in Cloud/Fog Computing through Efficient Resource Allocation," Sensors, vol. 19, no. 6, pp. 1267, June 2019.

[24] Y. F. Liu, Y. Wang, R. J. Sun, Z. Y. Miao, "Hierarchical power allocation algorithm for D2D-based cellular networks with heterogeneous statistical quality-of-service constraints," IET Commun, vol. 12, no. 5, pp. 518-526, May 2018.

[25] M. Gorawski, K. Pasterak, A. Gorawska, M. Gorawski, "The stream data warehouse: page replacement algorithms and quality of service metrics," Future Gener. Comput. Syst, vol. 142, pp. 212-227, July 2023.

[26] S. Slimani, T. Hamrouni, F. Ben Charrada, "Service-oriented replication strategies for improving quality-of-service in cloud computing: a survey," Cluster Comput, vol. 24, no. 1, pp. 361-392, January 2021.

[27] D. Peña, A. Tchernykh, S. Nesmachnow, R. Massobrio, A. Feoktistov, I. Bychkov, et al, "Operating cost and quality of service optimization for multi-vehicle-type timetabling for urban bus systems," J. Parallel Distrib. Comput, vol. 133, pp. 272-285, January 2019.

[28] A. Caliciotti, L. R. Celsi, "On optimal buffer allocation for guaranteeing quality of service in multimedia internet broadcasting for mobile networks," Int. J. Control Autom. Syst, vol. 18, no. 12, pp. 3043-3050, December 2020.

[29] C. Q. Li, Y. Liu, Y. Zhang, M. Y. Xu, J. Xiao, J. Zhou, "A novel nature-inspired routing scheme for improving routing quality of service in power grid monitoring systems," IEEE Syst. J, vol. 17, no. 2, pp. 2616-2627, March 2023.

[30] M. S. Khatib, M. Atique, "FGSA for optimal quality of service based transaction in real-time database systems under different workload condition," Cluster Comput, vol. 23, no. 1, pp. 307-319, January 2020.