Journal of Network and Computer Applications 211 (2023) 103564

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

Intelligent aerial video streaming: Achievements and challenges

The-Vinh Nguyen [a], Ngoc Phi Nguyen [b], Cheonshik Kim [a], Nhu-Ngoc Dao [a,*]

[a] Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea
[b] Department of Technology, Dong Nai Technology University, Bien Hoa 810000, Viet Nam

ARTICLE INFO

Keywords:
Aerial infrastructure
Artificial intelligence
Video streaming
Wireless communication

ABSTRACT

Deploying aerial infrastructure to support traditional terrestrial wireless communications has enhanced the maturity of 5G networks and beyond. This improvement has certainly benefited video streaming services whose data makes up the bulk of mobile traffic and continues to soar. Realising this fashion, artificial intelligence is considered as one of the important tools, which can effectively deal with various problems to improve and optimise many aspects of video streaming systems. However, there is still a lack of a systematic investigation of recent works in this field to facilitate researchers with a comprehensive reference framework to advance the development of the technology. Realising this shortcoming, this paper has dedicatedly investigated the recent achievements in using AI and the future challenges of the system. First, the basic knowledge and overall architecture of the aerial video streaming (AVS) system are given. Subsequently, achievable utilities are thoroughly evaluated by highlighting each system performance mathematically and analysing recent technical works to address them. Next, typical application scenarios are acquired to demonstrate the advantages to human life and the necessity of their implementation. Finally, open challenges are discussed to drive future technological developments towards video streaming over aerial infrastructure.

1. Introduction

The world is witnessing the tremendous growth of video streaming services in the Internet era. According to recent Cisco Annual Internet Report (Cisco, 2020), there will be 5.3 billion total Internet users in 2023, accounting for almost two-thirds of the world's current population. In compliance with a survey by Ericsson (Ericsson, 2018), video traffic is expected to account for 74% of the total data traffic, making it become the most dominant data category with the greatest number of user requests. It indicates a significant increase in the number of user equipment (UE) that drives up demand for video streaming services. In addition, the rise of video service providers such as Youtube, Tiktok, and Twitch has also highlighted the pervasiveness of video services in daily life. Consequently, video traffic has increased over the years and would continue to skyrocket in the future with an extensive amount of data.

In this circumstance, the current network infrastructures have been unfortunately considered incapable of serving all users as there are more than 30% of the global population does not have Internet accessibility. These prospective customers are largely concentrated in rural parts of the world where mobile coverage is insufficient or non-existent. The survey of Yaacoub and Alouini (Yaacoub and Alouini, 2020) have investigated in detail the necessity for employing Internet connectivity in rural areas. According to their findings, the challenge stems from the hostile climate, which has an impact on terrestrial facility building and system maintenance. To cope with this issue, the potential approach for delivering Internet connectivity under these limits in current mobile networking infrastructures would be to deploy aerial radio access networks (ARANs) (Dao et al., 2021a). ARAN is a system composed of multiple airborne devices (e.g., satellites, unmanned aerial vehicles (UAVs), and airplanes), each of which functions as an independent aerial base station (ABS) and operates on the air infrastructure. In particular, ARAN consists of three main aerial communication tiers: low Earth orbit (LEO) satellite constellation, high-altitude platform (HAP) topology, and low-altitude platform (LAP) swarm, which position at different altitudes from 0–1500 km height (Dao et al., 2021b). Along with these architecture, ARANs are expected to play a major role in the deployment of sixth-generation (6G) technology in the future thanks to their mobility, competitive computational capacity, and independence from traditional terrestrial infrastructures.

1.1. Motivation

Nevertheless, ARANs remain difficult in transferring massive amounts of data traffic in real-time; hence, implementing video streaming services within their architecture would be a challenging task.

* Corresponding author.
E-mail addresses: nguyenthevinh@sju.ac.kr (T.-V. Nguyen), nguyenngocphi@dntu.edu.vn (N.P. Nguyen), mipsan@sejong.ac.kr (C. Kim), nndao@sejong.ac.kr (N.-N. Dao).

https://doi.org/10.1016/j.jnca.2022.103564
Received 28 June 2022; Received in revised form 7 November 2022; Accepted 14 December 2022
Available online 17 December 2022
1084-8045/

Table 1
Summary on existing works on video streaming system and their contributions.

Ref.	Theme	Video source	Access network	AI approach	Contributions
Karaki et al. (2019)	Motion detection and tracking	UAV	Partially aerial		This work highlighted different existing solutions in aerial surveillance videos to improve quality and reduce image loss.
Marvasti-Zadeh et al. (2021)	Visual tracking	UAV	Partially aerial	✓	This work surveyed existing deep learning (DL) solutions on aerial object tracking based on various system characteristics.
Jiang et al. (2021a)	Resource allocation (RA)	Multiple	Terrestrial		This survey investigated recent solutions of RA optimisations in terms of communication, caching, and computing for video streaming over vehicular networks.
Yaqoob et al. (2020), Fan et al. (2019)	360-degree video	Multiple	Terrestrial		Viewport-independent, viewport-dependent, and title-based solutions are studied for video delivery problem in 360-degree video and VR streaming.
Lee et al. (2021)	Visual enhancement	Multiple	Terrestrial	✓	This paper examined existing technologies of content delivery system which works primarily on DL enhancement to improve responding time and visual quality of video.
Jiang et al. (2021b)	Edge computing utilisation	Multiple	Terrestrial		This paper surveyed enabling technologies and optimisations for intelligent video acceleration and video streaming analyses utilising multiaccess edge computing (MEC).
Ahmed et al. (2019)	D2D video transmission	Multiple	Terrestrial		Different challenges of D2D communication such as RA, cache clustering, energy insufficiency are discussed and the solutions are summarised.
Our survey	Aerial video streaming (AVS)	Multiple	Fully aerial	✓	We dedicatedly investigate video streaming services over the potential but challenging aerial infrastructure, where AI approaches have been proposed to optimise system performances in multiple aspects.

Although LAP devices like UAVs have the lowest latency among those of the three tiers, their diminutive design and limited energy make them sensitive to work for a long duration as well as to process large datasets in a video service transmission. Meanwhile, LEO satellites are capable of handling vast volumes of data, abundant power, and wide coverage, but they come with the main disadvantage of excessive delay. Whilst, HAP may be viewed as an average version of both tiers, it can only meet the needs of users to a limited extent. Furthermore, video streaming services commonly necessitate a high level of computing and storage capacity on edge servers and ABSs. The diversity of video contents and heterogeneous networks requires video servers to be able to cache popular videos and analyse networks condition for resolution, which consumes a significant number of cognitive resources. As user expectations on quality of service (QoS) and access bandwidth requirements are rising substantially, it becomes even more critical to overcome these unresolved challenges.

Fortunately motivated by artificial intelligence (AI) triumphs in several technical domains, the AI methods have been expected to be a potential enabler for the video streaming over aerial infrastructure challenge. The use of AI to overcome difficulties related to video service deployment has garnered a lot of interest from scientists. There have been a plethora of transaction papers published on the findings and enhancement on this research area. Notwithstanding, to the best of our knowledge, there has been only a limited amount of papers that comprehensively surveys the accomplishments and problems of AI techniques in video streaming via aerial infrastructure. Moreover, recent surveys are either out of the research scope, or lack of detail information in the area. For example, most of the surveys about video streaming on the aerial network are mainly focused on the topic of object tracking and computer vision detection using UAVs. As a result, HAP and LEO platforms are being left questionable, and the video sources are mostly generated from the UAVs but not from other sources, in particular, terrestrial user devices. On the other hand, articles focusing on AI approaches are leaning towards optimising traditional video streaming frameworks, rather than taking it onto aerial infrastructure. Hence, this survey has been dedicatedly conducted to summarise state-of-the-art studies about video streaming over aerial infrastructures,especially taking advantage of AI approaches, as distinct from other existing surveys.

1.2. Related works

Table 1 summarises existing survey studies on video streaming according to their themes and analysed them based on: *Video source* from which the video was generated, *Access network* wherein video streaming services broadcasted throughout the system, *AI approach* in case researches utilising AI techniques to resolve video streaming problems, and finally is their main *Contributions*.

In Karaki et al. (2019), authors provided a detailed examination on the advantages and challenges in aerial videos on surveillance, in particular motion detection and object tracking. The authors suggested that the three main problems which cause image quality loss or noises were parallax motion, brightness transformation, and height modification. Similarly, the work in Marvasti-Zadeh et al. (2021) surveyed recent deep learning (DL) solutions on aerial object tracking. The most common visual tracking benchmarks and their attributes are compared,

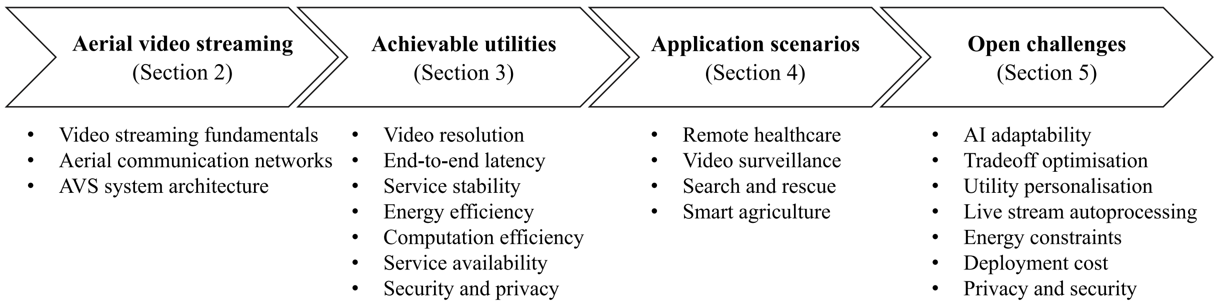

Fig. 1. Investigation framework and its components.

and their assessment metrics are also presented. However, these surveys were focusing on aerial-based video streaming, meaning videos were recorded on aerial devices themselves, rather than functioning as a networking component for video traffic forwarding. On the other hand, Jiang et al. (2021a) have recognised the usefulness of video streaming services in assisting safety driving and in-vehicle entertainment, as well as discussed difficulties in the technology deployments. Hence, cutting-edge solutions specific to the RA optimisation problem were introduced, focusing on the emerging AI approaches. The potential of integrating video streaming with edge-C3 (communication, computing, caching) was mentioned as an improvement for vehicular multimedia services. Finally, the authors have discussed some future challenges needed for further research.

Focusing on the specific 360-degree video streaming, the articles Yaqoob et al. (2020), Fan et al. (2019) surveyed network constraints and existing works on solving these problems. The solutions are divided into three categories: viewport-independent, viewport-dependent, and tile-based. The pros and cons of each category were summarised to compare the differences in these approaches. However, the research did not investigate feasible scenarios where 360-degree video streaming is transmitted over aerial communication networks, which is a challenging environment as the services necessitate a considerable number of processing resources and consistent connectivity. Another work leveraging AI techniques in 360-degree video streaming has been conducted by Lee et al. (2021). The authors examined cutting-edge technology of content delivery system which works primarily on deep neural network enhancement to improve responding time and visual quality of the video. Particularly, the challenges in content delivery systems can be tackled using deep neural network models. The advantages and difficulties in deploying these models were highlighted, as well as giving research directions for future developments.

Considering the advantages of edge computing techniques, the authors in Jiang et al. (2021b) performed a survey of multi-access edge computing (MEC) applications in video streaming services were conducted. Firstly, the fundamentals of edge technology and video streaming architecture were introduced, followed by a discussion on resource allocation (RA) problems. Next, enabling technologies such as blockchain, caching, computing, slicing engineering were surveyed in details. Despite its conveniences, MEC remains problematical in energy efficiency, bandwidth-QoE trade-off, security, and cooperation among multiple MECs. In addition to MEC, device-to-device (D2D) communications were also another promising technology for video streaming. The article Ahmed et al. (2019) pointed out that D2D communication can improve the increasing demand of QoE of users in video streaming services by using the proximity-based services between nearby devices to enhance spectrum/energy efficiencies, storage, and shorten the response time. Nevertheless, D2D is facing challenges in video quality degradation, RA, and signal management. To address these issues, the authors suggested several applicable approaches for future research to mature the technologies.

Although video streaming services have been investigated from multiple perspectives and various insights have been revealed, recent survey papers were not dedicated to properly cover video streaming over aerial infrastructure using AI as the focus but rather treated them separately. Motivated by this observation, this survey aims to comprehensively review video streaming over fully aerial infrastructure, in which video traffic from diverse sources can be streamed and transmitted by airborne devices. Moreover, recent AI solutions on system enhancement are investigated and carefully analysed.

1.3. Our contributions

The purpose of this survey is to present a comprehensive reference framework about cutting-edge research that has been done utilising AI techniques to improve performances of the video streaming services over aerial architectures. In this regard, the structure of this article, which illustrated in Fig. 1, is outlined as follows:

- **AVS systems:** An overview of the AVS system design is presented. Firstly, the foundations of video streaming are provided to assist readers without deep expertise in understanding the operation of a video streaming system. Subsequently, we sketch a prototype of aerial networks that link terrestrial users to the video streaming services through airborne networking components (e.g., UAVs, aircraft, and satellites). Finally, a demonstration of the working principle of video streaming in the aerial network is conducted. Section 2 provides the detail studies.
- **Achievable utilities:** This section analyses cutting-edge studies that exploit AI methods to tackle difficulties in video streaming services over aerial infrastructure. In particular, video resolution, end-to-end latency, service stability, energy efficiency, computing capabilities, service availability, and security/privacy are the seven important parameters being investigated. The constraints of each utility are examined referring to their objective and system constraints. Section 3 contains the specifics of the studies.
- **Application scenarios:** In this section, we investigated four typical application scenarios that well take advantages of video streaming services over aerial infrastructure such as remote healthcare, video surveillance, search and rescue, and smart agriculture. The detailed discussion is given in Section 4.
- **Open challenges:** Future challenges for video streaming services over aerial systems are highlighted in Section 5. Despite its promises, large-scale aerial network deployment retains several challenges for time-sensitive video streaming and require further efforts of technology development towards maturation.

Table 2 summarises key abbreviations used in this paper.

2. Aerial video streaming

The system architecture of AVS is presented in this section. The fundamentals of the video streaming system are described first in order to facilitate comprehension of the working principle, followed by research on the implementation of the aerial communication network. Finally, a complete model of aerial video service is presented as a composite of the former elements.

Table 2
List of abbreviations.

Abbreviation	Description
3GPP	3rd Generation Partnership Project
AI	Artificial Intelligence
AR	Augmented Reality
ABR	Adaptive bitrate
ABS	Aerial Base Station
ARAN	Aerial Radio Access Network
AVS	Aerial video streaming
BS	Base Station
CBR	Constant bitrate
CNN	Convolution neural network
CMDP	constrained Markov decision process
DL	Deep Learning
DNN	Deep neural network
DRL	Deep Reinforcement Learning
DASH	Dynamic adaptive streaming over HTTP
FL	Federated Learning
HAS	HTTP Adaptive Streaming
ISP	Internet service provider
LEO	Low Earth Orbit platform
HAP	High altitude platform
LAP	Low altitude platform
ML	Machine Learning
MEC	Multi-access edge computing
QoE	Quality of Experience
QoS	Quality of Service
RA	Resource Allocation
RL	Reinforcement Learning
SOA	State-of-the-art
SL	Split Learning
TCP	Transmission Control Protocol
UE	User equipment
UAV	Unmanned Aerial Vehicle
VR	Virtual Reality

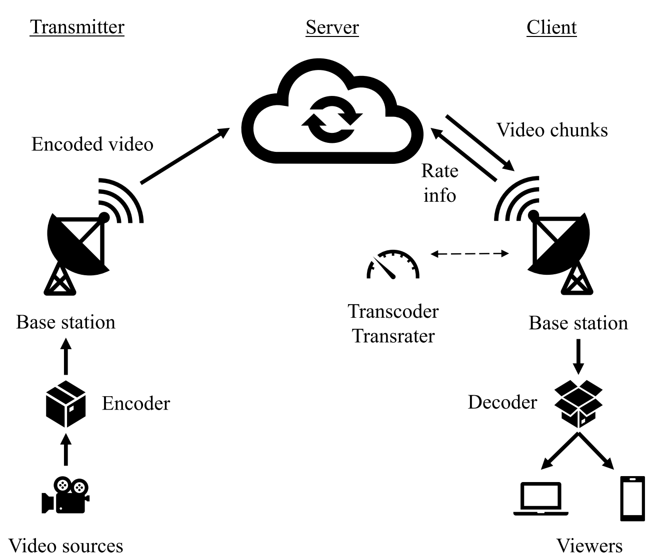

Fig. 2. Fundamental structure of a video streaming system.

2.1. Video streaming fundamentals

Video streaming system generally contain three main components: *Transmitter* refers to the content providers side, from whom the videos are obtained and pre-processed; *Clients* are viewers who receive videos and utilise streaming service; and *Server* which is in charge of distributing videos from the transmitter to clients (Jedari et al., 2020). Fig. 2 demonstrates a schematic structure of the system.

2.1.1. Transmitter

There are two main components worth noting inside the transmitter, which are *Video sources* and *Transmitter base station*, which contains an *Encoder*.

- **Video sources:** refers to the source of the videos, which records and displays frames in fast succession. There are two ways to generate a video: by recording the physical surroundings with a user device; or by creating a video synthetically (Jedari et al., 2020). In a video streaming system, live video streaming is the most prevalent sort of video, where users may directly record and generate video with an UE (i.e., camera or smartphone) (López et al., 2020). Video-on-demand (VoD), on the other hand, reflects the latter situation, with video types such as animation or altered videos.
- **Encoder:** The original form of videos after recording is relatively substantial for transmission across the system, which might cause computational burden for both the server and UE. A video compression is required to achieve a reasonable transmission size, decreases traffic load and redundant information, boosts storage efficiency, while still ensuring the integrity of the video at a local base station (BS). This process is known as *video encoding*. Similarly, on the client side, the compressed video must be decompressed in order to reproduce the video original appearance before it can be displayed by viewers, referred as *video decoding*. In a system-wide perspective, video streaming services are often distributed with the encoder–decoder configuration called *codec*. Barman and Martini (2021) have stated that H.264, HEVC/H.265, MPEG-4, and DivX are the most extensively used codecs recently, and utilise them as a threshold for their suggested integrated codec scheme performance evaluation.
  In Ivan Quinones (2020), Quinones claimed that there are two approaches for video encoding: *lossless encoding* and *lossly encoding*. Lossless encoding compresses video whilst attempting to keep the majority of its properties, ensuring that the compressed and original versions are identical. As a result, fewer modifications can be done, and larger compression sizes result in less space efficiency. Lossly video encoding, on the other hand, aims to delete or disregard some of the original video data in order to maximise storage capacity. Despite lossly encode being more prevail method, it is worth to consider the trade-off between video quality and video transmission sufficiency to select the most suitable encoding style.
- **Transmitter BS:** The base station closest to the video source is in charge of receiving video signals, encoding video, and sending it to the server for delivery. The BS can execute complex computing and pre-processing operations prior of transferring video to the server, minimising its strain Usman et al. (2019). Furthermore, instead of working on the user device at the video source, the encoder can be placed on the transmitter BS premise to exploit its computational resources.

2.1.2. Server

Encoded videos on the BS premise are forwarded towards the server for sorting and pre-processing. In particular, the server mutates an encoded video into numerous small video bitrate segments called *chunks*, and sorts these chunks into a serial line in the correct sequence of the video. A media presentation description (MPD) which contains information on video chunks (i.e., resolution, duration, and subtitles) is generated in the form of extensible markup language (XML) file. When a client sends a video retrieval request to the server, this MPD file is returned to the client so that the client can acknowledge which segment bitrate are accessible on the server (Han et al., 2019). Subsequently, the client selects the most appropriate bitrate described in the MPD to efficiently deliver maximum QoS/QoE for the user based on current network states and computational resources of the viewer. The client then sends the bitrate information back to the server, requesting that video chunks be sent based on the bitrate. This interaction between the server and the client has been considered the standard protocol in video streaming services for its effectiveness in delivering video contents (El Marai et al., 2017).

In addition to client–server collaboration, a server can also establish a bidirectional transmitter–server collaboration by adaptively uploading different frame qualities from the video source to raise the inference accuracy. In Du et al. (2020), a DL-based deep neural (DDS) nets-driven streaming was proposed adopting this model. Here, video sources can station on UEs, which are frequently low in computing capacity and have uncertain communication, hence insufficient for high video quality and bandwidth usage. By applying the DDS framework, the authors have noticed DDS can preserve bitrate prediction accuracy whilst decreasing bandwidth use by up to 59%, or increases the accuracy by up to 9% with no extra bandwidth usage.

2.1.3. Client

Similar to the transmitter side, a client consists of *Client BS* and *Viewers*. However, the emergence of adaptive bitrate (ABR) streaming and the client–server cooperation demands an additional *transcoder/transrater* along with appropriate codec, particularly a *decoder*.

- **Client BS:** Client BS is the closest access point to end-users, performs the most computing duties, and has been the focus of recent video streaming system development. The responsibilities of client BS include resource allocation, transmission optimisation, and transcoding (Liu et al., 2021a). Dynamic adaptive streaming over HTTP (DASH), one of the most commonly used HTTP adaptive streaming (HAS) protocols, has emerged as the preferred video streaming technique by adaptively maximising viewable bitrate from the server to enhance user QoE (Wainer et al., 2017). Client BS may analyse the network state of viewers and download relevant video chunks from the MPD file given by the server after deploying DASH on its infrastructure. In addition, transcoder can be installed either close to or inside client BS for transcoding tasks, which can considerably reduce network traffic throughout the core network and better respond to different circumstances of mobile users (Wang et al., 2018c). Client BS are also eligible for edge caching/computing technique. QoE of users can be improved by video edge caching on local client BSs to reduce latency in video transmission, as well as saving energy and computational resources (Li et al., 2017; Park and Song, 2018).
- **Transcoder/Transrater:** A *transcoder* is needed to meet the demand of video service users in multiple heterogeneous platforms and increase the accessibility for videos (Liu et al., 2020). Viewers may have different display platforms (e.g., computer and mobile devices), varied protocols and standards, and fluctuating network circumstances, all of which may differ from the intended requirements and cause video demonstration to be hampered. A transcoder receives encoded video chunks from the server, analyses user conditions, and decodes and re-encodes chunks depending on the condition analysis so that they are compatible with the receiver (Erfanian et al., 2021; Li et al., 2018).
  On the other hand, a *transrater* is used to lower the bitrate of video chunks while maintaining the same encoding standard. Transrating is beneficial for viewers who have intermittent network access and require lower bitrates for video delivery, ensuring user QoS/QoE (Zahran et al., 2017; Kim and Choi, 2019; Tran and Pompili, 2018; Jedari et al., 2020). Despite receiving fewer attention from researchers than transcoding, transrating remains a helpful option for distant users and emergency use-cases. In terms of aerial infrastructure for providing Internet access over suburban regions, the usage of transrating allows more customers with restricted Internet connectivity to utilise video streaming services with the highest QoS.
- **Decoder:** In contrast to the encoder, the decoder takes encoded video segments and reconstructs them in their original form. The integrity of video segments depend on which encoding method was used (lossly/lossless) and the stability of network connection (Jedari et al., 2020). Decoding can be computational resources and energy consuming, therefore having an efficient codec model in streaming service is critical for the transmission and circulation of video contents in the system. Yang et al. (2017) have stated the problem of low-bitrate video quality degradation in the HEVC model due to limited computational resources and proposed a decoder-side scalable convolution neural network (DSCNN) as a solution. Additionally, in Herglotz et al. (2017), Herglotz et al. proposed an extension standard of rate–distortion optimisation approach to cut down energy consumption in video decoding. The decoding energy can be saved up to 30% indicates the effectiveness of their solution.
- **Viewers:** End-users who see the video and utilise the associated services. Viewers may view videos via a variety of presenting platforms and UEs, including computer, mobile, tablet, and so on. There is a user-client BS interaction in which user devices transfer information about communication resource limits and user desires (i.e., resolution, particular content, and subtitles) to edge servers at BS for further analysis.

2.2. Aerial communication networks

The topic of deploying airborne BSs (ABSs) to support terrestrial BSs is becoming increasingly prevalent in the 5G/6G context. Its larger coverage area and independence from terrestrial buildings make it more versatile than traditional BSs, increasing its capability to deliver network services to a diverse range of customers (Ghanavi et al., 2018). The working principle of ABS is identical to traditional BS, as it has the ability to process network transmission signal, performing computational tasks, and distributing content to users. Technological advancement of aerial devices helps to open up the possibilities of performing computational acts in the field of telecommunication. In the previous decade, the most well-known airborne device which are capable on enabling of distributing Internet signal was the satellite of the LEO platform. However, as users are demanding higher requirements, satellites may not satisfy the requirement of fast response time from users and launching a satellite into the atmosphere was cost-burden, despite its large coverage. Hence, smaller devices such as drones, UAVs, airplanes equipped with computational capabilities were brought into use.

The integration of ABSs (LEO, HAP, and LAP) and ground BSs are commonly known as *space–air–ground* communication network. This implementation of multi-tier platforms allows them to collaborate in order to maximise their own benefits whilst minimising their limits. The report of Wang et al. (2021a) have demonstrated the ambitious vision on space–air–ground networks in 6G wireless environment. Currently, there have been a variety of large-scale projects such as O3b (SES) (SES, 2022) and Starlink (SpaceX) (Starlink, 2022), aim to optimise the LEO satellite constellation to shorten the round-trip transmission time from ground users to satellite BS, provide cheaper services, and improve storage capacity. It reveals that large electronics and telecommunications firms value the significance of LEO BS in supporting the Internet ecosystem. Furthermore, the use of space–air–ground offers up new opportunity for potential industries such as automated vehicles. In Niu et al. (2020), the authors suggested that with the use of effective RA strategy and AI approaches, autonomous cars can take advantages of aerial network resources to provide a real-time, reliable, and secure connection for exchanging data among vehicles, thereby enhancing the QoS of autonomous driving and Internet-of-Vehicles applications. Similarly, Sheng et al. (2021) recognised the potential of space–air–ground communication paradigm with AI for offering a seamless information services for High-speed Railways transportation. Aside from employing

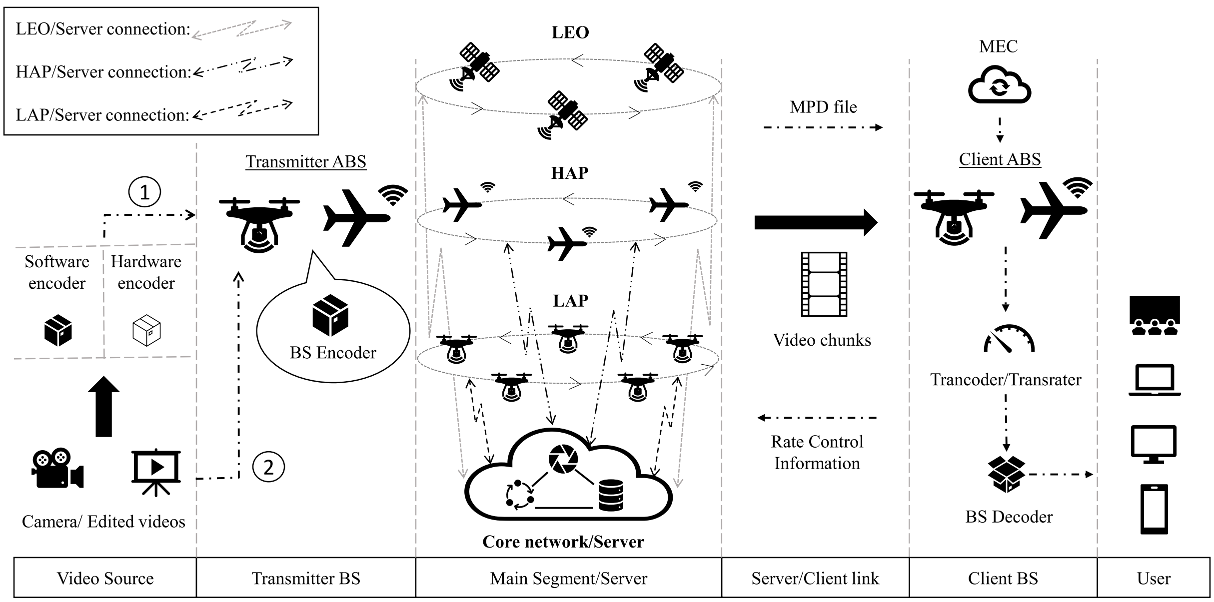

Fig. 3. System architecture of video streaming service over aerial infrastructure.

space–air–ground, it can be demonstrated that combining with AI technologies may enhance system performance.

Some recent works on developing ABS infrastructure were described in Cicek et al. (2019), where the authors surveyed existing studies on location optimisation of UAV-BS. They also proposed a general form of mathematical formulation of flying base station (FBS) localisation problems. In these problems, reliability of services provided by FBS and energy efficiency are the major concerns. Besides, there is still ability for better improvement on the localisation algorithms. For instance, Zhan and Huang (2020a) proposed an UAV scheme used as a ABS, providing connection to ground users with DASH to resolve the energy efficient problem. The result is, their proposed scheme have achieved better performance in terms of energy efficiency while maintaining QoE, compared to standard schemes. However, they only focus on the scenario of one UAV verses multiple GUs. The article Shakeri et al. (2019) introduced a prototype of cross-tier multi-UAV system and surveyed its current challenges in the implementation phase, mainly in five problems, coverage area, trajectory, computing/analysing capability, network design, and flight control. Despite being comprehensively giving a fairly comprehensive and detailed view of the system design, the article mainly focused on UAV-based video transmission, instead of video streaming service from independent providers to users.

2.3. AVS system architecture

Combining with the knowledge of the two previous sections, a complete system architecture of video streaming service over aerial infrastructure is presented as in Fig. 3. As demonstrated in the figure, some similarity to Fig. 2 can be spotted, such as the end-users of video source and viewers. Nonetheless, the mainstream components consist of transmitter BS, server, and client BS has been replaced by airborne devices which play a role as ABSs, similar to the work of Ferranti et al. (2020). Despite its replacement, their main functionalities are unchanged compared to terrestrial BSs. A detail description of the system working principle is given below.

Firstly, the video is recorded via a camera or a phone for live streaming, or being saved on user domain for edited video. After that, it is transferred to an encoder for video encoding. User have the option to choose where to encode video, it is either on their UE (route 1 in Fig. 3), or via BS encoder located inside the transmitter BS (route 2 in Fig. 3). For the first option, software encoder and hardware encoder are both available to be installed on user device. Software encoder often comes with streaming-assisted application, free and open source, offering user with easier operation process and use UE computational resources for their tasks. This is also the most notable vulnerability of software encoder, since they are creating resource burden for UE, hence degrading the performance of the device. Some of the most common software encoder are Streamlabs OBS, AWS Elemental MediaLive, and OBS Studio (Google Help, 2021). On the other hand, hardware encoder is represented as an individual device, can be separately connected to UE and use its own resources for performing tasks. Hardware encoder is capable of complying better encoding tasks, suitable for professional video content providers or limited resources UEs. On the negative side, hardware encoders are more complex for installation, payment required, and low mobility. AirServer, AWS Elemental Live, LiveU Solo are the examples of the most common hardware encoders being used nowadays (Google Help, 2021). The latter possibility is to upload raw video towards transmitter ABS to take advantage of its encoder. As ABSs are being equipped with better computational resources and abilities, they are capable of performing more tasks and providing various services for users, which includes encoding video. Despite various encoding approaches, they must comply to an uniform codec standard throughout the system, some of the most well-known codec standards include x264, x265 (Cai et al., 2021), and HEVC (Laitinen et al., 2020; Wieckowski et al., 2021).

The encoded video is complied by transmitter ABSs before being forwarded to the core network for chunk slicing. At this step, the working principle of the server is the same as have been mentioned in Section 2.1.2, in which the process of sorting, packaging, MPD file extracting, and transmitting are all included. The main difference compare to traditional video streaming model, is the interaction between the server and the aerial infrastructure to assist the transmission of video. In particular, airborne devices are connected to the server via wireless link, as well as forming a wireless connection among other devices. This ensures a tightly end-to-end connected network from source to viewers with a diverse topology. It is worth noticing that the transmitter ABS and client ABS are a part of the entire platform. When there is a video retrieval request from a viewer, it is transmitted throughout multiple devices and reaches the server. Subsequently, the server sends back the MPD file, receiving rate information, and transmitting video to UE on the same route. Depending on the geographic location of viewers and the scale of the streaming that different platform is used to ensure a satisfaction QoS/QoE. Devices on different infrastructures can also connect to each other, creating a multi-tier system to expand the operating range. For example, taking into account of a streamer in Japan and a viewer in the Republic of Korea, it would be insufficient to rely only on LAP or HAP devices due to the deployment requirement of a large quantity. The alternative solution is to integrate the connectivity of LAP, HAP, and satellites from LEO platform, leveraging their distinct advantages to achieve the working efficiency of the system.

Lastly, the client ABS receives the video chunks from the server and perform the final operations before passing them to the end users. On ABS premise consists of transcoder/transrater and a decoder, modifying the codec standards and video bitrate to match with the UE environment. A detail worth mentioning is that the client ABS can perform as an edge server MEC for video edge caching and computing (Nguyen et al., 2023). The development of MEC near end-user has been a remarkable solution for cutting down the energy and latency needed to cache popular videos, greatly improve the QoE for users. Recently, UAV-enabled MEC has gained popularity since it gives users with high QoE due to the availability of a highly probable line-of-sight (LoS) communication link (Nasir, 2021; Zhou et al., 2019; Yang et al., 2019b; Sun et al., 2020b). However, with limited resources, the implementation of edge services on aerial device is in need of an optimal RA optimisation. After the video chunks has been cached, modified, and decoded, it is forwarded to viewers for display.

2.4. Summary and discussion

In this section, a fundamental working principle of video streaming has been provided to give interested readers a glimpse of the main components in AVS systems. A typical video streaming system consists of three parts: *transmitter* which handles the video from the source, *server* for managing the transmission of video, and *client* to which the video is sent for display. The three components are tightly connected and always maintain an interaction via the server. Then, the current situation of aerial infrastructure deployment in modern communication networks was given to demonstrate the possibility of utilising airborne components into the streaming framework. Subsequently, a complete architecture of a video streaming over aerial infrastructure was sketched in the third section. Overall, AVS represents a special case of streaming structures, where aerial relay networks are exploited to conveniently exchange video traffic.

3. Achievable utilities

3.1. Video resolution

Video resolution is one of the most important aspects which greatly affects the experience of users when utilising video streaming services. Longer video viewing duration, more diverse content genres; hence, improving video resolution has been considered one of the key growth directions of video streaming services. The evolution of mobile telecommunication and video transmission protocols, in addition with the surge of Internet users, has raised up the QoE and increased the demand of users for QoS. Resolution downgrade may negatively affect the perceived QoE of users (Asan et al., 2017), hence an enormous amount of attempts has been made to improve the QoE via video resolution. The emergence of HAS and ABR has brought many improvements to the service of video streaming, one of which is improving video resolution. The work of Wassermann et al. (2019) has evaluated how HAS has enhanced video streaming resolution on Youtube platform. By encoding video with multiple resolutions and adaptively providing the best option based on network conditions of users, instead of a uniform dynamic resolution, video streaming services can avoid video stalling and degradation in QoE of users. In the designed AVS, video resolution is formulated as traditional QoE utility for DASH streaming (Zhang et al., 2020). Furthermore, super-resolution has been adapted for highlighting the optimal achievements and to disrupt the tight link between network condition and QoE of users (Zhang et al., 2021a). Super-resolution is a method of *reconstructing* the resolution of video, which replaces the base or low resolution with higher ones when the network condition is allowed for the upgrade.

As discussed in Section 2, after receiving a video from the source, the video is being divided into N chunks at the server. It is assumed that each chunk n in N has the duration of $\mathbf{T}$ seconds, and the duration is similar on every chunk n. Additionally, the video is encoded with K different resolutions, (e.g 144p, 240p, 360p, and so on). Denote q_n $(q_n \in K)$ as the standard resolution, i.e., the resolution which have been requested from the client at the beginning. Likewise, q'_n represents the desire resolution which the client wants to upgrade to. Hence, the reconstruction decision of super-resolution is (q_n, q'_n). Since different UEs have varied computational capabilities, the reconstruction time is not constant. Let $t_n(q_n, q'_n)$ as the reconstruction time, which is the time to reconstruct the resolution of chunk n from the standard resolution q_n to the desired q'_n. Similarly, the standard reconstruction time, i.e., the estimation of $t_n(q_n, q'_n)$, is denoted as $\overline{\mathrm{t_n}}(q_n, q'_n)$, and this value can be retrieved from the reconstruction time of user graphic card.

The video resolution utilisation is basically built on the synthetic of three important metrics: *Average Video Quality*, *Average Quality Variations*, and *Average Rebuffers*. Depending on use cases and optimisation objectives, there may be a vary of additional metrics. The three aforementioned metrics can be calculated as follow (Zhang et al., 2020)

- **Average Video Quality**: It is equivalent to the average Peak Signal to Noise Ratio (PSNR) for all video chunks.

$$Q_1 = \frac{1}{N}\sum_{n=1}^{N} \mathrm{PSNR}(n, q_n, q'_n) \tag{1}$$

  in which the PSNR represents the PSNR of chunk n when being reconstructed from resolution q_n to q'_n. The higher PSNR, the better video quality it gets. It is worth noticing that the MPD file also records the PSNR for each chunk for the clients for enhancing their reconstruct decisions.
- **Average Quality Variations**: The dramatic changes in video resolution among chunks cause unpleasant experiences for users (e.g. vertiginous). Hence, the following quality variation should remain stable during display process

$$Q_2 = \frac{1}{N-1}\sum_{n=2}^{N} \left| f(q'_n) - f(q'_{n-1}) \right| \tag{2}$$

  in which $f(q'_n)$ represents the resolution to visual experience of user. The above formula measures the average resolution to visual experience difference between the two chunks in series. Let $w_{q'_n}$ and $h_{q'_n}$ denotes the width and height of the n-th chunk after being reconstructed to the desire resolution q'_n, the resolution to visual experience can be calculated as $f(q'_n) = \ln(w_{q'_n}.h_{q'_n})$.
- **Average Rebuffers**: The rebuffering time refers to the waiting time for the downloading or reconstructing process to be completed, and pushing the newly reconstructed chunk into the buffer for displaying. The average rebuffering is formulated as

$$Q_3 = \frac{1}{N}\sum_{n=1}^{N} \tau_n \tag{3}$$

  in which τ_n is the rebuffering time.

In summary, the video resolution utilising super-resolution aims to maximise the following objective

$$\arg\max(Q_1 - \alpha_1 Q_2 - \alpha_2 Q_3) \tag{4}$$

where α_1 and α_2 are positive factors.

In order for the super-resolution mechanism to expose its maximum advantages, a double-buffer has been applied to avoid the conflict

Table 3
Prime examples of video resolution studies.

Ref.	AI Approach	Propose framework	Motivation	Merits	Drawbacks
Dasari et al. (2020)	DL	PARSEC	360-degree video streaming can consume large amount of bandwidth but still resulting in low quality video.	By utilising client-side bandwidth and trains micro-DL-models instead of large ones, PARSEC can improve video resolution while dropping bandwidth.	The test mostly focused on the viewport prediction, and advanced wireless network (5G/6G) and the connection speed have not been investigated.
Chen et al. (2020)	DRL	SR360	The prediction of FoV in 360-degree video streaming may inaccurate, leads to a drop in QoS.	The decision of reconstructing video tile was made jointly with FoV prediction, bitrate allocation, and super-resolution enhancement.	Other QoE metrics such as latency, service availability, aerial possibility, has not been mentioned.
Maggiori et al. (2017)	DL	CNN Semantic Label	Recent studies on CNN semantic labelling did not correctly address the usage of its properties.	Proposed a CNN framework specialised on leveraging semantic labelling for high-resolution image analysis, which can study and combine each feature to give better result with less computing resources.	Smaller resolution with larger scale videos are needed to be tested with CNN. The need of simplicity for using CNN is also notable.
Li et al. (2019)	Dynamic CNN saliency estimation	Spatiotemporal Knowledge Distillation (SKD)	The shortage of computational resources in cameras and airborne devices make it difficult to employ saliency models.	A dynamic saliency estimate technique for aerial videos was presented using spatiotemporal knowledge distillation. The approach can function at high speed, particularly 28,738 FPS on GPU and 1,490.5 on CPU, whilst performing similar to other saliency models.	The investigation on video streaming service via aerial devices rather than device-based are questionable.
Bronzino et al. (2019)	ML	Composite prediction model	ISPs are struggling to interfere encrypted videos to evaluate their quality.	Existing works are added some extensions to improve the predicting accuracy on performance metrics of video. In addition, a practical example was brought in by testing on 66 residents over 16 months to assess the effectiveness of the model.	The adaptability and generalisation of the model are problematic. Developing real-time prediction models are also a promising future extension.
Chen et al. (2019)	CNN	PMCNN	CNN can hardly be applied for motion compensation and alternative solutions are insufficient.	CNN predicts blocks by using previous knowledge of neighbouring blocks. The difference of pixels is analysed to perform smooth presentation and bitrate was utilised for storage or transmission.	The work is lack of entropy coding and configurations. Super-resolution has not been investigated.

between downloading and reconstructing stage. In double-buffer, the client would download the specified chunks into the downloading buffer during playback. Meanwhile, when the reconstruction is completed, the super-resolution module pulls the downloaded chunk from the downloading buffer and puts it to the playback buffer. The following constraints are given during the double-buffer process and the objective (4) must be optimised subject to:

- The requesting time for n-th video chunk should start immediately after the downloading time of the previous chunk, $(n-1)$, has finished, and the downloading buffer is free.
- If chunk n has been downloaded in the downloading buffer and the reconstruction of chunk $(n-1)$ has completed, the reconstruction of chunk n can begin when the playback buffer is not full.
- The chunk n can only be displayed after its reconstruction phase and the playback of chunk $(n-1)$ has been completed.
- The rebuffering time τ_n before the playback of n chunk must equal to $t_n - t_{n-1} - \mathbf{T}$, in which t_n is the display time of chunk n.

Understanding the optimisation objective of video resolution, Table 3 lists several current studies resolving the problem of video resolution, particularly using AI solutions. Super-resolution has proved to be effective for improving video resolution on multimedia platform such as Youtube (Belmoukadam et al., 2020), and currently receiving many attention as a promising solution for advanced high-resolution videos. In Dasari et al. (2020), Dasari et al. resolved the video resolution problem in 360-degree video streaming by proposing PARSEC, a streaming system that learns to harmonise the bandwidth with user's computational resources. Working based on super-resolution concept, PARSEC has appointed out key weaknesses of super-resolution systems such as large deep learning models, slow inference rate, and variance in video's quality. To overcome these challenges, PARSEC integrates with traditional video encoder, compress even smaller size of video segments and trains micro-models with super-resolution technique on these chunks. The final results indicate PARSEC performed outstandingly with most of the stats surpassed the existing streaming system. However, the experiment on PARSEC was performed under 4G/LTE network and only considered 360-degree video streaming. Similar to Dasari et al. (2020), Chen et al. (2020) proposed to boost super-resolution in 360-degree video streaming. Due to limited network bandwidth that the maximum video quality is not guaranteed, Therefore, one solution is to trade off the computational resources from user's side to increase the bandwidth. By using the deep reinforcement learning (DRL) approach for bitrate prediction and resource allocation, low resolution video tiles can be boosted. As a result, the scheme scored a 30% increase in performance efficiency compared to other existing schemes.

Individual works on maximising resolution on video should also be considered, as it poses potential solutions for aerial infrastructure deployment. The integration between aerial related and non-related which can benefit video streaming service are in need of further investigation. In Maggiori et al. (2017), Maggiori et al. highlighted the trend of using DL approach, convolution neural network (CNN) in particular, in aerial high-resolution video analysing. However, the authors realised the utilisation of properties of dense semantic labelling is being underestimated, hence proposing a CNN framework to tackle this issue. Through experiments and studies on the constraints of semantic label, the proposed framework proved to be able to both learn and combine the result features to provide an effective image labelling which can outperformed some traditional CNN models. Nevertheless, the scheme is needed to be generalised and reduce the complexity of the deployment. The video resolution optimisation using CNN requires an immense amount of computational resources, which is one of the weakness of aerial devices such as UAVs. To effectively tackle the RA for high-resolution video analysing, deep CNN saliency models are needed. Therefore, Li et al. (2019) have introduced a dynamic saliency estimation approach for aerial videos via spatiotemporal knowledge distillation consists of five components. Firstly, the resolution of the video was degraded for the ease of rapid extraction, whilst spatial and temporal saliency are being separated. Secondly, distilling and encoding of the two saliencies are used to train the appropriate spatiotemporal model. Through experiments, the proposed framework have shown its efficiency by improving the video processing time to 28,738 FPS on GPU and 1490.5 FPS on CPU, and still being able to main the high resolution of video.

In AVS perspective, especially in LAP-ABS devices, the limited resources require the system to have an efficient video encoder, which encode and encrypt video into a suitable size for transmission without having to spend many computing resources. However, after the video has been encrypted, it is challenging to the system itself to evaluate the integrity and quality of the video. To tackle this issue, Bronzino et al. (2019) proposed improved video streaming models which help the Internet service providers to infer the quality metrics of encrypted streaming services. The two main metrics being evaluated are startup delay and resolution of video. The work is an improvement from previous works, which resolved several contemporary issues such as platform flexibility, more detailed evaluate features, and inference precision. Models are supervised ML approaches, placed on network layer, transport layer, and application layer. The result have shown that the models are more robust to different platforms, more precise in predicting delays and resolution, and higher accuracy in quality prediction. They also found, surprisingly, higher Internet speed has only a small impact on video quality improvement. For future challenges, their models are bias, meaning they are vulnerable to new Internet services. Also the real-time prediction has not been considered. The compression of video also affects the resolution, as if the compression was not efficient, the decoding could potentially lost frames, leads to failed reconstruction. In Chen et al. (2019), the authors have modelled a spatiotemporal coherence based on PixelMotionCNN (PMCNN). Firstly, the framework tries to predict the blocks in a video segment using the frame knowledge from the previous block using CNN. Subsequently, it compares the pixel differences between two blocks in addition with bitrate to perform a smooth video display for users. The results show the model has similar performance with advanced codec H.264 and even superior than MPEG-2.

3.2. End-to-end latency

Latency is another important QoE metric in which users commonly prioritise when using video streaming services. Video latency denotes the time when a video is being processed until it is being displayed for user (Shuai and Herfet, 2018). Due to increasing number of video service users, low latency has always been the key issue for video service providers. There are an enormous amount of work to improve latency, most noticeable of which is the use of content delivery network (CDN) (Al-Abbasi et al., 2019). Leveraging its proximity to users, videos are being cached and stored to satisfy the demand of high responsive videos. In Luglio et al. (2019), a satellite–terrestrial integrated framework has been proposed to enhance CDN, expanding its coverage and increase the performance of services utilising it in 5G environment. Furthermore, edge technique (Yang et al., 2019a) and fog–cloud computing (Veillon et al., 2019) are also considered as potential options to reduce latency in video streaming. In addition with the emerging RL approaches, ABR algorithm promises better solutions to resolve the latency challenge (Zhao et al., 2019b).

Satellites constellations are used to be considered as obsolete and not suitable for modern latency-sensitive communication. However, with the release of VR/AR (Elbamby et al., 2018) and high-quality video streaming, the backhauling and high computational capacity of satellites are once again attracted attention from ISPs and researchers. In order to assure the performance of LEO and satellites in providing video streaming services is comparable to LAP and HAP, there has been an extensive work on studying the latency issue. The work of Ge et al. (2019), Wang et al. (2018b) are examples, in which the authors have analysed the QoE of HTTP-based video streaming in 5G using satellite communication. Satellite links are being used as backhaul for 5G core network, and MEC servers were brought in to reduce latency. The third-party factors like content providers are benefit from MEC by enhancing inadequate Transmission Control Protocol (TCP) performance through satellite links. Furthermore, the framework ensures the video chunks are being transmitted from the video source to UE via one unify multicast-based flow, guarantee the delivery phase for high-quality video.

Using the similar system model as in 3.1, let T_{trans} denotes the transmission duration of a video throughout the aerial infrastructure, starts when the encoded chunk transmitted from the server, to the closest ABS near user. Assuming the video is being encoded on user premise, the encode time is exempted from the codec interval and only consider the decoding time τ. Lastly is the video buffering time T_B. The latency objective is minimal function of the summation of the three variables

$$\begin{aligned} &\min \quad (T_{trans} + \tau + T_B) \\ &\text{where} \quad T_{trans} = \frac{\zeta_n}{\sigma} \\ &\qquad\quad\;\; T_B = \frac{\beta_{n-1} c}{f} \end{aligned} \tag{5}$$

in which the decoding time τ is a positive variable and can be different depending on the codec protocol used in the system. The transmission time T_{trans} includes ζ_n is the size of n-th chunk in bits and σ is the data transmission rate of the channel between video server and user in bits/s. Finally is the buffer time T_B in which β_{n-1} is the buffer size of the unprocessed previous chunk (bit), c is the channel complexity (Hz/bit), and f is the CPU frequency of the processing node (Hz/s).

Among the three aerial platforms, as has been mentioned in Section 1.1, LAP and its airborne devices are the most optimal option in latency perspective thanks to their close proximity to users. The authors in Sacoto-Martins et al. (2020) have realised the practicality of the UAV-based video streaming and have implemented an actual prototype. Another simulation study of Stornig et al. (2021) demonstrated ns3 simulation results of their study on the quality and latency of video streaming services over UAV infrastructure. Bigazzi et al. (2021) focused on the latency issue on UAV-based video streaming system affecting video quality. According to them, the lack of computational resources led to the limitation of frame rate, costing more time for frame features extraction. Besides simulation studies mentioned above, some of the recent studies using AI approaches for resolving the latency problem, both in UAVs signal communication latency and video streaming latency, have been surveyed in Table 4. Burhanuddin et al. (2022) have leveraged the usage of DRL to resolve the QoE of Live video streaming on UAV-BSs, which includes latency. In particular, a centralised UAV-BS scheme was considered to stream videos from other

Table 4
Prime examples of end-to-end latency studies.

Ref.	AI Approach	Propose framework	Motivation	Merits	Drawbacks
Burhanuddin et al. (2022)	DRL	UAV-to-UAV communication	Real-time video streaming are latency-sensitive, especially in disaster surveillance.	Latency with other QoS aspect were jointly formulated. Two DRL algorithms, AC and DQN, was used to resolved the problem. High resolution streaming can be obtained whilst maintaining low delays.	AC can be more complex and requires many parameters, DQN is more simple but not perform as good as AC.
Jiang and Li (2020)	RL	Peer and cache node selection	Latency in satellite communication are insufficient.	A selective relay system was designed to optimised latency for uplink, whilst optimal cache placement strategy were used for downlink.	The work has not mentioned on the application of video streaming.
Wang et al. (2020b)	Multi-agent DRL	MacoCache	Different edge dynamics can negatively affects the edge caching process.	Each edge acts as an agent, and multiple agents need to collaborate to achieve optimal caching decision. MacoCache integrates long short term memory network with AC to process time series.	Active content selection, online updates, scalability are some major drawbacks.
Sun et al. (2021)	MPC-Live and DRL-live	iLQR	Live video streaming is extremely sensitive to latency to maintain QoE for users.	DRL-Live and MPC-Live has achieved near optimal performance by reducing latency range to 2–5 s.	The experiment was implemented on client side, meaning the privacy issue needed for consideration.
Wang et al. (2019a)	RL	BitLat	The heterogeneous network conditions degrade QoE with latency.	BitLat increases adaptability for video streaming service and QoE around 20%–62% by leveraging RL to precisely predict and reduce delays.	BitLat can be overlapsed with the function of transrater/transcoder.
Pang et al. (2019)	DRL	iView	Recent works on improving 360-degree QoS are lack or did not consider different features.	Viewing prediction and video tiles optimisation was jointly resolved using multimodal DRL. DNN explores features, forming a correlation between features and objectives.	Practical implementation and multi-user use case are still under studying.
Perfecto et al. (2020)	DL	VR Multicast	Limited works towards putting reliable and fast-response limits on such wireless VR service problems.	The problem of proactive physical-layer multicast transmission was resolved with stochastic and game theory. A DRNN was used on VR headset for FoV prediction.	The utilisation of multiple ML algorithms may create large resource burden for the system.

UAV users, which requires a low delay and consistent transmission. The problem of long-term QoE optimisation was firstly formulated, then Actor Critic (AC) and Deep Q Network (DQN), two DRL algorithms, was utilised to resolved this NP-hardness problem. The bitrate selection was formed as Markov Decision Process (MDP), and DRL minimise the delay penalty of the system. AC achieved the best latency, as it can maintain a stream of minimum 1080p resolution with lowest latency.

One suggestion of reducing delay in video transmission, is to use caching technique at edge networks. Jiang and Li (2020) noticed the delay issue in satellite communication, hence proposing a caching strategy, in which the system attempt to optimise the cache placement to reduce latency for uplink and downlink. In particular, the system selects peers according to greedy algorithm to find the best caching node for uplink, whilst for downlink, optimal cache placement are search in respect to lowest latency. The simulation results show that the caching strategy proposed for downlink has similar performance as the optimal, and both uplink and downlink latency are being reduced. Another work from Wang et al. (2020b) investigated the application of DRL approach in intelligent video caching. An intelligent caching model called MacoCache was proposed to reduce the limitations for cooperative caching causing by the similarity of requests among nearby edges. The diversity of edge environments is adapted by multi-agent DRL, and AC was brought in to evaluate the caching actions. An interesting point is that each agent needs to consult and collaborate with each other, hence to provide optimal decisions and policies. When implemented in real use cases, MacoCache can reduce the latency and cost of the system by 21% and 26%, respectively.

Along with approaches which directly resolve the latency issue on aerial network, other works on traditional video streaming framework are also a potential source that can be applied into aerial scheme. The work of Sun et al. (2021) resolved the latency issue in live video streaming by carefully investigating the streaming model and took QoE metrics as core assets. The authors then proposed a live streaming algorithms based on the iterative Linear Quadratic Regulator (iLQR) using Model Predictive Control (MPC) and DRL approaches. The goal is to enhance QoE of users by adaptively retrieving the video bitrate whilst minimising the latency of the network. The results reviewed that the proposed schemes can reduce the latency to just 2–5 s, and the two promising strategies for achieving a high degree of user QoE in low-latency live streaming are chunk-based packaging/streaming and playback pace adaptation. Another work leveraging RL to reduce latency from Wang et al. (2019a), in which they introduced BitLat scheme for controlling the bitrate and latency of video. By using the reward function of RL algorithm, BitLat tries to distribute the bitrate of video chunks adaptively based on the network condition of users, therefore limit the possible delays during video transmission.

AVS system can take advantage from existing solutions for latency-sensitive streaming such as 360-degree video. The viewpoint in VR/AR video applications are critical because it directly affects the QoE. Real-time updates of viewpoints are expected by users, and video stalling or delayed updates will degrade the overall experiences. In Pang et al.

Table 5
Prime examples of energy efficiency studies.

Ref.	AI Approach	Propose framework	Motivation	Merits	Drawbacks
Zhang et al. (2021c)	safe-DQN	CMDP	Ensuring QoS for UAV video streaming service.	The proposed SDA enhanced the long-term energy efficiency whilst satisfying the secrecy timeout probability constraints.	The deployment of multiple UAVs is needed for further investigation.
Do et al. (2021)	DL	UAV surveillance	Video generated by UAVs are large, hence energy consuming and causing delays in transmission.	CNN extracts primarily targets on each frame, cut down redundant scenes and transmit it instead of sending the whole frame. UAVs can distribute tasks and share information to reduce workload.	The possibility of wrong object detection should be minimised, in addition with the use case in complex environments.
Zheng et al. (2020)	DRL	LSTMAE-DDPG	MEC can assist the immense energy constraint of VR streaming.	DRL algorithm minimised the tradeoff between latency and energy, as well as maximise the resource utilisation.	The framework requires a large computational resource capability for streaming system.
Liu et al. (2019b)	DRL	Predictive RA policy	To avoid video stalling due to video segments download, transmitting power needed to be increased.	The policy for RA prediction is determined using DRL to optimise the average energy consumption with respect to QoS constraints. The prediction accomplished in two different time frames, reducing communication overhead.	The topology of the experiment was a centralised network with BSs connecting with one central unit, which can be fatal if errors exist in this unit.
Liu et al. (2018)	DRL	DRL-EC^3	UAV-BSs are intolerable for long operations due to its limitations on energy storage.	The proposed solution optimises an innovative energy saving function while maintaining effective and equitable communication coverage and network connection.	A decentralised framework should be considered in case of unexpected errors at server, and the scalability needs more investigation.
Zhao et al. (2020)	DRL	DeepCA	High dynamics, limited resources, and a shortage of efficient energy solutions for LEO devices.	DeepCA introduced a sliding block scheme to model the dynamic feature of satellites. The channel allocation problem was formulated as MDP and solved with DRL algorithm.	For future extension, a battery model with the account of battery load and constraints can be implemented.

(2019), the authors presented iView, a latency-aware multi viewpoint system, integrates with DRL to efficiently predicts the next viewpoint based on the head movements of users, hence providing video tiles for display in a time manner. Likewise, Perfecto et al. (2020) used DL attempt to resolve the problem of proactive FoV-centric millimeter wave (mmWave) physical-layer multicast transmission in 360-degree video. The problem was then formulated as frame quality optimising under delay and stability constraints, divided into subproblems, and undertaken by a game theory approach. Furthermore, DRNN was integrated into VR headset for generating and analysing dataset from head movement tracking. As a result, the VR frame delay was cut down to at least 12% whilst maintaining high resolution HD rates above 98%. Finally, implementing strict latency restrictions reduces the delay-tail, as seen by 13% shorter delays in the 99th percentile.

3.3. Energy efficiency

The energy consumption for video streaming over aerial infrastructure consists of two main parts: the computational energy and the hovering energy. As ABSs are being deployed for its vast coverage area, the energy to keep the devices from operating on the air infrastructure is enormous. Unlike terrestrial BSs where the provision of energy is rather easier with ground wires, aerial devices operate on limited energy resources, and when the power source is exhausted, those devices need to be temporarily shut down for maintenance and recharging. This may cause interruptions during the data transmission process, leading to the decreasing of QoS. Furthermore, the emerging video data traffic with the increasing QoE require ABS to have higher computational capacity, hence creating more energy burden. For LEO and HAP platforms, the problem of energy consumption is less as a threat since satellites in LEO can leverage the unlimited solar energy source, whilst airplanes in HAP were created with huge energy capacity so that they are durable for long operation. The complication of energy seems to hit UAVs and other LAP devices the most, due to their small design compare to HAP and LEO, therefore unable to have large energy storage. To resolve the energy problem for UAVs, some approaches can be noted as using tethered connection (Kishk et al., 2020a,b), efficient caching (Zhang et al., 2017; Li et al., 2020b), or better resource allocation strategies (Yu et al., 2020; Zhan and Huang, 2020a; Zhan et al., 2019). It is worth mentioning that AI and ML methods have made a big contribution to the development of these ideas and demonstrated an outstanding performance.

In term of energy efficiency optimisation, AVS considers multiple airborne devices, in particular D devices, each device d ($d \in \{1, \dots, D\}$) is serving a number of G ground users. Assuming device d is flying at speed V and communicate with users whilst hovering. The time duration for device d to simultaneously communicating with ground users is T, divided into Q time slots, each slot $q \in 1, \dots, Q$ have equal length of ϖ. Firstly, the total amount of energy required for device d for travelling and flying, denote as E_{tr}, is calculated as in Zeng et al. (2019)

$$E_{tr} = E_0 S_{tr}(G) \tag{6}$$

in which $S_{tr}(G)$ is the total distance in meters, on which device d is required to travel to efficiently communicate with every ground user. E_0 is the energy consumption of the device per distance unit with velocity

$$E_0 \triangleq \frac{P(V)}{V} \tag{7}$$

where $P(V)$ is the propulsion power consumption in J/s unit and V is the velocity of device d in m/s. It is noted that E_0 is not represent energy, but rather energy per distance travelled, hence the unit of E_0 is J/m. Next, the device d maintains communication links and simultaneously transmitting video data to G ground users over the T time interval. The communication energy, or streaming energy, is derived as in Zhan and Huang (2020b), Liu et al. (2021), i.e.,

$$E_{co} = \sum_{q=1}^{Q} \sum_{g=1}^{G} p_g\,[q]\,\varpi + T\,P_b \tag{8}$$

in which $p_g[q]$ is the power transmitted to user g at time slot q, P_b is the amount of power used by the baseband and radio frequency circuits, as well as the cooling and power supply. The energy utility objective for AVS would be

$$\begin{aligned}\min_{p_g[q],x_g[q]} \quad & \sum_{d=1}^{D} E_{tr} + E_{co} \\ \text{s.t.} \quad & E_{tr} + E_{co} \leqslant E_{\max}, \\ & \sum_{g=1}^{G} p_g[q] \leqslant P_{\max}, \forall q, \\ & \sum_{g=1}^{G} x_g[q] \leqslant 1, \forall q, \\ & 0 \leqslant x_g[q] \leqslant 1, \forall g, q\end{aligned} \tag{9}$$

which is to minimise the total energy needed for AVS operation, following the constraints about the transmitting power $p_g[q]$ and the bandwidth allocation $x_g[q]$.

As has been previously mentioned, the issue of energy efficiency certainly impact mostly on LAP platform, hence there has been a dominant number of research on it compares to the other two platforms. Table 5 have listed some researches leveraging AI approaches to handle the energy problem. The recent advances of AI models, in particularly DL and DRL, can be a solution for AVS, as video data can be preprocessed prior to the transmission at ABSs. An example from the work of Zhang et al. (2021c), in which the authors have addressed the issue of energy efficiency in wireless UAVs by improving video level selection and power allocation simultaneously. The answer was to represent the problem as a constrained MDP and use DQN, which induces a set of safety regulations by building a Lyapunov function. As a consequence, the suggested algorithm functioned admirably and with excellent energy efficiency. The future development involves the deployment of several UAVs and macro BSs in the system, as well as the collaborative consideration of UAV path planning. Another work which AI was used for UAV surveillance from Do et al. (2021), in which the authors have highlighted the transmission of an enormous amount of video data generated from surveillance UAVs to server may drain out the energy storage. The proposed solution is to leverage AI, particularly CNN technique, to analyse the frame in video, extracting and sending moving objects only. By this attempt, instead of having to send the entire frames which can be redundant, the video system are able to send critical data with less capacity and energy needed. The UAVs can also form a decentralise connection to efficiently share and transmit data, increasing the latency and reduce workload for each other. Individual works on resolving the energy problem on each aspect, *video streaming* and *aerial infrastructure* are also worth to be mentioned. On the one hand, if the energy issue in traditional video streaming services can be resolved, the transition to an aerial infrastructure environment may be similar since the internal framework has been efficiently optimised. On the other hand, increasing energy storage of aerial devices means enhancing the endurance and operational duration of ABSs, therefore running services such as video streaming can also benefit from this.

Some of the work which utilised AI to increase energy efficient in traditional video streaming is Zheng et al. (2020), in which the minimisation problem of the tradeoff between energy and latency is formulated as MDP and being solved with DRL algorithm called long short-term memory auto-encoder deep deterministic policy gradient (LSTMAE-DDPG). Moreover, a joint dynamic caching-offloading strategy was introduced in order to minimise the possible resource waste. The numerical simulation results indicates that the proposed scheme is more comprehensive and achieved better energy efficiency. The work of Liu et al. (2019b) also used DRL approach for optimising the RA problem in mobile video streaming, in particular, maximise the average energy efficient subject to QoS constraints. The RA problem was formulated as RL problem, and the policy for predicting RA would be solved using deep deterministic policy gradient (DDPG). With this way, even when prediction errors occur, the RA predictive policy can still performs better than the prediction-based optimal policy. The two aforementioned works are compatible with large-scale video streaming system and has no effect on the transmission environment, hence the potential of employing these solution above the air is immense and completely feasible.

In contrast, articles focusing on increasing energy storage of aerial infrastructure includes the work of Liu et al. (2018) for UAVs-ABSs. As previously mentioned, UAVs are limited in energy resources and coverage area, hence it is impractical for them to hover for a long time or fly to cover the entire area. To this come, Liu et al. have proposed a DRL-based energy efficient control for coverage and connectivity (DRL-EC^3) framework. The method focuses on resolving the three key problem: maximising energy efficiency; learning the environment; and decision making using deep neural networks. Firstly, the scheme aims to reduce energy consumption in respect of coverage area, energy, connectivity, and fairness among users based on DPRG actor–critic method. Next, with consideration of the dynamics of UAVs and its environments, the decisions in DRL was made by consulting two deep neural networks (DNNs). For LEO platform, Zhao et al. (2020) investigated the energy problem at LEO satellites. Realised the lack of RA solutions for LEO satellites, the authors proposed a RL-based approach called DeepCA for energy efficient channel allocation. In DeepCA, a novel sliding block technique was offered for modelling LEO satellite feature. Additionally, the dynamic channel allocation problem was described as a MDP and being resolved using DRL algorithm. The size of action set and the user requests are also being minimised to speed up the learning process of DeepCA. Throughout simulation experiments, the results show that DeepCA can reduce the energy consumption to 67.86% when compared with other algorithms.

3.4. Service stability

In video streaming, service stability is used to demonstrate how steady the video service is being served to users. To achieve stability, the user should identify short-term bandwidth differences and avoid repeated short-term bitrate swaps (El Marai et al., 2017). Representative metrics of instability in a streaming service include video stalling, lagging, choppy audio, missing frames, etc. The video data traffic congestion and limited bandwidth are another factors causing the degradation in video service stability. Currently, the usage of DASH (Choi and Yoon, 2019; Huang et al., 2018; Zhou et al., 2017), with the advent of new techniques such as buffering (Huang et al., 2018; Kim et al., 2017) and transrating has dramatically improve the stability of video streaming service. Adaptive streaming and transrater targeting on balancing the video bitrate with the network condition of user, finding a tradeoff between QoS and QoE. Additionally, buffering creates a buffer between the actual playback time and the downloaded video segments to avoid interruptions during display. All of these efforts is to optimally adjusting different features of video streaming such that the QoE of users can achieve at least an acceptable rate. On aerial infrastructure perspective, the channel stability among ABSs are crucial to secure the transmission network from server to end users. Most of the channel connections among airborne devices are wireless, hence underlying some possible obstacles which interring the Line-of-Sight of these connections (e.g. geographical impediment and energy constraints). Overall, to evaluate the service stability of AVS, the two main elements to be discussed are the stability of streaming video and the cohesion between connection among ABSs.

It is assumed that the incoming video data rate are expected to be kept in a sufficient playout buffer at each user g. Let $r_g[q]$ denotes the video playback rate for user g at time slot q. The following information-causality constraints for video playing at each user was set up for preventing interruptions while playing videos

$$\sum_{m=1}^{q-1} f_g[m] R_g[m] \geq \sum_{m=2}^{q} r_g[m], \quad \forall g, \forall q = 2, \ldots, Q \tag{10}$$

Table 6
Prime examples of service stability studies.

Ref.	AI Approach	Propose framework	Motivation	Merits	Drawbacks
Bai et al. (2020)	RL	ABS Placement for minimum-delay	Large queuing delays can degrade the stability of low-buffer devices.	When the dynamics or statistical knowledge of ABSs are known, the placement problem can be resolved using backward induction, whilst RL is used when neither of them are available.	User mobility and scalability can be improved. DRL approaches are possible alternative solutions for trajectory.
Hao et al. (2020)	DRL	DDQN	Large-scale ABS with mobility from users make RL solutions less effective due to dimensionality.	The trajectory problem with different mobility constraints was resolved using DDQN, improving convergence speed and stability. The adaptability of ABS also increased.	The constraint of energy is being questionable, in addition with only one ABS are being considered in this work.
Luo et al. (2019)	DRL	Mobile-MEC video streaming	Energy and QoE enhancement for video streaming over MEC-Mobile network.	The problem was formulated as two MDP problems and resolved using A3C-DRL algorithm. Energy and QoE are improved compared to traditional frameworks.	The original problem was put into subproblems without constraints, hence the results may not be the most optimal.
Cui et al. (2020)	DRL	TCLiVi	Live video streaming is vulnerable to instability, which can directly decrease QoE of users.	TCLiVi uses DRL optimise four QoE parameters by adjusting bitrate and buffer level. Compare to other schemes, TCLiVi raises the QoE upto 40.84%.	The scalability has not been investigated and the improvements on video quality has not been clearly highlighted.
Zhou et al. (2020a)	DRL	HEVC Rate Control	Spatio-temporal HEVC rate control solutions are ineffective in fast moving video scenes.	The rate control was defined as MDP, CNN was used to find optimal parameters, and finally A3C was utilised to solve the problem. Video can have better quality, less fluctuations and buffer rate.	Users with different codec other than HEVC or new codecs may not be included in the scope of this research.
Wei et al. (2021)	RL	DASH Adaptive Bitrate	Rapid changes in video sequences make it challenging for stability.	Bitrate selection problem was formed as MDP and RL was used as a solution. Aiming for maximising quality and buffer level.	The scheme does not consider the constraint of energy and latency issues.

where $f_g[q]$ is the time fraction for ABS device d communicating with user g at time slot q and $R_g[q]$ is the achievable rate between the device d and user g.

In Eq. (10), the left-hand part shows the quantity of video data that has been acquired by user g from the ABS d, while the right-hand side indicates the data that has been played at the same time with acquiring data. It is satisfy the condition that the videos are needed to be downloaded at least the same amount with the displayed video. In practice, the UE always download an amount of video segment and display only a part of it, whilst the rest are called *buffer*, to avoid video stalling due to video downloading process. The video processing delay time (e.g. encoding–decoding) are often be considered as one time slot q worthy, and can be neglected if the total time slots Q are sufficiently large.

In addition, the degradation in QoS of video streaming system can be caused by the instability of video playback rate. Hence, the following constraint was introduced

$$\left|R_g[q] - \overline{R}_g[q]\right|^2 \le \lambda_g, \quad \forall g, q \tag{11}$$

in which the left part indicates the video quality variance for user g at time slot q, formulated as the square difference between the current video playback rate (a.k.a achievable rate) and the average transmission rate over time. The $\overline{R}_g[q]$ can be calculated by $\overline{R}_g[q] \triangleq \frac{\varpi \sum_{q=1}^{Q} R_g[q]}{Q}$. This variance must be less or equal to the maximum tolerable video quality variation λ_g.

With respect of the two aforementioned constraints, the service stability optimisation problem for AVS can be formulated as:

$$\begin{aligned} \max_{R_g[q], r_g[q], f_g[q]} \quad & T\sigma - \chi_1 \sum_{n=1}^{N} \tau_n - \chi_2 latency - \chi_3 switching \\ \text{s.t.} \quad & ((5), (9), (10), (11)) \end{aligned} \tag{12}$$

in which χ_1, χ_2 and χ_3 are the weights of rebuffering, latency and instability respectively.

The high mobility with different working environment at different altitudes have made the wireless communication among ABSs become unstable and progressively complex to maintain transmission stability (Amer et al., 2020; Wang et al., 2019b). Therefore, in order to enhance service communication, one of the most popular approach is optimising ABSs trajectory schemes to adaptively adjust their position, improving video transmission stability. The spectrum efficiency and the communication links, the two factors which largely affect the service stability, can be improved by decreasing the distance between ABS and user. Several studies on using AI to optimise the ABS trajectory and localisation have been carefully surveyed at Table 6 based on their framework, merits, and drawbacks. In addition, the service stability on video streaming represents in the buffer and quality changes. Therefore, many researches also used AI to increase the buffer and reduce the frequent change in video display. The work of Bai et al. (2020) aims to reduce the queuing delay by placing ABSs locations with respect of energy constraints. The problem is formulated as CMDP under three different scenarios: only the knowledge about wireless dynamics are fully know, only the statistical data available, and lastly is both of the knowledge are anonymous. For the first two cases, the authors have used the backward induction technique, whilst RL approach is leveraged for the last case. The framework was being compared with CSI-Only and MaxWeight schemes for performance evaluation, and has proved to be able to achieved better delay cut down. Nevertheless, with the mobility of users, it is questionable for the solution to be efficiently function. The work of Hao et al. (2020) would be a suitable supplement solution for the previous work. Particularly, Hao et al. realised the instability of RL approaches caused by the random movement of users, hence proposing DRL framework instead. The optimising goal for ABS trajectory is to maximise the sum-rate for users. The proposed double deep DQN has proved to be better in large ABS scale, where the dynamics of environments are learnt faster, shorter convergence and more stable than other schemes.

For improving the stability of video streaming service, Luo et al. (2019) considered a video streaming service in MEC-mobile integrated network and tried to jointly optimise both energy consumption and QoE, in specifically, stability. Similar to Bai et al. (2020), the instability of video streaming come from long queuing delay and short buffer. Therefore, the time-varying channel is formulated into two CMDP and MDP subproblems, and resolve them using asynchronous advantage actor–critic (A3C), which is one of the model-free DRL algorithm. Moreover, for tracking the fluctuation of video segment bitrate, a time average instability index was presented, therefore the system can monitor and adjusting to maintain stability. The simulation results from the experiment show that the instability index have decreased dramatically, increasing the overall QoE for users. Another work from Cui et al. (2020), where the problem of stability also lying on the bitrate and buffer size. A framework called TCLiVi was proposed utilising DRL to enhance QoE in live streaming with respect of video quality, stability, rebuffering and latency parameters. TCLiVi can observe the transmission dynamic depending on bitrate and buffer data, then combining with DNN and A3C for balancing the four parameters. Compared to DoubleDQN, MPC, and Buffer based schemes, TCLiVi outperforms and provide enhancement to live video streaming stability.

It can be seen that rate control plays an essential role in stabilising AVS, since the enormous video data traffic can exceed the bandwidth limitations of ground users. The transmission/download rate needed to be constant and improve when the network condition is allowed, hence rising a demand for efficient rate control. In Zhou et al. (2020a), the problem of nowadays rate control algorithms in HEVC codec is their dependence on spatio-temporal technique, which is vulnerable with the fluctuation of video scenes. To stable the video sequence, a DRL rate control method has been proposed to enhance coding phase. Similar to the above research, the problem was formulated as MDP, then being resolve with CNN and A3C in DRL. The performance evaluation indicates that the proposed framework has improved rate control, making video quality more stable, and buffer occupancy is minimised. An important discovery by Sunny et al. (2019) where they have spotted the instability of DASH was caused by the prolonged oscillations of user bandwidth and dynamics, therefore in Wei et al. (2021), a RL-based solution for DASH adaptive bitrate for enhancing QoE was proposed. Likewise, the bitrate selection problem was firstly formulated as MDP. Next, a RL approach with clients as RL agent was brought in with enhanced learning stability. Finally, the suggested algorithm is incorporated in the DASH framework with a focus on buffer and video quality.

3.5. Computation efficiency

Continuous video transmission over aerial infrastructures strains the wireless access infrastructures severely. In Wang et al. (2018a), it can be seen that live video streaming with high resolution are extremely computational-intensive and the need of latency may not be guaranteed. The constraint of processing capabilities, along with the high volume of video transmission, necessitates AVS to efficiently utilising computational resources. LAP devices are the most noteworthy airborne platform that suffers with inadequate RA. Similar energy restrictions, the compact nature and limited on-board storage of LAP devices makes it difficult to provide them with vast computing resources, hence RA is an important criterion for them (Kyrkou and Theocharides, 2020). The compute efficiency problem is referred to as the RA problem, implying that the resources in ABSs are fixed and that the utilisation should be optimised. However, the introduction of edge/cloud technology has increased options to enhance computing resources, provide more computational capabilities, and even improve AVS service performance (Liu et al., 2021b).

Table 7 illustrates studies working on reducing computational burden for ABS and video streaming. It can be seen that DRL is rising as the most prominent AI approach and can be applied in many different frameworks. The work of Hu et al. (2021) investigated the possibility of transmitting 360-degree VR video via UAV devices. Live video and AR/VR streaming are extremely computational expensive and latency-sensitive, especially in the scenario of large quantity of ground users. Therefore, aerial infrastructure needed to efficiently allocating resources in order to broadcast it. Leveraging cell-free multi-group broadcast (CF-MB), a VR video transmission with QoE optimised was deployed for sport event scenario. The problem of video tiles scheduling and dynamical association was firstly highlighted as the two aspects for RA in VR streaming. A centralised multi-agent DRL approach, in particular CNN, was given to generate decisions for resolving the first criteria, which have been transformed into semi-MDP problem, then a hierarchical algorithm was used to determined the optimal results for both aspects. For the latter one, the problem was formulated into subproblems using a networked-distributed Partially Observable MDP, and the solution was given using DRL. Another work investigating on AR from Chen and Liu (2021), where the authors studied the RA for AR in both single-MEC and multi-MEC. The optimisation problem of computing resources was formulated as a mixed multiuser competition and cooperation problem, and a DRL algorithm of multiagent deep deterministic policy gradient (MADDPG) was proposed to solve it, with consideration on its NP-hardness and dynamic environment.

Nowadays, edge networks are being mentioned as the potential resource which the system can leverage instead of using the computational capability of ABSs. These edge networks can be edge/cloud servers or user devices, and video streaming service can share their tasks to be executed on these edge to loosen the computing burden. This article (Liu et al., 2019a) solves the problem of computation efficiency at transcoding phase for blockchain-based MEC video streaming system by leveraging the edge computational resources. They distributed small BSs in a distributed and secure manner for computational and communications allocation. Improvement on operation efficiency were also discussed by using a series of smart contracts to a decentralised controller for video transcoding and delivery. The transcoding and delivery problem was formulated as a three-stage Stackelberg game. Simulation results show that the proposed approach could obtain the good performance in terms of average time to finality (TTF), average access delay and network cost. The work of Guo et al. (2020) has introduced a solution for the proposed video streaming scheme as shown in Fig. 3, which is to deploy transcoder at edge network near end-user. This is considered as a new approach to replace traditional video streaming system, where videos were fetched and transcoded remotely from the video server. By leveraging the close proximity and computing ability of edge server (or client BS), QoS of the system can be dramatically improved. The work considered an ABR streaming framework with MEC and enabling video transcoder at edge network, formulated the resource and video quality optimisation problem as stochastic, and proposing a DRL approach to solve the issue. The goal of the algorithm is to maximise the QoE whilst minimise the transcoding cost at edge server without having to have knowledge about network condition. Despite the results indicate the effectiveness of the proposed scheme, the authors also found out DRL approach are occasionally suboptimal in some special cases.

Nevertheless, in extreme remote areas where the edge network infrastructure are limited, the utilisation of edge/cloud servers can be limited. In those scenarios, ABS can also be considered as edge servers, and satellites or other ABS devices other than LAP are the most suitable alternative options. In Zhou et al. (2020b), Cheng et al. (2019), both of the researches considering a space–air–ground integrated network (SAGIN), in which LAP devices act as edge computing and satellites are cloud computing. The offloading decisions were classified as MDP problems with the consideration of energy constraints and network dynamics. Subsequently, DRL was used as a solution for learning optimal offloading policies, and actor–critic for speed up the learning rate. The simulation results demonstrated the effectiveness of these frameworks by decreasing the complexity, convergences time, total delay and increase RA efficiency. Caching is another aspect worth being mentioned, especially in the case of limited computational resources.

Table 7
Prime examples of computation efficiency studies.

Ref.	AI Approach	Propose framework	Motivation	Merits	Drawbacks
Hu et al. (2021)	DRL	CF-MB	Streaming computational-expensive and real-time video simultaneously with UAV can exceed bandwidth constraints.	The computational resources scheduling was turned into semi-MDP problem and a centralised DRL association approach with CNN was used for decision making. Furthermore, the association problem was reduced in complexity with ND-POMDP and resolved using DRL algorithm.	VR streaming creates large energy burden, hence this aspect need to be taken into account in the future. Moreover, a centralised method are more vulnerable than decentralised.
Chen and Liu (2021)	DRL	MADDPG	AR streaming consumes large amount of energy and computational resources, causing huge burden and high delays.	Joint optimisation problem of RA and task offloading for AR in MEC was formulated and solved using DRL-based MADDPG framework. Better convergence property and reward in energy and computational efficiency were achieved.	The stability and real-life implementation can be a great improvement and prove the practicality of this research.
Liu et al. (2019a)	Game theory	Blockchain in MEC	The heterogeneity in videos creates computational burdens for blockchain to satisfy users demands.	MEC can resolve the RA for blockchain distributively, and video transcoder and deliverer are eligible for self-functioning to increase operation efficiency. The proposed scheme achieve good performance in reducing delays, processing time, and cost.	The security challenges with malicious consensus nodes in blockchain will be studied, and the possibility to integrate with vehicular networks are questionable.
Guo et al. (2020)	DRL	MEC-based ABR streaming	Transcoding at edge server can improve MEC-based ABR video streaming instead of remote transcoder.	A system for combined video transcoding and video quality modification were suggested with the goal of maximising QoE and minimising transcoding cost. The combined computing resource assignment and video quality adaption decision is produced automatically at each time stage.	DRL is sometimes be outperformed by other domain based heuristic algorithm. Hierarchical DRL for multiuser and the integration of DRL and domain knowledge for ABR are investigated.
Zhou et al. (2020b), Cheng et al. (2019)	DRL	SAGIN	Remote areas where edge infrastructure are often unavailable makes it difficult to provide computational resources for ABSs.	The tasks offloading decision making was formulated as MDP with respect of network dynamics. DRL was used to learn optimal policy and artic-critic was used for accelerate learning speed.	The system model only consider single UAV-satellite scheme. Potential optimisation aspects can be communication stability and caching.
Zhang et al. (2019)	DRL	3D Proactive caching	Traditional methods cannot resolve the problems of proactive caching.	The problem was formulated as MDP with joint consideration of view selection and local RA. The state transition probability was learnt using DRL approach. Finally, a Dynamic k-Nearest Neighbour was embedded into DRL for action size changing.	The work considered the traditional BSs. Furthermore, the target users are vehicular, meaning the UEs can be used for leveraging the computational resources.

Caching in AVS means temporarily save contents which have high possibility of being retrieved by user at ABS, therefore ABS can cut down the computing task for requesting videos from server and cut down waiting time for user. Zhang et al. (2019) have discussed about the proactive caching of multi-view 3D video. Proactive caching means the system will learn from user habit and preference of videos, calculate the probability of popularity, and actively cache the contents that they think it would be fit and highly possible for being viewed by user. With 3D video, it is more complicated since the video also consist of viewport predicting and RA. Hence, the problem of caching is NP-hard, and being transformed as MDP problem. The caching policies will be determined using actor–critic algorithm, and k-Nearest Neighbour algorithm for the action state. Experiment was conducted under 5G networks, the proposed solution dramatically improved the QoE of users with high hit probability and quality provisioning.

3.6. Service availability

For video streaming service, it is important to maximise the satisfactions of individual users. Single users using HAS are often innately selfish and always aim for their own interests without considering of other users. However, in the scenario of multiple users utilising the service with limited bandwidth constraints, if one player inflates the bandwidth, other participants would converge to an unfair equilibrium which causes unfair QoE of users (Jiang et al., 2018). Therefore, the system have to consider a rational distribution in such a way that each user can receive an acceptable QoS and surpass the minimum QoE despite the limitations. This problem is commonly refers as *fairness* issue in video streaming. In AVS, the service availability denotes the ability of AVS system coordinating with ABSs to provide the service for different types of ground users with the goal of maximising the minimal QoE.

The optimisation problem of service availability is to maximise the minimum QoE utility of each user and increasing fairness. It is worth to note that the UAV trajectory can also play an important role in service availability, as ABS devices should be able to locate themselves at optimal positions in which they can provide services to ground users as much as possible.

The service availability problem can be shattered into many sub-problems such as fairness and trajectory, and using AI for the resolution. Table 8 surveyed some studies which used AI approached to find the solution for these sub-problems. As have been previously discussed, DASH video streaming can maximise the QoE for user, however, it is designed mostly for individual use. In a dense network where multiple users are having network constraints, DASH barely consulting others or cooperate, leads to a degradation in fairness. One promising solution is to use cross-layer optimisation which collaboratively optimises the physical layer transmission rate for each user, then adjust the request

Table 8
Prime examples of service availability studies.

Ref.	AI Approach	Propose framework	Motivation	Merits	Drawbacks
Tang et al. (2020)	DRL	Two-level QoE fairness optimisation	Recent cross-layer QoE techniques can only have short-term effective.	In physical level, beamforming weights are optimised to improve RA efficiency and fairness. In application level, a DRL approach was used to help user learn the optimal video bitrates.	The impact of transmission layer has not been investigated, the limitation of resource storage may affect the solution.
Chen et al. (2021)	Actor–critic RL	Fairness Actor–Critic	Traditional fairness optimise algorithms are ineffective, whereas RL approaches did not consider fairness optimising.	RL adjusts rewards optimally to guarantee to converge to at least a point. The proposed scheme is capable of learning an explicitly stochastic strategy by producing the optimum probability of alternative actions.	The scheme can be extended to examine fully distributed multi-agent algorithms and off-policy training.
Comşa et al. (2020)	ML	CACLA	ML can provide feasible solutions for bandwidth-hungry applications to improve QoS despite the heterogeneity of networks.	CACLA can adapt with different environment dynamics and generate optimal decisions to improve QoS. Radio resource are optimally managed, over-provisioning is adverted, and the fraction of time is boosted.	The test cases were deployed in near-perfect scenarios, hence the proposed solution may not be effective in real-life testing.
Qin et al. (2021)	DRL	MAUC	Existing works on UAV-BS trajectory did not consider user-level fairness.	Fairness was highlighted by setting time-varying weight for each user. The placement problem was formulated as Markov game and MAUC-DRL was designed to provide fairness communication.	The flexibility of the solution is not guaranteed as it only works in certain cases and required prior system knowledge.
Ghanavi et al. (2019)	Q-Learning	ABS-Assisted fairness	The placement problem are often NP-hard, and the network topology changes dramatically due to user mobility.	The solution was given for achieving user fairness in an ABS enabled terrestrial network with user mobility. The SA-Q-learning can help ABS to increase proportional fairness among users greatly.	The work have not considered multiple UAV-BSs scenario, in addition with the use of ground BS for backhauling.
Arani et al. (2021)	RL	UCB	UAV-BSs are vulnerable for backkhauling, RA, and efficient trajectory.	Deploying LEO satellites for backhaul links, optimising RA and trajectory for ABSs and ground BSs whilst maximising user fairness. The problem in access link was modelled as MAB and resolved by using RL approach with UCB policy.	The usage of DRL methods may be a promising future development.

video bitrates at application layer to satisfy with the optimal ones. However, the cross-layer approaches are often short-term and privacy-sensitive. In Tang et al. (2020), the authors recognised the fragility of cross-layer in long-term QoE fairness maximisation, therefore proposing a two-level decision framework. In physical layer at BSs (first level), an transmit beamformers are optimised with the consideration for long-term fair RA. On the other hand, in application layer at user terminal (second level), a DRL attempt called A3C was used for bitrate selection learning based on video complexity. Likewise, Chen et al. (2021) also utilised actor–critic RL method to optimise the fairness by adjusting the weights and rewards. It is proved that with suitable rewards, the convergence of policy gradient is guaranteed to at least a stationary optimal point. Compared to other algorithms such as SARSA, FEN, and OPT, the proposed fairness actor–critic has faster convergence speed, better fairness achieved, and more optimal. Another work on fairness in 360-degree video streaming with ML, Comşa et al. (2020) have spotted out the bandwidth-hungry essence of 360-degree video may cause selfishness and unfairness bandwidth consumption. A ML-based scheduling solution was proposed to enhance inter-class fairness and provide high QoS for high quality video streaming in high dynamic environments. Particularly, a RL framework called Continuous Actor–Critic Learning Automata (CACLA) was implemented with critic neural network calculates state values and examine actions, whilst the actor neural network sets the optimal prioritisation set and scheduling rule on each state. The performance of CACLA was observed by comparing with other recent solutions such as SP, RADS and FLS and proved to be more superior in achieving better fairness.

The location of ABS is crucial for providing fairness services on time and with appropriate QoS. As previously stated, ABS trajectory optimisation can improve energy and computational resource efficiency; nevertheless, there is a paucity of studies regarding the impact on service fairness. If an aerial device is capable of situating itself in an optimal place, its coverage might serve the majority of ground users without prodigally spending time and resources. In Qin et al. (2021), the authors proposed a distributed DRL approach with consideration of user-level proportional fairness schedule. The trajectory of UAV-BSs for maximising throughput was formulated as MDP, and used a DRL method called MAUC for resolving the problem. Similarly, Ghanavi et al. (2019) utilised Q-learning to resolve the ABS placement problem for better user fairness in mobile networks. Q-Learning approach proves to be able to fast-adapt with the changes in the environment, solving the NP-hardness essence in a timely manner, and endures with the user mobility. The simulation results show that the fairness can be obtained in any situation and have better performance in reaching optimal solutions. For LEO platform, a fairness-aware framework for air-to-ground network was introduced in Arani et al. (2021), as the authors have realised the vulnerability of UAV-BSs at backhauling and replace them with LEO satellites, which have more computing capability and larger coverage. First, the problem of backhaul link and access link were addressed, as LEO satellites providing backhaul for both UAV-BSs and terrestrial BSss, and the access link maximises fairness and minimises loads. The problem was formulated as a multi-armed bandit (MAB) and being resolved with upper confidence bound (UCB) policy method. The user fairness and spectrum efficiency are increased, proving the effectiveness of the proposed scheme compare to other benchmark algorithms.

3.7. Security and privacy

Video streaming may have access to user data to learn about their habits and behaviours and therefore be able to deliver better content

Table 9
Prime examples of security and privacy studies.

Ref.	AI Approach	Propose framework	Motivation	Merits	Drawbacks
Xiao et al. (2021)	RL	Anti jamming	Potential jamming attacks can degrade QoE of users.	Anti-jamming was formulated as MDP problem, and RL chooses the optimal transmission policies. CNN was later introduced to resize the state dimensions, accelerate the learning process.	Multiple jamming case, latency constraints, possible errors are some promising enhancement in the future.
Lu et al. (2022)	RL	SHDRL	Large action set slows learning speed and causing inaccurate rewards, leads to a degradation for RL approaches.	A safe RL technique for wireless security applications that optimises learning efficiency and safe exploration. A deep safe RL was also used to advance learning efficiency.	The experiment only consider single UAV scenario, with a slight amount of video traffic data.
Silva et al. (2019)	DL	RPTT-ReID	Object tracking in crowd and complex environments may challenging because of occlusion or target identifying.	MAPF algorithm was used to spot the target in dense crowd, then the UAVs scan and search whilst preserving privacy using MAPF. Lastly, an occlusion-awareness distributed DNN for identity sustain and predicting future target location.	The performance of the solution is not much superior than other existing algorithms, and the work only focus on vision perspective.
Li et al. (2021)	RL	RDC	Eavesdroppers can cause jamming signals and interfering the wireless data transmission.	RDC with the integration of RL was presented to avoid eavesdroppers and optimise BS network coding and transmission policy without requiring knowledge of drone eavesdropper channel conditions.	The case of large users with real-life services have not yet been investigated.
Challita et al. (2019)	ANN	Cellular-connected UAVs	The emergence of cellular networks enable a wide range of applications, as well as security challenges	The application scenarios of cellular UAVs were introduced, and solution based on AI were presented as a promising answer to tackle security threats.	Securing UAV swarm consensus is not possible for ANN. The authors, on the other hand, propose a federated method.
Sthapit et al. (2021)	DRL	DDPG	Computation offloading can contain possible threats to LEO satellites.	The security problem of computation offloading was formulated as a multi-objective problem. DDPG was used to optimise in Monte-Carlo simulations with real-time results and proved to be able to perform better than DQN.	In extremely high risks cases, DDPG can perform worse than DQN. The cooperation among satellites in swarm is poor.
Masood et al. (2021)	FL	HAP Content caching	Existing proactive caching schemes need access to local data of users, exposing privacy concerns.	A hierarchical FL was used to study user habits and predict contents popularity to cache. Prediction accuracy improved using DNN. The interaction between UEs and ABSs were discussed.	Other QoE aspects have not been investigated.
Wang et al. (2020a)	FL	SFAC	FL may vulnerable against unreliable contribution recording, central conservator, and inferior shared local models.	SFAC used blockchain, LDP, and RL. Blockchain omits central curator and verifies model exchange, LDP protects the privacy of ABS when updating with local models and RL enhances model sharing without prior knowledge of parameters.	Utilisation of FL in model selection and multi-agent DRL for accelerating learning process are some potential future challenges.
Ha et al. (2021)	SL	SASSL	Aerial surveillance may contain visual records of persons on the ground, exposing their identity.	SASSL protects privacy by processing first layer of DNN on UAV. It also consider the collaboration of other UAVs to provide sufficient images.	Possible future improvements include energy, RA, and joint scheduling.

depending on user preferences. However, this may expose private user information to third-party or unauthorised individuals, resulting in illicit exploitation and storage of video activities (Alarifi et al., 2020). Furthermore, assaults on video streaming services can impair QoS by creating delays, spamming, content pollution attacks, and password theft. On the other hand, significant risks exist in using ABSs, particularly LAP platform devices. Some primary security visions for LAP include GPS spoofing, de-authentication, jamming, man-in-the-middle attack, MESM gyroscope, camera spoofing, buffer overflow, and key loggers (Hamza et al., 2020; Sun et al., 2019; Fotouhi et al., 2019). For HAP or LEO satellites, the emergence of 5G/6G technologies, in addition to AVS and sensitive national information, has produced a significant quantity of data to be held on satellite premises, making them a lucrative target for hackers and attackers (Petrosino et al., 2022). From the AI perspective, one of the overarching challenges of the digital age is data privacy. Since data is the lifeblood of modern AI, data privacy issues play an important role in AI. Hence, privacy-preserving AI methods that allow AI models to learn from datasets without compromising user privacy are becoming an important pursuit.

Utilising advance AI algorithms to strengthen security level is currently the main direction for modern aerial service security solution. Table 9 presents several recent prime studies on privacy and security on AVS using AI. At first glance, popular AI models such as DL, RL and DRL can be applied on both ABS and video streaming aspect. Furthermore, some state-of-the-art AI models are being applied to further enhance the privacy on video data transmission. The countermeasure against security-related issues on UAV streaming has received much attention, e.g., Xiao et al. (2021) and Lu et al. (2022). In the first work, the article investigated the RL solutions for anti-jamming by choosing suitable policies and channel states based on video priorities. The authors also designed a DL algorithm to accelerate the learning process and reduce outage probability. In the latter, another RL algorithm was proposed to study the anti-jamming case. Similar to the

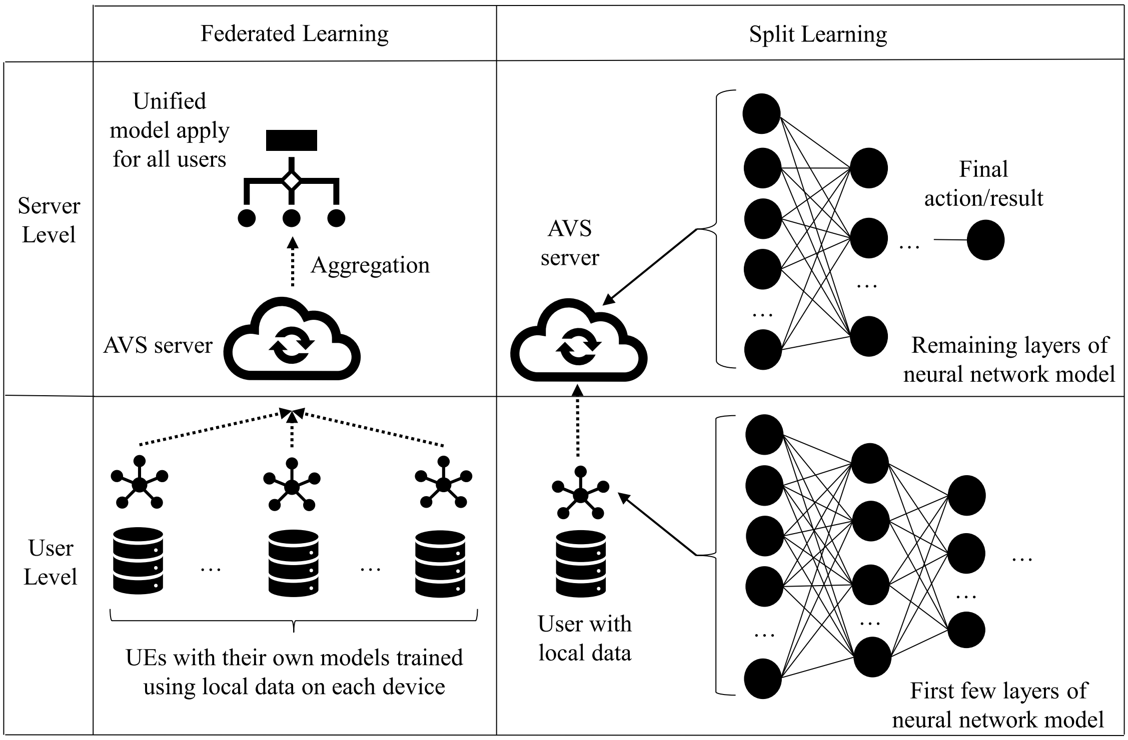

Fig. 4. Working principal of FL and SL.

previous work, the policy was selected priority-based but divided into three subsets to abbreviate the action set. In highly dynamic wireless networks, a deep-safe RL (SHDRL), which employs four CNNs at each level of sub-policy selection, was presented to stabilise exploration and expedite learning. SHDRL proved its effectiveness against jammer by reducing 60% of convergence time, 95.2% packet loss rate, 95.9% bit error rate, and 14.1% of energy consumption. In Silva et al. (2019), a distributed decentralised DL Real-Time Privacy-preserving Target Tracking Re-Identification (RPTT-ReID) approach was introduced to preserve privacy in human tracking. RPTT-ReID allows UAV surveillance to ensure the privacy of tracked objects but maintain close tracking despite re-identification. Each UAV scans and explores the defined regions revealed by the Multi-Agent Path Finding (MAPF) algorithm to identify targets, then observe targets with distributed DNN model.

For securing drones and UAV-BSs communication, Li et al. (2021) have designed a downlink transmission framework called RL-based drone-aided network coding (RDC), in which BSs leverage random linear network coding (RLNC) algorithm for packet encoding and RL to select network coding and transmission policy. The model virtually brings out secure transmission simulation to generate more policy learning. The suggested scheme beats FRP and CS benchmark systems according to simulation findings. In particular, RDC reduces intercept probability by 87.2%, 47.7% of latency, and outage probability by 84.9% after 2500 time slots while saving 41.4% BS energy. The article Challita et al. (2019) uses a machine learning approach to solve security problems such as interference management, mobility management and handover, cyber–physical attacks, and authentication for cellular-connected UAVs. Firstly, they addressed current security issues in UAV-based media streaming, then proposed an ANN-based scheme solution that allows UAVs to exploit system resources while maintaining security standards. The simulation has shown that ANN-based solutions can solve most of the wireless and security challenges when applied to UAV applications. However, securing a consensus on UAV swarms is unavailable for ANN. In contrast, the authors suggest a federated approach since they realised the suboptimal essence of the ANN approach. Securing wireless communication and offloading for satellites was discussed in Sthapit et al. (2021), as the main reasons for satellites' ease exposition to risks are the dynamic network, the heterogeneity of data, and massive traffic. Having learned about security issues, a security-aware algorithm for computing offloading was proposed based on the DDPG technique. The task queue and the jobs waiting for problems will be formulated as MDP to minimise the queuing time, energy, and risk levels and use DDPG as a solution.

Recently, to preserve the privacy of user data, *federated learning* (FL) and *split leaning* (SL) have been introduced to help servers can retrieve the required information they need without having to access directly to users. In AVS, it is essential to learn about each user's habits and video preferences; therefore, the streaming system can proactively cache contents, serve appropriate video suggestions, and provide relevantly advertises for ISPs. However, the system cannot violate user privacy by accessing UEs. Instead, FL allows each UE will train their model on their premises using data, then send only the trained model to a centralised server (Li et al., 2020a). Using FL, the server can aggregate all the models from every user without accessing user data and provide one standard model for all users it serves. In Masood et al. (2021), Masood et al. apply FL to the case of content caching in HAP devices, in which the edge ABS studies the video popularity to cache contents, cutting down access delays effectively. In Wang et al. (2020a), a secure FL framework called SFAC was proposed to tackle the vulnerabilities of aerial FL. SFAC integrates RL, blockchain, and local differential privacy (LDP) to train the AI model securely. Blockchain resolves the central curator issue, LDP secures the privacy of local models, and RL chooses the optimal strategies. Likewise, SL is being used primarily on neural networks, splitting layers into two subsets: to be trained locally on UEs and the others to be trained on the server (Vepakomma et al., 2018; Koda et al., 2020). Video surveillance, such as in Ha et al. (2021), takes advantage of SL to analyse frames effectively without exposing the initial figures. The visual video will be captured by UAVs, then only the first hidden layer of the neural network model for video analysis will be processed on UAV. After that, the feature map of the first learnt layer, which has been distorted, will be offloaded to the cloud server, in which the remaining layers of the neural network will continue to be trained based on the feature map. Based on the selected model, the server will be able to perform tasks such as object classification or detection on video. AVS system can use SL not only to preserve the privacy of users by not accessing original data but also to utilise the computational resources on UE and increase computational efficiency. Gao et al. (2020) have evaluated the usage, importance, and performance of both FL and SL in their work, and an overview principal of them is presented in Fig. 4.

3.8. Summary and discussion

In this section, an AVS prototype system has been modelled, and seven different utilities were examined. Each utility was given an overarching goal formula and limitations, which were then followed by recent AI attempts to resolve it. To better understand how AI was leveraged, a summary table was included under each utility to show which AI technique was used, what the suggested framework was, the purpose for building it, its benefits, and potential future obstacles. Throughout this section, it can be seen that there is a dearth of technical articles that jointly consider both *aerial* and *video streaming*. On the one hand, ABSs have gotten substantial interest since they are a potential research topic for 5G/6G deployment. Most work on optimising video streaming utilities, on the other hand, is only applicable to traditional terrestrial BSs infrastructure. As a response, this study has synthesised research works on each aspect in the spirit that if AI can optimise utilities in standard video streaming, likely, it can also operate when implemented on aerial infrastructure. Similarly, enhancing airborne devices with more powerful processing capability increases their potential to offer and distribute Internet services, which video streaming can benefit from.

In video resolution, the codec and transrater play critical roles in maximising user QoE. In this part, AI is used to improve the performance of the codec and flexibility of the transrater. Second, as previously described in Section 2, deploying the edge method and CDN on ABS can be used to reduce AVS delays. However, the amount of work utilising AI on latency for AVS is limited, most of them being simulation and assessment research. Third, the energy consumption of AVS was complicated by two issues: streaming energy and ABS hovering energy. Because of their dynamic settings, high data flow, and limited battery bank, LAP devices are the most vulnerable. AI may be used to either improve utilisation strategies in video streaming systems or to improve the overall energy storage of ABSs to boost performance. Following that, the integrity and continuity of video transmission are to designate service stability, preventing video stalling or server breakdown, which results in a decline in QoS. The recommended options for video service stabilisation include enhancing video buffers, optimising rate management, and reducing bandwidth restrictions. Furthermore, to efficiently serve users under resource limitations, AVS employs AI to on the RA problem. As represented in Table 7, edge caching, smart resource utility, and scheduling can increase the computation efficiency. The availability of services was also mentioned as the ability to provide services to as many people as feasible in the shortest time possible while maintaining the highest quality. Last but not least, the status of recent advancements in increasing AVS system security and privacy was addressed. Overall, the goals are to maximise video resolution utility, energy efficiency, service stability, computation efficiency, availability, and security while minimising the others.

At first glance at the summary tables, RL and DRL usage seem superior to any other AI approaches. This is not unexpected given that these two algorithms are regarded as the most effective optimisation algorithms for accomplishing long-term objectives and are ideal for addressing NP-hardness issues and learning from massive amounts of video data. However, because utility issues are challenging to solve using existing mathematical methodologies, it is usual practice to divide the original into numerous subproblems or to rephrase it as MDP before applying AI methods. Despite reducing its complexity, the final solutions may not be globally ideal due to the simplification of foremost concerns. Another thing worth mentioning is that the number of HAP devices is considerably less compared to LAP and LEO. The rationale is that LAP is more suited for latency-sensitive applications, LEO is ideally equipped for resources and extended coverage, and HAP is more akin to an average blend of the other two platforms. Because HAP does not stand out in any way, the demand for their use is modest, and hence the need for articles on them is likewise low.

4. Application scenarios

4.1. Remote healthcare

The current pandemic of COVID-19 has demonstrated critical impacts on the deployment of an effective healthcare system. The number of patients is increasing daily, in addition to external factors such as geographical constraints, medical equipment limitations, and the emergence of first aid, striving for a remote healthcare system. The mobility of aerial devices makes them an outstanding candidate for enabling the system. In Dong et al. (2021), an online eHealth system UAV-based for patient monitoring and offers edge computing services was presented. The Lyapunov optimisation approach is used to deconstruct the long-term optimisation issue into a series of instantaneous optimisation problems to decrease the health monitoring latency and ensure the resource usage efficiency of UAVs. Subsequently, the deconstructed subproblem is demonstrated as a minimal-cost maximum flow problem by constructing a bipartite graph that links medical analysis requests to target UAVs. Extensive experimental findings show that the technique is successful. Finally, system evaluations show that the system is stable and scalable and can deliver agile edge computing services to scattered patients.

Video streaming can assist the communication between doctors and patients in various aspects: surgery monitoring and remote guidance. In Podder et al. (2021), a sign language transformer (SLT) framework has been proposed to assist disabled patients in communicating with doctors. By leveraging computer vision study and image analysis, the system can study the doctor's movement, and speech guidance, then translates it back to the impaired patient with sign language. The model evaluates the existing computer vision technologies and selects the most suitable one. Another work from Rahim et al. (2021) leveraged the computational capability of BSs for image analysis on the premise of airborne devices. Remarkably, the videos and images are transmitted via BS to the doctor for a consultant; it simultaneously attempts to predict and detect the possible tumor inside the media data. With the DL approach, the authors have improved the existing tumor-detection scheme, You Only Look Once (YOLOv3-tiny). The results show that the improved version has provided competitive results compared to the original framework regarding precision, sensitivity, F1-score, and F2-score.

4.2. Video surveillance

Aerial video surveillance has been widely identified as one of AVS's most popular use cases. For instance, Yang et al. (2019c) focused on the ultra-high video resolution in object tracking with Atomic Visual Action (AVA) by proposing a multi-label AVA detection framework. The purpose of the study is to give a better multi-action prediction algorithm for ground objects which are located sparsely. The performance was evaluated using two metrics: detection performance and multi-label action recognition. The results show that the output accuracy of predicting action is significant and capable of multi-function usage. A drawback of this study is that the feeding images and videos have to be in ultra-high resolution for better learning, meaning low-and-medium videos are not suitable. The work of Zheng et al. (2021) has noted that for aerial surveillance, despite having more prominent and better image quality than LAP and HAP, LEO satellites are limited in the range of operations with a particular flight trajectory. However, LAP and HAP's diverse viewpoints have raised the stream's complexity. A prior sampling and sample check RANSAC (PSSC-RANSAC) was proposed to ease image registration. Samples used for the PSSC-RANSAC prior sampling incorporation were generated from three sample evaluation levels: texture magnitude, spatial consistency, and feature similarity. The compatibility of samples was checked using prior information on the sample, the quality of the sample subset, and subset invariability. The experiment results demonstrated the excellent performance of the given PSSC-RANSAC at a contamination level of 90%. For typical image pairings, the number of iterations is reduced by at least 16.67%, at least 11.01% reduces the evaluation computation, and the re-projection error is reduced by at least 4.44% and 6.31%, respectively when compared to RANSAC and SVH-RANSAC. It is capable of overcoming interference and is ideal for aerial picture registration.

The use of UAVs for smart city surveillance was investigated in the work of Jin et al. (2020), where the video streaming quality and other QoS metrics are concerned. UAV-based cluster optimisation algorithms and scheduling strategies were introduced to tackle these challenges. An end-to-end device coverage network was designed first, followed by formulating the scheduling problem as bi-objective fragile bin packing. Lastly, an optimal scheduling algorithm with a constant approximation performance ratio was given and tested via simulations. Compared to UAV-DG, the suggested algorithms can achieve a 50% system life cycle extension and a 20% improvement in video decodable ratio. Furthermore, the UAV flying time ratio is 10% smaller than that of UAV-DG. Throughput is at least 15% more than that of UAV-DG and 35% greater than that of UAV-NULL. Compared to UAV-DG, the delay is reduced by at least 150 percent.

4.3. Search and rescue

The independence from terrestrial infrastructure, flexibility, and high mobility have made aerial devices an excellent option for search and rescue missions. Precisely, airborne devices equipped with cameras can monitor a large area in the air, possibly to detect any hazards or find a target object on the ground that humans struggle to reach. For example, in McGee et al. (2020), UAVs are equipped with thermal detection video to detect any possible endangered person in the area. The dataset was taken at Kangaroo Point cliffs in Brisbane, Australia, with a total of 2751 original photos that were annotated, with the enhanced dataset including 10,380 images. The image analysing tool YOLOv3, like the work of Rahim et al. (2021) in the remote healthcare subsection, was used to study the images. In Lygouras et al. (2019), an autonomous UAV was used to detect possible endangered swimmers on the water surface in an unsupervised manner. The DL algorithms were applied to enhance the visual detection of UAVs. Furthermore, Alotaibi et al. (2019) proposed a multi-UAVs collaboration scheme to expand the scalability of UAVs monitoring. Each UAV collaboratively works with each other for efficient RA and ground person detection in disaster scenarios. Airborne devices can be used as ABSs to provide communication connectivity if terrestrial BSs have been damaged. This article (Zhao et al., 2019a) deploys a unified scheme for a UAV-assisted network in emergency disaster scenarios. The solution consists of 3 steps: first is to implement an optimised connection between UAVs and surviving BSs, second is to establish UAVs as a transceiver to ground devices, and last is to study the hovering position of UAVs that affects the connection.

Besides reconnaissance ability in search and rescue, AVS can be directly involved in the procedures. The work of Spurny et al. (2021) suggested that AVS on UAV can automatically detect fire in buildings or indoor situations. Furthermore, UAVs are armed with fire distinguish; therefore, based on the fire location being analysed through video streaming, UAVs can immediately activate the distinguisher onto the fire. Additionally, UAVs can be used to deliver supplies and materials for rescuers and victims in dangerous areas (Mardiansyah and Budi, 2018; Wenjian et al., 2020; Huang et al., 2021). Video streaming can be a connecting bridge between endangered people and rescuers by displaying video calls, providing aid instructions, or simply reassuring the victim in an emergency.

When the search area was enormous, HAP and LEO satellites were brought in to replace LAP devices. One of the most common use cases of satellites in searching and rescuing is the maritime rescue mission. Shimizu et al. (2019) proposed the application of LEO satellites in monitoring maritime ships in Japan, making it possible to react promptly if emergency relief is needed. In King (2021), the vital role of Cospas-Sarsat in rescuing missions has been highlighted, and more than 30 countries in the world use Cospas-Sarsat nowadays has demonstrated how effective it was. A famous example came from the mysterious case of MH370 when a Malaysian aircraft suddenly disappeared without traces in 2014 during its journey from Malaysia to China. Satellites have been mobilised to search for the remaining clues of the aircraft in the South East China Sea and the Strait of Malacca (Ivić et al., 2020; Gao et al., 2018; Saroni et al., 2019).

4.4. Smart agriculture

In the modern era, using AVS can increase the productivity of traditional tasks such as agriculture. Airborne devices are being used to monitor livestock and agricultural farming, as well as to generate visual data on the plant growth process and even to engage in cultivation directly. Due to the nature of engaging with property owners and that agriculture frequently has set scales, UAVs and LAP devices are the ideal match for the job. The surveys of del Cerro et al. (2021), Mukherjee et al. (2019), Erica et al. (2019), Yinka-Banjo and Ajayi (2019) conducted a brief overview of the use cases of UAVs in smart agriculture. According to them, farmers throughout the world have employed at least 2900 UAVs, resulting in an increase of service providers to 900 by 2020. The usage of aerial devices for farming dates back to 1906 for seed distribution and has continued to prosper with more advanced devices and tasks. Nowadays, UAVs equipped with video streaming services can be found in applying nutrient evaluation, health assessment, water analysis, biomass estimate, soil monitoring, weed detection, and other environmental issues monitoring. In some exceptional cases where the farming area is exceptionally remote, LEO satellites can also be leveraged for smart farming (Islam et al., 2021).

Video monitoring can detect possible hazards to crops and livestock, which can notify farmers promptly so that solutions can be given. For example, bugs and invasive insects can be detected in Stumph et al. (2019), DL for image analyses on UAVs for detecting spotted wing drosophila pest (Roosjen et al., 2020), crop health study in Pakistan using IoT and ML (Shafi et al., 2020), etc. Some traditional activities in agriculture, which used to be done only by human, is currently embracing a transition thanks to the advance of AVS. Mustering uses UAVs as sky shepherds (Yaxley et al., 2021) and animal fencing with video analysing (Sarwar et al., 2021; Sun et al., 2020a) are some of the trending use cases. In one interesting case in Chebrolu et al. (2019), the authors have pointed out that the likeness of landmass would confuse UAVs when locating their position on the field. By sufficiently analysing images taken from UAV's camera using ML and computer vision algorithms, UAVs can extract special features that help them locate themselves despite the landscape.

4.5. Summary and discussion

This section covered four typical scenarios in which AVS may be used: remote healthcare, video surveillance, search and rescue, and smart agriculture. It observed that video streaming through aerial infrastructure has many applications and plays an essential role in modernising human life. AVS has played a vital role in patient monitoring and delivering vital medical services to persons in rural places during the current COVID-19 outbreak. Airborne gadgets are also recognised for their capability to conduct aerial surveillance. UAVs are used as CCTV cameras and monitoring equipment in smart cities and metropolitan areas worldwide to provide public security and traffic safety. Furthermore, with the advancement of AVS in assisting hazard identification, and rescue under challenging areas, marine searching, search and rescue missions have become more accessible than before. Last but not least, conventional cultivation operations have been made more productive by using AVS on aerial equipment, improved farming techniques, greater crop yields, and enhanced agricultural damage prevention. To summarise, the potential of AVS is enormous, with a broad array of applications to human life. This demonstrates that the need for AVS will undoubtedly increase, especially with the introduction of sophisticated communications technologies. As a result, it is essential for researchers and ISPs to have a better understanding of AVS and to be able to provide optimal solutions for this system.

5. Open challenges

Despite the recent advances in the improvement of airborne devices and video streaming services, as well as the emergence of high-level communication technologies like 5G/6G, the popularisation of an end-to-end video streaming system over aerial infrastructure faces several obstacles. This section discusses these challenges and suggests feasible approaches to spur further investigation.

5.1. AI adaptability

The adoption of AI methods has undoubtedly improved AVS performance. Nonetheless, the adaptability of AI solutions to the AVS systems met several concerns due to the highly dynamic essences of video streaming environments with various diverse states. Extensive data training with limited resources and prior knowledge can easily persuade AI into poor execution. On the one hand, if the AI model were designed substantially with many parameters to adapt to the dynamic video streaming environment, it would result in significant resource loads, high latency, and degraded QoS. However, if the model is lightweight and only operates in hypothetical conditions, performing the tasks allocated in practice may be inefficient. Moreover, AI solutions must consider airborne devices' energy and resource limits to work correctly. As a result, developers/researchers must create a robust AI model capable of meeting output requirements and adapting to the AVS environment.

5.2. Tradeoff optimisation

Optimising a single utility of the AVS system does not completely optimise the system. This is attributable to the fact that a system is dependent on how each utility interacts with one another. However, each has its metrics and objectives, in addition to the constraint of computational resources. If one utility receives greater attention than the others, the system's performance is jeopardized because most resources have been devoted to optimising only the selected feature. A potential option is to balance the tradeoffs among utilities on the demands of users and operators. To some extent, this challenge is connected to the *fairness* problem, in which a video streaming system must consider a fair resource distribution among users. For example, to improve latency, video streaming might reduce video resolution to conserve energy and reduce delays.

5.3. Utility personalisation

Video streaming system tries to optimise every utility; however, it also means balancing the entire system. To some video viewers, this may be triggered since their desire for each utility is different. For example, some users may prioritise video resolution, while some may want to enjoy a video with the highest smoothness. To secure system performance, the video streaming system estimates the acceptable QoS rate without consulting the opinion of users. It provides an optimal service depending on network conditions (e.g., adaptive streaming). To this end, there should be more interaction between users and video streaming services, and also, the system should be able to adjust its goals and strategy to adapt to the objective changes.

5.4. Live stream autoprocessing

The autoprocessing of information has been a significant component in intelligent settings with AI capabilities. In these cases, the information is refined at networking devices by extracting, adding, removing, and fusing it to provide the needed knowledge. AI can automatically discover data patterns and learning contexts in this regard. Although the use of AI for this purpose has been demonstrated to be viable, more research needs to be conducted to build efficient and practical solutions. It is worth noting that when sophisticated AI algorithms are used, a balance between autoprocessing accuracy and resource consumption costs in terms of time, space, and energy must be addressed.

5.5. Energy constraints

When it comes to video streaming service deployment via aerial infrastructure, one of the most typical concerns is energy consumption, particularly for LAP devices, which are often prone to durability owing to their subtle design. Furthermore, to function as an ABS, devices must have substantial energy reserves to keep services operational. There has been an extensive effort on increasing energy efficiency; one popular solution is utilising tethered connectivity (Zhang et al., 2021b; Wang et al., 2021b; Yingst and Marojevic, 2021); however, these solutions frequently have limits in other parameters for the energy tradeoff. Another asset to consider is the RA issue since video traffic is rapidly expanding. Current video streaming services are combined with other services to improve QoE/QoS, such as edge-C3, CDN, transcoding/rating, etc. As a result, ABSs require an ample storage capacity and an effective resource sorting mechanism to have adequate resources for long-term operation. To overcome the energy problem, in addition to maximising the on-board energy storage of aerial devices, alternative solutions that externally charge energy for ABS energy replenishment, such as wireless power transfer, can be a viable option, as indicated in Le et al. (2020). LAP technologies, like LEO satellites, can use solar power as a recyclable energy source. Furthermore, innovative concepts such as employing ground vehicles as mobile charging stations for ABS devices should be studied.

5.6. Deployment cost

LEO equipment (e.g., satellites) has enormous installation and operational costs, but their maintenance costs are cheaper due to their capacity to operate for lengthy periods. On the other hand, UAVs for LAP are less expensive and easier to implement and operate. Nevertheless, many devices are required to deploy a comprehensive aerial system utilising drones and UAVs efficiently because of their restricted coverage area and computational capabilities. Furthermore, their lack of energy forces them to do maintenance regularly, which is insufficient when it comes to a large number of devices. In deploying Internet services, mainly video streaming services, each airborne device must be equipped with its processor, battery, and antenna on its premises rather than relying on a single centralised BS, which discourages ISPs from investing in low aerial infrastructures such as HAP or LAP.

5.7. Privacy and security

Although privacy and security have been investigated in numerous studies, as discussed in Section 3.7, the problem remains an open challenge and requires non-stopped efforts to overcome rising and novel security threats. In particular, video streaming activities may collect data and information about users, some of which may even be private data. From an aerial device perspective, a cyberattack on satellites can result in a national security threat since the majority of operators for LEO are often large corporations or national premises. For LAP devices, managing an enormous number of devices makes them more vulnerable to malware and can easily be attacked by hackers. Furthermore, to efficiently utilise and process the information and context in user data, system and user authorisation and accounting must be rigorously built to maintain user data privacy. To tackle this issue, recent AI advancements such as federated learning and split learning, in addition to blockchain technologies, are intriguing prospects for future study. Moreover, security and privacy should be addressed concurrently when enhancing system speed and user experience metrics.

6. Concluding remarks

Video streaming is expected to dominate data traffic in the future, putting a significant computational strain on traditional terrestrial communication systems. With the recent emergence of UAV systems and the application of AI solutions in various fields, AVS services will likely be one of the most prominent services in the coming years. This paper provides an in-depth assessment of current works on AI approaches for video streaming over aerial infrastructure to provide a comprehensive picture of the current state of research on AVS services. Overall, it can be seen that although there has been an intensive amount of research on both UAV systems and video streaming services individually, there is a lack of studies that investigate AVS as a whole complete system. Due to this limitation, we surveyed studies that leveraged AI in tackling problems on UAV and video streaming, then attempted to connect them as feasible solutions to be applied to a complete AVS system. To do that, we first begin by sketching a video streaming system deployment on an aerial communication network to illustrate each component's operating principle, system architecture, features, and roles. Subsequently, seven significant utilities of the AVS system were examined by establishing the general objective for each utility based on system modelling. By providing a summary of the most recent studies utilising AI to improve the performance of the aforementioned utilities, we help readers stay up to date with the most recent AI efforts on UAV systems and video streaming, hence providing an idea of how these individual studies can benefit future AVS. AVS application scenarios were provided for better highlights of the advantages of AVS on human lives, promoting the implementation of the system. However, although the efficacy of AI approaches on AVS, there are several difficulties in applying AI to tackle existing problems on streaming via aerial infrastructure. Several major disadvantages were discussed in the last section to accentuate these drawbacks and provide readers with potential directions for future research trends.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1G1A1008105).

References

Ahmed, I., Ismail, M.H., Hassan, M.S., 2019. Video transmission using device-to-device communications: A survey. IEEE Access 7, 131019–131038.
Al-Abbasi, A.O., Aggarwal, V., Ra, M.-R., 2019. Multi-tier caching analysis in CDN-based over-the-top video streaming systems. IEEE/ACM Trans. Netw. 27 (2), 835–847.
Alarifi, A., Sankar, S., Altameem, T., Jithin, K., Amoon, M., El-Shafai, W., 2020. A novel hybrid cryptosystem for secure streaming of high efficiency h. 265 compressed videos in IoT multimedia applications. IEEE Access 8, 128548–128573.
Alotaibi, E.T., Alqefari, S.S., Koubaa, A., 2019. Lsar: Multi-uav collaboration for search and rescue missions. IEEE Access 7, 55817–55832.
Amer, R., Saad, W., Marchetti, N., 2020. Mobility in the sky: Performance and mobility analysis for cellular-connected UAVs. IEEE Trans. Commun. 68 (5), 3229–3246.
Arani, A.H., Hu, P., Zhu, Y., 2021. Fairness-aware link optimization for space-terrestrial integrated networks: A reinforcement learning framework. IEEE Access 9, 77624–77636.
Asan, A., Robitza, W., Mkwawa, I.-h., Sun, L., Ifeachor, E., Raake, A., 2017. Impact of video resolution changes on QoE for adaptive video streaming. In: 2017 IEEE International Conference on Multimedia and Expo. ICME, IEEE, pp. 499–504.
Bai, T., Pan, C., Wang, J., Deng, Y., Elkashlan, M., Nallanathan, A., Hanzo, L., 2020. Dynamic aerial base station placement for minimum-delay communications. IEEE Internet Things J. 8 (3), 1623–1635.
Barman, N., Martini, M.G., 2021. User generated HDR gaming video streaming: dataset, codec comparison and challenges. IEEE Trans. Circuits Syst. Video Technol..
Belmoukadam, O., Khokhar, M.J., Barakat, C., 2020. On excess bandwidth usage of video streaming: when video resolution mismatches browser viewport. In: 2020 11th International Conference on Network of the Future (NoF). IEEE, pp. 159–167.
Bigazzi, L., Basso, M., Gherardini, S., Innocenti, G., 2021. Mitigating latency problems in vision-based autonomous UAVs. In: 2021 29th Mediterranean Conference on Control and Automation. MED, IEEE, pp. 1203–1208.
Bronzino, F., Schmitt, P., Ayoubi, S., Martins, G., Teixeira, R., Feamster, N., 2019. Inferring streaming video quality from encrypted traffic: Practical models and deployment experience. Proc. ACM Meas. Anal. Comput. Syst. 3 (3), 1–25.
Burhanuddin, L., Liu, X., Deng, Y., Challita, U., Zahemszky, A., 2022. Qoe optimization for live video streaming in UAV-to-UAV communications via deep reinforcement learning. IEEE Trans. Veh. Technol..
Cai, Y., Wang, R., Wang, Z., Han, B., Li, X., 2021. An efficient and open source encoder uavs3e for video compression. In: 2021 IEEE International Conference on Multimedia and Expo. ICME, IEEE, pp. 1–6.
Challita, U., Ferdowsi, A., Chen, M., Saad, W., 2019. Machine learning for wireless connectivity and security of cellular-connected UAVs. IEEE Wirel. Commun. 26 (1), 28–35.
Chebrolu, N., Lottes, P., Läbe, T., Stachniss, C., 2019. Robot localization based on aerial images for precision agriculture tasks in crop fields. In: 2019 International Conference on Robotics and Automation. ICRA, IEEE, pp. 1787–1793.
Chen, Z., He, T., Jin, X., Wu, F., 2019. Learning for video compression. IEEE Trans. Circuits Syst. Video Technol. 30 (2), 566–576.
Chen, J., Hu, M., Luo, Z., Wang, Z., Wu, D., 2020. SR360: boosting 360-degree video streaming with super-resolution. In: Proceedings of the 30th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video. pp. 1–6.
Chen, X., Liu, G., 2021. Energy-efficient task offloading and resource allocation via deep reinforcement learning for augmented reality in mobile edge networks. IEEE Internet Things J. 8 (13), 10843–10856.
Chen, J., Wang, Y., Lan, T., 2021. Bringing fairness to actor-critic reinforcement learning for network utility optimization. In: IEEE INFOCOM 2021-IEEE Conference on Computer Communications. IEEE, pp. 1–10.
Cheng, N., Lyu, F., Quan, W., Zhou, C., He, H., Shi, W., Shen, X., 2019. Space/aerial-assisted computing offloading for IoT applications: A learning-based approach. IEEE J. Sel. Areas Commun. 37 (5), 1117–1129.
Choi, W., Yoon, J., 2019. SATE: Providing stable and agile adaptation in HTTP-based video streaming. IEEE Access 7, 26830–26841.
Cicek, C.T., Gultekin, H., Tavli, B., Yanikomeroglu, H., 2019. UAV base station location optimization for next generation wireless networks: Overview and future research directions. In: 2019 1st International Conference on Unmanned Vehicle Systems-Oman. UVS, pp. 1–6.
Cisco, 2020. Global–2021 Forecast Highlights. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html, (Accessed on February 8, 2022).
Comşa, I.-S., Muntean, G.-M., Trestian, R., 2020. An innovative machine-learning-based scheduling solution for improving live UHD video streaming quality in highly dynamic network environments. IEEE Trans. Broadcast. 67 (1), 212–224.

Cui, L., Su, D., Yang, S., Wang, Z., Ming, Z., 2020. TCLiVi: Transmission control in live video streaming based on deep reinforcement learning. IEEE Trans. Multimed. 23, 651–663.

Dao, N.-N., Pham, Q.-V., Do, D.-T., Dustdar, S., 2021a. The sky is the edge–toward mobile coverage from the sky. IEEE Internet Comput. 25 (2), 101–108.

Dao, N.-N., Pham, Q.-V., Tu, N.H., Thanh, T.T., Bao, V.N.Q., Lakew, D.S., Cho, S., 2021b. Survey on aerial radio access networks: Toward a comprehensive 6G access infrastructure. IEEE Commun. Surv. Tutor. 23 (2), 1193–1225.

Dasari, M., Bhattacharya, A., Vargas, S., Sahu, P., Balasubramanian, A., Das, S.R., 2020. Streaming 360-degree videos using super-resolution. In: IEEE INFOCOM 2020-IEEE Conference on Computer Communications. IEEE, pp. 1977–1986.

del Cerro, J., Cruz Ulloa, C., Barrientos, A., de León Rivas, J., 2021. Unmanned aerial vehicles in agriculture: A survey. Agronomy 11 (2), 203.

Do, H.T., Truong, L.H., Nguyen, M.T., Chien, C.-F., Tran, H.T., Hua, H.T., Nguyen, C.V., Nguyen, H.T., Nguyen, N.T., 2021. Energy-efficient unmanned aerial vehicle (UAV) surveillance utilizing artificial intelligence (AI). Wirel. Commun. Mob. Comput. 2021.

Dong, P., Wang, X., Wang, S., Wang, Y., Ning, Z., Obaidat, M.S., 2021. Internet of UAVs based remote health monitoring: An online ehealth system. IEEE Wirel. Commun. 28 (3), 15–21.

Du, K., Pervaiz, A., Yuan, X., Chowdhery, A., Zhang, Q., Hoffmann, H., Jiang, J., 2020. Server-driven video streaming for deep learning inference. In: Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication. pp. 557–570.

El Marai, O., Taleb, T., Menacer, M., Koudil, M., 2017. On improving video streaming efficiency, fairness, stability, and convergence time through client–server cooperation. IEEE Trans. Broadcast. 64 (1), 11–25.

Elbamby, M.S., Perfecto, C., Bennis, M., Doppler, K., 2018. Toward low-latency and ultra-reliable virtual reality. IEEE Netw. 32 (2), 78–84.

Erfanian, A., Amirpour, H., Tashtarian, F., Timmerer, C., Hellwagner, H., 2021. LwTE: Light-weight transcoding at the edge. IEEE Access 9, 112276–112289.

Erica, P., Marivita, S., Güleç, K., Rosina, E., Grazia, T., et al., 2019. Aerial platforms (uav) surveys in the vis and tir range. Applications on archaeology and agriculture. In: Geores 2019. Vol. 42, COPERNICUS GESELLSCHAFT, pp. 945–952.

Ericsson, 2018. Streaming video – from megabits to gigabytes. https://www.ericsson.com/en/reports-and-papers/mobility-report/articles/streaming-video, (Accessed on February 10, 2022).

Fan, C.-L., Lo, W.-C., Pai, Y.-T., Hsu, C.-H., 2019. A survey on 360 video streaming: Acquisition, transmission, and display. ACM Comput. Surv. 52 (4), 1–36.

Ferranti, L., Bonati, L., D'Oro, S., Melodia, T., 2020. SkyCell: A prototyping platform for 5G aerial base stations. In: 2020 IEEE 21st International Symposium on" a World of Wireless, Mobile and Multimedia Networks"(WoWMoM). IEEE, pp. 329–334.

Fotouhi, A., Qiang, H., Ding, M., Hassan, M., Giordano, L.G., Garcia-Rodriguez, A., Yuan, J., 2019. Survey on UAV cellular communications: Practical aspects, standardization advancements, regulation, and security challenges. IEEE Commun. Surv. Tutor. 21 (4), 3417–3442.

Gao, Y., Kim, M., Abuadbba, S., Kim, Y., Thapa, C., Kim, K., Camtepe, S.A., Kim, H., Nepal, S., 2020. End-to-end evaluation of federated learning and split learning for internet of things. arXiv preprint arXiv:2003.13376.

Gao, J., Mu, L., Bao, X., Song, J., Ding, Y., 2018. Drift analysis of MH370 debris in the southern Indian ocean. Front. Earth Sci. 12 (3), 468–480.

Ge, C., Wang, N., Selinis, I., Cahill, J., Kavanagh, M., Liolis, K., Politis, C., Nunes, J., Evans, B., Rahulan, Y., et al., 2019. Qoe-assured live streaming via satellite backhaul in 5G networks. IEEE Trans. Broadcast. 65 (2), 381–391.

Ghanavi, R., Kalantari, E., Sabbaghian, M., Yanikomeroglu, H., Yongacoglu, A., 2018. Efficient 3D aerial base station placement considering users mobility by reinforcement learning. In: 2018 IEEE Wireless Communications and Networking Conference. WCNC, IEEE, pp. 1–6.

Ghanavi, R., Sabbaghian, M., Yanikomeroglu, H., 2019. Q-learning based aerial base station placement for fairness enhancement in mobile networks. In: 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP). IEEE, pp. 1–5.

Google Help, 2021. Create a live stream with an encoder. https://support.google.com/youtube/answer/2907883?hl=en#zippy=%2Csoftware-encoders, (Accessed on February 12, 2022).

Guo, Y., Yu, F.R., An, J., Yang, K., Yu, C., Leung, V.C., 2020. Adaptive bitrate streaming in wireless networks with transcoding at network edge using deep reinforcement learning. IEEE Trans. Veh. Technol. 69 (4), 3879–3892.

Ha, Y.J., Yoo, M., Park, S., Jung, S., Kim, J., 2021. Secure aerial surveillance using split learning. In: 2021 Twelfth International Conference on Ubiquitous and Future Networks. ICUFN, IEEE, pp. 434–437.

Hamza, A., Akram, U., Samad, A., Khosa, S.N., Fatima, R., Mushtaq, M.F., 2020. Unmaned aerial vehicles threats and defence solutions. In: 2020 IEEE 23rd International Multitopic Conference. INMIC, IEEE, pp. 1–6.

Han, S., Go, Y., Noh, H., Song, H., 2019. Cooperative server-client http adaptive streaming system for live video streaming. In: 2019 International Conference on Information Networking. ICOIN, IEEE, pp. 176–180.

Hao, G., Ni, W., Tian, H., Cao, L., 2020. Mobility-aware trajectory design for aerial base station using deep reinforcement learning. In: 2020 International Conference on Wireless Communications and Signal Processing. WCSP, IEEE, pp. 1131–1136.

Herglotz, C., Heindel, A., Kaup, A., 2017. Decoding-energy-rate-distortion optimization for video coding. IEEE Trans. Circuits Syst. Video Technol. 29 (1), 171–182.

Hu, F., Deng, Y., Aghvami, A.H., 2021. Cooperative multigroup broadcast 360° video delivery network: A hierarchical federated deep reinforcement learning approach. IEEE Trans. Wireless Commun..

Huang, Y., Han, H., Zhang, B., Su, X., Gong, Z., 2021. Supply distribution center planning in UAV-based logistics networks for post-disaster supply delivery. In: 2020 IEEE International Conference on E-Health Networking, Application & Services. HEALTHCOM, IEEE, pp. 1–6.

Huang, W., Zhou, Y., Xie, X., Wu, D., Chen, M., Ngai, E., 2018. Buffer state is enough: Simplifying the design of QoE-aware HTTP adaptive video streaming. IEEE Trans. Broadcast. 64 (2), 590–601.

Islam, N., Rashid, M.M., Pasandideh, F., Ray, B., Moore, S., Kadel, R., 2021. A review of applications and communication technologies for internet of things (Iot) and unmanned aerial vehicle (uav) based sustainable smart farming. Sustainability 13 (4), 1821.

Ivan Quinones, 2020. Common video file formats, codecs, and containers in 2020. https://www.borrowlenses.com/blog/video-file-formats/, (Accessed on February 22, 2022).

Ivić, S., Crnković, B., Arbabi, H., Loire, S., Clary, P., Mezić, I., 2020. Search strategy in a complex and dynamic environment: the MH370 case. Sci. Rep. 10 (1), 1–15.

Jedari, B., Premsankar, G., Illahi, G., Di Francesco, M., Mehrabi, A., Ylä-Jääski, A., 2020. Video caching, analytics, and delivery at the wireless edge: A survey and future directions. IEEE Commun. Surv. Tutor. 23 (1), 431–471.

Jiang, J., Hu, L., Hao, P., Sun, R., Hu, J., Li, H., 2018. Q-FDBA: improving QoE fairness for video streaming. Multimedia Tools Appl. 77 (9), 10787–10806.

Jiang, C., Li, Z., 2020. Decreasing big data application latency in satellite link by caching and peer selection. IEEE Trans. Netw. Sci. Eng. 7 (4), 2555–2565.

Jiang, X., Yu, F.R., Song, T., Leung, V.C., 2021a. Resource allocation of video streaming over vehicular networks: A survey, some research issues and challenges. IEEE Trans. Intell. Transp. Syst..

Jiang, X., Yu, F.R., Song, T., Leung, V.C., 2021b. A survey on multi-access edge computing applied to video streaming: Some research issues and challenges. IEEE Commun. Surv. Tutor. 23 (2), 871–903.

Jin, Y., Qian, Z., Yang, W., 2020. UAV cluster-based video surveillance system optimization in heterogeneous communication of smart cities. IEEE Access 8, 55654–55664.

Karaki, H.S.A., Alomari, S.A., Refai, M.H., 2019. A comprehensive survey of the vehicle motion detection and tracking methods for aerial surveillance videos. Int. J. Comput. Sci. Netw. Secur. 19 (1), 93.

Kim, G., Choi, H., 2019. Real-time ultra-wide viewing player for spatial and temporal random access. In: International Conference on Applied Computing and Information Technology. Springer, pp. 57–69.

Kim, M., Park, J., Chung, K., 2017. Content-aware rate adaptation scheme to improve stability in HTTP Adaptive Streaming. In: 2017 International Conference on Information Networking. ICOIN, IEEE, pp. 401–405.

King, J., 2021. SatCom today in Canada: Significant research: Overview of the cospas-sarsat satellite system for search and rescue. Online J. Space Commun. 2 (4), 15.

Kishk, M., Bader, A., Alouini, M.-S., 2020a. Aerial base station deployment in 6G cellular networks using tethered drones: The mobility and endurance tradeoff. IEEE Veh. Technol. Mag. 15 (4), 103–111.

Kishk, M.A., Bader, A., Alouini, M.-S., 2020b. On the 3-D placement of airborne base stations using tethered UAVs. IEEE Trans. Commun. 68 (8), 5202–5215.

Koda, Y., Park, J., Bennis, M., Yamamoto, K., Nishio, T., Morikura, M., Nakashima, K., 2020. Communication-efficient multimodal split learning for mmwave received power prediction. IEEE Commun. Lett. 24 (6), 1284–1288.

Kyrkou, C., Theocharides, T., 2020. Emergencynet: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 1687–1699.

Laitinen, J., Lemmetti, A., Vanne, J., 2020. Real-time implementation of scalable Hevc encoder. In: 2020 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 1166–1170.

Le, A.M., Truong, L.H., Quyen, T.V., Nguyen, C.V., Nguyen, M.T., 2020. Wireless power transfer near-field technologies for unmanned aerial vehicles (UAVs): A review. EAI Endorsed Trans. Ind. Netw. Intell. Syst. 7 (22), 162831.

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 Comput. Surv. 54 (8).

Li, J., Fu, K., Zhao, S., Ge, S., 2019. Spatiotemporal knowledge distillation for efficient estimation of aerial video saliency. IEEE Trans. Image Process. 29, 1902–1914.

Li, T., Sahu, A.K., Talwalkar, A., Smith, V., 2020a. Federated learning: Challenges, methods, and future directions. IEEE Signal Process. Mag. 37 (3), 50–60.

Li, X., Salehi, M.A., Joshi, Y., Darwich, M.K., Landreneau, B., Bayoumi, M., 2018. Performance analysis and modeling of video transcoding using heterogeneous cloud services. IEEE Trans. Parallel Distrib. Syst. 30 (4), 910–922.

Li, L., Shi, D., Hou, R., Chen, R., Lin, B., Pan, M., 2020b. Energy-efficient proactive caching for adaptive video streaming via data-driven optimization. IEEE Internet Things J. 7 (6), 5549–5561.

Li, C., Toni, L., Zou, J., Xiong, H., Frossard, P., 2017. Qoe-driven mobile edge caching placement for adaptive video streaming. IEEE Trans. Multimed. 20 (4), 965–984.

Li, H., Yu, S., Lu, X., Xiao, L., Wang, L.-C., 2021. Drone-aided network coding for secure wireless communications: A reinforcement learning approach. In: 2021 IEEE Global Communications Conference. GLOBECOM, IEEE, pp. 01–06.

Liu, C.H., Chen, Z., Tang, J., Xu, J., Piao, C., 2018. Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach. IEEE J. Sel. Areas Commun. 36 (9), 2059–2070.

Liu, Z., Li, Q., Chen, X., Wu, C., Ishihara, S., Li, J., Ji, Y., 2021a. Point cloud video streaming: Challenges and solutions. IEEE Netw. 35 (5), 202–209.

Liu, M., Teng, Y., Yu, F.R., Leung, V.C., Song, M., 2020. A mobile edge computing (MEC)-enabled transcoding framework for blockchain-based video streaming. IEEE Wirel. Commun. 27 (2), 81–87.

Liu, Y., Yu, F.R., Li, X., Ji, H., Leung, V.C., 2019a. Decentralized resource allocation for video transcoding and delivery in blockchain-based system with mobile edge computing. IEEE Trans. Veh. Technol. 68 (11), 11169–11185.

Liu, Z., Zhan, C., Cui, Y., Wu, C., Hu, H., 2021b. Robust edge computing in uav systems via scalable computing and cooperative computing. IEEE Wirel. Commun. 28 (5), 36–42.

Liu, D., Zhao, J., Yang, C., 2019b. Energy-saving predictive video streaming with deep reinforcement learning. In: 2019 IEEE Global Communications Conference. GLOBECOM, IEEE, pp. 1–6.

Liu, D., Zhao, J., Yang, C., Hanzo, L., 2021. Accelerating deep reinforcement learning with the aid of partial model: Energy-efficient predictive video streaming. IEEE Trans. Wireless Commun. 20 (6), 3734–3748.

López, R.R., Luengo, E.A., Orozco, A.L.S., Villalba, L.J.G., 2020. Digital video source identification based on container's structure analysis. IEEE Access 8, 36363–36375.

Lu, X., Xiao, L., Niu, G., Ji, X., Wang, Q., 2022. Safe exploration in wireless security: A safe reinforcement learning algorithm with hierarchical structure. IEEE Trans. Inf. Forensics Secur..

Luglio, M., Romano, S.P., Roseti, C., Zampognaro, F., 2019. Service delivery models for converged satellite-terrestrial 5G network deployment: A satellite-assisted CDN use-case. IEEE Netw. 33 (1), 142–150.

Luo, J., Yu, F.R., Chen, Q., Tang, L., 2019. Adaptive video streaming with edge caching and video transcoding over software-defined mobile networks: A deep reinforcement learning approach. IEEE Trans. Wireless Commun. 19 (3), 1577–1592.

Lygouras, E., Santavas, N., Taitzoglou, A., Tarchanidis, K., Mitropoulos, A., Gasteratos, A., 2019. Unsupervised human detection with an embedded vision system on a fully autonomous UAV for search and rescue operations. Sensors 19 (16), 3542.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. High-resolution aerial image labeling with convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 55 (12), 7092–7103.

Mardiansyah, D., Budi, A., 2018. UAV vision system for rescue payload delivery. In: IOP Conference Series: Materials Science and Engineering. Vol. 384, (1), IOP Publishing, 012005.

Marvasti-Zadeh, S.M., Cheng, L., Ghanei-Yakhdan, H., Kasaei, S., 2021. Deep learning for visual tracking: A comprehensive survey. IEEE Trans. Intell. Transp. Syst..

Masood, A., Nguyen, T.-V., Truong, T.P., Cho, S., 2021. Content caching in HAP-assisted multi-UAV networks using hierarchical federated learning. In: 2021 International Conference on Information and Communication Technology Convergence. ICTC, IEEE, pp. 1160–1162.

McGee, J., Mathew, S.J., Gonzalez, F., 2020. Unmanned aerial vehicle and artificial intelligence for thermal target detection in search and rescue applications. In: 2020 International Conference on Unmanned Aircraft Systems. ICUAS, IEEE, pp. 883–891.

Mukherjee, A., Misra, S., Raghuwanshi, N.S., 2019. A survey of unmanned aerial sensing solutions in precision agriculture. J. Netw. Comput. Appl. 148, 102461.

Nasir, A.A., 2021. Latency optimization of UAV-enabled MEC system for virtual reality applications under rician fading channels. IEEE Wirel. Commun. Lett..

Nguyen, T.-V., Tran, A.-T., Dao, N.-N., Moon, H., Cho, S., 2023. Information fusion on delivery: (a) survey on the roles of mobile edge caching systems. Inf. Fusion 89, 486–509.

Niu, Z., Shen, X.S., Zhang, Q., Tang, Y., 2020. Space-air-ground integrated vehicular network for connected and automated vehicles: Challenges and solutions. Intell. Converged Netw. 1 (2), 142–169.

Pang, H., Zhang, C., Wang, F., Liu, J., Sun, L., 2019. Towards low latency multi-viewpoint 360 interactive video: A multimodal deep reinforcement learning approach. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications. IEEE, pp. 991–999.

Park, G.S., Song, H., 2018. Cooperative base station caching and X2 link traffic offloading system for video streaming over SDN-enabled 5G networks. IEEE Trans. Mob. Comput. 18 (9), 2005–2019.

Perfecto, C., Elbamby, M.S., Del Ser, J., Bennis, M., 2020. Taming the latency in multi-user VR 360°: A QoE-aware deep learning-aided multicast framework. IEEE Trans. Commun. 68 (4), 2491–2508.

Petrosino, A., Piro, G., Grieco, L.A., Boggia, G., 2022. An optimal allocation framework of security virtual network functions in 6G satellite deployments. In: 2022 IEEE 19th Annual Consumer Communications & Networking Conference. CCNC, IEEE, pp. 917–920.

Podder, K.K., Tabassum, S., Khan, L.E., Salam, K.M.A., Maruf, R.I., Ahmed, A., 2021. Design of a sign language transformer to enable the participation of persons with disabilities in remote healthcare systems for ensuring universal healthcare coverage. In: 2021 IEEE Technology & Engineering Management Conference-Europe (TEMSCON-EUR). IEEE, pp. 1–6.

Qin, Z., Liu, Z., Han, G., Lin, C., Guo, L., Xie, L., 2021. Distributed UAV-BSs trajectory optimization for user-level fair communication service with multi-agent deep reinforcement learning. IEEE Trans. Veh. Technol. 70 (12), 12290–12301.

Rahim, T., Musaddiq, A., Kim, D.-S., 2021. E-health and resource management scheme for a deep learning-based detection of tumor in wireless capsule endoscopy videos. In: 2021 Twelfth International Conference on Ubiquitous and Future Networks. ICUFN, IEEE, pp. 48–52.

Roosjen, P.P., Kellenberger, B., Kooistra, L., Green, D.R., Fahrentrapp, J., 2020. Deep learning for automated detection of Drosophila suzukii: potential for UAV-based monitoring. Pest Manag. Sci. 76 (9), 2994–3002.

Sacoto-Martins, R., Madeira, J., Matos-Carvalho, J.P., Azevedo, F., Campos, L.M., 2020. Multi-purpose low latency streaming using unmanned aerial vehicles. In: 2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing. CSNDSP, IEEE, pp. 1–6.

Saroni, A.N., Abd Samat, M.A., Ibrahim, J., 2019. The case study of emergency response plan (ERP) implementation during the Malaysia airlines flight mh370 disappearance. Malays. J. Comput. 4 (2), 270–277.

Sarwar, F., Griffin, A., Chong, P.H.J., Pasang, T., 2021. Pasture fence line detection in UAV videos. In: 2021 36th International Conference on Image and Vision Computing New Zealand. IVCNZ, IEEE, pp. 1–6.

SES, 2022. https://www.ses.com/our-coverage/o3b-mpower, (Accessed on January 25, 2022).

Shafi, U., Mumtaz, R., Iqbal, N., Zaidi, S.M.H., Zaidi, S.A.R., Hussain, I., Mahmood, Z., 2020. A multi-modal approach for crop health mapping using low altitude remote sensing, internet of things (IoT) and machine learning. IEEE Access 8, 112708–112724.

Shakeri, R., Al-Garadi, M.A., Badawy, A., Mohamed, A., Khattab, T., Al-Ali, A.K., Harras, K.A., Guizani, M., 2019. Design challenges of multi-UAV systems in cyber-physical applications: A comprehensive survey and future directions. IEEE Commun. Surv. Tutor. 21 (4), 3340–3385.

Sheng, J., Cai, X., Li, Q., Wu, C., Ai, B., Wang, Y., Kadoch, M., Yu, P., 2021. Space-air-ground integrated network development and applications in high-speed railways: A survey. IEEE Trans. Intell. Transp. Syst..

Shimizu, S., Ishizawa, J., Sakamoto, H., Nakamura, K., 2019. Ship monitoring in japan using sar, ais and earth observation satellites. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 4731–4733.

Shuai, Y., Herfet, T., 2018. Towards reduced latency in adaptive live streaming. In: 2018 15th IEEE Annual Consumer Communications & Networking Conference. CCNC, IEEE, pp. 1–4.

Silva, S.H., Rad, P., Beebe, N., Choo, K.-K.R., Umapathy, M., 2019. Cooperative unmanned aerial vehicles with privacy preserving deep vision for real-time object identification and tracking. J. Parallel Distrib. Comput. 131, 147–160.

Spurny, V., Pritzl, V., Walter, V., Petrlik, M., Baca, T., Stepan, P., Zaitlik, D., Saska, M., 2021. Autonomous firefighting inside buildings by an unmanned aerial vehicle. IEEE Access 9, 15872–15890.

Starlink, 2022. https://www.starlink.com/, (Accessed on January 25, 2022).

Sthapit, S., Lakshminarayana, S., He, L., Epiphaniou, G., Maple, C., 2021. Reinforcement learning for security aware computation offloading in satellite networks. IEEE Internet Things J..

Stornig, A., Fakhreddine, A., Hellwagner, H., Popovski, P., Bettstetter, C., 2021. Video quality and latency for UAV teleoperation over LTE: A study with ns3. In: 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring). IEEE, pp. 1–7.

Stumph, B., Virto, M.H., Medeiros, H., Tabb, A., Wolford, S., Rice, K., Leskey, T., 2019. Detecting invasive insects with unmanned aerial vehicles. In: 2019 International Conference on Robotics and Automation. ICRA, IEEE, pp. 648–654.

Sun, X., Ng, D.W.K., Ding, Z., Xu, Y., Zhong, Z., 2019. Physical layer security in UAV systems: Challenges and opportunities. IEEE Wirel. Commun. 26 (5), 40–47.

Sun, Y., Yi, S., Hou, F., Luo, D., Hu, J., Zhou, Z., 2020a. Quantifying the dynamics of livestock distribution by unmanned aerial vehicles (UAVs): A case study of yak grazing at the household scale. Rangel. Ecol. Manag. 73 (5), 642–648.

Sun, S., Zhang, G., Mei, H., Wang, K., Yang, K., 2020b. Optimizing multi-UAV deployment in 3-D space to minimize task completion time in UAV-enabled mobile edge computing systems. IEEE Commun. Lett. 25 (2), 579–583.

Sun, L., Zong, T., Wang, S., Liu, Y., Wang, Y., 2021. Towards optimal low-latency live video streaming. IEEE/ACM Trans. Netw..

Sunny, A., El-Azouzi, R., Arfaoui, A., Altman, E., Poojary, S., Tsilimantos, D., Valentin, S., 2019. Enforcing bitrate-stability for adaptive streaming traffic in cellular networks. IEEE Trans. Netw. Serv. Manag. 16 (4), 1812–1825.

Tang, K., Kan, N., Zou, J., Li, C., Fu, X., Hong, M., Xiong, H., 2020. Multi-user adaptive video delivery over wireless networks: A physical layer resource-aware deep reinforcement learning approach. IEEE Trans. Circuits Syst. Video Technol. 31 (2), 798–815.

Tran, T.X., Pompili, D., 2018. Adaptive bitrate video caching and processing in mobile-edge computing networks. IEEE Trans. Mob. Comput. 18 (9), 1965–1978.

Usman, M., Jan, M.A., He, X., Chen, J., 2019. P2DCA: a privacy-preserving-based data collection and analysis framework for IoMT applications. IEEE J. Sel. Areas Commun. 37 (6), 1222–1230.

Veillon, V., Denninnart, C., Salehi, M.A., 2019. F-FDN: Federation of fog computing systems for low latency video streaming. In: 2019 IEEE 3rd International Conference on Fog and Edge Computing. ICFEC, IEEE, pp. 1–9.

Vepakomma, P., Gupta, O., Swedish, T., Raskar, R., 2018. Split learning for health: Distributed deep learning without sharing raw patient data. arXiv preprint arXiv:1812.00564.

Wainer, G., Fernandes, S., et al., 2017. Improving video streaming over cellular networks with DASH-based device-to-device streaming. In: 2017 International Symposium on Performance Evaluation of Computer and Telecommunication Systems. SPECTS, IEEE, pp. 1–8.

Wang, J., Feng, Z., Chen, Z., George, S., Bala, M., Pillai, P., Yang, S.-W., Satyanarayanan, M., 2018a. Bandwidth-efficient live video analytics for drones via edge computing. In: 2018 IEEE/ACM Symposium on Edge Computing. SEC, IEEE, pp. 159–173.

Wang, C., Guan, J., Feng, T., Zhang, N., Cao, T., 2019a. BitLat: Bitrate-adaptivity and latency-awareness algorithm for live video streaming. In: Proceedings of the 27th ACM International Conference on Multimedia. pp. 2642–2646.

Wang, N., Nouwell, N., Ge, C., Evans, B., Rahulan, Y., Boutin, M., Desmauts, J., Liolis, K., Politis, C., Votts, S., et al., 2018b. Satellite support for enhanced mobile broadband content delivery in 5G. In: 2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. BMSB, IEEE, pp. 1–6.

Wang, D., Peng, Y., Ma, X., Ding, W., Jiang, H., Chen, F., Liu, J., 2018c. Adaptive wireless video streaming based on edge computing: Opportunities and approaches. IEEE Trans. Serv. Comput. 12 (5), 685–697.

Wang, Y., Su, Z., Zhang, N., Benslimane, A., 2020a. Learning in the air: Secure federated learning for UAV-assisted crowdsensing. IEEE Trans. Netw. Sci. Eng. 8 (2), 1055–1069.

Wang, F., Wang, F., Liu, J., Shea, R., Sun, L., 2020b. Intelligent video caching at network edge: A multi-agent deep reinforcement learning approach. In: IEEE INFOCOM 2020-IEEE Conference on Computer Communications. IEEE, pp. 2499–2508.

Wang, H., Xia, X., Song, T., Xing, Y., 2021a. Survey on space-air-ground integrated networks in 6G. In: 2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops). IEEE, pp. 315–320.

Wang, J., Zhang, H., Guo, S., Yuan, D., 2021b. Trajectory design and resource allocation for tethered-UAV assisted wireless networks. In: 2021 IEEE/CIC International Conference on Communications in China. ICCC, IEEE, pp. 647–652.

Wang, T., Zheng, Z., Lin, Y., Yao, S., Xie, X., 2019b. Reliable and robust unmanned aerial vehicle wireless video transmission. IEEE Trans. Reliab. 68 (3), 1050–1060.

Wassermann, S., Seufert, M., Casas, P., Gang, L., Li, K., 2019. Let me decrypt your beauty: Real-time prediction of video resolution and bitrate for encrypted video streaming. In: 2019 Network Traffic Measurement and Analysis Conference. TMA, IEEE, pp. 199–200.

Wei, X., Zhou, M., Kwong, S., Yuan, H., Wang, S., Zhu, G., Cao, J., 2021. Reinforcement learning-based qoe-oriented dynamic adaptive streaming framework. Inform. Sci. 569, 786–803.

Wenjian, Z., Sidong, Z., RongJie, C., Jingchang, X., Yeqian, L., Huiru, C., 2020. Design of a relief materials delivery system based on UAV. In: IOP Conference Series: Materials Science and Engineering. Vol. 715, (1), IOP Publishing, 012049.

Wieckowski, A., Brandenburg, J., Hinz, T., Bartnik, C., George, V., Hege, G., Helmrich, C., Henkel, A., Lehmann, C., Stoffers, C., et al., 2021. Vvenc: An open and optimized vvc encoder implementation. In: 2021 IEEE International Conference on Multimedia & Expo Workshops. ICMEW, IEEE, pp. 1–2.

Xiao, L., Ding, Y., Huang, J., Liu, S., Tang, Y., Dai, H., 2021. UAV anti-jamming video transmissions with QoE guarantee: A reinforcement learning-based approach. IEEE Trans. Commun. 69 (9), 5933–5947.

Yaacoub, E., Alouini, M.-S., 2020. A key 6G challenge and opportunity—connecting the base of the pyramid: A survey on rural connectivity. Proc. IEEE 108 (4), 533–582.

Yang, P., Lyu, F., Wu, W., Zhang, N., Yu, L., Shen, X.S., 2019a. Edge coordinated query configuration for low-latency and accurate video analytics. IEEE Trans. Ind. Inform. 16 (7), 4855–4864.

Yang, Z., Pan, C., Wang, K., Shikh-Bahaei, M., 2019b. Energy efficient resource allocation in UAV-enabled mobile edge computing networks. IEEE Trans. Wireless Commun. 18 (9), 4576–4589.

Yang, F., Sakti, S., Wu, Y., Nakamura, S., 2019c. A framework for knowing who is doing what in aerial surveillance videos. IEEE Access 7, 93315–93325.

Yang, R., Xu, M., Wang, Z., 2017. Decoder-side HEVC quality enhancement with scalable convolutional neural network. In: 2017 IEEE International Conference on Multimedia and Expo. ICME, IEEE, pp. 817–822.

Yaqoob, A., Bi, T., Muntean, G.-M., 2020. A survey on adaptive 360 video streaming: Solutions, challenges and opportunities. IEEE Commun. Surv. Tutor. 22 (4), 2801–2838.

Yaxley, K.J., McIntyre, N., Park, J., Healey, J., 2021. Sky shepherds: a tale of a UAV and sheep. In: Shepherding UxVs for Human-Swarm Teaming. Springer, pp. 189–206.

Yingst, A.L., Marojevic, V., 2021. Tethered UAV with high gain antenna for BVLOS CNPC: A practical design for widespread use. In: 2021 IEEE 22nd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM). IEEE, pp. 323–328.

Yinka-Banjo, C., Ajayi, O., 2019. Sky-farmers: Applications of unmanned aerial vehicles (UAV) in agriculture. Auton. Veh. 107–128.

Yu, J., Vandanapu, A., Qu, C., Wang, S., Calyam, P., 2020. Energy-aware dynamic computation offloading for video analytics in multi-UAV systems. In: 2020 International Conference on Computing, Networking and Communications. ICNC, IEEE, pp. 641–647.

Zahran, A.H., Quinlan, J.J., Ramakrishnan, K., Sreenan, C.J., 2017. SAP: Stall-aware pacing for improved DASH video experience in cellular networks. In: Proceedings of the 8th ACM on Multimedia Systems Conference. pp. 13–26.

Zeng, Y., Xu, J., Zhang, R., 2019. Energy minimization for wireless communication with rotary-wing UAV. IEEE Trans. Wireless Commun. 18 (4), 2329–2345.

Zhan, C., Hu, H., Wang, Z., Fan, R., Niyato, D., 2019. Unmanned aircraft system aided adaptive video streaming: A joint optimization approach. IEEE Trans. Multimed. 22 (3), 795–807.

Zhan, C., Huang, R., 2020a. Energy efficient adaptive video streaming with rotary-wing UAV. IEEE Trans. Veh. Technol. 69 (7), 8040–8044.

Zhan, C., Huang, R., 2020b. Energy efficient adaptive video streaming with rotary-wing UAV. IEEE Trans. Veh. Technol. 69 (7), 8040–8044.

Zhang, A., Li, Q., Chen, Y., Ma, X., Zou, L., Jiang, Y., Xu, Z., Muntean, G.-M., 2021a. Video super-resolution and caching–an edge-assisted adaptive video streaming solution. IEEE Trans. Broadcast..

Zhang, S., Liu, W., Ansari, N., 2021b. On tethered UAV-assisted heterogeneous network. IEEE Trans. Veh. Technol..

Zhang, H., Rengasamy, P.V., Zhao, S., Nachiappan, N.C., Sivasubramaniam, A., Kandemir, M.T., Iyer, R., Das, C.R., 2017. Race-to-sleep+ content caching+ display caching: A recipe for energy-efficient video streaming on handhelds. In: Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture. pp. 517–531.

Zhang, Z., Yang, Y., Hua, M., Li, C., Huang, Y., Yang, L., 2019. Proactive caching for vehicular multi-view 3D video streaming via deep reinforcement learning. IEEE Trans. Wireless Commun. 18 (5), 2693–2706.

Zhang, Z., Zhang, Q., Miao, J., Yu, F.R., Fu, F., Du, J., Wu, T., 2021c. Energy-efficient secure video streaming in UAV-enabled wireless networks: A safe-DQN approach. IEEE Trans. Green Commun. Netw. 1–14.

Zhang, Y., Zhang, Y., Wu, Y., Tao, Y., Bian, K., Zhou, P., Song, L., Tuo, H., 2020. Improving quality of experience by adaptive video streaming with super-resolution. In: IEEE INFOCOM 2020-IEEE Conference on Computer Communications. IEEE, pp. 1957–1966.

Zhao, B., Liu, J., Wei, Z., You, I., 2020. A deep reinforcement learning based approach for energy-efficient channel allocation in satellite internet of things. IEEE Access 8, 62197–62206.

Zhao, N., Lu, W., Sheng, M., Chen, Y., Tang, J., Yu, F.R., Wong, K.-K., 2019a. UAV-assisted emergency networks in disasters. IEEE Wirel. Commun. 26 (1), 45–51.

Zhao, Y., Shen, Q.-W., Li, W., Xu, T., Niu, W.-H., Xu, S.-R., 2019b. Latency aware adaptive video streaming using ensemble deep reinforcement learning. In: Proceedings of the 27th ACM International Conference on Multimedia. pp. 2647–2651.

Zheng, C., Liu, S., Huang, Y., Yang, L., 2020. MEC-enabled wireless VR video service: A learning-based mixed strategy for energy-latency tradeoff. In: 2020 IEEE Wireless Communications and Networking Conference. WCNC, IEEE, pp. 1–6.

Zheng, J., Peng, W., Wang, Y., Zhai, B., 2021. Accelerated RANSAC for accurate image registration in aerial video surveillance. IEEE Access 9, 36775–36790.

Zhou, C., Lin, C.-W., Zhang, X., Guo, Z., 2017. TFDASH: A fairness, stability, and efficiency aware rate control approach for multiple clients over DASH. IEEE Trans. Circuits Syst. Video Technol. 29 (1), 198–211.

Zhou, Y., Pan, C., Yeoh, P.L., Wang, K., Elkashlan, M., Vucetic, B., Li, Y., 2019. Secure communications for UAV-enabled mobile edge computing systems. IEEE Trans. Commun. 68 (1), 376–388.

Zhou, M., Wei, X., Kwong, S., Jia, W., Fang, B., 2020a. Rate control method based on deep reinforcement learning for dynamic video sequences in HEVC. IEEE Trans. Multimed. 23, 1106–1121.

Zhou, C., Wu, W., He, H., Yang, P., Lyu, F., Cheng, N., Shen, X., 2020b. Deep reinforcement learning for delay-oriented IoT task scheduling in SAGIN. IEEE Trans. Wireless Commun. 20 (2), 911–925.

The-Vinh Nguyen is currently pursuing his M.S. degree at the Department of Computer Science and Engineering, Sejong University, Republic of Korea. He received his B.S. degree in Electrical and Automation Engineering at HAMK University of Applied Sciences, Finland, in 2020. His research interests include machine Learning, edge caching/computing, and wireless communication.

Ngoc Phi Nguyen received the B.S. and M.S. degrees in mechatronics engineering in the HCMC University of Technology and Education and the Vietnamese-German University, Vietnam, in 2012 and 2015, respectively. He is currently a visiting lecturer with the Department of Technology, Dong Nai Technology, Vietnam. His research interests include UAV applications and Nonlinear control.

Cheonshik Kim received his M.S. degree in Computer Engineering from Hankuk University of Foreign Studies (HUFS), Korea, in 1997; and his Ph.D. degree in Computer Engineering from HUFS in 2003. He is currently a Professor with the Department of Computer Science and Engineering, Sejong University, Seoul, Republic of Korea. His research interests include coding theory, Data Hiding, and Watermarking.

Nhu-Ngoc Daois an Assistant Professor at the Department of Computer Science and Engineering, Sejong University, Seoul, Republic of Korea. He received his M.S. and Ph.D. degrees in computer science at the School of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea, in 2016 and 2019, respectively. He received the B.S. degree in electronics and telecommunications from the Posts and Telecommunications Institute of Technology, Hanoi, Viet Nam, in 2009. Prior to joining the Sejong University, he was a visiting researcher at the University of Newcastle, NSW, Australia, in 2019 and a postdoc researcher at the Institute of Computer Science, University of Bern, Switzerland, from 2019 to 2020. He currently serves as an Editor of the Scientific Reports. His research interests include network softwarization, mobile cloudization, intelligent systems, and the Intelligence of Things. Dr. Dao is a Senior Member of IEEE and a Professional Member of ACM.