A Unified Framework for Integrating Semantic Communication and AI-Generated Content in Metaverse

Yijing Lin, Zhipeng Gao, Hongyang Du, Dusit Niyato, Jiawen Kang, Abbas Jamalipour, and Xuemin Sherman Shen

Abstract

As the Metaverse continues to grow, the need for efficient communication and intelligent content generation becomes increasingly important. Semantic communication focuses on conveying meaning and understanding from user inputs, while AI-Generated Content utilizes artificial intelligence to create digital content and experiences. Integrated Semantic Communication and AI-Generated Content (ISGC) has attracted a lot of attentions recently, which transfers semantic information from user inputs instead of raw data, generates digital content, and renders graphics for Metaverse. In this paper, we introduce a unified framework that captures ISGC's two primary benefits: integration gain for optimized resource allocation and coordination gain for goal-oriented high-quality content generation to improve immersion from both communication and content perspectives. We also classify existing ISGC solutions, analyze the major components of ISGC, and present several use cases. We then construct a case study based on the diffusion model to identify a near-optimal resource allocation strategy for performing semantic extraction, content generation, and graphic rendering in the Metaverse. Finally, we discuss several open research issues, encouraging further exploring the potential of ISGC and its related applications in the Metaverse.

Introduction

The concept of Metaverse, originally introduced in the scientific novel Snow Crash, has attracted considerable interest from academia and industries. Metaverse refers to a virtual environment that seamlessly integrates with the physical world, allowing for the existence of digital avatars to engage in various activities, interact with other users, and access virtual objects and experiences. The construction of Metaverse is supported by all virtual reality (VR), augmented reality (AR), and the Internet of Things to create a comprehensive and interconnected digital ecosystem.

The continuous advancement of technologies such as semantic communication (SemCom) and AI-Generated Content (AIGC) has prompted the Metaverse to increase demands for efficient communication and intelligent content generation. Semantic communication focuses on the associated meanings rather than simply transmitting raw data to enable effective communication. In this context, semantic information, which refers to the meaningful content or knowledge derived from user inputs, plays a crucial role. It is extracted using a semantic encoder, composed of neural networks. This encoder functions to transform raw data into semantic information, transmitting the comprehension and interpretation of information in a coherent and meaningful manner. AIGC generates digital content automatically with the assistance of AI technologies to improve efficiency and provide personalized and relevant content tailored to the preferences and needs of the users. These technologies lead to the emergence of a new integration technology: integrated SemCom and AIGC (ISGC) for improving immersion from both communication and content perspectives. ISGC combines the benefits of SemCom and AIGC to extract autonomously relevant information from raw data, enabling the generation of high-quality digital content in the Metaverse without direct human intervention. Additionally, There are certain difficulties that may arise without tight integration between SemCom and AIGC for Metaverses:

- **Inefficient Use of Resources:** Recognizing the challenges of the collaborative execution of AIGC tasks across a multitude of devices and the diverse access requirements of users [1], there is a lack of integration in the allocation of computing and communication resources for semantic extraction, AIGC, and graphic rendering tasks. Due to this disconnected structure, each single function is unable to fully utilize resources, leading to sub-optimal performance and inefficiencies in the overall system.
- **Low-Quality of Content:** Without effective coordination between SemCom and AIGC, the generated content may not meet the desired quality standards [2]. This can lead to poor user experiences and dissatisfaction, ultimately affecting the adoption and success of Metaverse.

As a consequence, ISGC has emerged as a promising technology that combines the advantages of the aforementioned technologies. By

Digital Object Identifier: 10.1109/MNET.2023.3321539
Date of Current Version: 15 July 2024
Date of Publication: 6 October 2023

Yijing Lin and Zhipeng Gao (corresponding author) are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China; Hongyang Du and Dusit Niyato are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798; Jiawen Kang is with the Guangdong University of Technology, Guangzhou 510006, China; Abbas Jamalipour is with The University of Sydney, Sydney, NSW 2006, Australia; Xuemin Sherman Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

0890-8044/23©2023IEEE

Authorized licensed use limited to: University of Waterloo. Downloaded on August 08,2024 at 19:27:07 UTC from IEEE Xplore. Restrictions apply.

combining AIGC and SemCom, ISGC enables the production of content that is not only visually appealing but also contextually relevant and meaningful, enhancing user experiences in the Metaverse. It also ensures that the right resources are allocated to the right tasks at the right time through joint computing and communication resource optimization. Additionally, it can adapt to user preferences, contextual information, and real-time interactions. In summary, ISGC can provide two main benefits over the above functionalities by obtaining the integration and coordination gains [3] for optimized resource allocation and high-quality content generation.

This paper provides a conceptual overview and concrete use cases of ISGC as well as its role in the Metaverse. Specifically, we present the related works and the key benefits of ISGC. We propose a unified framework for ISGC that utilizes the advantages of integration and coordination gains. Furthermore, we provide a case study employing the diffusion model to determine the near-optimal strategy for resource allocation in ISGC. It is shown that the diffusion model is capable of handling the effects of randomness and noise, and promoting exploration to enhance policy flexibility. To the best of our knowledge, this work is the first to comprehensively explore the potential integration and coordination benefits of ISGC for improving the efficiency and intelligence of the Metaverse. Our main contributions are summarized as follows:

- We propose a comprehensive overview of ISGC, including an investigation of related works, an explanation of why SemCom and AIGC integration is necessary, and the reasons why ISGC is needed in the Metaverse.
- We present a unified framework for ISGC, which includes a step-by-step workflow for capturing the integration and coordination gains, as well as several potential use cases.
- To further explore the benefits of integration gains, we conduct a case study that analyzes the effects of ISGC on resource allocation. Specifically, we use the diffusion model to derive and promote near-optimal strategies for utilities of resource allocation.

Why Integrate SemCom and AIGC

ISGC is a design paradigm in which SemCom and AIGC are integrated to provide efficient communication and goal-oriented content generation. A notable surge in research activities pertaining to ISGC has been observed, as shown in Fig. 1. We collect several papers *from IEEE Xplore and arXiv in April 2023* and identify research trends and directions from Fig. 1 that depict the current research of ISGC.

1) SemCom and AIGC. The integration of SemCom and AIGC primarily aims to leverage AIGC technologies such as Generative Adversarial Networks to develop semantic decoders that address the out-of-distribution problem between transmitters and receivers [4]. To compute the loss function, a variational autoencoder is employed to calculate the lower bound of semantic distortion, while diffusion models are combined with deep reinforcement learning to determine the near-optimal decisions in semantic communication [5].

2) SemCom and Metaverse. The integration of SemCom and Metaverse aims to circulate meaningful information with fewer symbols within the Metaverse, thereby reducing communication overheads. In order to mitigate privacy concerns arising from this integration, federated learning is introduced to preserve user data privacy [6].

3) AIGC and Metaverse. To improve the integration of AIGC and Metaverse, the focus is on generating high-quality digital content to create immersive virtual environments and construct economic systems, such as autonomous driving simulations and customized content. Additionally, the integration utilizes diffusion models to efficiently manage and optimize network and resource allocation [7].

4) SemCom, AIGC and Metaverse. The integration of SemCom, AIGC, and Metaverse is still in its early stages and is primarily focused on improving the efficiency of Metaverse through the application of SemCom and AIGC. GAN is utilized for the extraction of semantic information to improve the transmission efficiency in Metaverse [8].

Layers of Integration. Figure 2 depicts the various layers involved in the integration of ISGC. Data collected from sensors is extracted and transformed into semantic information such as image

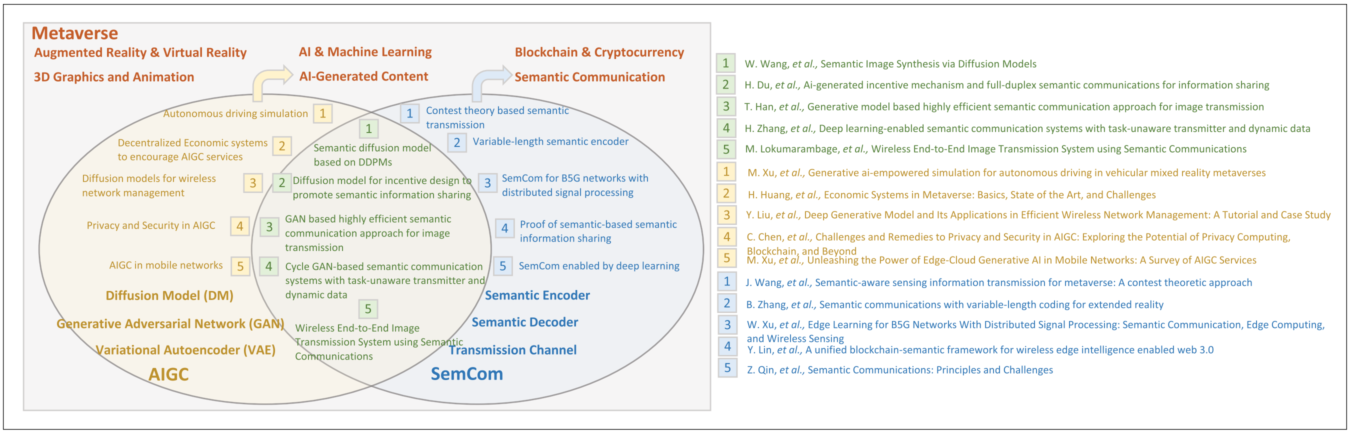

FIGURE 1. A review of recent research studies and emerging trends across SemCom, AIGC, and Metaverse, inspiring a unified framework for integrated SemCom and AIGC in the Metaverse.

Authorized licensed use limited to: University of Waterloo. Downloaded on August 08,2024 at 19:27:07 UTC from IEEE Xplore. Restrictions apply.

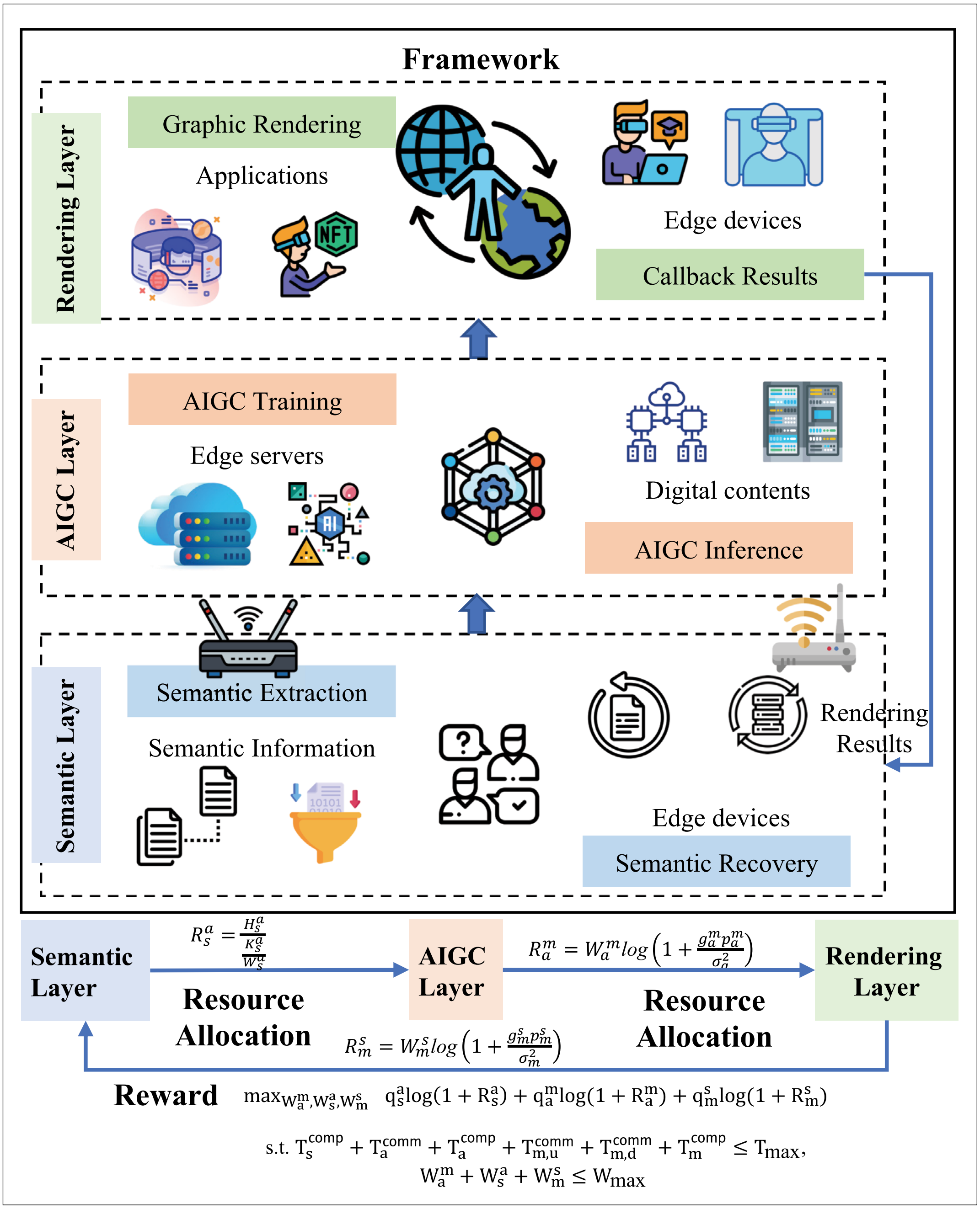

FIGURE 2. A layered view of ISGC-aided Metaverse: the upper segment comprises the semantic, AIGC, and rendering layers. The lower segment embodies the reward function, which facilitates strategic resource allocation across these layers. This reward function is influenced by the pricing and data rates associated with each layer and is constrained by the tolerable transmission time and overall bandwidth.

segments or model features, which is transmitted using semantic communication. AIGC inference is then applied to generate digital content from this information. The generated content is then fused through rendering graphics to create virtual environments that can be used by various applications and users within the Metaverse ecosystem.

Research Trends. To identify research trends related to ISGC, the research activities are classified into different layers based on their integration, as shown in Table 1. The figure highlights that current research solutions predominantly focus on addressing issues caused by individual layers, such as the out-of-distribution problem between the semantic encoder and decoder, efficient incentive mechanism for sharing semantic information, decentralized semantic sharing, and resource allocation for tasks within the same layers. However, these solutions may not fully reflect the benefits of the integration, and there is a need for more research efforts to explore the potential of ISGC as a whole.

ISGC Unified Framework

Framework Overview

ISGC includes semantic, inference, and rendering modules to capture the benefits of the integration of SemCom, AIGC, and Metaverse.

Authorized licensed use limited to: University of Waterloo. Downloaded on August 08,2024 at 19:27:07 UTC from IEEE Xplore. Restrictions apply.

1) Semantic Module: To optimize the data processing stage and reduce communication overhead, data collection, data processing, and semantic extraction should be performed at the edge devices. The semantic module is specifically designed to process data generated by edge devices and extract semantic information from raw data simultaneously. The extracted semantic information is then transmitted to Metaverse Service Providers (MSPs) that control the AIGC and render modules via edge servers.

2) Inference Module: Semantic information is fed into semantic decoder to recover useful information. Since the recovered images are low-quality or incomplete, MSPs should utilize AIGC to generate high-quality digital content to improve user experiences. The inference module employs pre-trained models to generate high-quality images with depth maps from multiple angles via the latent diffusion model, which employs forward and reverse diffusion processes to add and remove noise from images.

3) Rendering Module: Empowered by the above modules, the render module can synthesize massive and conditioned information from real-world or imaginary scenarios to enable immersive and interactive virtual environments.

Major Issues in Separated Functionalities

When the functionalities of SemCom, AIGC, and Metaverse are separated without ISGC, several significant issues may arise.

1) Resource Underutilization: Current resource allocation solutions tend to focus on individual modules, rather than considering the integrated ISGC as a whole. For instance, Wang et al. [9] proposed using contest theory to incentive users in the semantic module to contribute more valuable information. However, this approach may lead to the overuse of certain resources in one module while leaving others idle, resulting in inefficient resource allocation and decreased performance.

2) Limited Flexibility: The information provided by individual modules may suffer from high transmission latency or be affected by noise from the communication channel, leading to a decrease in the quality of user experience in Metaverse. For instance, Zhang et al. [10] proposed a variable-length channel coding method to highly compress unimportant semantic information to improve transmission efficiency. However, this approach may generate low-quality content in Metaverse. Additionally, the content generated by AIGC within Metaverse may require meaningful information from users to enhance the quality of particular applications.

Technical Gains of ISGC in Metaverse

Integration Gain. It can be achieved through resource allocation and sharing, specifically in terms of computing, communication, and dataset sharing among SemCom, AIGC, and Metaverse. A strategic allocation or balance of resources can be implemented based on environmental conditions and user requirements to maximize overall utilities. In cases where the channel conditions are unfavorable, it becomes impractical to allocate excessive resources to AIGC and the Metaverse, as they would be limited by the performance of SemCom. Instead, allocating more resources to SemCom can yield optimal utilities. This process can be seen as maximizing minimum utilities among SemCom, AIGC, and Metaverse. The workflow for dynamically coupling resources of ISGC consists of the following three steps. More details are shown in the next section.

- **Step 1: Design the joint optimization problem.** Given computing and communication resources, ISGC needs to consider both the usage of resources and latency in each module. To achieve this, ISGC can construct joint resource allocation optimization problems to maximize utility.
- **Step 2: Learn the policy via training.** Diffusion model-based Deep Reinforcement Learning (DRL) is utilized to solve the joint optimization problem and learn the policy since the diffusion model could mitigate the effect of randomness and noises [11].
- **Step 3: Generate the near-optimal strategy via inference.** The trained model can generate near-optimal strategies based on dynamic inputs to improve efficiency of the integration.

Coordination Gain. The coordination gain achieved by ISGC is essential in achieving goal-oriented content generation within Metaverse, which can couple the functionalities of semantic communication, AIGC inference, and graphic rendering more tightly. For coordination gain, we can customize SemCom based on the AIGC algorithm and Metaverse user requirements. For example, if a user is participating in virtual driving, SemCom should focus on vehicular network semantic information.

Layers	Solutions	Open challenges
Rendering Layer	Semantic diffusion model based on DDPMs	Identity authentication
	Diffusion model for incentive design to promote semantic information sharing [11]	Economic systems and decentralized autonomous
	GAN based highly efficient semantic communication approach for image transmission [8]	Redefine quality of experience and quality of service metrics
	Cycle GAN-based semantic communication systems with task-unaware transmitter and dynamic data [4]	Fraud, financial risks and crimes
	Wireless end-to-end image transmission system using semantic communications	
AIGC Layer	Autonomous driving simulation [7]	Content generation from multimodal inputs Content authenticity and incentives Content regulation and provenance Content caching and privacy. Interoperability standards
	Decentralized economic systems to encourage AIGC services	
	Diffusion models for wireless network management [12]	
	Privacy and security in AIGC	
	AIGC in mobile networks	
Semantic Layer	Contest theory based semantic transmission [9]	Training-based targeted semantic attacks and defenses Task-relevant semantic extraction Invariant semantic extraction across virtual environments Unified evaluation metrics Multi-user cooperative approaches
	Variable-length semantic encoder [10]	
	SemCom for B5G networks with distributed signal processing	
	Proof of semantic-based semantic information sharing	
	SemCom enabled by deep learning	

TABLE 1. ISGC-aided Metaverse: solutions, descriptions, and their integration challenges.

Authorized licensed use limited to: University of Waterloo. Downloaded on August 08,2024 at 19:27:07 UTC from IEEE Xplore. Restrictions apply.

By integrating SemCom, AIGC, and Metaverse, ISGC can extract semantic information efficiently, generate high-quality content with AI, and seamlessly integrate it into the Metaverse ecosystem. Unlike separate functionalities, the integration of ISGC ensures to perform mutual assistance. To provide a more concrete illustration of the coordination gain achieved by ISGC, a use case involving a virtual campus is presented in Fig. 3.

- **Step 1: Capture the environment.** In the scenario of a university campus, sensors, e.g., camera sectors of VR/AR/XR devices, capture the environmental settings from the real campus, like animals running around or airplane flying in the sky.
- **Step 2: Learn useful representations of input data.** Semantic information, e.g., feature vectors, is extracted from images in the semantic modules and transmitted to the inference module controlled by MSPs.
- **Step 3: Generate depth maps from representations.** MSPs first use feature vectors to reconstruct low-quality images and then generate depth maps with multiple angles of the environmental settings in the inference module.
- **Step 4: Render virtual campus with personalized feedback from depth maps.** The rendering model could provide personalized feedback to devices based on the above depth maps to simulate real-world settings.

Case Study: Exploring Integration Gain

Within the ISGC framework, to explore the integration gain, given limited available computing and communication resources, they need to be allocated to semantic extraction, AIGC inference, and graphic rendering modules to maximize the end-to-end utility. In this section, we first formulate the utility joint optimization problem of ISGC, as shown in Fig. 2. then achieve an effective resource allocation mechanism to obtain near-optimal strategies, and depict the simulation results of the proposed mechanism.

Problem Formulation

To simplify the notation, we use subscripts and superscripts of s, a, and m to represent the semantic, AIGC, and rendering modules, respectively. Additionally, comm represents the communication time, and comp represents the computation time.

1) Semantic Extraction. Edge devices employ the semantic module to extract semantic information from raw data, which reduces the amount of data transmitted by using fewer symbols. As mentioned in [12], the computation time T_s^{comp} for semantic extraction depends on the available computational resources of edge devices. Specifically, it is determined by the ratio of the required computational resources Z_s of semantic extraction to the total resources available C_s on edge devices.

2) Semantic Module to Inference Module. The semantic rate R_s^a refers to the amount of semantic information transmitted per second [13]. It is determined by considering the proportion of the approximate semantic entropy H_s^a, a measure of the uncertainty or randomness associated with semantic information, to the available bandwidth W_s^a between edge devices and MSPs and the average number of transmitted symbols K_s^a. Consequently, the time T_a^{comm} required to transmit semantic information D_s from the semantic module on edge devices to the inference module on edge servers is the ratio of the extracted semantic information to the semantic rate R_s^a.

3) AIGC Inference. Upon receiving semantic information from edge devices, MSPs carry out AIGC inference tasks that are guided by the semantic information to conditionally generate digital content in edge servers. The time required for AIGC inference T_a^{comp} is influenced by the computational resources available on edge servers containing the inference module. In particular, this time is dictated by the proportion of the necessary computational resources Z_a for AIGC inference to the overall resources C_a^s present on edge servers managed by MSPs.

4) Inference Module to Rendering Module. The transmission rate R_a^m from the inference module to the rendering module is computed as the product of the bandwidth available W_a^m between edge servers and MSPs and the channel capacity, referring to [7]. The channel capacity is influenced by the channel gain g_a^m, transmit power p_a^m, and

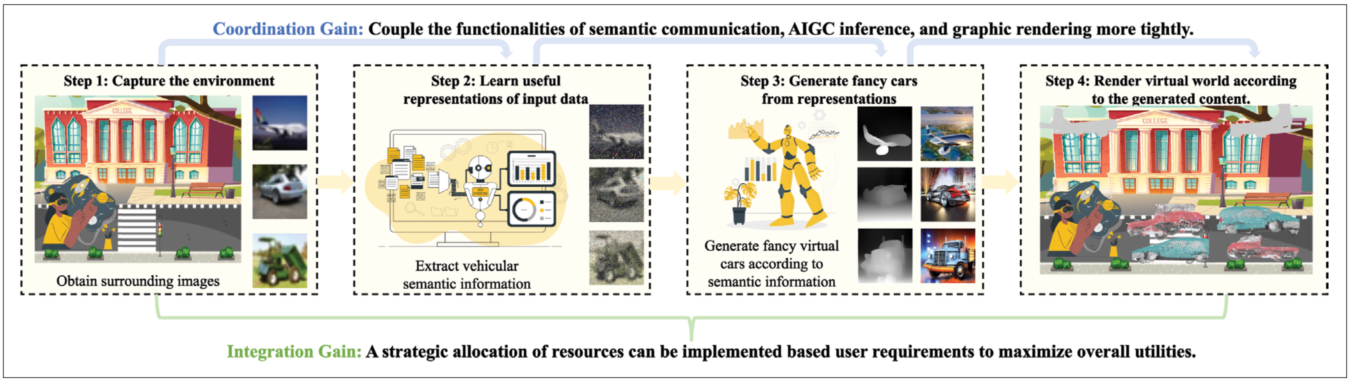

FIGURE 3. An illustration of the coordination gain of ISGC: Initially, edge devices capture the environment, obtaining images of the surroundings. From these images, they extract relevant semantic information. This information is then used to produce content. This content is subsequently utilized to render a virtual world, which in turn provides feedback to the edge devices. The synergy among the semantic, AIGC, and render modules constitutes the coordination gain. Meanwhile, the integration gain refers to the strategic resource allocation that takes these modules into account.

Authorized licensed use limited to: University of Waterloo. Downloaded on August 08,2024 at 19:27:07 UTC from IEEE Xplore. Restrictions apply.

the additive Gaussian noise σ_a^2. The transmission time $T_{m,u}^{\text{comm}}$ is determined by the ratio of the data size of the generated AIGC digital content to the transmission rate.

5) Graphics Rendering. Once the digital content is received from the corresponding edge servers running the inference module, MSPs equipped with the rendering module undertake graphics rendering tasks. These tasks involve leveraging digital content to augment and enrich virtual environments. The computation time T_m^{comp} required for graphics rendering is contingent upon the available computational resources of the edge servers deploying the rendering module. Precisely, this time is derived from the proportion of the required computing resources Z_m to the aggregate resources accessible C_m^a on these edge servers.

6) Rendering Module to Users. The transmission rate R_m^s between the rendering module and the semantic module can be analogous to that between the AIGC and rendering module with the bandwidth available W_m^s, channel gain g_m^s, transmit power p_m^s, and the additive Gaussian noise σ_m^2. The transmission time $T_{m,d}^{\text{comm}}$ is determined by the ratio of the data size of the rendering feedback to the transmission rate.

7) MSP Utility. MSPs impose charges on edge devices for the transmission and execution of tasks on edge servers. Referring to [12] and [13], the utility of MSPs can be determined by the products of the price q_s^a, q_a^m, q_m^s and transmit rate R_s^a, R_a^m, R_m^s of semantic, AIGC, and rendering modules. The utility is limited by the tolerable transmission time and the given bandwidth resources among the three modules, as shown in Fig. 2.

Diffusion Model-Based Joint Resource Allocation

Inspired by [14], in this paper, we present the diffusion model-based joint resource allocation mechanism. This mechanism is characterized as a Markov decision process consisting of state spaces, action spaces, environment dynamics, a reward function, a discount factor, and an initial state distribution. The reward is calculated by the utility function. The primary objective of this mechanism is to learn a policy that maximizes the cumulative discounted reward, thereby optimizing the utilities for MSPs within the ISGC framework.

AI-Generated Resource Allocation: The resource (i.e., bandwidth) allocation problem is solved by the diffusion model, which is composed of forward and reverse processes. The processes are designed to add and remove noises from samples, ultimately yielding generative outputs. The diffusion model can be further extended to include conditional models to represent the policy that optimizes the rewards for MSPs [14]. The conditional diffusion model is integrated with DRL to iteratively denoise the initial distribution and produce a near-optimal utility function for MSPs.

- **Step 1: Design state spaces.** Based on the MSPs' utility derived in the previous section, the near-optimal strategy $\pi(a^0 | s \in S)$ is influenced by a variety of factors, denoted as *state spaces* S. These state spaces $[H_s^a, \sigma_a, \sigma_m, g_a^m, p_a^m, g_m^s, p_m^s, K_s^a, C_a^s, C_m^a]$ include the approximate semantic entropy and the average transmitted symbols of the semantic module, channel gains and transmit power from the inference module to the rendering module, channel gains and transmit power from the rendering module to the inference module, as well as the computing resources and additive Gaussian noises present at the AIGC and rendering modules.
- **Step 2: Construct action spaces.** Given the state spaces, the *action spaces* $a^0 \in \mathcal{A}$ are associated with several factors, including the available bandwidth from the semantic, AIGC, and rendering modules, respectively. Consequently, the diffusion model that establishes a mapping between states S as the condition and action $\mathcal{A}$ as outputs represent the near-optimal policy $\pi(a^0 | s \in S)$. This policy yields a deterministic resource allocation strategy, which aims to maximize the expected cumulative reward over a series of steps.
- **Step 3: Explore the training policy in the forward process.** Initiating the training step involves providing hyper-parameters, including diffusion steps T, batch size, and exploration noise. The diffusion model is then initialized, incorporating two critic networks along with their corresponding target networks with different weights. In each iteration, the method initializes a random Gaussian distribution c^T for resource allocation exploration, followed by entering a loop of multiple steps. During each step, the method initially observes the current environment and its associated states, then sets the current actions as Gaussian noise. Subsequently, it generates the next action by denoising the current action $p(a^i | a^{i+1}, s)$ through the reverse diffusion process and adds exploration noise to the generated action. Once the action is executed, the method obtains the corresponding reward based on the utility function and stores the environment record in the replay buffer. To further refine the model, it samples a random mini-batch of records from the replay buffer, updates the critic networks by computing the loss and policy gradient, and finally updates the target networks.
- **Step 4: Generate near-optimal resource allocation strategy in the reverse process.** In the inference step, the environment with its associated states is input into the networks. Subsequently, the near-optimal resource allocation strategy $\pi(a^0 | s \in S)$ is generated by denoising Gaussian noise through the reverse diffusion process. This step focuses on utilizing the trained model to produce effective resource allocation strategies based on the given environmental conditions.

Simulation Results

The experimental platform utilized for executing the bandwidth resource allocation is built on a generic Ubuntu 20.04 system, featuring an AMD Ryzen Threadripper PRO 3975WX 32-Core CPU and an NVIDIA RTX A5000 GPU. The approximate semantic entropy, average transmitted symbols, channel gain and transmit power

Authorized licensed use limited to: University of Waterloo. Downloaded on August 08,2024 at 19:27:07 UTC from IEEE Xplore. Restrictions apply.

between the AIGC and rendering modules, as well as channel gain and transmit power between edge servers and devices, are randomly sampled from uniform distributions (1, 2), (0, 0.8), (0, 1), (3, 5), (0, 1), and (3, 5), respectively. The additive Gaussian noise at the AIGC and rendering modules is randomly sampled from normal distributions (0, 1) and (0, 1), respectively. The constraints of the total interaction time, the available bandwidth between the semantic, AIGC, and rendering modules, respectively. The above parameters are set as indicated in [7], [12], and [13].

In the simulation experiment, the diffusion model (Diffusion) and the Proximal Policy Optimization (PPO) [15] algorithm with learning rates, 3e-7 and 3e-6 are used to determine the near-optimal allocation of the available bandwidth among the semantic module, AIGC module, and rendering module. Unless otherwise specified, these methods are assumed to operate under identical parameters and environments. PPO is a model-free, on-policy actor-critic algorithm that uses the clipped surrogate objective to improve the stability and efficiency of learning. The training process is set to run for 3,000 epochs with buffer size 1,000,000, exploration noise 0.01, 10 steps per epoch, and 100 steps per collection, providing sufficient iterations for these methods to learn and adapt to the given environment and parameters.

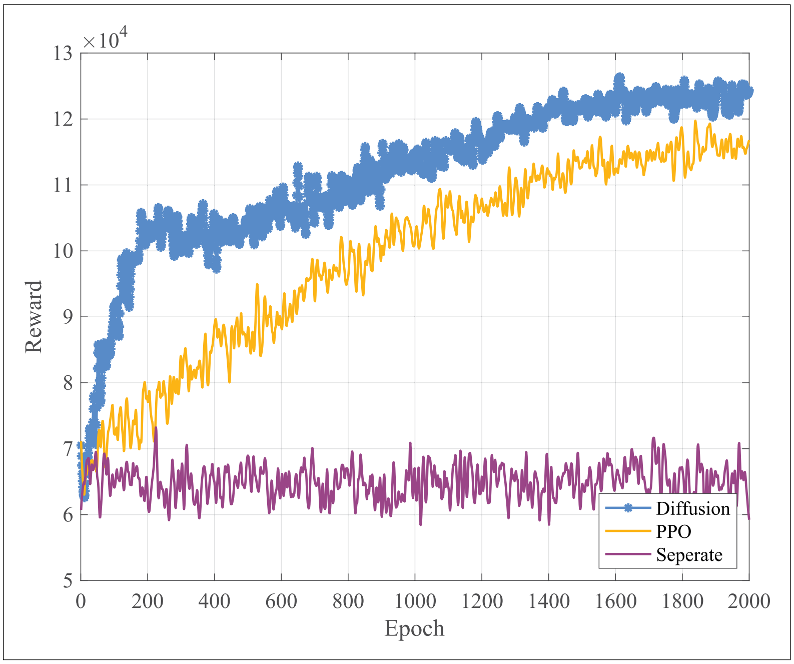

FIGURE 4. Training curves of the joint resource allocation.

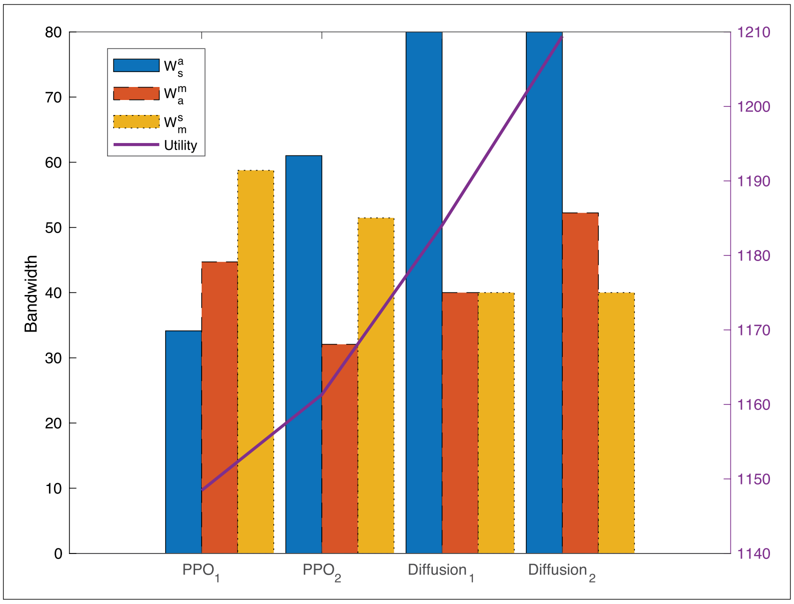

FIGURE 5. Generated utility of diffusion compared with PPO.

Figure 4 illustrates the reward comparison among our proposed mechanism, PPO, and Separate (the non-integrated baseline that randomly allocates resources among these modules). The training process demonstrates the reward values of Diffusion are significantly higher than those of PPO and Separate. Figure 5 compares the utilities computed by the near-optimal actions under different network states $[H_s^a, \sigma_a, \sigma_m, g_a^m, p_a^m, g_m^s, p_m^s, K_s^a, C_a^s, C_m^a]$, i.e., PPO_1 with [1.17, 0.66, 1.97, 0.30, 0.24, 4.76, 4.46, 0.91, 8.03, 15.28], PPO_2 with [1.40, 0.30, 0.65, 0.58, 0.16, 3.42, 4.69, 0.89, 7.49, 15.24], $Diffusion_1$ with [1.80, 1.05, 0.47, 0.05, 0.0004, 4.23, 4.58, 0.91, 5.02, 19.73], and $Diffusion_2$ with [1.52, 0.12, 1.23, 0.03, 0.14, 4.83, 3.86, 0.96, 9.93, 18.42], in the Diffusion and PPO methods. The PPO_1, PPO_2, $Diffusion_1$, and $Diffusion_2$ strategies are generated under various network states to evaluate the utility produced by the resource allocation mechanism based on the diffusion model. The near-optimal strategies of $Diffusion_1$ and $Diffusion_2$ found by Diffusion is better than that PPO_1 and PPO_2 of PPO. The underlying cause for these outcomes is that the diffusion model-based resource allocation mechanism can adapt outputs by fine-tuning given the diffusion steps and promoting exploration, thereby enhancing flexibility and mitigating the impact of uncertainty and noise encountered during the training process. This allows the proposed mechanism to achieve superior results in comparison to the other tested algorithms.

Real-World Example

Gaming and Virtual Reality (VR) is one of the concrete examples of the proposed framework in a practical context. Considering the complex environments in Massively Multiplayer Online Role-Playing Games or immersive VR simulations, our framework's semantic module can extract significant semantic information from raw data sensed by devices. The AIGC module then generates content tailored to this semantic information, while the rendering module brings game characters and their immediate surroundings to life. By leveraging diffusion model-based resource allocation, communication resources among these modules are optimized, guaranteeing the best bandwidth utilization for an enhanced user experience.

Future Directions

Several open challenges arise from the use of ISGC in the Metaverse as shown in Table 1. We elaborate on several of them in this section.

a) Invariant Semantic Extraction Across Virtual Environments: The diverse and heterogeneous nature of devices within the Metaverse

Authorized licensed use limited to: University of Waterloo. Downloaded on August 08,2024 at 19:27:07 UTC from IEEE Xplore. Restrictions apply.

may lead to the unintentional absorption of irrelevant environmental information during semantic extraction. This can result in the retrieval of meaningless information, potentially causing inaccuracies in content generation. It underscores the need to consider the impact of out-of-distribution data and the importance of extracting invariant semantic information across various virtual environments.

b) Content Authenticity and Incentives: The Metaverse's limited computational resources necessitate reliance on MSPs for content generation. Designing fair incentive mechanisms to verify content authenticity and incentivize MSPs is a crucial aspect of ensuring trust and collaboration within the system.

c) Practical Implementation Difficulties: Managing computing and communication functions across different layers and service providers presents substantial challenges. These arise from limited infrastructure access, diverse deployment environments, and specific requirements. Addressing interoperability, resource allocation, and performance management highlights the complexity of integrating compute and communication functions in real-world scenarios.

d) Communication Security and Privacy Preserving: With users required to upload their semantic information for customized AIGC-based immersive experiences within the Metaverse, safeguarding communication security and privacy becomes paramount. Collaborative techniques such as federated learning, differential privacy, and secure multi-party computation may offer solutions without exposing users' sensitive information [6].

Conclusion

In conclusion, this paper has provided a comprehensive overview of ISGC in the context of the growing Metaverse. By integrating SemCom and AIGC, ISGC offers significant benefits in terms of efficient communication and intelligent content generation. The proposed unified framework captures the integration and coordination gains of ISGC, optimizing resource allocation and enhancing the quality of digital content in the Metaverse. The case study utilizing the diffusion model demonstrates an improvement of 8.3% in rewards compared to PPO. However, there are still open research issues that need to be explored, such as privacy concerns and advanced techniques for resource allocation optimization. Overall, this paper contributes to the understanding and potential of ISGC, paving the way for immersive and intelligent experiences in the Metaverse.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 62072049 and Grant 62102099, in part by the Beijing Natural Science Foundation under Grant 4232029, in part by the Guangzhou Basic Research Program under Grant 2023A04J1699, in part by the National Research Foundation, Singapore, and Infocomm Media Development Authority under its Future Communications Research and Development Programme, DSO National Laboratories under the AI Singapore Programme under AISG Award AISG2-RP-2020-019, and in part by the Energy Research Test-Bed and Industry Partnership Funding Initiative, Energy Grid (EG) 2.0 Programme, DesCartes and the Campus for Research Excellence and Technological Enterprise (CREATE) Programme, and MOE Tier 1 under Grant RG87/22.

References

[1] H. Du et al., "Exploring collaborative distributed diffusion-based AI-generated content (AIGC) in wireless networks," *IEEE Netw.*, early access, Jul. 3, 2023, doi: 10.1109/MNET.006.2300223.

[2] D. Huang et al., "Deep learning-based image semantic coding for semantic communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.

[3] Y. Cui et al., "Integrating sensing and communications for ubiquitous IoT: Applications, trends, and challenges," *IEEE Netw.*, vol. 35, no. 5, pp. 158–167, Sep./Oct. 2021.

[4] H. Zhang et al., "Deep learning-enabled semantic communication systems with task-unaware transmitter and dynamic data," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 170–185, Jan. 2023.

[5] H. Du et al., "Beyond deep reinforcement learning: A tutorial on generative diffusion models in network optimization," 2023, *arXiv:2308.05384*.

[6] J. Chen et al., "Trustworthy semantic communications for the Metaverse relying on federated learning," *IEEE Wireless Commun.*, vol. 30, no. 4, pp. 18–25, Aug. 2023.

[7] M. Xu et al., "Generative AI-empowered simulation for autonomous driving in vehicular mixed reality Metaverses," 2023, *arXiv:2302.08418*.

[8] T. Han et al., "Generative model based highly efficient semantic communication approach for image transmission," 2022, *arXiv:2211.10287*.

[9] J. Wang et al., "Semantic-aware sensing information transmission for Metaverse: A contest theoretic approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5214–5228, Aug. 2023.

[10] B. Zhang, Z. Qin, and G. Y. Li, "Semantic communications with variable-length coding for extended reality," 2023, *arXiv:2302.08645*.

[11] H. Du et al., "AI-generated incentive mechanism and full-duplex semantic communications for information sharing," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2981–2997, Sep. 2023.

[12] Y. Liu et al., "Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11158–11168, Nov. 2019.

[13] L. Yan et al., "QoE-aware resource allocation for semantic communication networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2022, pp. 3272–3277.

[14] Z. Wang, J. J. Hunt, and M. Zhou, "Diffusion policies as an expressive policy class for offline reinforcement learning," 2022, *arXiv:2208.06193*.

[15] J. Schulman et al., "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.

Biographies

Yijing Lin (yjlin@bupt.edu.cn) is currently pursuing the Ph.D. degree with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China.

Zhipeng Gao (gaozhipeng@bupt.edu.cn) is a Professor with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China.

Hongyang Du (hongyang001@e.ntu.edu.sg) is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Nanyang Technological University, Singapore.

Dusit Niyato (Fellow, IEEE) (dniyato@ntu.edu.sg) is a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore.

Jiawen Kang (kavinkang@gdut.edu.cn) is a Professor with the School of Automation, Guangdong University of Technology, China.

Abbas Jamalipour (Fellow, IEEE) (a.jamalipour@ieee.org) is a Professor with The University of Sydney, Sydney, NSW, Australia.

Xuemin Sherman Shen (Fellow, IEEE) (sshen@uwaterloo.ca) is a professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada.

Authorized licensed use limited to: University of Waterloo. Downloaded on August 08,2024 at 19:27:07 UTC from IEEE Xplore. Restrictions apply.