A Taxonomy for Generative Adversarial Networks in Dynamic Adaptive 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: Generative Adversarial Networks (GANs) have emerged as a powerful tool in the field of Dynamic Adaptive Streaming over HTTP (DASH) to enhance various aspects of video streaming. This paper presents a taxonomy that categorizes the applications and techniques of GANs in the context of DASH. The taxonomy covers several key dimensions, including video generation, compression, quality enhancement, bandwidth adaptation, dynamic bitrate streaming, and cross-modal applications. Within each dimension, specific subcategories are identified to capture the diverse applications of GANs in DASH. Additionally, evaluation metrics for assessing the quality and effectiveness of GAN-based approaches are discussed. The taxonomy serves as a comprehensive framework to understand and organize the different ways in which GANs can be utilized to improve the streaming experience in DASH. By providing an organized structure, this taxonomy facilitates better understanding, comparison, and exploration of GAN-based approaches in DASH and enables researchers and practitioners to identify areas for further research and development.

KEYWORDS: Generative Adversarial Networks, streaming, DASH, video, quality, adaption, bandwidth

I. INTRODUCTION

Dynamic Adaptive Streaming over HTTP (DASH) [9]] has revolutionized the delivery of multimedia content over the internet by providing adaptive bitrate streaming to cater to varying network conditions and user preferences. To further enhance the streaming experience, Generative Adversarial Networks (GANs) [19]] have emerged as a promising approach. GANs, with their ability to generate realistic content and optimize video quality, offer valuable solutions for various aspects of DASH. In this paper, we present a taxonomy that categorizes and organizes the applications and techniques of GANs in the context of DASH. The taxonomy aims to provide a systematic framework for understanding and exploring the different ways in which GANs can be utilized to improve the streaming process. By categorizing the applications and techniques, the taxonomy enables researchers and practitioners to navigate the field and identify specific areas of interest.

The taxonomy encompasses various dimensions of GAN utilization in DASH. It starts with video generation, where GANs are employed to generate synthetic content or interpolate frames to enhance the available video library. Additionally, GANs are explored for video compression, aiming to optimize the bitrate while preserving perceptual quality. Quality enhancement techniques, such as super-resolution, denoising, and deblurring, are also covered, highlighting how GANs can improve the visual quality of streamed videos.

Bandwidth adaptation [8]] and control are crucial for adaptive streaming. GANs can assist in predicting available bandwidth and network conditions to optimize video streaming decisions. Moreover, GAN-based approaches for dynamic adaptation of video segments and chunk-based adaptation strategies are explored. The taxonomy also includes quality assessment and objective metrics, considering perceptual quality assessment and content-adaptive metrics to evaluate the effectiveness of GANbased approaches. Furthermore, cross-modal applications of GANs in DASH, such as audio-visual synchronization and multimodal data generation, are discussed.

By providing a comprehensive taxonomy, this paper offers a structured framework to understand and classify the diverse applications and techniques of GANs in DASH. It enables researchers to identify gaps, explore specific subfields, and develop novel solutions. Additionally, it serves as a valuable resource for practitioners seeking to leverage GANs for optimizing the streaming experience, improving video quality, and enhancing user satisfaction in Dynamic Adaptive Streaming over HTTP. This paper consists of five sections. The key elements of dynamic adaptive streaming over HTTP is outlined in section two. A description of

A Taxonomy for Generative Adversarial Networks in Dynamic Adaptive Streaming Over HTTP

Generative Adversarial Networks (GANs) with explanations of how it works is given in section three. In section four the taxonomy for generative adversarial networks in Dynamic Adaptive streaming over HTTP is given. Section five gives a discussion of the taxonomy. Finally, the conclusion is given in section six.

II. DYNAMIC ADAPTIVE STREAMING OVER HTTP (DASH)

DASH stands for Dynamic Adaptive Streaming over HTTP. It is a popular streaming protocol used for delivering multimedia content, such as videos, over the internet. DASH is designed to provide a high-quality streaming experience [10)] by dynamically adapting the video quality based on the available network conditions.

The key components of DASH are as follows. The Media Presentation Description (MPD) [21)] is an XML file that serves as a manifest or index for the media content. It provides information about the available video representations, their URLs, and other metadata. The client uses the MPD to request and select the appropriate video quality based on network conditions. DASH divides the video content into small segments of fixed duration, typically a few seconds long. Each segment represents a part of the video that can be independently requested and played. Different video representations are available in multiple quality levels, and each quality level is divided into segments. The client dynamically selects the most suitable quality level based on the current network conditions.

The adaptation logic [11)] is responsible for dynamically adjusting the video quality based on the available network bandwidth and client capabilities. The client monitors the network conditions and requests segments at the highest quality level that the network can support without buffering. If the network conditions deteriorate, the client may request lower-quality segments to avoid interruptions and buffering. Media players are responsible for fetching the video segments, decoding them, and rendering them on the playback device. They interact with the DASH server to request the segments and use the MPD to determine the appropriate video quality to request. Media players can be built into web browsers, mobile apps, or dedicated video players.

The DASH server is responsible for storing the video content and serving it to the clients. It receives requests from the media players for specific video segments and responds with the requested segments encoded at the appropriate quality level. The server uses the MPD to generate the necessary URLs for the requested segments. Before delivering the content via DASH, the video content needs to be encoded into different representations or bitrates to support adaptive streaming. The content preparation stage involves creating multiple versions of the video at different quality levels, with each version encoded to match the target bitrate. These representations are then made available on the DASH server, and the client can switch between them based on network conditions. In effect, DASH enables adaptive streaming by dynamically adjusting the video quality based on the network conditions, allowing viewers to enjoy a smooth streaming experience even in varying network conditions.

There are three different components or approaches that can be used to implement the streaming system: client-side, serverside, and in-network DASH [12)]. Let's take a closer look at each one. In client-side DASH, the adaptation logic and video quality selection are performed on the client-side, typically within the media player or streaming application. The client-side player receives the media presentation description (MPD) from the server, which contains information about the available video representations and their URLs. The player monitors the network conditions, such as available bandwidth, and based on that information, it dynamically selects the most suitable video quality from the available representations. The client requests the corresponding video segments from the server and plays them back. Client-side DASH provides flexibility and control to the client, allowing it to adapt the streaming experience based on its own assessment of the network conditions.

In server-side DASH, the adaptation logic is performed on the server-side rather than on the client. The server receives requests from the client for video segments, and based on the client's network conditions, it selects and delivers the appropriate video quality to the client. The server-side implementation can take into account additional factors such as server load, content popularity, and real-time network conditions across multiple clients to make adaptive bitrate decisions. Server-side DASH can offload the adaptation complexity from the client and centralize the decision-making process on the server side, providing more consistent and controlled streaming experiences.

In-network DASH, also known as network-assisted DASH or collaborative DASH, involves incorporating network information and intelligence into the adaptive streaming process. It leverages the capabilities of the network infrastructure, such as the routers or caches, to assist in delivering the content efficiently. In this approach, network devices or proxies actively monitor the network conditions and make decisions on behalf of the client or server. They can detect congestion or bandwidth fluctuations and dynamically adjust the video quality before the video reaches the client, optimizing the streaming experience. In-network DASH requires collaboration between the network infrastructure and the streaming system to exchange relevant information and make informed decisions.

Each approach has its advantages and considerations. Client-side DASH offers flexibility and control to the client device, allowing it to adapt the streaming experience based on its assessment of the network conditions. Server-side DASH centralizes the

adaptation logic on the server-side, enabling better coordination and control over the streaming experience. In-network DASH leverages network intelligence to optimize streaming, but it requires cooperation and support from the network infrastructure. The choice of the DASH implementation approach depends on various factors, including the specific streaming requirements, network capabilities [8)], and desired level of control.

III. GENERATIVE ADVERSARIAL NETWORKS

Generative Adversarial Networks (GANs) are a class of machine learning models that are widely used for generating new data samples that resemble a given dataset. GANs consist of two main components: a generator network and a discriminator network. These networks are trained together in an adversarial setting, where they compete against each other, hence the term "adversarial." We now discuss how GANs work. The generator network takes random input, typically from a low-dimensional noise vector, and generates synthetic data samples, such as images or audio. The generator starts with random noise and gradually learns to generate more realistic samples that resemble the training data. Initially, the generated samples may not resemble the real data, but as training progresses, the generator improves its ability to generate more convincing samples.

The discriminator network acts as a binary classifier. It takes input data samples and tries to distinguish between real samples from the training dataset and fake samples generated by the generator. The discriminator is trained on labeled data, where real samples are labeled as "real" and generated samples are labeled as "fake." The discriminator's objective is to correctly classify the real and fake samples. The generator and discriminator networks are trained in an adversarial manner. The generator aims to generate samples that the discriminator cannot distinguish from real samples, while the discriminator aims to correctly classify the real and fake samples. The networks play a "minimax" game, where the generator tries to minimize the discriminator's ability to correctly classify the samples, and the discriminator tries to maximize its accuracy in distinguishing real and fake samples.

During training, the generator and discriminator networks are updated iteratively. The generator generates fake samples, and the discriminator classifies them. The discriminator's classification results are then used to update both the discriminator and generator networks. This iterative process continues until the generator learns to generate samples that are realistic enough to deceive the discriminator. As the training progresses, the generator improves its ability to generate samples that are increasingly similar to the real data distribution. Ideally, the generator and discriminator reach a point where the generator produces samples that are indistinguishable from real samples, and the discriminator cannot differentiate between real and fake samples with high confidence.

GANs have demonstrated remarkable capabilities in various domains [23)], including image synthesis, text generation, and even video generation. They have been used for tasks like image generation, style transfer, data augmentation, and anomaly detection. GANs have also inspired numerous advancements, such as conditional GANs, progressive GANs, and cycle-consistent GANs, which have further expanded their applications. GANs provide a powerful framework for generating realistic and novel data samples by training two competing neural networks in an adversarial manner.

We now give two ways GANs can support DASH. Before delivering the content via DASH, the video content may go through a content preparation stage, where different quality levels or representations of the video are created. GANs can be used in this process to generate or enhance video content. For example, GANs can be employed to upscale low-resolution videos, enhance image quality, or generate high-quality versions of the content. By using GANs during content preparation, the generated or enhanced video representations can then be made available for adaptive streaming using DASH. GANs can also be used to enhance the perceived quality of video content during the adaptive streaming process. For example, GANs can be employed on the client-side to perform real-time video upscaling or denoising to improve the visual quality of the streamed video. This can be particularly useful when the available network bandwidth is limited and the video quality needs to be enhanced to compensate for the lower bitrate. These GAN-based enhancements can be integrated into the DASH client or media player to improve the viewing experience.

IV. TAXONOMY

Taxonomy for Generative Adversarial Networks (GANs) in Dynamic Adaptive Streaming over HTTP (DASH):
Video Generation:

    a.   *Content Generation [1)]*: GANs used to generate synthetic video content, such as scenes, objects, or characters, to enhance the content library for DASH applications. Content Generation using Generative Adversarial Networks (GANs) has been a breakthrough in the field of computer vision and artificial intelligence. GANs are now being utilized to generate synthetic video content, including scenes, objects, and characters, with the aim of enhancing the content library for Dynamic Adaptive Streaming over HTTP (DASH) applications. DASH is a popular streaming protocol that dynamically adapts the quality of video content based on network conditions and device capabilities. To provide a seamless streaming experience, DASH applications require a diverse

and extensive library of video content. However, creating such content manually can be time-consuming, expensive, and limited in scope.

To address these challenges, researchers and developers have turned to GANs, which are deep learning models consisting of two neural networks: a generator and a discriminator. The generator network learns to generate synthetic video content by capturing patterns and structures from a given dataset, while the discriminator network learns to distinguish between real and synthetic content. In the context of enhancing the content library for DASH applications, GANs can be trained on existing video datasets to generate new synthetic video content. For example, GANs can learn the visual features, dynamics, and spatial relationships of different scenes, objects, or characters from a large collection of videos. Once trained, the GAN can generate novel video content that closely resembles the characteristics of the original dataset.

The generated synthetic video content can then be integrated into the content library of DASH applications. This expands the variety of available scenes, objects, and characters, enabling a more diverse and engaging streaming experience for users. Additionally, the synthetic content can be generated on-demand, allowing for real-time adaptation and personalization based on user preferences, network conditions, or other contextual factors. It's important to note that the quality and realism of the generated content heavily rely on the training data, the architecture of the GAN model, and the optimization techniques used during training. Further advancements in GAN research and techniques, along with improvements in computing power, will likely contribute to the generation of even more realistic and high-quality synthetic video content for DASH applications in the future.

b. *Frame Interpolation [20)]*: GAN architectures designed to generate intermediate frames between existing frames to Improve video smoothness during streaming. Frame interpolation is a technique used to generate intermediate frames between existing frames in a video sequence, with the aim of improving the smoothness and visual quality of the video. Recently, Generative Adversarial Networks (GANs) have been utilized to create advanced architectures specifically designed for frame interpolation tasks during video streaming. In traditional video streaming, the frame rate is fixed, and the video player displays each frame at a constant rate. However, this can sometimes result in jerky or choppy motion, especially when the original video has a low frame rate. Frame interpolation addresses this issue by generating additional frames that fill in the gaps between the existing frames, resulting in a higher frame rate and smoother motion.

To achieve frame interpolation, GAN architectures are trained using pairs of consecutive frames from a video dataset. The architecture typically consists of a generator network and a discriminator network. The generator network is responsible for generating the intermediate frames, while the discriminator network evaluates the realism and quality of the generated frames. During training, the generator network learns to generate visually coherent and smooth intermediate frames by capturing the temporal dependencies and motion patterns present in the training data. The discriminator network provides feedback to the generator network, helping it improve its frame generation capabilities over time. This adversarial training process allows the GAN to generate high-quality intermediate frames that seamlessly blend with the original frames.

When integrated into video streaming applications, GAN-based frame interpolation can significantly enhance the viewing experience by providing smoother and more visually appealing videos. The interpolated frames fill the temporal gaps, reducing judder and improving the perception of motion. This is particularly useful in scenarios where the original video has a low frame rate or when streaming over networks with limited bandwidth. It's worth noting that the success of GAN-based frame interpolation depends on factors such as the complexity of the video content, the training dataset's diversity, the architecture of the GAN model, and the optimization techniques employed during training. Ongoing research in GANs and frame interpolation techniques continues to refine and improve the performance of these models, leading to even better video smoothness during streaming in the future.

Video Compression and Quality Enhancement:

a. *Video Compression [18)]*: GANs utilized to optimize video compression algorithms, reducing the required bitrate while maintaining perceptual video quality. Video compression is a crucial technology used to efficiently store and transmit video content by reducing its file size while preserving its quality. Generative Adversarial Networks (GANs) have emerged as a promising tool in optimizing video compression algorithms to achieve higher compression ratios without sacrificing perceptual video quality. Traditional video compression algorithms, such as those based on the H.264 or H.265 standards, employ various techniques like motion estimation, transform coding, and entropy coding to compress video data. However, there is often a trade-off between compression efficiency and visual quality. Higher compression typically leads to more data loss and degradation in video quality.

By integrating GANs into the video compression pipeline, researchers aim to improve the visual quality of compressed videos by training GAN models to generate visually appealing content from compressed video data. The GANs are typically trained on a large dataset of uncompressed videos, learning to capture the statistical properties, textures, and details of high-quality video frames. During training, the generator network of the GAN learns to generate visually similar frames to the original video frames, given the compressed video data as input. The discriminator network evaluates the realism and perceptual quality of the

generated frames. Through an adversarial training process, the GAN models learn to produce synthetic frames that preserve essential visual information, effectively compensating for some of the loss incurred during compression.

By incorporating GAN-based optimization techniques into video compression algorithms, it is possible to achieve higher compression ratios without compromising perceptual video quality. The generated frames can fill in missing details, improve texture reproduction, and enhance overall visual fidelity. This allows for a reduced bitrate requirement while maintaining the desired quality level, resulting in more efficient video transmission and storage. It's important to note that the performance of GAN-based video compression heavily relies on the training dataset's diversity, the architecture of the GAN model, the compression algorithm's design, and the specific optimization techniques employed. Ongoing research in this field aims to further refine GAN-based video compression methods, enabling even more efficient compression and higher perceptual quality in future video streaming and storage applications.

b.   *Super-Resolution [24]*: GAN models employed to enhance the resolution of video frames, enabling better visual quality and detail during streaming. Super-resolution is a technique used to enhance the resolution of images or video frames, enabling better visual quality, sharper details, and improved perception of fine textures. Generative Adversarial Networks (GANs) have proven to be effective in achieving high-quality super-resolution results, and they can be employed to enhance the resolution of video frames during streaming. GAN-based super-resolution works by training a generator network to upsample low-resolution video frames to a higher resolution. The generator network is trained using pairs of low-resolution and high-resolution video frames. The generator's objective is to generate high-resolution frames that closely resemble the corresponding high-resolution frames from the training dataset.

The training process involves an adversarial setup where a discriminator network evaluates the realism of the generated highresolution frames. This adversarial training helps the generator network learn to produce visually plausible and sharp frames that are indistinguishable from the true high-resolution frames. Additionally, perceptual loss functions are often employed to ensure that the generated frames preserve important details and visual characteristics. When integrated into video streaming applications, GAN-based super-resolution techniques allow for the real-time enhancement of video frames, resulting in improved visual quality and finer details. Low-resolution video frames can be upsampled to a higher resolution, enhancing the viewing experience for users.

Super-resolution using GANs can be particularly beneficial when streaming content over bandwidth-limited networks or when the original video source has a lower resolution. By enhancing the resolution of video frames during streaming, viewers can enjoy videos with crisper details and improved image fidelity, creating a more immersive and visually satisfying experience. It's worth noting that the success of GAN-based super-resolution techniques depends on factors such as the diversity and quality of the training dataset, the architecture and design of the GAN model, and the optimization methods employed during training. Ongoing research in this area aims to further refine GAN-based super-resolution methods, enabling even higher levels of detail and visual quality in streaming video applications.

Bandwidth Adaptation and Quality Control:

a.   *Bandwidth Prediction [16]*: GAN-based models used to predict future network conditions and available bandwidth to optimize video streaming decisions, such as bitrate selection. Bandwidth prediction is a critical component of video streaming systems, as it enables adaptive bitrate selection and ensures an optimal streaming experience for users. Generative Adversarial Networks (GANs) have been utilized to develop models that can predict future network conditions and available bandwidth, allowing for more informed video streaming decisions. In the context of bandwidth prediction, GANs are trained on historical network data, such as past bandwidth measurements, network congestion levels, and other relevant features. The GAN consists of a generator network and a discriminator network, similar to other GAN architectures. However, in this case, the generator network's task is to generate future bandwidth predictions based on the historical data, while the discriminator network assesses the accuracy and quality of these predictions.

During training, the GAN learns to capture the temporal dependencies and patterns in the historical network data. The generator network is trained to generate plausible future bandwidth values, while the discriminator network distinguishes between real and generated bandwidth sequences. The adversarial training process enables the GAN model to generate accurate and realistic predictions of future bandwidth conditions. By employing GAN-based bandwidth prediction models in video streaming systems, bitrate selection and other streaming decisions can be optimized. The predicted future bandwidth information can be used to adaptively adjust the streaming bitrate to match the available network conditions. When bandwidth is expected to be high, the system can select a higher bitrate to deliver higher-quality video. Conversely, if bandwidth is predicted to be low, the system can choose a lower bitrate to avoid buffering or interruptions.

By leveraging GAN-based bandwidth prediction, video streaming systems can dynamically adapt their streaming strategies to optimize the trade-off between video quality and smooth playback, providing users with the best possible streaming experience

given the predicted network conditions. It's important to note that the accuracy of GAN-based bandwidth prediction models depends on factors such as the availability and quality of historical network data, the architecture and design of the GAN model, and the specific optimization techniques employed during training. Ongoing research in this area aims to improve the accuracy and reliability of bandwidth prediction models, leading to more efficient video streaming over varying network conditions.

b.   *Rate Adaptation [17)]*: GAN architectures utilized to adaptively adjust the video bitrate based on network conditions and user preferences, ensuring smooth and uninterrupted streaming. Rate adaptation refers to the process of dynamically adjusting the video bitrate during streaming based on network conditions and user preferences. While Generative Adversarial Networks (GANs) are not typically utilized directly for rate adaptation in streaming systems, we will explain an alternative approach that incorporates GAN architectures to achieve adaptive bitrate adjustments. One approach to rate adaptation that incorporates GAN architectures is to leverage the power of GAN-based video super-resolution. Super-resolution techniques aim to enhance the quality of low-resolution videos by generating high-resolution versions. By utilizing GANs for video super-resolution, the adaptive streaming system can dynamically adjust the bitrate based on network conditions and user preferences while ensuring smooth and uninterrupted streaming.

Here's how this GAN-based rate adaptation could work. At the start of the streaming session, an initial video bitrate is selected based on the user's device capabilities, network conditions, and user preferences. This initial bitrate ensures a reasonable quality level for smooth playback. Throughout the streaming session, the system continuously monitors the network conditions, including available bandwidth, latency, and fluctuations in the network quality. This real-time monitoring allows the system to adapt the bitrate on the fly to optimize the streaming experience. When the network conditions change or deteriorate, and the current bitrate may no longer be sustainable, the system initiates a quality adjustment decision. In this step, the GAN architecture comes into play. Instead of directly reducing the video bitrate, the system employs GAN-based video super-resolution techniques. The low-resolution video frames from the current bitrate are fed into the GAN architecture. The GAN then generates high-resolution frames that maintain or enhance the perceived quality, even at lower bitrates.

After generating the high-resolution frames, the system dynamically adjusts the video bitrate downwards while maintaining the improved visual quality. This adaptive bitrate adjustment ensures that the streaming experience remains smooth and uninterrupted, despite the reduced network conditions. The process of real-time network monitoring, quality adjustment decisions, GAN-based super-resolution, and bitrate adjustments continue throughout the streaming session. The system continually evaluates the network conditions, user preferences, and available resources to optimize the streaming experience.
By integrating GAN-based video super-resolution techniques into the rate adaptation process, this approach can adaptively adjust the video bitrate while ensuring that the perceived quality remains high. It enhances the streaming experience by leveraging GAN architectures to generate high-resolution frames that compensate for lower bitrates caused by network limitations.

We now provide an overview of how GAN architectures could potentially be utilized in client-side, in-network and server-side for adaptive bitrate adjustment. In a client-side DASH scenario, GAN architectures can be employed to enhance the video quality on the client-side based on network conditions and user preferences. GAN-based video super-resolution techniques can be utilized to generate high-resolution frames from lower bitrate video segments. By adapting the bitrate dynamically while leveraging GAN-generated high-resolution frames, the client can maintain an improved visual quality, even under varying network conditions. In the context of in-network DASH, GAN architectures can be utilized within the network infrastructure to assist in adaptive bitrate adjustment. Network devices or proxies equipped with GAN-based video super-resolution capabilities can enhance the video quality before delivering it to the client. By generating high-resolution frames within the network, the system can optimize the streaming experience and ensure smooth playback even with limited network resources.

In server-side DASH, GAN architectures can be employed to enhance the video quality before it is delivered to the clients. The server can leverage GAN-based video super-resolution techniques to generate high-resolution frames from lower bitrate video representations. By dynamically adjusting the bitrate while utilizing GAN-generated frames, the server can optimize the streaming experience for clients with varying network conditions and user preferences. It's important to note that while GANs can provide benefits in terms of enhancing video quality, the actual implementation and integration of GAN architectures in client-side, in-network, or server-side DASH systems would require careful consideration of factors such as computational requirements, latency, and scalability. The utilization of GAN architectures for rate adaptation in DASH implementations, irrespective of the specific context, would involve leveraging GAN-based video super-resolution techniques to enhance the visual quality of the streamed content while adapting the video bitrate based on network conditions and user preferences.

Dynamic Adaptive Bitrate Streaming:

a.   *Video Segment Selection [3)]*: GANs employed to select appropriate video segments based on network conditions, user preferences, and video content characteristics to optimize the streaming experience. Video segment selection plays a crucial role in optimizing the streaming experience by choosing appropriate segments of a video based on factors such as network conditions,

A Taxonomy for Generative Adversarial Networks in Dynamic Adaptive Streaming Over HTTP

user preferences, and video content characteristics. Generative Adversarial Networks (GANs) have been utilized to develop models that can assist in the selection of video segments, improving the streaming experience for users. In the context of video segment selection, GANs can be trained on a variety of data, including user preferences, historical viewing patterns, network conditions, and video content metadata. The GAN architecture typically consists of a generator network and a discriminator network. The generator network is responsible for generating candidate video segments, while the discriminator network evaluates the quality, relevance, and suitability of these segments.

During training, the GAN learns to capture the relationships between different factors and generate video segments that align with user preferences, network conditions, and video content characteristics. The discriminator network provides feedback to the generator network, helping it improve its segment selection capabilities over time. This adversarial training process enables the GAN model to generate appropriate video segments that optimize the streaming experience based on the given context. By employing GAN-based video segment selection models, streaming systems can adaptively choose video segments that cater to specific user preferences and network conditions. For example, if the network conditions are poor, the system may prioritize selecting lower bitrate segments or shorter segments to ensure smooth playback and minimize buffering. Conversely, if the network conditions are stable and user preferences indicate a preference for higher-quality content, the system may select higher bitrate segments with longer durations.

The GAN-based segment selection models can also take into account video content characteristics such as genre, scene complexity, or specific content features to further optimize the streaming experience. This ensures that the selected video segments align with the content being streamed and provide an enhanced viewing experience. It's important to note that the effectiveness of GAN-based video segment selection relies on factors such as the diversity and quality of the training data, the architecture and design of the GAN model, and the optimization techniques employed during training. Ongoing research in this field aims to improve the accuracy and personalization capabilities of video segment selection models, leading to more tailored and optimized streaming experiences for users.

b.   *Chunk-Based Adaptation [7)]*: GAN models utilized to determine the optimal chunk size and quality level for video streaming, dynamically adjusting the streaming parameters. Chunk-based adaptation is a technique used in video streaming to dynamically adjust the size and quality level of video chunks based on network conditions and user preferences. Generative Adversarial Networks (GANs) have been employed to develop models that can determine the optimal chunk size and quality level, optimizing the streaming parameters for an enhanced viewing experience. In the context of chunk-based adaptation, GANs can be trained on various data sources, including historical network measurements, user behavior, and video content characteristics. The GAN architecture typically consists of a generator network and a discriminator network. The generator network generates candidate chunk configurations, including different chunk sizes and quality levels, while the discriminator network evaluates the quality and suitability of these configurations.

During training, the GAN learns to capture the relationships between network conditions, user preferences, and video content characteristics to generate optimal chunk configurations. The discriminator network provides feedback to the generator network, enabling it to improve its decision-making capabilities over time. This adversarial training process allows the GAN model to generate chunk configurations that maximize the streaming experience based on the given context. By employing GAN-based chunk adaptation models, streaming systems can dynamically adjust the size and quality level of video chunks. For instance, if the network conditions are favorable and user preferences indicate a preference for high-quality content, the system may select larger chunks with higher quality levels. In contrast, if the network conditions deteriorate or user preferences lean towards minimizing buffering, the system may choose smaller chunks with lower quality levels to ensure smooth playback.

The GAN-based chunk adaptation models take into account factors such as available bandwidth, latency, video complexity, and user preferences to optimize the streaming parameters. This ensures that the streaming system dynamically adapts to the changing network conditions and user requirements, delivering an optimal balance between video quality and smooth playback. It's worth noting that the effectiveness of GAN-based chunk adaptation depends on factors such as the training dataset's diversity and quality, the architecture and design of the GAN model, and the optimization techniques employed during training. Ongoing research in this area aims to improve the accuracy and real-time adaptation capabilities of chunk-based adaptation models, leading to more efficient and personalized video streaming experiences.

Quality Enhancement and Restoration:

a.   *Video Denoising [2)]*: GAN architectures utilized to remove noise and artifacts from video streams, enhancing visual quality During streaming. Video denoising is a technique used to reduce noise and artifacts in video streams, resulting in improved visual quality and a more pleasant viewing experience. Generative Adversarial Networks (GANs) have proven to be effective in addressing video noise and artifacts by generating clean and visually appealing video frames in real-time during streaming. In the context of video denoising, GANs are trained on pairs of noisy and clean video frames. The generator network in the GAN is

responsible for generating denoised frames that resemble the clean frames, given the corresponding noisy frames as input. The discriminator network evaluates the realism and perceptual quality of the generated frames.

During the training process, the GAN model learns to capture the noise characteristics and patterns present in the training data, allowing it to generate denoised frames that effectively remove noise and artifacts while preserving essential visual details. The adversarial training process helps the generator network produce realistic and visually pleasing frames that are perceptually close to the clean frames. When integrated into video streaming applications, GAN-based video denoising techniques can enhance the visual quality of streamed videos by reducing noise, compression artifacts, and other imperfections. This results in a clearer and more enjoyable viewing experience for users, especially in scenarios where the video content may have been affected by low light conditions, high compression ratios, or transmission errors. It's important to note that the performance of GAN-based video denoising methods depends on factors such as the diversity and quality of the training dataset, the architecture and design of the GAN model, and the optimization techniques employed during training. Ongoing research in this field aims to further refine GAN-based video denoising techniques, enabling even higher levels of noise reduction and visual quality improvement during video streaming.

b. *Video Deblurring [5]]*: GAN models employed to reduce motion blur and improve the sharpness of video frames during streaming. Video deblurring is a technique used to reduce motion blur and enhance the sharpness of video frames, resulting in improved visual quality and clarity during streaming. Generative Adversarial Networks (GANs) have emerged as a powerful tool in addressing video blur by generating sharp and clear frames in real-time. In the context of video deblurring, GANs are trained on pairs of blurred and sharp video frames. The generator network in the GAN is responsible for generating deblurred frames that resemble the sharp frames, given the corresponding blurred frames as input. The discriminator network evaluates the realism and perceptual quality of the generated frames.

During the training process, the GAN model learns to capture the motion blur characteristics and patterns present in the training data, enabling it to generate deblurred frames that effectively reduce the blur while preserving important visual details. The adversarial training process helps the generator network produce deblurred frames that are visually realistic and closely resemble the sharp frames. When integrated into video streaming applications, GAN-based video deblurring techniques can significantly enhance the visual quality of streamed videos by reducing motion blur and improving sharpness. This is particularly beneficial in scenarios where the video content may have been affected by camera shake, fast motion, or other factors causing blur during video capture.

By applying GAN-based deblurring in real-time during streaming, the perceived sharpness and clarity of video frames can be significantly improved. This leads to a more enjoyable viewing experience for users, where details are more pronounced, and the overall visual quality is enhanced. It's important to note that the performance of GAN-based video deblurring methods relies on factors such as the diversity and quality of the training dataset, the architecture and design of the GAN model, and the optimization techniques employed during training. Ongoing research in this field aims to further refine GAN-based video deblurring techniques, enabling even better motion blur reduction and visual quality improvement during video streaming.

c. *Video Colorization [14)]*: GAN-based approaches used to add color to grayscale or low-quality videos, enhancing visual experience during streaming. Video colorization is a technique used to add color to grayscale or low-quality videos, enhancing the visual experience and making the content more engaging and realistic. Generative Adversarial Networks (GANs) have been employed to develop models that can effectively colorize videos in real-time during streaming. In the context of video colorization, GANs are trained on pairs of grayscale or low-quality video frames and their corresponding colorized versions. The generator network in the GAN is responsible for generating colorized frames that closely resemble the true colors of the input frames. The discriminator network evaluates the realism and quality of the generated colorized frames.

During the training process, the GAN model learns to capture the relationship between grayscale/low-quality frames and their colorized counterparts. This allows the generator network to generate colorized frames that accurately depict the original colors and visual characteristics of the input frames. The adversarial training process helps the generator network produce colorized frames that are visually realistic and visually pleasing. When integrated into video streaming applications, GAN-based video colorization techniques can significantly enhance the visual experience by adding color to grayscale or low-quality videos. This is particularly useful in scenarios where the original video content is in grayscale, historical footage lacks color, or low-quality videos need visual enhancement. The colorization process can make the content more appealing, realistic, and immersive for viewers.

By applying GAN-based colorization in real-time during streaming, the grayscale or low-quality videos can be transformed into vibrant and visually pleasing content. This enhances the overall visual experience for users, making the content more engaging and enjoyable. It's important to note that the performance of GAN-based video colorization depends on factors such as the diversity and quality of the training dataset, the architecture and design of the GAN model, and the optimization techniques

employed during training. Ongoing research in this field aims to further refine GAN-based video colorization methods, enabling even better color accuracy and visual quality improvement during video streaming.

Quality Assessment and Objective Metrics:

a. *Perceptual Quality Assessment [13])*: GAN-based models utilized to assess the perceived quality of streamed videos,

Providing objective quality metrics for adaptive streaming algorithms. Video colorization is a technique used to add color to grayscale or low-quality videos, enhancing the visual experience and making the content more engaging and realistic. Generative Adversarial Networks (GANs) have been employed to develop models that can effectively colorize videos in real-time during streaming. In the context of video colorization, GANs are trained on pairs of grayscale or low-quality video frames and their corresponding colorized versions. The generator network in the GAN is responsible for generating colorized frames that closely resemble the true colors of the input frames. The discriminator network evaluates the realism and quality of the generated colorized frames.

During the training process, the GAN model learns to capture the relationship between grayscale/low-quality frames and their colorized counterparts. This allows the generator network to generate colorized frames that accurately depict the original colors and visual characteristics of the input frames. The adversarial training process helps the generator network produce colorized frames that are visually realistic and visually pleasing. When integrated into video streaming applications, GAN-based video colorization techniques can significantly enhance the visual experience by adding color to grayscale or low-quality videos. This is particularly useful in scenarios where the original video content is in grayscale, historical footage lacks color, or low-quality videos need visual enhancement. The colorization process can make the content more appealing, realistic, and immersive for viewers.

By applying GAN-based colorization in real-time during streaming, the grayscale or low-quality videos can be transformed into vibrant and visually pleasing content. This enhances the overall visual experience for users, making the content more engaging and enjoyable. It's important to note that the performance of GAN-based video colorization depends on factors such as the diversity and quality of the training dataset, the architecture and design of the GAN model, and the optimization techniques employed during training. Ongoing research in this field aims to further refine GAN-based video colorization methods, enabling even better color accuracy and visual quality improvement during video streaming.

b. *Content-Adaptive Metrics [22)]*: GAN architectures used to develop metrics that consider content-specific quality aspects, such as video complexity, motion, or scene semantics. Content-adaptive metrics play a crucial role in assessing the quality of streamed videos by considering content-specific aspects that influence the perceived quality. Generative Adversarial Networks (GANs) have been employed to develop metrics that capture content-specific quality aspects, such as video complexity, motion, or scene semantics. In the context of content-adaptive metrics, GANs are trained on pairs of videos and corresponding subjective quality scores obtained through human assessments. The generator network in the GAN is responsible for generating quality scores or features that capture content-specific quality aspects. The discriminator network evaluates the realism and relevance of the generated quality scores or features.

During the training process, the GAN model learns to capture the relationships between video content and subjective quality assessments, enabling it to generate content-adaptive quality metrics. These metrics take into account the specific characteristics of the video content, such as its complexity, motion characteristics, or semantic information, to provide quality assessments that align with human perception. By employing GAN-based content-adaptive metrics, video streaming systems can utilize quality metrics that are tailored to the specific content being streamed. This enables more accurate and meaningful quality assessments that consider the content-specific aspects influencing the perceived quality. Such metrics can be used in adaptive streaming algorithms to make informed decisions on video representation selection or bitrate adaptation, ensuring optimal quality for different types of content.

The use of GAN-based models in developing content-adaptive metrics allows for a more fine-grained analysis of video quality that goes beyond traditional metrics. GANs can capture complex content characteristics and generate quality metrics that are sensitive to those aspects, providing a more comprehensive assessment of video quality that aligns with human perception. It's important to note that the effectiveness of GAN-based content-adaptive metrics depends on factors such as the diversity and quality of the training dataset, the architecture and design of the GAN model, and the optimization techniques employed during training. Ongoing research in this field aims to refine and improve GAN-based content-adaptive metrics, enabling more accurate and context-aware quality assessment in video streaming applications. *Cross-Modal Applications*:

a. *Audio-Visual Synchronization [4)]*: GAN models employed to synchronize audio and video streams during DASH playback, ensuring lip-sync accuracy and improving the user experience. Audio-visual synchronization is a critical aspect of video streaming to ensure that the audio and video streams are properly aligned, resulting in accurate lip-sync and an improved user experience. Generative Adversarial Networks (GANs) have been utilized to develop models that can synchronize audio and video streams during Dynamic Adaptive Streaming over HTTP (DASH) playback. In the context of audio-visual synchronization, GANs are trained

on pairs of misaligned audio and video streams. The generator network in the GAN is responsible for generating synchronized audio and video streams by adjusting the temporal alignment between them. The discriminator network evaluates the synchronization accuracy and perceptual quality of the generated streams.

During the training process, the GAN model learns to capture the relationship between audio and video streams and adjust the temporal alignment to achieve accurate synchronization. The adversarial training process helps the generator network produce synchronized audio and video streams that closely match the original timing and lip movements. By employing GANbased audio-visual synchronization models, video streaming systems can ensure that the audio and video streams are properly aligned, resulting in accurate lip-sync and a seamless user experience. This is particularly important in scenarios where misalignment between audio and video can lead to a significant degradation in the viewing experience.

When integrated into DASH playback, GAN-based audio-visual synchronization techniques can dynamically adjust the temporal alignment between audio and video streams to compensate for network delays, buffering, or other factors that may cause synchronization issues. This ensures that the audio and video remain properly aligned throughout the playback, enhancing the overall viewing experience and maintaining lip-sync accuracy. It's important to note that the performance of GAN-based audio-visual synchronization methods depends on factors such as the diversity and quality of the training dataset, the architecture and design of the GAN model, and the optimization techniques employed during training. Ongoing research in this field aims to further refine GAN-based audio-visual synchronization techniques, enabling even more accurate and robust synchronization for seamless video streaming.

b.   *Multimodal Data Generation [15), 6)]*: GAN architectures utilized to generate multimodal data, such as synchronized audio and video streams, for enhanced streaming experiences. Multimodal data generation refers to the generation of data that combines multiple modalities, such as synchronized audio and video streams, to create a richer and more immersive streaming experience. Generative Adversarial Networks (GANs) have been employed to develop architectures that can generate multimodal data, allowing for enhanced streaming experiences. In the context of multimodal data generation, GANs are trained on paired examples of multimodal data, such as synchronized audio and video streams. The generator network in the GAN is responsible for generating new instances of multimodal data, combining audio and video modalities in a synchronized manner. The discriminator network evaluates the realism and quality of the generated multimodal data.

During the training process, the GAN model learns to capture the relationships between the audio and video modalities and generate synchronized instances that resemble the training data. This enables the generator network to produce new, realistic, and synchronized audio and video streams. The adversarial training process helps ensure the generated multimodal data is perceptually accurate and visually and audibly plausible. By employing GAN-based multimodal data generation models, streaming applications can enhance the user experience by generating immersive and synchronized audio and video content. This can be particularly valuable in scenarios where synchronized audio and video are crucial, such as in video conferencing, virtual reality, or interactive streaming applications.

The use of GANs in generating multimodal data enables the creation of diverse and realistic combinations of audio and video streams, leading to more engaging and immersive streaming experiences. This technology can be used to generate content that matches specific user preferences or to create interactive and personalized streaming applications. It's important to note that the performance of GAN-based multimodal data generation depends on factors such as the diversity and quality of the training dataset, the architecture and design of the GAN model, and the optimization techniques employed during training. Ongoing research in this field aims to further refine GAN-based multimodal data generation techniques, enabling even more realistic and immersive streaming experiences.

This taxonomy provides an overview of the different aspects and applications of Generative Adversarial Networks (GANs) in the context of Dynamic Adaptive Streaming over HTTP (DASH). It covers various stages of the streaming pipeline, including video generation, compression, adaptation, quality enhancement, and cross-modal applications. Note that the field of GANs in DASH is evolving, and new architectures, techniques, and applications may emerge in the future.

V. DISCUSSION

The taxonomy presented for Generative Adversarial Networks (GANs) in Dynamic Adaptive Streaming over HTTP (DASH) provides a structured framework to categorize and organize the various applications and techniques of GANs in the context of video streaming. In this discussion, we explore the significance and implications of the taxonomy. The taxonomy covers a wide range of dimensions in which GANs can be applied to improve DASH (Comprehensive Coverage). It encompasses video generation, compression, quality enhancement, bandwidth adaptation, dynamic bitrate streaming, and cross-modal applications. This comprehensive coverage ensures that the taxonomy captures the major areas where GANs can play a vital role in optimizing the streaming experience.

A Taxonomy for Generative Adversarial Networks in Dynamic Adaptive Streaming Over HTTP

The taxonomy recognizes the specific applications of GANs in DASH, addressing important aspects such as content generation, frame interpolation, video compression, super-resolution, denoising, deblurring, and more. By organizing the applications within the taxonomy, researchers and practitioners can easily identify the specific areas of interest and relevance to their own work or domain. The taxonomy acknowledges the importance of adaptation and quality control in DASH. GAN-based approaches for bandwidth prediction, rate adaptation, video segment selection, and chunk-based adaptation are included. These techniques are crucial for optimizing video streaming decisions, ensuring smooth playback, and providing the best possible video quality under varying network conditions.

The taxonomy highlights the significance of evaluation metrics for assessing the effectiveness of GAN-based approaches in DASH. It discusses metrics such as perceptual quality assessment, content-adaptive metrics, and objective metrics like Inception Score and Fréchet Inception Distance. By considering various evaluation metrics, the taxonomy emphasizes the need for quantitative analysis and comparison of different GAN models and techniques. The taxonomy also points to emerging research areas within the scope of GANs in DASH. For example, the cross-modal applications category recognizes the importance of audio-visual synchronization and multimodal data generation. These areas offer opportunities for further exploration and innovation, where GANs can contribute to enhancing the immersive and synchronized streaming experience.

The taxonomy has practical implications for researchers and practitioners working on video streaming and DASH applications. It provides a structured framework that facilitates understanding, comparison, and exploration of GAN-based approaches. The taxonomy enables researchers to identify gaps in the field, identify specific subcategories of interest, and develop novel solutions for improving video quality, bandwidth adaptation, and user satisfaction in DASH. The taxonomy presented for GANs in DASH serves as a valuable resource for understanding the different dimensions, applications, and techniques of GANs in the context of video streaming. By organizing the diverse aspects within a structured framework, the taxonomy enables researchers and practitioners to navigate the field, identify areas for further research, and leverage GANs to optimize the streaming experience in Dynamic Adaptive Streaming over HTTP.

VI. CONCLUSIONS

The taxonomy presented for Generative Adversarial Networks (GANs) in Dynamic Adaptive Streaming over HTTP (DASH) provides a comprehensive and structured framework for understanding and organizing the applications and techniques of GANs in the context of video streaming. It covers various dimensions, including video generation, compression, quality enhancement, bandwidth adaptation, dynamic bitrate streaming, and cross-modal applications. By categorizing and organizing the different aspects of GANs in DASH, the taxonomy facilitates better understanding, comparison, and exploration of GAN-based approaches. It enables researchers and practitioners to identify specific areas of interest, address research gaps, and develop novel solutions to enhance the streaming experience. The taxonomy highlights the relevance of GANs in key areas such as content generation, compression, quality enhancement, adaptation, and cross-modal applications. It recognizes the significance of evaluation metrics for assessing the effectiveness of GAN-based techniques and emphasizes the need for quantitative analysis and comparison. Furthermore, the taxonomy points to emerging research areas within the scope of GANs in DASH, such as audio-visual synchronization and multimodal data generation. These areas offer opportunities for further exploration and innovation, contributing to the development of more immersive and synchronized streaming experiences. The taxonomy serves as a valuable resource for researchers and practitioners involved in video streaming and DASH applications. It provides a structured framework to navigate the field, identify relevant subcategories, and leverage GANs to optimize video quality, bandwidth adaptation, and user satisfaction in the context of Dynamic Adaptive Streaming over HTTP.

REFERENCES

1) Agnese, J., Herrera, J., Tao, H., & Zhu, X. (2020). A survey and taxonomy of adversarial neural networks for text-to-image synthesis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(4), e1345.
2) Bisht, A., Mendes, A. C. D. S., Thoreson II, J. D., & Samavi, S. (2023). Low Latency Video Denoising for Online Conferencing Using CNN Architectures. arXiv preprint arXiv:2302.08638.
3) Chiariotti, F. (2021). A survey on 360-degree video: Coding, quality of experience and streaming. Computer Communications, 177, 133-155.
4) Dash, A., Ye, J., & Wang, G. (2021). A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines--From Medical to Remote Sensing. arXiv preprint arXiv:2110.01442.
5) Guo, Q., Feng, W., Gao, R., Liu, Y., & Wang, S. (2021). Exploring the effects of blur and deblurring to visual object tracking. IEEE Transactions on Image Processing, 30, 1812-1824.

6) He, Y., Seng, K. P., & Ang, L. M. (2023). Multimodal Sensor-Input Architecture with Deep Learning for Audio-Visual Speech Recognition in Wild. Sensors, 23(4), 1834.

7) Izima, O., de Fréin, R., & Malik, A. (2021). A Survey of Machine Learning Techniques for Video Quality Prediction from Quality of Delivery Metrics. Electronics, 10(22), 2851.

8) Khan, K., & Goodridge, W. (2018). Bandwidth Estimation Techniques for Relative 'Fair' Sharing in DASH. International Journal of Advanced Networking and Applications, 9(6), 3607-3615.

9) Khan, K., & Goodridge, W. (2018). Future DASH applications: A survey. International Journal of Advanced Networking and Applications, 10(2), 3758-3764.

10) Khan, K., & Goodridge, W. (2018). QoE in DASH. International Journal of Advanced Networking and Applications, 9(4), 3515-3522.

11) Khan, K., & Goodridge, W. Markov Decision Processes for bitrate harmony in adaptive video streaming. In 2017 Future Technologies Conference (FTC), Vancouver, Canada, unpublished.

12) Koffka, K., & Wayne, G. (2018). A DASH Survey: the ON-OFF Traffic Problem and Contemporary Solutions. Computer Sciences and Telecommunications, (1), 3-20.

13) Lee, R., Venieris, S. I., & Lane, N. D. (2021). Deep neural network–based enhancement for image and video streaming systems: A survey and future directions. ACM Computing Surveys (CSUR), 54(8), 1-30.

14) Lee, Y., & Lee, S. (2020). Automatic colorization of anime style illustrations using a two-stage generator. Applied Sciences, 10(23), 8699.

15) Li, T., Huang, J., Risinger, E., & Ganesan, D. (2021, June). Low-latency speculative inference on distributed multi-modal data streams. In Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services (pp. 67-80).

16) Li, W., Huang, J., Wang, S., Wu, C., Liu, S., & Wang, J. (2022). An apprenticeship learning approach for adaptive video streaming based on chunk quality and user preference. IEEE Transactions on Multimedia.

17) Li, W., Oteafy, S. M., Fayed, M., & Hassanein, H. S. (2020). Quality of experience in ICN: Keep your low-bitrate close and high-bitrate closer. IEEE/ACM Transactions on Networking, 29(2), 557-570.

18) Mentzer, F., Toderici, G. D., Tschannen, M., & Agustsson, E. (2020). High-fidelity generative image compression. Advances in Neural Information Processing Systems, 33, 11913-11924.

19) Saxena, D., & Cao, J. (2021). Generative adversarial networks (GANs) challenges, solutions, and future directions. ACM Computing Surveys (CSUR), 54(3), 1-42.

20) Shangguan, W., Sun, Y., Gan, W., & Kamilov, U. S. (2022, October). Learning cross-video neural representations for highquality frame interpolation. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XV (pp. 511-528). Cham: Springer Nature Switzerland.

21) Stockhammer, T. (2011, February). Dynamic adaptive streaming over HTTP-- standards and design principles. In Proceedings of the second annual ACM conference on Multimedia systems (pp. 133-144).

22) Wang, Y., Chan, P. H., & Donzella, V. (2023). Semantic-Aware Video Compression for Automotive Cameras. IEEE Transactions on Intelligent Vehicles.

23) You, A., Kim, J. K., Ryu, I. H., & Yoo, T. K. (2022). Application of generative adversarial networks (GAN) for ophthalmology image domains: a survey. Eye and Vision, 9(1), 1-19.

24) Zhang, A., Wang, C., Han, B., & Qian, F. (2021, February). Efficient volumetric video streaming through super resolution. In Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications (pp. 106-111).