Journal of
Sensor and Actuator Networks

Review

Mobile Edge Computing in Space-Air-Ground Integrated Networks: Architectures, Key Technologies and Challenges

Yuan Qiu [1,2], Jianwei Niu [1], Xinzhong Zhu [2], Kuntuo Zhu [2], Yiming Yao [1], Beibei Ren [3] and Tao Ren [1,*]

[1] School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[2] Shanghai Aerospace Electronic Technology Institute, Shanghai 201109, China
[3] STL in Shanghai Institute for Integrated Application of Network Technology, Shanghai 201103, China
* Correspondence: rentaobh@buaa.edu.cn

Abstract: Space-air-ground integrated networks (SAGIN) provide seamless global coverage and cross-domain interconnection for the ubiquitous users in heterogeneous networks, which greatly promote the rapid development of intelligent mobile devices and applications. However, for mobile devices with limited computation capability and energy budgets, it is still a serious challenge to meet the stringent delay and energy requirements of computation-intensive ubiquitous mobile applications. Therefore, in view of the significant success in ground mobile networks, the introduction of mobile edge computing (MEC) in SAGIN has become a promising technology to solve the challenge. By deploying computing, cache, and communication resources in the edge of mobile networks, SAGIN MEC provides both low latency, high bandwidth, and wide coverage, substantially improving the quality of services for mobile applications. There are still many unprecedented challenges, due to its high dynamic, heterogeneous and complex time-varying topology. Therefore, efficient MEC deployment, resource management, and scheduling optimization in SAGIN are of great significance. However, most existing surveys only focus on either the network architecture and system model, or the analysis of specific technologies of computation offloading, without a complete description of the key MEC technologies for SAGIN. Motivated by this, this paper first presents a SAGIN network system architecture and service framework, followed by the descriptions of its characteristics and advantages. Then, the MEC deployment, network resources, edge intelligence, optimization objectives and key algorithms in SAGIN are discussed in detail. Finally, potential problems and challenges of MEC in SAGIN are discussed for future work.

Keywords: space-air-ground integrated network; mobile edge computing; resource scheduling; service framework

check for updates

Citation: Qiu, Y.; Niu, J.; Zhu, X.; Zhu, K.; Yao, Y.; Ren, B.; Ren, T. Mobile Edge Computing in Space-Air-Ground Integrated Networks: Architectures, Key Technologies and Challenges. *J. Sens. Actuator Netw.* **2022**, *11*, 57. https://doi.org/10.3390/jsan11040057

Academic Editor: Mingjun Xiao

Received: 10 August 2022
Accepted: 14 September 2022
Published: 22 September 2022

1. Introduction

The rapid development of next-generation communication technologies, integrated with popular information technologies, e.g., Internet, big data and artificial intelligence, has spawned many new mobile applications, such as mobile payment, online games, telemedicine, and unmanned driving. Moreover, the communication technologies have also been applied in industry, transportation, medical care, education, etc., effectively improving informatization and digital transformation in all aspects of life [1,2]. In addition, with the rapid development of terrestrial mobile communication, Internet of things (IoTs) with intelligent sensing capabilities have also undergone unprecedented development. According to Cisco's estimation [3], there will be 29.3 billion devices connected to networks worldwide by 2023, of which about 14.7 billion IoT devices will be connected to the Internet, roughly accounting for 50% of network devices. However, the large-scale IoT devices generate large amounts of data all the time, resulting in a large burden of real-time connecting and processing on their hosts. In short, the interaction and transmission of large information put forward high requirements for reliable and stable communication networks.

Although the ground communication network can be equipped with powerful resources to provide high-speed communication, it is still unavoidable that the communication blockage caused by geological disasters or abnormal connectivity, due to harsh natural conditions, will eventually lead to blind spots in the communication coverage. Therefore, space and satellite integrated communication with wide-area range and broadband communication capability becomes a prospective technology to provide seamless coverage and connectivity without geographical restrictions. Both domestic and foreign companies, including SpaceX, OneWeb, and China Satellite Network Group, have launched constellation plans to launch giant low earth orbit (LEO) satellites, and their services will be established in the future, which is crucial for Internet access [4]. However, bandwidth and energy constraints lead to limited computing and communication resources compared to ground base stations, and periodic orbital flights often require inter-satellite communication to achieve coverage relays. Therefore, to more effectively meet the application requirements of ground users, building a multi-level network architecture with multi-domain interconnection and collaboration is of great significant to provide heterogeneous spatial services.

The space-air-ground integrated network (SAGIN) integrates satellite constellation, air network, and ground communication as a cross-domain heterogeneous system, providing flexible node access, high-speed data transmission, and seamless coverage to users' terminals in various domains. The interconnection performance of SAGIN is beyond the capability of the traditional single network form, which has abstracted wide attention from worldwide researchers [5]. For example, the European Space Agency (ESA) recently announced the first successful trial of intercontinental connectivity between Europe and Japan through the fifth-generation (5G) telecommunications link via satellites and are also planning to develop technologies to connect multiple nodes, such as satellites, unmanned aerial vehicles (UAVs), and high altitude platforms (HAPs), to obtain more ubiquitous connectivity [6].

Although SAGIN provides users with elastic, flexible, and reliable access and communication, it still faces some challenges, especially the contradiction between users' increasingly complex applications and the quality of services. For instance, for emerging network services, such as intelligent interaction, game rendering, and 4K/8K video transmission, the limited battery capacity and computing power of local devices result in long-time task delays and high energy consumption, which is far away from users' demands. To solve the above problems, mobile edge computing (MEC) is proposed as a promising technology to provide users with both satisfactory computing capabilities and task latency. MEC deploys computing and storage resources at the edge of the mobile network, which provides cloud computing capabilities for mobile users with largely reduced latency [7]. In view of this, MEC is also introduced into SAGIN, significantly improving the serving ability of SAGIN for ground users' computing requirements, reducing the task latency and energy consumption, and meeting the demands for service of quality.

The terrestrial mobile network brings cloud computing capabilities to the network edge through MEC, improving the quality of service for multi-tasking of multi-users and reducing system costs [8]. However, it is worth noting that in the case of ground communication interruption, such as in oceans, deserts, or disaster environments, it is hard to deploy terrestrial networks and provide MEC services to ground users. In this case, it is necessary to build SAGIN to offload computing tasks to space-based MEC or air-based MEC nodes to realize real-time computation offloading, on-board data processing, and other network tasks. The deployment of MEC will be explained in detail later.

In summary, according to specific application scenarios and users' demands, SAGIN deploys MEC resources to provide fast-response and seamless coverage of computing offloading to improve the efficiency of network resource utilization in the whole domain, as well as the quality of users' complex services, which has become a research hotspot in the global academic and industrial community and has achieved many representative results in recent years. However, existing survey papers only provide SAGIN's network architecture, protocol and system design, or summarize specific offloading problems or optimization algorithms, without comprehensively combing or reviewing the key MEC technologies for various SAGIN application scenarios. The main contributions of this paper are summarized as follows:

1. Considering the complex and varying application scenarios of the SAGIN, this paper presents a three-layer network architecture and service framework, and analyzes its advantages in solving the following three challenges: inaccessibility, optimization difficulty, and incompleteness.
2. To the best of our knowledge, we are the first to classify and summarize the MEC technology of SAGIN from the aspects of MEC deployment, network resources, optimization objectives and decision algorithms.

The rest of this paper is organized as follows. Section 2 introduces the basic concepts of edge computing in SAGIN. In Section 3, we present the network architecture, service framework, characteristics, and advantages, followed by the detailed explanations of critical technologies in SAGIN, such as network MEC deployments, service resources, user demand objectives, and related algorithms. Section 5 provides a detailed analysis and discussion of some open issues and future research directions. Finally, Section 6 concludes the paper.

The rest of this paper is organized as follows. Section 2 introduces the basic concepts of edge computing in SAGIN. In Section 3, we present the network architecture, service framework, characteristics, and advantages. In Section 4, we discuss critical technologies in SAGIN, such as network MEC deployments, service resources, edge intelligence, user demand objectives and related algorithms. Section 5 provides a detailed analysis and discussion of some open issues. Section 6 presents possible research directions in the future. Finally, Section 7 concludes the paper.

2. Background

2.1. Related Networks

2.1.1. Terrestrial Mobile Network

Since the 1980s, the terrestrial mobile wireless communication network has undergone tremendous innovation and has now developed to the fifth generation, which is referred to as 5G. Each generation of mobile wireless networks has unique standards, technologies, and characteristics compared to the previous generation. For example, first-generation (1G) uses analog voice modulation technology, which only provides local voice communication; second-generation (2G) began to use digital communication technology to support text transmission; third-generation (3G) provides multimedia services along with higher data rates and greater capacity; fourth-generation (4G) introduced orthogonal frequency division multiplexing (OFDM) and multiple input multiple output (MIMO) technology, which significantly improves mobile broadband service capabilities. The first four generations of mobile communications were all oriented toward human-centric application scenarios. At the same time, 5G extends the application to people and things, realizing the network architecture of the Internet of everything. The main communication scenarios of 5G include enhanced mobile broadband (eMBB), ultra-reliable low-latency communication (URLLC), and massive machine-type communication (mMTC), which can provide end users with transmission rates up to 20Gb/s [9]. In addition, 5G provides more flexible services, greater capacity, and higher efficiency for new network applications, such as virtual reality, autonomous driving, and smart cities, which have now entered the stage of deployment and commercial usage. However, due to the small coverage area of base stations, large-scale 5G network deployment requires many infrastructures, such as base stations and backhaul networks, resulting in high infrastructure costs and maintenance costs. The coverage of 5G base stations is concentrated within 10 km of the land surface. In the 5G era, more than 80% of the land area and more than 95% of the ocean area are not covered by the mobile network signal. There are still many communication blind spots for the construction of network applications covering the world. Therefore, while 5G is accelerating its entry into large-scale commercial applications, the IEEE 802.11ax standard for next-generation wireless networks has gradually become a research hotspot.

Based on 5G, the upcoming sixth-generation (6G) massive IoT network architecture will consist of space, air, ground, underwater/sea networks, and edge computing [10], which provides information assurance for any user with access to any subnet. 6G network performance should achieve greater connection density, transmission bandwidth, lower end-to-end latency, and higher reliability and intelligence, supporting the deep integration and long-term development of mobile communication networks and vertical industries [11]. 6G is expected to cover about 100% of the Earth's surface, providing sub-centimeter-level positioning accuracy and millisecond-level positioning update rates, and will be a hybrid network of fixed, mobile cellular, high-altitude platforms satellite, and others yet to be defined [12]. Diverse heterogeneous networks, rich communication scenarios, ultra-high bandwidth, and other factors will always generate large amounts of data. Therefore, 6G will realize a series of new intelligent applications with the help of artificial intelligence and machine learning technologies. They improve network performance in terms of quality of service (QoS) and security [13].

The International Telecommunication Union has not yet developed a 6G standard. However, ITU-T 2030, a study group established by the International Telecommunication Union, believes that 6G networks are designed to provide a revolutionary user experience, the connection speed on the order of Tb/s, and new sensory information, such as touch, taste, and smell. It is expected that the application scenarios of 6G include enhanced applications based on 5G, such as further enhanced mobile broadband (FeMBB) and ultra-massive machine-type communication (umMTC) in the future, as well as a large number of new applications that can only be realized under 6G networks, such as holographic communication, etc.

2.1.2. Air-Ground Network

In recent years, with the sharp increase in mobile users, the original network topology can no longer meet the communication needs of users. The single-layer macro-cellular network is evolving into a multi-layer heterogeneous macro-cellular/small-cell network. This heterogeneous network networking method provides a reference for the combined air-ground integration network. The air-ground integrated communication network refers to the introduction of ground-based communication methods within the coverage of UAVs, HAPs, and other nodes, to strengthen the scope of hotspot areas in a targeted manner and to form a complex and heterogeneous network structure with overlapping coverage. On the one hand, through air-to-air communication using heterogeneous radio interfaces, such as IEEE 802.15.4 or WiFi, air nodes can form mobile ad hoc networks, which are more effectively used for wide-area coverage, which is beyond the range of ground-based networks. Users are provided with broadband wireless access services, thereby increasing network capacity [14]. On the other hand, users can independently choose to access the most suitable space-based or ground-based wireless network to meet their flexible business needs. In addition, the diversity and complexity of the battlefield environment also make data collection and transmission extremely challenging. In these unreliable communication environments, in order to enable these systems to operate in the network to provide reliable and timely information exchange, it is necessary to study innovative solutions, combining multiple technologies to achieve network- and information-centric objectives. The full-dimensional battlefield sensor network system uses the air-ground integrated network as the communication infrastructure to provide relay communication for sensor nodes of various forms and to organically integrate with all the sensors installed on satellites and ships.

2.1.3. Space-Ground Network

The space-ground network is between satellites and the ground and, between satellite constellations. It includes the communication network between satellites, terrestrial infrastructure, and satellite systems with different orbital planes, types, and architectures [15]. Compared with the traditional network, the space-ground network structure is more dynamic and three-dimensional, and the links between satellites are more flexible, enabling more extensive network coverage, including deserts, oceans, deep space, and other unexplored places. As a result, it can be widely used in navigation, communication, broadcasting, remote sensing, and other fields, which combine with different needs to provide safe, fast, and efficient services to different types of users.

According to [16], the current space-ground integrated network can be roughly divided into two categories. One type is based on geostationary earth orbit (GEO) satellite nodes. Theoretically, three GEO satellite nodes can achieve global coverage. The International Mobile Satellite (Inmarsat) system is the world's first commercial satellite mobile communication service operator to provide global coverage. Inmarsat has grown to a fifth-generation system, with a total of 13 GEO satellite nodes in orbit, 5 fourth-generation Inmarsat satellite nodes using the L-band, 5 fifth-generation satellite nodes using the Ka-band, and a few other satellite nodes serving a specific area.

Another type is based on LEO satellite nodes. The LEO satellite constellation has the characteristics of low delay, strong signal, mass production, and low cost. At present, many foreign companies have planned low-orbit Internet constellation plans. For instance, Starlink of the SpaceX Company has entered the stage of large-scale deployment. Currently, the number of satellites in orbit has exceeded 2750, which is expected to reach 42,000.

The satellite constellation network is the inevitable trend of future network infrastructure development. A satellite constellation is a system composed of several satellites with the same functions, distributed in satellite orbits according to specific rules to achieve cooperation [17]. Typical constellation designs include the following two categories: polar orbit constellations and inclined orbit constellations. All satellites in the polar orbit constellation have the same orbital inclination and orbital height. A constellation consists of multiple orbits with the same and evenly distributed satellites on each orbital plane. The preference of satellite orbits in the inclined orbit constellation is less than $\pi/2$, does not pass through the poles, and the ascending nodes of all orbits are evenly distributed within the longitude of the equatorial reference plane. Under the condition that the satellite orbit height and coverage are the same, the total number of satellites required by the inclined orbit constellation is usually less than that of the polar orbit constellation. However, the coverage area of the inclined orbit satellite is irregular, and the changing relationship between the links is relatively complex.

The future development direction of the space-ground network includes, but is not limited to, the following:

- Multi-layer constellation networks. The space-based networks that have been built so far are all single-layer constellations. However, multi-layer constellation networks have more robust performance than single-layer constellation networks in terms of all-around performance, network anti-blocking, and survivability.
- Space-based network expands to multi-function. With the development of technologies such as communication and satellite payloads, the functions of space-based networks will be extended from the original single communication function to the multi-functional expansion of communication, navigation enhancement, earth observation, and IoT.
- Deep integration of space and ground. It includes the integration of heterogeneous networks in space and the ground and the smoother inter-satellite link connection between constellations.

2.2. Mobile Edge Computing

Mobile edge computing is a new computing paradigm developed based on cloud computing. The characteristic of cloud computing is that the data generated by the network edge device are centralized on the cloud server for processing. As a result, the edge device must frequently exchange information with the cloud server; however, the edge device is usually far away from the cloud server. Hence, it has higher requirements for real-time performance. In the scenario of cloud computing, the inadequacy of cloud computing limits its application. Due to the proposal of mobile edge computing, the tension was partly alleviated by offloading computation-intensive and delay-critical tasks to edge servers, which are deployed with powerful computing and energy resources close to ground users [18].

According to [19], edge computing is a federated network structure that extends cloud services by introducing edge devices between end nodes and cloud computing. The cloud-edge collaboration framework is usually divided into the terminal layer, the edge layer, and the cloud computing layer. The terminal layer consists of all the devices connected to the edge network, including mobile terminals and IoT devices. The terminal layer is both a data generator and a data user. The edge layer is the core of the three-layer architecture, usually including base stations, access points, routers, switches, gateways, etc. The edge layer supports the access of terminal devices, stores and processes data uploaded by terminal apparatus, and connects to the cloud computing layer. Since the edge layer is closer to the user, real-time data processing and intelligent analysis can be performed more efficiently. The cloud computing layer includes high-performance servers and storage devices with powerful computing and storage capabilities. The cloud computing layer can permanently store the data generated by the edge layer. It can also handle the analysis tasks that the edge layer cannot manage and information processing tasks that involve the entire network. In addition, the cloud computing layer can dynamically adjust the deployment of edge layer devices.

3. MEC in SAGIN

3.1. Network Architecture

In previous research, the SAGIN integrated MEC technology has developed rapidly. Taking advantage of its characteristics such as low latency, high bandwidth, and ubiquitous coverage, it has begun to be deployed in practical applications. To meet the challenges of system applications, scientific research institutions and academia have carried out relevant research. Feng et al. in [20] proposed a flexible network structure regarded as HetNet, which can effectively integrate heterogeneous satellite-terrestrial networks. It also uses software-defined networking and network function virtualization technology to improve the network's resilience, but the UAV edge nodes in the air segment are not considered. Lu et al. in [21] considered both the time-varying ground channel and the line-of-sight air-ground channel and a robust optimization scheme for efficient cooperative double-MEC resource scheduling between air and ground was proposed. Still, the space-based network nodes are not incorporated into the architecture. However, Pfandzelter et al. in [22] only focused on the low-orbit constellation edge platform. A simulation and verification test bed were also designed to evaluate mission migration and space-ground integration. Many scholars have designed various types of SAGIN edge computing frameworks and proposed the joint scheduling optimization problem of edge service fusion or cloud-edge collaboration framework for various user terminals, such as 5G/6G IoT, Internet of vehicles, and ships, to achieve user service quality requirements, such as minimum total system energy consumption or minimum delay under power consumption constraints. However, the proposed network architecture applies to relatively limited application scenarios, and network performance objectives, for example, reliability and robustness are missing. Key technologies such as the software-defined network (SDN), network function virtualization (NFV) and network components have not been considered comprehensively [23–26]. To solve the application challenges faced by SAGIN and to analyze the differences of network layers, functions of components at each layer, and application scenarios in a more comprehensive and detailed way, this paper designs a heterogeneous cross-domain network architecture driven by multi-type resource services, which is composed of hybrid orbital constellation, aerial UAV, ground terminals and other infrastructures. It should be pointed out that the network architecture and application scenarios concerned in this paper are composed of cross-domain nodes in SAGIN, which inevitably include air-ground networks and space-ground networks. Figure 1 shows the proposed network architecture.

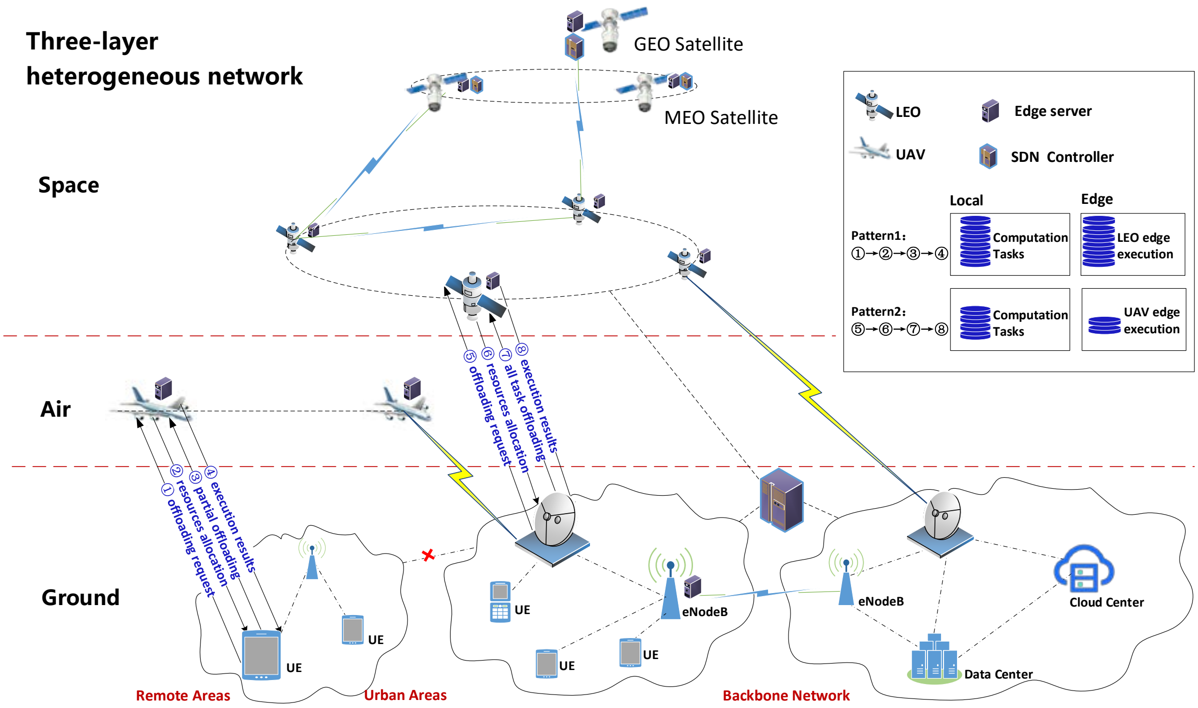

Figure 1. SAGIN system architecture.

The composition and functions of each layer network are summarized as follows.

(1) Space segment

The space segment includes satellite constellations of LEO, medium earth orbit (MEO), GEO and other different orbital heights. Among them, the propagation delay of the GEO satellite is 120 ms, that of the MEO satellite is 10 ms, and that of the LEO satellite is several ms or even less than 1ms [27]. With the development of lightweight satellite technology, phased array multi-beam antenna, onboard real-time processing, inter-satellite links, and other technologies, LEO satellites have become more economical, miniaturized, and well-resourced. Many research institutions and academic organizations have carried out LEO constellation construction and put it into use, including Globalstar, SpaceX, OneWeb, and so on. LEO satellites can provide round-the-clock global coverage through constellation design, which cannot be achieved by ground base stations. In the SAGIN system, we use LEO satellites with sufficient computing power as edge server nodes to provide multiple types of edge processing services. It is worth noting that the satellite constellation can adopt the master-slave formation flight mode to ensure the link stability when the network topology is relatively fixed, or it can adopt the self-organizing Ad-hoc network architecture to improve the system flexibility and robustness, which serve as an important basis for satellite mission coordination and resource scheduling. Considering the propagation time of the satellite and ground, it is more suitable for delay tolerance and large bandwidth requirements under other similar constraints.

The control function in the SAGIN control layer based on SDN is usually deployed in the MEO or GEO satellites. As the core component of SAGIN, the control layer is responsible for controlling and managing multiple resources in the data layer [28]. In particular, NFV technology pools all types of resources to achieve network resource scheduling and collaborative management.

(2) Air segment

In the aerial segment, data collection, transmission, and task processing are mainly carried out by UAVs, HAPs, and other carriers. Their transmission delay is shorter than that of satellites, their coverage is broader than that of ground base stations, and they have the characteristics of flexible deployment, low cost, fast response, and constantly improving endurance. Flying UAVs, such as the Facebook Aquila, can fly for months without being charged by solar panels [29]. Furthermore, UAVs can be used as network relay and forwarding nodes, edge service nodes, or aerial Internet of things nodes. Specifically, UAV can provide a relay channel for space-ground data transmission and task offloading, realizing signal enhancement and anti-interference. Also, UAV can directly forward data to the ground cloud computing center. In addition, the UAV platform loaded with sufficient computing power and storage resources can now be used as an edge computing node to provide network services without satellite transmission. Finally, in environmental monitoring or emergency rescue scenarios, UAVs can be used as IoT devices to be dynamically deployed in the covered area and complete characteristic collection tasks to bring back to the ground or offload to the satellite for execution. The UAV flies according to the predetermined trajectory or adjusts the course dynamically according to the optimized target to improve the system efficiency.

The aerial network can achieve low latency and extensive area coverage, but the capacity is limited, and the link is unstable.

(3) Ground segment

The broad ground segment mentioned in this paper includes remote areas, disaster zones, ocean, polar regions, and even underwater. The ground segment includes IoT devices, user terminals, communication stations, ground base stations, and other infrastructure supporting ground services. The ground base station has a strong computing capacity and sufficient power consumption. The edge server configured in the ground base station can provide low-latency and high-efficiency nearby services, but it is vulnerable to terrain limitations and deployment constraints. According to user density, the ground coverage area is divided into dense and remote areas. In dense areas, with a large number of people and devices and a large amount of information, resources are usually deployed at the edge of the cellular network to provide nearby services to reduce task delay and to improve service quality. However, in the case of saturated flow, network overload will need to uninstall missions to air or space segment to meet users' needs. In addition, ground network edge servers cannot be deployed in remote areas, such as deserts and uninhabited islands. In this case, specific ground terminals need to access air or space networks to obtain computing, storage, communication, and other edge services. According to [30], the ground SDN control center equipped with virtualization pools and other facilities will also become the control layer components of SAGIN, including a logically centralized and physically distributed multi-controller system.

SAGIN is a comprehensive network of heterogeneous integration and cross-domain interconnection of multi-space and multi-network domains, covering space, air, land, sea, etc., breaking through the restriction of terrain and surface, and truly realizing the "ubiquitous connection" of the whole world. The three network segments of SAGIN can work independently or interoperate with each other. Due to the different channel characteristics and network architecture of each network segment, the three network segments may adopt different access modes and network protocols. Therefore, cross-domain communication or unified protocol conversion is required to construct a layered broadband wireless network. For SAGIN, the actual resource constraints of each network segment are different, and the limited network resources need to be used to obtain the best performance of data transmission and information exchange, especially for the interoperation of different network segments. Therefore, its basic feature is comprehensiveness. Through the unified management of multiple network system resources, the global resource utilization efficiency is improved, so as to provide global users with uninterrupted and consistent information services.

In the SAGIN architecture, whether it is the user terminal, IoT, Internet of vehicles (IoV), the base station on the ground, UAV, hot air balloon, or satellite in space, they can be abstracted as "nodes", integrating computing, storage and communication resources. Constrained by physical and electrical parameters, such as volume, weight, and power consumption, network nodes have different resource types and capabilities. Under the particular scenario when nodes produce computationally intensive, delay-sensitive, multi-node cooperative tasks, its local implementation may not be able to meet the quality requirements due to its resource capacity limit. Therefore, according to the decision policy, edge computing services need to be requested nearby. Therefore, participants were respectively defined as “service requester” (user node) and “resource provider” (edge node) in the task-driven SAGIN edge computing scenario. Generally, by loading edge servers (in general, they have spare resource capacity) to the edge of the end, their affordable computing, storage, communication, and other resources are provided to the nearest user to meet the service quality requirements, such as low power consumption and low latency, in specific task scenarios. It is worth noting that user node QoS requirements may differ in different business and application scenarios. For example, in the same space-ground collaboration scenario, the targets such as transmission power and transmission delay are respectively focused on [19,31,32].

With the cross-domain connectivity of SAGIN, the edge nodes of the traditional terrestrial cellular network will also be extended to the air and space to provide better quality, network performance, and resource services. However, despite such comprehensive coverage of edge computing, it may fail to meet users’ needs due to the lack of maximum computing power, storage capacity, or communication bandwidth. Therefore, SAGIN will also introduce a cloud-edge collaboration framework based on a cloud center with more sufficient resources and powerful capabilities, which will be deployed in MEO/GEO or ground cloud computing centers and provide sufficient computing and processing communication resources [33–35]. However, since this paper mainly focuses on edge computing technology, the cloud-edge collaboration model will not be described in detail.

3.2. Resource Service Framework

Although high-density users have diverse application requirements, the basic process of edge computing is that edge nodes will serve user nodes in the nearby network domain under specific scheduling and arrangement of their surplus resources to respond to their business needs. Therefore, from the perspective of resource supply and acquisition, this paper abstracts the edge computing service process of SAGIN to form a layered framework [36–38]. The framework consists of the infrastructure, data, control, service, and application layers, as shown in Figure 2.

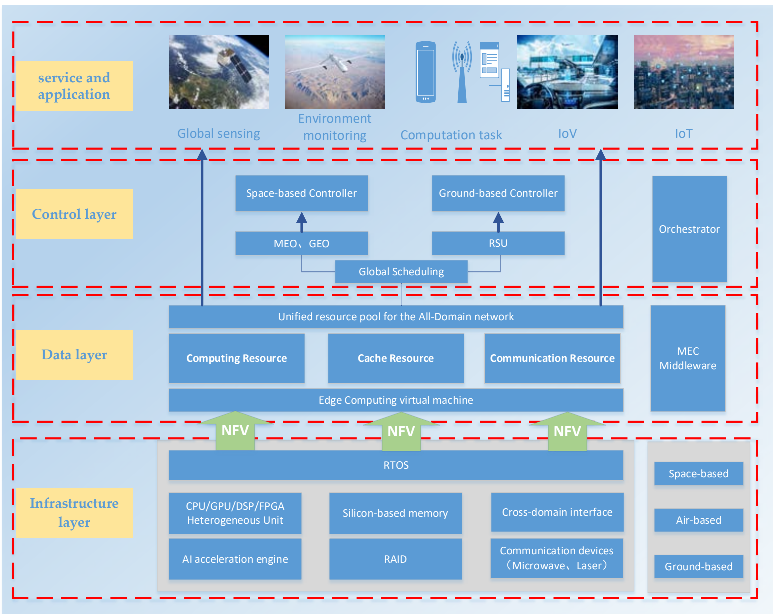

Figure 2. SAGIN resource service framework.

The infrastructure layer is an important guarantee for SAGIN to realize heterogeneous mobile edge computing with multiple users and resources. The facility can be divided into the following three types: computing, storage, and communication. Computing facilities are mainly constructed by various integrated circuits, such as the central processing unit (CPU), graphics processing unit (GPU), digital signal processor (DSP), field programmable gate array (FPGA), etc., with different topological forms. For example, a multi-level redundant backup method can be adopted to improve the reliability and service life of hardware devices and shield a single point of failure. To further enhance edge intelligence, artificial intelligence (AI) acceleration modules can also be loaded. For example, the satellite edge platform supports complex orbital applications, due to its high-performance computing architecture and functional modules [39–41]. Storage facilities mainly rely on non-volatile flash memory to achieve large-capacity storage space and support the storage and application of massive data. To improve the data security, redundant arrays of independent disks (RAID) can also improve disaster recovery performance. Communication facilities realize SAGIN cross-domain interconnection, and users can obtain flexible links through the access interface. Considering the radical goals of the next-generation cellular network in terms of 1000 times the capacity, 100 times the energy efficiency, and 10 times the delay, new access technologies such as MIMO and non-orthogonal multiple access (NOMA) could be adopted to improve system performance [42]. Meanwhile, the communication link always pursues a large bandwidth and high speed to achieve low-latency data services. For example, according to [43], the inter-satellite laser communication link can achieve 100 Gbps data rate capability. In addition, the SAGIN infrastructure is deployed in space, air, and on the ground according to different network domains. In order to provide high-quality edge services to multi-user terminals, different edge nodes in each domain of SAGIN are configured with the above heterogeneous resources. However, due to the limitations of the volume, weight, and energy consumption of each platform, the resource types and capabilities of each node are different, which leads to the imbalance of network resources. Therefore, it is necessary to offload and allocate resources among heterogeneous edge nodes according to the QoS requirements of user tasks to achieve the optimal overall network performance.

The data layer virtualizes the hardware infrastructure devices of each network segment using NFV technology, and forms virtual resources of computing resources, storage resources and communication resources based on the assistance of real-time operating systems, and finally forms a unified resource pool of the global network. It forwards information from the control plane through the southbound interface and forwards tasks to the infrastructure layer. NFV virtualizes the data plane and implements hardware functions of edge computing node devices in the software. Each virtual machine container can carry a resource type and will serve as an application under the intelligent orchestration of the controller. The joint scheduling of distributed parallel resource virtual machines will meet the specific needs of SAGIN edge computing. Virtualization technology realizes decoupling from the facility hardware, making resource management and control easier.

The control layer is mainly divided into space-based controllers (including MEO satellites and GEO satellites) and ground-based controllers. They are the core management layers of the resource service framework and are responsible for controlling and scheduling multiple resources in the data layer under global scheduling, as well as the generation of internal exchange paths and independent planning strategies of the network. In particular, the control layer can dynamically allocate various resources in the unified resource pool according to application requirements through the southbound interface. Computation offloading decisions and resource scheduling schemes can be generated in distributed or centralized modes. For example, according to the characteristics of application scenarios, the unified resource scheduling of the whole network can be completed by the space-based controller. In this case, in order to solve the high mobility problem of low-orbit satellites, a dynamic resource monitor needs to be added to the orchestrator. In this way, the monitor can grasp the link status, node resource utilization, terminal service requests and other information, and quickly complete resource registration and various operations.

In addition, SAGIN is highly dynamic and resource-limited, so network information is sent to the MEC server to immediately adjust its policies when the available resources change.

The service and application layer provides a wide range of applications to all users in the network domain, including global remote sensing, environmental monitoring, emergency rescue, intelligent processing, smart city, and so on. The edge computing technology is extended to the SAGIN, which results in its promotion and improvement for the user application service guarantee. Each type of application in SAGIN has its own resource requirements and service quality evaluation criteria. Therefore, the research goal of SAGIN edge computing is to maximize service coverage efficiency, optimize user service quality, and maximize service robustness by optimizing the scheduling strategy and allocating limited network resources [44,45]. However, emerging services and higher QoS requirements also pose greater challenges to the future development of SAGIN edge computing.

3.3. Characteristics

Compared with the mature application of edge computing in terrestrial mobile networks, edge computing in SAGIN is still in the research and exploration stage. However, it is considered to have new features that are incomparable on the ground, such as easy deployment, expansive space, and new services. Therefore, it is thought to be the most prospective technology to solve the new network challenges in the global interconnection period [46–49].

With regard to resource deployment, although terrestrial base stations can be configured with powerful computing servers and massive storage arrays at the edge of the cell, the deployment cost leads to limited density. In addition, once the deployment location is fixed, the resource service mode is also determined. However, this bottleneck can be alleviated by the flexible deployment characteristics of space/air resources. Considering users' requirements, the UAV edge resources can adjust and optimize the track to adapt to the application. The orbit edge resources can also adjust the satellite attitude to extend the service time or dynamically switch the power amplifier to obtain more bandwidth to improve the service quality.

With regard to spatial expansion, SAGIN expands to multiple physical spaces, such as the space, air, ground, and sea. The edge of the network, which has the characteristics of geomorphic independence, also breaks through the limitations of traditional physical distance and area and expands to the air and space. Efficient cross-domain interconnection ensures that in the scenario of ground communication blocking, space-based, and space-based access provides fast, flexible, and high-bandwidth access to meet user service quality requirements. Using the gridded low-orbit constellation coverage capability, SAGIN can almost realize node coordination and resource scheduling on a global scale.

With regard to emerging tasks, in the air-ground integration network application scenario, users' tasks will also generate new challenges. They will no longer be limited to the user's local business requests to provide services at the edge, but will also explore new tasks by utilizing wide-area coverage and seamless connectivity features, such as online information distribution of remote sensing data, and ocean-going ship status monitoring, etc. The expanded users' tasks still follow the essence of edge computing, that is, to provide services nearby, but it grows and supplements the scope of resource scheduling and service deployment.

3.4. Neural Network Progress

Neural networks (NN) have been used as a computational approach to simulate and solve engineering problems, and are widely studied in the field of MEC in SAGIN. NN predicts the desired property based on previous learning cycles or training [50]. Multiple factors need to be considered in SAGIN. Therefore, as shown in Figure 3, the data set is generally composed of the available energy, generated task, computation resource and other parameters of each device. The training dataset optimizes the weights of the interconnection between nodes so that the neural network has the ability to predict accurate output results for a given input set, such as whether to offload or not and the number of resources allocated to each device.

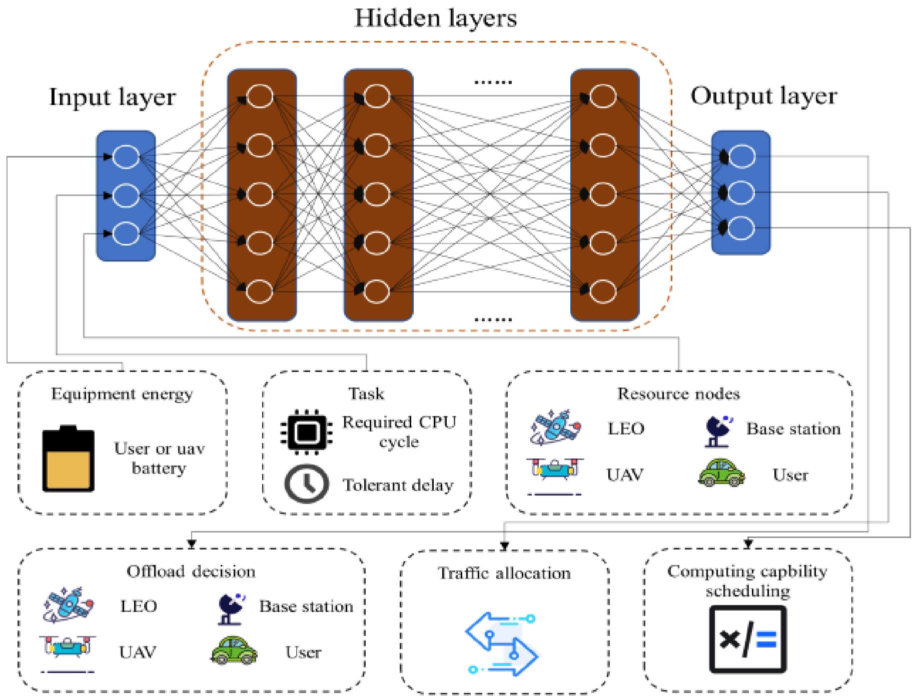

Figure 3. Neural network process flow in SAGIN.

Neurons or nodes are the essential processing elements of neural networks. In the mathematical model of a neuron, the connection weight represents the influence of synapses on the input signal in the form of a matrix, and the nonlinear characteristic of the neuron is represented by a transfer function. One of the essential steps in building the structure of a neural network is to select the best dataset for training the network.

The number of hidden layers and nodes in each hidden layer of the neural network model cannot be obtained directly [51]. There is no rule to estimate the exact number of hidden layers and nodes. Therefore, the number of hidden layers and nodes is essentially influenced by the network application. Since the optimization problem in the air-space integration network is usually complex, it can be solved more easily by using multiple hidden layers.

There are a variety of algorithms suitable for neural networks. The factors that determine which algorithm is most appropriate for a given problem are the complexity of the problem, the number of data training sets, the size of matrix weights and biases, the maximum error between actual data and neural network predictions, and the ability of neural networks to predict patterns (regression approximation) [52].

3.5. Advantages

The unique advantages of SAGIN edge computing technology can be analyzed from the following three aspects.

1. Solving the problem of "Ground edge service inaccessible"

In general, there are two scenarios in which the ground edge service is unreachable; one is that the ground edge cannot be connected to another ground edge due to the destruction of geological disasters, deserts, oceans, etc., under the condition of ground communication interruption, such as beyond the line of sight, and the edge service is suspended. The other is that the density of user is too low in remote areas. Even if the terrestrial edge resources can be deployed, the cost would be too high for practical implementation. Faced with the application scenarios where the above two types of edge services are unreachable, the introduction of space/air edge computing nodes has become the only option to solve the problem. Network architecture that incorporates space-based or air-based edge computing brings "Qualitative Change" to users. Traditional terrestrial networks have been unable to meet the global demands for "ubiquitous connection" because it is difficult to fully cover some complex terrains, such as mountains and oceans. Moreover, the terrestrial network infrastructure is vulnerable to damage by natural disasters, such as earthquakes and hurricanes, which can interrupt users' communications [53]. Edge computing-enhanced IOV is a network that relies heavily on connectivity and interaction between vehicles and transitional engineering infrastructure, and will break down in some remote areas that lack infrastructure, such as deserts, isolated islands and disaster areas. SAGIN solves this problem with ubiquitous links and global area coverage [54].

Cross-domain connectivity is a prominent advantage of the SAGIN. However, users can still choose to offload the satellite directly, or directly offload the UAV, remotely collect and forward the satellite by the UAV, or bring it back to the ground for processing, even if the ground network is not visible.

2. Solving the problem of "Single service pattern unoptimizable"

The development of network technology always seems to be driven by decreasing the user cost and improving service quality, and the two objectives are not opposed. The demand for traditional telecom services is not high. Even if the network communication bandwidth is insufficient, the execution of services on the local terminal still meets the requirements. Therefore, edge services do not demonstrate key capabilities in addressing users' demands. With the emergence of 5G emerging business applications, the user experience, such as delay, cannot be satisfied, and the computing power of the cloud center is sunk to the edge of the cell to satisfy users' requirements. Unfortunately, in the face of the rapid development of increasingly complex users' emerging applications, even a single edge resource such as the cell edge of a terrestrial cellular network can barely provide services, but it still cannot guarantee users' QoS due to its high cost (such as excessive delay or power consumption).

The introduction of space-air edge service to realize joint resource scheduling and node coordination will produce a "Quantitative Change" effect on the resource service. In Internet of vehicles (IoV), architectures only with terrestrial networks and existing resource allocation strategies may not provide satisfactory and real-time quality of service due to the deployment, coverage, and capacity issues of the roadside units. Therefore, it is necessary to use emerging technologies and improved resource allocation strategies to enhance users' experience. SAGIN (including satellite networks, aerial networks, and terrestrial networks) can tackle the problems of network coverage and data transmission in IoV. SAGIN in IoV can provide flexible and reliable services for vehicles by taking advantage of different networks [55]. For example, UAVs and satellites provide wirelessly powered edge computing and cloud computing services for user devices, respectively. The UAVs regularly fly on a predetermined orbit and hover at several locations within a fixed time to provide communication and computing services for the covered user devices, which can effectively reduce the delay of user devices. Unlike the discrete connectivity provided by UAVs, LEO satellites considered in the space layer can provide continuous services, which can effectively reduce the computational overhead of user devices [56].

For example, adopting SAGIN with MEC will lead to the following applications:

- For users with high cell users' density and saturated ground edge services, SAGIN edge servers will replace remote cloud computing centers to provide services with low latency, which is suitable for delay-sensitive service types;
- When local resources are limited or communication is blocked, space-based or space-based edge services are used to optimize user quality of service.

3. Solving the problem of "System optimization objective incompleteness".

Although introducing edge computing services improves system efficiency, the optimization objectives may not be complete due to different scenarios. This will also include the aspect of SAGIN's continuous enhancement of advantages. For example, terrestrial terminal users ignore the particular purpose of the application and do not consider the reliability, security, and fairness requirements. For example, the bottleneck of multi-service saturation of users occurs, and fairness under limited resources will be the primary issue that needs attention. For example, the traction of network security and data privacy issues must rely on the distributed deployment of SAGIN with MEC, which reduces the concentration of user information data in the network and the chance of data exposure, thereby reducing the risk of private data loss or leakage. There are also optimization objectives, such as service reliability and low error rate, which should also be the research focus. For example, MEC can provide abundant cloud-like services and resources in closer proximity to IoT devices. MEC can not only reduce communication latency but also mitigate the computational burden of aerial computing network, which serves as an effective method to optimize objectives, including reducing energy cost, time cost, and improving resource utilization for IoT devices [57].

Furthermore, the optimization objectives proposed by the traditional network architecture cannot be solved purely on the ground, or the user nodes themselves are in the air or space. Therefore, such users will focus on optimization objectives in special services, such as data fusion decisions, space-ground collaborative observation, and orbital mission planning. Usually, this can only be achieved by relying on network resources provided by multi-orbit satellite constellations or aerial clusters. In [58], a satellite edge computing (SEC) method with computing resources placed in low-Earth orbit satellite constellation is proposed to study how to achieve robust knowledge service coverage with limited resources.

To sum up, compared with the edge computing of terrestrial networks, SAGIN features can solve three-dimensional problems in which services may not be able to be accessed, may not be necessarily optimal when accessed, and may not be necessarily complete when optimal, which highlights its application advantages.

4. Key Technologies

4.1. Deployment of MEC

Mobile edge computing, with its key feature of providing computing, storage, communication, and other service capabilities to users nearby, meets increasingly complex application tasks and service quality requirements. However, how to effectively deploy edge servers in heterogeneous networks, flexibly and conveniently provide resource services, and achieve the balance between resources supply and demand of users is still a hot research issue in academia.

In various application scenarios of the SAGIN, there are MEC deployment modes with different network domains and edge numbers for multiple user needs and optimization objectives. Therefore, it can be found that the differentiated deployment schemes and effect characteristics will generate a multi-style network resource service framework, which affects system performance. In this section, the relevant applications of three application modes, i.e., single-edge computing, double-edge computing, and multi-edge computing in SAGIN, are combed and analyzed in detail.

4.1.1. Single MEC

UAV serves the edge nodes. Considering the application scenarios in remote areas, the work [59] has provided seamless communication coverage and computing services for power IoT (PIoT) devices. UAVs provide high transmission rates and sufficient power, with the support of edge servers and computing ability. A learning-based queue-aware task offloading and resource allocation algorithm is proposed to minimize the long-term average power consumption of all PIoT devices. The Lyapunov optimization method is used to decompose the joint optimization problem. The simulation results show that compared with the two existing algorithms, the energy consumption of this method is reduced by 22.36% and 23.13%, respectively. In the complex and changeable oceans, the service quality of marine IoT delay-sensitive applications with limited computing resources is challenging. The authors in [60] proposed a new type of marine-oriented SAGIN architecture with limited resources. Marine IoT devices can offload computing tasks to nearby UAVs equipped with edge servers, thus meeting the QoS requirements of computing-intensive marine IoT devices in high-traffic hotspots. The multi-armed bandit machine learning algorithm is introduced to optimize the edge offloading strategy, which can minimize the system energy consumption and delay. The work [61] enhances the computing power of remote IoT devices in a three-layer SAGIN through the MEC server deployed on the UAV. It proposed a random resource online algorithm that combines CPU cycle frequency, power control, and UAV trajectory planning, which maximizes the total computing rate of the system. Although UAVs that deploy MEC servers can provide broadband connections and sufficient computing power, edge services may also be unavailable due to their mobility and trajectory uncertainty. Therefore, in a seamlessly connected, stable, and visible SAGIN, users can also offload computing tasks to a remote cloud computing center through a low-orbit satellite backhaul link. This mode is most suitable for time-tolerance scenarios but has high computing power requirements.

LEO satellites have a wide coverage area and use their onboard computing capabilities to provide edge services (orbit edge computing or satellite edge computing), which has become the focus of non-terrestrial networks (NTN), and especially large-scale and broadband-connected ultra-dense LEO constellations will be more advantageous. Although the processing capacity is not as good as that of an excellent-performance ground base station, the communication payload and computing unit (edge server) equipped on the satellite can still provide fast and distributed computing services to the ground or air nodes. The work [62] considers constellation scenarios with 4 orbital parameters and deployment scales of 66, 180, 360, and 1584. It uses round-robin (RR), full offloading (FO), and fuzzy strategies to analyze and calculate the impact of delay, CPU utilization and bit rate caused by mission satellite offloading. The inclusion of a satellite communication network truly makes SAGIN edge computing possible. The layered space-ground collaborative network architecture of [63] consists of the following three network nodes: remote cloud center, satellite edge computing server, and data node. This paper takes advantage of satellite edge computing, and low-orbit satellite network communication to deal with the coordination problem of the parameter transfer process. An asynchronous adaptive collaborative aggregation algorithm (AFLS) is proposed, which achieves 95% accuracy in image classification scenarios when the nodes are in good collaboration. The satellite-based orbital edge computing framework is also developed with users' needs. The work [64] proposes a space-based cloud-fog computing architecture, which consists of an essential resource layer, a lightweight virtualization layer, and an edge service layer. Satellite or multi-satellite collaboration provides cluster service capabilities and supports business applications, such as computing offloading, task coordination, and content distribution, in space-based information network systems. In the 5G wireless communication scenario, the system architecture proposed in this paper can improve the primary network with 60%~70% throughput. In addition, considering the limited transmission power or energy of ground users, UAVs can also be used as data collection and forwarding nodes in SAGIN to offload computing tasks to satellite edge servers or base dtation (BS) for execution or to transfer the execution results back to ground users [65].

4.1.2. Double MEC

Although SAGIN directly provides edge services through cross-domain interconnection, the edge terminals such as ground-based MEC, air-based MEC, and space-based MEC still have limitations, such as high mobility, limited resources, and unstable connectivity. Moreover, the lower edge service is unavailable and even needs to be forwarded to the remote ground cloud computing center through the bend-pipe, thus affecting the QoS of users. Therefore, the pattern of deploying double edge computing nodes in different network domains has also been studied by many scholars.

As an essential supplement to space-based MEC or air-based MEC, in the SAGIN edge computing architecture, the computing and communication service resources of ground base stations are introduced to form a double-edge computing paradigm of space-ground and air-ground. In a wide-area complex environment, SAGIN also plays a crucial role in search and rescue [66,67]. To improve the execution efficiency of edge computing at sea under the constraints of latency and energy consumption, in complex marine search and rescue missions, the computationally intensive tasks generated by the IoT of sea surface sensors need to be offloaded to nearby edge servers, such as large ships or UAVs [66]. An intelligent task offloading algorithm based on reinforcement learning (RL) is proposed in [66]. The simulation results show that the algorithm has advantages in terms of time delay and energy efficiency.

Combining the advantages of edge computing technology and artificial intelligence technology, the authors in [68] designed a satellite-terrestrial double-edge intelligent computing system architecture, including terrestrial MEC servers and satellite MEC servers. Among them, the MEC server deployed in the cellular base station can sense the environment and context, cache popular local files, and process data and other intelligent management functions, while the MEC server as a satellite payload device also provides satellite edge smart management, senses global traffic and network conditions, caches popular files and on-orbit real-time processing. Simulations based on the Iridium constellation network show that the network architecture proposed in this paper performs better in cache distribution and task offloading. The work [69] has introduced edge computing technology in the high-speed space-ground network as an essential means to improve QoS. Based on the offloading location and satellite-ground link, computation offloading in SAGIN may have the following possibilities: no edge computing, proximal terrestrial offloading, satellite-borne offloading, and remote terrestrial offloading. A collaborative computing offload (CCO) model is proposed to realize parallel computing in the space-ground network. The simulation results show that the proposed model can significantly reduce the user-perceived delay and system energy consumption. Satellites also have multiple roles in double-edge computing architecture. For example, the work [70] has proven that satellites have both task relaying and on-board computing capabilities, and MEC servers deployed on low-orbit satellites can not only complete the offloading tasks of ground users by themselves, but it is also possible to transfer tasks to base stations in adjacent areas when resources are insufficient. That paper believes that the space-ground cooperation double-edge architecture has broad prospects.

However, ground base stations are usually unreachable in vast natural spaces, such as deserts and oceans, or emergency scenarios of ground communication blocking. Therefore, considering the UAV's characteristics of easy deployment and flexible access, as well as the advantages of extensive satellite coverage and broadband access, a space-aerial double-edge computing pattern is designed to improve the service efficiency of the system. As a result, the characteristics and application scenarios of space-based MEC and air-based MEC are different [71], and the double-edge will play a more significant role. The work [72] designed two new types of satellite and UAV frameworks in the space-ground network, namely intelligent enhanced satellites (ieSat) and intelligent enhanced UAVs (ieUAV). Satellites' seamless coverage characteristics with UAVs' enhanced processing capabilities provides edge computing power for IoT devices in complex ground networks. That paper designs the following three different types of application scenarios and double-edge collaborative architecture in detail: (1) long-term wide-area surveillance-type delay-sensitive IoT tasks, in which data collection is completed by flexible UAVs and forwarded to satellites to complete the task offloading; (2) computing-intensive IoT tasks in sudden hotspot areas, and delay-sensitive and delay-tolerant offloading by neighboring UAVs and neighboring satellites, respectively; (3) delay-tolerant tasks for highly reliable and secure networks and the aviation network composed of dense UAVs to realize fast-response collection and calculation and to provide edge services that meet mission requirements.

The authors in [73] designed a new satellite-UAV double-edge architecture based on a time-slot system. To minimize the average task processing delay within a specific period, the delay was subdivided into transmission, wait, and computation, and the optimization problem was modeled as Markov decision process (MDP). However, due to the dynamics and complexity of the multi-user generation task on the ground per slot, an algorithm based on curriculum reinforcement learning was proposed to obtain the optimal offloading strategy. The extensive simulation results show that the algorithm can effectively utilize the satellite-UAV edge resources. The work [74] expands edge computing nodes to HAPs and LEO satellites in SAGIN. High-altitude platforms and low-orbit satellites, equipped with on-board computing resources, provide edge computing services for ground users' devices. The edge node has multi-channel antennas to realize communication offloading, and its performance is better than the single-antenna transmission method to improve the offload efficiency. To minimize the total weighted energy consumption of the system, this paper proposes a joint scheduling algorithm of joint ground equipment allocation, multi-user multiple-input multiple-output (MU-MIMO) transmission precoding, computing task allocation, and resource allocation. However, that paper only considers single satellite and static offloading scenarios.

4.1.3. Multi MEC

Although the static deployment of MEC in fixed or dual specific network domains can enrich computing resources and improve service quality, due to the rapid growth of user equipment and ubiquitous application requirements in the SAGIN heterogeneous space with huge physical boundaries and huge information capacity, it also makes the ubiquity and robustness of edge services critical. To alleviate the co-channel interference caused by the multi-user multi-edge irregular topology sharing spectrum in the marine network of SAGIN, the work [75] proposes a system that can not only sense the spectrum state, but also identify the ship's position information in the cognitive framework to optimize the inter-cluster/intra-cluster power allocation strategy and obtain considerable benefits. The authors in [76] designed a collaborative service framework with the following three modes: fine-grained, medium-grained, and coarse-grained. In the scenario of space-ground ubiquitous edge server deployment, a service coordination method based on pre-selection and threshold update was proposed to achieve low effectiveness of cost reduction in service delay, which was verified by an emergency search and rescue example. Although ubiquitous multi-edge computing provides services for universal applications, the complex network environment poses severe challenges to essential equipment, such as communication terminals and antennas. Therefore, the authors in [77] introduced the reconfigurable intelligent surface (RIS) antenna to enhance wireless coverage and improve the wireless communication environment, transmission, and computing quality. In the network scenario of cloud-edge collaboration, the work [78] deploys ubiquitous edge services in big cyberspace composed of heterogeneous subnets, such as space, air, ground, and sea. It provides a personalized intelligent network service through brilliant orchestration and task-oriented networking methods. The work [55] builds an edge cloud computing cluster based on SDN and NFV, providing edge services that can be reached anywhere and anytime. In addition, this paper proposes an improved Two_Arch2, algorithm using an angle-based diversity measurement strategy to meet application requirements in the high security and real-time performance of the Internet of vehicles environment.

Table 1 provides a summary of MEC deployment. The analysis shows that from the relative logical relationship between service requests and resource provision, any node or combination of nodes in SAGIN may provide edge services in specific scenarios. Edge deployment solutions can meet differentiated user service requirements.

Table 1. Summary of related works on MEC deployment.

Deployment Type		Objective	Considered Factors	Advantages	Disadvantages	Ref.
Single MEC	UAV	Minimizing the latency and energy consumption	Complex marine environment	Flexible deployment in hot spots	Lower height, less resource	[60]
		Maximizing computing performance	Frequency division duplexing, CPU cycle, power control, UAV trajectory, joint stochastic scheme	Wide coverage for remote IoT	Limited computation capacities, high mobility	[61]
	Satellite	Better exploit the overall available distributed resources	Orbital edge offloading, mega-LEO satellite constellations	Better exploitation, more homogeneous distribution	More complex among different layers, long propagation delay	[62]
		Minimize the long-term delay of all tasks	UAV collect and relay, task scheduling	Wider coverage, low delay under energy constraints	Path loss, high mobility	[65]
Double MEC	UAV-BS	Minimizing energy and delay consumption	Maritime IoT, intelligent task offloading	Low delay, flexible deployment	Unstable link, finite resouces, intermittent service	[66]
		Minimizing power consumption	SAGIN resource allocation	Seamless coverage, high rate	Dynamic topology, uncertainty link	[67]
	Satellite-BS	Minimizing delay and energy, maximizing efficiency	Satellite MEC, high-speed network	Multi cooperative computation offloading	Not suitable for high cost task	[69]
		Minimizing the completion delay of all users' tasks	LEO edge server, or BS server sent by LEO	More flexible edge decision, global coverage	Rare inter-satellite cooperation	[70]
	Satellite-UAV	Enhancing edge service capbility	Intelligent-enhanced UAV, intelligent enhanced satellites	Flexible offloading options, seamless global coverage	High dynamic, unstable interactionlink	[72]
		Minimizing delay of all tasks	Propagation time, transmit time, compute time	High data rate, less delay	Finite energy	[73]
		Minimizing the weighted sum energy consumption	Transmit precoding, task assignment, resource allocation	MIMO, more users, higher efficiency	Doppler effects	[74]
Multi-MEC		Minimizing overall service delay and cost	Service coordination	Flexibly integrates and manages services	High allocation complexity	[76]
		Maximizing data rate	SAG MEC, double benefits of comp. and comm.	More flexible offloading options	More security risks	[77]
		Four kind of objectives	Edge–cloud resource scheduling, SDN, NFV	Rich resources, joint optimization	Higher optimization difficulty	[55]

4.1.4. Offloading Schemes

In addition to providing global ubiquitous connectivity for mobile users, SAGIN also provides a variety of computing services. Usually, mobile users can offload computing tasks to other nodes with abundant computing resources for processing to compensate for the limited computing and storage resources of mobile user devices [79]. In general, the most important issue of mobile edge computing is how to decide task offloading, i.e., whether to offload, how much to offload, and the offloading destination. For offload destinations, the distance between cloud data center and mobile users is often far apart, and cloud center forwarding via satellites can be achieved over the ultra-visible distance through SAGIN, regardless of geographical restrictions. The downside is the large latency, which leads to the processing of mobile services needing to experience large latency, making it difficult to meet the needs of emerging applications for end-to-end latency down to the millisecond level.

The destination of task offloading can be local devices, edge servers, or cloud centers. First of all, the task can be executed on the local terminal device without offloading to the edge or cloud. However, this is limited by the energy consumption and computing capability of the device, which has many constraints in practical applications. Second, tasks can be considered offloaded to edge (e.g., UAV, LEO) execution, which is the focus of this article. This section describes where edge servers can be deployed in the SAGIN. Thirdly, when the edge resources still cannot meet the user's QoS, the tasks can also be offloaded to the cloud. However, this paper does not focus on how the cloud data center performs computation tasks. It is also worth mentioning that since the SAGIN is not restricted by a region, it can also use satellites, UAVs or HAPs to relay tasks to the cloud center for execution [80–82].

In terms of the single task offloading ratio, the offloading scheme can be 0/1 pattern or partial pattern. In order to simplify the model of computation task offloading, the strategy of 0/1 offloading is usually considered in SAGIN's application scenario [83,84]. In addition, to improve the efficiency of task execution, sometimes the task can be arbitrarily partitioned into serval parts, where the data are bit-wise independent and can also be partially offloaded to different edge servers for simultaneous execution [85,86]. For example, Yu et al. in [54] divide the computational task into fine-grained slices, according to constraints on delay and energy consumption. However, it also leads to the process of task reconfiguration and migration, which increases the system complexity. The multi-task offloading schemes can be divided into all-local execution, or all-edge execution, or optimization offloading decisions based on the channel quality, resource constraints, users' QoS and other variables, or random offloading.

4.2. Network Resources Services

Due to advanced communication technologies, such as interface protocol conversion, dynamic sensing routing, and multi-beam phased array antennas, accessing and efficiently interconnecting nodes in different domains in SAGIN is possible. However, physical and electrical performance constraints, such as weight and power consumption, result in additional resource capabilities. To meet the business and service requirements of user equipment in an unbalanced and unequal network resource environment, a SAGIN edge computing technology is proposed to evaluate the resource level of network nodes, schedule global network resources, respond to user application requests, and provide services nearby to realize the efficient coordination of the entire network nodes in the SAGIN. Network nodes are loaded with various resource capabilities, such as detection sensors, signal processing, transmission, store-and-forward, etc., to serve user nodes under unified scheduling control. Furthermore, there are particular types of network resources. For example, in [61], UAVs provide computing offloading, while wirelessly transmitting energy to ground remote IoT devices through radio frequency energy transmitters. As is the case in most edge computing research, this paper focuses on the scheduling and offloading of three types of network resources, including computing, communication, and storage between nodes.

4.2.1. Computation Offloading

The task offloading of computation and signal processing is the most common edge-end node collaboration scenario. Based on the differentiated system models in SAGIN, the computation offloading problems are optimized to satisfy different QoS, such as scheduling virtualized network resources to minimize energy consumption [81]. As a result, many scholars have conducted in-depth research on edge computing tasks of typical computation offload. In this paper, it is summarized as a "4W1H" five-dimensional computation offloading decision problem of "Whether to offload," "Where to offload," "How many to offload," "When to offload," and "in What order."

Different domain nodes of SAGIN have potential requirements for offloading computing tasks. Chen et al. in [79] use UAVs as the computational offloading decision-making object. UAVs are deployed flexibly and are not restricted by ground traffic. Ground IoT data can also be collected in remote areas without cellular networks. However, due to their limited computing power, appropriate offloading decision-making is required. The computing tasks are offloaded to the ground BS or sent back to the cloud center through satellites. To avoid unnecessary system overhead or higher task abandonment rates, the work [79] proposes a distributed robust delay optimization algorithm, which gives decisions such as where to offload and the proportion of offloaded tasks to make sure that the delay in the expected energy constraint system is minimized under the worst-case probability distribution.

Based on satellite IoT, Wei et al. in [87] proposed a novel inter-satellite edge computing offloading architecture, which uses remote sensing satellites integrated with visible light camera loads to perform on-orbit real-time image processing as a scene. To solve the data transmission bottleneck caused by the high delay, due to satellite-to-ground transmission and the surge of raw data, the need for real-time processing of remote sensing satellite images in orbit is increasingly urgent [88,89]. However, in-orbit processing, such as identification, classification, and reasoning of payload raw data based on deep learning-based satellite edge intelligent computing, usually requires gigabit floating point arithmetic or even higher computing power, in which single-satellite resources cannot cover. The work [87] chooses to deploy the deep learning framework on the satellite IoT cloud node cluster with more robust computing power. It selects part of the processing or offloads to nearby satellites for cooperative processing according to their status. Then, the authors present the performance of different neural network models in the satellite edge intelligent computing architecture. The simulation shows that the lightweight neural network model is more suitable for satellite IoT scenarios.

Considering the time slot allocation system with different priorities, Wang et al. in [90] introduced a time slot allocation scheme based on customized service priority in the space-ground collaborative double-edge computing architecture and analyzed the time slot allocation of computing task offloading. By showing the performance of co-simulation of three-level time slot allocation and three-priority service scheduling in a 66-satellite constellation and ground coordination, the results show that the new network structure and offloading scheme can effectively improve the service efficiency and reduce service delay.

4.2.2. Communication Traffic Offloading

Traffic offloading and network capacity allocation are essential forms of communication resource scheduling for SAGIN to provide edge services. It is common in end-to-end inter-satellite and satellite-to-BS scenarios or UAV relay and forwarding scenarios. Finally, it realizes ultra-reliable communication (URC) [91], energy efficiency [92], reduced packet loss, and enhanced communication in wireless networks with low latency [93,94].

Unlike the basic offloading strategy adopted in traditional fixed communication state networks, the traffic offloading workflow cannot guarantee the bandwidth continuity of forwarding hops and drop hops in SAGIN with high mobility, resulting in potential risks of traffic accumulation. Furthermore, there is no guarantee that offloaded targets will remain visible within the communication range, so a traffic offloading model with state intelligence awareness must be considered. Tang et al. in [95] proposed a traffic offloading method based on the double Q-learning algorithm and an improved delay-sensitive replay memory algorithm (DSRPM), which considers the high dynamics of network nodes and frequently changing network traffic and link states. Local and adjacent historical information continuously trains nodes to decide on traffic offloading strategies, and the algorithm obtains advantages in signaling overhead, dynamic adaptation, packet loss rate, and transmission delay.

Due to the extended distance from the base station, Lyu et al. in [96] proposed an algorithm for offloading cellular network traffic to solve the performance bottleneck of mobile terminals at the edge of traditional terrestrial cellular systems in the air along the cell edge by using UAV. The work [96] jointly optimizes the UAV flight trajectory, the bandwidth allocation between the UAV and the base station, and the user division to maximize the minimum throughput of all mobile terminals in a single cell to achieve the most significant degree of users' fairness. The simulations show that this hybrid network with optimized spectrum sharing and cyclic multiple access design significantly improves communication throughput. Combining space-based UAV-assisted communication with MEC is a prospective paradigm for enhancing space communication. To cope with the surge in application data in the UAV-assisted IoT, Guo et al. in [97] designed a joint optimization problem of QoS and energy consumption, in which the block coordinate descent method and successive convex approximation techniques effectively improve the network's overall energy consumption and communication performance.

Although satellite network construction has been accelerated with the rapid development of the reusable space transportation system and mass satellite manufacturing technology, inter-satellite communication resources are still precious, meaning that the transmission capacity is still limited. Therefore, it is necessary to improve network system performance through inter-satellite capacity allocation and scheduling. Jiang et al. in [98] proposed a low-complexity inter-satellite capacity calculation method based on time-expansion graphs in the three-layer heterogeneous satellite network model of the high, medium and low orbits. According to the communication needs of users, the capacity is allocated among satellite users. Then, the authors design a long-term optimal capacity allocation algorithm based on Q-learning to improve the long-term utility of inter-satellite capacity allocation. At the same time, to obtain better system performance, satellites are usually equipped with multiple antennas to cover omni-directional space to achieve inner-orbital and inter-orbital communication, which is considered as a new trend.

4.2.3. Cache Resource Distribution

Content caching and file distribution at the edge nodes of SAGIN are prospective technologies that can effectively reduce data traffic and improve user experience. In particular, remote sensing satellites use their wide-area coverage characteristics to broadcast sensor detection result cache far beyond the line-of-sight to end users on the ground or in the air, which can realize the user's global awareness. Mobile edge caching can improve the quality of service for end users on this basis. Extensive data analysis shows that in the case of limited cache size, active caching can provide 100% user satisfaction, while offloading 98% of backhaul traffic [99]. Mobile edge caching will also reduce the traffic load on the backhaul link. Chen et al. in [100] propose a multi-base station agent cooperative edge caching algorithm based on deep reinforcement learning, in which nodes make caching decisions based on local and global states. This algorithm improves the cooperation between edge caches and the hit rate of edge caches.

Li et al. in [101] introduced LEO satellites with storage capabilities into the radio access network (RAN) and proposed a cooperative cache distribution architecture to respond to user requests. In [101], a request of file user and satellite/access point (AP) distribution model based on Zipf probability distribution is designed in detail. Considering the limitations of satellite energy, a nonlinear fractional programming problem for the joint optimization objective of traffic offloading and energy efficiency is proposed. Numerical simulations show that the space-ground coordinated data distribution scheme significantly improves energy efficiency.

The distributed content in the satellite-to-ground cache distribution scenario is proposed in [102] and includes IoT gateway configuration information, new docker applications, and edge computing scripts. Since satellite bandwidth resources are limited, to save space-air-ground communication resources and to reduce frequent communication between satellites and gateways, ground nodes are clustered based on the idea of Node2Vec, and ground nodes communicate with satellites to obtain tasks or files based on the sorting results. The clustered edge cloud dispatching task collaboration method can effectively improve the data transmission efficiency by comparing the performance of four groups of test data with different task sizes. This advantage will become more evident as the number of nodes increases.

Jia et al. [103] proposed a file distribution architecture to collaborate between LEO satellites and the high-altitude platform HAP to achieve full coverage and to provide data services in remote areas. HAP collects data and forwards it to satellites, and satellites distribute data to the ground data center through inter-satellite links or by themselves. A Benders decomposition optimization algorithm is proposed to solve the mixed integer linear programming problem based on the time expanding graph (TEG) joint optimization, with limited complexity to maximize the total distribution of LEO to the ground data center. The simulation results of the suboptimal solution obtained by this accelerated algorithm show that the method proposed in this paper has obvious advantages in the total data performance under various network parameters.

The space-ground edge cache system should consider reliability issues, such as unstable transmission under high dynamics or data errors under complex link conditions. Gu et al. in [104] proposed a SAGIN mobile edge caching IoT system composed of satellites and UAVs. The satellites distribute the cached data to the ground IoT storage sensors. In order to reduce the potential risk of loss or error of unstable ground sensors when receiving broadcast data from LEO satellites, fault-tolerant coding with intelligent optimization parameters is adopted to improve data reliability. Compared with maximum distance separable (MDS) codes and regeneration codes, the adaptive minimum memory regeneration (AMSR) code proposed in this paper can significantly reduce the system's total communication cost and maintain the system's availability.

4.2.4. Joint Resource Service

Similar to the ground system using joint resource scheduling to better adapt to the growing demand for data caching and computing services [105], user nodes in SAGIN usually have computing offloading and cache acquisition, network communication, and other task requirements to improve the business experience. Therefore, the SAGIN edge nodes can use 3C (computing, caching, communication) combination resources to realize joint scheduling optimization and converged services.

SAGIN's large information throughput, scattered users, and complex resource requirements led to a novel multi-resource management system architecture [106]. Different from other scholars who focus on maximizing system throughput, the objective of [106] is balancing multi-user fairness and improving data security, based on the three-step workflow of data perception acquisition, block-chain computing, and wireless transmission between the satellites and the ground. A 3C allocation joint optimization model of the Nash negotiation game is proposed and solved by dual decomposition. A large number of numerical simulations verify the performance advantages of the proposed system architecture. Fu et al. in [107] also pay attention to the integrated management of 3C resources caused by the surge of network data traffic, which considers the intermittent periodic characteristics and network dynamics of satellite communication. In [107], the downlink relay forwarding model of a terrestrial relay (TR) and air relay (AR) is introduced, and the optimization problem of minimizing transmission power consumption is proposed. Many simulation results verify the improvement of system throughput performance and the joint scheduling effect of 3C resources. Performance and multi-user experience are also significantly improved.

The future objective of 6G advanced networks is to build a seamlessly integrated network to fill the large coverage gap on a global scale and to solve the bottleneck problem of mission-critical services [108]. Therefore, 6G networks are also committed to jointly optimizing communication, computing, and cache of multiple types of resources. The authors of this paper design a service-centric ultra-reliable and low-latency edge intelligent architecture for mission-critical scenarios with diverse requirements and propose several critical technologies for efficient 3C resource scheduling. The service decomposition strategy of dividing each service into multiple micro-services for distributed processing can significantly improve the system's resource efficiency and experience quality. To solve the high dynamics of network topology, a multi-semantic addressing method is adopted to decouple users, content, and resources from the network topology, which reduces the delay and reliability challenges brought by traditional addressing methods. Finally, considering the whole life cycle of services, a service-centric 3C resource scheduling method is proposed. A dynamic adaptive resource scheduling framework based on a knowledge graph and real-time monitoring is designed, which utilizes the diverse computing and storage resources at the edge of the network to support critical services and improve the three challenges of heterogeneity, time variability, and reliability.

Table 2 provides a summary of the service resources provided by the network edge. Common types of edge resources include (1) computing resources, (2) communication re-sources, and (3) cache resources. In addition, the joint service of multiple resources is of-ten investigated. Different papers have analyzed the performance improvement of the SAGIN brought by the differentiated service resources provided by the edge server, including network communication enhancement, high system reliability, service guarantee, etc. Since the target of the scenario we are dealing with is the "nearby service", this article focuses on edge side resource services rather than remote cloud center resources.

Table 2. Summary of related works on network resources services.

Resource Types	Objective	Key Issues	Advantages	Disadvantages	Ref.
Computation offloading	Minimizing system delay under energy constraint	Data-driven approach describes the uncertainty of task arrival	High robustness of application completion	High energy consumption	[79]
	Minimizing inter-satellites process delay	Distributed intelligent on-board computing	On-board realtime image process	High inter-satellite link dependence	[87]
	Maximizing resources allocation efficiency	Double-edge, customized service priority	High space-ground services efficiency	Uncertainty of connectivity	[90]
Communication traffic offloading	Improving end-to-end energy efficiency	Stochastic geometry, interference/no interference scenario	Cross-domain communication enhacing	Large path loss	[93]
	Maximizing the network transmission rate	Dual timeslot cooperative communication scheme	High space-ground signal quality	Complex inter-node interference	[94]
	Maximizing all users'throughput	Double Q-learning traffic offloading	Better load balancing capability	Higher link dynamic	[95]
	Minimizing overall energy consumption	Joint optimization of QoS and energy consumption	Higher overall network performance	Multi-layer unbalancedness	[97]

Table 2. *Cont.*

Resource Types	Objective	Key Issues	Advantages	Disadvantages	Ref.
Cache distribution	Minimizing the cost of acquiring content	Cooperative content sharing, multi-agent data exchange	Higher data utilization of MEC	Excess communication load	[100]
	Maximizing system content availability	Caching system with fault-tolerant codes	Higher data reliability	Higher network cost	[104]
Joint resource service	maximizing joint objectives	Information-centric virtualized resources	Higher network overall utility	More optimization constraints	[105]
	Maximizing link time, minimizing energy cost	Block-chain, data security, double MEC	Higher throughput fairness	More difficult optimizations	[107]
	Improving the service's reliability	Adaptive resource scheduling framework, mission-critical services	More complete service guarantee	More complex channel effects	[108]

4.3. Edge Intelligence

Another kind of important computation resources in networks is artificial intelligence (AI), which is now developing rapidly due to recent advances in models, algorithms, processing, power, and big data. AI has made substantial breakthroughs in a wide spectrum of applications, ranging from computer vision, natural language processing to automatic driving and robotics. It is widely recognized that these intelligent applications are significantly enriching people's lifestyle, as well as science. In the traditional architecture for AI, the input data generated by mobile devices are sent to the cloud for processing, and results are then sent back to mobile devices. However, with such a cloud-centric approach, large amounts of data are uploaded to the remote cloud via a long wide-area network data transmission, resulting in high end-to-end latency and energy consumption of the mobile devices. Considering the necessity of quick analysis, there exists a strong demand to integrate AI and edge computing, which gives rise to edge intelligence (EI). With a large number of concepts and technologies interwoven together, the subject of EI is enormously sophisticated. Generally, EI could be distinguished into AI for edge and AI on edge [66,108–110].

4.3.1. AI for Edge

AI for edge aims to provide a better solution to optimization problems in edge computing with the help of AI. A typical example about AI for edge is edge caching.

With the rise of various smart terminal devices, services such as multimedia applications, mobile games, and social applications have also grown rapidly. This trend brings a new characteristic that the same content is often repeatedly requested by devices in the same area, which leads to the demand for edge caching. The core idea of edge caching is to cache and reuse the task results at the network edge, reducing the querying latency of EI application. Two problems are important in edge caching [111]. On the one hand, the distribution of popular content within the coverage of edge nodes is difficult to estimate and may vary with space and time. On the other hand, the complex heterogeneous network makes the design of content caching strategy more difficult. Applying deep learning to accurately predict the future popularity of cached contents is now the focus of edge caching. Deep neural networks (DNN) consist of an encoder for data regularization and a hidden layer behind it, which can be trained and deployed with solutions generated by optimal or heuristic algorithms to determine the caching policy, and thus avoid online optimization iterations [112,113]. Recurrent neural networks (RNN) can predict the user's preference. Based on the preference, the contents are then prefetched and cached in advance to each predicted edge node at each predicted location. Deep reinforcement learning (DRL) can learn key features from raw observations and can optimize cache management policies for EC networks directly from high-dimensional observations [99,114,115].

4.3.2. AI on Edge

AI on edge refers to the circumstance of running AI models on edge, which performs training and inference of AI models with device-edge-cloud synergy to extract insights from massive data. AI on edge could be divided into model training, model inference and processor acceleration [40].

The frameworks of model training include federated training (FL) and knowledge distillation. FL is a practical deep learning training mechanism between the end-edge-cloud [116]. In the framework of FL, mobile devices are considered as clients performing local training. Meanwhile, end devices, edge nodes and servers in the cloud can be equivalently considered as clients in FL under certain conditions. FL does not need to upload data to the central cloud for training, and edge devices only need to train local models using local data and then upload the updated model parameters. There are usually three types of training approaches in FL, which are as follows: (1) end-edge cooperation, where the edge nodes replace the cloud as the server and the end side as the client; (2) edge-cloud cooperation, where the edge side participates in FL as the client, while the cloud acts as the aggregation server; (3) end-edge-cloud cooperation, where both sides of the end side participate in FL as the client, while the cloud acts as the aggregation server, and this approach can combine the advantages of the former two approaches [117]. Another important strategy in model training is knowledge distillation, which is a method of transferring knowledge from complex AI models to compact AI models [118]. In general, complex AI models are powerful, while compact AI models are more flexible and efficient. Knowledge distillation can use a complex AI model to train a compact AI model with similar performance to a complex AI model. These methods can be applied to different types of DNNs or combinations to optimize complex edge AI models [119].

The frameworks of model inference include model optimization and model splitting [110]. Model optimization is used to handle AI tasks that require a large memory footprint. At the edge, there are not enough resources to support raw large-scale AI models. Optimizing AI models and quantifying their weights can reduce resource costs. Some researchers have optimized AI models for parameter pruning and sharing, caching intermediate data between adjacent layers to reduce data movement [120]. Another way to optimize the parameters is to design specially structured convolutional filters, which is only applicable to convolutional layers. Model splitting can decompose a large number of computational tasks into different parts, and different devices can work together to solve the problem. One of the most commonly used segmentation methods is to segment AI models horizontally, i.e., along the end-edge-cloud. The process of data analysis is usually divided into two parts, one is processed at the edge while the other one is processed in the cloud, reducing the network traffic between the edge and the cloud [121]. Another model partitioning method is vertical partitioning, especially for CNNs. As opposed to horizontal partitioning, vertical partitioning fuses the layers and partitions them vertically in a grid fashion, dividing each CNN layer into independently distributed computational tasks.

4.4. Optimization Objective

In the SAGIN cross-domain network system, even if the network resource types of MEC service deployment and scheduling are the same, due to the difference in QoS requirements and working modes of user nodes, different optimization objectives will still be pursued. For example, in time-sensitive task scenarios with relatively sufficient computing power and energy, application services pay more attention to the impact of delay characteristics on system tasks and usually regard minimizing delay as the optimization objective of edge computing.

This section mainly focuses on optimization objectives such as power consumption, delay, and their combined states and analyzes their application scenarios and constraints in detail. At the same time, with the continuous expansion of the network scale, the network environment becomes more complex, and network security has gradually become the focus of many researchers, which is more critical for cross-domain SAGIN. Therefore, this section will also provide the research progress of particular optimization objectives, such as privacy communication, security perception, and anti-jamming.

4.4.1. Minimized Energy Consumption

With the proliferation of IoT devices, many computing-intensive applications are bound to be generated. Despite the rapid development of lightweight energy technologies, high-efficiency energy utilization is still required to prolong the run time of instruments, due to the limitations of battery capacity and device weight [122,123]. Meanwhile, SAGIN has a vast physical space and a relatively long communication distance, which brings additional transmission power consumption, posing a severe challenge to minimize system power consumption. In terms of optimization objects, the power consumption minimization of different roles of the network system may be a concern for different application scenarios. For example, the ground service offloading task mainly pursues the lowest energy consumption of ground IoT or terminal equipment. At the same time, it involves air-based data collection, data-forwarding, and traffic offloading tasks focused on obtaining the optimal UAV power consumption and performance. Some scholars have studied the lowest sum of computing and transmission energy consumption in terms of quantitative values. At the same time, the work [101] takes the ratio of data throughput to total power consumption as the optimization target to highlight the efficiency.

In the scenario of air-borne MEC, to reduce energy consumption, the computationally intensive tasks of ground mobile user terminals can be offloaded to UAV or BS for execution. Lu et al. in [124] proposed a joint optimization problem to minimize the computational tasks of UAVs and terminals, as well as the energy consumption caused by sending data. Considering the time-varying, random ground channels, and line-of-sight air-ground channels, a weighted robust iterative optimization method, combined with the mean square error method and S-process, is proposed to solve this problem. The numerical results show that, by adjusting and optimizing power allocation, CPU frequency, and offloading data volume, the algorithm can effectively reduce energy consumption under computationally intensive tasks compared to other schemes and is more conducive to air-ground collaborative MEC. In addition, compared with the non-robust algorithm, the algorithm can reduce the energy consumption of the combined air-ground system more stably.

Jia et al. in [125] proposed a SAGIN scenario based on UAV sampling carrying or offloading satellite backhaul. However, since the UAV's energy capacity is minimal, while its flight and data transmission require a lot of energy, they focused on minimizing the total energy consumption of the UAV, while meeting the needs of ground equipment and various constraints. Then, the authors divided the total energy consumption into take-off power consumption, flight propulsion power consumption, descent power consumption, and power consumption for sending data to satellites. Finally, a joint optimization problem of typical UAV trajectory and data transmission is proposed to solve this problem efficiently. The algorithm's effectiveness, shown by the numerical simulation, verifies the LEO-assisted UAV trajectory design and data transmission advantages.

The SAGIN hybrid communication network enhances coverage by deploying flexible and maneuverable UAVs, which shows excellent prospects in random communication. However, UAV-assisted air edges are not self-contained, relying on existing satellite/terrestrial systems for spectrum sharing and efficient backhaul. Li et al. in [126] introduced a dedicated automatic identification system, denoted as AIS, to obtain ship position information as a priori knowledge, so that only the large-scale channel state information (CSI) related to the position is considered available. In [126], the joint optimization problem of UAV flight trajectory and in-flight transmit power was proposed. The numerical solution of the UAV data rate proportional to share passion is carried out through problem decomposition and successive convex optimization. The simulation results show that the algorithm will obtain better UAV system performance under total energy limitations. The UAV can achieve better mobile user service according to the optimized trajectory and launch power.

Considering the importance of a good resource allocation strategy for solving the contradiction between the sudden increase in the needs of disaster victims and the shortage of wireless resources in emergencies, the space-air-ground heterogeneous system is considered as a prospective candidate system to meet the communication needs of emergency rescue. Moreover, the QoS requirements of users in emergency scenarios are fully assembled, and energy efficiency (EE) is a better design criterion for evaluating the performance of hybrid networks.

The work [127] has studied the problem of maximizing energy efficiency (the ratio of system communication rate to energy consumption) in an orthogonal frequency-division multiple access (OFDMA)-based emergency communication SAGIN hybrid cooperative network. In [127], the problem of joint relay selection and power allocation is studied under various total power constraints, quality of service requirements, and backhaul capacity constraints. To solve this complex optimization problem, relaxation of binary variables and dual decomposition methods are used. The simulation results illustrate the impact of total power constraints and backhaul capacity on energy efficiency and system capacity. In addition, the choice of the relay will also affect the system's performance.

4.4.2. Minimizing the Delay

In disaster-stricken areas, e.g., medical IoT, autonomous driving intelligent computing, 4K/8K high-definition video transmission, forest fire monitoring, etc. [128,129], most emergency and intelligent application scenarios are delay-sensitive, and it is necessary to propose optimal target solutions with minimum delays. The time of computing task at the edge of the SAGIN is mainly divided into the following four types: computing processing time, propagation time, transmission time, and waiting for non-value-creating time (queue, slot gap, and so on). To analyze the influence of different time types on the system in detail, many scholars have conducted optimization research on various time types. For example, in [130], the time for coordinated offloading of the space-ground network is subdivided into nine segments for evaluation. The inter-satellite propagation delay is regarded as the main communication delay in the highly dynamic satellite network system. Some pioneering routing mechanisms ensure the minimum end-to-end transmission delay between satellites [131]. The intelligent edge computing of satellite networks will also exert on-orbit real-time processing capabilities. For example, in the task of satellite Internet of things remote sensing image target detection, the deep learning method proposed in [109] can reduce the delay of satellite image acquisition and target detection, thereby making it easier to shorten the processing delay.

Mao et al. in [132] proposed a cloud-edge collaboration SAGIN framework, which uses the flexible deployment of UAVs and wide-area satellite coverage characteristics to give full play to the computing resources of each node in a heterogeneous network to meet the needs of delay-sensitive emergency applications. In [132], the time models of local execution, UAV offloading, and satellite offloading are designed in detail. Under the constraints of power consumption and maximum tolerance time, the problem of minimizing the time overhead of the top delay user in the ground terminal is derived. Descent and successive convex approximation, an alternating optimization algorithm with a convergence guarantee, is proposed to solve this problem, and the simulations show that this algorithm can significantly reduce the delay.

By deploying edge computing servers at satellite and terrestrial gateways, efficient offloading of compute-intensive tasks is achieved, resulting in significant performance improvements in network computing power, at the expense of complex resource management and high mobility. In [133], a weighting coefficient is defined for the delay of each user to ensure the fairness of multiple users. Latency is subdivided as calculations that can be performed locally or offloaded to satellite or terrestrial gateways. Then, the authors proposed an approximate optimal solution based on game theory and matching theory. The numerical results show that this method can obtain almost the same weights and delay as the Brute-force method at the cost of lower complexity.

To improve the parallel ability of multi-user utilization of edge resources in remote areas, a specific virtual machine is deployed for each computing task in the UAV edge server scenario [134]. The computing resources of the UAV are virtualized as virtual machines, and each virtual machine is used for a specific application [135]. Quan et al. in [135] propose a joint resource allocation and task scheduling method to allocate computing resources to virtual machines in order to pursue better latency performance. In [135], the offloading decision is formulated as a Markov decision process that considers the network dynamics in the system state. The computational offloading method based on deep reinforcement learning is used to learn the optimal offloading strategy in real-time. The simulation results show that the proposed edge virtual machine allocation and task scheduling method can achieve near-optimal performance with extremely low complexity. In addition, the convergence speed is faster, and the total cost is lower.

If only the system cost caused by transmission delay is considered, it can also be converted into the pursuit of maximizing system throughput and transmission rate. To maximize the average throughput among users, the work [136] proposes a space-ground scenario, where multiple UAVs and BSs are deployed under the premise of a satellite-ground link. Considering different constraints and QoS, the authors in [136] proposed a joint optimization problem of user association, transmission power, and UAV flight trajectory. By decomposing this iterative problem and solving it sequentially, the proposed scheme outperforms different benchmark schemes in terms of the average users' throughput.

4.4.3. Multi-Dimensional Joint Optimization

Due to the complexity and constraints of SAGIN networks, a single optimization objective may lead to insufficient system performance improvement. Therefore, many scholars consider the possibility of multi-dimensional system cost optimization objectives, including the two weighted metrics (delay and energy consumption, security and power consumption) and the three weighted metrics (reliability, energy efficiency, and load balancing) [57,137]. According to users' requirements, the weights of different targets may be continuously adjusted, thereby improving the adaptability of multi-user differentiated service applications. However, pursuing multi-dimensional optimization objectives of the system will bring mixed problems with high computational complexity. Therefore, to reduce the difficulty of solving, the problem is usually decomposed without losing the accuracy in the numerical simulation.

Li et al. in [138] propose a cloud-edge collaborative on-orbit edge computing architecture, in which indivisible tasks can be performed locally, whether on satellite edge servers, or forwarded to ground cloud centers. To minimize the energy consumption and delay weighted total cost of the ground terminal under the joint optimization of offloading decision and resource allocation, the collaborative satellite-terrestrial network distributed offloading algorithm based on a parallel neural network is designed to speed up computation and convergence. Similar application scenarios and two-dimensional optimization objectives have been proposed in [82]. However, the authors relax the binary variables of the optimization problem based on linear reconstruction technology, transforming the original non-convex problem into a convex problem, and also achieve an excellent result of minimization of the joint optimization of the delay and energy consumption of ground terminals.

Zhu et al. in [139] consider new application scenarios, such as industrial automation and environmental monitoring deployed in remote areas. Resource-constrained terminals cannot meet the delay requirements without experiencing the local execution mode. The work [139] proposes a novel offloading cost, which is accumulated by all ground users and cloud edge nodes in the system. The offloading cost consists of two parts, delay and energy consumption, and considers multiple constraints, such as offloading location, current channel state, and bandwidth allocation. In addition, to speed up the optimization, the deep reinforcement learning-based task offloading (DRTO) algorithm is proposed in this paper continuously to adjust the number of candidate positions for space-ground networks with fast fading channels, thereby achieving near-optimal offloading performance with less time consumption. In some specific scenarios, resource services may be multi-faceted. For example, the work [140] roughly divides resource services for high-speed rail into secure and non-secure services according to reliability requirements. Different services will also pursue other optimization objectives. For example, safety services related to operation control and scheduling of high-speed trains have high requirements on bit error rates and end-to-end delay performance. In contrast, non-safety services, such as video surveillance and passenger information systems (PIS), usually require higher data rates. Under the architecture of diversified network resources in [140], a safety-oriented resource allocation scheme is proposed to deliver security services, which can always meet the multi-objective application requirements of high-speed rail security services.

4.4.4. Specific Optimization Objectives

Under the complex network environment and multiple time-varying constraints, many scholars also pay attention to particular types of system objectives, such as high security, high reliability, and low cost-effectiveness ratio, to cope with the diverse needs of the user.

In the complex environment of SAGIN, data communication security has received attention due to its unpredictability and aggressiveness. It must be protected by methods such as physical waveform processing, encryption and decryption, and trusted transmission. Although mobile edge computing provides nearby resource services to improve user service quality, it is still challenging to provide edge computing customized services to meet the various personalized needs of vehicles for computing-intensive applications in the heterogeneous Internet of vehicles, to ensure the QoS of vehicle customization quality of experience. Hui et al. in [141] analyzed the attack model, including malicious edge nodes and malicious vehicle nodes, in detail, designed a new secure edge computing service framework based on the computing resources of different network infrastructures, and introduced the Nash game equilibrium to guide the network deployment. The collaborative computing resource scheduling algorithm, based on the optimal bidding strategy between the network nodes, can satisfy the individual experience of different users. In particular, the security analysis verifies that the scheme can effectively defend against attacks. In some terrestrial or maritime edge computing service scenarios with uneven distribution density and surges in data volume, such as military applications, it is usually necessary to pay more attention to high-security data transmission and traffic offloading, in addition to the basic requirements of high speed, low latency and comprehensive coverage [142]. Especially when the satellite or airborne network is connected, the challenges in the complex network environment are more severe. To solve the security problems caused by cross-domain data interaction and frequent link switching, this paper uses the block-chain of the high-trust mechanism to realize distributed and lightweight node authentication, as well as information transmission of heterogeneous network nodes, to obtain a higher security level, lower computational and communication costs.

The space-ground joint edge computing network faces many challenges, such as high user service quality requirements, high mobility coordination, multi-task scheduling, etc. For instance, the work [143] analyzes the satellite-ground coordination failure mode and fault recovery in detail. In computing offloading and collaborative processing, the FDIR policy is introduced to improve system reliability to reduce the impact of faults caused by dynamic channel state changes, insufficient resources, and data corruption. Multiple fault recovery mechanisms are established, including computing task migration and task re-instantiation. The probability of failure of SAGIN edge computing is relatively high risk for applications with high-reliability requirements, such as aviation and aerospace. Therefore, the work [57] focuses on the problem of computing offloading for delay-sensitive applications under reliability constraints and designs a two-stage reliability-aware computational offloading method. Different from the task migration or reconstruction after failure in the previous paper, the reliability constraint defined in [57] is the dual constraint of the maximum time consumption constraint and the maximum energy consumption constraint, and an improved lazy shadowing scheme is designed to further enhance the reliability of the application. Future satellite networking collaborative edge offloading will also further study the above strategies.

In addition, many scholars also use the ratio of two target variables (which can be called the "efficiency-cost ratio") to measure the capability of edge service systems, such as the ratio of power consumption rate, the ratio of error rate, the ratio of data volume to power consumption, etc., which can be defined as efficiency objectives. For example, to realize the efficient utilization of spectrum resources by IoT devices in SAGIN, Ruan et al. in [144] comprehensively consider the energy efficiency and security requirements of satellite communication and propose a ratio of achievable secure communication rate to total power consumption to maximize energy efficiency to solve the optimization problem of the target. A cooperative cognitive beam-forming scheme is designed to facilitate secure and energy-efficient IoT communications. The simulation results verify the superiority of the collaborative resource management algorithm's implementation efficiency priority in this field.

When the users' object under consideration is tens of billions of IoT devices, improving user service density per unit area will become the objective pursued [145]. To make up for the shortage of ground network communication blockage and inability to deploy, adding SAGIN of satellites and UAVs to provide seamless coverage has excellent advantages. Meanwhile, when the perspective turns to the space-based edge server of the giant low-orbit constellation, the more stable visibility and lower inter-satellite distance will also bring system performance improvements to satellite edge computing, so that access users can obtain better real-time services [146]. Based on the sensitive quantity matrix of access user delay, communication rate, and visible time, the work [146] uses an extended graph model and an improved BFST algorithm to establish an inter-satellite link to achieve a highly stable and robust space-based edge server network.

Table 3 summarizes the optimization objectives for different user QoS requirements in different application scenarios. Lower energy consumption and lower latency are the most common optimization pursuits, which come from the basic requirements of the user terminal equipment for energy and time. Combining multi-dimensional optimization goals can lead to better performance improvements, but it can also lead to complex system models. In addition, special edge service objectives, such as high reliability, personalized service and network security, are also discussed in this paper.

Table 3. Summary of related works on SAGIN objectives.

Resource Types	Objective	Key Issues	Advantages	Disadvantages	Ref.
Minimum energy consumption	Minimizing computation energy of UAVs and UEs	Efficient and robust optimization problem	More stable energy reduction performance	Time-varying and random link channel	[124]
	Minimizing energy consumption of UAVs	Joint optimization of UAV trajectory and data transmission	Lower ground equipment requirements	High mobility, more constraints	[125]
	Maximizing the energy efficiency	Joint relay selection and power allocation	Better collaboration performance	More complex connections	[127]

Table 3. *Cont.*

Resource Types	Objective	Key Issues	Advantages	Disadvantages	Ref.
Lowest latency	Minimizing on-board image processing delay	Orbital edge intelligent framework, remote sensing	Lower backhaul load, higher bandwidth utilization of inter-satellite link	Limited application scenarios	[109]
	Minimizing the total weighted delay of uesrs	Joint computing and communication allocation	Higher fairness of multi users	Uncertain energy factor	[133]
	Maximizing the average inter-user throughput	Joint user association, power optimization and trajectory control	Higher users' throughput	More susceptible to interferences	[136]
Joint optimization	Minimizing weighted power consumption and latency	LEO edge computing system, joint computation offloading and resource allocation	Lower system average cost	Limited on-board resources, low versatility	[137]
	Improving reliability, energy efficiency, and load balancing	A two-stage reliability-aware offloading method	Higher network services reliability	More constraints, higher complexity	[57]
	Maximizing the normalized value of weighted data rate, error rate and delay	Resource allocation priorities, network handover costs	High safety, more suitable for delay sensitive applications	Larger state space	[140]
Specific objectives	Optimizing the overall scheduling	An optimal bidding strategy by Nash game	Higher personalized service experience	Worse environmental impact	[141]
	Improving robustness and security	Using high trust mechanism to realize data transmission	Higher security level, lower network cost	More complex channel state information	[142]
	Maximizing the number of users in the coverage area	Wide area connection, increasing user density	More stable continuity of service	Not suitable for high bandwidth applications	[145]

4.5. Key Algorithms

SAGIN MEC can improve user service quality and network resource utilization efficiency. Many scholars use different algorithms to conduct in-depth research on achieving the optimization objective. For example, the deep reinforcement learning method that integrates neural networks and reinforcement learning can use the high-dimensional state and operation space to solve the complex system analysis problems of caching, offloading, network and transmission coupling in MEC scenarios, which effectively promotes the research of mobile edge computing technology [147].

4.5.1. Reinforcement Learning

In the edge computing of SAGIN, energy consumption is a significant constraint. Xu et al. in [148] propose a joint resource allocation problem of a satellite-ground-sea network and use the classical reinforcement learning algorithm deep Q-network (DQN) to solve this collaborative communication and computing resource allocation problem and provide a better QoS guarantee. The algorithm defines the state space and action space.

The state space stores the channel quality state and computing capability state. The action space stores the selection of access points that provide users with network access services and the choice of MEC servers that provide computing assistance services. The computing offloading scheme proposed in [124] only considers the state of computing resources, without considering the constraints of communication resources. To address this issue, the works [56,149] proposed a new strategy. Luis et al. in [149] offer a power allocation scheme for multi-beam satellite systems based on deep reinforcement learning and use the near-end proximal policy optimization algorithm (PPO) to optimize the allocation strategy that does not meet the system requirements' lowest power. Optimizing the power allocation for each beam while keeping other satellite resource parameters fixed, the algorithm uses a pessimistic estimate of the policy's performance, which does not allow it to make continuous significant changes. In this way, it prevents the strategy's performance from deteriorating significantly in some cases, making the implementation process more stable and less volatile. In [56], a deep learning-based offloading strategy optimization strategy is proposed, which considers the dynamics of energy consumption performance, using a long short-term memory (LSTM) model to predict future energy. Then, the system uses the available energy information of the next time slot to optimize the computing offloading strategy for different IoT devices. The optimization algorithm can improve the computational performance of the system, considering the energy dynamics and other communication conditions, and maximize the completed computational tasks, while experiencing the energy and latency constraints of IoT devices. The algorithm proposed in [149] can minimize the power consumption but does not consider the delay problem in edge computing. Gao et al. in [150] propose a new offloading algorithm to take into account both the delay and energy consumption, providing edge computing services for ships on the ocean. Under the framework of the multi-armed bandit, which considers the choice of the UAV offloading to the edge server and takes into account the delay demand and energy consumption, a new optimal edge server offloading decision algorithm is proposed. Considering the changing marine environment, ship IoT users may not understand relevant prior information. Through the history of the algorithm, IoT users can analyze the reward and cost of each selected UAV to decide which UAV to choose next time to offload the mission.

Considering the delay and energy consumption in the above literature, Yu et al. in [54] divided the tasks into fine-grained slices. Fine-grained space-ground coordination-based joint offloading and caching algorithms were proposed to minimize the task completion time and satellite resource usage. Since the action space grows exponentially with the number of subtasks, it is difficult to obtain the optimal solution in polynomial time, so the optimization algorithm adopts a pre-classified offloading and caching scheme, as well as a decision-making scheme based on deep imitation learning. While significantly reducing the action space, this algorithm makes fine-grained offloading and cache placement decisions for tasks to minimize task completion time and satellite resource usage. However, fine-grained division of duties will inevitably lead to increasing loads. Therefore, Tang et al. in [151] proposed a cooperative offloading algorithm for LEO satellite networks based on a three-tier computing architecture. First, considering that LEO satellites can exchange information through inter-satellite links, this network's offloading strategy of inter-satellite cooperative computing is designed. Under this framework, the computing tasks of heavy-load LEO satellites can be forwarded to other light-load LEO satellites for processing, which can balance the computing load of the LEO satellite network and achieve better resource utilization. Second, a distributed deep learning-based collaborative computing offloading (DDLCCO) algorithm is proposed for the real-time computing offloading problem of the LEO satellite network in a time-varying environment. The algorithm can dynamically adjust the offloading decision according to the needs of ground users. Compared with the traditional optimization algorithm, the algorithm has low computational complexity and is more suitable for computational offloading in the actual network environment. Again, to make full use of the computing resources of the LEO satellite network, not only the horizontal cooperation between LEO satellites but also the vertical cooperation among ground users, LEO satellites and cloud servers are considered. Some scholars conducted research based on such algorithms for different scenarios and optimization objectives, such as in the high-speed rail safety business field [140].

4.5.2. Mathematical Programming

In addition to reinforcement learning, some other algorithms can play a role in SAGIN MEC. The work [53] proposes a three-layer computing architecture hybrid cloud and edge computing LEO satellite network to provide terrestrial users with heterogeneous computing resources, to enable terrestrial users to access computing services on a global scale. Based on the architecture, satisfying each LEO satellite's coverage time and computing capability constraints, the offloading calculation decision of the minimum ground user total energy consumption is studied. The problem is discrete and non-convex because both the objective function and conditions contain binary variables, making the problem difficult to solve. The authors transform the original non-convex problem into a linear programming problem, using a binary-variable relaxation method to address this challenging problem. Then, a distributed algorithm using the alternate direction method of multipliers is proposed to approach the optimal solution with low computational complexity. Unlike other MEC solutions, this paper fully uses the powerful resources of cloud and edge servers to provide heterogeneous computing services for ground users. In addition, the limited computing power and coverage time of each LEO satellite are also considered. For MEC-enhanced satellite IoT networks with multiple satellites and satellite gateways, we need joint optimization of coupled user association, offload decisions, computing, and communication resource allocation to reduce latency and energy costs. Therefore, the work [152] defines the delay and energy optimization problem of MEC-enhanced satellite IoT networks as a dynamic mixed integer programming problem, where it is difficult to obtain an optimal solution. To solve this problem, the complex problem is decomposed into two sub-problems. One is computing and communication resource allocation based on fixed user association and offloading decisions, and the other is user federation association and offloading based on optimal resource allocation. For the sub-problem of resource allocation, the optimal solution to the problem is proved by using the Lagrange multiplier method. On this basis, the second sub-problem is further formulated as a Markov decision process. A joint user association and offload decision with optimal resource allocation, based on deep reinforcement learning, is proposed. The simulation results show that this method can achieve better long-term latency and energy consumption returns. When the optimization problem is non-convex, Lagrangian duality theory is a good solution.

Therefore, the work [153] mainly studies the resource allocation problem of cells in the space-air-ground integrated vehicular networks (SAGVN), considering user association to optimize the connection between the BS and the vehicle. A low-complexity user association method is designed, and the car selects the base station connection that can obtain the maximum channel gain for reference. Since the objective function is non-convex, one can relax the constraints of the sub-channel allocation index and convert the objective function into a convex function. On this basis, a sub-channel and power allocation method is designed, considering the QoS of aircraft in the cell and the interference of UAVs and satellites. The proposed sub-channel allocation scheme ensures that the user obtains the maximum gain on the sub-channel. The Lagrangian duality theory is introduced to solve the power allocation problem. To reduce the delay in the offloading process, the delay time as a constraint can optimize the communication performance and ensure the low-latency QoS of vehicles in SAGVN.

4.5.3. Game Theory

The competition of users' equipment for computing resources in SAGIN edge computing can be considered as a classic game theory problem. Game theory is a powerful tool for designing distributed mechanisms that allow mobile device users in a system to make local decisions based on policy interactions and to achieve mutually satisfactory computational offloading solutions. This helps relieve the heavy burden of complex centralized management, such as collecting vast amounts of information from mobile device users for cloud operators. Furthermore, since different mobile devices are often owned by other individuals who may pursue different interests, game theory provides a practical framework to analyze the interactions among multiple mobile device users to their benefit and to design incentives that are compatible with computing loading mechanisms, so that no mobile user has an incentive to deviate unilaterally. Wang et al. in [80] propose a game-theoretic approach for computing offloading strategy optimization in satellite edge computing. A satellite edge computing offloading system model is established, and the intermittent problem of ground-satellite communication caused by satellite in-orbit operation is considered. A computational offloading game framework is set, and queuing theory is used as the optimization indicator to calculate task response time and energy consumption. Each device selfishly chooses the strategy that minimizes its cost. The response time and energy consumption of the task are calculated based on queuing theory. They are indicators of optimization performance, theoretically prove the existence and uniqueness of Nash Equilibrium, and propose an iterative algorithm. The game-based offloading strategy can significantly reduce the average cost of equipment. Under actual network conditions, satellites will have not only intermittent communication, but also cause mutual interference between multiple channels. Based on this, Chen et al. in [154] study the multi-user computing offloading problem of mobile edge cloud computing in the multi-channel wireless interference environment. The optimal solution is NP-hard in the computational set, so game theory methods are used to achieve efficient distributed offload computation. The distributed computing offloading decision problem among mobile device users is formulated as a multi-user computing offloading game. The decision's structural properties are analyzed, and it is proved that the decision has a Nash equilibrium and limited improvement properties. A distributed computing offloading algorithm that can achieve the Nash equilibrium is designed, the upper bound of the convergence time is derived, and its efficiency ratio relative to the centralized optimal solution is quantified from two essential performance indicators. The Nash game equilibrium method also achieves performance gains in multi-user customized service requirements, such as in [141].

4.5.4. Other Algorithms

In addition to the above three types of classical algorithms applied to SAGIN edge computing, some other algorithms can also solve the problem of SAGIN edge offloading. Wang et al. in [155] studied the resource scheduling problem in edge computing satellite networks. Considering the resource allocation strategy of edge computing satellites and the establishment of edge server collaborative networks in emergencies, and taking the different sensitivities of terminals, i.e., delay, bandwidth, and connection time, as resource allocation factors, a K-means algorithm is proposed to guide resource partitioning in edge servers. The spanning tree algorithm based on the breadth-first search is improved, and data transmission links are established to realize the dynamic scheduling of information and resources in emergencies. This dramatically reduces the need for the edge server to reallocate computing resources and realizes continuous terminal control based on dynamic adjustment. Although the work [155] can directly learn continuous control of the terminal based on dynamic adjustment, the memory consumption of the spanning tree based on the breadth-first search is significant. Therefore, the work [58] adopts an online learning method to study how to effectively deploy services on satellite edge computing nodes to achieve robustness-aware service coverage under limited resources. The problem can be formulated as a stochastic optimization problem with a long-term average objective function and constraints. Facing the challenges of space-time system dynamics and service coverage-robustness conflicts, a new online algorithm is proposed, which uses Lyapunov optimization theory, Markov approximation method, and Gibbs sampling algorithm to transform the long-term averaging problem into a real-time online optimization problem for each time point. The algorithm can converge to a near-optimal result, and the optimal gap has a theoretical boundary. Although online learning is in real-time, the training process is unstable and tends to go in the wrong direction. Therefore, to minimize the weighting and energy consumption of mobile devices, the work [156] decomposes the problem into the problem of reducing the delay of the space segment and the problem of minimizing the uncertainty of the ground side. The sequential fractional programming algorithm is used to solve the underlying sub-problem of the minimum space segment delay, and the first-order optimal solution is obtained. The upper sub-problems are decoupled and solved using convex structures and Lagrangian dual decomposition methods. Based on solving these two hierarchical sub-problems, an energy-efficient computation offloading and resource allocation algorithm (E-CORA) is proposed. In addition, the work [146] uses the extended graph model and the breadth-first search-based spanning tree (BFST) algorithm to realize a satellite edge service collaborative network.

Table 4 is a summary of common algorithms of SAGIN MEC. Because the optimization goal of edge computing is often achieved by finding the optimal offloading and scheduling decisions, deep reinforcement learning, Nash game equilibrium and other algorithms are often used. In addition, the particle swarm algorithm [144], fine-grained heuristic algorithm [44], spanning tree and other methods are also used in SAGIN to pursue better service guarantee.

Table 4. Summary of related works on SAGIN key algorithms.

Algorithm Types	Objective	Key Issues	Advantages	Disadvantages	Complexity	Ref.
Reinforcement learning methods	Minimizing the delay of computation and transmission	Edge-cloud collaborate, ratio of the service reward to the resouces renting cost	Easier to solve high-dimension problems	Slow convergence rate	N/A	[148]
	Improving dynamic energy Distribution of multi-beam satellites	Interaction with the environment to alternate sampling data	More stable policy implementation	Uncertain optimal strategy	N/A	[149]
	Minimizing the cumulative regret value of marine users	The reward and cost of decisions, upper bound of the confidence interval	Better performance under different QoS	Harder to solve huge state space problems	N/A	[150]
	Minimizing mission completion time and satellite resources	Learning optimal policies through behavioral cloning	Less action space, lower energy consumption for training	High requirements for training data	$o\left(4^{\|V\|}\right)$ $\|V\|: task\ number$	[54]
Mathematical programming	Minimizing the overall energy consumption	Relax binary variables, the alternating direction method of multipliers	Low computational complexity	Large communication overhead	$o(I^3)$ $I: user\ number$	[53]
	Maximizing the sum rate of IoVs	Optimize using the Lagrangian duality theory	Low system complexity	Low sample efficiency	N/A	[153]

Table 4. *Cont.*

Algorithm Types	Objective	Key Issues	Advantages	Disadvantages	Complexity	Ref.
Game theory	Minimizing the value of cost function	A computation offloading game framework, Nash equilibrium	Lower average energy consumption, high resource utilization	High balancing complexity	$O(KMN(log(I/\varepsilon)))$ [1]	[80]
	Improving offloading performance under interference environment	Distributed Nash equilibrium offloading	Higher computational efficiency	More complex with high mobility	$O(CM \log M)$ [2]	[154]
Others	Minimizing the maximum standard deviation of all clusters	Euclid distance, advanced K-means, breadth-first-search-based spanning tree	Stable continuity of control	Large memory consumption	N/A	[155]
	Improving service coverage and robustness	Lyapunov optimization, Gibbs sampling	Online fast optimization	Unstable training process	$O(I \quad L \max_{i \in S} \vert \mathcal{X}_i \vert)$ [3]	[58]

[1] K is the number of iterations, M is the number of satellites, N indicates the number of mobile devices, I represents the maximum iteration length, ε indicates the precision requirement. [2] M stands for the wireless channel, C indicates the number of slots that terminate the algorithm. [3] I denotes the iteration required for the outer loop to converge, L represents the number of iterations in the middle loop; $\max_{i \in S} |\mathcal{X}_i|$ represents the greatest number of iterations in each satellite edge computing node.

5. Challenges

It has become the consensus that SAGIN is the most prospective technology to provide users with edge services with low latency, comprehensive coverage, and high-reliability system characteristics. Many scholars have explored various application schemas based on their respective scenarios [157,158] and use different algorithms, such as deep reinforcement learning, the Lagrange multiplier method, and Nash game equilibrium to verify their feasibility for optimization problems. However, to fully exploit the advantages of SAGIN edge services and promote practical applications, some potential challenges and research issues still need to be further explored [143], which are summarized as follows.

5.1. High Dynamicity

The introduction of multi-layer heterogeneous constellations and aerial drone clusters in high, medium, and low orbits makes SAGIN more dynamic, which also brings more uncertainty to the user service experience, and can be mainly divided into the following three aspects: (1) uncertainty visibility, (2) unsecured resources, and (3) discontinuous services. First, although the satellite flies according to the preset orbit, considering factors such as inter-satellite communication, payload tasks, and on-orbit failures, the satellite's attitude may be adjusted to affect the satellite-ground link. Any position of the mobile user within the satellite coverage area will also constitute the uncertainty of the satellite-ground visibility. In addition, the trajectory of UAVs also involves dynamics. In conclusion, there is an emotional problem of connection between the two sides of the edge service in the SAGIN network. Second, although the development of large-scale processor circuits and lightweight memory chips has improved the resource capabilities of UAVs and satellites, compared with the almost unlimited power and resources of edge base stations in cellular networks, the edge service resources they can provide are still limited, and they may migrate to other satellites or backhaul to the remote cloud center. Therefore, there is still service uncertainty in the edge computing of the SAGIN network. Third, UAVs or satellite nodes with wide-area coverage will inevitably have multi-user access requirements. However, due to antenna layout and frequency constraints, it is usually necessary to use time-division multiplexing and multi-user methods to share frequencies and channels, which means that the service time at the edge of the SAGIN network is also different and discontinuous. It is also challenging to ensure user fairness and service density in a highly dynamic network [159–162].

5.2. Random Access Requirements

In traditional satellite-to-ground communication links, satellites and ground stations usually know the location and direction information, and the access process is predictable. At the same time, conventional terrestrial wired networks also have fixed connection characteristics, and network status changes are controllable. However, the SAGIN environment is constantly changing, and the number and types of users within its coverage are also evolving. To improve the system performance, users always seek the optimal offloading strategy, which leads to changes in the communication topology. Furthermore, in the face of emerging business applications and the emergency needs of multiple users, it is necessary to build flexible user access capabilities to edge end nodes (such as LEO satellites or UAVs) [163,164]. However, how to ensure flexible and reliable access for a large number of users, while maintaining a stable high-speed bandwidth after entry to ensure the reliability of edge computing services, and solving the global shortage of spatial telemetry, tracking, and command (TT&C) resources and the high complexity of planning and scheduling remain the focus of further research, which involves multi-user access and exit management, secure access authentication, global user uplink policy and downlink addressing policy and so on.

5.3. Task Relay and Migration

Based on the inter-satellite link technology, the gridded low-orbit mega-constellation can better maintain continuous coverage and communication with terrestrial users to achieve edge services in high mobility environments. However, it is limited by the degree of mission saturation, the number of users, and services. Influenced by factors such as resources, the highly dynamic satellite network topology and sudden changes in node access traffic will significantly impact the quality of on-orbit service. Therefore, carrying out mission relay and migration to meet users' needs is still a research challenge [165,166]. It includes the following two aspects: first, problems such as excessive computing tasks or discontinuous visibility cause a single edge computing node to be unable to be fully executed, and the remaining tasks or slices need to be transferred to the next pair of user-visible nodes for continued execution to achieve task relay. For example, the LEO satellite that undertakes the edge computing task offloaded by the user is about to leave the visible arc, and the remaining functions can be handed over to the backward satellite in the same orbit, which will complete the computing task and send the result to the user. Second, due to the imbalance of the network state, the space/air edge nodes in the connectable area are saturated with tasks and cannot execute user requests. The lessons can be migrated to low-load nodes. For example, if the satellites or UAVs visible in the hotspot area of the ground service are saturated with computing power, the application can migrate to adjacent low-load nodes for execution.

5.4. Network Security and Reliability

SAGIN is a wide-area open network with a wireless connection. The network covers a wide range of airspace and ground. In addition, the network covers are vulnerable to frequency interference and network attacks. Therefore, while the edge server provides flexible and convenient edge services to network users, it must be highly concerned with its anti-interference ability and information security to prevent illegal intrusion or malicious attacks from causing users' information leakage, data tampering, or edge service interruption. However, SAGIN challenges network security due to its complex and heterogeneous network composition. Many scholars mainly focus on communication interference and data security [167–170]. The SAGIN network environment is complex, which means that unintentional or malicious frequency interference from other devices will affect the MEC optimization objective [171]. Therefore, interference management and network scheduling are important research issues. In addition, malicious nodes' interception or eavesdropping of communication links will also challenge user information security. From the perspective of network reliability, the satellite-ground link has weather effects, such as rain attenuation, which leads to the deterioration of the channel quality and becomes more obvious with the increase in frequency bands. Therefore, the adaptive capability of the channel must be considered to obtain maximum communication quality. In addition, the single event flipping effect caused by space radiation will lead to data errors and even abnormal functions. Therefore, in the edge services of space-ground and space-air computation offloading and traffic scheduling, the reliability challenges brought by space environment radiation should also be paid attention to.

6. Future Research Directions

Combined with the introduction of the research branches of the key technologies and challenges of edge computing for SAGIN, the future trend of this research direction mainly includes the following aspects.

6.1. Wider Range of Emerging Businesses

Global ubiquitous users will generate large amounts of data and various service demands in real time, which will put greater operational pressure on traditional terrestrial cellular mobile networks. With the construction of national low-orbit satellite constellations, space-air cross-domain interconnection, multimode communication terminals, and other infrastructure, the integrated SAGIN with edge computing will drive the development of large user applications and complex new services, including space-air information services, space situational awareness, and real-time processing, with its seamless global coverage and user-access-on-demand characteristics.

6.2. Space-Air Information Service

The huge amount of network users will generate huge information data, which will bring unprecedented pressure on traffic control. Meanwhile, in the application scenario of SAGIN, the task is usually intensive, delay-sensitive and requires faster communication rates and higher information distribution efficiency, which is undoubtedly a greater challenge to network traffic management. In traditional network architectures, traffic generated by user terminals accesses the core network and further accesses cloud servers through satellites, UAVs or other access devices. If these services can be catered to at the network edge, the burden on the core network can be greatly reduced, thus improving channel bandwidth utilization. Not only telecom network operators, but also cloud application service providers face the same challenges. For example, if the data generated by IoT sensors (e.g., smart homes) are processed at the nearest edge node, the demand for computing resources at the remote data center is reduced. Therefore, edge computing in the SAGIN can effectively address the traffic pressure and congestion in the core network and data centers. Since space-air edge nodes in the SAGIN network are necessary for traffic transfer and offloading, attention is also paid to areas such as long propagation delay, link selection, and channel assignment [47].

6.3. Better Guaranteed QoS of Users

Space-air edge nodes gives SAGIN networks a stronger guarantee of users' QoS, but in the face of more access and more complex applications in the future, continuously improving users' QoS guarantee is still a research trend and a goal. The key point is to focus on the delay characteristics and energy consumption. On the one hand, it remains a problem of how to effectively allocate computing resources, reduce computing waits, and match users' computing demands for numerous users in a specific region at the remote end, so as to achieve the goal of multi-user total latency optimization. On the other hand, it is still necessary to consider the balance between communication overhead and edge computing in space-air scenarios, which concerns the strategy of selecting the appropriate edge nodes and performing task allocation in space-air dual mobility scenarios.

6.4. Satellite Networks Assistance

Undoubtedly, the construction of satellite network infrastructure will greatly promote the development of edge computing in the SAGIN and realize the task collaboration and efficient scheduling of the whole area network. Software-defined networks, network function virtualization, resource pooling characterization, inter-satellite dynamic routing and other related research are also advancing rapidly, which is an important guarantee and support for edge computing technology. Compared with a single satellite, satellite star cluster, same orbit ring satellite network, etc., a multi-layer multi-orbital satellite network has the advantages of high spatial spectrum utilization, low link congestion and high robustness. However, due to the rapid change in multi-satellite network topology and complex topology, QoS guarantee, dynamic migration, switching management, load balancing and other issues among the orbiting satellites remain key research directions.

6.5. Higher Security

Through edge computing in SAGIN, computation and traffic are distributed at the edge of the network for storage and processing. The distributed architecture is less susceptible to network attacks, thus improving the reliability of the entire network. At the same time, edge computing provides proximity services, which also shortens the in-network transmission time of information and reduces security risks. However, the large number of users, emerging services, access to multiple heterogeneous terminals, and application network deployment all pose new challenges to the subsequent SAGIN network security and reliability [172]. Systems integrated with military applications generate and deliver large amounts of sensitive data, and resources require security, reliability, and real-time. Therefore, there is an increasingly urgent research trend to effectively resist interference, message tampering and malicious attacks.

7. Conclusions

SAGIN edge computing technology is regarded as the most prospective technology by researchers worldwide, due to its low latency, high bandwidth, and ubiquitous coverage characteristics. As a result, it is gradually being pushed into practical applications. This paper aims to study the MEC architecture, key technologies, and challenges of SAGIN heterogeneous networks. First, we briefly review the development of related network technologies in detail, design SAGIN network system architecture and service framework, and analyze the characteristics and advantages of SAGIN edge computing. Then, we describe the critical technologies of SAGIN, including MEC deployment, resource scheduling, edge intelligence, optimization objectives, and critical algorithms. Finally, we discuss some problems and challenges that exist in SAGIN edge computing technology, hoping to put forward some new ideas for future applications in this field.

Author Contributions: Conceptualization, J.N. and X.Z.; resources, Y.Q.; writing—original draft preparation, Y.Q., K.Z., Y.Y. and B.R.; writing—review and editing, Y.Q.; supervision, T.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mitra, R.N.; Agrawal, D.P. 5G mobile technology: A survey. *Ict. Express* **2015**, *1*, 132–137. [CrossRef]
2. Panwar, N.; Sharma, S.; Singh, A.K. A survey on 5G: The next generation of mobile communication. *Phys. Commun.* **2016**, *18*, 64–84. [CrossRef]
3. Cisco Annual Internet Report (2018–2023) White Paper[EB/OL]. 2020. Available online: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html (accessed on 1 July 2022).
4. Duan, T.; Dinavahi, V. Starlink space network-enhanced cyber–physical power system. *IEEE Trans. Smart Grid.* **2021**, *12*, 3673–3675. [CrossRef]
5. Chien, W.; Lai, C.; Hossain, M.S.; Muhammad, G. Heterogeneous space and terrestrial integrated networks for IoT: Architecture and challenges. *IEEE Netw.* **2019**, *33*, 15–21. [CrossRef]
6. Space-Enabled 5G links Japan and Europe[EB/OL]. 2022. Available online: https://www.esa.int/Applications/Telecommunications_Integrated_Applications/Space-enabled_5G_links_Japan_and_Europe (accessed on 1 July 2022).
7. Mach, P.; Becvar, Z. Mobile edge computing: A survey on architecture and computation offloading. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 1628–1656. [CrossRef]
8. Elgendy, I.A.; Zhang, W.; Zeng, Y.; He, H.; Tian, Y.; Yang, Y. Efficient and secure multi-user multi-task computation offloading for mobile-edge computing in mobile IoT networks. *IEEE Trans. Netw. Serv. Manag.* **2020**, *17*, 2410–2422. [CrossRef]
9. Samdanis, K.; Taleb, T. The road beyond 5G: A vision and insight of the key technologies. *IEEE Netw.* **2020**, *34*, 135–141. [CrossRef]
10. Guo, F.; Yu, F.R.; Zhang, H.; Li, X.; Ji, H.; Leung, V.C.M. Enabling massive IoT toward 6G: A comprehensive survey. *IEEE Internet Things J.* **2021**, *8*, 11891–11915. [CrossRef]
11. Saarnisaari, H.; Dixit, S.; Alouini, M.; Chaoub, A.; Giordani, M.; Kliks, A.; Matinmikko-Blue, M.; Zhang, N.; Agrawal, A.; Andersson, M.; et al. A 6G white paper on connectivity for remote areas. *arXiv* **2020**, arXiv:2004.14699.
12. Latva-aho, M.; Leppänen, K.; Clazzer, F.; Munari, A. *Key Drivers and Research Challenges for 6G Ubiquitous Wireless Intelligence*; Oulu University Library: Oulu, Finland, 2020.
13. Saad, W.; Bennis, M.; Chen, M. A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Netw.* **2019**, *34*, 134–142. [CrossRef]
14. Yao, H.; Wang, L.; Wang, X.; Lu, Z.; Liu, Y. The space-terrestrial integrated network: An overview. *IEEE Commun. Mag.* **2018**, *56*, 178–185. [CrossRef]
15. Wang, P.; Zhang, J.; Zhang, X.; Yan, Z.; Evans, B.G.; Wang, W. Convergence of satellite and terrestrial networks: A comprehensive survey. *IEEE Access* **2019**, *8*, 5550–5588. [CrossRef]
16. Niephaus, C.; Kretschmer, M.; Ghinea, G. QoS provisioning in converged satellite and terrestrial networks: A survey of the state-of-the-art. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 2415–2441. [CrossRef]
17. Kodheli, O.; Lagunas, E.; Maturo, N.; Sharma, S.K.; Shankar, B. Satellite communications in the new space era: A survey and future challenges. *IEEE Commun. Surv. Tutor.* **2020**, *23*, 70–109. [CrossRef]
18. Ren, T.; Niu, J.; Qiu, Y. Enhancing generalization of computation offloading policies in novel mobile edge computing environments by exploiting experience utility. *J. Syst. Archit.* **2022**, *125*, 102444. [CrossRef]
19. Wei, X.; Lv, R.; Yu, S.; Hu, Y.; Lu, R. Edge Computing Empowered Satellite-Territorial Networks: Service Architecture, Use Case, and Open Issues. In Proceedings of the International Conference on Wireless and Satellite Systems; Springer: Cham, Switzerland, 2020; pp. 308–323.
20. Feng, B.; Zhou, H.; Zhang, H.; Li, G.; Li, H.; Yu, S.; Chao, H. HetNet: A flexible architecture for heterogeneous satellite-terrestrial networks. *IEEE Netw.* **2017**, *31*, 86–92. [CrossRef]
21. Wang, Y.; Ru, Z.Y.; Wang, K.; Huang, P. Joint deployment and task scheduling optimization for large-scale mobile users in multi-UAV-enabled mobile edge computing. *IEEE Trans. Cybern.* **2019**, *50*, 3984–3997. [CrossRef]
22. Pfandzelter, T.; Hasenburg, J.; Bermbach, D. Towards a computing platform for the LEO edge. In Proceedings of the 4th International Workshop on Edge Systems, Analytics and Networking, Edinburgh, UK, 26 April 2021; pp. 43–48.
23. Wang, X.; Lin, H.; Zhang, H.; Miao, D.; Miao, Q.; Liu, W. Intelligent Drone-assisted Fault Diagnosis for B5G-enabled Space-Air-Ground-Space Networks. *IEEE Trans. Netw. Sci. Eng.* **2020**, *8*, 2849–2860. [CrossRef]
24. Hui, Y.; Cheng, N.; Huang, Y.; Chen, R.; Xiao, X.; Li, C.; Mao, G. Personalized Vehicular Edge Computing in 6G. *IEEE Netw.* **2021**, *35*, 278–284. [CrossRef]
25. Shang, B.; Yi, Y.; Liu, L. Computing over space-air-ground integrated networks: Challenges and opportunities. *IEEE Netw.* **2021**, *35*, 302–309. [CrossRef]
26. Zhang, P.; Wang, C.; Kumar, N.; Liu, L. Space-air-ground integrated multi-domain network resource orchestration based on virtual network architecture: A DRL method. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 2798–2808. [CrossRef]
27. Xu, C.; Ishikawa, N.; Rajashekar, R.; Sugiura, S.; Maunder, R.G.; Wang, Z.; Yang, L.; Hanzo, L. Sixty years of coherent versus non-coherent tradeoffs and the road from 5G to wireless futures. *IEEE Access* **2019**, *7*, 178246–178299. [CrossRef]
28. Qiu, C.; Yao, H.; Yu, F.R.; Xu, F.; Zhao, C. Deep Q-learning aided networking, caching, and computing resources allocation in software-defined satellite-terrestrial networks. *IEEE Trans. Veh. Technol.* **2019**, *68*, 5871–5883. [CrossRef]
29. Marchiori, A. Maximizing coverage in low-power wide-area IoT networks. In Proceedings of the 2017 IEEE International Conference on Pervasive computing and Communications Workshops (Percom Workshops), Kona, HI, USA, 13–17 March 2017; pp. 467–472.

30. Qiu, C.; Tian, T.; Song, Q.; Zhao, C.; Xu, F. Sleeping mode of multi-controller in green software-defined networking. *EURASIP J. Wirel. Commun. Netw.* **2016**, *2016*, 282. [CrossRef]
31. Evans, B.G. The role of satellites in 5G. In Proceedings of the 2014 7th Advanced Satellite Multimedia Systems Conference and the 13th Signal Processing for Space Communications Workshop (ASMS/SPSC), Livorno, Italy, 8–10 September 2014; pp. 197–202.
32. Bi, Y.; Han, G.; Xu, S.; Wang, X.; Lin, C.; Yu, Z.; Sun, P. Software defined space-terrestrial integrated networks: Architecture, challenges, and solutions. *IEEE Netw.* **2019**, *33*, 22–28. [CrossRef]
33. Suzhi, C.; Junyong, W.; Hao, H.; Yi, Z.; Shuling, Y.; Lei, Y.; Shaojun, W. Space edge cloud enabling network slicing for 5G satellite network. In Proceedings of the 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), Gong Yongsheng Tangier, Morocco, 24–28 June 2019; pp. 787–792.
34. Liu, M.; Wang, Y.; Li, Z.; Lyu, X.; Chen, Y. Joint optimization of resource allocation and multi-UAV trajectory in space-air-ground IoRT networks. In Proceedings of the 2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Virtual conference, Seoul, Korea, 25–28 May 2020; pp. 1–6.
35. Wang, Z.; Zhou, Z.; Zhang, H.; Zhang, G.; Ding, H. AI-Based Cloud-Edge-Device Collaboration in 6G Space-Air-Ground Integrated Power IoT. *IEEE Wirel. Commun.* **2022**, *29*, 16–23. [CrossRef]
36. Zhou, Z.; Feng, J.; Tan, L.; He, Y.; Gong, J. An air-ground integration approach for mobile edge computing in IoT. *IEEE Commun. Mag.* **2018**, *56*, 40–47. [CrossRef]
37. Lyu, F.; Wu, F.; Zhang, Y.; Xin, J.; Zhu, X. Virtualized and micro services provisioning in space-air-ground integrated networks. *IEEE Wirel. Commun.* **2020**, *27*, 68–74. [CrossRef]
38. Cheng, N.; Xu, W.; Shi, W.; Zhou, Y.; Lu, N.; Zhou, H.; Shen, X. Air-ground integrated mobile edge networks: Architecture, challenges, and opportunities. *IEEE Commun. Mag.* **2018**, *56*, 26–32. [CrossRef]
39. Geist, A.; Brewer, C.; Davis, M.; Franconi, N.; Heyward, S.; Wise, T.; Crum, G.; Petrick, D.; Ripley, R.; Wilson, C.; et al. *SpaceCube v3. 0 NASA Next-Generation High-Performance Processor for Science Applications*; National Aeronautics and Space Administration: Washington, DC, USA, 2019.
40. Furano, G.; Meoni, G.; Dunne, A.; Moloney, D.; Ferlet-Cavrois, V.; Tavoularis, A.; Byrne, J.; Buckley, L.; Psarakis, M.; Voss, K.; et al. Towards the use of artificial intelligence on the edge in space systems: Challenges and opportunities. *IEEE Aerosp. Electron. Syst. Mag.* **2020**, *35*, 44–56. [CrossRef]
41. Denby, B.; Lucia, B. Orbital edge computing: Machine inference in space. *IEEE Comput. Archit. Lett.* **2019**, *18*, 59–62. [CrossRef]
42. Liu, Y.; Li, X.; Yu, F.R.; Ji, H.; Zhang, H.; Leung, V.C.M. Grouping and cooperating among access points in user-centric ultra-dense networks with non-orthogonal multiple access. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 2295–2311. [CrossRef]
43. Laser Communication in Space[EB/OL]. Available online: https://mynaric.com/products/space/ (accessed on 1 July 2022).
44. Li, T.; Zhou, H.; Luo, H.; Yu, S. SERvICE: A software defined framework for integrated space-terrestrial satellite communication. *IEEE Trans. Mob. Comput.* **2017**, *17*, 703–716. [CrossRef]
45. Qu, H.; Xu, X.; Zhao, J.; Yue, P. An SDN-based space-air-ground integrated network architecture and controller deployment strategy. In Proceedings of the 2020 IEEE 3rd International Conference on Computer and Communication Engineering Technology (CCET), Beijing, China, 14–16 August 2020; pp. 138–142.
46. Guo, H.; Li, J.; Liu, J.; Tian, N.; Kato, N. A survey on space-air-ground-sea integrated network security in 6G. *IEEE Commun. Surv. Tutor.* **2021**, *24*, 53–87. [CrossRef]
47. Liu, J.; Shi, Y.; Fadlullah, Z.M.; Kato, N. Space-air-ground integrated network: A survey. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 2714–2741. [CrossRef]
48. Zhou, S.; Wang, G.; Zhang, S.; Niu, Z.; Shen, X.S. Bidirectional mission offloading for agile space-air-ground integrated networks. *IEEE Wirel. Commun.* **2019**, *26*, 38–45. [CrossRef]
49. ETSI, G. Network functions virtualisation (nfv): Architectural framework. *ETsI Gs NFV* **2013**, *2*, V1.
50. Alwattar, T.A.; Mian, A. Development of an elastic material model for BCC lattice cell structures using finite element analysis and neural networks approaches. *J. Compos. Sci.* **2019**, *3*, 33. [CrossRef]
51. Alwattar, T.A.; Mian, A. Developing an equivalent solid material model for BCC lattice cell structures involving vertical and horizontal struts. *J. Compos. Sci.* **2020**, *4*, 74. [CrossRef]
52. Al-wattar, T.A.A. Developing Equivalent Solid Model for Lattice Cell Structure Using Numerical Approaches. Ph.D. Thesis, Wright State University, Dayton, OH, USA, 2020.
53. Tang, Q.; Fei, Z.; Li, B.; Han, Z. Computation offloading in leo satellite networks with hybrid cloud and edge computing. *IEEE Internet Things J.* **2021**, *8*, 9164–9176. [CrossRef]
54. Yu, S.; Gong, X.; Shi, Q.; Wang, X.; Chen, X. EC-SAGINs: Edge-Computing-Enhanced Space–Air–Ground-Integrated Networks for Internet of Vehicles. *IEEE Internet Things J.* **2021**, *9*, 5742–5754. [CrossRef]
55. Cao, B.; Zhang, J.; Liu, X.; Sun, Z.; Cao, W.; Nowak, R.M.; Lv, Z. Edge–Cloud Resource Scheduling in Space–Air–Ground-Integrated Networks for Internet of Vehicles. *IEEE Internet Things J.* **2021**, *9*, 5765–5772. [CrossRef]
56. Mao, B.; Tang, F.; Kawamoto, Y.; Kato, N. Optimizing computation offloading in satellite-UAV-served 6G IoT: A deep learning approach. *IEEE Netw.* **2021**, *35*, 102–108. [CrossRef]
57. Peng, K.; Zhao, B.; Bilal, M.; Xu, X. Reliability-Aware Computation Offloading for Delay-Sensitive Applications in MEC-Enabled Aerial Computing. *IEEE Trans. Green Commun. Netw.* **2022**, *6*, 1511–1519. [CrossRef]

58. Li, Q.; Wang, S.; Ma, X.; Sun, Q.; Wang, H.; Cao, S.; Yang, F. Service coverage for satellite edge computing. *IEEE Internet Things J.* **2021**, *9*, 695–705. [CrossRef]
59. Liao, H.; Zhou, Z.; Zhao, X.; Wang, Y. Learning-based queue-aware task offloading and resource allocation for space–air–ground-integrated power IoT. *IEEE Internet Things J.* **2021**, *8*, 5250–5263. [CrossRef]
60. Yang, T.; Gao, S.; Li, J.; Qin, M.; Sun, X.; Zhang, R.; Wang, M.; Wang, M.; Li, X. Multi-Armed Bandits Learning for Task Offloading in Maritime Edge Intelligence Networks. *IEEE Trans. Veh. Technol.* **2022**, *71*, 4212–4224. [CrossRef]
61. Liu, M.; Wang, Y.; Chen, Y.; Jia, H. Joint stochastic computational resource and UAV trajectory for wireless-powered space-air-ground IoRT networks. *IEEE Access* **2020**, *8*, 193728–193743. [CrossRef]
62. Cassará, P.; Gotta, A.; Marchese, M.; Patrone, F. Orbital Edge Offloading on Mega-LEO Satellite Constellations for Equal Access to Computing. *IEEE Commun. Mag.* **2022**, *60*, 32–36. [CrossRef]
63. Zhao, M.; Chen, C.; Liu, L.; Lan, D.; Wan, S. Orbital collaborative learning in 6G space-air-ground integrated networks. *Neurocomputing* **2022**, *497*, 94–109. [CrossRef]
64. Cao, S.; Han, H.; Wei, J.; Yang, Y.Z.S.; Yan, L. Space Cloud-Fog Computing: Architecture, Application and Challenge. In Proceedings of the 3rd International Conference on Computer Science and Application Engineering, Sanya, China, 22–24 October 2019; pp. 1–7.
65. Zhou, C.; Wu, W.; He, H.; Yang, P.; Lyu, F.; Cheng, N.; Shen, X. Deep reinforcement learning for delay-oriented IoT task scheduling in SAGIN. *IEEE Trans. Wirel. Commun.* **2020**, *20*, 911–925. [CrossRef]
66. Wang, Z.; Lin, B.; Sun, L.; Wang, Y. Intelligent Task Offloading for 6G-Enabled Maritime IoT Based on Reinforcement Learning. In Proceedings of the 2021 International Conference on Security, Pattern Analysis and Cybernetics (SPAC), Chengdu, China, 18–20 June 2021; pp. 566–570.
67. Yang, J.; Zhang, T.; Wu, X.; Liang, T.; Zhang, Q. Efficient Scheduling in Space-Air-Ground Integrated Localization Networks. *IEEE Internet Things J.* **2022**, *9*, 17689–17704. [CrossRef]
68. Zhang, J.; Zhang, X.; Wang, P.; Liu, L.; Wang, Y. Double-edge intelligent integrated satellite terrestrial networks. *China Commun.* **2020**, *17*, 128–146. [CrossRef]
69. Zhang, Z.; Zhang, W.; Tseng, F.H. Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques. *IEEE Netw.* **2019**, *33*, 70–76. [CrossRef]
70. Tong, M.; Wang, X.; Li, S.; Peng, L. Joint Offloading Decision and Resource Allocation in Mobile Edge Computing-Enabled Satellite-Terrestrial Network. *Symmetry* **2022**, *14*, 564. [CrossRef]
71. Nguyen, V.; Khanh, T.T.; van Nam, P.; Thu, N.T.; Hong, C.S.; Huh, E. Towards flying mobile edge computing. In Proceedings of the 2020 International Conference on Information Networking (ICOIN), Barcelona, Spain, 7–10 January 2020; pp. 723–725.
72. Ghobaei-Arani, M.; Souri, A.; Rahmanian, A.A. A space-air-ground enabled edge computing architecture for the internet of things. In Proceedings of the 2021 IEEE 4th International Conference on Electronics Technology (ICET), Chengdu, China, 7–10 May 2021; pp. 752–757.
73. Wang, Z.; Yu, H.; Zhu, S.; Yang, B. Curriculum Reinforcement Learning-Based Computation Offloading Approach in Space-Air-Ground Integrated Network. In Proceedings of the 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP), Virtual meeting, Hunan, China, 20–22 October 2021; pp. 1–6.
74. Ding, C.; Wang, J.; Zhang, H.; Lin, M.; Li, G.Y. Joint optimization of transmission and computation resources for satellite and high altitude platform assisted edge computing. *IEEE Trans. Wirel. Commun.* **2021**, *21*, 1362–1377. [CrossRef]
75. Fang, X.; Wang, Y.; Feng, W.; Chen, Y.; Ai, B. Power Allocation for Maritime Cognitive Satellite-UAV-Terrestrial Networks. In Proceedings of the 2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), Beijing, China, 26–28 September 2020; pp. 139–143.
76. Guo, Y.; Li, Q.; Li, Y.; Zhang, N.; Wang, S. Service Coordination in the Space-Air-Ground Integrated Network. *IEEE Netw.* **2021**, *35*, 168–173. [CrossRef]
77. Cao, X.; Yang, B.; Huang, C.; Yuen, C.; Zhang, Y.; Niyato, D.; Han, Z. Converged reconfigurable intelligent surface and mobile edge computing for space information networks. *IEEE Netw.* **2021**, *35*, 42–48. [CrossRef]
78. Liu, J.; Du, X.; Cui, J.; Pan, M.; Wei, D. Task-oriented intelligent networking architecture for the space–air–ground–aqua integrated network. *IEEE Internet Things J.* **2020**, *7*, 5345–5358. [CrossRef]
79. Chen, Y.; Ai, B.; Niu, Y.; Zhang, H.; Han, Z. Energy-constrained computation offloading in space-air-ground integrated networks using distributionally robust optimization. *IEEE Trans. Veh. Technol.* **2021**, *70*, 12113–12125. [CrossRef]
80. Wang, Y.; Yang, J.; Guo, X.; Qu, Z. A game-theoretic approach to computation offloading in satellite edge computing. *IEEE Access* **2019**, *8*, 12510–12520. [CrossRef]
81. Ding, C.; Wang, J.; Zhang, H.; Zhang, H.; Wang, J.; Lin, M. Joint Optimization of Radio and Computation Resources for Satellite-Aerial Assisted Edge Computing. In Proceedings of the ICC 2021-IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; pp. 1–6.
82. Wei, K.; Tang, Q.; Guo, J.; Zeng, M.; Fei, Z.; Cui, Q. Resource Scheduling and Offloading Strategy Based on LEO Satellite Edge Computing. In Proceedings of the 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), Virtual conference, Norman, OK, USA, 27–30 September 2021; pp. 1–6.

83. Salmani, M.; Sohrabi, F.; Davidson, T.N.; Yu, W. Multiple Access Binary Computation Offloading via Reinforcement Learning. In Proceedings of the 2019 16th Canadian Workshop on Information Theory (CWIT), the campus of McMaster University, Hamilton, ON, Canada, 30 May–2 June 2019; pp. 1–6.
84. Bi, S.; Zhang, Y.J. Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 4177–4190. [CrossRef]
85. Ji, Z.; Wu, S.; Jiang, C.; Hu, D.; Wang, W. Energy-efficient data offloading for multi-cell satellite-terrestrial networks. *IEEE Commun. Lett.* **2020**, *24*, 2265–2269. [CrossRef]
86. Zhang, Y.; Di, B.; Wang, P.; Lin, J.; Song, L. HetMEC: Heterogeneous multi-layer mobile edge computing in the 6 G era. *IEEE Trans. Veh. Technol.* **2020**, *69*, 4388–4400. [CrossRef]
87. Wei, J.; Han, J.; Cao, S. Satellite IoT Edge Intelligent Computing: A Research on Architecture. *Electronics* **2019**, *8*, 1247. [CrossRef]
88. Willburger, K.; Schwenk, K.; Brauchle, J. AMARO—An On-Board Ship Detection and Real-Time Information System. *Sensors* **2020**, *20*, 1324. [CrossRef]
89. Li, D.; Wang, M.; Dong, Z.; Shen, X.; Shi, L. Earth observation brain (EOB): An intelligent earth observation system. *Geo-Spat. Inf. Sci.* **2017**, *20*, 134–140. [CrossRef]
90. Wang, P.; Zhang, J.; Zhang, X.; Liu, L.; Wang, Y.; Ouyan, L. Performance evaluation of double-edge satellite terrestrial networks on OPNET platform. In Proceedings of the 2018 IEEE/CIC International Conference on Communications in China (ICCC Workshops), Beijing, China, 16–18 August 2018; pp. 37–42.
91. Leyva-Mayorga, I.; Soret, B.; Röper, M.; Wübben, D.; Matthiesen, B.; Dekorsy, A.; Popovski, P. LEO small-satellite constellations for 5G and beyond-5G communications. *IEEE Access* **2020**, *8*, 184955–184964. [CrossRef]
92. Huang, Q.; Lin, M.; Wang, J.; Tsiftsis, T.A.; Wang, J. Energy efficient beamforming schemes for satellite-aerial-terrestrial networks. *IEEE Trans. Commun.* **2020**, *68*, 3863–3875. [CrossRef]
93. Pan, G.; Ye, J.; Zhang, Y.; Alouini, M. Performance analysis and optimization of cooperative satellite-aerial-terrestrial systems. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 6693–6707. [CrossRef]
94. Akiyoshi, T.; Okamoto, E.; Tsuji, H.; Miura, A. Performance improvement of satellite/terrestrial integrated mobile communication system using unmanned aerial vehicle cooperative communications. In Proceedings of the 2017 International Conference on Information Networking (ICOIN), Da Nang, Vietnam, 11–13 January 2017; pp. 417–422.
95. Tang, F.; Hofner, H.; Kato, N.; Kaneko, K.; Yamashita, Y.; Hangai, M. A Deep Reinforcement Learning-Based Dynamic Traffic Offloading in Space-Air-Ground Integrated Networks (SAGIN). *IEEE J. Sel. Areas Commun.* **2021**, *40*, 276–289. [CrossRef]
96. Lyu, J.; Zeng, Y.; Zhang, R. Spectrum sharing and cyclical multiple access in UAV-aided cellular offloading. In Proceedings of the GLOBECOM 2017-2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; pp. 1–6.
97. Guo, H.; Liu, J. UAV-enhanced intelligent offloading for Internet of Things at the edge. *IEEE Trans. Ind. Inform.* **2019**, *16*, 2737–2746. [CrossRef]
98. Jiang, C.; Zhu, X. Reinforcement learning based capacity management in multi-layer satellite networks. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 4685–4699. [CrossRef]
99. Zhong, C.; Gursoy, M.C.; Velipasalar, S. A deep reinforcement learning-based framework for content caching. In Proceedings of the 2018 52nd Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 21–23 March 2018; pp. 1–6.
100. Chen, S.; Yao, Z.; Jiang, X.; Yang, J.; Hanzo, L. Multi-agent deep reinforcement learning-based cooperative edge caching for ultra-dense next-generation networks. *IEEE Trans. Commun.* **2020**, *69*, 2441–2456. [CrossRef]
101. Li, J.; Xue, K.; Wei, D.S.L.; Liu, J.; Zhang, Y. Energy efficiency and traffic offloading optimization in integrated satellite/terrestrial radio access networks. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 2367–2381. [CrossRef]
102. Li, L.; Wang, H.; Su, W.; Cui, E.; Lu, Z. SAG-EC: Edge Computing Platform Based on Space-Air-Ground Integrated Network. In Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12–14 March 2021; Volume 5, pp. 2685–2689.
103. Gu, S.; Wang, Y.; Wang, N.; Wu, W. Towards data collection and transmission in 6G space-air-ground integrated networks: Cooperative HAP and LEO satellite schemes. *IEEE Internet Things J.* **2021**, *9*, 10516–10528.
104. Jia, Z.; Sheng, M.; Li, J.; Han, Z. Intelligent optimization of availability and communication cost in satellite-UAV mobile edge caching system with fault-tolerant codes. *IEEE Trans. Cogn. Commun. Netw.* **2020**, *6*, 1230–1241.
105. Zhou, Y.; Yu, F.R.; Chen, J.; Kuo, Y. Resource allocation for information-centric virtualized heterogeneous networks with in-network caching and mobile edge computing. *IEEE Trans. Veh. Technol.* **2017**, *66*, 11339–11351. [CrossRef]
106. Fu, S.; Gao, J.; Zhao, L. Integrated resource management for terrestrial-satellite systems. *IEEE Trans. Veh. Technol.* **2020**, *69*, 3256–3266.
107. Fu, S.; Wu, B.; Wu, S.; Fang, F. Multi-resources management in 6G-oriented terrestrial-satellite network. *China Commun.* **2021**, *18*, 24–36.
108. Hou, X.; Wang, J.; Fang, Z.; Ren, Y.; Chen, K.; Hanzo, L. Edge Intelligence for Mission-Critical 6G Services in Space-Air-Ground Integrated Networks. *IEEE Netw.* **2022**, *36*, 181–189. [CrossRef]
109. Wei, J.; Cao, S. Application of edge intelligent computing in satellite Internet of Things. In Proceedings of the 2019 IEEE International Conference on Smart Internet of Things (SmartIoT), Tianjin, China, 9–11 August 2019; pp. 85–91.
110. Deng, S.; Zhao, H.; Fang, W.; Yin, J.; Dustdar, S.; Zomaya, A.Y. Edge intelligence: The confluence of edge computing and artificial intelligence. *IEEE Internet Things J.* **2020**, *7*, 7457–7469. [CrossRef]

111. Yao, J.; Han, T.; Ansari, N. On mobile edge caching. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2525–2553. [CrossRef]
112. Chang, Z.; Lei, L.; Zhou, Z.; Mao, S.; Ristaniemi, T. Learn to cache: Machine learning for network edge caching in the big data era. *IEEE Wirel. Commun.* **2018**, *25*, 28–35. [CrossRef]
113. Yang, J.; Zhang, J.; Ma, C.; Wang, H.; Zhang, J.; Zheng, G. Deep learning-based edge caching for multi-cluster heterogeneous networks. *Neural Comput. Appl.* **2020**, *32*, 15317–15328. [CrossRef]
114. Dulac-Arnold, G.; Evans, R.; van Hasselt, H.; Sunehag, P.; Lillicrap, T.; Hunt, J.; Mann, T.; Weber, T.; Degris, T.; Coppin, B. Deep reinforcement learning in large discrete action spaces. *arXiv* **2015**, arXiv:1512.07679.
115. Zhong, C.; Gursoy, M.C.; Velipasalar, S. Deep reinforcement learning-based edge caching in wireless networks. *IEEE Trans. Cogn. Commun. Netw.* **2020**, *6*, 48–61. [CrossRef]
116. Brendan McMahan, H.; Moore, E.; Ramage, D.; Hampson, S.; Agüera y Arcas, B. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Ft. Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
117. Wang, X.; Han, Y.; Wang, C.; Zhao, Q.; Chen, X.; Chen, M. In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning. *IEEE Netw.* **2019**, *33*, 156–165. [CrossRef]
118. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531.
119. Wang, F.; Zhang, M.; Wang, X.; Ma, X.; Liu, J. Deep learning for edge computing applications: A state-of-the-art survey. *IEEE Access* **2020**, *8*, 58322–58336. [CrossRef]
120. Alwani, M.; Chen, H.; Ferdman, M.; Milder, P. Fused-layer CNN accelerators. In Proceedings of the 2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), Taipei, Taiwan, 15–19 October 2016; pp. 1–12.
121. Li, G.; Liu, L.; Wang, X.; Dong, X.; Zhao, P.; Feng, X. Auto-tuning neural network quantization framework for collaborative inference between the cloud and edge. In Proceedings of the International Conference on Artificial Neural Networks; Springer: Cham, Switzerland, 2018; pp. 402–411.
122. Zhang, Y.; Miao, Y.; Pan, S.; Chen, S. Efficient and Energy-Saving Computation Offloading Mechanism with Energy Harvesting for IoT. *Secur. Commun. Netw.* **2021**, *2021*, 8167796. [CrossRef]
123. Tong, Z.; Cai, J.; Mei, J.; Li, K.; Li, K. Dynamic Energy-Saving Offloading Strategy Guided by Lyapunov Optimization for IoT Devices. *IEEE Internet Things J.* **2022**. [CrossRef]
124. Lu, Y.; Huang, Y.; Hu, T. Robust Resource Scheduling for Air-Ground Cooperative Mobile Edge Computing. In Proceedings of the 2021 IEEE/CIC International Conference on Communications in China (ICCC), Xiamen, China, 28–30 July 2021; pp. 764–769.
125. Jia, Z.; Sheng, M.; Li, J.; Niyato, D.; Han, Z. LEO-satellite-assisted UAV: Joint trajectory and data collection for Internet of remote things in 6G aerial access networks. *IEEE Internet Things J.* **2020**, *8*, 9814–9826. [CrossRef]
126. Li, X.; Feng, W.; Chen, Y.; Wang, C.; Ge, N. Maritime coverage enhancement using UAVs coordinated with hybrid satellite-terrestrial networks. *IEEE Trans. Commun.* **2020**, *68*, 2355–2369. [CrossRef]
127. Xu, Y.; Wang, Y.; Sun, R.; Zhang, Y. Joint relay selection and power allocation for maximum energy efficiency in hybrid satellite-aerial-terrestrial systems. In Proceedings of the 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain, 4–7 September 2016; pp. 1–6.
128. Wang, D.; He, Y.; Yu, K.; Srivastava, G.; Nie, L.; Zhang, R. Delay-Sensitive Secure NOMA Transmission for Hierarchical HAP–LAP Medical-Care IoT Networks. *IEEE Trans. Ind. Inform.* **2021**, *18*, 5561–5572. [CrossRef]
129. Li, J.; Liang, W.; Xu, W.; Xu, Z.; Jia, X.; Zhou, W.; Zhao, J. Maximizing user service satisfaction for delay-sensitive IoT applications in edge computing. *IEEE Trans. Parallel Distrib. Syst.* **2021**, *33*, 1199–1212. [CrossRef]
130. Boero, L.; Marchese, M.; Patrone, F. The impact of delay in software-defined integrated terrestrial-satellite networks. *China Commun.* **2018**, *15*, 11–21. [CrossRef]
131. Xu, H.; Li, D.; Liu, M.; Han, G.; Huang, W.; Xu, C. A hybrid routing algorithm in terrestrial-satellite integrated network. In Proceedings of the 2020 IEEE/CIC International Conference on Communications in China (ICCC), Chongqing, China, 9–11 August 2020; pp. 90–95.
132. Mao, S.; He, S.; Wu, J. Joint UAV position optimization and resource scheduling in space-air-ground integrated networks with mixed cloud-edge computing. *IEEE Syst. J.* **2020**, *15*, 3992–4002. [CrossRef]
133. Zhang, S.; Cui, G.; Long, Y.; Wang, W. Joint computing and communication resource allocation for satellite communication networks with edge computing. *China Commun.* **2021**, *18*, 236–252. [CrossRef]
134. Cheng, N.; Lyu, F.; Quan, W.; Zhou, C.; He, H.; Shi, W.; Shen, X. Space/aerial-assisted computing offloading for IoT applications: A learning-based approach. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1117–1129. [CrossRef]
135. Quan, W.; Liu, Y.; Zhang, H.; Yu, S. Enhancing crowd collaborations for software defined vehicular networks. *IEEE Commun. Mag.* **2017**, *55*, 80–86. [CrossRef]
136. Pervez, F.; Zhao, L.; Yang, C. Joint User Association, Power Optimization and Trajectory Control in an Integrated Satellite-Aerial-Terrestrial Network. *IEEE Trans. Wirel. Commun.* **2021**, *21*, 3279–3290. [CrossRef]
137. Wang, B.; Feng, T.; Huang, D. A joint computation offloading and resource allocation strategy for LEO satellite edge computing system. In Proceedings of the 2020 IEEE 20th International Conference on Communication Technology (ICCT), Nanning, China, 28–31 October 2020; pp. 649–655.

138. Li, H.; Chen, C.; Li, C.; Liu, L.; Gui, G. Aerial Computing Offloading by Distributed Deep Learning in Collaborative Satellite-terrestrial Networks. In Proceedings of the 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP), Hunan, China, 20–22 October 2021; pp. 1–6.
139. Zhu, D.; Liu, H.; Li, T.; Sun, J.; Liang, J.; Zhang, H.; Geng, L.; Liu, Y. Deep reinforcement learning-based task offloading in satellite-terrestrial edge computing networks. In Proceedings of the 2021 IEEE Wireless Communications and Networking Conference (WCNC), Nanjing, China, 29 March–1 April 2021; pp. 1–7.
140. Yan, L.; Fang, X.; Hao, L.; Fang, Y. Safety-oriented resource allocation for space-ground integrated cloud networks of high-speed railways. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 2747–2759. [CrossRef]
141. Hui, Y.; Cheng, N.; Su, Z.; Huang, Y.; Zhao, P.; Luan, T.H.; Li, C. Secure and personalized edge computing services in 6G heterogeneous vehicular networks. *IEEE Internet Things J.* **2021**, *9*, 5920–5931. [CrossRef]
142. Pang, Y.; Wang, D.; Wang, D.; Guan, L.; Zhang, C.; Zhang, M. A space-air-ground integrated network assisted maritime communication network based on mobile edge computing. In Proceedings of the 2020 IEEE World Congress on Services (SERVICES), Beijing, China, 18–23 October 2020; pp. 269–274.
143. Xie, R.; Tang, Q.; Wang, Q.; Liu, X.; Yu, F.R.; Huang, T. Satellite-terrestrial integrated edge computing networks: Architecture, challenges, and open issues. *IEEE Netw.* **2020**, *34*, 224–231. [CrossRef]
144. Ruan, Y.; Li, Y.; Zhang, R.; Cheng, W.; Liu, C. Cooperative resource management for cognitive satellite-aerial-terrestrial integrated networks towards IoT. *IEEE Access* **2020**, *8*, 35759–35769. [CrossRef]
145. Gineste, M.; Deleu, T.; Cohen, M.; Chuberre, N.; Saravanan, V.; Frascolla, V.; Mueck, M.; Strinati, E.C.; Dutkiewicz, E. Narrowband IoT service provision to 5G user equipment via a satellite component. In Proceedings of the 2017 IEEE Globecom Workshops (GC Wkshps), Singapore, 4–8 December 2017; pp. 1–4.
146. Wang, F.; Jiang, D.; Qi, S.; Qiao, C.; Song, H. Fine-grained resource management for edge computing satellite networks. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6.
147. Luong, N.C.; Hoang, D.T.; Gong, S.; Niyato, D.; Wang, P.; Liang, Y.; Kim, D.I. Applications of deep reinforcement learning in communications and networking: A survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3133–3174. [CrossRef]
148. Xu, F.; Yang, F.; Zhao, C.; Wu, S. Deep reinforcement learning based joint edge resource management in maritime network. *China Commun.* **2020**, *17*, 211–222. [CrossRef]
149. Luis, J.J.G.; Guerster, M.; del Portillo, I.; Crawley, E.; Cameron, B. Deep reinforcement learning for continuous power allocation in flexible high throughput satellites. In Proceedings of the 2019 IEEE Cognitive Communications for Aerospace Applications Workshop (CCAAW), Cleveland, OH, USA, 25–26 June 2019; pp. 1–4.
150. Gao, S.; Yang, T.; Ni, H.; Zhang, G. Multi-armed bandits scheme for tasks offloading in MEC-enabled maritime communication networks. In Proceedings of the 2020 IEEE/CIC International Conference on Communications in China (ICCC), Chongqing, China, 9–11 August 2020; pp. 232–237.
151. Tang, Q.; Fei, Z.; Li, B. Distributed deep learning for cooperative computation offloading in low earth orbit satellite networks. *China Commun.* **2022**, *19*, 230–243. [CrossRef]
152. Cui, G.; Li, X.; Xu, L.; Wang, W. Latency and energy optimization for MEC enhanced SAT-IoT networks. *IEEE Access* **2020**, *8*, 55915–55926. [CrossRef]
153. Liu, Y.; Zhang, H.; Zhou, H.; Long, K.; Leung, V.C.M. User Association, Subchannel and Power Allocation in Space-Air-Ground Integrated Vehicular Network with Delay Constraints. *IEEE Trans. Netw. Sci. Eng.* **2022**. [CrossRef]
154. Chen, X.; Jiao, L.; Li, W.; Fu, X. Efficient multi-user computation offloading for mobile-edge cloud computing. *IEEE/ACM Trans. Netw.* **2015**, *24*, 2795–2808. [CrossRef]
155. Wang, F.; Jiang, D.; Qi, S.; Qiao, C.; Shi, L. A dynamic resource scheduling scheme in edge computing satellite networks. *Mob. Netw. Appl.* **2021**, *26*, 597–608. [CrossRef]
156. Song, Z.; Hao, Y.; Liu, Y.; Sun, X. Energy-efficient multiaccess edge computing for terrestrial-satellite Internet of Things. *IEEE Internet Things J.* **2021**, *8*, 14202–14218. [CrossRef]
157. Wu, Q.; Xu, J.; Zeng, Y.; Ng, D.W.K.; Al-Dhahir, N.; Schober, R.; Swindlehurst, A.L. A comprehensive overview on 5G-and-beyond networks with UAVs: From communications to sensing and intelligence. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 2912–2945. [CrossRef]
158. Xianghang, B.; Yongxing, C.; Chuang, D.; Chi, W.; Ke, J. Integrated Space-Air-Ground Early Detection Technologies and Applicationfor Potential Landslide of Transmission Line Corridor. In Proceedings of the 2021 4th International Conference on E-Business, Information Management and Computer Science, Hong Kong, China, 29–31 December 2021; pp. 513–519.
159. Wang, Y.; Zhang, J.; Zhang, X.; Wang, P.; Liu, L. A computation offloading strategy in satellite terrestrial networks with double edge computing. In Proceedings of the 2018 IEEE international conference on communication systems (ICCS), Chengdu, China, 19–21 December 2018; pp. 450–455.
160. Seyedi, Y.; Safavi, S.M. On the analysis of random coverage time in mobile LEO satellite communications. *IEEE Commun. Lett.* **2012**, *16*, 612–615. [CrossRef]
161. Yang, Z.; Li, Y.; Yuan, P.; Zhang, Q. Tcsc: A novel file distribution strategy in integrated leo satellite-terrestrial networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 5426–5441. [CrossRef]

162. Bertaux, L.; Medjiah, S.; Berthou, P.; Abdellatif, S.; Hakiri, A.; Gelard, P.; Planchou, F.; Bruyere, M. Software defined networking and virtualization for broadband satellite networks. *IEEE Commun. Mag.* **2015**, *53*, 54–60. [CrossRef]
163. Zhen, L.; Bashir, A.K.; Yu, K.; Al-Otaibi, Y.D.; Foh, C.H.; Xiao, P. Energy-efficient random access for LEO satellite-assisted 6G internet of remote things. *IEEE Internet Things J.* **2020**, *8*, 5114–5128. [CrossRef]
164. Zhang, Z.; Li, Y.; Huang, C.; Guo, Q.; Liu, L.; Yuen, C.; Guan, Y.L. User activity detection and channel estimation for grant-free random access in LEO satellite-enabled Internet of Things. *IEEE Internet Things J.* **2020**, *7*, 8811–8825. [CrossRef]
165. Ren, C.; Hou, J. Edge Network Extension Based on Multi-domains Fusion and LEO Satellite. In Proceedings of the International Conference on Internet of Things as a Service; Springer: Cham, Switzerland, 2020; pp. 633–643.
166. Han, H.; Wang, H.; Cao, S. Space Edge Cloud Enabling Service Migration for On-Orbit Service. In Proceedings of the 2020 12th International Conference on Communication Software and Networks (ICCSN), Chongqing, China, 12–15 June 2020; pp. 233–239.
167. Di, B.; Zhang, H.; Song, L.; Li, Y.; Li, G.Y. Ultra-dense LEO: Integrating terrestrial-satellite networks into 5G and beyond for data offloading. *IEEE Trans. Wirel. Commun.* **2018**, *18*, 47–62. [CrossRef]
168. Vassaki, S.; Poulakis, M.I.; Panagopoulos, A.D.; Constantinou, P. Power allocation in cognitive satellite terrestrial networks with QoS constraints. *IEEE Commun. Lett.* **2013**, *17*, 1344–1347. [CrossRef]
169. Li, B.; Fei, Z.; Chu, Z.; Zhou, F.; Wong, K.; Xiao, P. Robust chance-constrained secure transmission for cognitive satellite–terrestrial networks. *IEEE Trans. Veh. Technol.* **2018**, *67*, 4208–4219. [CrossRef]
170. Asheralieva, A.; Niyato, D. Distributed dynamic resource management and pricing in the IoT systems with blockchain-as-a-service and UAV-enabled mobile edge computing. *IEEE Internet Things J.* **2019**, *7*, 1974–1993. [CrossRef]
171. Wang, Y.; Su, Z.; Ni, J.; Zhang, N.; Shen, X. Blockchain-empowered space-air-ground integrated networks: Opportunities, challenges, and solutions. *IEEE Commun. Surv. Tutor.* **2021**, *24*, 160–209. [CrossRef]
172. Kang, J.; Yu, R.; Huang, X.; Wu, M.; Maharjan, S.; Xie, S.; Zhang, Y. Blockchain for secure and efficient data sharing in vehicular edge computing and networks. *IEEE Internet Things J.* **2018**, *6*, 4660–4670. [CrossRef]