*Journal of Network Intelligence* ©2022 ISSN 2414-8105 (Online)
**Taiwan Ubiquitous Information** Volume **7**, Number **1**, February **2022**

# A Suvey on Edge Intelligent Video Surveillance with Deep Reinforcement Learning

Yan Li

Nanchang Institute of Technology
No. 289, Tianxiang Road, Nanchang, Jiangxi, 330099, China
Jiangxi University of Finance and Economics
No. 665, West Yuping Road, Nanchang, Jiangxi, 330032, China
yli@nit.edu.cn

Received October 2021; revised December 2021

ABSTRACT. *Intelligent video surveillance technology is one of the key technologies in many application fields such as public safety surveillance, intelligent transportation, and live broadcast of large-scale events. With the rapid development of artificial intelligence, many intelligent video surveillance model architectures have been produced, especially intelligent video surveillance technology based on deep learning. As one of the key technologies of 5G communication technology, edge computing has also been widely used in intelligent video surveillance system to solve low-latency, high-precision and real-time computing tasks. In this paper, the mobile edge computing based (MEC-based) deep learning intelligent video surveillance architecture is introduced and reviewed. This paper discusses the MEC-based intelligent video surveillance architecture in detail, and proposes a new federal deep reinforcement learning based cluster MEC intelligent video surveillance system. Furthermore, this paper discusses application scenarios of various architectures, and summarizes the main research directions in recent years. The content summarized in this paper can provide a more comprehensive perspective for related researchers to understand the main research content of this direction, and inspire new research methods.*

**Keywords:** edge intelligence, clustered mobile edge computing, deep reinforcement learning, intelligent video surveillance

1. **Introduction.** With the continuous development of the intelligent industry, video surveillance is widely used in many occasions due to its intuitive, accurate, timely and rich information content. It is especially important in security systems and becomes the most powerful means of technical security. Nowadays, data is showing explosive growth. Security monitoring has higher and higher requirements for high-definition, intelligence, networking, and digitization. The security field has gradually changed from a single system deployment in the past to a comprehensive intelligent system for big data application analysis. It has to be admitted that the era of intelligent has arrived.

As a typical data-dependent business, video surveillance services rely on data processing to speak naturally. Through big data transmission and processing technology, further mining the value information behind the massive video surveillance data, and quickly feeding back the connotative knowledge to assist decision-making and judgment, is the golden key to good use of video surveillance. At present, with the rapid development of smart transportation, smart gardens, public safety, and visual management in smart cities, the application of video surveillance in smart cities has already begun.

Smart city is a brand-new urban form, and it is also the main application scenario that scholars have researched and explored in recent years [1, 2]. In the construction of smart cities, the construction of smart gardens permeates all aspects of smart environment, smart residence and smart management [3, 4]. And intelligent video surveillance is an indispensable core technology in the construction of smart gardens. In intelligent video surveillance service, it generally requires low latency and high recognition accuracy. In the overall mobile data traffic as summarized in Cisco Visual Networking

Index [5], mobile video contents account for more than half of this data traffic, and will be predicted to further grow by 2022, accounting for 82% of the total data traffic. At present, the development of garden construction mainly has the following problems: (i) the lack of garden design talents and maintenance personnel, (ii) the garden market is relatively small and the demand for ornamental plants is not large, (iii) The combination of gardens and modern technology is insufficient, (iv) the garden safety monitoring and disaster prevention system is not sound. In particular, the issues of the insufficient combination of gardens and modern technology and the garden safety monitoring and disaster prevention systems have severely restricted the efficient development of urban garden construction. Therefore, an efficient intelligent video surveillance system in the construction of smart gardens is a important problem that needs to be solved urgently.

The intelligent video surveillance system adopts image processing, pattern recognition and computer vision technology, by adding an intelligent video analysis module to the surveillance system. Then, the system uses the computer's powerful data processing capabilities to filter out the useless or disturbing information of the video screen, automatically identify different objects, analyze and extract the key useful information from the video source. Then it quickly and accurately locate the accident site, and judge the abnormal situation in the monitoring screen. After that, the fastest and best way to issue an alarm or trigger other actions, so as to effectively carry out pre-warning, processing during the event, and after the event, a fully automatic, all-weather, real-time monitoring intelligent system.

In intelligent video surveillance system, accurate decision-making based on multi-dimensional heterogeneous data information and real-time data processing are key issues to be solved urgently. In 2014, mobile edge computing (MEC) [6,7] was proposed by the European Telecommunications Standards Institute (ETSI) as an emerging paradigm to solve real-time computing tasks. It is a technology based on the 5G evolution architecture and deeply integrates mobile access networks and internet services. With the rapid development of machine learning algorithms, especially deep learning algorithms, more and more target recognition algorithms are implemented based on deep learning (DL) [8, 9]. For the identification and scheduling problems in the network environment, due to the variable factors in the complex and changeable environment, deep reinforcement learning (DRL) is often used to solve complex joint optimization problems [10–12]. Therefore, both MEC and DRL have been gradually applied to intelligent video surveillance technology in recent years to promote the low latency and accuracy of surveillance system.

In this paper, we mainly did the following innovative works:

(i)The paper introduced and discussed the MEC-based intelligent video surveillance architecture in detail, and analyzes the shortcomings of its architecture.

(ii)This paper proposed an intelligent surveillance system based on Clustered Mobile Edge Computing and Blockchain empowerment (CMEC-BC-IVS), which not only promotes surveillance video data security and privacy protection, but also ensures collaborative resource allocation and optimization work among edge cloud, edge nodes and terminals.

(iii)This paper designed a new federated deep reinforcement learning framework (FDRL-method) in the CMEC-BC-IVS network. The DRL algorithm is used to realize joint decision-making of surveillance video caching, transcoding, real-time processing, and response operations.

In this paper, we will briefly review the development process, core issues, latest challenges and opportunities of intelligent video surveillance technology. The remainder of this paper is organized as follows. The related work of MEC-based intelligent video surveillance with deep reinforcement learning will present in Section 2. Section 3 gives the CMEC-based and blockchain-empowered intelligent video surveillance system and Section 4 describes the trends and challenges. Finally, the conclusions presents in Section 5. The structure of this paper can be shown in Figure 1.

## 2. **MEC-based Intelligent Video Surveillance with Deep Reinforcement Learning.**

2.1. **Development of Intelligent Video Surveillance.** The intelligent video surveillance system has been developed in just over two decades. From the earliest analog surveillance to the fiery digital

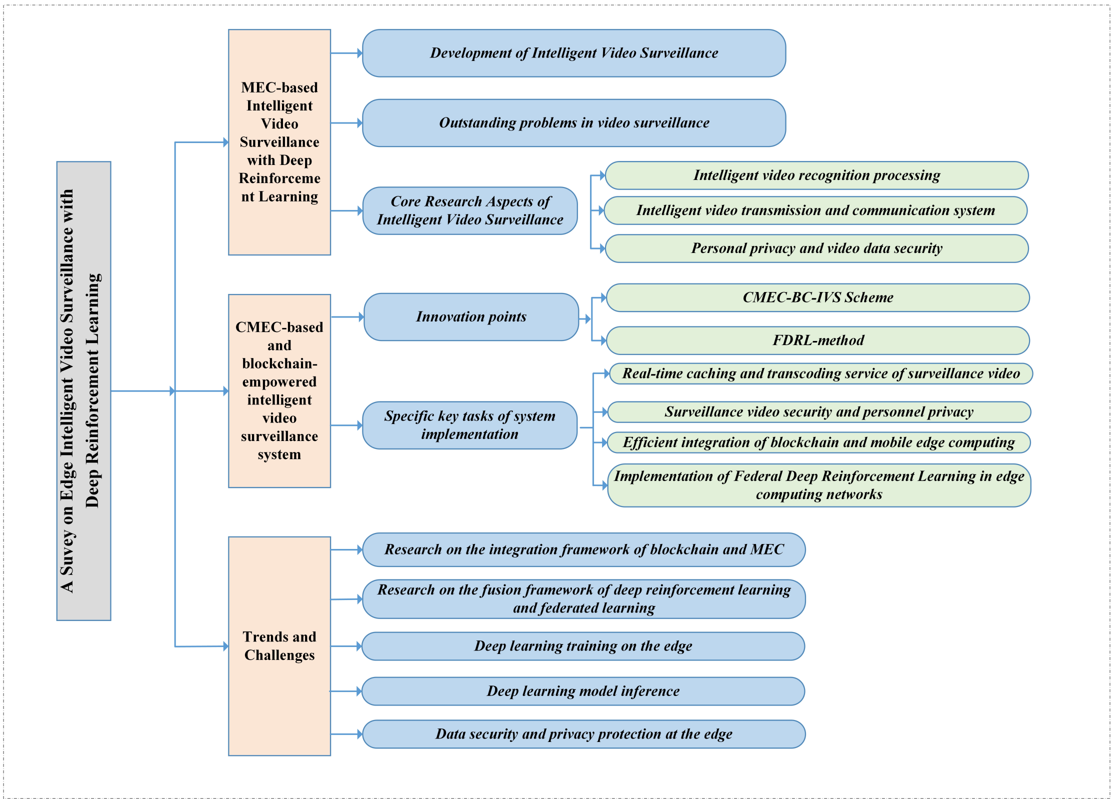


FIGURE 1. The structure of this paper

surveillance in previous years and then to the now in the ascendant network video surveillance, earth-shaking changes have taken place.

From a technical point of view, the development of video surveillance systems can be divided into three generations. The first generation is an analog video surveillance system, the second generation is a digital video surveillance system based on PC+Multimedia Card, and the third generation is a video surveillance system based entirely on IP networks [13–17].

The first generation of video surveillance was the traditional Closed-Circuit Television (CCTV). The system consists of two parts: front-end equipment and monitoring center. The front-end equipment includes cameras, dome cameras, pan-tilts, decoders, etc. Monitoring center equipment includes monitors, video splitters, switching matrix, control keyboards, video recorders, and so on. The two parts of equipment are connected by video cables, control cables, etc. Analog video surveillance systems are usually suitable for small-scale regional surveillance. As we all know, traditional audio and video signals are analog signals, which are usually transmitted by means of coaxial cables. Analog monitoring is only suitable for single buildings, small residential areas and other small areas. Limited by analog video cable transmission length and cable amplifier, it can only support local monitoring.

The analog-digital monitoring system is a semi-analog-semi-digital solution based on Digital Video Recorder (DVR). From the camera to the DVR, the coaxial cable is still used to output the video signal. The DVR supports both recording and playback, and can support limited IP network access. Since DVR products are diverse and there is no standard, this generation system is a non-standard closed system, and the DVR system still has a lot of limitations. For example, limited remote monitoring and control capabilities. In this system, any camera cannot be accessed from any client, but the camera can only be accessed indirectly through the DVR. At the same time, the analog-digital program video has no protection and is easy to lose.

Compared with the previous two schemes, IP Video Surveillance (IPVS) is significantly different. The advantage of this system is that the camera has a built-in web server and directly provides an Ethernet port. These cameras generate JPEG or MPEG4 data files, which can be accessed, monitored, recorded and printed by any authorized client from any location on the network, rather than generating images in the form of continuous analog video signals. The great advantages of the IP video surveillance system include simplicity, powerful central control, easy upgrade and comprehensive scalability, comprehensive remote monitoring, and rugged redundant memory. Currently, more and more monitoring systems are based on edge computing architecture. The surveillance video system architecture based on edge computing is shown in Figure 2.

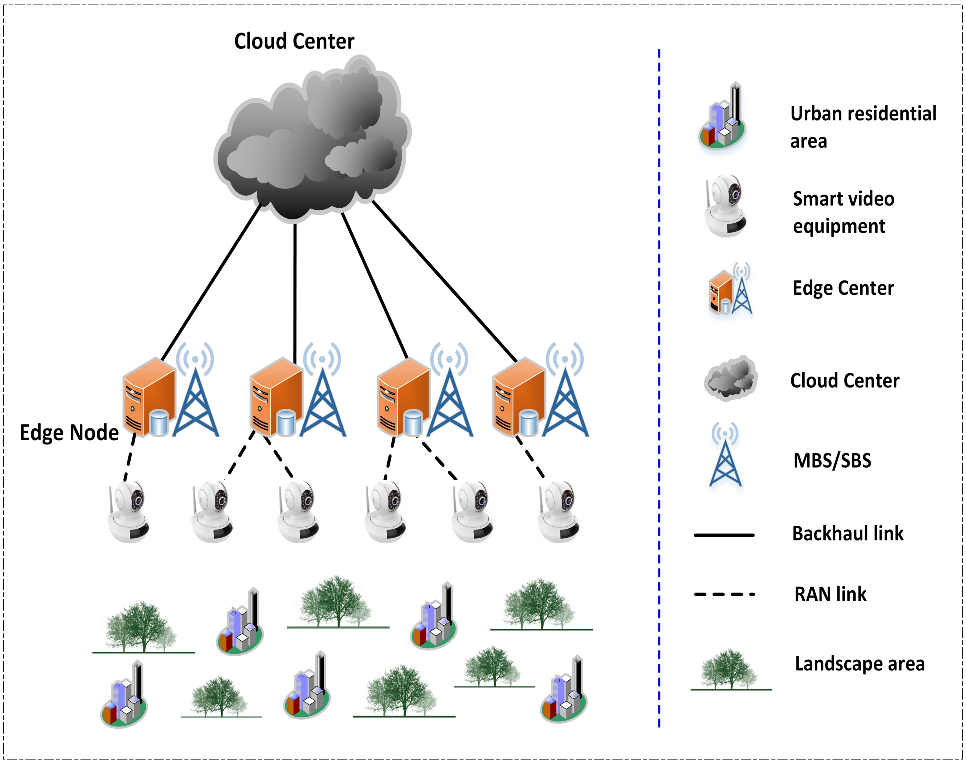


FIGURE 2. Surveillance video system architecture based on edge computing

## 2.2. **Outstanding problems in video surveillance.**

2.2.1. *Low latency and high accuracy of surveillance video task processing.* Generally speaking, in many application scenarios, video surveillance requires low latency and high recognition accuracy for public safety monitoring, real-time warning of emergencies and disasters. In intelligent video surveillance, the delay is mainly affected by the task processing time, which the task processing can be implemented at the end, edge or cloud. Furthermore, with the proposed of MEC, edge computing can be divided into near-edge, far-edge, and end-edge according to the understanding of different scenarios. Different types of edge nodes will have different intelligent processing capabilities, including computing capabilities, storage capabilities, and collaboration capabilities. Based on this issue, MEC presents an unique opportunity to implement edge processing, and is widely used in many low-latency application scenarios [18–21]. Furthermore, the guarantee of high accuracy depends entirely on the efficiency of the recognition algorithm in video surveillance. With the rapid development of machine learning algorithms, especially deep learning algorithms, more and more target recognition algorithms are implemented based on DL. Due to the variable factors in the complex and changeable environment, DRL is often used to solve complex joint optimization problems. Low latency and high accuracy are a pair of contradictions. How to find the best balance between the two is a difficult point that needs to be explored and studied.

2.2.2. *Personal privacy and video data security of surveillance video.* Privacy protection and data security issues in video surveillance have always been one of the main issues that need to be resolved [22–24]. In particular, privacy protection and data security in the edge computing network environment require in-depth research to find an effective solution suitable for the designed network architecture. In recent years, a new application model called blockchain has developed rapidly. The core content of blockchain includes encryption algorithm, distributed storage, P2P transmission and consensus mechanism [25, 26]. As a effective distributed management framework, blockchain has been widely used in many fields to realize the security monitoring and privacy protection of data resources for intelligent data processing. However, in different application scenarios, it is difficult to achieve a unified efficient integration of blockchain and edge computing, which makes data security and privacy protection in the edge computing environment much more challenging.

2.2.3. *Other problems in intelligent video surveillance.* In recent years, there has been a lot of exploration and research on video surveillance systems [27–31], but these video surveillance systems still have some shortcomings, mainly as follows: (i) Failure to detect abnormalities and intelligent early warning in time. Most current video surveillance systems cannot perform real-time and efficient intelligent analysis and processing of surveillance video images, and also cannot find problems in time. (ii) For emergencies that occur, although early warning can be given, real-time automated operations cannot be implemented to solve the problem, so as to prevent the problem from getting worse. (iii) The data security and privacy protection of video surveillance systems are not guaranteed. The detailed research progress of intelligent video surveillance will be elaborated in the section of 2.3.

2.3. **Core Research Aspects of Intelligent Video Surveillance.** The detailed research progress of intelligent video surveillance was elaborated in Table 1. The research of intelligent video surveillance mainly focuses on three aspects. Firstly, it is intelligent video recognition processing, which refers to the interpretation of video scenes in intelligent video surveillance. In video surveillance, it detects, recognizes, and tracks specific objects from the sequence of frame images, then understands and interprets the behavior of the objects. This part of the research requires image processing or computer vision technology based on video surveillance. Secondly, it is the optimization design of intelligent video transmission and communication system, which refers to how to carry out effective video collection in intelligent video surveillance. And the optimization of wireless video transmission and communication system is designed to achieve low-cost, low-latency real-time video intelligence processing and decision-making feedback. This part of the research needs to be based on wireless video transmission theory, cloud computing and edge computing. Thirdly, it is user privacy protection and video data security in the intelligent video surveillance system to ensure the security of the surveillance network system and the security of sensitive data.

2.3.1. *Intelligent video recognition processing.* Target detection and recognition algorithms in video surveillance include moving target detection algorithms, traditional template-based recognition methods, and statistical learning recognition methods. Among them, moving target detection algorithms can be divided into three types: inter-frame difference method, background difference method, and optical flow method. With the rapid development of machine learning algorithms, especially deep learning algorithms, more and more target recognition algorithms have been implemented in combination with deep learning.

The latest trends in object recognition research focus on designing lightweight deep learning models to reduce computational costs and designing different loss functions to further enhance the capabilities of neural networks [32–39]. A Deep Q Learning Network (DQN) was developed to locate anomalies in videos by enabling agents to learn how to detect and recognize anomalies in videos used in smart city [40]. A monitoring framework of Industrial Internet of Things based on artificial intelligence, called VD-Network, is proposed in [41]. Its recognition method is based on deep learning of ConvLSTM. In order to implement the faster real-time object detection, Ren et al. [42] introduced a Region Proposal Network(RPN) into the Fast R-CNN for the detection network, leading to deep-learning-based object detection system can run at 5-17 fps and also improving the overall object detection accuracy. Yeh et al. [43] proposed a lightweight deep-water target detection network, which is used to jointly learn the color conversion and target detection of underwater images to achieve the effectiveness of underwater object detection. In the literature [44], a deep separable convolution strategy is proposed to build a lightweight deep neural network, and applied it to the intelligent edge

TABLE 1. Research Progress of Representative Literature of Intelligent Video Surveillance

| Ref. | Research Direction | Focus | Method | Application scenario |
|---|---|---|---|---|
| [8] | Intelligent video recognition processing | A multi-task framework was proposed to jointly estimate 3D human pose from monocular color images and classify human actions from video sequences. | Multi-task CNN | Human action recognition |
| [10] | Intelligent video recognition processing | Proposed an efficient visual tracker by means of sequential actions learned using deep neural networks. | DRL | Visual object tracking |
| [11] | Intelligent video recognition processing | A novel deep reinforcement learning architecture for active learning and estimation of engagement from video data was proposed. | DRL | Engagement estimation |
| [12] | Intelligent video recognition processing | A unified space-time depth Q network (ST-DQN) is proposed to solve the problem of human activity positioning. | DRL(DQN) | Human activity localization |
| [14] | Intelligent video recognition processing | Developed an automatic image denoising system in video surveillance. | DL | Smart city |
| [19] | Intelligent video transmission and communication systemg | A tailor-made Edge Intelligent Video Surveillance (EIVS) system is proposed, which is a scalable edge computing architecture. | DL,MEC | Video surveillance system |
| [20] | Intelligent video recognition processing | Using YOLO algorithm and deep neural network to achieve multi-class target detection. | YOLO | Target detection |
| [28] | Intelligent video transmission and communication system | Focusing on multi-target detection for real-time monitoring in intelligent IoT systems, the deep neural network model A-YONet is proposed to be deployed in the end-side-cloud monitoring system. | YOLO, MEC, Cloud computing | Video surveillance system |
| [29] | Intelligent video transmission and communication system | Proposed a distributed intelligent video surveillance (DIVS) system using deep learning (DL) algorithms. | DL, MEC | Video surveillance system |
| [22] | Personal privacy and video data security | A system for privacy-preserving cloud video surveillance was proposed. | Cloud computing | Video surveillance system |
| [23] | Personal privacy and video data security | Proposed a cloud-based hierarchical video surveillance system considering the security of 5G networks. | MEC, Cloud computing | City-wide surveillance system |
| [37] | Intelligent video recognition processing | Developed a video surveillance algorithm based on deep reinforcement learning. | DRL(DQN) | Visual surveillance |
| [38] | Intelligent video recognition processing | A Deep Q Learning Network (DQN) was developed to locate anomalies in videos by enabling agents to learn how to detect and recognize anomalies in videos. | DRL(DQN) | Smart city |
| [39] | Intelligent video recognition processing | A method that can detect fires by analyzing the video obtained by surveillance cameras is proposed. | Transfer learning | Fire detection |
| [40] | Intelligent video transmission and communication system | A novel form of real-time human detection network was presented. | YOLOv2, MEC | Smart video surveillance |
| [41] | Intelligent video transmission and communication system | A monitoring framework of Industrial Internet of Things based on artificial intelligence, VD-Network, is proposed. | ConvLSTM, MEC | Public and industrial security |
| [44] | Intelligent video transmission and communication system | In the Intelligent Internet of Things (IIoT), an Intelligent Edge Surveillance (INES) method based on deep learning is proposed. | Lightweight deep learning, MEC | Intelligent Internet of Things |
| [58] | Personal privacy and video data security | Proposed federated learning empowered end-edge-cloud cooperation-based framework for enhancing 5G HetNet security. | Federated learning, MEC | 5G heterogeneous networks |
| [50] | Intelligent video transmission and communication system | A reinforcement learning based approach of video surveillance on mobile edge networks was proposed. | DRL, MEC | Video surveillance system |
| [51] | Intelligent video transmission and communication system | A video surveillance network based on mobile edge computing (MEC) is proposed, and the DRL algorithm is used to solve the joint optimization of computing and communication resources. | DQN+NN, MEC | Industrial Internet of Things |
| [52] | Intelligent video transmission and communication system | Given a reliable parking surveillance system with edge artificial intelligence on IoT devices. | YOLO-V3, MEC | Smart city |
| [54] | Intelligent video transmission and communication system | Proposed an edge-based solution of video surveillance in smart construction site assisted by Graph Neural Network. | Graph-assisted hierarchical RL (DQN), MEC | Smart construction |
| [55] | Intelligent video transmission and communication system | A collaborative Cloud-Edge architecture is proposed to analyze surveillance videos and extract video key frames, and use deep reinforcement learning algorithms to perform task scheduling to perform adaptive offloading in the cloud or edge. | DRL (A3C), MEC, Cloud computing | Video surveillance system |

surveillance (INES) method to reduce its computational cost and obtain the guarantee of detection accuracy. In face recognition, F. Schroff et al. designed FaceNet [45], triplet loss has been introduced into face recognition and effectively improves the recognition accuracy. In order to solve the problem that video streams captured from foggy environments are often unsatisfactory, Muhammad et al. [46] proposed a light-weight structure of deep convolutional neural networks for edge intelligent assisted smoke detection in foggy surveillance environments method. This method makes the early smoke detection in foggy day detection have better performance than the most advanced methods. Deep learning is well known as the innovator in the field of machine learning, especially in the field of computer vision. Similar to deep learning crushing traditional models in the field of graph classification, deep learning models are now the best method in the field of object recognition, pursuing fast real-time processing while ensuring high recognition accuracy. In the face of complex heterogeneous problems, deep reinforcement learning has been more widely used.

2.3.2. *Intelligent video transmission and communication system.* Video acquisition and intelligent transmission processing in intelligent video surveillance mainly involve sensing technology, wireless video transmission and intelligent processing technology of video information. Among them, wireless video transmission and real-time intelligent processing are even more important. The network environment in which wireless video transmission is located has experienced three stages: wireless access-wired core, wireless multi-hop network and heterogeneous wireless network. And currently we are in the age of heterogeneous wireless network. At present, the real-time processing of intelligent video in intelligent video surveillance is a difficult point. Furthermore, the main research trends focus on end computing processing, edge computing processing, cloud computing processing, and collaborative processing between them [47–49].

In order to achieve high recognition accuracy and low recognition time, Hu et al. [50] builds a video surveillance system based on mobile edge computing (MEC), designs image recognition algorithms for camera sensors and MEC servers, and uses reinforcement learning to solve joint optimization problems to obtain efficient and intelligent video surveillance. Kunpeng et al. [51] proposed a MEC-based video surveillance network for face recognition applications, and solved tasks such as task offloading, wireless channel allocation, and image compression rate selection through a two-layer analysis. The layer learning framework performs joint decision-making to obtain high average recognition accuracy and low average processing delay, and also the algorithm performs better in terms of convergence and training efficiency. Ke et al. [52] designed an intelligent parking monitoring task that uses edge computing for parking occupancy detection depending on real-time video sources. By implementing an enhanced single shot multibox (SSD) to achieve artificial intelligence at the edge, the system takes into account of flexibility, detection accuracy and system reliability. The final detection method achieves an accuracy of more than 95% in real scenes with high efficiency and high reliability. In order to solve the problems of high latency, high bandwidth, high energy consumption, the Internet of Video Things (IoVT) was proposed by Sultana and Wahid [53]. The video surveillance system under the IoVT framework provides multiple layers (including end, edge, Fog and cloud) communication and decision-making by capturing and analyzing rich context and behavior information. Then in [54], the authors proposed a graph neural network-assisted intelligent construction video surveillance system based on edge computing, and developed a graph-assisted hierarchical reinforcement learning algorithm to effectively obtain the characteristics of the mobile edge network. Thereby it can effectively maintain reliable accuracy and lower latency. In [55], a collaborative Cloud-Edge architecture is proposed to analyze surveillance videos and extract video key frames, and use deep reinforcement learning algorithms to perform task scheduling to perform adaptive offloading in the cloud or edge. And an automatic camera control method based on DQN in the video surveillance system is proposed in [56].

The ultimate goal of intelligent video transmission and communication system design is to achieve faster real-time processing while ensuring the accuracy of surveillance video recognition, and can automatically give intelligent decision feedback and response actions. So, many deep reinforcement learning methods (such as A3C, DQN) have been used in the intelligent decision-making process at the edge networks.

2.3.3. *Personal privacy and video data security.* In intelligent video surveillance system, since surveillance video data needs to be uploaded to a cloud center or edge cloud center, it is easy to cause video data security issues. At the same time, user privacy is easily leaked during network access [57].

Wei et al. [58] proposed federated learning empowered end-edge-cloud cooperation-based framework for enhancing 5G HetNet security. Federated learning is a machine learning framework that can effectively help multiple institutions to perform data usage and machine learning modeling under the requirements of user privacy protection, data security, and government regulations. Kim et al. [23] presented a cloud-based hierarchical video surveillance system considering the security of 5G networks. The authors eliminates the risk of privacy and personal information leakage through a cloud-based hierarchical video surveillance system. And a system for privacy-preserving cloud video surveillance was proposed in [22]. In this article, the PatronuS system is developed using conceptual diagrams to determine privacy areas that meet privacy and security requirements.

In intelligent video surveillance system, user privacy protection and data security can be achieved by setting security protection algorithms, hierarchical security authorization, and the security design of the surveillance system. And more and more video surveillance systems tend to be based on mobile edge computing networks. At the same time, it is combined with federated learning with distributed characteristics.

## 3. CMEC-based and blockchain-empowered intelligent video surveillance system.

3.1. **Innovation points.** Based on the core characteristics of distributed data storage, point-to-point transmission, consensus mechanism, encryption algorithm and other core features, blockchain technology is integrated into the mobile edge computing network to promote data security and traceability [59–61]. In this paper, we intends to propose a blockchain-empowered edge intelligent security monitoring and disaster prevention system in smart gardens, which integrate blockchain technology into Clustered MEC Networks proposed in our previous papers [59, 62]. At the same time, we explore and study the end-edge-cloud integration mechanism and the edge-edge, end-end collaboration method in the edge cluster. We aim at solving the network resource joint optimization problem of surveillance video caching, surveillance video transcoding, and real-time intelligent video processing in the intelligent surveillance system. Specifically, the main scheme of this edge intelligent surveillance system can be consist of CMEC-BC-IVS Scheme and FDRL-method.

3.1.1. *CMEC-BC-IVS Scheme.* A CMEC-based and blockchain-empowered intelligent surveillance system (CMEC-BC-IVS) has been proposed, which promotes surveillance video data security and privacy protection and ensures cooperation among edge cloud, edge node and end. And the CMEC-BC-IVS scheme is given in Figure 3. The architecture can achieve the following functions:

(i)Blockchain empowers data security and privacy protection. The blockchain established in the edge cluster can effectively protect the security of monitoring data and privacy information. In this framework, edge nodes are data producers and consumers, maintaining a permitted blockchain to ensure the security of cached multimedia data and the protection of user privacy. Through the integration of blockchain and edge computing, the blockchain can ensure the security of the entire edge computing network system through a strict identity verification mechanism and a cryptographic data encryption mechanism. At the same time, the integrity, authenticity and traceability of video data are guaranteed through blockchain technology.

(ii)The hierarchical network structure realizes high efficiency and accuracy processing of computing tasks. In the CMEC-BC-IVS system, the network is divided into multi-level hierarchical structures such as edge cloud, edge area, edge cluster, cluster head, and cluster element. Furthermore, the vertical integration among the edge cloud, edge nodes and terminals and the horizontal collaboration between edges, terminals can improve the real-time processing efficiency of edge surveillance video while maintaining low cost and less bandwidth traffic. At the same time, in an edge cluster, the collaboration and integration between edge nodes within the cluster can effectively promote the improvement of the computing processing capacity and recognition calculation accuracy of the edge cloud.

In the implementation of CMEC-BC-IVS, when there is a video surveillance computing task, the real-time requirement level of the task and the task calculation amount level are first determined according to the task demand, so as to determine the execution mode of the task according to the computing resources, storage resources and transmission resources. Generally speaking, the system includes four modes, namely end processing, edge processing, cloud processing, and end-edge-cloud fusion processing. Tasks with low computing requirements and high real-time requirements can choose end-processing or edge processing mode first, while tasks with large computing requirements and low real-time requirements can choose the operation processing mode. In addition, according to the needs of the task, you can flexibly choose joint processing modes such as end-to-edge fusion, edge-to-side fusion, and edge-to-cloud fusion.

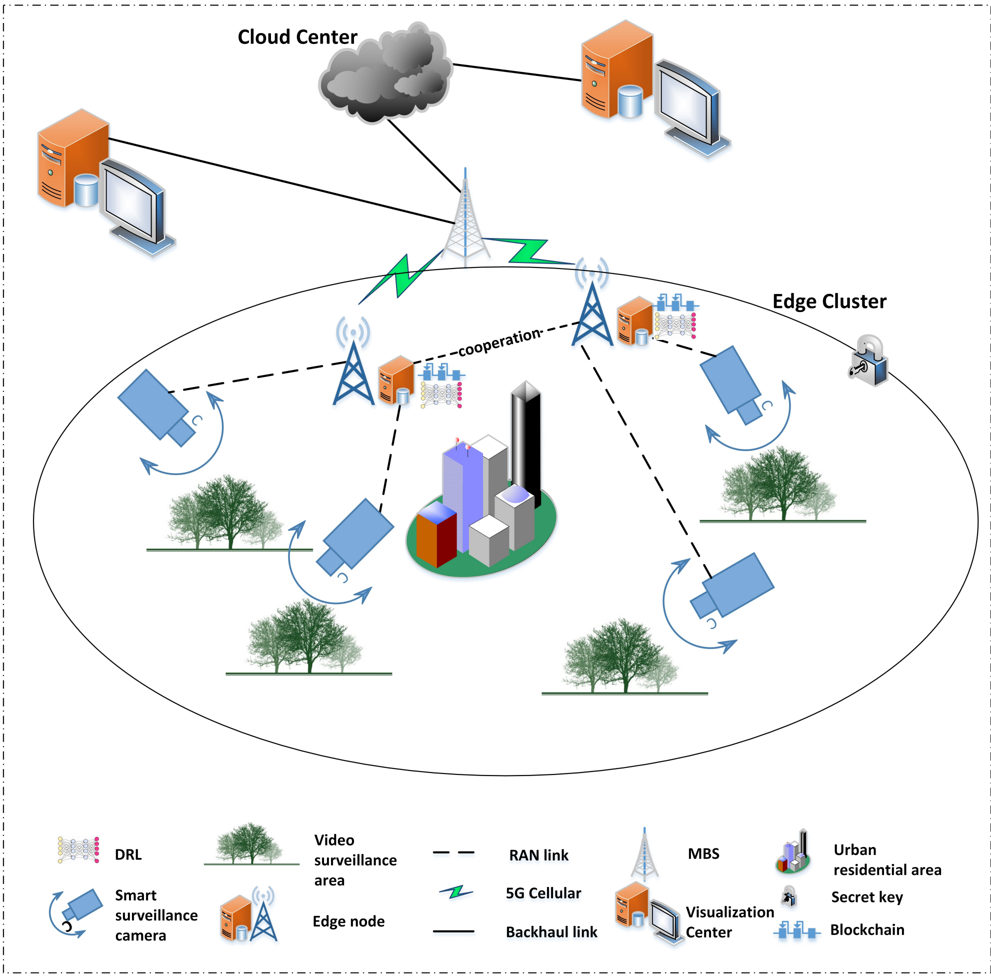


FIGURE 3. CMEC-based and blockchain-empowered intelligent video surveillance system

3.1.2. *FDRL-method.* CMEC-BC-IVS scheme can promote surveillance video data security and privacy protection and ensures cooperation among edge cloud, edge node and end. Then aiming at exploiting network resource joint optimization problem of surveillance video caching, surveillance video transcoding, and real-time intelligent video processing in the intelligent surveillance services, we focus on solving the complex joint optimization problems by using the FDRL-method algorithm in CMEC-BC-IVS model. The specific implementation mode of deep federated reinforcement learning is shown in Figure 4.

The FDRL-method in CMEC-BC-IVS networks has been designed. Based on federated learning that can effectively satisfy user privacy protection and data security, and combined with deep reinforcement learning that can make decisions about multi-dimensional, complex and heterogeneous problems, we propose a federal deep reinforcement learning model to be applied to edge intelligent video surveillance systems.

The network resource joint optimization problem of surveillance video caching, surveillance video transcoding, and real-time intelligent video processing in the edge intelligent surveillance system can be formulated as a stochastic MDP (Markov decision process). So we can solve the joint optimization problem by using the federal deep reinforcement learning to maximize the cumulative reward. Based on the network environment conditions and the needs of safety protection and management, the DRL algorithm will be used to implement the decision of surveillance video caching, transcoding, real-time processing, and response operations. Furthermore, deep learning training is realized through a federated learning model of distributed local training.

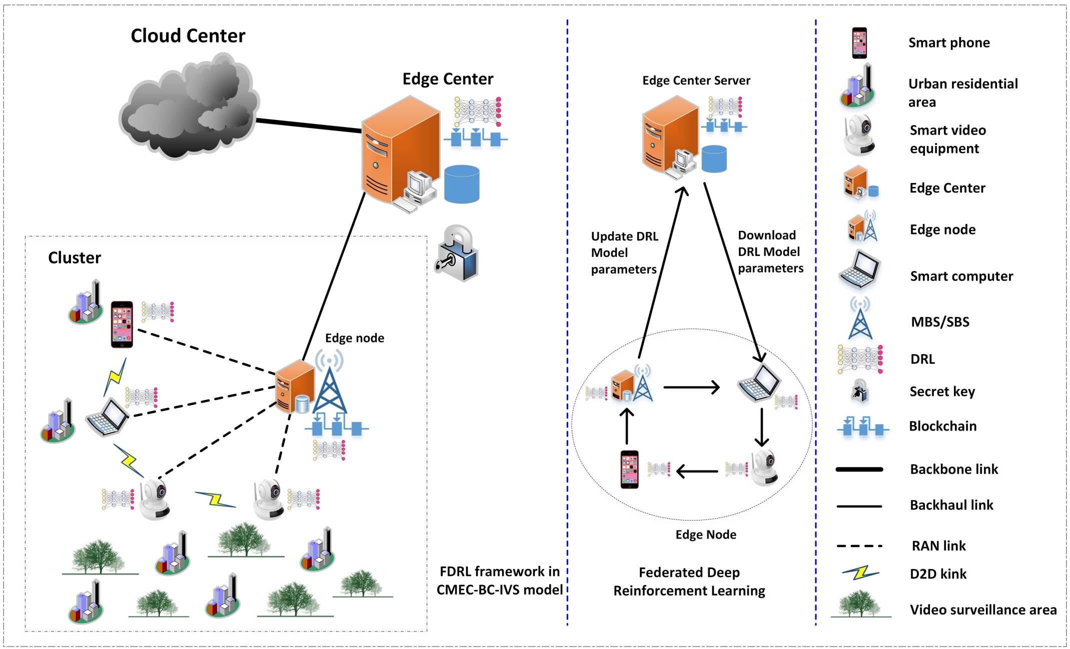


FIGURE 4. Federal deep reinforcement learning framework (FDRL-method) in CMEC-BC-IVS networks

The implementation of FDRL-method mainly includes two parts of the architecture and training of federated deep reinforcement learning algorithm, and the aggregation method of federated learning model. Federated deep reinforcement learning is proposed by integrating federated learning into the process of implementing deep reinforcement learning, which further protects the privacy and security of training data and effectively guarantees the security of data. In the process of federated deep reinforcement learning, the deep learning training location is selected at the edge node, and the edge node may be deployed in small base stations, macro base stations, or smart terminals with strong computing and storage capabilities. The update method of the model in federated learning intends to find a model aggregation method suitable for the current federated deep reinforcement learning architecture from the three directions of local update method, model compression scheme, and decentralized training.

### 3.2. **Specific key tasks of system implementation.**

3.2.1. *Real-time caching and transcoding service of surveillance video.* Intelligent video surveillance requires real-time intelligent processing such as timely extraction, identification and feedback decision-making on the collected video information. The amount of surveillance video data is huge, and it is impossible to store all surveillance videos in smart cameras or edge nodes. This requires designing surveillance video caching strategies according to requirements under constraints of the lowest costs. In order to obtain less time delay, due to the variability of the network environment, the video collected by the smart camera needs to determine the bitrate, resolution and other information of the surveillance video according to specific application requirements and network environmental conditions. Therefore, real-time caching and transcoding services are needed in video surveillance to meet the easy changing network environment and the needs of various applications, which is implemented by edge intelligence.

3.2.2. *Surveillance video security and personnel privacy.* In video surveillance, due to the limited storage and computing capabilities of smart terminals, surveillance video often needs to be transmitted to edge nodes or even to cloud centers for processing. As a result, the surveillance video data is exposed to the network, which can easily cause related problems such as data security and privacy protection. Furthermore, during the implementation of deep learning, data needs to be transmitted to computing capability centers such as edge clouds and cloud centers, also leading to security and privacy risks. Therefore, in order to solve these problems, blockchain and federated learning have been introduced into the framework of CMEC-BC-IVS, while research on this aspect in intelligent video surveillance has rarely been seen.

3.2.3. *Efficient integration of blockchain and mobile edge computing.* Both blockchain and MEC have distributed characteristics, and the integration between them has been explored and studied in many application scenarios. The smart video surveillance architecture in the smart gardens requires an effective and mature pattern of blockchain-empowered MEC model to support intelligent video processing and surveillance security protection.

3.2.4. *Implementation of Federal Deep Reinforcement Learning in edge computing networks.* Edge intelligence is a major trend in artificial intelligence research. How to implement deep learning inference and training at edge nodes is the core content of edge intelligence research. At the same time, the noval federal deep reinforcement learning, with its unique advantages, plays an important role in solving complex joint optimization problems in edge computing network scenarios.

4. **Trends and Challenges.** Based on the research situation of intelligent video surveillance and video streaming service in recent years [63, 64], combined with the CMEC-based and blockchain-empowered intelligent video surveillance system proposed in this article, we have given some research trends and challenges for your reference only.

4.1. **Research on the integration framework of blockchain and MEC.** Blockchain and mobile edge computing are both distributed structures. According to the task requirements in different application scenarios, how to effectively integrate blockchain technology and mobile edge computing network is a difficult problem to solve. Generally speaking, it can be done from two angles. On the one hand, it is based on the mobile edge computing network and integrates blockchain technology into the network. On the other hand, it is based on the blockchain technology architecture to apply mobile edge computing to solve blockchain tasks.

4.2. **Research on the fusion framework of deep reinforcement learning and federated learning.** In the intelligent video surveillance system network, how to apply deep reinforcement learning and federated learning to the edge is a difficult problem to be solved urgently. Especially when the resources of edge nodes are limited, how to implement deep learning training tasks well and obtain the required task processing accuracy? Among them, the most critical is the actual deployment of deep learning agents at the edge.

4.3. **Deep learning training on the edge.** The main research is how to use distributed storage and computing devices to train edge intelligence models under the constraints of resources and privacy. In general, it can be divided into two categories, distributed deep learning implementation in edge networks and federated deep learning implementation in edge networks. In the specific implementation process, the impact of factors such as data privacy, actual low relevance of data, aggregation efficiency, node heterogeneity and discrimination, etc., on the training effect should also be considered.

4.4. **Deep learning model inference.** This direction focuses on the actual deployment and reasoning of deep learning in edge computing infrastructure to meet different requirements, such as latency and accuracy. For example, streamlining deep learning model input, deep learning model segmentation in the edge, and deep learning model structure optimization, etc.

4.5. **Data security and privacy protection at the edge.** Due to the complexity and real-time nature of the edge computing service model, the multi-source heterogeneity and perception of data, and the resource-constrained characteristics of the terminal, the data security and privacy protection mechanisms in the traditional cloud computing environment are no longer suitable for the protection of massive data at the edge . Data storage security, sharing security, computing security, dissemination and control, and privacy protection are becoming more and more prominent.

5. **Conclusions.** This paper systematically has reviewed the MEC-based intelligent video surveillance architecture, and has given a comparison of the research content and methods of representative papers. The research content of the edge intelligent video surveillance system based on MEC is divided into three aspects, namely intelligent video recognition processing, intelligent video transmission and communication system, personal privacy and video data security. Based on the above analysis, this paper has proposed a new clustered-MEC intelligent video surveillance system based on federal deep reinforcement learning, and has given the innovation points of the system and specific key tasks of system implementation. Finally, the main research directions and trends of intelligent video surveillance research are given.

Based on the CMEC-based and blockchain-empowered intelligent video surveillance system and FDRL-method proposed in the article, follow-up research will continue to build an edge intelligent video surveillance simulation system to gradually improve the effectiveness and application of the proposed CMEC-BC-IVS scheme.

**Acknowledgment.** The authors would like to thank the associate editor and the anonymous reviewers for their outstanding comments and suggestions, which greatly helped to improve the technical quality and presentation of this paper. The work was supported by Science and Technology Project of Jiangxi Education Department (No. GJJ211917), National Natural Science Foundation of China (No. 61961021), and Innovation Special Fund for Individual Graduate Student of Jiangxi University of Finance and Economics (2020 Annual, No. 24).

## REFERENCES

[1] Y. Zeng, S. Zhou and K. Xiang, Online-offline interactive urban crowd flow prediction toward IoT-based smart city, *IEEE Transactions on Services Computing*,2021,https://doi.org/10.1109/TSC.2021.3099781.

[2] C.M. Chen, L. Chen, Y. Huang, S. Kumar, J.M. Wu, Lightweight authentication protocol in edge-based smart grid environment, *EURASIP Journal on Wireless Communications and Networking*,vol.2021:68,2021.

[3] J. Yang, T.Y. Lee and W. Zhang, Smart cities in china: a brief overview, *IT Professional*, vol. 23, no. 3, pp. 89-94, 2021.

[4] F. Viegas, J.L.V. Barbosa, R. Kunst and W.F. Heckler, UFollower: a model for smart cities based on ubiquitous security and surveillance, *IEEE Latin America Transactions*, vol. 19, no. 12, pp. 2019-2027, 2021.

[5] Cisco, Cisco visual networking index: Global mobile data traffic forecast update, *2017-2022 white paper*, Available: https://www.cisco.com/c/en/us/solutions/collateral/serviceprovider/visual-networking-index-vni/white-paper-c11-738429.html,2019.

[6] Y. Mao, C. You, J. Zhang, K. Huang, and K.B. Letaief, A survey on mobile edge computing: the communication perspective, *IEEE Communications Surveys & Tutorials*, vol. 19, pp. 232-2358, 2017.

[7] Y.C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young. Mobile edge computing-a Key technology towards 5G, *ETSI White Paper*, vol. 11, 2015.

[8] D.C. Luvizon, D. Picard and H. Tabia, Multi-task deep learning for real-time 3D human pose estimation and action recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2752-2764, 2021.

[9] J. Oh, G.Y. Youm and M. Kim, SPAM-Net: a CNN-based SAR target recognition network with pose angle marginalization learning, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 2, pp. 701-714, 2021.

[10] S. Yun, J. Choi, Y. Yoo, K. Yun and J.Y. Choi, Action-driven visual object tracking with deep reinforcement learning, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2239-2252, 2018.

[11] O. Rudovic, H.W. Park, J. Busche, B. Schuller, C. Breazeal and R.W. Picard, Personalized estimation of engagement from videos using active learning with deep reinforcement learning, *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 217-226, 2019.

[12] W. Xu, J. Yu, Z. Miao, L. Wan and Q. Ji, Spatio-temporal deep Q-networks for human activity localization, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 9, pp. 2984-2999, 2020.

[13] H.H. Nguyen, T.N. Ta, Turnstile jumping detection in real-time video surveillance, *Pacific-Rim Symposium on Image and Video Technology*, 2019, https://doi.org/10.1007/978-3-030-34879-3-30.

[14] A. Khasanova, A. Makhmutova and I. Anikin, Image denoising for video surveillance cameras based on deep learning techniques, *International Conference on Industrial Engineering, Applications and Manufacturing*, pp. 713-718, 2021.

[15] X. Xu, Q. Wu, L. Qi, W. Dou, S.B. Tsai and M.Z.A. Bhuiyan, Trust-aware service offloading for video surveillance in edge computing enabled internet of vehicles, *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1787-1796, 2021.

[16] C. Xue, P. Liu and W. Liu, Studies on a video surveillance system designed for deep learning, *IEEE International Conference on Imaging Systems and Techniques*, pp. 1-5, 2019.

[17] H. Luo, Y. Guo, T.S. Pan. Target detection and recognition for wide area border defense intelligent surveillance. *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 11, No. 4, pp. 199-208, 2020.

[18] H. Hu,H. Shan,Z. Zheng,Z. Huang,C. Cai,C. Wang,X. Zhen,L. Yu,Z. Zhang,and Q. Tony, Intelligent video surveillance based on mobile edge networks,*IEEE International Conference on Communication Systems*, pp. 286-291, 2018.

[19] J. Li, Z. Zheng, Y. Li, R. Ma and S.T. Xia, Multitask deep learning for edge intelligence video surveillance system, *IEEE 18th International Conference on Industrial Informatics*, pp. 579-584, 2020.

[20] S. More, A. Thanawala, T. Shaikh and N. Prabhu, Bird's Eye: analysis of video surveillance data using deep learning, *6th International Conference for Convergence in Technology*, pp. 1-5, 2021.

[21] R. Khurana and A.K.S. Kushwaha, Deep learning approaches for human activity recognition in video surveillance - a survey, *First International Conference on Secure Cyber Computing and Communication*, pp. 542-544, 2018.

[22] H. Du, L. Chen, J. Qian, J. Hou, T. Jung and X. Li, PatronuS: a system for privacy-preserving cloud video surveillance, *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 6, pp. 1252-1261, 2020.

[23] H. Kim, Y. Cha, T. Kim and P. Kim, A study on the security threats and privacy policy of intelligent video surveillance system considering 5G network architecture, *International Conference on Electronics, Information, and Communication*, pp. 1-4, 2020.

[24] A. Shifa, M.N. Asghar, M. Fleury, N. Kanwal, M.S. Ansari, B. Lee, M. Herbst, Y. Qiao, MuLViS: multi-level encryption based security system for surveillance videos, *IEEE Access*, vol. 8, pp. 177131-177155, 2020.

[25] Z. Xiong, Y. Zhang, D. Niyato, P. Wang and Z. Han, When mobile blockchain meets edge computing, *IEEE Communications Magazine*, vol. 56, no. 8, pp. 33-39, 2018.

[26] Y. Liu, F. R. Yu, X. Li, H. Ji and V. C. M. Leung, Blockchain and machine learning for communications and networking systems. *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1392-1431, 2020.

[27] A. Gumaei, M. Al-Rakhami, M.M. Hassan, P. Pace, G. Alai, K. Lin, G. Fortino, Deep learning and blockchain with edge computing for 5G-enabled drone identification and flight mode detection, *IEEE Network*, vol. 35, no. 1, pp. 94-100, 2021.

[28] X. Zhou, X. Xu, W. Liang, Z. Zeng and Z. Yan, Deep-learning-enhanced multitarget detection for end edge cloud surveillance in smart IoT, *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 12588-12596, 2021.

[29] J. Chen, K. Li, Q. Deng, K. Li and P. S. Yu, Distributed deep learning model for intelligent video surveillance systems with edge computing, *IEEE Transactions on Industrial Informatics*, https://doi.org/10.1109/TII.2019.2909473.

[30] L. Cui, D. Su, Y. Zhou, L. Zhang, Y. Wu and S. Chen, Edge learning for surveillance video uploading sharing in public transport systems, *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2274-2285, 2021.

[31] J. Huang, A. Huang and L. Wang, Intelligent video surveillance of tourist attractions based on virtual reality technology, *IEEE Access*, vol. 8, pp. 159220-159233, 2020.

[32] J. Xie, Y. Zheng, R. Du, W. Xiong, and J. Guo, Deep learning-based computer vision for surveillance in ITS: evaluation of state-of-the-art methods, *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3027-3042, 2021.

[33] A. A. Ahmed and M. Echi, Hawk-Eye: an AI-powered threat detector for intelligent surveillance cameras, *IEEE Access*, vol. 9, pp. 63283-63293, 2021.

[34] J. Zhang, C. Xu, Z. Gao, J.J.P.C. Rodrigues and V.H.C. de Albuquerque, Industrial pervasive edge computing-based intelligence IoT for surveillance saliency detection, *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 5012-5020, 2021.

[35] L. Kong and R. Dai, Object-detection-based video compression for wireless surveillance systems, *IEEE MultiMedia*, vol. 24, no. 2, pp. 76-85, 2017.

[36] X. Zhou, X. Xu, W. Liang, Z. Zeng and Z. Yan, Deep-learning-enhanced multitarget detection for end edge cloud surveillance in smart IoT, *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 12588-12596, 2021.

[37] K.-H. Choi and J.-E. Ha, Visual surveillance using deep reinforcement learning, *20th International Conference on Control Automation and Systems*, pp. 289-291, 2020.

[38] H.H. Nguyen, T.N. Ta, N.C. Nguyen, V.T. Bui, H.M. Pham and D.M. Nguyen, YOLO based real-time human detection for smart video surveillance at the edge, *IEEE Eighth International CONference on Communications and Electronics*, pp. 439-444, 2021.

[39] A. Bari, T. Saini and A. Kumar, Fire detection using deep transfer learning on surveillance videos, *Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks*, pp. 1061-1067, 2021.

[40] S. Aberkane and M. Elarbi, Deep reinforcement learning for real-world anomaly detection in surveillance videos, *6th International Conference on Image and Signal Processing and their Applications*, pp. 1-5, 2019.

[41] F.U.M. Ullah, K. Muhammad, I.U. Haq, N. Khan, A.A. Heidari, S.W. Baik, V. Albuquerque, AI assisted edge vision for violence detection in IoT based industrial surveillance networks, *IEEE Transactions on Industrial Informatics*, 2021, https://doi.org/10.1109/TII.2021.3116377.

[42] S. Ren, K. He, R. Girshick and J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no.6, pp. 1137-1149, 2017.

[43] C.H. Yeh,C.H. Lin, L.W. Kang, C. Huang, M. Lin, C. Chan, C.C. Wang, Lightweight deep neural network for joint learning of underwater object detection and color conversion, *IEEE Transactions on Neural Networks and Learning Systems*, 2021, https://doi.org/10.1109/TNNLS.2021.3072414.

[44] Y. Zhao, Y. Yin and G. Gui, Lightweight deep learning based intelligent edge surveillance techniques, *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 4, pp. 1146-1154, 2020.

[45] F. Schroff, D. Kalenichenko, and J. Philbin, FaceNet: a unified embedding for face recognition and clustering, *Proc. IEEE Comput. Vis. Pattern Recognit.*, pp. 815-823, 2015.

[46] K. Muhammad, S. Khan, V. Palade, I. Mehmood and V.H.C. de Albuquerque, Edge intelligence-assisted smoke detection in foggy surveillance environments, *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1067-1075, 2020.

[47] J.-Y. Yu, Y. Kim and Y.-G. Kim, Intelligent video data security: a survey and open challenges, *IEEE Access*, vol. 9, pp. 26948-26967, 2021.

[48] K. Wang, S.P. Xu, C.-M. Chen, S.H. Islam, M.M. Hassan, C. Savaglio, P. Pace, G. Aloi, A trusted consensus scheme for collaborative learning in the edge AI computing domain, *IEEE Network*, vol. 35, no. 1, pp. 204-210, 2021.

[49] Q. Ramirez, L. Sequeira and J. Ruiz-Mas, An edge-cloud approach for video surveillance in public transport vehicles, *IEEE Latin America Transactions*, vol. 19, no. 10, pp. 1763-1771, 2021.

[50] H. Hu, H. Shan, CH. Wang, T. Sun, X. Zhen, K. Yang, L. Yu, ZH.Y. Zhang, T.Q.S. Quek, Video surveillance on mobile edge networks: a reinforcement learning based approach, *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 4746-4760, 2020.

[51] Y. Kunpeng, H. Shan, T. Sun, R. Hu, Y. Wu, L. Yu, ZH.Y. Zhang, T.Q.S. Quekk, Reinforcement learning-based mobile edge computing and transmission scheduling for video surveillance, *IEEE Transactions on Emerging Topics in Computing*, 2021, https://doi:10.1109/TETC.2021.3073744.

[52] R. Ke, Y. Zhuang, Z. Pu and Y. Wang, A smart, efficient, and reliable parking surveillance system with edge artificial intelligence on IoT devices, *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 4962-4974, 2021.

[53] T. Sultana and K. A. Wahid, Choice of application layer protocols for next generation video surveillance using internet of video things, *IEEE Access*, vol. 7, pp. 41607-41624, 2019.

[54] Z. Ming, J. Chen, L. Cui, S. Yang, Y. Pan, W. Xiao, L. Zhou, Edge-based video surveillance with graph-assisted reinforcement learning in smart construction, *IEEE Internet of Things Journal*, 2021, https://doi.org/10.1109/JIOT.2021.3090513.

[55] B. Hou and J. Zhang, Real-time surveillance video salient object detection using collaborative cloud-edge deep reinforcement learning, *International Joint Conference on Neural Networks*, pp. 1-8, 2021.

[56] D. Kim, K. Kim and S. Park, Automatic PTZ camera control based on deep Q network in video surveillance system, *International Conference on Electronics, Information, and Communication*, pp. 1-3, 2019.

[57] L. Li, Z. Wang, Q. Hu and Y. Dong, Adaptive nonconvex sparsity based background subtraction for intelligent video surveillance, *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4168-4178, 2020.

[58] Y. Wei, S. Zhou, S. Leng, S. Maharjan and Y. Zhang, Federated learning empowered end-edge-cloud cooperation for 5G hetNet security, *IEEE Network*, vol. 35, no. 2, pp. 88-94, 2021.

[59] Y. Li, ZH. Wan, Blockchain-enabled intelligent video caching and transcoding in clustered MEC networks, *Security and Communication Networks*, vol. 2021, 7443260, 2021.

[60] Y. Liu, F. R. Yu, X. Li, H. Ji and V. C. M. Leung, Blockchain and machine learning for communications and networking systems. *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1392-1431, 2020.

[61] D.C. Nguyen, P.N. Pathirana, M. Ding, A. Seneviratne, Blockchain for 5G and beyond networks: a state of the art survey. *Journal of Network and Computer Applications*, Vol. 166, 102693, 2020.

[62] ZH. Wan, Y. Li, Deep reinforcement learning based collaborative video caching and transcoding in clustered and intelligent edge B5G networks, *Wireless Communications and Mobile Computing*, vol. 2020, 6684293, 2020.

[63] X. Wang, Y. Han, V.C.M. Leung, D. Niyato, X. Yan and X. Chen, Convergence of edge computing and deep learning: a comprehensive survey. *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 869-904, 2020.

[64] B. Jedari, G. Premsankar, G. Illahi, M.D. Francesco, A. Mehrabi and A. Yla Jaaski, Video caching, analytics, and delivery at the wireless edge: A survey and future directions, *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 431-471, 2021.