INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS

ISSN(print): 2643-9840, ISSN(online): 2643-9875

Volume 06 Issue 05 May 2023

DOI: 10.47191/ijmra/v6-i5-60, Impact Factor: 7.022

Page No. 2324-2332

A Taxonomy for Deep Learning in Dynamic Adaptive Video Streaming Over HTTP

Koffka Khan

Department of Computing and Information Technology, the University of the West Indies, St Augustine, Trinidad and Tobago, W.I.

ABSTRACT: Deep Learning (DL) has become a fundamental technology in the field of Dynamic Adaptive Video Streaming over HTTP (DASH), enabling significant advancements in video streaming systems. This taxonomy presents a novel framework for categorizing and organizing the diverse applications and methodologies of DL in DASH. The taxonomy encompasses various aspects of DL, including video representation, quality of experience (QoE) estimation, bitrate adaptation, buffer management, content- and context-aware adaptation, and network optimization. By providing a comprehensive overview of DL in DASH, this taxonomy serves as a valuable resource for researchers and practitioners, facilitating a better understanding of the different DL techniques and their applications in enhancing video streaming performance and user experience.

KEYWORDS: Deep Learning, DASH, QoE, network, taxonomy, streaming

I. INTRODUCTION

Dynamic Adaptive Streaming over HTTP (DASH) is a streaming protocol that allows for the adaptive delivery of multimedia content, such as videos, over HTTP (Hypertext Transfer Protocol) [1), 13), 11)]. It is a standardized technology developed by the Moving Picture Experts Group (MPEG) [18)] and the International Organization for Standardization (ISO) [19)]. DASH works by dividing the multimedia content into small segments and storing them on a server. These segments are then dynamically selected and delivered to the client device based on its current network conditions, device capabilities, and other factors. In client-based DASH the client device can request different segments of the content at varying bit rates or quality levels, depending on the available network bandwidth and device capabilities. This adaptive streaming approach enables a seamless playback experience without interruptions or buffering. In server-based DASH the server makes the decision on segment selection whereas in network-based DASH the network device takes control of the chunk or segment selection decision.

DASH is an industry-standard protocol, which means it is widely adopted by content providers, streaming platforms, and device manufacturers. This standardization ensures interoperability and promotes a consistent streaming experience across different services and devices. DASH allows for real-time adaptation of the streaming quality based on the viewer's network conditions. It dynamically adjusts the bitrate and resolution of the video segments to optimize the viewing experience and avoid buffering or stuttering. In addition, DASH is compatible with a wide range of devices, including smartphones, tablets, smart TVs, and computers [20)]. It supports various operating systems, web browsers, and streaming platforms. Some DASH-based optimizations are designed to scale efficiently, allowing content providers to deliver their multimedia content to a large number of viewers simultaneously. By adapting the streaming quality to each viewer's device and network conditions, DASH optimizes bandwidth usage and minimizes server load.

Optimization in Dynamic Adaptive Streaming over HTTP (DASH) [12), 14)] involves various techniques aimed at improving the streaming experience and resource utilization. DASH employs rate adaptation algorithms [29)] that determine the appropriate bitrate for video segments based on various factors. These algorithms consider the estimated network bandwidth, buffer occupancy, video quality assessment, and playback deadlines. By selecting the optimal bitrate for each segment, the algorithms balance video quality and uninterrupted playback, ensuring an optimal streaming experience. Buffer management techniques optimize the utilization of the playback buffer in DASH [30)]. This involves controlling the buffer size, playout delay, and buffer occupancy to minimize buffering while maintaining high-quality video playback. Efficient buffer management algorithms ensure that the buffer is adequately filled with video segments, reducing the likelihood of interruptions. In addition, pre-fetching techniques [15)] in DASH involve fetching and storing video segments in the buffer ahead of time. By predicting the viewer's behavior and network conditions, pre-fetching helps reduce start-up delay and minimize interruptions during playback. Effective pre-fetching algorithms optimize the delivery of video segments, ensuring a seamless streaming experience.

Optimization in DASH includes monitoring and assessing quality of service (QoS) metrics to ensure high user quality of Experience (QoE) [28)]. These QoS metrics may include video quality (e.g., PSNR, SSIM), rebuffering ratio, start-up delay, smoothness, and stalling events. By analyzing these metrics, streaming systems can adapt and optimize their parameters to deliver an improved user experience. In addition, DASH can employ congestion control mechanisms to adapt to network congestion and prevent network congestion collapse. Congestion control algorithms help manage the allocation of available bandwidth among concurrent video streams and ensure fair resource utilization. These mechanisms prevent excessive buffering or quality degradation during network congestion. Optimization in DASH can also leverage context-aware and content-aware adaptation techniques [25)]. Context-aware adaptation considers factors such as device type, location, time of day, or user preferences to tailor the streaming parameters for individual viewers. Content-aware adaptation takes into account the characteristics of the video content, such as scene complexity or motion intensity, to optimize the streaming decisions. These optimization techniques mentioned above in DASH aim to enhance the streaming experience by dynamically adjusting parameters, adapting to network conditions, and considering viewer preferences. By continuously optimizing the streaming process, DASH ensures smooth playback, minimized buffering, and improved video quality, ultimately delivering an optimal user experience.

Deep learning can be used as a component within an optimization process or to enhance optimization techniques in various domains, including DASH [22), 23)]. In the context of DASH, deep learning can be applied as a tool to improve optimization. For example, deep learning models can be trained to predict network bandwidth, buffer occupancy, or other relevant variables, which can aid in the optimization of bitrate selection and adaptive streaming decisions [3)]. Deep learning can also be used to develop models for video quality assessment, enabling better selection of video segments or bitrate levels during adaptive streaming [8)]. Further, deep learning can help model user behavior and preferences, allowing for personalized adaptation and optimization of the streaming experience [10)]. Furthermore, deep learning can be employed to analyze video content characteristics and assist in content-aware adaptation [7)], optimizing bitrate selection or segment prioritization.

Deep learning models can be trained to assess the quality of video content automatically [24)]. By analyzing video frames or perceptual features, these models can predict the subjective quality of videos. Quality assessment helps in selecting the appropriate bitrate and video segment during adaptive streaming. In addition, deep learning can be used to predict network bandwidth or available capacity. By analyzing historical data or network conditions, deep learning models can estimate future bandwidth variations [5)]. This information is valuable for rate adaptation algorithms in selecting the optimal bitrate for video streaming. Deep learning models can be trained to predict buffer occupancy based on various factors such as network bandwidth, bitrate, segment duration, and viewer behavior [4)]. By forecasting buffer occupancy, the rate adaptation algorithms can make informed decisions to avoid buffering or underutilization of the buffer. Also, deep learning techniques can be used to model user behavior patterns and preferences. By analyzing historical data of user interactions, deep learning models can predict user preferences, such as preferred video quality or viewing duration [9)]. This information can help optimize the streaming experience by adapting the streaming parameters to match user preferences.

Deep learning can be used to incorporate context-awareness into the streaming process [27)]. For example, deep learning models can learn to adapt streaming decisions based on contextual factors like device type, location, time of day, or user demographics. This personalized adaptation improves the user experience by tailoring the streaming parameters to individual viewer characteristics. Furthermore, deep learning techniques can also be utilized to analyze the content characteristics of videos [21)]. Models can learn to identify scene complexity, motion intensity, or other visual features to determine the appropriate bitrate or segment selection during adaptive streaming. Content-aware adaptation helps to allocate resources efficiently and deliver the best possible quality to viewers. These are just a few examples of how deep learning is employed in DASH to optimize various aspects of the streaming process. Deep learning techniques provide powerful tools for analyzing data, learning patterns, and making informed decisions, ultimately improving the streaming experience for viewers.

This paper consists of four sections. In section two the six steps of the methodology is given. The taxonomy is given in section three and the conclusion in section four.

II. METHODOLOGY

In this paper the development of the taxonomy for Deep Learning in Dynamic Adaptive Video Streaming over HTTP involved the following steps:

1. Literature Review: A thorough review of relevant research papers, academic articles, conference proceedings, and industry reports was conducted to gain a comprehensive understanding of the field. This step helped identify the key concepts, techniques, and applications of DL in DASH.
2. Conceptual Framework: A conceptual framework was developed to establish the main categories and subcategories within the taxonomy. This involved identifying the major areas of DL application in DASH and their interrelationships.
3. Iterative Classification: Through an iterative process, different DL applications and methodologies were classified and organized into appropriate categories. This step involved carefully analyzing the literature to ensure that all significant aspects of DL in DASH were captured.
4. Hierarchical Structure: The taxonomy was structured hierarchically to depict the relationships between different categories and subcategories. This structure aimed to provide a clear and logical organization of the various aspects of DL in DASH.
5. Validation and Refinement: The developed taxonomy was reviewed and validated by domain experts in the fields of DL and video streaming. Feedback from experts was collected and incorporated to ensure the taxonomy's accuracy, completeness, and relevance.
6. Iterative Improvement: The taxonomy underwent multiple iterations of refinement based on feedback and discussions with experts. The goal was to achieve a taxonomy that was comprehensive, coherent, and representative of the key aspects of DL in DASH.

The steps employed a combination of literature analysis, expert input, and iterative refinement to develop the taxonomy. The process involved a systematic and rigorous approach to ensure the taxonomy's reliability and usefulness in organizing and understanding the diverse applications of DL in Dynamic Adaptive Video Streaming over HTTP. Each step is explained in detail in the following subsections.

A. Literature Review

The literature review [16)] conducted to develop the taxonomy for Deep Learning in Dynamic Adaptive Video Streaming over HTTP involved an extensive search across various sources, including research papers, academic articles, conference proceedings, and industry reports. The review aimed to identify and analyze relevant studies and advancements in the field. Here are some key findings from the literature review.

The first is DL Models for Video Representation which includes Convolutional Neural Networks, Recurrent Neural Networks and Spatiotemporal Networks. Convolutional Neural Networks (CNNs) have been widely used for extracting spatial features from video frames, enabling accurate content analysis and classification. Recurrent Neural Networks (RNNs) have been employed to capture temporal dependencies in video sequences, facilitating tasks such as action recognition and video captioning. Spatiotemporal Networks, combining CNNs and RNNs, have demonstrated improved performance in modeling both spatial and temporal aspects of videos.

Secondly, is DL-based QoE Estimation. Single-Stream Models have been developed to predict video quality based on the content of a single video stream, utilizing features extracted by DL models. Multi-Stream Models leverage multiple video streams, incorporating quality information from various representations or viewpoints to assess overall video quality accurately. In addition, reinforcement Learning approaches have been applied to optimize video streaming decisions based on QoE metrics, enabling adaptive and personalized video delivery. DL models have been used to predict network conditions, such as bandwidth availability, to optimize video streaming parameters and enhance QoE.

Thirdly, is DL-based Bitrate Adaptation. Reinforcement Learning techniques have been utilized to train DL agents to learn optimal bitrate adaptation policies based on feedback and network conditions. Rate-Distortion Optimization approaches have been employed to optimize video quality while considering the trade-off between bitrate and distortion, leveraging DL models for efficient video encoding and delivery.

Fourthly, is DL-based Buffer Management. Predictive Buffer Models have been developed using DL techniques to forecast future buffer occupancy, enabling proactive adaptation to avoid underflow or overflow situations. Adaptive Play out Mechanisms, driven by DL algorithms, dynamically adjust play out strategies based on buffer conditions, ensuring smooth and uninterrupted video streaming experiences.

Fifthly, is DL for Content- and Context-aware Adaptation. DL models have been applied to capture user preferences and behavior, facilitating personalized video streaming experiences. Context-aware Adaptation leverages DL techniques to adapt video delivery based on contextual information such as device type, network conditions, and user location.

Finally, is DL-based Network Optimization. DL models have been used for traffic prediction, accurately forecasting network traffic patterns to optimize video delivery and resource allocation. Quality-aware Routing techniques employ DL algorithms to optimize video routing decisions based on network conditions and QoE metrics, ensuring efficient and reliable video streaming.

A Taxonomy for Deep Learning in Dynamic Adaptive Video Streaming Over HTTP

The literature review provided a comprehensive understanding of the state-of-the-art techniques and applications of DL in Dynamic Adaptive Video Streaming over HTTP, serving as a foundation for the development of the taxonomy. It enabled the identification of key research areas, methodologies, and their interrelationships, thereby ensuring the taxonomy's relevance and accuracy.

B. Conceptual Framework

The conceptual framework [2)] used to develop the taxonomy for Deep Learning in Dynamic Adaptive Video Streaming over HTTP involved identifying the major categories and subcategories based on the key aspects and applications of DL in the field. The framework aimed to provide a structured organization of the diverse areas of DL in DASH and their interrelationships. Here is the conceptual framework:

DL Models for Video Representation
Frame-level DL Models
Chunk-level DL Models
DL-based QoE Estimation
Video Quality Assessment
QoE Optimization
DL-based Bitrate Adaptation
DL-based Buffer Management
DL for Content- and Context-aware Adaptation
DL-based Network Optimization

The conceptual framework provided a hierarchical structure, with each category representing a significant aspect of DL in DASH. Within each category, the subcategories further delineate the specific techniques and methodologies employed. The framework allows for a systematic and organized classification of the various applications of DL in video streaming, ensuring comprehensive coverage of the field. The interrelationships between the categories highlight the interconnected nature of DL in DASH. For example, DL Models for Video Representation form the foundation for subsequent categories, such as DL-based QoE Estimation, Bitrate Adaptation, and Buffer Management. Additionally, DL-based Network Optimization can leverage insights from QoE Estimation and Bitrate Adaptation to optimize video delivery across networks. By establishing this conceptual framework, the taxonomy provides a structured approach to understanding the diverse applications of DL in Dynamic Adaptive Video Streaming over HTTP, facilitating effective organization, exploration, and analysis of the field.

C. Iterative Classification

The development of the taxonomy for Deep Learning in Dynamic Adaptive Video Streaming over HTTP involved an iterative classification process. This process aimed to categorize and organize the different applications and methodologies of DL in DASH into appropriate categories and subcategories. Here is an overview of the iterative classification approach used.

Firstly, the major categories and subcategories were identified through initial classification based on the initial understanding gained from the literature review. Initial classification was performed by assigning relevant DL applications and methodologies to their corresponding categories. Secondly, the initial classification was reviewed and refined iteratively to ensure accuracy, completeness, and coherence by iterative refinement. Additional literature review was conducted to gather more insights and identify any missed or emerging aspects of DL in DASH. Discussions and consultations with domain experts were held to validate the classification and incorporate their feedback.

Thirdly, by reorganization and restructuring during the refinement process, the organization and structure of the taxonomy were evaluated and adjusted as necessary. Categories and subcategories were reorganized to improve the logical flow and hierarchical relationships within the taxonomy. New subcategories were added or existing ones were merged or split based on the emerging patterns and insights. Fourthly, cross-category connections are made. The iterative classification also involved identifying and establishing connections between different categories and subcategories. Connections were established to represent the interdependencies and relationships between various aspects of DL in DASH. Cross-category connections helped to highlight the interactions and influences between different areas of research and application.

Finally, expert validation involved the refined taxonomy was reviewed and validated by experts in the field of DL and video streaming. Feedback from domain experts was collected and incorporated to ensure the taxonomy's accuracy, relevance, and coverage of the field. The iterative classification approach ensured that the taxonomy evolved and improved through a systematic and iterative refinement process. It allowed for the inclusion of new insights and emerging trends while maintaining a coherent and comprehensive organization of DL in Dynamic Adaptive Video Streaming over HTTP.

D. Hierarchical Structure

The taxonomy for Deep Learning in Dynamic Adaptive Video Streaming over HTTP follows a hierarchical structure, organizing the various categories and subcategories in a logical and hierarchical manner. Here is the hierarchical structure of the taxonomy:

1. DL Models for Video Representation:
   1.1. Frame-level DL Models:
   1.1.1. Convolutional Neural Networks (CNNs)
   1.1.2. Recurrent Neural Networks (RNNs)
   1.1.3. Spatiotemporal Networks
   1.2. Chunk-level DL Models:
   1.2.1. 3D Convolutional Neural Networks (3D CNNs)
   1.2.2. Transformer-based Models
2. DL-based QoE Estimation:
   2.1. Video Quality Assessment:
   2.1.1. Single-Stream Models
   2.1.2. Multi-Stream Models
   2.2. QoE Optimization:
   2.2.1. Reinforcement Learning
   2.2.2. Bandwidth Prediction
3. DL-based Bitrate Adaptation:
   3.1. Reinforcement Learning 3.2. Rate-Distortion Optimization
4. DL-based Buffer Management:
   4.1. Predictive Buffer Models
   4.2. Adaptive Playout Mechanisms
5. DL for Content- and Context-aware Adaptation:
   5.1. User Preference Modeling
   5.2. Context-aware Adaptation
6. DL-based Network Optimization:
   6.1. Traffic Prediction
   6.2. Quality-aware Routing

The hierarchical structure represents the levels of categorization within the taxonomy, where each category represents a major aspect of DL in DASH, and the subcategories provide more specific areas or techniques within those categories. The taxonomy's organization allows for a clear understanding of the different dimensions and applications of DL in Dynamic Adaptive Video Streaming over HTTP, from video representation to QoE estimation, bitrate adaptation, buffer management, content- and context-aware adaptation, and network optimization. The hierarchical structure also signifies the relationships and dependencies between categories. For example, DL Models for Video Representation form the foundation for subsequent categories, such as DL-based QoE Estimation and Bitrate Adaptation. The structure ensures a systematic and coherent organization of the taxonomy, facilitating effective navigation and exploration of the different aspects of DL in DASH.

E. Validation and Refinement

The validation and refinement process [17)] played a crucial role in ensuring the accuracy, completeness, and relevance of the taxonomy for Deep Learning in Dynamic Adaptive Video Streaming over HTTP. Here's an explanation of the validation and refinement steps. Firstly, through expert validation the developed taxonomy was shared with domain experts in the fields of DL, video streaming, and related areas. Experts reviewed the taxonomy to evaluate its structure, coverage, and accuracy in representing the different aspects of DL in DASH. Feedback and suggestions from experts were collected, including any additions, modifications, or clarifications they proposed.

Secondly, through feedback incorporation the feedback received from domain experts was carefully analyzed and incorporated into the taxonomy. Suggestions and recommendations were considered to improve the taxonomy's organization, clarity, and representation of the research field. In cases where multiple experts provided conflicting feedback, discussions were held to reach a consensus or to refine the taxonomy based on the most substantial and relevant inputs. Thirdly, there was iterative refinement. The taxonomy underwent multiple iterations of refinement based on the expert feedback and internal review. During each iteration, the taxonomy's structure, categorization, and hierarchy were evaluated and adjusted to enhance its coherence and completeness. Ambiguous or overlapping areas were clarified, and new subcategories were added or existing ones were merged or split as necessary.

Fourthly, there was the literature review updates where as part of the refinement process, additional literature review was conducted to ensure the taxonomy captured the latest advancements and emerging trends in DL in DASH. New research papers, articles, and relevant sources were reviewed to identify any missed aspects or new subdomains that needed to be incorporated into the taxonomy. Lastly, there was final evaluation. The refined version of the taxonomy was evaluated against the initial objectives and criteria set during its development. The taxonomy was assessed for its coverage of DL applications in DASH, its clarity of categorization, and its ability to provide a comprehensive overview of the field. Any remaining inconsistencies or gaps were addressed, and the final version of the taxonomy was prepared. The validation and refinement process ensured that the taxonomy accurately represented the different aspects and applications of DL in Dynamic Adaptive Video Streaming over HTTP. It leveraged expert insights, feedback, and up-to-date literature review to refine and improve the taxonomy's structure, ensuring its relevance and usefulness to researchers and practitioners in the field.

F. Iterative Improvement

The iterative improvement process [6)] was employed to refine and enhance the taxonomy for Deep Learning in Dynamic Adaptive Video Streaming over HTTP. It involved multiple iterations of assessment, feedback collection, and adjustments to ensure the taxonomy's accuracy, completeness, and effectiveness. Here's an explanation of the iterative improvement steps. In step 1, there is initial taxonomy development. The taxonomy was initially developed based on the literature review and conceptual framework. The major categories and subcategories were identified, and the hierarchical structure was established. In step 2, is the evaluation and feedback collection step. The initial version of the taxonomy was evaluated to identify any gaps, inconsistencies, or areas that needed improvement. Feedback was collected from domain experts, researchers, and practitioners in the field of DL and video streaming. Feedback was also sought from individuals with diverse perspectives to gather comprehensive insights.

Feedback analysis and adjustments is step 3. The feedback collected was carefully analyzed and categorized based on common themes, suggestions, and criticisms. Adjustments were made to the taxonomy based on the feedback, such as adding new subcategories, merging or splitting existing categories, or refining the descriptions. Step 4 is iterative refinement. The taxonomy underwent several iterations of refinement based on the feedback and adjustments made. Each iteration involved reviewing and revising the taxonomy to address the identified issues and incorporate the suggested improvements. The hierarchical structure, category definitions, and subcategory relationships were refined iteratively to enhance the taxonomy's clarity and organization.

In step 5 expert validation and review is performed. The refined version of the taxonomy was shared with domain experts for validation and review. Experts provided additional feedback, suggestions, and insights based on their expertise and knowledge. The taxonomy was adjusted further based on the expert input to ensure its accuracy and alignment with the latest developments in the field. Step 6 is final evaluation and publication. The taxonomy underwent a final evaluation to ensure that it met the desired objectives and criteria. It was reviewed against the initial goals of providing a comprehensive and coherent classification of DL in DASH. The refined taxonomy was prepared for publication or dissemination to the research community. The iterative improvement process allowed for continuous refinement and enhancement of the taxonomy, incorporating diverse perspectives and expert insights. It ensured that the taxonomy evolved through each iteration, addressing any identified limitations and accurately representing the applications and methodologies of DL in Dynamic Adaptive Video Streaming over HTTP.

III. TAXONOMY

In this taxonomy, we propose a novel framework to classify and organize the various applications and methodologies of DL in DASH systems.

1. DL Models for Video Representation:

1.1. Frame-level DL Models:

- Convolutional Neural Networks (CNNs): Used for extracting spatial features from video frames.
- Recurrent Neural Networks (RNNs): Captures temporal dependencies in video sequences.
- Spatiotemporal Networks: Combines CNNs and RNNs to model both spatial and temporal aspects.

1.2. Chunk-level DL Models:

- 3D Convolutional Neural Networks (3D CNNs): Analyzes video chunks to capture spatial and temporal features simultaneously.
- Transformer-based Models: Explores self-attention mechanisms to model relationships between video chunks.

2. DL-based QoE Estimation:

2.1. Video Quality Assessment:

- Single-Stream Models: Predicts video quality based on the content of a single video stream.
- Multi-Stream Models: Incorporates multiple video streams to assess overall video quality.

2.2. QoE Optimization:

- Reinforcement Learning: Utilizes RL algorithms to optimize video streaming decisions based on QoE metrics.
- Bandwidth Prediction: DL models predict network conditions to optimize video streaming parameters.

3. DL-based Bitrate Adaptation:

- Reinforcement Learning: DL agents learn optimal bitrate adaptation policies based on feedback and network conditions.
- Rate-Distortion Optimization: DL models optimize video quality while considering the trade-off between bitrate and distortion.

4. DL-based Buffer Management:

- Predictive Buffer Models: Deep models predict future buffer occupancy to avoid underflow or overflow situations.
- Adaptive Playout Mechanisms: DL algorithms dynamically adjust playout strategies based on buffer conditions.

5. DL for Content- and Context-aware Adaptation:

- User Preference Modeling: DL models capture user preferences to personalize video streaming experiences.
- Context-aware Adaptation: Explores DL techniques to adapt video delivery based on contextual information (e.g., device type, location).

6. DL-based Network Optimization:

- Traffic Prediction: DL models predict network traffic patterns to optimize video delivery.
- Quality-aware Routing: DL algorithms optimize video routing decisions based on network conditions and QoE metrics.

The taxonomy for Deep Learning in Dynamic Adaptive Video Streaming over HTTP provides a structured and comprehensive classification of the various applications and methodologies of deep learning in the context of video streaming. Let's analyze the taxonomy's key characteristics. The first characteristic is hierarchical structure. The taxonomy follows a hierarchical structure, with major categories representing different aspects of deep learning in video streaming, and subcategories providing more specific areas or techniques within those categories. The hierarchical structure facilitates easy navigation and understanding of the taxonomy, enabling researchers and practitioners to explore specific subdomains or broader categories of interest. Coverage of DL applications is the second characteristic. The taxonomy covers various aspects of deep learning in video streaming, including video representation, quality estimation, bitrate adaptation, buffer management, content- and context-aware adaptation, and network optimization. It addresses both frame-level and chunk-level DL models, representing the different approaches for video representation using convolutional neural networks (CNNs), recurrent neural networks (RNNs), spatiotemporal networks, 3D CNNs, and transformer-based models.

The third characteristic is application-specific categories. The taxonomy includes categories dedicated to specific applications, such as QoE estimation, bitrate adaptation, buffer management, content- and context-aware adaptation, and network optimization. These application-specific categories highlight the importance of deep learning techniques in addressing key challenges and improving the performance of dynamic adaptive video streaming over HTTP. The fourth characteristic is interconnections and relationships. The taxonomy establishes connections between categories, representing the interdependencies and relationships between different areas of research and application. For example, DL models for video representation serve as the foundation for subsequent categories such as QoE estimation, bitrate adaptation, and buffer management. The interconnections provide insights into how different aspects of DL in video streaming interact and influence each other, facilitating a holistic understanding of the field.

The fifth characteristic is methodological diversity. The taxonomy encompasses a diverse range of methodologies, including reinforcement learning, rate-distortion optimization, predictive buffer models, adaptive playout mechanisms, user preference modeling, context-aware adaptation, traffic prediction, and quality-aware routing. This methodological diversity reflects the various approaches and techniques employed to leverage deep learning in optimizing video streaming quality, adaptability, and user experience. Incorporation of User Experience is the sixth characteristic. The taxonomy addresses the importance of user experience by including categories such as QoE estimation, content- and context-aware adaptation, and user preference modeling. This highlights the integration of deep learning techniques to understand user preferences, optimize video delivery, and enhance the overall viewing experience.

A Taxonomy for Deep Learning in Dynamic Adaptive Video Streaming Over HTTP

The taxonomy in this paper provides a comprehensive and structured framework for understanding the diverse applications and methodologies of deep learning in Dynamic Adaptive Video Streaming over HTTP. It captures the essential aspects of video representation, quality estimation, adaptation mechanisms, buffer management, content- and context-awareness, and network optimization. The taxonomy's hierarchical structure, interconnections, and methodological diversity contribute to its effectiveness in organizing and exploring the field, making it a valuable resource for researchers and practitioners in this domain.

VI. CONCLUSIONS

This taxonomy provides a comprehensive framework for organizing the various applications of DL in DASH systems. It encompasses DL models for video representation, QoE estimation, bitrate adaptation, buffer management, content- and context-aware adaptation, as well as network optimization. Understanding this taxonomy can guide researchers and practitioners in effectively applying DL techniques to enhance dynamic adaptive video streaming over HTTP.

REFERENCES

1) Ashok Kumar, P. M., Arun Raj, L. N., Jyothi, B., Soliman, N. F., Bajaj, M., & El-Shafai, W. (2022). A Novel Dynamic Bit Rate Analysis Technique for Adaptive Video Streaming over HTTP Support. Sensors, 22(23), 9307.
2) Baía Reis, A., & Ashmore, M. (2022). From video streaming to virtual reality worlds: an academic, reflective, and creative study on live theatre and performance in the metaverse. International Journal of Performance Arts and Digital Media, 18(1), 7-28.
3) Behravesh, R., Rao, A., Perez-Ramirez, D. F., Harutyunyan, D., Riggio, R., & Boman, M. (2022). Machine Learning at the Mobile Edge: The Case of Dynamic Adaptive Streaming over HTTP (DASH). IEEE Transactions on Network and Service Management.
4) Behravesh, R., Rao, A., Perez-Ramirez, D. F., Harutyunyan, D., Riggio, R., & Boman, M. (2022). Machine Learning at the Mobile Edge: The Case of Dynamic Adaptive Streaming over HTTP (DASH). IEEE Transactions on Network and Service Management.
5) Biernacki, A. (2022). Improving Streaming Video with Deep Learning-Based Network Throughput Prediction. Applied Sciences, 12(20), 10274.
6) Cong, J., Zheng, P., Bian, Y., Chen, C. H., Li, J., & Li, X. (2022). A machine learning-based iterative design approach to automate user satisfaction degree prediction in smart product-service system. Computers & Industrial Engineering, 165, 107939.
7) Dao, N. N., Tran, A. T., Tu, N. H., Thanh, T. T., Bao, V. N. Q., & Cho, S. (2022). A contemporary survey on live video streaming from a computation-driven perspective. ACM Computing Surveys, 54(10s), 1-38.
8) de Sousa, N. F. S., Islam, M. T., Mustafa, R. U., Perez, D. A. L., Rothenberg, C. E., & Gomes, P. H. (2022). Machine learningassisted closed-control loops for beyond 5g multi-domain zero-touch networks. Journal of Network and Systems Management, 30(3), 46.
9) Hsu, C. F., Hung, T. H., & Hsu, C. H. (2022). Optimizing immersive video coding configurations using deep learning: A case study on TMIV. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 18(1), 1-25.
10) Huang, T., Zhou, C., Zhang, R. X., Wu, C., & Sun, L. (2022). Learning tailored adaptive bitrate algorithms to heterogeneous network conditions: A domain-specific priors and meta-reinforcement learning approach. IEEE Journal on Selected Areas in Communications, 40(8), 2485-2503.
11) Khan, K., & Goodridge, W. (2018). QoE in DASH. International Journal of Advanced Networking and Applications, 9(4), 3515-3522.
12) Khan, K., & Goodridge, W. (2020). Reinforcement Learning in DASH. International Journal of Advanced Networking and Applications, 11(5), 4386-4392.
13) KHAN, K., & GOODRIDGE, W. (2022). Ultra-HD Video Streaming in 5G Fixed Wireless Access Bottlenecks.
14) Koffka, K., & Wayne, G. (2018). A DASH Survey: the ON-OFF Traffic Problem and Contemporary Solutions. Computer Sciences and Telecommunications, (1), 3-20.
15) Li, Z., Li, J., Wu, Q., Tyson, G., & Xie, G. (2022). A Large-Scale Measurement and Optimization of Mobile Live Streaming Services. IEEE Transactions on Mobile Computing.
16) Lim, W. M., Kumar, S., & Ali, F. (2022). Advancing knowledge through literature reviews:'what','why', and 'how to contribute'. The Service Industries Journal, 42(7-8), 481-513.

17) Lin, L., Di, L., Zhang, C., Guo, L., Di, Y., Li, H., & Yang, A. (2022). Validation and refinement of cropland data layer using a spatial-temporal decision tree algorithm. Scientific Data, 9(1), 63.
18) Liu, T., Xu, M., Li, S., Chen, C., Yang, L., & Lv, Z. (2023, June). Learnt Mutual Feature Compression for Machine Vision. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.
19) Liu, X. (2023). Quality of service for video stream over IP networks.
20) Loseto, G., Scioscia, F., Ruta, M., Gramegna, F., & Bilenchi, I. (2023). Semantic-based Adaptation of Quality of Experience in Web Multimedia Streams.
21) Motaung, W., Ogudo, K. A., & Chabalala, C. (2022, August). Real-time monitoring of video quality in a dash-based digital video broadcasting using deep learning. In 2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD) (pp. 1-6). IEEE.
22) Murad, T., Nguyen, A., & Yan, Z. (2022, October). DAO: Dynamic Adaptive Offloading for Video Analytics. In Proceedings of the 30th ACM International Conference on Multimedia (pp. 3017-3025).
23) Naik, S., Khan, O., Katre, A., & Keskar, A. (2022, May). ARMPC-ARIMA based prediction model for Adaptive Bitrate Scheme in Streaming. In 2022 1st International Conference on the Paradigm Shifts in Communication, Embedded Systems, Machine Learning and Signal Processing (PCEMS) (pp. 107-112). IEEE.
24) Sanborn, K., Richardson, A., & Sprinkle, J. (2022, May). Semantic Tagging of CAN and Dash Camera Data from Naturalistic Drives. In 2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS) (pp. 312-313). IEEE.
25) Seng, J. K. P., Ang, K. L. M., Peter, E., & Mmonyi, A. (2022). Artificial Intelligence (AI) and Machine Learning for Multimedia and Edge Information Processing. Electronics, 11(14), 2239.
26) Sharaf Addin, E. H., Admodisastro, N., Mohd Ashri, S. N. S., Kamaruddin, A., & Chong, Y. C. (2022). Customer mobile behavioral segmentation and analysis in telecom using machine learning. Applied Artificial Intelligence, 36(1), 2009223.
27) Shishkov, B., Ivanova, K., Verbraeck, A., & van Sinderen, M. (2022, December). Combining context-awareness and data analytics in support of drone technology. In Telecommunications and Remote Sensing: 11th International Conference, ICTRS 2022, Sofia, Bulgaria, November 21–22, 2022, Proceedings (pp. 51-60). Cham: Springer Nature Switzerland.
28) Viola, R., Zorrilla, M., Angueira, P., & Montalbán, J. (2022). Multi-access Edge Computing video analytics of ITU-T P. 1203 Quality of Experience for streaming monitoring in dense client cells. Multimedia Tools and Applications, 81(9), 1238712403.
29) Wang, N. (2022). Application of DASH client optimization and artificial intelligence in the management and operation of big data tourism hotels. Alexandria Engineering Journal, 61(1), 81-90.
30) Wei, B., Song, H., Nguyen, Q. N., & Katto, J. (2022, January). DASH Live Video Streaming Control Using Actor-Critic Reinforcement Learning Method. In Mobile Networks and Management: 11th EAI International Conference, MONAMI 2021, Virtual Event, October 27-29, 2021, Proceedings (pp. 17-24). Cham: Springer International Publishing.
31) Zhao, Jia, Jiangchuan Liu, Haiyang Wang, Changqiao Xu, and Hongke Zhang. "Multipath Congestion Control: Measurement, Analysis, and Optimization From the Energy Perspective." IEEE Transactions on Network Science and Engineering (2023).