UAV Anti-Jamming Video Transmissions With QoE Guarantee: A Reinforcement Learning-Based Approach

Liang Xiao, *Senior Member, IEEE*, Yuzhen Ding, *Student Member, IEEE*, Jinhao Huang, Sicong Liu, *Member, IEEE*, Yuliang Tang, *Member, IEEE*, and Huaiyu Dai, *Fellow, IEEE*

***Abstract*— Unmanned aerial vehicles (UAVs) that are widely utilized for video capturing, processing and transmission have to address jamming attacks with dynamic topology and limited energy. In this paper, we propose a reinforcement learning (RL)-based UAV anti-jamming video transmission scheme to choose the video compression quantization parameter, the channel coding rate, the modulation and power control strategies against jamming attacks. More specifically, this scheme applies RL to choose the UAV video compression and transmission policy based on the observed video task priority, the UAV-controller channel state and the received jamming power. This scheme enables the UAV to guarantee the video quality-of-experience (QoE) and reduce the energy consumption without relying on the jamming model or the video service model. A safe RL-based approach is further proposed, which uses deep learning to accelerate the UAV learning process and reduce the video transmission outage probability. The computational complexity is provided and the optimal utility of the UAV is derived and verified via simulations. Simulation results show that the proposed schemes significantly improve the video quality and reduce the transmission latency and energy consumption of the UAV compared with existing schemes.**

***Index Terms*— Unmanned aerial vehicles, video transmission, quality-of-experience, jamming, reinforcement learning.**

I. Introduction

UNMANNED aerial vehicles (UAVs) equipped with high definition cameras and sensors support booming multimedia services such as video surveillance, geographical photography and patrol, owing to the high mobility, line-of-sight channel with few obstacles, low cost, convenience and safety of UAVs [1]. With limited processor performance and storage, UAVs have to process the captured video streams and transmit them to the control stations (CSs) on the ground while satisfying the quality-of-experience (QoE) of the multimedia services in a dynamic network [2]. The broadcast nature and the line-of-sight dominant UAV-ground channel make UAV video transmissions more vulnerable to jamming attacks, causing security challenges [3]–[5]. Especially, smart jammers equipped with intelligent programmable radio devices, which can optimize the waveforms and power of the jamming signals are more detrimental to the UAV video service quality, and might even launch denial-of-service attacks [6].

Video compression coding, channel coding and modulation are essential to guarantee the reliability and efficiency of the UAV video service against jamming. For instance, the widely applied H.264 standard utilizes encoding techniques controlled by the video quantization parameter (QP) to compress the UAV video data to overcome the limited processor performance and storage [7]. Most UAV video transmission systems apply the forward error check channel coding such as the capacity-approaching low density parity check (LDPC) code and modulate the data before transmission to improve the reliability, quality and efficiency [8]. However, these systems usually use constant compression and coding parameters, such as the QP, the channel coding rate and the modulation type, which is unable to meet the various requirements of the video QoE with dynamic UAV-controller channel conditions in the presence of jamming attacks. For adaptive modulation schemes, the number of bits per symbol can be set adaptively to make a tradeoff between the reliability and the spectrum efficiency [9]. However, existing adaptive modulation techniques usually apply a fixed modulation type for each level of signal-to-interference-plus-noise ratio (SINR), which is not always the optimal policy in the dynamic UAV network.

UAV cooperative relay communication, trajectory optimization, power control and other resource allocation techniques are helpful to resist jamming attacks and eavesdropping, and improve the security performance [10]–[12]. For example, a UAV anti-jamming scheme proposed in [13] uses Q-learning to obtain the optimal power control strategy, which makes a tradeoff between the power consumption and the SINR of the received signal, but the video service QoE is not addressed.

Manuscript received January 9, 2021; revised May 2, 2021; accepted June 3, 2021. Date of publication June 9, 2021; date of current version September 16, 2021. This work was supported by the National Natural Science Foundation of China under Grants 61901403, 61971366, 61731012 and 61971365, in part by the Youth Innovation Fund of Xiamen under grant 3502Z20206039, in part by the Natural Science Foundation of Fujian Province under grant 2019J05001, and in part by the Fundamental Research Funds for the central universities No. 20720200077. The associate editor coordinating the review of this article and approving it for publication was V. Aggarwal. *(Corresponding author: Sicong Liu.)*

Liang Xiao, Yuzhen Ding, Jinhao Huang, Sicong Liu, and Yuliang Tang are with the Department of Information and Communication Engineering, Xiamen University, Xiamen 361005, China (e-mail: liusc@xmu.edu.cn).

Huaiyu Dai is with the Department of Electrical and Computer Engineering, NC State University, Raleigh, NC 27695 USA.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2021.3087787.

Digital Object Identifier 10.1109/TCOMM.2021.3087787

In addition, many existing UAV anti-jamming schemes use machine learning and deep learning to improve detection accuracy and data quality [14], [15], [15], [16]. For example, a UAV-based real-time multimedia streaming delivery scheme proposed in [15] uses long-short term memory and recurrent neural networks to optimize its beamwidth, tilt angle and trajectory to improve the SINR. However, some conventional model-based machine learning and deep learning methods are dependent on the accurate models of the system and the attackers. Without sufficient data from practice, the generalization performance might degrade in dynamic UAV-based applications in practice.

In this paper, a UAV anti-jamming video transmission scheme based on model-free reinforcement learning (RL) techniques is proposed to address the problems of the existing methods and reduce the energy consumption with the video QoE guarantee. To be specific, the video encoder of the UAV selects a proper QP to compress the captured video after receiving a task request with a certain priority from the CS. After channel coding and modulation, the optimal transmit power is determined for the delivery of the processed video to the CS. The anti-jamming video transmission process can be formulated as a Markov decision process (MDP), where the RL technique can be applied to determine the optimal transmission policy based on the observed state via trial-and-error without being aware of the specific video service model or the attack model. Moreover, the safe RL technique is introduced to reduce transmission outage and guarantee the video QoE, where convolutional neural networks (CNNs) are utilized to compress the high-dimensional and continuous state space, and a modified Boltzmann distribution is exploited to determine the transmission policy.

The main contributions of this paper are outlined as follows:

- A UAV anti-jamming video transmission framework is proposed, and the RL technique is applied to determine the optimal QP, channel coding rate, modulation type and transmit power to improve the video QoE and reduce the energy consumption when the video service model and the attack model are difficult to obtain. Transfer learning is used for initialization, which reduces the initial random exploration and accelerate the learning process.
- A safe RL-based approach utilizing deep learning and CNNs is proposed to compress the high-dimensional state space and further improve the video transmission performance. This scheme introduces safe RL and modifies Boltzmann distribution to avoid dangerous action exploration and reduce the transmission outage probability.
- The optimal value regarding the utility of the UAV and the computational complexity of the proposed schemes are derived. Simulations verify that the proposed schemes significantly improve the video quality, reduce the transmission latency and energy consumption compared to the benchmark schemes, and the safe RL-based scheme is asymptotically approaching the theoretical optimal value.

The structure of this paper is shown as follows. First, we review related work in Section II and present the system model in Section III. Then the proposed RL-based UAV video transmission algorithm against jamming is devised in Section IV and the safe RL-based approach is proposed in Section V. The optimal value and the computational complexity of the proposed safe RL-based algorithm are given in Section VI, and the simulation results are presented in Section VII. Finally, we conclude this paper in Section VIII.

II. Related Work

Coding parameter selection schemes used in video coding can improve the video quality. For instance, a block level adaptive quantization algorithm as proposed in [17] obtains the proper QP for each block according to the distortion costs to improve the efficiency and reduce the computational complexity of video coding. A QP adaptation approach for groups of frames proposed in [18] applies a prediction method based on the ratio of the non-zero coefficients to estimate the encoding time and bit rate accurately subject to time constraints, while the performance is dependent on the prediction accuracy. An adaptive initial QP determination scheme for H.264-based video transcoding as proposed in [19] obtains the most suitable initial QP value by formulating the R-QP model to achieve the target bit rate without increasing complexity, while this method relies on the specific model between rate and QP.

Adaptive modulation and coding techniques have been applied to make better use of the channel conditions, which is investigated in plenty of works. For example, a modulation switching scheme as proposed in [20] applies a switching criterion based on fixed signal-to-noise ratio (SNR) thresholds to satisfy the bit-error-rate (BER) requirement, while the proper thresholds are difficult to obtain. An adaptive modulation scheme proposed in [21] changes the modulation type according to the calculated desired-to-undesired signal ratio in unmanned aircraft systems. An adaptive layer switching algorithm for scalable video coding changes the modulation type and code rate according to the channel quality indicator associated with SNR to realize adaptive modulation and coding [22], while it needs to construct the mapping from SNR to channel quality indicator.

Power control, game theory and trajectory optimization are conducted in UAV networks to resist jamming [23]–[30]. A UAV-aided mobile relaying scheme proposed in [24] exploits difference-of-concave programming and a water-filling-based solution to optimize the transmit power and maximize the secrecy rate. A worst-case optimal energy allocation scheme in [26] calculates the jamming-to-signal-power ratio thresholds to provide an upper bound of video transmission distortion in a mobile cognitive radio system in the presence of an intelligent adversary, which relies on the accuracy of threshold calculation. A Stackelberg game approach for anti-jamming defence in wireless networks is proposed in [27] to choose the optimal transmission power and channel to resist jamming and improve the transmission rate. However, the unknown parameters and the inaccurate or incomplete information of the system might result in difficulty in game theoretic analysis. A UAV trajectory design algorithm is proposed in [30] to optimize the UAV movements

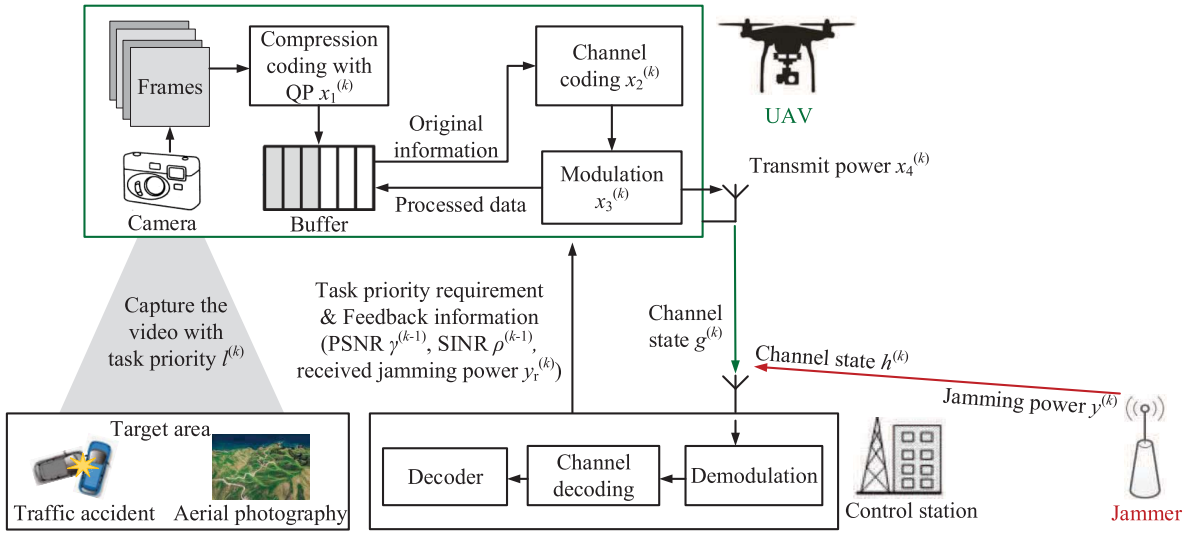


Fig. 1. Anti-jamming video transmission process in the UAV network, where the UAV receives the feedback information and task requirement from the CS and sends the processed video of the target area to the CS, and a smart jammer aims to block the UAV-CS transmission.

by using multi-agent deep Q-networks, which improves the probability of successful data transmission in the presence of interference.

RL has been applied in different applications including video coding and UAV networks. A Q-learning based UAV multimedia transmission scheme is proposed in [31] to obtain the optimal video compression and power control strategy to improve the received video quality and reduce the energy consumption of the UAV. However, the security issues in video transmission are not taken into consideration. A UAV-aided relay strategy [32] applies a hotbooting policy hill climbing algorithm to resist smart jammers and improve the BER performance, but the performance will deteriorate when the network consists of a large amount of UAVs and the state-action space is large. Deep RL and multi-agent RL have been applied in trajectory design and power control in UAV networks [33]–[39]. For example, a joint trajectory design and power control scheme proposed in [37] uses multi-agent Q-learning to maximize the instantaneous sum transmit rate while satisfying the user rate requirements. A multi-agent deep deterministic policy gradient based scheme [39] is proposed to maximize the secure capacity by jointly optimizing the trajectory of UAVs, the transmit power from UAV transmitter and the jamming power from the UAV jammers.

III. System Model

A. Network Model

The model of a UAV network for video compression and transmission is shown in Fig. 1, where a complete task of video capturing, processing and transmission is accomplished by the UAV-side compression coding, channel coding and modulation, the data transmission on the UAV-CS link, and the CS-side demodulation and decoding. In this network, a UAV equipped with camera sensors, which is located at altitude h_{U} and moving in any direction at a speed of v_{U}, can be dispatched by the CS to capture the video of the target area. The UAV compresses the captured video and transmits it to the CS after channel coding and modulation. Different tasks have different requirements of video quality and realtime performance. Thus it is important to realize a dynamic trade-off between the video quality and transmission latency.

More specifically, assuming in a time-slotted system, the CS sends the task requirement to the UAV at time slot k to obtain the video of the target area which is difficult for humans to reach. The emergency level of different video transmission tasks may vary. For instance, realtime or low latency transmission is more important in the scenario of traffic accident monitoring than aerial photography, whose requirement for video quality is higher. Thus an indicator of priority $l^{(k)}$ $(1 \leq l^{(k)} \leq L)$ is used to represent the emergency level of a task, which is divided into L levels. A higher value of priority $l^{(k)}$ indicates a stricter requirement of low latency.

The video encoder embedded in the UAV compresses the video frames to relieve the pressure of the UAV with limited storage capacity and processing capability. The video encoder quantizes the video data using the selected QP $x_1^{(k)} \in \{0, 1, \ldots, N\}$ related to the quantization interval in video compression coding as specified by the standards such as H.264, where N is the maximum value of QP. A higher value of $x_1^{(k)}$ represents a larger quantization interval and leads to a larger quantization error and compression loss. The channel encoder of the UAV applies adaptive channel coding using LDPC codes with the code rate $x_2^{(k)}$ selected from a certain range, i.e. $x_2^{(k)} \in [R_1, R_2]$. Then adaptive modulation is applied, where different modulation types such as binary phase shift keying (BPSK), quadrature phase shift keying (QPSK) and 16-quadrature amplitude modulation (QAM) can be chosen with $x_3^{(k)} \in \{1, 2, \ldots, M\}$ representing the modulation type index, with M the total number of the feasible modulation types. A higher value of the modulation parameter $x_3^{(k)}$ represents a higher-order modulation type. The number of

bits transmitted per symbol is associated with the modulation parameter $x_3^{(k)}$ and can be modeled by $b^{x_3^{(k)}-1}$, where $b > 1$ is a constant. Thus, channel coding and modulation jointly determine the amount of video data to be transmitted adaptively.

The UAV buffer stores the compressed video bitstreams waiting for the processing of channel coding and modulation. Although the buffer size is limited, it is assumed to be large enough for general video tasks with a moderate source bitstream rate and thus no congestion occurs. In an extreme case when the shot video is of very high definition and the source bitstream is of very high rate, the processing latency will increase, but the UAV can choose a larger QP to compress the source video stream or use a lower definition to shoot the video to relieve the congestion. The data coming from the compression block should be stored in the buffer in a first input first output manner, and the data rate of the original information leaving the buffer for channel coding is denoted by R_{P}, which is increased to $R_{\mathrm{P}}/x_2^{(k)}$ after channel coding. The processed data leaves the modulation block at a rate of $R_{\mathrm{B}} b^{x_3^{(k)}-1}$, where R_{B} is the baud rate, with the remaining processed data saved back in the buffer. The processing throughput denoted by $E^{(k)}$ can be measured by the difference between the amount of data stored in the buffer before and after channel coding and modulation given by

$$E^{(k)} = R_{\mathrm{P}} \frac{x_2^{(k)} - 1}{x_2^{(k)}} + R_{\mathrm{B}} b^{x_3^{(k)}-1}. \tag{1}$$

The front-end transmitter of the UAV selects a transmit power $x_4^{(k)}$ from a certain range $[X_1, X_2]$ to transmit the processed video in the modulated signal to the CS, where demodulation and decoding are conducted to reconstruct the received video upon receiving the signal from the UAV. The CS also measures the peak signal-to-noise ratio (PSNR) using no reference estimation method similar to [40] via extracting the transform coefficients by parsing the received bitstream, and then send the feedback information to the UAV.

B. Attack Model

The video transmission between the UAV and the CS is vulnerable to jamming and interference due to the mobility of the UAV and the time-varying UAV-CS link [41]–[43]. In this work, consider a smart jammer close to the CS on the ground emitting jamming signals to block the legal video transmission of the UAV. Because the CS in charge of receiving videos and controlling is fixed on the ground and the UAV located long away from the jammer can be moving, it takes the jammer a much lower cost to attack the CS receiver than the UAV without exposing itself. Thus, the jammer usually tries to get illegal reward by attacking the CS receiver rather than the UAV. The smart jammer equipped with intelligent programmable radio devices can change its geometric locations and optimize the waveforms and jamming power to adapt to the dynamic UAV networks. For the convenience of analysis, the smart jammer transmits jamming signals with a randomly selected jamming power $y^{(k)} \in [0, y_{\max}]$, where $y_{\max}$ is the maximum jamming power. If the jammer keeps silent, then $y^{(k)} = 0$. It should be noted that the random jammer is considered in this paper without loss of generality, while the proposed method is also applicable to other types of jammers, such as sweep jammer, comb jammer, single jammer, multiple jammers, and cooperative jammers. The jammer can change the jamming power to adapt to different scenarios according to the estimated channel condition. The channel state of the jammer-CS link denoted by $h^{(k)}$ can be modeled by two-ray path loss model [44], which mainly consists of a line-of-sight transmission signal and a ground reflection signal, and is often used to model the channel between two users on the ground. The proposed scheme is not limited to the two-ray model and also applicable to other channel models. The cost of the jammer is the energy consumption related to the jamming power. To quantify its utility, the jammer obtains a corresponding instant payoff, which is calculated by the difference between the illegal reward and the energy consumption, every time it selects a jamming power.

C. Video Transmission Model

In the video transmission model in the UAV network, specifically, the channel state of the UAV-CS link denoted by $g^{(k)}$ can be modeled by the typical air-to-ground channel model [45]. The video transmission is also affected by noise usually modeled by the additive white Gaussian noise (AWGN) with the power of σ^2. The SINR of the received signal at the CS denoted by $\rho^{(k)}$ is usually used to indicate the transmission quality of the signal. The BER $P_e^{(k)}$ of the received signal is used to indicate the transmission reliability, which is monotonically increasing with the channel coding rate $x_2^{(k)}$ and modulation type index $x_3^{(k)}$, and decreasing with the SINR of the received signal $\rho^{(k)}$, i.e. $P_e^{(k)} = f_{\mathrm{P}}\left(x_2^{(k)}, x_3^{(k)}, \rho^{(k)}\right)$, where $f_{\mathrm{P}}(\cdot)$ is a mapping function. A packet loss occurs when the BER is larger than a certain threshold T, so the packet error probability $P_r^{(k)}$ can be defined as $P_r^{(k)} = \mathrm{P}\left\{P_e^{(k)} > T\right\}$.

At the receiver of the CS, the received video is reconstructed through demodulation and decoding. According to [46], a fixed-point search algorithm combined with successive convex approximation optimizes the UAV trajectory and transmit beamforming to minimize the energy consumption by considering the hardware impairment effects, and simulation results show that the performance is not much different under different hardware impairment factors. Thus, this paper mainly considers the transmission loss and ignores other hardware impairments. The mean square error distortion $D^{(k)}$ between the original video and the reconstructed video expressed as $D^{(k)} = f_{\mathrm{D}}\left(x_1^{(k)}, P_r^{(k)}\right)$ consists of two components, i.e. the source distortion related to the compression ratio reflected by the QP and the channel loss distortion due to the packet error [26]. More specifically, the distortion is monotonically increasing with the QP $x_1^{(k)}$ and packet error probability $P_r^{(k)}$. A larger QP is related to a larger quantization interval. When using a larger QP for compression coding and thus a larger quantization interval is used, a larger quantization error will be introduced and the image precision is reduced, leading to more severe video compression distortion. The PSNR denoted by $\gamma^{(k)}$ is used to measure the video quality, which is related to

the mean square error distortion as $\gamma^{(k)} = 20\lg\left(G/\sqrt{D^{(k)}}\right)$ according to [47]–[49]. G denotes the dynamic range of the gray value of the image and is usually given by $G = 2^{n-1}$, where n represents the number of bits per pixel.

The QoE consists of the video quality measured by PSNR, and the transmission latency $\tau^{(k)}$, which is calculated by the total time duration from capturing the video to having successfully sent all the information by the UAV. As another important metric of the UAV video transmission, the energy consumption $C^{(k)}$ of the UAV mainly consists of two sources, i.e. energy consumed by the communication equipment for data transmission and the energy consumption of power engines for UAV movements [50]. The motor consumption of the UAV is inevitable and usually can only be adjusted by improving its own performance or optimizing its flying trajectory. In our scenario, the UAV usually files at a constant velocity at the same altitude, which has fixed motor energy consumption. Thus, it is necessary to optimize the communication energy consumption and this paper mainly considers the energy consumed by video compression, channel coding and data transmission. Since the energy resources of the UAV are quite limited, the total energy consumption should be taken care of in the video transmission scheme, in which energy efficient strategies such as power control that can adjust the transmit power of UAV and adapt to the varying channel conditions are required [51]. Hence, it is in great need to find an appropriate trade-off between the video quality, transmission latency and energy consumption in the dynamic time-varying wireless environment.

Frequently used symbols in this article are listed in Table I for the ease of reference.

TABLE I
SUMMARY OF SYMBOLS AND NOTATIONS

Symbol	Notation
$l^{(k)}$	task priority at time slot k
N	maximum value of QP
$R_{1/2}$	minimum/maximum code rate
M	number of feasible modulation types
$X_{1/2}$	minimum/maximum transmit power
$y^{(k)}$	jamming power
$y_r^{(k)}$	received jamming power by CS
$y_{\max}$	maximum jamming power
$h^{(k)}$	channel state of jammer-CS link
$g^{(k)}$	channel state of UAV-CS link
σ^2	power of the AWGN
$E^{(k)}$	processing throughput
$\rho^{(k)}$	SINR of the CS received signal
$P_e^{(k)}$	BER of the CS received signal
T	BER threshold of packet error
$P_r^{(k)}$	packet error probability
$\gamma^{(k)}$	PSNR of the reconstructed video
$\tau^{(k)}$	transmission latency
$C^{(k)}$	total energy consumption
R_{P}	rate of the data leaving the buffer
R_{B}	baud rate
N_{B}	number of bits of the processed video
d	UAV-CS distance
θ	UAV-CS elevation angle

IV. REINFORCEMENT LEARNING-BASED ANTI-JAMMING SCHEME FOR UAV VIDEO TRANSMISSION

During the interactions between the UAV and the smart jammer in the UAV video transmission process, the UAV optimizes the transmission action, and the jammer also determines its jamming action dynamically. The smart jammer may determine its jamming power according to the transmission policy of the UAV. On the other hand, the UAV has to adapt to the variation of the jamming policy in the next time slot. The current decision of the UAV and the jammer is only dependent on the latest state, so this anti-jamming video transmission process can be formulated as an MDP. The key elements in MDP are listed as follows:

State: The UAV receives the feedback information including the PSNR $\gamma^{(k-1)}$ from the CS, and measures the processing throughput $E^{(k-1)}$ through the difference of the amount of data in the buffer, the transmission latency $\tau^{(k-1)}$ via calculating the total time duration from capturing the video to having successfully sent all the information, and the energy consumption $C^{(k-1)}$ through the difference of the battery level of the previous video transmission after accomplishing a task. Upon receiving a task request with priority $l^{(k)}$ and the received jamming power $y_r^{(k)}$ from the CS, the UAV estimates the UAV-CS channel state $g^{(k)}$ and formulates the system state $\mathbf{s}^{(k)}$ as given by

$$\mathbf{s}^{(k)} = \left[l^{(k)}, g^{(k)}, y_r^{(k)}, \gamma^{(k-1)}, E^{(k-1)}, \tau^{(k-1)}, C^{(k-1)}\right]. \quad (2)$$

Action: The UAV anti-jamming video transmission policy is viewed as the action, which consists of the QP $x_1^{(k)}$, code rate $x_2^{(k)}$, modulation type $x_3^{(k)}$ and transmit power $x_4^{(k)}$. The action is chosen from the action space denoted by $\mathbf{X}$, which is the set of all feasible actions, i.e. $\mathbf{x}^{(k)} = [x_i^{(k)}]_{1\le i\le 4} \in \mathbf{X}$.

Reward: The reward or utility is the optimization object in MDP. The video encoder embedded in the UAV transmitter compresses the captured video with the selected QP $x_1^{(k)}$. The compressed video is processed through channel coding with code rate $x_2^{(k)}$ and then modulated using the modulation type specified by the parameter $x_3^{(k)}$. The processed video is then sent to the CS using the selected transmit power $x_4^{(k)}$. The CS measures the PSNR $\gamma^{(k)}$ of the reconstructed video in the received signal, and then sends feedback information to the UAV. The transmission latency $\tau^{(k)}$ and the energy consumption $C^{(k)}$ are measured. Then, the utility of the UAV denoted by $u^{(k)}$ can be given by

$$u^{(k)} = \gamma^{(k)} - \alpha_0 l^{(k)} \tau^{(k)} - \alpha_1 C^{(k)}, \quad (3)$$

where α_0 is the effective delay coefficient and α_1 is the coefficient denoting the cost per unit energy consumption. These two coefficients can be used to make a tradeoff among video quality, transmission latency and energy consumption. It can be noted from (3) that the utility is dependent on the PSNR $\gamma^{(k)}$, transmission latency $\tau^{(k)}$ and energy consumption $C^{(k)}$. The extent to which the latency deteriorates the utility is

determined by the level of the event priority $l^{(k)}$. The goal of the UAV is to find the optimal action maximizing the reward.

In practical scenarios, it is infeasible to directly obtain the optimal video transmission policy since much of the necessary information, such as the channel state, the transmission and attack models, and the jamming power, is unknown or not perfectly known. Besides, the transmission environment keeps changing and the parameters and conditions are also time-varying. It is difficult to use dynamic programming or game-theoretic methods to solve this problem while RL techniques do not need to know such information. Thus, we propose an RL-based anti-jamming (RL-AJ) scheme for UAV video transmission to achieve an optimal policy for the UAV in the MDP mentioned above, in which the UAV plays a role of a learning agent and determines the optimal transmission policy based on the observed system state to maximize the Q-function, without being aware of the video service model or the attack model. The Q-function therein is the expectation of the cumulative rewards from taking actions in the current state. Based on the transfer learning technique, the Q-function can be initialized with previous experiences consisting of the action, the resulting utility and the new state under a given state and Q-values in similar anti-jamming video transmission scenarios obtained through pre-training. Thus the previous experiences rather than all zero values can be adopted to reduce the initial random exploration of the learning process.

The RL-based anti-jamming video transmission algorithm for UAV networks is summarized in **Algorithm 1**. The UAV observes the state $\mathbf{s}^{(k)}$ and chooses the action $\mathbf{x}^{(k)}$ according to the ϵ-greedy method as described in Line 7 in **Algorithm 1**. Then the actions determined by the RL-based algorithm are executed and the utility of the UAV is evaluated by (3). As described in Line 17 in **Algorithm 1**, the Q-function will be updated according to the Bellman equation iteratively using the current Q-function $Q(\mathbf{s}^{(k)}, \mathbf{x}^{(k)})$, the instantaneous utility $u^{(k)}$, and the value function of the next state denoted by $V\left(\mathbf{s}^{(k+1)}\right) = \max_{\mathbf{x}\in\mathbf{X}} Q\left(\mathbf{s}^{(k+1)}, \mathbf{x}\right)$, which represents the maximum Q-value for the given state $\mathbf{s}^{(k+1)}$ among all possible actions. The learning rate denoted by $\alpha \in (0, 1]$ represents the weight of the current Q-value on the future Q-function, and the discount factor denoted by $\delta \in [0, 1]$ represents the weight of uncertainty about the future utility in the learning process.

Algorithm 1 RL-Based Anti-Jamming (RL-AJ) Algorithm for UAV Video Transmission

1 Initialize learning rate α, discount factor δ, initial Q-function $\mathbf{Q} \leftarrow \mathbf{Q}^*$ according to the previous similar experiences, initial value function $\mathbf{V} \leftarrow \mathbf{0}$
2 **for** $k = 1, 2, \ldots$ **do**
3 UAV receives a task requirement with priority $l^{(k)}$
4 Estimate the channel state of UAV-CS link $g^{(k)}$
5 Obtain the received jamming power $y_r^{(k)}$ from the CS
6 Observe and formulate the current state $\mathbf{s}^{(k)} = \left[l^{(k)}, g^{(k)}, y_r^{(k)}, \gamma^{(k-1)}, E^{(k-1)}, \tau^{(k-1)}, C^{(k-1)}\right]$
7 Select the transmission policy $\mathbf{x}^{(k)} = [x_i^{(k)}]_{1\le i\le 4}$ according to the ϵ-greedy method, i.e. $\Pr\left(\mathbf{x}^{(k)} = \mathbf{x}^*\right) = \begin{cases} 1-\epsilon, & \mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathbf{X}} Q\left(\mathbf{s}^{(k)}, \mathbf{x}\right) \\ \frac{\epsilon}{|\mathbf{X}|-1}, & \text{otherwise.} \end{cases}$
8 Video compression with the selected QP $x_1^{(k)}$
9 Channel coding with the code rate $x_2^{(k)}$
10 Modulation with the type of $x_3^{(k)}$
11 Measure the processing throughput $E^{(k)}$
12 Transmit the processed video to the CS with the selected transmit power $x_4^{(k)}$
13 Receive the feedback information including the PSNR $\gamma^{(k)}$ and SINR $\rho^{(k)}$ from the CS
14 Measure the transmission latency $\tau^{(k)}$ and energy consumption $C^{(k)}$
15 Formulate the next state $\mathbf{s}^{(k+1)}$ after taking action $\mathbf{x}^{(k)}$
16 Obtain the utility $u^{(k)}$ via (3)
17 $Q\left(\mathbf{s}^{(k)}, \mathbf{x}^{(k)}\right) \leftarrow (1-\alpha) Q\left(\mathbf{s}^{(k)}, \mathbf{x}^{(k)}\right) + \alpha\left(u^{(k)} + \delta \max_{\mathbf{x}\in\mathbf{X}} Q\left(\mathbf{s}^{(k+1)}, \mathbf{x}\right)\right)$
18 **end**

V. Safe Reinforcement Learning-Based Anti-Jamming Scheme for UAV Video Transmission

We further propose a safe RL-based anti-jamming (SRL-AJ) video transmission scheme for the tasks with more stringent requirements of video QoE to accelerate the learning process and reduce the transmission outage probability in case the SINR is too low, thus increasing the security of the UAV video transmission. Different from the proposed RL-AJ scheme using classical Q-learning, the safe RL-based scheme models the risk of transmission outage in an explicit manner using a risk network, determines the transmission policy based on the observed state using a modified Boltzmann distribution. CNNs with convolution kernels are utilized to extract the inherent features of the system and compress the high-dimensional state-action space to reduce the complexity and improve the performance. Some pre-training such as transfer learning techniques can be used to reduce the burden of neural network training and accelerate the process of convergence. In addition, CNNs can be mapped on a lightweight embedded processing platform by CNN parameter compression and quantization, which do not necessarily require excessive complexity and can be implemented on UAVs with simple hardware and limited resources [52], [53].

More specifically, in the framework of safe RL, a risk level of taking the action $\mathbf{x}^{(k)}$ in the state $\mathbf{s}^{(k)}$, i.e. $r(\mathbf{s}^{(k)}, \mathbf{x}^{(k)})$, is defined and utilized to represent the probability of transmission outage. A certain threshold σ_T is configured according to the prior knowledge of the system to determine whether a state-action pair is safe or not. The risk level is determined by the relation between the SINR $\rho^{(k)}$ and the given threshold σ_T, i.e. $r(\mathbf{s}^{(k)}, \mathbf{x}^{(k)}) = \mathrm{I}(\rho^{(k)} < \sigma_T)$. The long-term risk level denoted by $R(\mathbf{s}^{(k)}, \mathbf{x}^{(k)})$ for λ time slots in a risk network (R-network) is given by

$$R(\mathbf{s}^{(k)}, \mathbf{x}^{(k)}) = \sum_{j=0}^{\lambda} \beta^j r(\mathbf{s}^{(k+j)}, \mathbf{x}^{(k+j)}), \tag{4}$$

where $0 < \beta < 1$ is a risk discount factor.

The pseudo-code of the proposed safe RL-based algorithm is shown in **Algorithm 2**. The state $\mathbf{s}^{(k)}$ is formulated in a similar way as in (2), and reshaped into a $z_0 \times z_0$ matrix to serve as the input of the CNN. A typical CNN architecture mainly consists of Convolution (Conv) layers, pooling layers and fully-connected (FC) layers. The inherent features of the data to be trained are extracted layer-by-layer, and the classification is completed by several FC layers. The operation of Conv layers is inspired by the concept of local receptive field, while a pooling layer is mainly adopted to reduce the high data dimension. As shown in Fig. 2, the CNN consists of two Conv layers and two FC layers. A Q-network and an R-network are used to estimate the Q-values $Q(\mathbf{s}^{(k)}, \mathbf{x})$ and the long-term risk levels $R(\mathbf{s}^{(k)}, \mathbf{x})$ of the feasible policies $\mathbf{x}$ in the current state $\mathbf{s}^{(k)}$, respectively. The first Conv layer has f_1 filters with size $z_1 \times z_1$ and stride s_1, and the second Conv layer has f_2 filters with size $z_2 \times z_2$ and stride s_2. The first FC layers in the Q-network and the R-network, i.e. FC 1 and FC 3, have n_1 and n_2 neurons, respectively. The output layers of these two networks both have $|\mathbf{X}|$ neurons, with each neuron representing one of the feasible policies. The hyper-parameters related to the CNN configuration are assembled to a hyper-parameter vector denoted by $\mathbf{F} = [f_1, f_2, z_1, z_2, s_1, s_2, n_1, n_2]$.

The video transmission policy is determined according to the conditional probability distribution $\pi(\mathbf{x}|\mathbf{s}^{(k)})$ to select the action $\mathbf{x}$ in the state $\mathbf{s}^{(k)}$, which is obtained through the modified Boltzmann distribution [54] given by

$$\pi\left(\mathbf{x}|\mathbf{s}^{(k)}\right) = \frac{\exp\left(\frac{Q\left(\mathbf{s}^{(k)},\mathbf{x}\right)}{R\left(\mathbf{s}^{(k)},\mathbf{x}\right)+1}\right) \mathrm{I}\left(R\left(\mathbf{s}^{(k)},\mathbf{x}\right) < \xi\right)}{\sum_{\mathbf{x}'} \exp\left(\frac{Q\left(\mathbf{s}^{(k)},\mathbf{x}'\right)}{R\left(\mathbf{s}^{(k)},\mathbf{x}'\right)+1}\right) \mathrm{I}\left(R\left(\mathbf{s}^{(k)},\mathbf{x}'\right) < \xi\right)}, \tag{5}$$

where $R(\mathbf{s}^{(k)}, \mathbf{x})$ can be regarded as a temperature parameter in the modified Boltzmann distribution used to adjust the randomness of the decisions, and ξ is the threshold of the risk level tolerance to avoid transmission outage.[1] It can be observed from the conditional probability distribution given in (5) that, the action with a higher Q-value and a lower risk level within the risk level tolerance is more likely to be selected. It is implied that the actions that cause the risk level to exceed the threshold ξ will hardly be selected, so the transmission outage probability can be reduced.

Similar to the RL-AJ algorithm, the UAV evaluates the utility u based on (3), observes the next state $\mathbf{s}^{(k+1)}$, and evaluates the risk level $r(\mathbf{s}^{(k)}, \mathbf{x}^{(k)})$ based on the feedback SINR $\rho^{(k)}$. Then the UAV formulates and stores a memory transition denoted by $\mathbf{e}^{(k)} = (\mathbf{s}^{(k)}, \mathbf{x}^{(k)}, u^{(k)}, \mathbf{s}^{(k+1)}, r(\mathbf{s}^{(k)}, \mathbf{x}^{(k)}))$ in the memory pool $\mathcal{D}$, where a memory transition is a piece of record including the current state and action, the utility, the next state, and the risk level. The technique of experience replay is exploited, where a minibatch $\mathcal{B}$ is sampled from $\mathcal{D}$ to update the CNN parameters using stochastic gradient descent (SGD). More specifically, the UAV samples a mini-batch data to train the deep neural networks and calculate the loss function in a forward propagation manner. Then the gradient with respect to each component or weight of the CNN is calculated by the method of back propagation, which is used to update the network weights and parameters iteratively using the training data set. The weight of the Q-network θ_{Q} is updated by minimizing the loss function given by

$$\mathcal{L}(\theta_{\mathrm{Q}}) = \mathbb{E}_{\mathbf{e}^{(i)}\in\mathcal{B}}\Bigg[\Bigg(u^{(k)} + \delta \max_{\mathbf{x}\in\mathbf{X}} Q\left(\mathbf{s}^{(i+1)}, \mathbf{x}\right) - Q\left(\mathbf{s}^{(i)}, \mathbf{x}^{(i)}\right)\Bigg)^2\Bigg], \tag{6}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator. The weight of the R-network θ_{R} is updated by minimizing the loss function given by

$$\mathcal{L}(\theta_{\mathrm{R}}) = \mathbb{E}_{\mathbf{e}^{(i)}\in\mathcal{B}}\Bigg[\Bigg(\sum_{j=0}^{\lambda} \beta^j r\left(\mathbf{s}^{(i+j)}, \mathbf{x}^{(i+j)}\right) - R\left(\mathbf{s}^{(i)}, \mathbf{x}^{(i)}\right)\Bigg)^2\Bigg]. \tag{7}$$

Algorithm 2 Safe RL-Based Anti-Jamming (SRL-AJ) Algorithm for UAV Video Transmission

1 Initialize discount factor δ, CNN weights θ_{Q} and θ_{R} and initial risk level $r(\mathbf{s}, \mathbf{x}) = \mathbf{0}$
2 **for** $k = 1, 2, \ldots$ **do**
3 UAV receives a task requirement with priority $l^{(k)}$
4 Estimate the channel state of UAV-CS link $g^{(k)}$
5 Obtain the received jamming power $y_r^{(k)}$ from the CS
6 Observe and formulate the current state $\mathbf{s}^{(k)} = \left[l^{(k)}, g^{(k)}, y_r^{(k)}, \gamma^{(k-1)}, E^{(k-1)}, \tau^{(k-1)}, C^{(k-1)}\right]$
7 Select the transmission policy $\mathbf{x}^{(k)} = [x_i^{(k)}]_{1\le i\le 4}$ according to the conditional probability distribution $\pi(\mathbf{x}|\mathbf{s}^{(k)})$ in (5)
8 Conduct procedures same to the lines 8-15 in **Algorithm 1**
9 Obtain the utility $u^{(k)}$ via (3)
10 Observe the feedback SINR $\rho^{(k)}$ and evaluate the risk level $r(\mathbf{s}^{(k)}, \mathbf{x}^{(k)})$
11 Formulate a memory transition $\mathbf{e}^{(k)} = (\mathbf{s}^{(k)}, \mathbf{x}^{(k)}, u^{(k)}, \mathbf{s}^{(k+1)}, r(\mathbf{s}^{(k)}, \mathbf{x}^{(k)}))$
12 Store the transition into memory pool: $\mathcal{D} \leftarrow \mathcal{D} \cup \mathbf{e}^{(k)}$
13 Obtain a minibatch $\mathcal{B}$ sampled from $\mathcal{D}$
14 Update the weight of Q-network θ_{Q} via (6)
15 Update the weight of R-network θ_{R} via (7)
16 **end**

VI. Performance Evaluations

In this section, we analyze the theoretical optimal value of the proposed safe RL-based anti-jamming algorithm in the UAV video transmission network regarding the UAV utility with different constraints. The computational complexity of the proposed RL-AJ and SRL-AJ scheme is derived, respectively. The superscript k for the time slot is omitted for

[1]There is high probability that there exists at least one action $\mathbf{x}$ such that $\pi\left(\mathbf{x}|\mathbf{s}^{(k)}\right)$ is non-zero due to the high dimensionality of the action space. Even in the extreme case, random exploration can be adopted to enter a new state with a different Boltzmann probability distribution. If the extreme case keeps occurring for a long time, the UAV can change its position to reach a new distribution.

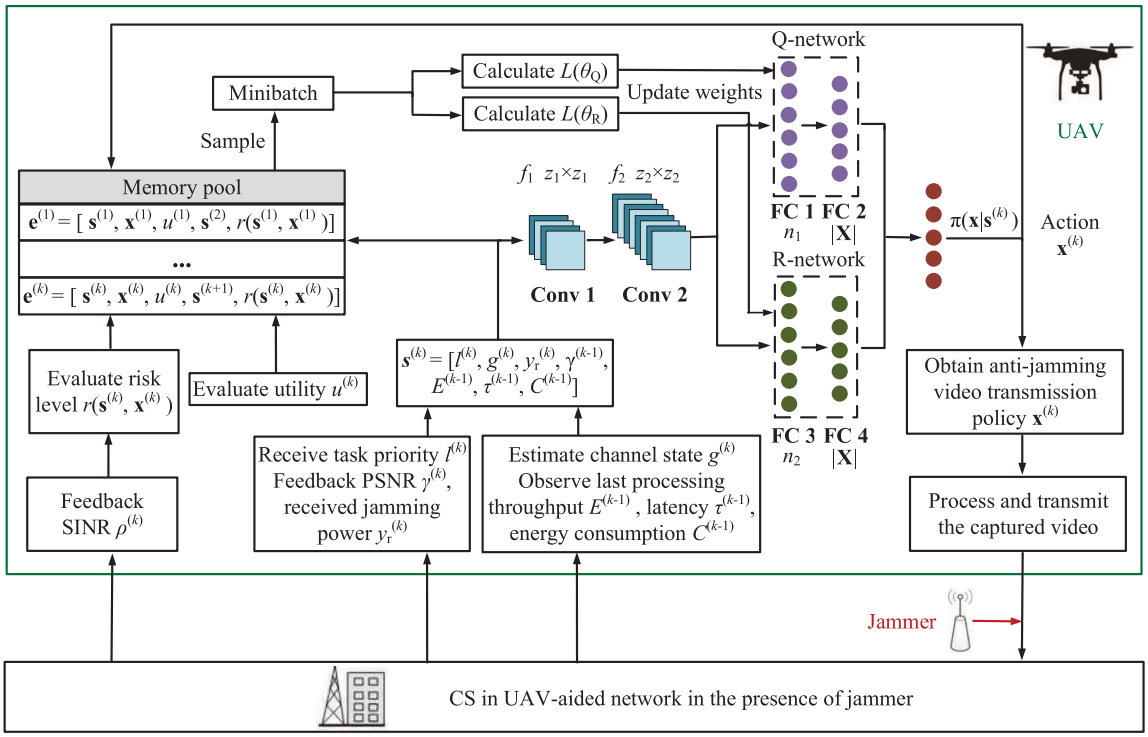


Fig. 2. Illustration of safe RL-based anti-jamming (SRL-AJ) video transmission policy selection algorithm.

notation simplicity when no confusion is incurred in the following content.

When the channel state of the UAV-CS link g and the jamming power y are estimated or measured, the SINR ρ and the BER P_e can be obtained by $\rho = x_4 g^2/(\sigma^2 + yh^2)$ and $P_e = A\text{erfc}(\sqrt{B\rho})\sqrt{2x_2 b^{x_3-1}}$, respectively, where A and B are parameters related to the modulation type, and $\text{erfc}(\cdot)$ is the complementary error function [9]. According to [47] and [55], the PSNR γ in dB is given by

$$\gamma\ (\text{dB}) = \omega_1 x_1 + \omega_2 - \nu \text{I}\left(A\text{erfc}\left(\sqrt{\frac{Bx_4 g^2}{\sigma^2 + yh^2}}\right)\sqrt{2x_2 b^{x_3-1}} > T\right), \quad (8)$$

where $\omega_1 < 0$ and $\omega_2 > 0$ are fitting parameters, and T is a certain threshold of the BER to determine packet loss, with the coefficient ν indicating the sensitivity to packet loss. The indicator function $\text{I}(\cdot)$ equals 1 when the condition therein is true and 0 otherwise. The first two terms in (8) represent the influence of source distortion caused by video compression, and the third term represents the packet loss distortion.

According to [7], the transmission latency can be given by

$$\tau = \varphi x_1^2 + \eta x_1 + \frac{N_\text{B}}{R_\text{B} b^{x_3-1}}, \quad (9)$$

where the first two terms represent the time consumed by compression coding with $\varphi > 0$ and $\eta < -2\varphi N$ being fitting parameters. The third term is the transmission delay, in which N_B denotes the bits of the processed video data needed to be transmitted, and R_B and $R_\text{B} b^{x_3-1}$ denote the baud rate and the data rate emitted out of the UAV transmitter after modulation.

The energy consumption can be modeled by $C = (a_1 x_1 + a_0) + (b_1 x_2 + b_0) + c x_4$, where the coefficients a_1 related to the video compression and b_1 related to the energy consumption of the UAV transmitter are negative, i.e. $a_1 < 0$ and $b_1 < 0$, since less processing overhead is required for video compression when the QP is larger and for channel coding when the code rate is higher. The coefficients $a_0 > 0$ and $b_0 > 0$ are modifying factors, and $c > 0$ is a time factor.

Recalling that the action $\mathbf{x}$ can be chosen from the range of $x_1 \in [0, N]$, $x_2 \in [R_1, R_2]$, $x_3 \in [1, M]$, and $x_4 \in [X_1, X_2]$, the optimal value regarding the utility of the UAV can be derived by the following theorem:

Theorem 1: The optimal value regarding the utility of the UAV in the proposed safe RL-based anti-jamming algorithm for UAV video transmission is given by (10), as shown at the bottom of the next page, with different constraints given by (11) as

$$\begin{cases} \Pi_1 : \omega_1 < \alpha_0 L\eta + \alpha_1 a_1 & (11a) \\ \Pi_2 : \alpha_0\eta + \alpha_1 a_1 < \omega_1 < \alpha_0 L(2\varphi N + \eta) + \alpha_1 a_1 & (11b) \\ \Pi_3 : \omega_1 > \alpha_0(2\varphi N + \eta) + \alpha_1 a_1 & (11c) \\ \Pi_4 : A\text{erfc}\left(\sqrt{\frac{BX_1 g^2}{\sigma^2 + y_{\max} h^2}}\right)\sqrt{2R_2 b^{M-1}} < T & (11d) \end{cases}$$

Proof: See Appendix A. ■

Remark 1: The safe RL-based algorithm can converge to the optimal video anti-jamming transmission policy with a probability of one in the dynamic environment modeled as an MDP after a sufficiently long time according to [56]. If the video quality fitting parameter ω_1 is relatively low as shown in (11a), a slight increase in QP will cause a significant

decrease in PSNR and thus the utility decreases with QP, so the minimum QP should be selected in order to compress the video without too much source distortion. If the fitting parameter ω_1 has a moderate value as shown in (11b), the utility will increase at first and then decrease with the QP, so the UAV should search for an optimal QP that maximizes the utility in the process of video compression. If ω_1 satisfies (11c), a larger QP will not cause a significant decrease in utility while the processing latency can be much smaller since the quantization interval is larger. If the UAV-CS link is in a good channel state as shown in (11d), the maximum code rate with low computational complexity can be applied for channel coding, and a higher modulation order can be applied since packet loss is less likely to occur compared with lower modulation order. Moreover, the minimum transmit power X_1 can be adopted to reduce the energy consumption of the UAV and meanwhile the utility and video QoE can be guaranteed.

The computational complexity of **Algorithm 1** denoted by $\mathcal{O}(\Gamma_1)$ is mainly contributed by the number of steps required for the algorithm to converge. Let K denote the number of steps required for the algorithm to converge in each episode, and let $N_{\rm F}$ denote the number of episodes. The computational complexity of **Algorithm 1** under the assumption that $KN_{\rm F} > \text{poly}\{|\mathbf{S}|, |\mathbf{X}|, K\}$ is given by the following theorem, where $\text{poly}\{\cdot\}$ is an operator that calculates the characteristic polynomial of an equation with vector $[|\mathbf{S}|, |\mathbf{X}|, K]$ as the solution, and $|\mathbf{S}|$ and $|\mathbf{X}|$ is the size of the state and action sets, respectively. In addition, the computational complexity of **Algorithm 2** denoted by $\mathcal{O}(\Gamma_2)$ is mainly contributed by the computational complexity in the CNN. The number of the input channels of the CNN is denoted by f_0, and m_ψ denotes the size of the output features of the convolutional layer Conv ψ. The size of the output features of Conv 1 and Conv 2 is $(z_0 - z_1)/s_1 + 1$ and $(z_0 - z_1)/(s_1 s_2) - (z_2 - 1)/s_2 + 1$ according to [57], respectively. The computational complexity of **Algorithm 2** under the assumption that $f_0 = s_1 = s_2 = 1$ is given by the following theorem.

Theorem 2: The computational complexity of the proposed RL-based anti-jamming video transmission algorithm is given by

$$\mathcal{O}(\Gamma_1) = \mathcal{O}(KN_{\rm F}). \tag{12}$$

The computational complexity of the proposed safe RL-based anti-jamming video transmission algorithm is given by

$$\mathcal{O}(\Gamma_2) = \mathcal{O}\left(f_1 f_2 z_0^2 z_2^2\right). \tag{13}$$

Proof: See Appendix B. ■

Remark 2: The computational complexity of **Algorithm 1** increases with the total number of steps required for convergence. The proposed RL-based scheme uses transfer learning to initialize the Q-values with prior experiences and reduce the initial random exploration, which can reduce the number of steps required for convergence and the computational complexity of the algorithm. The computational complexity of **Algorithm 2** increases with the number of filters of the Conv layers, i.e. f_1 and f_2, the CNN input size z_0, and the filter size of Conv 2 z_2. However, the system performance improves with the CNN input size and the number of filters, because the CNN with a larger input size is able to extract more learning experiences, and a larger number of filters can represent more features captured. Thus the CNN parameters should be properly selected to achieve a good tradeoff between the computational complexity and the learning performance.

VII. Simulation Results

Simulations are performed to evaluate the performance of the proposed RL-based and safe RL-based anti-jamming schemes for video transmission in a UAV network, which consists of a UAV located at the coordinate of $(10, 10, 200)$ m at the initial time, flying horizontally with the moving speed $v_{\rm U} = 10$ m/s, a CS located at $(0, 0, 0)$ m, and a jammer at $(100, 0, 0)$ m. The task priority is divided into 4 levels from 1 to 4. The widely applied compression coding standard H.264 is adopted to compress the captured video with the QP discretized uniformly into 6 levels from 0 to 50 with the step of 10. The LDPC code is adopted and the feasible code rate ranges from 0.5 to 0.9 with the step of 0.2 according to [58]. The number of modulation types is set as $M = 3$, and BPSK, QPSK and 16-QAM are adopted as the feasible modulation types. It should be noted that these coding and modulation types are considered as an example in our simulations, and the proposed schemes can be applied for other compression coding, channel coding and modulation modes. As described in Eq.(1) in Section III, the rate of the data leaving the buffer for channel coding is set to $R_{\rm P} = 700$ kbps according to [22], and the baud rate $R_{\rm B}$ is set to 6×10^5 symbols/s according to [59].

In the simulations, the channel state of the jammer-CS link h can be modeled by the two-ray path loss model [44]. The channel state of the UAV-CS link g can be modeled by the air-to-ground channel model [45], which consists of two parts, i.e. the line-of-sight link with probability $P_{\rm LoS} = 1/(1 + \varpi \exp(-\mu(\theta - \varpi)))$ and non-line-of-sight link with probability $P_{\rm NLoS} = 1 - P_{\rm LoS}$, where θ is the elevation angle between the UAV and the CS, and $\varpi = 0.136$ and $\mu = 11.95$ are constant parameters related to the environment. Thus, the pathloss can be modeled by $\Lambda = |d|^\kappa P_{\rm LoS} + \varepsilon |d|^\kappa P_{\rm NLoS}$, where d is the distance between the UAV and the CS, $\kappa = 3$

$$u^* = \begin{cases} \omega_2 - \dfrac{\alpha_0 N_{\rm B}}{R_{\rm B} b^{M-1}} - \alpha_1(a_0 + b_0 + b_1 R_2 + cX_1), & \Pi_1 \cap \Pi_4 \quad (10a) \\ \omega_2 + \dfrac{(\omega_1 - \alpha_0\eta - \alpha_1 a_1)^2}{4\alpha_0\varphi} - \dfrac{\alpha_0 N_{\rm B}}{R_{\rm B} b^{M-1}} - \alpha_1(a_0 + b_0 + b_1 R_2 + cX_1), & \Pi_2 \cap \Pi_4 \quad (10b) \\ \omega_2 + \big(\omega_1 - \alpha_0(\varphi N + \eta) - \alpha_1 a_1\big)N - \dfrac{\alpha_0 N_{\rm B}}{R_{\rm B} b^{M-1}} - \alpha_1(a_0 + b_0 + b_1 R_2 + cX_1), & \Pi_3 \cap \Pi_4 \quad (10c) \end{cases}$$

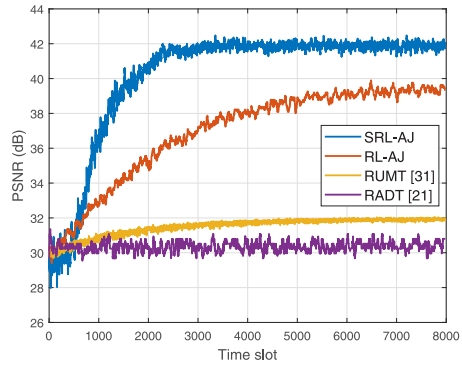
is the path loss exponent and $\varepsilon = 20$ is the additional path loss factor of the non-line-of-sight link. The channel gain g is then given by

$$g = \frac{1 + \varpi \exp\big(-\mu(\theta - \varpi)\big)}{|d|^{\kappa}\Big(1 + \varepsilon\varpi \exp\big(-\mu(\theta - \varpi)\big)\Big)}. \tag{14}$$

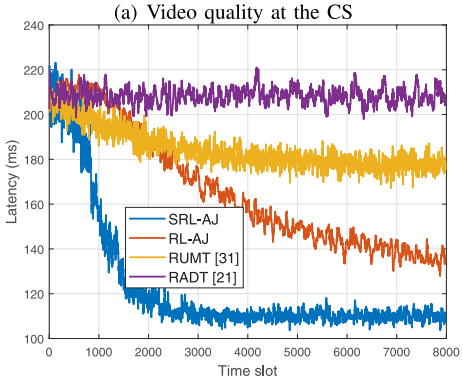
The power of the AWGN at the receiver of the CS is set to 1.2 mW. After estimating the UAV-CS channel state and measuring the jamming power in the range between 100 and 120 mW, the processed video with the size of $N_{\mathrm{B}} = 480$ kbits is transmitted to the CS using the transmit power from 100 to 180 mW with the step of 20 mW. The BER threshold T related to the packet error probability was set as 5×10^{-3} through setting different values and obtaining one maximizing the performance.

The learning rate α and the discount factor δ are set to 0.7 and 0.4 in **Algorithm 1**, respectively, unless explicitly specified otherwise. The risk discount factor β in **Algorithm 2** is set to 0.1 and the CNN hyper-parameter vector is set to $\mathbf{F} = [20, 40, 3, 2, 1, 1, 180, 180]$ through grid search, i.e. setting multiple possible values and selecting a set of values to achieve a good tradeoff between the computational complexity and the learning performance. If the setup such as the UAV trajectory or channel models has changed, appropriate hyper-parameters can still be obtained through offline tuning. The resource allocation based data transmission (RADT) scheme [21] is evaluated as a benchmark scheme, in which the code rate and the transmit power are invariant and set to 0.5 and 1 W in the simulations, respectively, which does not adapt to the dynamic jamming environment. The RADT scheme also applies adaptive modulation and employs the desired-to-undesired signal ratio as a threshold to change between different modulation types. The method in prior work called RL-based UAV media transmission (RUMT) in [31] without utilizing the hot-booting technique is also evaluated as another benchmark.

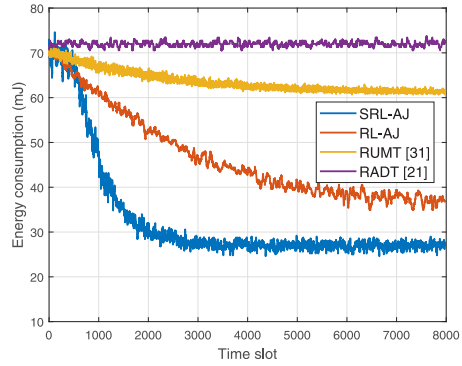
The performance of the RL-based and safe RL-based anti-jamming schemes for UAV video transmission is reported in Fig. 3. The performance is calculated using the average value over 100 episodes, each of which has 8000 time slots. It is shown by Fig. 3 that both the proposed RL-AJ and SRL-AJ schemes can jointly improve the PSNR and the utility, while reducing the transmission latency and energy consumption. For instance, the RL-AJ scheme improves the PSNR by 31.7% from 30 dB to 39.5 dB, and the SRL-AJ scheme is able to further improve it by 40.1% after 3000 time slots. The value of PSNR optimized ranging from 25 dB to 45 dB is common practice and realistic in literature [60] and [40]. For example, the bit rate can be set as 50, 200, 400 or 1000 kbps when the PSNR ranges from 25 to 45 dB according to [40], which indicates that the improvement of PSNR will not curtail the bitrates and degrade the performance. The RL-AJ scheme decreases the transmission latency by 35.6% from 208 ms to 134 ms while the SRL-AJ scheme decreases the latency by 47.6% at the 3000-th time slot. It can be observed from Fig. 3 (c) that the energy consumption of the RL-AJ scheme is decreased by 47.3% from 70 mJ to 36.9 mJ at the 8000-th time slot. The energy consumption of the SRL-AJ scheme is further decreased by 62.2% and converges much more rapidly, which reaches the optimal level at the 3000-th time slot. The utility of the UAV is improved from 1.5 to 22.3 using the RL-AJ scheme and from 1.5 to 28 using the SRL-AJ scheme.



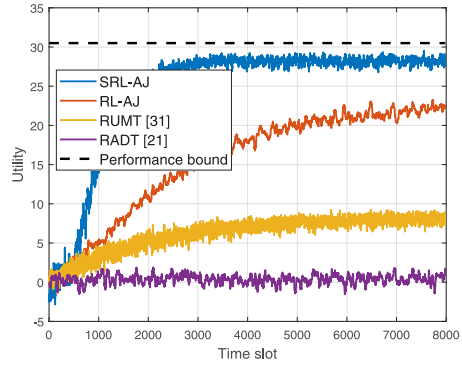
(a) Video quality at the CS



(b) Transmission latency



(c) Energy consumption of the UAV



(d) Utility of the UAV

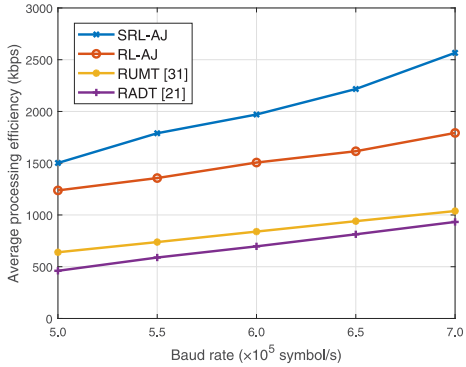
Fig. 3. Performance of the RL-based and safe RL-based anti-jamming schemes for UAV video transmission networks.

In addition, it is noted from Fig. 3 (d) that the proposed safe RL-based algorithm is approaching the theoretical optimal value given by (10) after 3000 time slots. The simulation results validate that, the proposed SRL-AJ scheme applying deep learning to extract features and compress the state-action space can accelerate the learning process, and the required time of convergence is 62.5% less than that of the RL-AJ scheme.

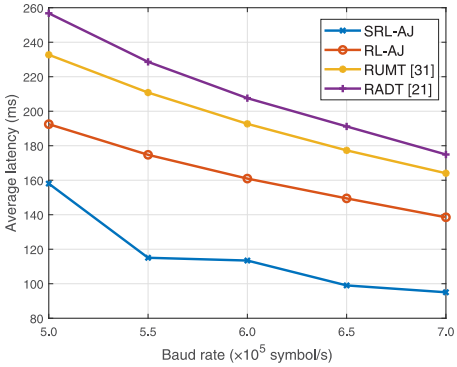
It can also be noted from the simulation results that, the performance of the proposed schemes is significantly improved compared with that of the RADT and RUMT schemes. As shown in Fig. 3 (a), the RL-AJ scheme improves the PSNR by 31.7% and 23.4% compared to the RADT and RUMT schemes, respectively, after 8000 time slots. The SRL-AJ scheme further improves the PSNR by 6.1% compared to the RL-AJ scheme. As shown in Fig. 3 (b) and (c), the SRL-AJ scheme has the lowest transmission latency and energy consumption, and the RL-AJ scheme also has a relatively low latency. More specifically, the RL-AJ scheme reduces the transmission latency by 35.6% and 24.3%, and reduces the energy consumption by 47.3% and 39.5%, compared to the RADT and RUMT schemes, respectively, at the 8000-th time slot. The SRL-AJ scheme consumes 22.9% less latency and 27.9% less energy compared to the RL-AJ scheme at convergence. In addition, according to Fig. 3 (d), the utility of the UAV using the RL-AJ scheme is 13.9 and 1.7 times larger than that using the RADT and RUMT schemes, respectively, and the SRL-AJ scheme has a 25.6% even larger utility than the RL-AJ scheme.

The performance of the proposed schemes with respect to the baud rate in the range between 5×10^5 and 7×10^5 symbols/s is reported in Fig. 4. The performance is calculated using the average value over 100 episodes, each of which has 8000 time slots. It is shown by Fig. 4 (a) and (b) that, the processing throughput of the proposed schemes of SRL-AJ and RL-AJ is increasing with the baud rate while the average transmission latency is decreasing with the baud rate. It can also be noted that, the processing throughput of the proposed SRL-AJ scheme is more than 2.5 times that of the conventional RUMT and RADT methods, and the latency of the SRL-AJ scheme is about half of that of the conventional methods, when the baud rate is 7×10^5 symbols/s. Thus, the throughput is greatly improved and the latency is significantly reduced. The SRL-AJ scheme has a superior performance because convolution neural networks are utilized to compress the state space to accelerate the learning process. Besides, the risk of transmission outage is modeled in an explicit manner using R-network and is formulated in probability distribution using the modified Boltzmann distribution, which can help to reduce the probability of the actions encountering transmission outage. In addition, as shown in Fig. 4 (c), the utility of the UAV using the SRL-AJ scheme is 1.6, 4.4 and 19.3 times larger than those using the RL-AJ, RADT and RUMT schemes, respectively, at the baud rate of 6.5×10^5 symbols/s.

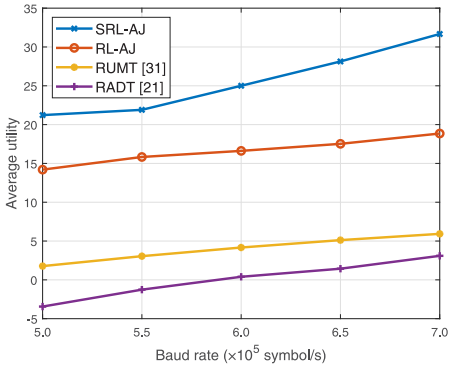
In order to show the overall performance versus the UAV altitude, we assume that the UAV camera over the target area has a maximum video shooting angle θ_{U}, and thus it has different shooting ranges at different altitudes h_{U}. The maximum shooting area is denoted by $S_{\mathrm{U}} = \pi\theta_{\mathrm{U}}^2 \tan^2\theta_{\mathrm{U}}$.



(a) Average processing throughput



(b) Average transmission latency

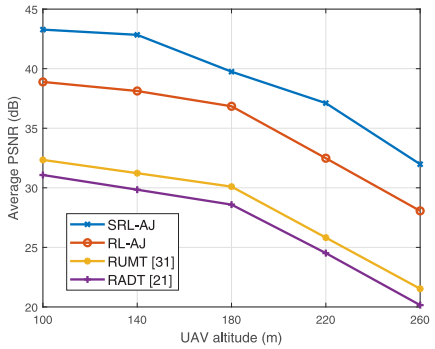


(c) Average utility of the UAV

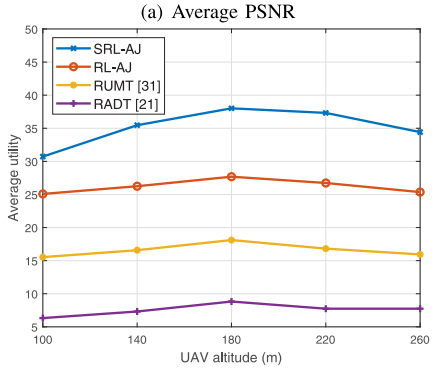
Fig. 4. Average performance of the dynamic anti-jamming process in UAV video transmission networks with respect to the baud rate.

In order to consider the influence of UAV altitude on the learning performance, the utility can be extended as $u' = u + \alpha_2 S_{\mathrm{U}}$, where α_2 is the reward coefficient. The performance of the proposed schemes with respect to the UAV altitude in the range between 100 and 260 m is reported in Fig. 5. The maximum shooting area increases with the altitude, and the monitoring range can be improved by increasing the altitude. On the other hand, if the UAV altitude is higher, the distance from the CS and the channel loss is larger. SINR and PSNR are reduced, thus having an impact on the received video quality as shown in Fig. 5 (a). As shown in Fig. 5 (b), we can find an optimal altitude, i.e. 180 m, to make a best tradeoff between the shooting area and the video quality.

Consequently, the simulation results validate that the proposed RL-AJ and SRL-AJ schemes significantly improve the PSNR, and reduce the transmission latency and energy consumption compared with benchmark schemes including RADT in [21] and RUMT in [31]. Furthermore, it can be noted



(a) Average PSNR



(b) Average utility

Fig. 5. Average performance of the dynamic anti-jamming process in UAV video transmission networks with respect to the UAV altitude.

from Fig. 3 (d) that the proposed safe RL-based algorithm is asymptotically approaching the theoretical optimal value derived in **Theorem 1**, and the learning process is further accelerated in the framework of deep RL. Moreover, the effectiveness of the proposed schemes is also verified by the average performance including the processing throughput, transmission latency and utility of the UAV as shown in Fig. 4.

VIII. Conclusion

In this paper, we have presented an RL-based UAV video transmission scheme to guarantee the QoE of the video capturing services while reducing the energy consumption against smart jamming attacks. Our proposed scheme enables UAVs to optimize their video compression and transmission policies without being aware of the jamming model or the video service model. Moreover, we have developed a safe RL-based scheme to reduce the UAV transmission outage probability, further accelerate the UAV policy learning process. Simulation results verify the derived optimal value of the UAV utility and show that the proposed schemes significantly improve the video transmission QoE with lower transmission latency and energy consumption in the presence of smart jamming. In future research, the scenario of multiple cooperative jammers can be taken into consideration and evaluated. The delay and error caused by many various components in the video transmission system can be taken into consideration to further improve the applicability of the proposed scheme, especially in system implementation in practice.

Appendix A
Proof of Theorem 1

We assume that x_1 is continuous from 0 to N and x_3 is continuous from 1 to M for simplicity of presentation, so that u is differentiable for $x_i (1 \le i \le 4)$. According to (3), (8) and (9), the utility of the UAV can be rewritten as

$$\begin{aligned} u = {} & \omega_1 x_1 + \omega_2 \\ & - \nu \mathrm{I}\left(A\mathrm{erfc}\left(\sqrt{\frac{B x_4 g^2}{\sigma^2 + y h^2}} \right) \sqrt{2 x_2 b^{x_3 - 1}} > T \right) \\ & - \alpha_0 l \left(\varphi {x_1}^2 + \eta x_1 + \frac{N_{\mathrm{B}}}{R_{\mathrm{B}} b^{x_3 - 1}} \right) \\ & - \alpha_1 (a_1 x_1 + a_0 + b_1 x_2 + b_0 + c x_4). \end{aligned} \tag{15}$$

If (11d) holds, we have

$$\begin{aligned} & A\mathrm{erfc}\left(\sqrt{\frac{B x_4 g^2}{\sigma^2 + y h^2}} \right) \sqrt{2 x_2 b^{x_3 - 1}} \\ & \quad \le A\mathrm{erfc}\left(\sqrt{\frac{B X_1 g^2}{\sigma^2 + y_{\max} h^2}} \right) \sqrt{2 R_2 b^{M-1}} < T. \end{aligned} \tag{16}$$

The utility now is given by

$$\begin{aligned} u = {} & \omega_1 x_1 + \omega_2 - \alpha_0 l \left(\varphi {x_1}^2 + \eta x_1 + \frac{N_{\mathrm{B}}}{R_{\mathrm{B}} b^{x_3 - 1}} \right) \\ & - \alpha_1 (a_1 x_1 + a_0 + b_1 x_2 + b_0 + c x_4). \end{aligned} \tag{17}$$

Then if (11a) holds, we have

$$\begin{aligned} \frac{\partial u}{\partial x_1} &= -2\alpha_0 l \varphi x_1 + \omega_1 - \alpha_0 l \eta - \alpha_1 a_1 \\ &\le \omega_1 - \alpha_0 L \eta - \alpha_1 a_1 < 0, \end{aligned} \tag{18}$$

$$\frac{\partial u}{\partial x_2} = -\alpha_1 b_1 > 0, \tag{19}$$

$$\frac{\partial u}{\partial x_3} = \frac{\alpha_0 N_{\mathrm{B}} \ln b}{R_{\mathrm{B}} b^{x_3 - 1}} > 0, \tag{20}$$

$$\frac{\partial u}{\partial x_4} = -\alpha_1 c < 0, \tag{21}$$

indicating that the utility u increases with x_2 and x_3, and decreases with x_1 and x_4.

Thus, the gradient

$$\nabla u = \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \frac{\partial u}{\partial x_3}, \frac{\partial u}{\partial x_4} \right) \neq \mathbf{0}, \tag{22}$$

indicating that the utility u has no extreme point in the feasible domain of $\mathbf{x}$. According to the gradient ascent algorithm, $\mathbf{x}$ is updated along the direction of the gradient ascent, which is given by

$$\mathbf{x} = \mathbf{x} + \mathbf{q} \nabla u, \tag{23}$$

where $\mathbf{q} = [q_i]_{1 \le i \le 4}$ is the iteration step which can be varying with different x_i or identical.

Keep using the gradient ascent algorithm in (23) and finally since $x_1 \in [0, N]$, $x_2 \in [R_1, R_2]$, $x_3 \in [1, M]$ and $x_4 \in [X_1, X_2]$, we have the optimal solution as $\mathbf{x}^* = \arg\max_{\mathbf{x} \in \mathbf{X}} u = [0, R_2, M, X_1]$. This result still holds when x_1 and x_3 are discrete. By substituting $\mathbf{x} = \mathbf{x}^*$ and $l = 1$ into the utility in (17), the optimal value regarding the utility (10a) can be obtained. Eqs. (10b) and (10c) can also be proved in a similar way.

APPENDIX B
PROOF OF THEOREM 2

The total steps required for **Algorithm 1** to converge is KN_{F}. According to [61] and [62], if the lower order terms can be ignored, i.e. $KN_{\mathrm{F}} > \mathrm{poly}\{|\mathbf{S}|, |\mathbf{X}|, K\}$, where $\mathrm{poly}\{\cdot\}$ is an operator that calculates the characteristic polynomial of an equation with vector $[|\mathbf{S}|, |\mathbf{X}|, K]$ as the solution, and $|\mathbf{S}|$ and $|\mathbf{X}|$ is the size of state and action sets, respectively, then the computational complexity $\mathcal{O}(\Gamma_1)$ can be given by $\mathcal{O}(\Gamma_1) = \mathcal{O}(KN_{\mathrm{F}})$.

As far as **Algorithm 2** is concerned, the computational complexity of the FC layers can be ignored compared with the Conv layers because the operations are linear and thus the computational complexity is sufficiently small compared to the quadratic operations in convolutional layers. In most cases, the conditions $z_0 > z_1 > z_2$ and $z_1 + z_2 - 2 \approx 0$ hold. According to [63], we have

$$z_1^2(z_0 - z_1 + 1)^2 \ll f_2 z_2^2 (z_0 - z_1 - z_2 + 2)^2. \tag{24}$$

According to (24) and [64], we have

$$\begin{aligned}
\mathcal{O}(\Gamma_2) &= \mathcal{O}\left(\sum_{\psi=1}^{2} f_{\psi-1} z_\psi^2 f_\psi m_\psi^2 + \sum_{i=1}^{2} n_i \big(m_2 + |\mathbf{X}|\big)\right) \\
&= \mathcal{O}\Bigg(f_0 z_1^2 f_1 \left(\frac{z_0 - z_1}{s_1} + 1\right)^2 \\
&\quad + f_1 z_2^2 f_2 \left(\frac{z_0 - z_1}{s_1 s_2} - \frac{z_2 - 1}{s_2} + 1\right)^2 \\
&\quad + (n_1 + n_2)\left(\frac{z_0 - z_1}{s_1 s_2} - \frac{z_2 - 1}{s_2} + 1 + |\mathbf{X}|\right)\Bigg) \\
&= \mathcal{O}\Bigg(f_1 z_1^2 (z_0 - z_1 + 1)^2 + f_1 f_2 z_2^2 (z_0 - z_1 - z_2 + 2)^2 \\
&\quad + (n_1 + n_2)\big(z_0 - z_1 - z_2 + 2 + |\mathbf{X}|\big)\Bigg) \\
&= \mathcal{O}\left(f_1 f_2 z_2^2 (z_0 - z_1 - z_2 + 2)^2\right) \\
&= \mathcal{O}\left(f_1 f_2 z_0^2 z_2^2\right).
\end{aligned}$$

REFERENCES

[1] Y. Mi, C. Luo, G. Min, W. Miao, L. Wu, and T. Zhao, "Sensor-assisted global motion estimation for efficient UAV video coding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 2237–2241.

[2] J. Zhang *et al.*, "REMT: A real-time end-to-end media data transmission mechanism in UAV-aided networks," *IEEE Netw.*, vol. 32, no. 5, pp. 118–123, Sep. 2018.

[3] A. Fotouhi *et al.*, "Survey on UAV cellular communications: Practical aspects, standardization advancements, regulation, and security challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3417–3442, 4th Quart., 2019.

[4] B. Li, Z. Fei, Y. Zhang, and M. Guizani, "Secure UAV communication networks over 5G," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 114–120, Oct. 2019.

[5] H. Wang, J. Wang, G. Ding, Z. Xue, L. Zhang, and Y. Xu, "Robust spectrum sharing in air-ground integrated networks: Opportunities and challenges," *IEEE Wireless Commun.*, vol. 27, no. 3, pp. 148–155, Jun. 2020.

[6] L. Xiao, C. Xie, M. Min, and W. Zhuang, "User-centric view of unmanned aerial vehicle transmission against smart attacks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3420–3430, Apr. 2018.

[7] D. Marpe, T. Wiegand, and G. J. Sullivan, "The H.264/MPEG4 advanced video coding standard and its applications," *IEEE Commun. Mag.*, vol. 44, no. 8, pp. 134–143, Aug. 2006.

[8] H. Lu, Y. Gui, X. Jiang, F. Wu, and C. W. Chen, "Compressed robust transmission for remote sensing services in space information networks," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 46–54, Apr. 2019.

[9] M. Li, "Queueing analysis of unicast IPTV with adaptive modulation and coding in wireless cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9241–9253, Oct. 2017.

[10] Q. Wu, W. Mei, and R. Zhang, "Safeguarding wireless network with UAVs: A physical layer security perspective," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 12–18, Oct. 2019.

[11] H.-M. Wang, X. Zhang, and J.-C. Jiang, "UAV-involved wireless physical-layer secure communications: Overview and research directions," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 32–39, Oct. 2019.

[12] X. Sun, D. W. K. Ng, Z. Ding, Y. Xu, and Z. Zhong, "Physical layer security in UAV systems: Challenges and opportunities," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 40–47, Oct. 2019.

[13] S. Lv, L. Xiao, Q. Hu, X. Wang, C. Hu, and L. Sun, "Anti-jamming power control game in unmanned aerial vehicle networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.

[14] N. Mowla, N. H. Tran, I. Doh, and K. Chae, "Federated learning-based cognitive detection of jamming attack in flying ad-hoc network," *IEEE Access*, vol. 8, pp. 4338–4350, 2020.

[15] U. Challita, A. Ferdowsi, M. Chen, and W. Saad, "Machine learning for wireless connectivity and security of cellular-connected UAVs," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 28–35, Feb. 2019.

[16] N. Gao, Z. Qin, X. Jing, Q. Ni, and S. Jin, "Anti-intelligent UAV jamming strategy via deep Q-networks," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 569–581, Jan. 2020.

[17] M. Wang, K. N. Ngan, H. Li, and H. Zeng, "Improved block level adaptive quantization for high efficiency video coding," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Lisbon, Portugal, May 2015, pp. 509–512.

[18] A. S. Dias, S. Huang, S. G. Blasi, M. Mrak, and E. Izquierdo, "Time-constrained video delivery using adaptive coding parameters," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 7, pp. 2082–2095, Jul. 2019.

[19] Z. Wu, H. Yu, B. Tang, and C. W. Chen, "Adaptive initial quantization parameter determination for H.264/AVC video transcoding," *IEEE Trans. Broadcast.*, vol. 58, no. 2, pp. 277–284, Jun. 2012.

[20] A. E. A. A. Abdulla, Z. M. Fadlullah, H. Nishiyama, N. Kato, F. Ono, and R. Miura, "Toward fair maximization of energy efficiency in multiple UAS-aided networks: A game-theoretic methodology," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 305–316, Jan. 2015.

[21] Y. Kawamoto, H. Nishiyama, N. Kato, F. Ono, and R. Miura, "An efficient throughput-aware resource allocation technique for data transmission in unmanned aircraft systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[22] S. Chen, J. Yang, Y. Ran, and E. Yang, "Adaptive layer switching algorithm based on buffer underflow probability for scalable video streaming over wireless networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 6, pp. 1146–1160, Jun. 2016.

[23] H. Hu, C. Zhan, J. An, and Y. Wen, "Optimization for HTTP adaptive video streaming in UAV-enabled relaying system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–6.

[24] Q. Wang, Z. Chen, W. Mei, and J. Fang, "Improving physical layer security using UAV-enabled mobile relaying," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 310–313, Jun. 2017.

[25] N. Zhao *et al.*, "Caching UAV assisted secure transmission in hyper-dense networks based on interference alignment," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2281–2294, May 2018.

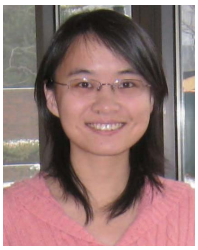
[26] M. Soysa, P. C. Cosman, and L. B. Milstein, "Disruptive attacks on video tactical cognitive radio downlinks," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1411–1422, Apr. 2016.

[27] L. Jia, Y. Xu, Y. Sun, S. Feng, and A. Anpalagan, "Stackelberg game approaches for anti-jamming defence in wireless networks," *IEEE Wireless Commun.*, vol. 25, no. 6, pp. 120–128, Dec. 2018.

[28] J. Xu, Y. Zeng, and R. Zhang, "UAV-enabled wireless power transfer: Trajectory design and energy optimization," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5092–5106, Aug. 2018.

[29] S. Zhang, Y. Zeng, and R. Zhang, "Cellular-enabled UAV communication: A connectivity-constrained trajectory optimization perspective," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2580–2604, Mar. 2019.

[30] F. Wu, H. Zhang, J. Wu, and L. Song, "Cellular UAV-to-device communications: Trajectory design and mode selection by multi-agent deep reinforcement learning," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4175–4189, Jul. 2020.
[31] Y. Ding *et al.*, "QoE-aware power control for UAV-aided media transmission with reinforcement learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.
[32] L. Xiao, X. Lu, D. Xu, Y. Tang, L. Wang, and W. Zhuang, "UAV relay in VANETs against smart jamming with reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4087–4097, May 2018.
[33] J. Cui, Y. Liu, and A. Nallanathan, "Multi-agent reinforcement learning-based resource allocation for UAV networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 729–743, Feb. 2020.
[34] U. Challita, W. Saad, and C. Bettstetter, "Interference management for cellular-connected UAVs: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2125–2140, Apr. 2019.
[35] Q. Wang, W. Zhang, Y. Liu, and Y. Liu, "Multi-UAV dynamic wireless networking with deep reinforcement learning," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2243–2246, Dec. 2019.
[36] W. Zhang, Q. Wang, X. Liu, Y. Liu, and Y. Chen, "Three-dimension trajectory design for multi-UAV wireless network with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 20, no. 1, pp. 600–612, Jan. 2020.
[37] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Trajectory design and power control for multi-UAV assisted wireless networks: A machine learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7957–7969, Aug. 2019.
[38] Y. Zeng, X. Xu, S. Jin, and R. Zhang, "Simultaneous navigation and radio mapping for cellular-connected UAV with deep reinforcement learning," *IEEE Trans. Wireless Commun.*, early access, pp. 1–16, Feb. 2021, doi: 10.1109/TWC.2021.3056573.
[39] Y. Zhang, Z. Mou, F. Gao, J. Jiang, R. Ding, and Z. Han, "UAV-enabled secure communications by multi-agent deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 11599–11611, Oct. 2020.
[40] K. Hossain, C. Mantel, and S. Forchhammer, "No reference prediction of quality metrics for H.264 compressed infrared image sequences for UAV applications," *Electron. Imag.*, vol. XV, pp. 108-1–108-6, Jan. 2018.
[41] J. Zhao, Q. Dong, Y. Zhao, B. Wang, and F. Gao, "Time varying channel tracking for multi-UAV wideband communications with beam squint," 2019, *arXiv:1911.09433*. [Online]. Available: http://arxiv.org/abs/1911.09433
[42] X. Cheng and Y. Li, "A 3-D geometry-based stochastic model for UAV-MIMO wideband nonstationary channels," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1654–1662, Apr. 2019.
[43] D. Darsena, G. Gelli, I. Iudice, and F. Verde, "Equalization techniques of control and non-payload communication links for unmanned aerial vehicles," *IEEE Access*, vol. 6, pp. 4485–4496, 2018.
[44] C. Sommer and F. Dressler, "Using the right two-ray model? A measurement-based evaluation of PHY models in VANETs," in *Proc. ACM Int. Mobile Ad Hoc Netw. Comput.*, Nevada, LV, USA, Sep. 2011, pp. 1–3.
[45] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.
[46] J. Hou, Z. Yang, and M. Shikh-Bahaei, "Hardware impairment-aware data collection and wireless power transfer using a MIMO full-duplex UAV," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.
[47] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.
[48] Z. Guan, T. Melodia, and D. Yuan, "Optimizing cooperative video streaming in wireless networks," in *Proc. 8th Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Netw.*, Salt Lake City, UT, USA, Jun. 2011, pp. 503–511.
[49] X. Zhu, E. Setton, and B. Girod, "Congestion–distortion optimized video transmission over ad hoc networks," *Signal Process., Image Commun.*, vol. 20, no. 8, pp. 773–783, Sep. 2005.
[50] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3747–3760, Jun. 2017.
[51] J. Yao and N. Ansari, "QoS-aware power control in Internet of drones for data collection service," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6649–6656, Jul. 2019.
[52] C. Kyrkou, G. Plastiras, T. Theocharides, S. I. Venieris, and C.-S. Bouganis, "DroNet: Efficient convolutional neural network detector for real-time UAV applications," in *Proc. Design, Autom. Test Eur. Conf. Exhib. (DATE)*, Mar. 2018, pp. 967–972.
[53] S. Tripathi, B. Kang, G. Dane, and T. Nguyen, "Low-complexity object detection with deep convolutional neural network for embedded systems," *Proc. SPIE*, vol. 10396, Sep. 2017, Art. no. 103961M.
[54] E. Parisotto, L. Ba, and R. Salakhutdinov, "Actor-mimic: Deep multitask and transfer reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Juan, Puerto Rico, May 2017, pp. 1–16.
[55] J. Tian, H. Zhang, D. Wu, and D. Yuan, "Interference-aware cross-layer design for distributed video transmission in wireless networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 5, pp. 978–991, May 2016.
[56] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.
[57] C. Mendes, V. Frémont, and D. F. Wolf, "Exploiting fully convolutional neural networks for fast road detection," in *Proc. IEEE Int. Conf. Robot. Automat.*, Stockholm, Sweden, May 2016, pp. 3174–3179.
[58] T. Richardson and S. Kudekar, "Design of low-density parity check codes for 5G new radio," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 28–34, Mar. 2018.
[59] Z. Hong, Q. Yan, Z. Li, T. Zhan, and Y. Wang, "Photon-counting underwater optical wireless communication for reliable video transmission using joint source-channel coding based on distributed compressive sensing," *Sensors*, vol. 19, no. 5, p. 1042, Mar. 2019.
[60] X. Tang, X. Huang, and F. Hu, "QoE-driven UAV-enabled pseudo-analog wireless video broadcast: A joint optimization of power and trajectory," *IEEE Trans. Multimedia*, early access, pp. 1–15, Jul. 2020, doi: 10.1109/TMM.2020.3011319.
[61] C. Jin, Z. Allen-Zhu, S. Bubeck, and M. I. Jordan, "Is Q-learning provably efficient?" 2018, *arXiv:1807.03765*. [Online]. Available: http://arxiv.org/abs/1807.03765
[62] M. Kearns and S. Singh, "Near-optimal reinforcement learning in polynomial time," *Mach. Learn.*, vol. 49, nos. 2–3, pp. 209–232, 2002.
[63] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Jan. 2015.
[64] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 5353–5360.



Liang Xiao (Senior Member, IEEE) received the B.S. degree in communication engineering from the Nanjing University of Posts and Telecommunications, China, in 2000, the M.S. degree in electrical engineering from Tsinghua University, China, in 2003, and the Ph.D. degree in electrical engineering from Rutgers University, NJ, USA, in 2009. She is currently a Professor with the Department of Communication Engineering, Xiamen University, Fujian, China. She was a recipient of the Best Paper Award for 2016 INFOCOM Big Security WS and 2017 ICC. She has served as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and IEEE TRANSACTIONS ON COMMUNICATIONS.

Yuzhen Ding (Student Member, IEEE) received the B.S. degree in communication engineering from Wuhan University, Wuhan, China, in 2018. She is currently pursuing the M.S. degree with the Department of Information and Communication Engineering, Xiamen University, Xiamen, China. Her research interests include network security and wireless communications.

Jinhao Huang received the B.S. degree in electronic information engineering from Shantou University, Shantou, China, in 2018. He is currently pursuing the M.S. degree with the Department of Information and Communication Engineering, Xiamen University, Xiamen, China. His research interests include network security and wireless communications.

Sicong Liu (Member, IEEE) received the B.S.E. and Ph.D. degrees (Hons.) in electronic engineering from Tsinghua University, Beijing, China, in 2012 and 2017, respectively. He was a Visiting Scholar with the City University of Hong Kong in 2010. He served as a Senior Research Engineer at Huawei Technologies before joining Xiamen University in 2018, where he is currently an Assistant Professor. His current research interests lie in compressed sensing, vehicular networks, smart grid communications, and visible light communications. He has served as an editor, the TPC chair, and the publication chair for several IEEE and other academic journals and conferences.

Yuliang Tang (Member, IEEE) received the M.S. degree from the Beijing University of Posts and Telecommunications, China, in 1996, and the Ph.D. degree in information and communication engineering from Xiamen University in 2009. He is currently a Professor with the Department of Information and Communication Engineering, Xiamen University. He has published more than 90 articles in journals and international conferences. He has been granted over 20 patents in his research areas. His research interests include wireless communication, 5G and beyond, and vehicular *ad-hoc* networks.

Huaiyu Dai (Fellow, IEEE) received the B.E. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2002.

He was with Bell Labs, Lucent Technologies, Holmdel, NJ, in Summer 2000, and with AT&T Labs Research, Middletown, NJ, in Summer 2001. He is currently a Professor of electrical and computer engineering with NC State University, Raleigh, holding the title of the University Faculty Scholar. His research interests are in the general areas of communications, signal processing, networking, and computing. His current research focuses on machine learning and artificial intelligence for communications and networking, multilayer and interdependent networks, dynamic spectrum access and sharing, and security and privacy issues in the above systems.

Dr. Dai is a member of the Executive Editorial Committee for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was a co-recipient of best paper awards at 2010 IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS 2010), 2016 IEEE INFOCOM BIGSECURITY Workshop, and 2017 IEEE International Conference on Communications (ICC 2017). He has served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is an Area Editor in charge of wireless communications for IEEE TRANSACTIONS ON COMMUNICATIONS.