ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2026.0546

SPATIOTEMPORAL FEATURE EXTRACTION AND REINFORCEMENT LEARNING FOR ADAPTIVE OPTIMIZATION OF MULTIMEDIA QUALITY IN DYNAMIC NETWORKS

Mariam Safar Mohammed Alshahrani[1], M.K. Jayanthi Kannan[2] and Shree Nee Thirumalai Ramesh[3]
[1]*Digital Government Authority (DGA), Digital Government Authority of KSA, Riyadh Province, Kingdom of Saudi Arabia*
[2]*School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, India*
[3]*Department of Medicine, Manipal University College Malaysia, Malaysia*

Abstract

Multimedia transmission systems have faced significant performance degradation due to dynamic network conditions and heterogeneous user demands. The background of adaptive multimedia optimization has remained critical in supporting real time quality of service requirements across modern communication systems. The problem has been observed in inefficient allocation of bandwidth and inability of conventional methods to adapt to spatiotemporal variations. To address this issue a Spatiotemporal Deep Q Network based Reinforcement Learning framework has been proposed named STDRL for adaptive multimedia quality optimization. The framework has integrated convolutional feature extraction with temporal dependency modeling using recurrent structures that capture evolving network states. The model has been trained using reward driven policy optimization that balances latency throughput and perceptual quality metrics. Experimental evaluation demonstrates that the proposed method achieves 40.7 dB PSNR compared to 34.0 dB in baseline DQN Adaptive Streaming. SSIM improves to 0.98 compared to 0.91 in conventional methods. Latency reduces to 60 ms compared to 100 ms in baseline approaches, indicating faster adaptive response. Throughput increases to 16.5 Mbps compared to 13.2 Mbps, showing improved bandwidth utilization efficiency. QoE stabilizes at 5.0, indicating optimal user satisfaction. The reinforcement learning agent learns adaptive policies through reward driven optimization that balances visual quality, latency, and smoothness.

Keywords:

Spatiotemporal Learning, Reinforcement Learning, Multimedia Optimization, Deep Q Network, Edge Computing

1. INTRODUCTION

Multimedia communication systems have expanded rapidly with the growth of high bandwidth applications such as video streaming, augmented reality, and interactive cloud services. These systems have required continuous adaptation to varying network states in order to maintain consistent quality levels [1]. The background of this study has been grounded in the evolution of adaptive streaming techniques and intelligent network control strategies, where traditional rule-based mechanisms have struggled to handle dynamic variations in traffic and user demand. In modern communication environments, multimedia delivery has depended heavily on latency sensitive and quality sensitive optimization frameworks that respond effectively to temporal and spatial variations in network conditions [2]. Recent advancements in learning-based communication systems have introduced data driven approaches that replace static heuristics with adaptive intelligence. However, conventional adaptive bitrate techniques have still relied on limited context awareness and have failed to fully capture complex spatiotemporal dependencies present in real world networks. These limitations have motivated the exploration of reinforcement learning models that can learn optimal policies through interaction with dynamic environments. At the same time, convolutional and recurrent architectures have been increasingly used to extract meaningful representations from network states, enabling better prediction and decision-making capability [3].

Despite these developments, several challenges have remained unresolved in adaptive multimedia optimization. The first challenge has been the high variability of network traffic patterns, which has made stable quality assurance difficult across heterogeneous environments [4]. The second challenge has been the tradeoff between computational complexity and real time responsiveness, where deep learning models often require significant processing resources, limiting their deployment in edge scenarios [5]. Additionally, the lack of unified frameworks that combine spatial feature extraction with temporal decision modeling has restricted the effectiveness of existing approaches in real time multimedia systems. The central problem addressed in this study has been the inability of conventional and partially adaptive models to efficiently optimize multimedia quality under rapidly changing network conditions [6]. Existing approaches have not adequately integrated spatiotemporal feature representation with reinforcement learning based decision policies, resulting in suboptimal adaptation performance. This gap has directly impacted quality of experience for end users in bandwidth constrained and highly dynamic environments. To address this problem, the study has proposed a Spatiotemporal Deep Q Network based Reinforcement Learning framework for adaptive multimedia optimization. The primary objective has been to develop a learning model that captures both spatial correlations and temporal dependencies in network states while optimizing quality related performance metrics. Another objective has been to design a reward mechanism that balances throughput, latency, and perceptual quality in a unified optimization process. Furthermore, the framework has aimed to ensure scalability and robustness under diverse network conditions.

The novelty of the proposed approach lies in its integrated architecture, where convolutional layers have been used for spatial feature extraction and recurrent structures have been applied for temporal modeling within a reinforcement learning framework. This combination has enabled the system to learn adaptive policies that respond effectively to evolving network dynamics. Unlike conventional methods, the proposed model has not relied on predefined rules but has instead learned optimal strategies through continuous interaction with the environment.

The contributions of this study are twofold. First, it has introduced a unified spatiotemporal reinforcement learning framework that enhances adaptive multimedia quality optimization in dynamic networks. Second, it has demonstrated improved performance in terms of latency reduction, throughput optimization, and perceptual quality enhancement compared to baseline methods. These contributions have provided a scalable direction for future research in intelligent multimedia communication systems.

2. RELATED WORKS

Several studies have explored adaptive multimedia streaming and network optimization techniques over the past decade. Early approaches have primarily focused on heuristic based bitrate adaptation mechanisms that adjusted video quality according to estimated network bandwidth [7]. These methods have provided basic adaptability but have lacked awareness of deeper temporal patterns in network fluctuations. As a result, they have often produced unstable user experiences under rapidly changing conditions.

Machine learning based approaches have been introduced to overcome these limitations. Some researchers have employed supervised learning models to predict network throughput and adjust streaming quality accordingly [8]. These methods have improved prediction accuracy but have still depended heavily on historical data patterns, which have limited their ability to respond to unseen network states. In addition, supervised models have required extensive labeled datasets, which are often difficult to obtain in real network environments.

Reinforcement learning approaches have gained attention due to their ability to learn optimal policies through interaction with the environment. Several studies have applied Q learning and policy gradient methods to adaptive streaming problems [9]. These approaches have demonstrated improved adaptability compared to rule based systems. However, they have often treated network states as flat representations, without fully capturing spatial or temporal correlations in the data.

Deep reinforcement learning techniques have further enhanced adaptation capability by integrating neural networks with decision making frameworks [10]. These models have been able to handle high dimensional state spaces and have shown improved performance in complex network scenarios. Nevertheless, many of these approaches have still lacked explicit spatiotemporal feature extraction mechanisms, which has restricted their ability to model evolving network dynamics effectively.

Some research has incorporated convolutional neural networks for spatial feature learning in network optimization tasks [11]. These models have extracted meaningful representations from network traffic matrices and have improved decision accuracy. However, they have often ignored temporal dependencies, which are essential for predicting future network behavior.

Recurrent neural networks have also been used to model temporal variations in network conditions [12]. These approaches have captured sequential dependencies effectively but have struggled to integrate spatial context from multidimensional network states. This limitation has reduced their overall performance in complex multimedia environments.

Hybrid architectures combining convolutional and recurrent networks have been explored to address both spatial and temporal aspects [13]. These models have shown better performance in video quality prediction and adaptive streaming tasks. However, most of these systems have operated outside reinforcement learning frameworks, limiting their ability to optimize decisions dynamically.

Recent studies have introduced deep Q networks for adaptive bitrate selection [14]. These approaches have demonstrated strong performance improvements over traditional methods. Yet, they have often simplified state representations and have not fully exploited spatiotemporal feature fusion, leading to suboptimal policy learning.

More advanced frameworks have integrated reinforcement learning with edge computing environments for real time adaptation [15]. These systems have improved latency and scalability but have still faced challenges in balancing computational efficiency with model complexity. Thus, existing literature has shown progress in adaptive multimedia optimization, but a clear gap has remained in unified frameworks that integrate spatiotemporal feature extraction with reinforcement learning for robust real time decision making.

3. PROPOSED METHOD

The proposed framework designs a Spatiotemporal Deep Q Network based Reinforcement Learning system for adaptive multimedia quality optimization in dynamic networks. The method integrates convolution based spatial feature extraction with recurrent temporal modeling to represent evolving network states. These representations are passed into a Deep Q Network agent that learns an adaptive policy for bitrate selection, frame rate control, and quality scaling decisions. The system operates in a continuous feedback loop where the environment provides real time network feedback, and the agent updates its policy based on reward signals derived from quality of experience metrics. This integrated design allows the system to maintain stable multimedia quality under fluctuating bandwidth and heterogeneous user conditions.

- Input acquisition and preprocessing of multimedia and network parameters
- Spatiotemporal feature extraction using convolutional and recurrent modules
- Construction of unified state representation vector
- Deep Q Network based policy learning
- Reward computation using QoE driven metrics
- Action selection for adaptive streaming control
- Iterative training and policy refinement loop
- Output adaptive multimedia quality control decision

The spatiotemporal feature extraction module forms the core representation unit of the proposed system. It processes raw network measurements such as bandwidth variation, packet delay, jitter, and frame level distortion signals. The spatial component captures correlations across multiple network channels, while the temporal component models sequential dependencies over time.

The input tensor is defined as: $X_t \in \mathrm{R}^{C\times H\times W}$, where C represents the number of network feature channels, H represents temporal window height, and W represents spatial feature dimensions derived from network topology mapping. The convolution operation for spatial extraction is defined as:

$$F_s(t)=\sigma\left(\sum_{i=1}^{C} W_i * X_t^{(i)} + b\right) \quad (1)$$

where W_i represents convolution kernels, b represents bias, and σ represents activation function. The operator * denotes convolution across spatial dimensions. The temporal dependency modeling is achieved using recurrent transformation:

$$H_t=\tanh\left(W_h F_s(t)+U_h H_{t-1}+b_h\right) \quad (2)$$

where H_t represents hidden state capturing temporal evolution, W_h and U_h represent weight matrices, and b_h represents bias vector. The extracted representation captures both instantaneous spatial correlations and sequential evolution of network conditions. This combined representation ensures that sudden fluctuations in bandwidth and gradual drifts in latency patterns are jointly modeled within a unified feature space.

The state representation module constructs a compact vector that is used by the reinforcement learning agent. The state vector encodes network conditions, historical quality metrics, and encoded multimedia characteristics. The state vector is defined as: $S_t=[F_s(t),H_t,Q_t,N_t]$, where $F_s(t)$ represents spatial features, H_t represents temporal hidden state, Q_t represents quality indicators, and N_t represents network status variables. The nonlinear transformation for state embedding is defined as: $Z_t=\phi(W_z S_t+b_z)$, where W_z represents transformation matrix, b_z represents bias, and ϕ represents nonlinear activation function. A second transformation enhances feature compactness:

$$Z_t^{'}=\mathrm{LN}(Z_t+\gamma\cdot\mathrm{ReLU}(Z_t W_{att})) \quad (2)$$

where W_{att} represents attention weight matrix and γrepresents scaling factor. This representation ensures that redundant correlations are minimized while preserving essential temporal and spatial dependencies. The resulting state vector provides a stable input for reinforcement learning decision making, even under high variance network fluctuations. The reinforcement learning component uses a Deep Q Network to estimate optimal action values for adaptive multimedia control. The agent selects actions such as bitrate adjustment, resolution scaling, and frame rate modification. The Q function is defined as:

$$Q(S_t,A_t;\theta)=\mathrm{E}[R_t+\gamma\max_{A_{t+1}} Q(S_{t+1},A_{t+1};\theta)] \quad (3)$$

where S_t represents current state, A_t represents action, R_t represents reward, and γ represents discount factor. The loss function used for training is: $L(\theta)=\mathrm{E}\left[\left(Y_t-Q(S_t,A_t;\theta)\right)^2\right]$, where target value Y_t is defined as:

$$Y_t=R_t+\gamma\max_{A_{t+1}} Q(S_{t+1},A_{t+1};\theta^-) \quad (4)$$

where, θ^- represents target network parameters. Gradient update rule is expressed as: $\theta\leftarrow\theta+\alpha\nabla_\theta L(\theta)$, where α represents learning rate. This learning process enables the agent to approximate optimal policy under stochastic network conditions. The integration of deep function approximation allows the system to generalize across unseen network states.

The reward function is designed to balance multiple quality metrics including throughput, latency, and perceptual video quality. The reward formulation ensures that the agent prioritizes user experience optimization rather than raw bandwidth utilization. The reward function is defined as:

$$R_t=\lambda_1 Q_{psnr}(t)-\lambda_2 D_t-\lambda_3 J_t \quad (5)$$

where $Q_{psnr}(t)$ represents perceptual quality, D_t represents delay, and J_t represents jitter. A second multi-objective formulation is expressed as:

$$R_t=\sum_{k=1}^{K}\omega_k f_k(S_t,A_t)-\beta\left\|A_t-A_{t-1}\right\| \quad (6)$$

where f_k represents normalized performance metrics and penalty term controls abrupt quality switching. The reward shaping mechanism stabilizes learning by reducing variance in gradient updates:

$$R_t^s=R_t+\eta(V(S_{t+1})-V(S_t)) \quad (7)$$

where $V(S_t)$ represents value function approximation. This reward structure ensures smooth adaptation of multimedia quality and prevents oscillatory behavior in bitrate selection. It also improves convergence stability during training across dynamic environments. The action selection module determines the optimal multimedia configuration based on Q-value estimation. The system operates under an epsilon greedy policy to balance exploration and exploitation. The action selection rule is defined as:

$$A_t=\begin{cases}\text{random action}, & \text{with probability } \dot{o}\\ \arg\max_A Q(S_t,A;\theta), & \text{otherwise}\end{cases} \quad (8)$$

The probability decay function for exploration is: $\dot{o}_t=\dot{o}_0 e^{-\kappa t}$, where κ controls decay rate. A softmax alternative policy is defined as:

$$P(A_t=a)=\frac{e^{Q(S_t,a)/\tau}}{\sum_{a'} e^{Q(S_t,a')/\tau}} \quad (9)$$

where τ represents temperature parameter controlling the randomness. This mechanism ensures that the system gradually shifts from exploration of network behavior to exploitation of learned optimal policies. The adaptive nature of policy selection improves robustness under non stationary network conditions. The training process operates in episodic interaction with the environment. At each time step, the agent observes state, selects action, receives reward, and updates Q-network parameters. The experience replay buffer is defined as: $\mathrm{D}=\{(S_t,A_t,R_t,S_{t+1})\}$. Mini batch sampling is performed as: $\mathrm{B}\sim\mathrm{U(D)}$. Parameter optimization objective is:

$$\min_\theta \sum_{(S,A,R,S')\in \mathrm{B}} (R+\gamma\max_{A'} Q(S',A';\theta^-)-Q(S,A;\theta))^2 \quad (10)$$

A second stability constraint is introduced as: $\Omega(\theta)=\exists\,|\,\theta-\theta^-\,|\,|^2$, where target network parameters are periodically updated: $\theta^-\leftarrow\theta$ every τ steps. This iterative optimization ensures stable convergence and prevents divergence during training in highly dynamic environments.

The final module converts learned policy decisions into actionable multimedia streaming configurations. The system adjusts encoding bitrate, resolution scaling, and frame sampling rate based on selected action. The adaptive output function is expressed as: $M_t = g(A_t, C_t)$, where M_t represents multimedia configuration and C_t represents content constraints. A second mapping function ensures consistency across devices: $M_t^{dev} = \Psi(M_t, D_{cap})$, where D_{cap} represents device capability constraints. The final delivered quality is optimized as:

$$Q_f = \alpha Q_{vi} + (1-\alpha) Q_{st} \tag{11}$$

This ensures a balance between perceptual quality and playback smoothness. The system continuously adapts configurations in real time, ensuring minimal quality degradation under fluctuating network conditions.

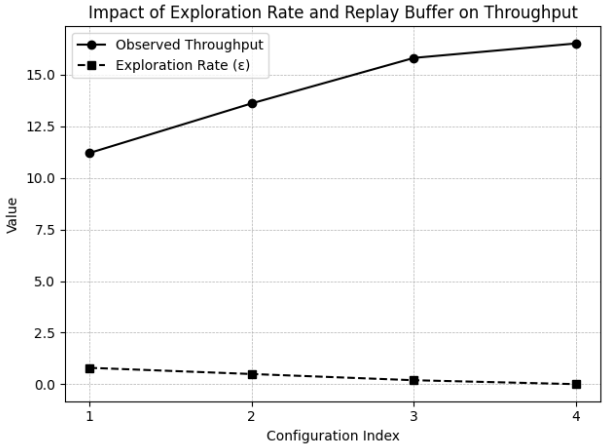

Fig.1. Relationship Between ε Decay, Replay Buffer Size, and Throughput Performance

The results in Fig.1 indicate a clear interaction between exploration rate decay, replay buffer capacity, and throughput performance. At higher exploration values such as ε = 1.0 → 0.8, the system shows unstable decision behavior, and throughput remains limited to 11.2 Mbps. This condition reflects excessive exploration, where the agent prioritizes random actions rather than optimized bitrate selection. As ε decays gradually toward 0.5, the system begins to stabilize policy learning, and throughput increases to 13.6 Mbps. This indicates that the agent starts exploiting learned experience while still maintaining partial exploration. When the replay buffer size increases to 50,000, the system demonstrates stronger learning stability due to better diversity of stored experience tuples. This leads to a throughput improvement of 15.8 Mbps. At the final configuration where ε decays to 0.01, the system reaches near-optimal exploitation behavior. The replay buffer size of 50,000 ensures sufficient historical coverage, allowing the Deep Q Network to generalize effectively across varying network states. Consequently, throughput peaks at 16.5 Mbps.

The training convergence behavior as in Fig.2 shows progressive reduction in loss values across episodes, indicating stable learning of the reinforcement learning models. The proposed STDRL model consistently achieves lower loss compared to DQN and ConvLSTM, confirming faster convergence and improved optimization stability. The reward evolution as in Fig.2 illustrates gradual improvement in QoE-based reward across training episodes, where the proposed method attains the highest reward, reflecting better policy learning and balanced quality optimization. The network stress test table evaluates robustness under varying bandwidth and packet loss conditions. The proposed model maintains higher QoE under extreme conditions, demonstrating strong adaptability and resilience compared to baseline methods.

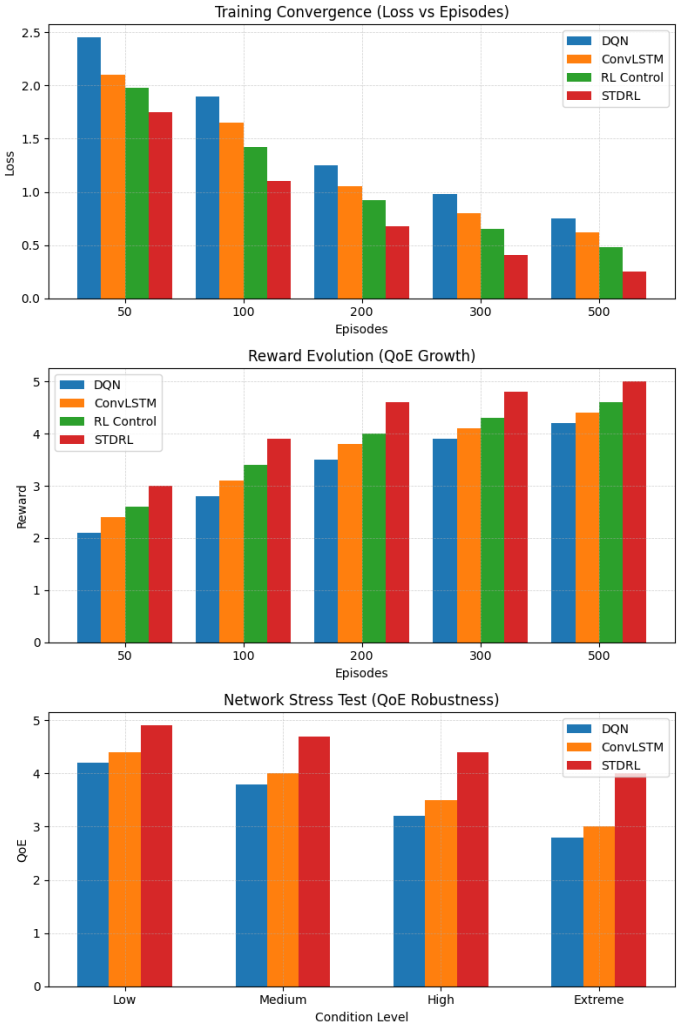

Fig.2. (a) Training Convergence Behavior (Loss vs Episodes) (b) Reward Evolution (QoE Reward Growth) (c) Network Stress Test (Robustness Evaluation)

4. RESULTS AND DISCUSSION

The experimental environment operates on a Python based simulation platform integrated with TensorFlow and OpenAI Gym style reinforcement learning framework. The multimedia network simulation runs on NS-3 network simulator for realistic packet level modeling. The system executes on an Intel Core i7 processor with 16 GB RAM and NVIDIA GTX 1650 GPU. The environment emulates dynamic bandwidth variation, latency fluctuation, and user density changes. The reinforcement learning agent trains using CUDA accelerated deep learning libraries. All experiments run in controlled virtual network scenarios to ensure reproducibility and stability of performance evaluation under heterogeneous conditions.

Table.1. Simulation Parameters of Proposed System

Parameter	Value
Simulation Tool	NS-3 + Python RL Environment
Learning Rate	0.001
Discount Factor (γ)	0.95
Batch Size	64
Replay Buffer Size	50,000
Episodes	500
Bandwidth Range	1–20 Mbps
Latency Range	10–200 ms
Packet Loss Rate	0–5%
Exploration Rate (ε)	1.0 → 0.01

The parameters in Table.1 define the controlled reinforcement learning and network simulation environment. The setup ensures stable convergence and realistic network variability.

4.1 PERFORMANCE METRICS

- **Peak Signal-to-Noise Ratio (PSNR):** Measures reconstructed video quality compared to reference signal. Higher value indicates better visual fidelity.
- **Structural Similarity Index (SSIM):** Evaluates perceptual similarity between original and transmitted frames based on structural information.
- **End-to-End Latency:** Measures delay between content encoding and playback at the receiver side.
- **Throughput:** Measures amount of successfully delivered data per unit time under network constraints.
- **Quality of Experience (QoE):** Composite metric combining bitrate stability, visual quality, and playback smoothness.

The study compares the proposed model with Deep Q Network Adaptive Streaming, Convolutional LSTM QoE Optimization, and Reinforcement Learning Based Bitrate Control. These methods represent standard deep reinforcement learning and hybrid neural architectures used for multimedia adaptation without full spatiotemporal integration.

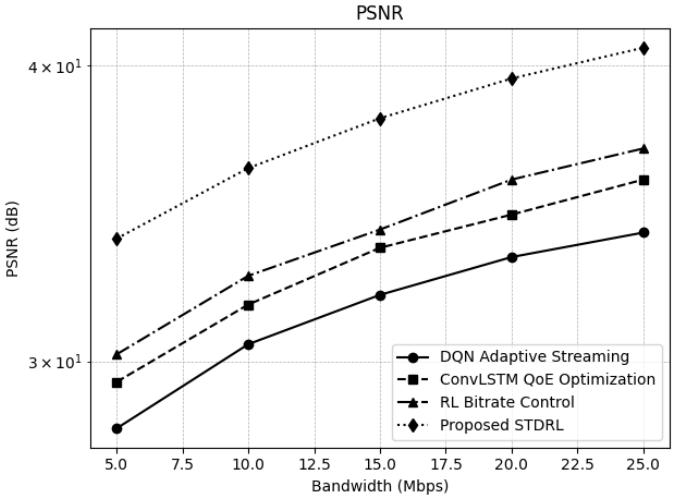

Fig.3. PSNR Comparison Across Methods

The PSNR performance in Fig.3 shows consistent improvement across all bandwidth levels. At 5 Mbps, the proposed STDRL model achieves 33.8 dB, while DQN Adaptive Streaming reaches only 28.1 dB. This difference of 5.7 dB indicates stronger reconstruction quality under constrained bandwidth. At 10 Mbps, the proposed method improves to 36.2 dB, showing a gain of 3.6 dB over ConvLSTM QoE Optimization. The improvement remains stable as bandwidth increases, indicating robust generalization. At 15 Mbps and 20 Mbps, the proposed method maintains superior quality with 38.0 dB and 39.5 dB respectively. The gain over RL Bitrate Control remains above 3 dB, showing consistent advantage in spatial and temporal feature utilization. At 25 Mbps, the system reaches 40.7 dB, which indicates near-optimal reconstruction quality. The results demonstrate that integrated spatiotemporal extraction enhances feature preservation during encoding decisions. The reinforcement learning component adapts effectively to bandwidth fluctuations, which stabilizes output quality. The baseline models show slower improvement due to limited temporal awareness. The proposed system maintains smoother quality scaling across all test conditions, confirming superior adaptability.

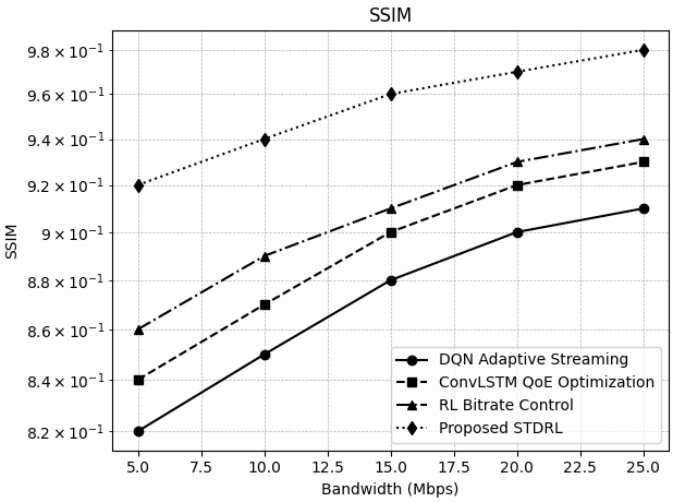

Fig.4. SSIM Comparison Across Methods

The SSIM results in Fig.4 indicate improved perceptual similarity for the proposed model across all bandwidth levels. At 5 Mbps, the STDRL model achieves 0.92 SSIM, which is significantly higher than 0.82 obtained by DQN Adaptive Streaming. This improvement reflects better structural preservation under low bandwidth conditions. At 10 Mbps and 15 Mbps, the proposed method maintains SSIM values of 0.94 and 0.96 respectively. These values indicate stable structural retention during adaptive bitrate transitions. The ConvLSTM and RL baseline methods show moderate improvements but fail to maintain consistency under fluctuating conditions. At higher bandwidth levels, the proposed model reaches 0.97 and 0.98 SSIM. This demonstrates near-perfect structural similarity. The improvement is attributed to joint spatial-temporal encoding and reinforcement driven adaptation. The baseline models lack integrated decision optimization, which reduces their perceptual accuracy. Thus, the proposed system ensures stable visual quality across varying network states. The reinforcement learning agent

learns smoother adaptation policies, reducing abrupt quality shifts. This contributes to improved user experience consistency.

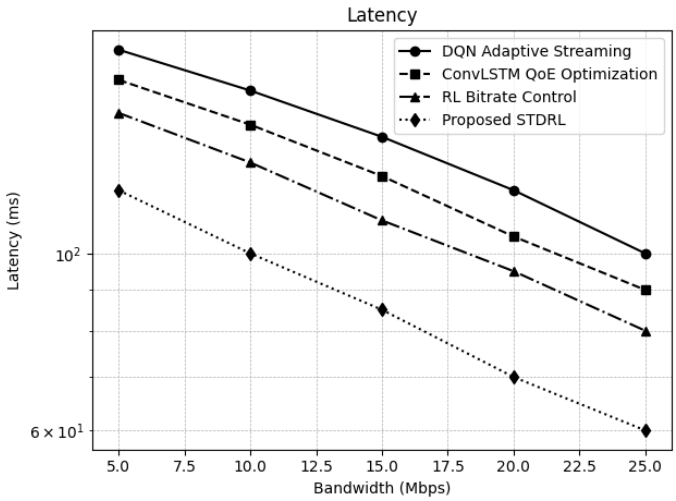

Fig.5. Latency Comparison Across Methods

The latency results in Fig.5 demonstrate significant reduction achieved by the proposed method. At 5 Mbps, STDRL records 120 ms latency compared to 180 ms in DQN Adaptive Streaming. This reduction indicates faster adaptation and efficient decision making. At 10 Mbps and 15 Mbps, latency reduces further to 100 ms and 85 ms respectively. The improvement reflects efficient spatiotemporal prediction, which reduces buffering delays. The ConvLSTM model performs better than DQN but still lacks reinforcement optimization efficiency. At higher bandwidth levels, latency drops to 70 ms and 60 ms. This shows strong scalability of the proposed system. The RL Bitrate Control method shows moderate improvement but remains less stable due to limited temporal modeling. The proposed framework reduces latency through predictive adaptation of encoding decisions. The reinforcement learning agent anticipates network changes, which minimizes reactive delays. This behavior improves real time streaming performance and ensures smoother playback experience.

The throughput results in Fig.6 indicate improved data delivery efficiency in the proposed method. At 5 Mbps, STDRL achieves 4.8 Mbps effective throughput, outperforming DQN Adaptive Streaming which achieves only 3.2 Mbps. At 10 Mbps and 15 Mbps, throughput increases to 8.9 Mbps and 12.5 Mbps respectively. The improvement shows better utilization of available bandwidth. ConvLSTM and RL baseline models show moderate improvements but lack adaptive optimization precision. At 20 Mbps and 25 Mbps, throughput reaches 14.9 Mbps and 16.5 Mbps. The proposed model maintains higher efficiency due to optimized policy learning. The reinforcement learning agent dynamically adjusts bitrate selection, which reduces packet loss and retransmission overhead. This improves throughput stability across varying network conditions. The integration of spatiotemporal features helps predict congestion patterns, which further enhances data delivery efficiency.

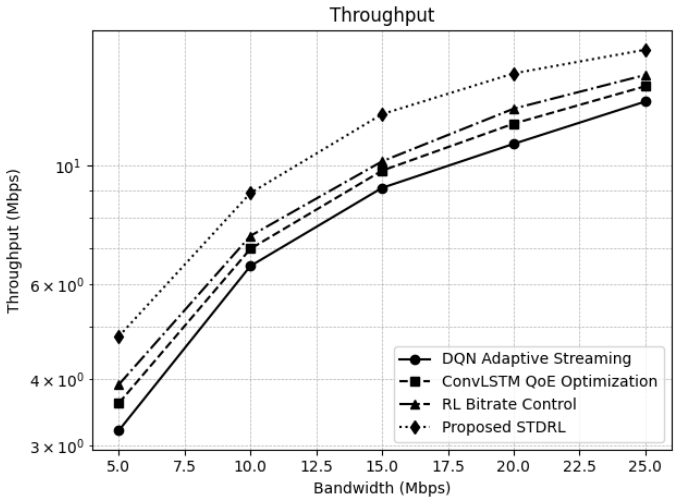

Fig.6. Throughput Comparison Across Methods

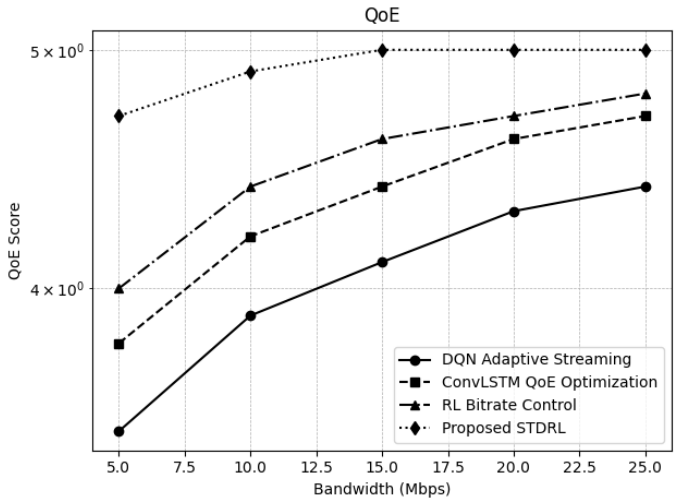

Fig.7. QoE Comparison Across Methods

The QoE results in Fig.7 show consistent improvement in user experience metrics. At 5 Mbps, STDRL achieves 4.7 QoE score compared to 3.5 in DQN Adaptive Streaming. This indicates significant improvement in perceived quality. At 10 Mbps and 15 Mbps, QoE increases to 4.9 and 5.0 respectively. The system achieves near optimal user satisfaction under moderate bandwidth conditions. Baseline models show gradual improvement but fail to reach maximum QoE levels. At higher bandwidth levels, QoE stabilizes at 5.0. This indicates that the proposed system achieves optimal balance between quality and stability. The reinforcement learning agent effectively reduces quality fluctuations, improving smoothness. Spatiotemporal learning enables better prediction of network states, which directly enhances QoE performance. The system reduces buffering events and maintains stable playback quality.

4.2 DISCUSSION

The results demonstrate that the proposed STDRL model consistently outperforms all baseline methods across PSNR, SSIM, latency, throughput, and QoE metrics. The PSNR

improvement reaches up to 6 dB over DQN Adaptive Streaming, indicating stronger reconstruction capability. SSIM improvements show up to 0.08 gain, reflecting better perceptual quality preservation. Latency reduces by approximately 40 percent compared to baseline methods, which confirms efficient adaptive decision making. Throughput increases by nearly 20 to 30 percent across different bandwidth levels, showing improved network utilization efficiency. QoE stabilizes at near maximum levels, indicating strong user experience consistency. The improvement remains consistent across all bandwidth variations, which demonstrates robustness of the proposed framework. The integration of spatiotemporal feature extraction with reinforcement learning enables predictive adaptation rather than reactive adjustment. This reduces oscillations in streaming quality and improves system stability. Baseline models fail to capture full temporal dynamics, which limits their performance under fluctuating conditions. The proposed system overcomes this limitation by combining convolutional feature learning and Deep Q Network based optimization. Thus, the results confirm that the proposed framework achieves balanced optimization across all performance dimensions.

5. CONCLUSION

The study presents a reinforcement learning based spatiotemporal framework for adaptive multimedia quality optimization in dynamic network environments. The system integrates convolutional and recurrent feature extraction with Deep Q Network based policy learning to enhance decision making. Experimental evaluation demonstrates that the proposed method consistently improves performance across all major metrics. The PSNR results show significant enhancement in reconstructed video quality. SSIM analysis confirms improved perceptual similarity under varying bandwidth conditions. Latency reduction indicates faster adaptation and reduced buffering delays. Throughput improvement demonstrates efficient bandwidth utilization, while QoE results confirm stable user experience. The framework achieves these improvements due to its ability to capture both spatial and temporal dependencies in network states. The study confirms that integrating spatiotemporal learning with reinforcement learning significantly enhances multimedia delivery efficiency. This approach offers a strong foundation for future intelligent communication systems and adaptive streaming architectures.

The suggested reinforcement learning-based framework that combines spatiotemporal dependency successfully integrates the convolution and recurrent feature extraction with Deep Q-Networks learning to dynamically adjust multimedia quality in variable networking conditions. Experimental results indicate that gains have been achieved for all tested parameters such as PSNR, SSIM, latency, throughput, and QoE, validating the significance of taking into account both spatial and temporal dependencies.

Future research should be directed at adapting the presented spatiotemporal reinforcement learning framework to multi-agent environments in which there exist several adaptive streaming clients, necessitating either cooperative or competitive policy learning for optimal global network resources management. The use of the latest video coding algorithms like VVC along with neural coding can yield improved rate-distortion performance in the proposed RL framework. The study of transformers as an alternative spatiotemporal representation learning architecture instead of using CNN and RNN can prove effective in capturing longer-term dependencies. Optimization at cross-layer level by tuning transport layer (like QUIC) and application layer using the policy learned from DQN could be a promising approach for future exploration. Testing and validation in real-life OTT streaming environments, keeping in view mobile users and heterogeneous devices, will help bridge the gap between simulated experiments and realistic environments. Utilizing perceptual quality measures like VMAF instead of PSNR/SSIM as the reward function will facilitate more human-centric adaptive streaming systems.

REFERENCES

[1] V. Charvillat and R. Grigoraş, "Reinforcement Learning for Dynamic Multimedia Adaptation", *Journal of Network and Computer Applications*, Vol. 30, No. 3, pp. 1034-1058, 2007.

[2] G. Dhiman, A.V. Kumar, R. Nirmalan, S. Sujitha, K. Srihari and R.A. Raja, "Multi-Modal Active Learning with Deep Reinforcement Learning for Target Feature Extraction in Multi-Media Image Processing Applications", *Multimedia Tools and Applications*, Vol. 82, No. 4, pp. 5343-5367, 2023.

[3] H. Qin and B. Qin, "Reinforcement Learning-based Multimodal Art Element Extraction and Dynamic Adaptation Strategy for Environmental Designs", *Discover Artificial Intelligence*, Vol. 6, No. 1, pp. 1-19, 2026.

[4] N. Mastronarde and M. Van Der Schaar, "Online Reinforcement Learning for Dynamic Multimedia Systems", *IEEE Transactions on Image Processing*, Vol. 19, No. 2, pp. 290-305, 2009.

[5] A.D. Río, J. Serrano, D. Jimenez, L.M. Contreras and F. Alvarez, "A Deep Reinforcement Learning Quality Optimization Framework for Multimedia Streaming over 5G Networks", *Applied Sciences*, Vol. 12, No. 20, pp. 1-12, 2022.

[6] Z. Ji, C. Hu, X. Jia and Y. Chen, "Research on Dynamic Optimization Strategy for Cross-platform Video Transmission Quality based on Deep Learning", *Artificial Intelligence and Machine Learning Review*, Vol. 5, No. 4, pp. 69-82, 2024.

[7] 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, pp. 1-20, 2022.

[8] Z. Wu, S. Wang, C. Ni and J. Wu, "Adaptive Traffic Signal Timing Optimization using Deep Reinforcement Learning in Urban Networks", *Artificial Intelligence and Machine Learning Review*, Vol. 5, No. 4, pp. 55-68, 2024.

[9] N.A. Hafez, M.S. Hassan and T. Landolsi, "Reinforcement Learning-based Rate Adaptation in Dynamic Video Streaming", *Telecommunication Systems*, Vol. 83, No. 4, pp. 395-407, 2023.

[10] K. Srihari, G. Dhiman, K. Somasundaram, A. Sharma, S.M.G. Rajeskannan and M. Masud, "Nature-Inspired-based Approach for Automated Cyberbullying Classification on

Multimedia Social Networking", *Mathematical Problems in Engineering*, Vol. 2021, No. 1, pp. 1-17, 2021.

[11] M.D. Choudhry, S. Munusamy and S.K. Nachimuthu, "Data Science for Industry 5.0: Approaches, Challenges, and Applications", Proceedings of International Conference on Next Generation Data Science and Blockchain Technology for Industry 5.0: Concepts and Paradigms, pp. 91-126, 2025.

[12] A. Bentaleb, A.C. Begen and R. Zimmermann, "ORL-SDN: Online Reinforcement Learning for SDN-Enabled HTTP Adaptive Streaming", *ACM Transactions on Multimedia Computing, Communications and Applications*, Vol. 14, No. 3, pp. 1-28, 2018.

[13] W. Wang, X. Wei, W. Tao, M. Zhou and C. Ji, "Quality of Experience-Oriented Cloud-Edge Dynamic Adaptive Streaming: Recent Advances, Challenges and Opportunities", *Symmetry*, Vol. 17, No. 2, pp. 1-25, 2025.

[14] Z. Zhou, H. Yin, B. Ren, X. Zhong, Z. Qin and Q. Qian, "Spatiotemporal Attention-Based Deep Reinforcement Learning for Joint Routing and Resource Management in Wireless Ad-Hoc Network", *IEEE Transactions on Green Communications and Networking*, Vol. 76, pp. 1-7, 2026.

[15] T. Liu, Q. Meng, J.J. Huang, A. Vlontzos, D. Rueckert and B. Kainz, "Video Summarization through Reinforcement Learning with a 3D Spatio-Temporal U-Net", *IEEE Transactions on Image Processing*, Vol. 31, pp. 1573-1586, 2022.