Deep Reinforcement Learning for Adaptive Bitrate Streaming over Wireless Networks: Advances, Challenges, and Open-Source Simulation

Thanh Thien-An Dang*, Anh-Tien Tran*, Nhu-Ngoc Dao†, and Sungrae Cho*

*School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea.
†Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea.
Email: {attdang, attran}@uclab.re.kr, nndao@sejong.ac.kr, srcho@cau.ac.kr.

***Abstract*—Adaptive Bitrate (ABR) streaming is the de facto mechanism for delivering video over heterogeneous and time-varying wireless networks. However, volatile bandwidth, interference, mobility, and resource contention complicate bitrate selection and buffer control, often degrading users' Quality of Experience (QoE). Traditional rule-based ABR algorithms struggle to generalize across diverse conditions. Recent work demonstrates that Deep Reinforcement Learning (DRL) can jointly reason over network dynamics and player status to optimize QoE, including quality, rebuffering, smoothness, and latency. This paper surveys DRL-based ABR advances from 2022 to 2025 in wireless contexts and emphasizes the role of open and reproducible research using ns-3 for network simulation and FFmpeg for video processing and objective quality assessment (e.g., PSNR, SSIM, VMAF). We first formalize the DRL problem for ABR (state, action, and reward design), then discuss how ns-3 models cellular and Wi-Fi environments and how FFmpeg enables practical bitrate ladders, decoding, trace generation, and automated quality evaluation. We synthesize findings across recent works on access-point-assisted ABR, meta-RL bitrate guidance, cross-layer optimization for real-time communication, fairness-aware multiuser bandwidth allocation at the mobile edge, and throughput prediction for ABR. We highlight reported gains, modeling choices, and, critically, open-source availability where indicated. We conclude by outlining open challenges—generalization, QoE modeling, and real-world deployment—and advocating for open-source benchmarks, datasets, and truly end-to-end ns-3+FFmpeg pipelines to accelerate impactful and reproducible research in DRL-driven ABR for wireless video streaming.**

***Index Terms*—Adaptive Bitrate Streaming, Deep Reinforcement Learning, ns-3, FFmpeg, Quality of Experience, Wireless Communication**

I. Introduction

Video traffic dominates mobile network usage and continues to surge in bitrate and interactivity requirements. Delivering high Quality of Experience (QoE) in fluctuating wireless environments is challenging due to fast timescale capacity variations, congestion, contention, handovers, and device heterogeneity. Adaptive Bitrate (ABR) streaming mitigates these variations by selecting per-segment quality based on observed network and player conditions.

While classical heuristic ABR approaches (e.g., throughput- or buffer-based rules) are simple and widely deployed, they often fail under previously unseen conditions and cannot jointly optimize multiple QoE objectives. Recent advances indicate that Deep Reinforcement Learning (DRL) can learn robust bitrate control policies by observing throughput, buffer, client status, and content features to directly optimize QoE [1]–[3]. Wireless-specific constraints such as high-density Wi-Fi, cellular mobility, and shared spectrum, further motivate learning-based control integrated with the network and edge [4], [5].

This paper surveys DRL-based ABR advances from 2022 to 2025 with an emphasis on open and reproducible research using ns-3 for network simulation and FFmpeg for video processing. We: (i) formalize the DRL problem for ABR; (ii) explain how ns-3 and FFmpeg underpin practical, reproducible experimentation; and (iii) synthesize findings from recent literature covering access-point-assisted ABR, meta-RL server guidance, cross-layer control, fairness-aware multiuser allocation, and throughput prediction, with attention to code availability.

The rest of the paper is organized as follows. Section II introduces DRL for ABR. Section III motivates open-source simulation with ns-3 and FFmpeg. Section IV reviews recent advances. Section V discusses challenges and future directions. Section VI concludes.

II. Related Work

Several recent works apply machine learning and DRL to ABR control, and surveys of learning-based adaptive streaming continue to evolve [6]. In addition to bitrate selection, complementary lines of work (e.g., throughput prediction and edge-assisted scheduling) provide inputs or constraints to ABR controllers [5], [7], [8].

Meta reinforcement learning and on-policy/off-policy DRL have been used to guide bitrate choices in various reinforcement learning approaches. Server-side meta-RL guidance (e.g., Ahaggar) leverages Common Media Client/Server Data (CMCD/SD) to provide bitrate hints to heterogeneous clients [9], [10]. Value-based methods such as Deep Q-Network (DQN) [11] and its variants improve decision stability and sample efficiency for ABR [1], [2]. Actor-critic methods, par-

ticularly soft actor-critic (SAC), are attractive when optimizing multi-objective QoE or fairness [4].

In the context of 5G and edge computing solutions, edge-assisted strategies coordinate bandwidth allocation and ABR, either through prediction-guided prefetching at the mobile edge computing (MEC) [12] or joint optimization of QoE and fairness across users [4], [5]. Related quality-aware streaming and scheduling in device-to-device and wireless caching systems have also been explored [13], [14]. Cross-layer RL aligns transport and codec control for real-time video communication (RTVC), reducing stalls and delay [15]. Infrastructure-assisted vehicular streaming over mmWave further leverages DRL for proactive delivery and bitrate decisions [16].

When comparing existing approaches and identifying gaps, throughput prediction methods target cellular networks [7], [8], [17], yet production deployment faces issues such as data sampling, distribution shift, and on-device constraints. Many works do not publicly release complete code and simulation scripts, limiting reproducibility and fair comparison.

III. Deep Reinforcement Learning for ABR

We outline the DRL formulation commonly used in ABR.

A. State Space

Typical observations include: recent measured/estimated throughput; download time of prior segments; player buffer occupancy and its trend; history of selected bitrates; rebuffering events; viewport/device resolution; and content features (e.g., complexity proxies) [1]–[3], [9]. Wireless-aware contexts further add PHY/MAC indicators (e.g., RSSI/SINR, modulation/coding), number of contending clients, and AP scheduling hints [3]. Specifically, these include signal-to-noise ratio (SNR) for channel quality assessment, airtime utilization (AU) representing the percentage of time the wireless channel is busy, and modulation and coding scheme (MCS) index that determines transmission parameters. Some works treat ABR as a partially observable Markov decision process (POMDP) due to limited observability (e.g., per-client local views in multiuser settings) [9].

B. Action Space

The agent typically selects the next segment's bitrate (from the available ladder). Extended actions include adjusting buffer targets, switching aggressiveness, or requesting server-side bitrate guidance where supported [9], [15].

C. Reward Function

QoE-oriented rewards trade off: (i) delivered quality (via bitrate or perceptual metrics like VMAF/PSNR [18]); (ii) rebuffering penalties; (iii) smoothness penalties for large quality steps; and (iv) latency, especially for live/RTVC settings [1], [2], [15]. Weights reflect service goals. Multiuser settings may incorporate fairness (e.g., Jain's index) [4].

D. Algorithms

Recent ABR works span value-based methods, actor–critic policy-gradient approaches, and meta-RL guidance. On the value-based side, DQN and the DQNReg [19] variant learn discrete bitrate choices from QoE-shaped rewards [2], and, similarly, AP-assisted Wi-DASH employs a DQN model to exploit global network status for server-side bitrate decisions [3]. In contrast, actor–critic methods optimize multi-objective QoE and handle continuous controls: Palette uses asynchronous advantage actor-critic (A3C) [20] to couple transport and encoder parameters for real-time communication [15], and MEC-side bandwidth allocation leverages SAC to improve both QoE and fairness across users [4]. Meanwhile, meta-RL shifts learning to the server: Ahaggar provides bitrate guidance using an advantage Actor-Critic (A2C) backbone with distributed proximal policy optimization (DPPO) [21], [22] updates and Model Agnostic Meta-Learning (MAML)-style [23] fast adaptation, interoperating with heterogeneous clients via CMCD/SD [9], [10]. Additionally, Pensieve-inspired DRL remains a strong baseline in 5G/UHD settings [24], and long short-term memory network (LSTM)-based [25] DRL variants explicitly moderate quality switches [1]. Overall, value-based methods are simple and sample-efficient, actor–critic approaches better navigate QoE trade-offs and continuous actions, and meta-RL offloads computation and generalization to the server; moreover, cross-layer agents reduce stalls and latency by jointly tuning transport and compression.

E. State-Action-Reward Design Patterns

Table I presents concrete examples of how the theoretical DRL framework translates into practical implementations, comparing state space, action space, and reward function designs across different algorithmic approaches. The table shows diverse state representations (from basic network metrics to wireless-aware indicators like SNR/AU/MCS), action spaces (bitrate selection to cross-layer controls), and reward functions (balancing QoE objectives through perceptual metrics or operational constraints). This systematic analysis helps researchers understand the trade-offs between different design choices and provides a foundation for developing new DRL-based ABR algorithms.

It is worth noting that several works in Table II are excluded as they employ non-DRL approaches (throughput prediction, supervised caching/prefetch) rather than RL control algorithms.

IV. The Role of Open-Source Simulation with ns-3 and FFmpeg

Simulation enables safe, controllable, and repeatable evaluation of ABR under diverse wireless conditions.

A. ns-3 for Network Simulation

ns-3 offers detailed models for Wi-Fi and cellular (LTE/NR) stacks, mobility, interference, and application traffic, allowing researchers to configure challenging scenarios (variable RSSI/SINR, handovers, multi-AP contention, user mobility,

TABLE I
DRL FORMULATIONS BY PAPER: STATE/OBSERVATION, ACTION, AND REWARD.

Paper	Algorithm	State	Action	Reward
Xiao et al. [4]	SAC MEC	Buffer T^{buff}, bits, bitrate, multiuser	Allocate bandwidth	$\sum \log(\text{QoE})$ + fairness
Bentaleb et al. [10]	Meta-RL	Throughput, buffer, quality, device, content	Select bitrate	VMAF + rebuffer + switch penalties
Wu et al. [3]	AP-DQN	SNR/AU/MCS, throughput, buffer, prefs	Select bitrate	QoE + rebuffer/switch penalties
Li et al. [15]	Cross-layer RL	RTT, stalling, bandwidth, complexity	CRF + pacing	Quality + low stalling/RTT
Hafez et al. [2]	DQNReg	Bitrate, throughput, buffer, time, available	Select bitrate	QoE - rebuffer - switch penalties
Bentaleb et al. [9]	Meta-RL	Throughput, buffer, resolution, content, POMDP	Select bitrate	VMAF + rebuffer + switch penalties
Arunruangsirilert et al. [24]	Pensieve-5G	Throughput, buffer, chunk sizes, 5G/UHD	Select bitrate	QoE: quality + rebuffer + smoothness
Souane et al. [1]	DRL ABR	Quality, bandwidth, buffer, classes	Select quality	$f(q, \text{rb}, \Delta q)$ + penalties

and background traffic). ABR experiments typically emulate HTTP-based segment downloads over TCP/QUIC, capturing segment-level throughput and latency consistent with player behavior. Prior work used ns-3-based platforms to explore multiuser sharing and rate control effects for video streaming [26]. Beyond platform studies, ns-3/mmWave is also used to synthesize 5G streaming traces for throughput-prediction pipelines; for example, Sen et al. generate 5G mmWave traces with DASH traffic in ns-3/mmWave to evaluate their predictor and downstream ABR QoE [17].

B. FFmpeg for Video Processing

FFmpeg is widely used to prepare and analyze video content for ABR: encode and decode a bitrate ladder across resolutions/bitrates with consistent GOP structure; segment into HLS/DASH chunks; and optionally generate per-segment size/bitrate traces used to emulate downloads in ns-3. In addition, FFmpeg can compute objective quality metrics such as PSNR, SSIM, and VMAF either inline through built-in filters or offline, enabling both evaluation and reward shaping.

C. End-to-End Pipeline Architecture

An open-source workflow integrates: (i) FFmpeg to produce representations, perform decoding checks, and compute objective quality metrics (e.g., PSNR, SSIM, VMAF) in addition to emitting segment metadata; (ii) a trace-driven ABR client generating segment requests; and (iii) ns-3 scenarios for Wi-Fi/cellular with realistic channel and mobility models. Together, ns-3 and FFmpeg form a truly end-to-end simulation and evaluation pipeline. We advocate releasing configuration files, content preparation scripts, and ABR agent code for community reuse.

D. ns-3 and Gymnasium Integration

A practical way to make the simulation loop RL-ready is to expose the ns-3 experiment as an environment compliant with the OpenAI Gym/Gymnasium API. The ns3-gym framework maps simulator state and control to observation/action spaces and implements the standard `reset`/`step` interaction, letting an external agent train and evaluate over ns-3 with minimal glue code [27]. Adopting the modern Gymnasium interface [28] ensures consistent semantics (termination/truncation, seeding, vectorized rollouts) and interoperability with common RL libraries. For ABR, the wrapper typically publishes observations (e.g., recent throughput, round-trip time (RTT), buffer level, selected bitrate history, optional content features) and exposes actions as bitrate (and optionally buffer target or switching aggressiveness); the reward encodes QoE (quality, rebuffering, smoothness, latency). This integration decouples environment mechanics from learning, so researchers can focus solely on designing and comparing optimization algorithms while reusing the same ns-3 scenarios, content, logging, and evaluation harness.

E. Ready-to-Use Datasets and Content Suites

Ready-to-use datasets substantially accelerate research by eliminating the heavy upfront effort of content preparation and enabling fair, apples-to-apples comparisons across studies. The multi-profile UHD DASH datasets and scripts by Quinlan and Sreenan provide 4K AVC/HEVC content, bitrate ladders, and generation tooling suitable for both real-time testbeds and trace-based simulation (including ns-3) [29], and the UVG dataset offers diverse 4K sequences with varied motion and texture characteristics that are valuable for codec and ABR analysis [30]. Because video scene types (e.g., static/low-motion, animation, interview/talking-head, action/sports, nature) materially affect segment sizes, achievable bitrate, and QoE under ABR/DRL optimization, benchmarking suites should deliberately span multiple scene categories to avoid overfitting to a single content type. Their availability reduces barrier to entry for new ABR researchers, standardizes evaluation across resolutions/bitrates, and supports reproducible QoE assessment. When coupled with ns-3 scenarios and FFmpeg-based traces, such datasets allow rapid ablation studies, robust hyperparameter sweeps, and cross-paper benchmarking without confounding differences in content encoding pipelines.

V. REVIEW OF RECENT ADVANCES

We synthesize common trends and key differentiators across recent work that targets ABR for wireless video. Most papers cast bitrate control or its enablers as a learning problem, but they differ by control location (client vs. AP/edge/server), action granularity (bitrate vs. cross-layer knobs vs. bandwidth allocation), and evaluation methodology (trace-driven, ns-3, testbed, field).

A. Commonalities and Differences

Recent approaches exhibit diverse learning paradigms and control scopes. Client-side DRL for bitrate selection remains a central theme [1], [2], complemented by server-side meta-RL that provides guidance to heterogeneous clients [9], [10]. Some works expand the control scope through cross-layer

RL, extending actions beyond bitrate to include encoder and transport parameters [15], while edge-side RL addresses the complexities of multiuser resource allocation and fairness [4]. The state representations adapt accordingly: while most models use throughput history, buffer level, and bitrate history, wireless-aware solutions incorporate PHY/MAC indicators or AP-side congestion signals [3]. Similarly, multiuser and edge schemes add per-user buffer, demand, and fairness signals [4], and server-guidance systems leverage CMCD/CMCD-SD telemetry and content descriptors [10].

The objectives and evaluation methodologies also vary. QoE is commonly encoded as a reward function balancing delivered quality against penalties for rebuffering and quality switching, with live settings adding latency as a key factor [15]. In multiuser scenarios, fairness is often incorporated, for instance, through Jain's index [4]. Evaluation practices are equally diverse, ranging from trace-driven simulations to more complex ns-3 environments and real-world systems. Throughput prediction studies, for example, often target cellular traces and consider on-device constraints [7], [8], [17], while other works explore complementary techniques like edge caching and prefetching to support ABR control [5]. Furthermore, adaptations of established frameworks like Pensieve for 5G and UHD content highlight the importance of using practical QoE metrics [24], and survey papers play a crucial role in consolidating these varied taxonomies and toolchains [6].

B. Representative Works by Theme

Representative works span AP/edge-assisted ABR, meta-RL guidance, client-side DRL, cross-layer RL, throughput prediction, caching/prefetch, and 5G/UHD adaptations. AP-assisted DRL for dense Wi-Fi integrates network indicators with player status to stabilize bitrate, while MEC-side SAC allocates bandwidth to improve QoE and fairness under multiuser contention [3], [4]. Server-side meta-RL (Ahaggar) issues bitrate hints via CMCD/SD, enabling rapid generalization across devices and ABR clients [9], [10]. On the client, value-based (DQN) and actor-critic variants learn bitrate policies with QoE-shaped rewards, and constraining quality steps improves smoothness [1], [2]. Cross-layer control for real-time video communication jointly tunes encoder compression and transport pacing to reduce stalls and delay [15]. Throughput-prediction enablers — including multistage context-aware models and device-based (including federated) predictors — help stabilize ABR in cellular networks [7], [8], [17]. In particular, Sen et al. [17] develop a collaborative multi-device, multi-network predictor (FedPut) and report QoE and its components (average bitrate, bitrate variation, rebuffering) as downstream outcomes when used with ABR; they do not directly optimize QoE, hence only the Throughput column is checked in Table II. At the edge, supervised prefetching of likely next representations increases cache hits and reduces backhaul [5]. Practical adaptations for 5G/UHD (e.g., Pensieve-5G) report QoE gains in 5G Standalone (SA) and 5G New Radio (NR)-NR Dual Connectivity (NR-DC) networks [24]. Finally, ns-3 platforms facilitate controlled evaluation [17], [26].

C. Systematic Comparison of Recent Works

Building on the general trends identified above, we now provide a systematic comparison of recent works across multiple dimensions. The Table II and analysis offer a structured evaluation of how different approaches address key aspects of ABR optimization, from throughput prediction to fairness considerations.

The "Throughput" column in Table II indicates whether a paper explicitly builds and evaluates a network throughput predictor. Entries marked with ✓ include such a predictor; entries marked with × do not. Clarifications for some × entries: Arunruangsirilert et al. [24] train and test Pensieve-5G using measured 5G throughput traces; the ABR algorithm is RL-based and does not contain a predictor. Bentaleb et al. [9] (Ahaggar) consume measured throughput as input for guidance without learning a throughput model. Behravesh et al. [5] predict segment bitrates for MEC prefetching (content-side), not network throughput.

Beyond throughput prediction, the added columns highlight which QoE/Quality of Service (QoS) dimensions each work explicitly optimizes or evaluates. In terms of quality, several papers compute or target perceptual measures: Bentaleb et al. [9], [10] report VMAF, and Li et al. [15] improves perceived quality by jointly tuning encoder parameters (e.g., CRF/QP) with transport control. Xiao et al. [4] incorporate bitrate level into a QoE model that drives bandwidth allocation, while Nolan et al. [8] use bitrate utility as part of a QoE formulation when assessing prediction impact. Raca et al. [7] optimize device-based throughput prediction and report video bitrate, quality switches, and stall metrics as downstream outcomes when the predictor is used by an ABR client; they do not directly optimize QoE, hence only the Throughput column is checked in Table II. Similarly, Sen et al. [17] evaluate QoE and components (average bitrate, bitrate variation, rebuffering) when their FedPut predictor drives ABR, but do not optimize QoE directly; thus only the Throughput column is checked for this paper.

Rebuffering and stall avoidance are central across DRL-based ABR works: Hafez et al. [2] and Souane et al. [1] include explicit penalties for interruptions, Bentaleb et al. [10] measure total rebuffering duration, and Pensieve-5G [24] targets smoother UHD playback under 5G variability. AP/edge-assisted solutions such as Wu et al. [3] also account for stall sensitivity in multi-client settings.

Smoothness (quality switching) is explicitly addressed in Hafez et al. [2] and Souane et al. [1] via penalties on large quality steps, and appears as a switching penalty in Nolan et al. [8] when quantifying QoE improvements from better prediction.

Latency is emphasized in real-time and edge-caching contexts. Li et al. [15] directly minimizes interaction delay (RTT) alongside stall rate in RTVC scenarios, and Behravesh et al. [5] reduce segment access delay through MEC prefetching and caching, improving user-perceived responsiveness and backhaul efficiency.

TABLE II
COMPARISON OF RECENT WORKS: PAPER, SETTING, APPROACH, QUALITY (BITRATE/PERCEPTUAL), REBUFFERING, SMOOTHNESS (QUALITY SWITCHES), LATENCY/RTT, AND MULTIUSER FAIRNESS.

Paper	Year	Setting	Approach	Throughput	Quality	Rebuffer	Smooth.	Latency	Fairness
Xiao et al. [4]	2025	MEC multiuser	SAC allocation	×	✓	✓			✓
Bentaleb et al. [10]	2024	Mixed ABR ecosystems	DPPO + MAML	×	✓	✓			
Nolan et al. [8]	2024	Cellular	Multistage DL predictor (non-DRL)	✓	✓	✓	✓		
Wu et al. [3]	2024	Dense Wi-Fi	AP-assisted DRL	×	✓	✓	✓		
Raca et al. [7]	2024	Smartphones	On-device prediction (non-DRL)	✓					
Li et al. [15]	2024	RTVC	Cross-layer RL (Palette)	×	✓	✓		✓	
Sen et al. [17]	2023	Edge, multi-device	Federated prediction (FedPut) (non-DRL)	✓					
Hafez et al. [2]	2023	Wi-Fi/5G traces	DQNReg (value-based)	×	✓	✓	✓		
Bentaleb et al. [9]	2023	Heterogeneous clients	Meta-RL guidance (Ahaggar)	×	✓	✓			
Arunruangsirilert et al. [24]	2023	5G SA/NR-DC, UHD	Pensieve-5G	×	✓	✓			
Souane et al. [1]	2023	Wi-Fi/cellular traces	DRL for ABR	×	✓	✓	✓		
Behravesh et al. [5]	2022	MEC caching	Supervised prefetch (non-DRL)	×				✓	

Finally, fairness is explicitly modeled by Xiao et al. [4], who maximize the sum of log(QoE) across users and couple ABR selection with resource allocation to maintain stable buffers and equitable QoE in multiuser MEC deployments. For Li et al. [15], fairness is reported (Jain index) as a measurement in multi-flow experiments rather than as a direct optimization target; thus the Fairness column remains unchecked.

As a brief conclusion to this section, recent advances indicate that learning-based ABR now spans client, AP/edge, and server: client-side DRL improves QoE and is strengthened by network-side cues and server guidance; explicitly modeling fairness and latency alongside quality and rebuffering yields policies better suited to multiuser and real-time video communication; yet methodological rigor remains uneven, with few works releasing end-to-end artifacts (ns-3 scenarios, content pipelines, and agents), limiting reproducibility and comparability—underscoring calls for open benchmarks and shared tooling [6], [17], [26].

VI. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Several key challenges and promising research directions remain for DRL-based ABR in wireless networks. A primary concern is the generalization and robustness of learned policies; those trained on specific traces or environments may not perform well on unseen networks, devices, or content. To address this, domain randomization, meta-learning, and uncertainty-aware decision-making are promising avenues for future work. Another critical area is improving QoE modeling and personalization. Since QoE depends on perceptual quality, context, and individual user preferences, incorporating advanced perceptual metrics (e.g., VMAF) and personalized utility models can better align reward functions with the actual human experience.

Furthermore, wireless ABR is an inherently multiuser problem, making coordination and fairness essential. While joint bandwidth allocation and ABR (e.g., SAC-based allocation) show promise, developing scalable and fair coordination mechanisms for heterogeneous clients remains an open challenge [4]. Practical real-world deployment also presents significant hurdles. Deployments must confront measurement noise, distribution shifts, and limited computational resources on client devices. Lightweight inference models, server-side guidance, and continual learning on-device or at the edge are practical strategies to overcome these limitations [7], [10].

Finally, to accelerate progress and ensure fair, reproducible comparisons, the community must embrace open-source benchmarks and pipelines. We urge researchers to release end-to-end ns-3 scenarios (for both Wi-Fi and cellular), FFmpeg bitrate ladders with segment metadata, standardized content sets, and ABR agents with documented training and evaluation scripts. The availability of ready-to-use content suites, such as the UHD DASH datasets [29], helps make results more comparable and reproducible by fixing the content distribution and bitrate ladder across studies. Establishing common benchmarks is crucial for accelerating progress and ensuring the validity of research findings.

VII. CONCLUSION

DRL presents a powerful framework for ABR control in challenging wireless environments, enabling policies that optimize QoE across quality, smoothness, rebuffering, and latency. Recent advances span AP-assisted control, meta-RL guidance, cross-layer optimization, fairness-aware multiuser allocation, and throughput prediction. Open-source simulation with ns-3 and FFmpeg is critical to reproducibility and impact. We call for community efforts to provide standardized, open pipelines and datasets so that future DRL-based ABR research can be compared rigorously and deployed confidently.

ACKNOWLEDGMENT

This work was supported in part by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) (IITP-2025-RS-2022-00156353, 50%) grant funded by the Korea government (Ministry of Science and ICT) and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00209125).

REFERENCES

[1] N. Souane, M. Bourenane, and Y. Douga, "Deep Reinforcement Learning-Based Approach for Video Streaming: Dynamic Adaptive Video Streaming over HTTP," *Applied Sciences*, vol. 13, no. 21, p. 11697, Jan. 2023, number: 21 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2076-3417/13/21/11697

[2] 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, Aug. 2023. [Online]. Available: https://doi.org/10.1007/s11235-023-01031-3

[3] W. Wu, J. Yuan, S. Ma, and M. Yang, "AP-assisted adaptive video streaming in wireless networks with high-density clients," *Computer Communications*, vol. 219, pp. 53–63, Apr. 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366424000811

[4] A. Xiao, S. Wu, Y. Ou, N. Chen, C. Jiang, and W. Zhang, "QoE-Fairness-Aware Bandwidth Allocation Design for MEC-Assisted ABR Video Transmission," *IEEE Transactions on Network and Service Management*, vol. 22, no. 1, pp. 499–515, Feb. 2025. [Online]. Available: https://ieeexplore.ieee.org/document/10701003

[5] R. Behravesh, A. Rao, D. F. Perez-Ramirez, D. Harutyunyan, R. Riggio, and M. Boman, "Machine Learning at the Mobile Edge: The Case of Dynamic Adaptive Streaming Over HTTP (DASH)," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 4779–4793, Dec. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9841468

[6] H. Amer, M. S. Hassan, and M. H. Ismail, "A Review of Learning-Based Methods for Adaptive Video Streaming Over HTTP," *IEEE Access*, vol. 13, pp. 111 134–111 162, Jun. 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11048852

[7] D. Raca, A. H. Zahran, C. J. Sreenan, R. K. Sinha, E. Halepovic, and V. Gopalakrishnan, "Device-Based Cellular Throughput Prediction for Video Streaming: Lessons From a Real-World Evaluation," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, pp. 318–334, Mar. 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10457536

[8] K. Nolan, D. Raca, G. Provan, and A. Zahran, "MATURE: Multistage Throughput Prediction for Adaptive Video Streaming in Cellular Networks," in *Proceedings of the 34th Workshop on Network and Operating System Support for Digital Audio and Video*, ser. NOSSDAV '24. New York, NY, USA: Association for Computing Machinery, Apr. 2024, pp. 15–21. [Online]. Available: https://dl.acm.org/doi/10.1145/3651863.3651878

[9] A. Bentaleb, M. Lim, M. N. Akcay, A. C. Begen, and R. Zimmermann, "Meta Reinforcement Learning for Rate Adaptation," in *IEEE INFOCOM 2023 - IEEE Conference on Computer Communications*, May 2023, pp. 1–10, iSSN: 2641-9874. [Online]. Available: https://ieeexplore.ieee.org/document/10228951

[10] ——, "Bitrate Adaptation and Guidance With Meta Reinforcement Learning," *IEEE Transactions on Mobile Computing*, vol. 23, no. 11, pp. 10 378–10 392, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10470394

[11] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/nature14236

[12] A.-T. Tran, N.-N. Dao, and S. Cho, "Bitrate Adaptation for Video Streaming Services in Edge Caching Systems," *IEEE Access*, vol. 8, pp. 135 844–135 852, 2020, conference Name: IEEE Access. [Online]. Available: https://ieeexplore.ieee.org/document/9146640

[13] J. Kim, G. Caire, and A. F. Molisch, "Quality-Aware Streaming and Scheduling for Device-to-Device Video Delivery," *IEEE/ACM Transactions on Networking*, vol. 24, no. 4, pp. 2319–2331, Aug. 2016, conference Name: IEEE/ACM Transactions on Networking. [Online]. Available: https://ieeexplore.ieee.org/document/7174557

[14] M. Choi, J. Kim, and J. Moon, "Wireless Video Caching and Dynamic Streaming Under Differentiated Quality Requirements," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1245–1257, Jun. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8374957

[15] Y. Li, H. Chen, B. Xu, Z. Zhang, and Z. Ma, "Improving adaptive real-time video communication via cross-layer optimization," *IEEE Transactions on Multimedia*, vol. 26, pp. 5369–5382, 2024, publisher: IEEE. [Online]. Available: https://ieeexplore.ieee.org/document/10316603

[16] W. J. Yun, D. Kwon, M. Choi, J. Kim, G. Caire, and A. F. Molisch, "Quality-Aware Deep Reinforcement Learning for Streaming in Infrastructure-Assisted Connected Vehicles," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2002–2017, Feb. 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9647911

[17] A. Sen, A. Zunaid, S. Chatterjee, B. Palit, and S. Chakraborty, "Revisiting Cellular Throughput Prediction over the Edge: Collaborative Multi-device, Multi-network in-situ Learning," in *Proceedings of the 2023 International Conference on Embedded Wireless Systems and Networks*, ser. EWSN '23. New York, NY, USA: Association for Computing Machinery, Dec. 2023, pp. 72–83, event-place: Rende, Italy. [Online]. Available: https://dl.acm.org/doi/10.5555/3639940.3639950

[18] N.-N. Dao, A.-T. Tran, N. H. Tu, T. T. Thanh, V. N. Q. Bao, and S. Cho, "A Contemporary Survey on Live Video Streaming from a Computation-Driven Perspective," *ACM Comput. Surv.*, vol. 54, no. 10s, pp. 202:1–202:38, Nov. 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/3519552

[19] J. D. Co-Reyes, Y. Miao, D. Peng, E. Real, S. Levine, Q. V. Le, H. Lee, and A. Faust, "Evolving Reinforcement Learning Algorithms," Nov. 2022, arXiv:2101.03958 [cs]. [Online]. Available: http://arxiv.org/abs/2101.03958

[20] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous Methods for Deep Reinforcement Learning," in *Proceedings of The 33rd International Conference on Machine Learning*. PMLR, Jun. 2016, pp. 1928–1937, iSSN: 1938-7228. [Online]. Available: https://proceedings.mlr.press/v48/mniha16.html

[21] N. Heess, D. TB, S. Sriram, J. Lemmon, J. Merel, G. Wayne, Y. Tassa, T. Erez, Z. Wang, S. M. A. Eslami, M. Riedmiller, and D. Silver, "Emergence of Locomotion Behaviours in Rich Environments," Jul. 2017, arXiv:1707.02286 [cs]. [Online]. Available: http://arxiv.org/abs/1707.02286

[22] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," Aug. 2017, arXiv:1707.06347 [cs]. [Online]. Available: http://arxiv.org/abs/1707.06347

[23] C. Finn, P. Abbeel, and S. Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks," in *Proceedings of the 34th International Conference on Machine Learning*. PMLR, Jul. 2017, pp. 1126–1135, iSSN: 2640-3498. [Online]. Available: https://proceedings.mlr.press/v70/finn17a.html

[24] K. Arunruangsirilert, B. Wei, H. Song, and J. Katto, "Pensieve 5G: Implementation of RL-based ABR Algorithm for UHD 4K/8K Content Delivery on Commercial 5G SA/NR-DC Network," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2023, pp. 1–6, iSSN: 1558-2612. [Online]. Available: https://ieeexplore.ieee.org/document/10118834

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997, publisher: MIT press.

[26] G. Liu and L. Kong, "Simulation of Video Streaming Over Wireless Networks with NS-3," Feb. 2023, arXiv:2302.14196 [cs]. [Online]. Available: http://arxiv.org/abs/2302.14196

[27] P. Gawłowicz and A. Zubow, "ns-3 meets OpenAI Gym: The Playground for Machine Learning in Networking Research," in *Proceedings of the 22nd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWIM '19. New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 113–120. [Online]. Available: https://dl.acm.org/doi/10.1145/3345768.3355908

[28] M. Towers, A. Kwiatkowski, J. Terry, J. U. Balis, G. D. Cola, T. Deleu, M. Goulão, A. Kallinteris, M. Krimmel, A. KG, R. Perez-Vicente, A. Pierré, S. Schulhoff, J. J. Tai, H. Tan, and O. G. Younis, "Gymnasium: A Standard Interface for Reinforcement Learning Environments," Nov. 2024, arXiv:2407.17032 [cs]. [Online]. Available: http://arxiv.org/abs/2407.17032

[29] J. J. Quinlan and C. J. Sreenan, "Multi-profile ultra high definition (UHD) AVC and HEVC 4K DASH datasets," in *Proceedings of the 9th ACM Multimedia Systems Conference*, ser. MMSys '18. New York, NY, USA: Association for Computing Machinery, Jun. 2018, pp. 375–380. [Online]. Available: https://dl.acm.org/doi/10.1145/3204949.3208130

[30] A. Mercat, M. Viitanen, and J. Vanne, "UVG dataset: 50/120fps 4K sequences for video codec analysis and development," in *Proceedings of the 11th ACM Multimedia Systems Conference*, ser. MMSys '20. New York, NY, USA: Association for Computing Machinery, May 2020, pp. 297–302. [Online]. Available: https://dl.acm.org/doi/10.1145/3339825.3394937