IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 February 2024, accepted 1 April 2024, date of publication 8 April 2024, date of current version 19 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3386176

RESEARCH ARTICLE

A Deep Reinforcement Learning-Based Optimal Transmission Control Method for Streaming Videos

YAWEN YANG[1,2] AND YUXUAN XIAO[3]
[1]College of Humanities and Arts, Hunan International Economics University, Changsha 410205, China
[2]School of Journalism and Communication, Hunan University, Changsha 410082, China
[3]College of Design Arts, Changsha University of Science and Technology, Changsha 410114, China

Corresponding author: Yuxuan Xiao (xyy_3535@163.com)

This work was supported in part by the Project of Hunan Provincial Social Science Achievement Evaluation Committee under Grant XSP2023WXC035; in part by the Department of Education Scientific Research Project under Grant 22C0654; in part by Hunan Provincial Social Science Planning Project under Grant 22wtc09; and in part by the Teaching Reform Research Project of Hunan Province, Exploration and Practice Research on the Training Mode of Network and New Media Professional Innovative Talents in the "Era of Digital Intelligence" under Grant HNJG-20231260.

ABSTRACT Time delay and image quality degradation are main challenges faced by streaming video transmission. How to make adaptive planning for transmission schemes according to dynamic change of transmission environment, always remains a technical concern. As a consequence, this paper proposes a deep reinforcement learning-based optimal transmission control method for streaming videos. Firstly, edge buffer task allocation is combined with quality of experience (QoE)-oriented deep reinforcement learning algorithm, in order to develop a resource allocation method for streaming videos. Secondly, an actively coordinated streaming data streaming transmission mechanism is established to construct a specific optimal transmission control method that satisfies environment requirement. Finally, a set of experiments are conducted to verify effectiveness and performance on public video transmission datasets. And the proposal is compared with several traditional transmission methods. The experiments show that the proposal in this work can effectively reduce delay and startup time and improve the QoE. This shows that the proposal is able to bring better stability and transmission quality.

INDEX TERMS Deep reinforcement learning, streaming videos, optimal transmission control, quality of experience.

I. INTRODUCTION

In today's digital age, streaming video has become an indispensable part of people's daily lives, and its applications in fields such as education, entertainment, and communication are becoming increasingly widespread [1]. However, with the popularization of streaming video services and the increasing demand for high-definition and low latency from users, how to effectively transmit high-quality video content has become a highly concerned issue [2]. In this context, the optimization control method for streaming video transmission based on deep reinforcement learning is proposed to address this challenge.

Traditional video transmission systems often have some problems, such as video stream lag, degraded image quality, or video loss in network congestion [3], [4], which directly affect the viewing experience of users. As a technology that combines deep learning and reinforcement learning [5], deep reinforcement learning has the potential to automatically learn and optimize system control strategies [6], providing new ideas and methods for solving optimization control problems in video transmission processes [7].

The purpose of this study is to explore an optimization control method for streaming video transmission based on deep

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

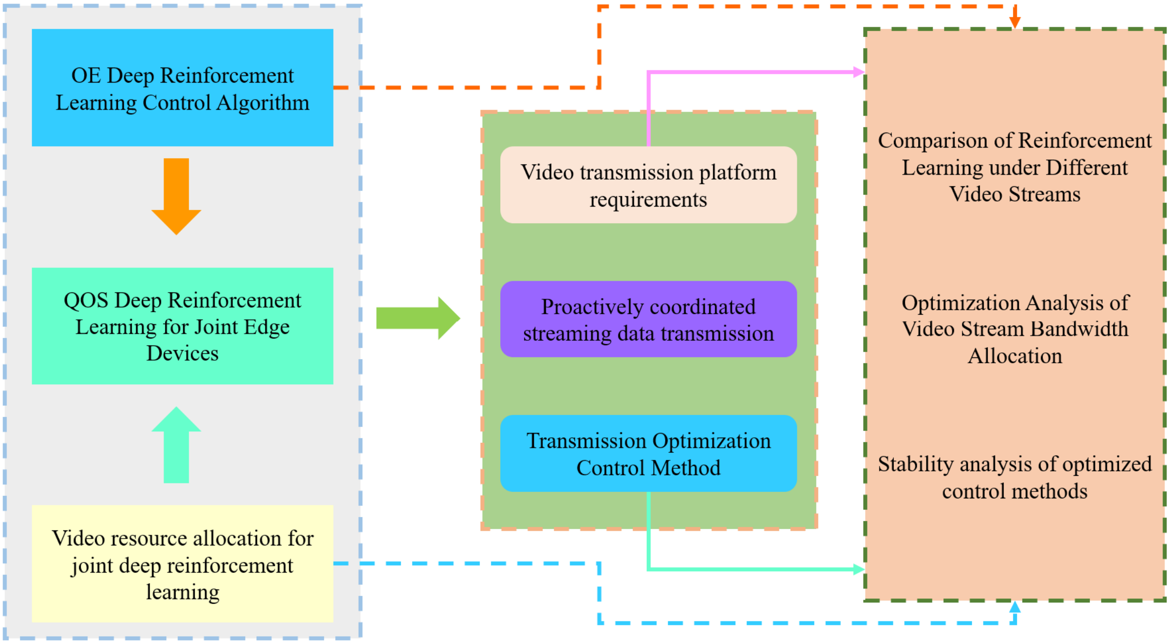


FIGURE 1. Research framework of this paper.

reinforcement learning. We will combine deep learning technology and reinforcement learning algorithms to design an intelligent decision-making system for real-time and dynamic control of various parameters and strategies during video transmission. By optimizing the encoding, transmission, and decoding process of video frames, we aim to improve the clarity, stability, and response speed of videos, and maximize the satisfaction of users' needs for streaming media videos. In this study, we will build an experimental environment based on existing open source platforms and datasets, and propose an effective optimization control method based on the characteristics and related technologies of streaming video transmission. The research framework is shown in Figure 1. We will validate the effectiveness and advantages of the proposed method through extensive experiments and comparative analysis, and compare it with traditional video transmission methods to evaluate its performance in various scenarios.

The main contributions of the study are as follows:

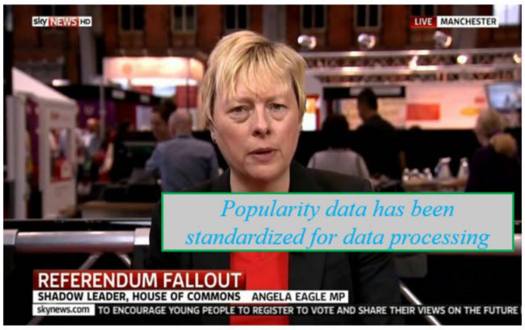
1) The difference between this study and previous studies is that we used QOE deep reinforcement learning method to control edge video tasks. Unlike previous research, we use reinforcement learning to process multiple edge tasks to determine whether they can meet video transmission requirements.

2) The study also considered the quality perception and latency issues of video transmission. We allocate video transmission tasks in more detail in edge tasks through more efficient algorithms, optimize control methods, and improve video quality while ensuring low latency video playback.

Structure of the article is organized as follows. Section I analyzes the main problems and challenges faced in streaming video transmission. The second section summarizes some of the relevant research of predecessors. Section III combines the caching task allocation of edge devices and the control algorithm of QOE deep reinforcement learning to complete the adjustment of video resources. Section IV proposes an actively coordinated streaming data transmission method based on the requirements of the video output platform, and establishes a complete M-QOE video transmission optimization control method. Section V collected public data from the network and completed the analysis of reinforcement learning comparison, bandwidth allocation, and stability. Section VI is a summary of the entire text and reflections on subsequent research.

II. LITERATURE REVIEW

This study is a research topic involving deep learning and reinforcement learning techniques in the field of streaming video transmission optimization. In the research of relevant scholars, many are committed to using deep learning algorithms to optimize various aspects of video transmission. They usually explore how to use deep learning technology to improve the efficiency and quality of video transmission from aspects such as video encoding, transmission protocols, and network topology. Among them, some scholars are concerned about how to use deep learning algorithms to optimize video encoding and decoding, in order to reduce the bandwidth consumption of video streams while maintaining high visual quality. They optimize video encoding by training neural networks to adapt to different network conditions and terminal devices, thereby improving the efficiency of video transmission. In addition, some scholars are committed to using reinforcement learning algorithms to optimize control



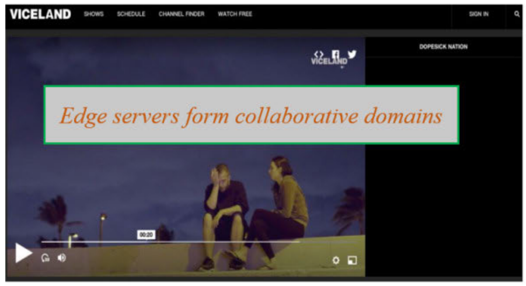


FIGURE 2. Streaming video transmission optimization control.

strategies for video transmission. They establish a model for video transmission systems and design corresponding reward functions, allowing reinforcement learning algorithms to dynamically adjust transmission parameters based on real-time network status and video quality feedback, in order to achieve optimal transmission control in different network environments.

These scholars and organizations are representatives of those who have made certain achievements in the field of optimizing control methods for streaming video transmission based on deep reinforcement learning. Of course, many other scholars and organizations are also conducting in-depth research in this field. In summary, the optimization control method for streaming video transmission based on deep reinforcement learning has become a research hotspot, attracting the attention and investment of many excellent scholars and organizations at home and abroad.

III. QOS-ORIENTED DEEP REINFORCEMENT LEARNING FOR JOINT EDGE DEVICES

A. EDGE DEVICE CACHE TASK ALLOCATION

Edge device cache task allocation refers to the allocation of video content delivery tasks to different cache devices according to certain strategies in the edge computing environment to optimize the transmission quality and user experience of streaming video [16]. The optimization control method for streaming video transmission based on deep reinforcement learning can effectively solve the problem of edge device caching task allocation [17]. Edge devices typically have limited storage capacity, so it is necessary to allocate video content delivery tasks to various devices in a reasonable manner to maximize storage utilization [18].

Video popularity is an important indicator of live video content, with a focus on considering the impact on user viewing before the start of the live video. However, in addition to video popularity, several other important factors are also considered comprehensively, and there are significant differences in the order of magnitude between popularity indicators and other factors. If popularity data is directly used, the importance of popularity will be overly emphasized. To balance this difference and ensure the reliability of the results, This article has standardized the popularity data, and the standardization formula is as follows [19]:

$$pop\,(v) = \frac{curpop - dop}{SD} \tag{1}$$

Among them, *pop(v)* is the standardized popularity value of live video v, *curpop* is the original popularity value, *dop* is the mean popularity value, and *SD* is the standard deviation of live video popularity.

Each server forms a collaborative domain with 10 regular edge servers, and to simplify the experiment, the link bandwidth is set to a fixed value. The delay between the edge server and the server is 10ms, and the delay between the edge server and the content delivery networks (CDN) is 200ms. The request response time is as follows:

$$Delay = T_{send} + T_c + T_d \tag{2}$$

Among them, T_{send} is the transmission delay, T_C is the service request delay from the cluster head server, and T_d is the service request delay from the CDN. If a direct request is made to the edge server and the requested resource exists on the edge server, then T_C and T_d values are 0.

B. QOE-ORIENTED DEEP REINFORCEMENT LEARNING

The optimization control method for streaming video transmission based on deep reinforcement learning has the potential to improve the quality of user experience (QoE). In this article, you may use deep reinforcement learning control algorithms to optimize key parameters in the process of streaming video transmission, in order to improve the viewing experience for users. The QOE deep reinforcement learning control algorithm is based on a deep reinforcement learning framework, which selects appropriate actions from a given state through interactive learning with the environment to maximize the predefined reward function [21], [22]. In streaming video transmission, QoE is influenced by many factors, such as bandwidth, latency, packet loss rate, etc. Deep reinforcement learning can optimize the video transmission process by monitoring and adjusting these parameters in real-time, thereby improving the viewing experience for users.

Value based reinforcement learning is used to handle tasks with discrete action space and limited action space.

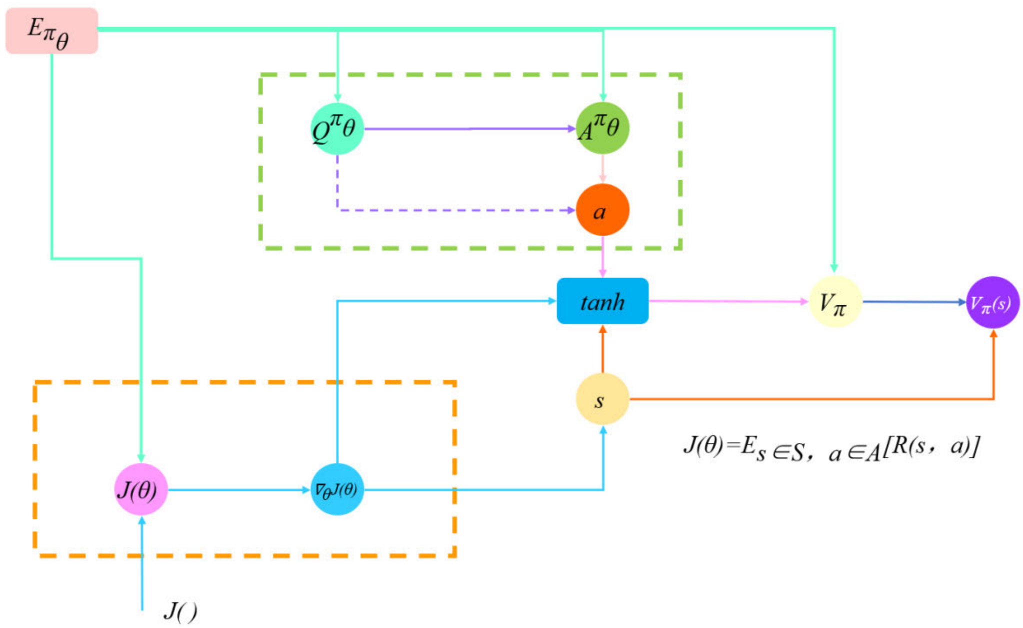


FIGURE 3. Internal structure of QOE deep reinforcement learning.

By learning the value function, the value of the action is obtained, and the action with the highest value is selected for execution. This is an indirect way to select the optimal strategy. Taking the update method of the Q-learning value function as an example [23]:

$$Q'(s,a) = Q(s,a) + \alpha\left[R(s,a) + \gamma \max Q'\left(s',a'\right) - Q(s,a)\right] \quad (3)$$

The execution process of the entire system is as follows: at the initial moment of the t time slot, the user reports their current observed relevant information $s_j(t)$, including channel quality $q_{jj}(t)$, buffer data $T_j(t)$, requested video block number m_j, to the central controller i. The central controller stores $s_j(t)$ in the user's cache queue Q_j, and the resource allocation module pulls the global environment state required for QoS control decisions from the state storage module.

The optimization goal of the strategic network is to maximize the expected cumulative return $J(\theta)$ to obtain strategies that maximize returns [24]:

$$J(\theta) = E_{s\in S, a\in A}\left[R(s,a)\right] \quad (4)$$

Directly targeting policy parameters θ The formula for derivative and strategy gradient update is as follows:

$$\nabla_\theta J(\theta) = E_{\pi_\theta}\left[\nabla_\theta \log \pi_\theta(a/s)\, Q^{\pi_\theta}(s,a)\right] \quad (5)$$

At the same time, Critical updates the network parameters based on the time difference error, making the network's value estimation of the state more accurate and better guiding the update of the Actor network. The state value function of reinforcement learning is written in an iterative form as follows:

$$V_\pi(s) = E_\pi\left[r + \gamma V_\pi\left(s'\right)\right] \quad (6)$$

Due to the fact that traditional value-based reinforcement learning uses the original cumulative expected return Rt as the state action value function Q π (s, a) to guide the degree of policy update, it can lead to a high contrast, which can be understood as the advantage of executing action a relative to the average performance of state s. The formula for the parameters of the policy gradient network can be rewritten as equation 7, and its internal structure is shown in Figure 3.

$$\nabla_\theta J(\theta) = E_{\pi_\theta}\left[\nabla_\theta \log \pi_\theta(a/s)\, A^{\pi_\theta}(s,a)\right] \quad (7)$$

C. VIDEO RESOURCE ALLOCATION FOR JOINT DEEP REINFORCEMENT LEARNING

Joint deep reinforcement learning has great potential in video resource allocation [25]. In the video transmission optimization control method based on deep reinforcement learning, video resource allocation is an important link that can determine the transmission quality and user experience of the video stream. Through joint deep reinforcement learning, the system can achieve intelligent decision-making and optimize the allocation process of video resources [26]. Deep reinforcement learning can help systems make intelligent decisions based on factors such as current network conditions, user needs, and video characteristics. In video resource allocation, the system can learn the optimal strategy through deep reinforcement learning to achieve optimal performance during video transmission.

Specifically, the intelligent agent selects the optimal video quality level at each time step and sends its options to the client. These decisions will also be passed back to the server and used to improve the next decision. According to the previous definition of QoS for user viewing, the return r_k

is composed of video quality and quality fluctuations in the spatiotemporal domain, and is affected by prediction accuracy. Set Set_k to represent the set of code blocks that need to be downloaded after the selected action updates the highest level of L_k. For any element C, the return value can be expressed as:

$$r_k = \sum_{C \in Set_k} \Delta G(C) \cdot \left(1 - P_{Set_k} \cdot C\right) \tag{8}$$

All encoding blocks in the set with playback or encoding time ranking first are marked as Set_k^*, and calculated Δ The probability of C being viewed, which $\Delta G(C)$ depends on, is replaced by the predicted viewing probability of the FoV where it is located. The video parameters for streaming media transmission are shown in Table 1.

TABLE 1. Collection and display of streaming video parameters.

Collection and display of streaming video parameters			
Parameter Source	NTHU	UTD	THU
Required quantity	8+4	8+8	5+3
Content classification	3	5	5
Resolving power	2K	2K	1080P

In addition, this joint deep reinforcement learning video resource allocation method also has great flexibility and can adapt to different types of streaming media services and network conditions. For example, with higher bandwidth and more stable network connectivity, this method can automatically improve the video quality level and provide higher quality video streams; In situations where the network is more congested, the video quality level will be automatically reduced to avoid video buffering and lag. The video resource allocation of joint deep reinforcement learning is an efficient, flexible, and personalized video transmission control method that can automatically adjust the video quality level according to real-time changes in network conditions, improve user experience and satisfaction, which will be verified in subsequent experiments.

IV. M-QOE: OPTIMAL TRANSMISSION CONTROL METHOD FOR STREAMING VIDEOS

A. VIDEO TRANSMISSION PLATFORM REQUIREMENTS

The streaming video transmission platform needs to ensure that video content is transmitted to users with high quality and stability to provide a smooth viewing experience. The transmission platform needs to have excellent encoding and decoding capabilities and transmission technology, while utilizing deep reinforcement learning algorithms to optimize transmission control and improve the quality and performance of video transmission. The research will focus on the key requirements of the video transmission platform required in the optimization control method of streaming video transmission based on DDPG deep reinforcement learning. Video transmission needs to meet efficient speed requirements, which require efficient network bandwidth and transmission protocols to ensure smooth video transmission [27]. In addition, reliable service quality and stable connectivity are also required to avoid issues such as interruption or lag in video transmission. Stability analysis can also help us evaluate the performance stability of optimized control methods. The goal of optimizing control methods is to maximize bandwidth utilization and user experience while maintaining video transmission quality. Stability analysis can help us determine the performance stability of optimization control methods under different network conditions, as well as their sensitivity to different optimization objectives. DDPG is a reinforcement learning method based on the Actor-Critic network framework. It is a deterministic strategy approach, in which the learned strategy outputs only one deterministic action under the same state input [28]. In the DDPG algorithm, not only the Actor-Critic method is used, but also a dual network structure is used to distinguish the current network from the target network. The task of the current network is to use greedy algorithms to select corresponding actions for the current input state s α And execute, during the training process of the Critical network, as it is a value based method, it is similar to the loss function of DQN [29]:

$$J(\omega) = \frac{1}{N} \sum_{i=1}^{N} \{y_i - Q[\phi(s), a, \omega]\} \tag{9}$$

The actual meaning of the loss corresponding to the Actor network is that under the same input state, if the policy outputs two different actions α_1 and α_2. Through the Critical network, the corresponding values can be calculated as Q_1 and Q_2, respectively. For a layer of convolutional layers, N samples are first taken from the training samples and input into the neural network to calculate the N feature maps ycov output by the convolutional layer. The weight values in the convolutional kernel of this layer are sequentially set to 0, and the new output feature map y'cov is recalculated. The mean square error of the feature map before and after pruning is calculated:

$$L_w\left(y_{fc}, y'_{fc}\right) = \frac{1}{N} \sum_{i=1}^{N} \left(y_{cov} - y'_{cov}\right)^2 \tag{10}$$

For video content with high real-time requirements, such as live streaming activities, the transmission platform needs to have low latency transmission capabilities to ensure real-time and stability. By optimizing control methods through deep reinforcement learning, transmission delay can be effectively reduced and user experience can be improved.

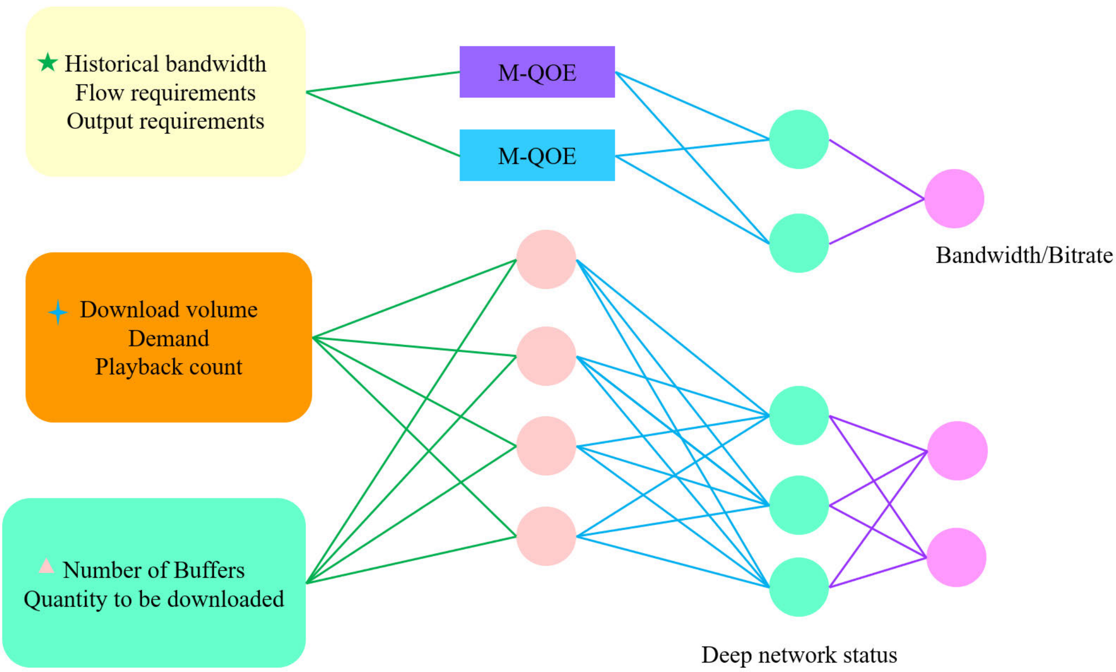


FIGURE 4. M-QOE video transmission optimization control algorithm framework.

B. PROACTIVELY COORDINATED STREAMING DATA TRANSMISSION

Proactively coordinated streaming data transmission refers to the use of deep reinforcement learning techniques and proactive measures taken to optimize data transmission during the streaming video transmission process. In traditional streaming media data transmission, video data is often sent according to fixed transmission rules and strategies, but this static transmission method often cannot adapt to changes in the network environment and differences in user needs. The intelligent agent can learn the optimal transmission optimization strategy to provide better video transmission quality and user experience. This method based on deep reinforcement learning has certain adaptability and generalization ability, and can adapt to dynamic changing environments under different network conditions.

In order to cope with instantaneous wireless network channel fluctuations, the network channel conditions in a short period of time are accurately estimated by periodically sampling the physical layer channel to track the instantaneous network channel fluctuations. Take t as the sampling start time, $t+n.\Delta T$ is the end time of the sample, Δt is the sampling interval, and equal interval sampling is performed on the physical layer transmission blocks of the wireless network. The specific calculation formula is as follows [30]:

$$BW_i\left(t\right) = \frac{1}{\Delta t}\sum_{t(i-1)\cdot\Delta t}^{t+i\cdot\Delta t} TBSize_j \tag{11}$$

$$BW_i\left(t\right) = \frac{1}{n}\sum_{i=1}^{N} BW_i\left(t\right) \tag{12}$$

$$BW_t = (1-\alpha)\cdot BW\left(t\right) + \alpha\cdot BW_{t-1} \tag{13}$$

Among them, $BW_i\left(t\right)$represents the physical layer bandwidth sampling value at the i equal time interval starting from sampling time t, $BW\left(t\right)$ represents the arithmetic mean of the physical layer bandwidth sampling at sampling time t, BW_i represents the estimated user available network bandwidth at time t, and BW_{t-1} is the estimated available network bandwidth from the previous sampling, ΔT is set to 50ms, the number of samples is set to 20, and the attenuation factor a is set to 01, which means the sampling covers the channel conditions within a time range of approximately 05 seconds.

The optimization control method for streaming video transmission based on deep reinforcement learning can achieve better data transmission performance through real-time network state perception and active adjustment strategies based on user needs learning. Stability analysis can help us evaluate the robustness of optimized control methods. In practical applications, the streaming media transmission environment may be affected by various factors, such as network latency, bandwidth changes, etc [35]. Stability analysis can help us evaluate the performance of optimized control methods under these changing conditions and determine their robustness and adaptability. Adopting reinforcement learning algorithms for training and optimization in real environments, continuously adjusting strategies through interaction with the environment to maximize the long-term reward function and provide a better user experience. In real-time video transmission, based on the current network environment and video quality requirements, the strategy obtained through deep reinforcement learning model adjusts the video bit rate and quality according to priority to provide the best viewing experience [36]. Through this proactive and coordinated streaming data transmission method, the system can automatically adjust according to real-time requirements in complex network environments, providing a better video playback experience, reducing buffering time and image lag.

C. OPTIMAL TRANSMISSION CONTROL

In the process of streaming video transmission, there are many factors that may affect video quality, such as bandwidth fluctuations, network latency, packet loss, etc. To ensure a good viewing experience for users, it is necessary to adjust parameters such as bit rate, resolution, and buffering strategy in real-time during the transmission process. Traditional methods often adopt fixed strategies or adjust based on certain rules, lacking flexibility and adaptability. This model takes the current network state and video features as inputs and outputs an optimal transmission parameter sequence. This sequence can maximize the quality of video transmission and user experience, while avoiding network congestion and resource waste [31]. When the media plays smoothly, the behavior of the timestamp in the RTP header is similar to that of the transmission timestamp, which follows the following approximate equation [32]:

$$\frac{T_r - T_{r1}}{T_r - T_{r2}} = \frac{T_e - T_{m1}}{T_e - T_{m2}} \tag{14}$$

Among them, T_r represents the timestamp in the RTP header, and T_m represents the timestamp of the sent packet. T_e is the expected time to send a packet. When maintaining the latest two pairs of T_r(T_{r1} and T_{r2}) and T_m(T_{m1} and T_{m2}), T_e can be calculated using the following equation.

$$T_e = \frac{(T_r - T_{r1})\,T_{m2} - (T_r - T_{r2})\,T_{m1}}{T_{r2} - T_{r1}} \tag{15}$$

This section will introduce the transmission packet loss analysis model. Let $R = \{1, 2, \cdots, r\}$ represent the set of subflows in a multiplexing connection. For any sub stream i in R, its transmission loss analysis model can be expressed as follows:

$$\varphi_i = \frac{1}{1 + \mu^{v(1-p_i)\cdot\sigma_i\cdot c_i}} \tag{16}$$

Among them φ_i is the transfer factor, σ_i is a parameter that reflects the characteristics of RTT, and ci is a parameter that describes the swnd state σ_i and c_i evaluate the transmission status from two different perspectives ρ_i is a parameter that reflects the network packet loss rate. Parameters μ and v are constants that determine the offset and sensitivity of the TFE model, respectively. The optimization control algorithm framework is shown in Figure 4. This work constructs an appropriate model based on factors such as network topology, transmission protocols, and video encoding, and use deep reinforcement learning algorithms for training and optimization. In addition, the bandwidth allocation problem in multi-user scenarios can also be considered, and overall performance can be improved through collaborative learning and resource allocation strategies.

This transmission optimization control method based on deep reinforcement learning has strong adaptability and adaptability, and can dynamically adjust transmission parameters according to different network environments and video content to achieve the best transmission effect [33]. In addition, this method can also improve transmission performance and user satisfaction through continuous learning and optimization [34], [35].

The method based on deep reinforcement learning can automatically learn the optimal decision by training an intelligent agent to optimize the process of video transmission. The core idea of this method is to model the problem as a Markov Decision Process (MDP), where the agent perceives the network state and environmental feedback based on the current state, selects the best action to achieve transmission optimization.

Algorithm 1 Deep Reinforcement Learning-Based Optimal Transmission For Streaming Videos

```
Input: The standardized popularity value pop (v)
    of live video data, maximizing expected cumulative
    returns as optimization goal J(θ), any element i, and
    training sample N.
Standardized the popularity data
for all r = 1 to n do
  Calculate the value of pop (v) for eq-1
      Request and the edge server has the
    requested resource
      for J(θ)1: n
      Strategy Gradient Update
          State value function iteration
        if the quality and spatiotemporal
    fluctuations of the video
          Predict viewing probability instead
        else
          Reset the weight values in the
    convolutional kernel of the layer
Recalculate until the threshold is
    met
  end for
end for
```

V. EXPERIMENTS AND ANALYSIS

A. COMPARISON OF REINFORCEMENT LEARNING UNDER DIFFERENT VIDEO STREAMS

Different types of video streams may have their own unique characteristics, such as dynamism, complexity, and image quality requirements, which can affect the effectiveness of reinforcement learning algorithms. Therefore, it is necessary to train and adjust reinforcement learning models for different situations to maximize the quality of video transmission and user experience. The comparison of decisions made by

QE-MCC, DQN, MPC, and M-QOE in different iteration sequences is shown in Figure 5, and their delivery patterns are shown in Figure 6.

Based on the provided data, we can see the impact of different optimization control algorithms on video stream quality (QOE score) under different iteration orders. When the iteration order is 0, the QOE score of M-QOE optimization control is the highest, at 0.8, followed by the QOE score of MPC optimization control at 0.7, QE-MCC optimization control at 0.4, and DQN optimization control at 0.3. As the iteration order increases, the QOE scores of all optimization control algorithms have improved. At iteration order 1, the QOE score of M-QOE optimization control is the highest, at 1.1, MPC optimization control has a QOE score of 0.8, QE-MCC optimization control has a score of 0.8, and DQN optimization control has a score of 0.5. In iteration order 2, the QOE score of M-QOE optimization control is the highest, at 1.6, MPC optimization control has a QOE score of 1.3, QE-MCC optimization control has a score of 1.0, and DQN optimization control has a score of 0.8. As the iteration order increases, the QOE scores of various optimization control algorithms show an increasing trend, and M-QOE optimization control performs best in all iteration orders.

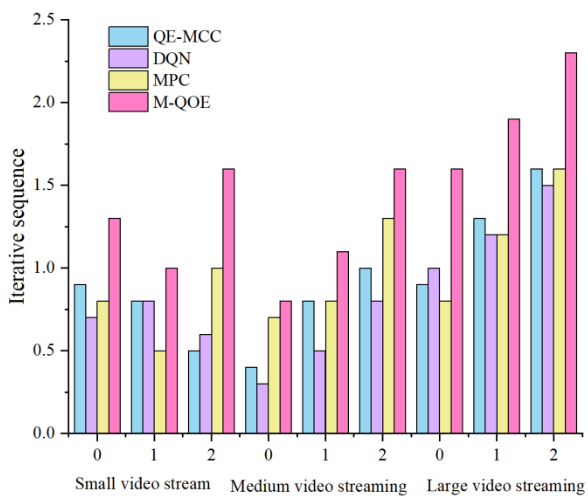


FIGURE 5. Comparison of decision values for several reinforcement learning methods.

B. OPTIMIZATION ANALYSIS OF BANDWIDTH ALLOCATION IN VIDEO STREAMS

The bandwidth allocation in video streaming transmission mainly involves the rational utilization of network resources and the balanced allocation of video quality. Traditional methods often rely on static rules to partition and allocate bandwidth, but this approach is difficult to cope with the dynamic changes in the network environment and the diversity of video content. Deep reinforcement learning can model key performance indicators (such as latency, throughput, and video quality) during video streaming transmission, construct optimization problems, and obtain optimal strategies through

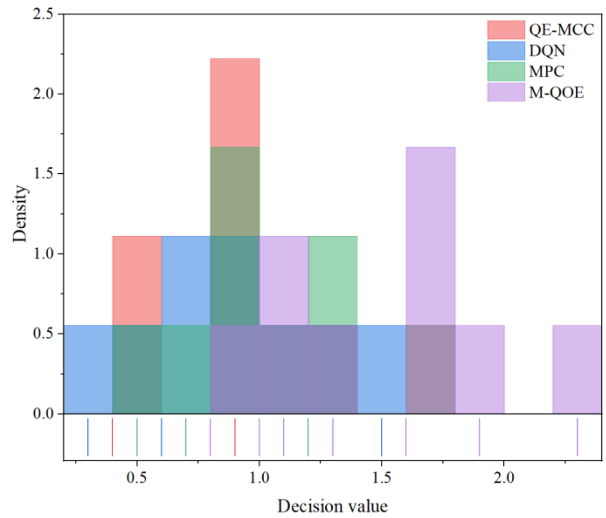


FIGURE 6. Display of decision value delivery pattern.

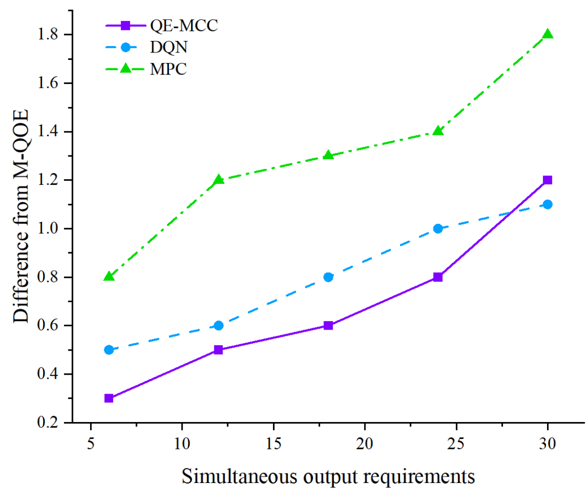


FIGURE 7. Video stream output requirements M-QOE differential summary.

training. Intelligent agents can choose appropriate bandwidth allocation methods based on the current network status and video characteristics, thereby maximizing user experience and video quality. These constraints will affect the allocation scheme of video stream bandwidth. The viewing experience of users is one of the key factors in evaluating video stream bandwidth allocation schemes. The viewing experience can be measured by many indicators, such as video retention rate, buffer waiting time, start playback delay, etc. The M-QOE difference required for video stream output during the same period is shown in Figure 7.

According to the data analysis in Figure 7, Based on the provided data, we can see that the difference in QOE scores between the optimization control algorithms and the M-QOE algorithm varies for different video stream output requirements at different time periods. When the required number of outputs is 6 video streams, the QOE score difference of QE-MCC algorithm is 0.3, the QOE score difference of

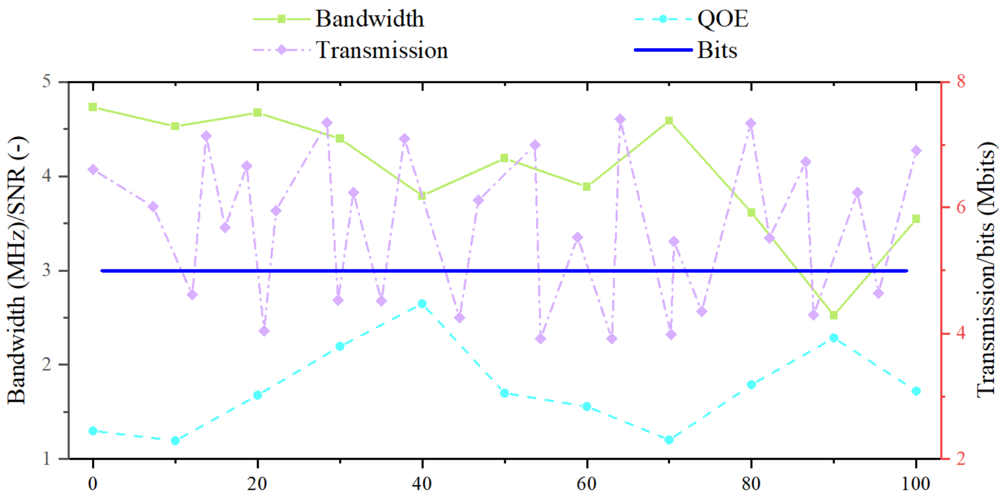


FIGURE 8. Bandwidth allocation and bit rate changes.

DQN algorithm is 0.5, and the QOE score difference of MPC algorithm is 0.8. As the number of video stream output requirements increases, the difference in QOE scores of each algorithm also gradually increases, indicating that in these situations, the M-QOE algorithm has better performance compared to other algorithms. It should be noted that the QOE (Quality of Experience) score difference is used to evaluate the performance differences of different optimization control algorithms in terms of video stream output quality. Based on the given data, it can be seen that as the number of video stream output requirements increases, the difference in QOE scores of each algorithm will also increase. This may mean that in higher demand video stream output situations, better algorithms are needed to ensure the quality of user experience.

Based on the above factors, obtain video quality levels, bandwidth constraints, and viewing experience data under different scenarios through experiments or real world user data collection. These data will provide a basis for subsequent analysis. Clean and preprocess the collected data to remove outliers and noise. Ensure the accuracy and consistency of data. Extract useful features from the collected data, such as video quality levels, bandwidth constraints, and indicators of viewing experience. These features will be used to construct an optimization model, as shown in Figure 8.

According to the data in Figure 8, we can see that the relationship between bandwidth (MHz) and QOE value is not a simple positive relationship. When the bandwidth is 0, the QOE value is 1.3003. This may be because there is no available bandwidth, the network connection is very poor, and the user experience is very low. As the bandwidth gradually increases from 10 to 70, the QOE value gradually increases from 1.1957 to 1.20413. This indicates that increasing bandwidth can improve user experience, but the improvement is relatively small. When the bandwidth is 80, the QOE value drops to 1.78954, which may be due to channel limitations or other network factors leading to a decrease in network quality. At a bandwidth of 90, the QOE value sharply drops to 2.28553. This may be due to signal interference or other technical issues caused by excessive bandwidth, which in turn reduces the user experience. At a bandwidth of 100, the QOE value drops back to 1.72206, indicating a decrease in user experience compared to the QOE value at higher bandwidths. When the bandwidth is small, the QOE value is low, but when the bandwidth reaches a certain level, the QOE value reaches its peak, and then decreases as the bandwidth increases. This may be because when the bandwidth is small, the data transfer rate is slow, which can affect the user experience; When the bandwidth is too high, signal interference and channel limitations can lead to a decrease in network quality. In addition, we can also see that the QOE value is relatively high when the SNR is high, indicating that the signal-to-noise ratio has a significant impact on user experience.

C. STABILITY ANALYSIS OF OPTIMIZED CONTROL METHODS

In research, stability analysis of optimization control methods is a very important aspect. Stability analysis mainly evaluates the stability of the proposed deep reinforcement learning based streaming video transmission optimization control method in the application process of the system. The balance between video quality and network bandwidth is crucial for user experience in streaming video transmission. Optimization control methods can improve video transmission quality by monitoring network status and adjusting video encoding parameters. Stability analysis can help us determine whether the performance of optimized control methods is stable and predict their performance in different environments. Frequency domain analysis can provide useful information about system stability and performance, as shown in Figure 9.

The data in Figure 9 shows that from the relationship between the completion times of M-QOE and the number of requirements, as the number of requirements increases, the completion times of M-QOE also increase, indicating

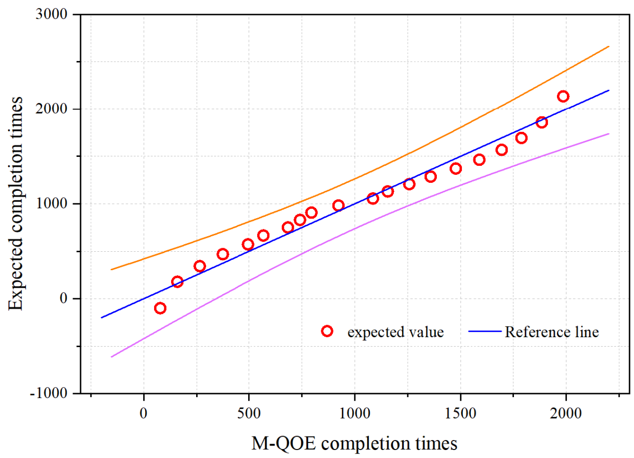


FIGURE 9. Delivery of stability analysis results.

that the algorithm is effective to a certain extent. However, the number of M-QOE completed in the observed data is not a monotonically increasing trend. Especially when the demand frequency is 900, the completion frequency of M-QOE slightly decreases, which may be due to certain specific conditions or abnormal situations. From the perspective of stability, the stability performance of algorithms is inconsistent. In the case of low demand frequency, the algorithm has greater volatility; After the number of demands exceeds 900, the volatility of the algorithm decreases and tends to stabilize. In summary, from the perspective of stability of the M-QOE optimization control algorithm, it has higher robustness in data preprocessing, outlier handling, parameter tuning, and data monitoring.

VI. CONCLUSION

This study is based on a deep reinforcement learning optimization control method for streaming video transmission. Through the combination of deep learning algorithms and reinforcement learning theory, the optimization control problem in the streaming video transmission process is deeply studied. We have demonstrated the effectiveness of deep reinforcement learning algorithms in optimizing control of streaming video transmission through research and experimental verification. Deep reinforcement learning can intelligently adjust the transmission parameters of videos, such as bit rate and resolution, through learning and training, in order to achieve the best video transmission effect. Compared with traditional manual adjustment methods, deep reinforcement learning can automatically optimize the video transmission process, improve the quality of video transmission and user experience.

Our research also focuses on the impact of network environment instability on video transmission. Through the training and learning of deep reinforcement learning algorithms, we can adjust video transmission strategies based on the dynamic situation of the network, and improve the transmission quality of videos in unstable network environments. The experimental results show that in unstable network environments, control methods based on deep reinforcement learning can effectively reduce video lag and image quality loss.

This study achieved significant results in optimizing video transmission performance using a deep reinforcement learning based streaming video transmission optimization control method. Through the application of deep reinforcement learning algorithms, we can intelligently adjust video transmission strategies, reduce video latency, and improve video transmission quality and user experience. In the future, we will further explore the application of deep reinforcement learning in the field of optimizing and controlling streaming video transmission, and continuously improve and perfect existing methods to better meet the needs of users for high-quality video transmission.

REFERENCES

[1] Z. Guo, K. Yu, A. K. Bashir, D. Zhang, Y. D. Al-Otaibi, and M. Guizani, ''Deep information fusion-driven POI scheduling for mobile social networks,'' *IEEE Netw.*, vol. 36, no. 4, pp. 210–216, Jul. 2022.

[2] L. Cui, D. Su, S. Yang, Z. Wang, and Z. Ming, ''TCLiVi: Transmission control in live video streaming based on deep reinforcement learning,'' *IEEE Trans. Multimedia*, vol. 23, pp. 651–663, 2021.

[3] J. Yang, F. Lin, C. Chakraborty, K. Yu, Z. Guo, A.-T. Nguyen, and J. J. P. C. Rodrigues, ''A parallel intelligence-driven resource scheduling scheme for digital twins-based intelligent vehicular systems,'' *IEEE Trans. Intell. Vehicles*, vol. 8, no. 4, pp. 2770–2785, Jun. 2023.

[4] T. Zhang and S. Mao, ''Smart power control for quality-driven multiuser video transmissions: A deep reinforcement learning approach,'' *IEEE Access*, vol. 8, pp. 611–622, 2020.

[5] X. Zhu, F. Ma, F. Ding, Z. Guo, J. Yang, and K. Yu, ''A low-latency edge computation offloading scheme for trust evaluation in finance-level artificial intelligence of things,'' *IEEE Internet Things J.*, vol. 11, no. 1, pp. 114–124, Jan. 2024.

[6] C. Fang, H. Xu, Y. Yang, Z. Hu, S. Tu, K. Ota, Z. Yang, M. Dong, Z. Han, F. R. Yu, and Y. Liu, ''Deep-reinforcement-learning-based resource allocation for content distribution in fog radio access networks,'' *IEEE Internet Things J.*, vol. 9, no. 18, pp. 16874–16883, Sep. 2022.

[7] M. Abbasi, A. Shahraki, M. J. Piran, and A. Taherkordi, ''Deep reinforcement learning for QoS provisioning at the MAC layer: A survey,'' *Eng. Appl. Artif. Intell.*, vol. 102, Jun. 2021, Art. no. 104234.

[8] Z. Tian, L. Zhao, L. Nie, P. Chen, and S. Chen, ''Deeplive: QoE optimization for live video streaming through deep reinforcement learning,'' in *Proc. IEEE 25th Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2019, pp. 827–831.

[9] D. Liu, X. Tan, Y. Liu, Y. He, and Q. Zheng, ''Reinforcement learning based dynamic adaptive video streaming for multi-client over NDN,'' in *Proc. 4th Int. Conf. Hot Inf.-Centric Netw. (HotICN)*, Nov. 2021, pp. 68–73.

[10] J. Luo, F. R. Yu, Q. Chen, and L. Tang, ''Adaptive video streaming with edge caching and video transcoding over software-defined mobile networks: A deep reinforcement learning approach,'' *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1577–1592, Mar. 2020.

[11] N. Hu, X. Cen, F. Luan, L. Sun, and C. Wu, ''A novel video transmission optimization mechanism based on reinforcement learning and edge computing,'' *Mobile Inf. Syst.*, vol. 2021, pp. 1–10, Oct. 2021.

[12] Y. Guo, F. R. Yu, J. An, K. Yang, Y. He, and V. C. M. Leung, ''Buffer-aware streaming in small-scale wireless networks: A deep reinforcement learning approach,'' *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6891–6902, Jul. 2019.

[13] N. Kan, J. Zou, K. Tang, C. Li, N. Liu, and H. Xiong, ''Deep reinforcement learning-based rate adaptation for adaptive 360-degree video streaming,'' in *Proc. ICASSP IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4030–4034.

[14] D. Liu, J. Zhao, and C. Yang, ''Energy-saving predictive video streaming with deep reinforcement learning,'' in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[15] F. Shah-Mohammadi and A. Kwasinski, "Deep reinforcement learning approach to QoE-driven resource allocation for spectrum underlay in cognitive radio networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[16] H. Lin, B. Zhang, Y. Cao, Z. Liu, and X. Chen, "A deep reinforcement learning approach for point cloud video transmissions," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2021, pp. 1–5.

[17] J. Li, H. Zhang, and H. Ma, "DRL-based transmission control for QoE guaranteed transmission efficiency optimization in tile-based panoramic video streaming," *Multimedia Syst.*, vol. 29, no. 5, pp. 2761–2777, Oct. 2023.

[18] K. Tang, N. Kan, J. Zou, C. Li, X. Fu, M. Hong, and H. Xiong, "Multi-user adaptive video delivery over wireless networks: A physical layer resource-aware deep reinforcement learning approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 798–815, Feb. 2021.

[19] L. Liu, H. Hu, Y. Luo, and Y. Wen, "When wireless video streaming meets AI: A deep learning approach," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 127–133, Apr. 2020.

[20] Q. Lan, B. Lv, R. Wang, K. Huang, and Y. Gong, "Adaptive video streaming for massive MIMO networks via approximate MDP and reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5716–5731, Sep. 2020.

[21] P.-Y. Chou, W.-Y. Chen, C.-Y. Wang, R.-H. Hwang, and W.-T. Chen, "Deep reinforcement learning for MEC streaming with joint user association and resource management," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–7.

[22] T. Huang, R.-X. Zhang, and L. Sun, "Self-play reinforcement learning for video transmission," in *Proc. 30th ACM Workshop Netw. Operating Syst. Support Digit. Audio Video*, Jun. 2020, pp. 7–13.

[23] X. Zhang, Y. Hu, and Z. Li, "Live video streaming optimization based on deep reinforcement learning," in *Proc. 12th Int. Conf. Mach. Learn. Comput.*, Feb. 2020, pp. 116–120.

[24] X. Zhou, W. Liang, K. Yan, W. Li, K. I. Wang, J. Ma, and Q. Jin, "Edge-enabled two-stage scheduling based on deep reinforcement learning for Internet of everything," *IEEE Internet Things J.*, vol. 10, no. 4, pp. 3295–3304, Feb. 2023.

[25] Z. Ming, J. Chen, L. Cui, S. Yang, Y. Pan, W. Xiao, and L. Zhou, "Edge-based video surveillance with graph-assisted reinforcement learning in smart construction," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9249–9265, Jun. 2022.

[26] M. Al Jameel, T. Kanakis, S. Turner, A. Al-Sherbaz, and W. S. Bhaya, "A reinforcement learning-based routing for real-time multimedia traffic transmission over software-defined networking," *Electronics*, vol. 11, no. 15, p. 2441, Aug. 2022.

[27] Z. Guo, Y. Shen, A. K. Bashir, M. Imran, N. Kumar, D. Zhang, and K. Yu, "Robust spammer detection using collaborative neural network in Internet-of-Things applications," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9549–9558, Jun. 2021, doi: 10.1109/JIOT.2020.3003802.

[28] X. He, K. Wang, H. Huang, T. Miyazaki, Y. Wang, and S. Guo, "Green resource allocation based on deep reinforcement learning in content-centric IoT," *IEEE Trans. Emerg. Topics Comput.*, vol. 8, no. 3, pp. 781–796, Jul. 2020.

[29] T. Huang, R.-X. Zhang, and L. Sun, "Zwei: A self-play reinforcement learning framework for video transmission services," *IEEE Trans. Multimedia*, vol. 24, pp. 1350–1365, 2022.

[30] P. Amaral and D. Simues, "Deep reinforcement learning based routing in IP media broadcast networks: Feasibility and performance," *IEEE Access*, vol. 10, pp. 62459–62470, 2022.

[31] D. Nallaperuma, R. Nawaratne, T. Bandaragoda, A. Adikari, S. Nguyen, T. Kempitiya, D. De Silva, D. Alahakoon, and D. Pothuhera, "Online incremental machine learning platform for big data-driven smart traffic management," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 12, pp. 4679–4690, Dec. 2019.

[32] Y. Zhang, Z. Wang, J. Liu, H. Du, Q. Zheng, and W. Zhang, "Deep reinforcement learning based adaptive 360-degree video streaming with field of view joint prediction," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2022, pp. 1–8.

[33] Q. Li, L. Liu, Z. Guo, P. Vijayakumar, F. Taghizadeh-Hesary, and K. Yu, "Smart assessment and forecasting framework for healthy development index in urban cities," *Cities*, vol. 131, Dec. 2022, Art. no. 103971.

[34] X. Wei, M. Zhou, S. Kwong, H. Yuan, S. Wang, G. Zhu, and J. Cao, "Reinforcement learning-based QoE-oriented dynamic adaptive streaming framework," *Inf. Sci.*, vol. 569, pp. 786–803, Aug. 2021.

[35] Z. Guo, K. Yu, K. Konstantin, S. Mumtaz, W. Wei, P. Shi, and J. J. P. C. Rodrigues, "Deep collaborative intelligence-driven traffic forecasting in green Internet of Vehicles," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 1023–1035, Jul. 2023.

[36] D. Meng, Y. Xiao, Z. Guo, A. Jolfaei, L. Qin, X. Lu, and Q. Xiang, "A data-driven intelligent planning model for UAVs routing networks in mobile Internet of Things," *Comput. Commun.*, vol. 179, pp. 231–241, Nov. 2021.

YAWEN YANG received the master's degree from Hunan University, in 2013. She is currently with Hunan International Economics University, majoring in internet and new media. Her research interests include multimedia computing and video technology.

YUXUAN XIAO received the master's degree from Hunan University, in 2009. She is currently with Changsha University of Science and Technology, majoring in information systems and smart design. Her research interests include video processing and multimedia systems.