

Semantic Communications for Artificial Intelligence Generated Content (AIGC) Toward Effective Content Creation

Guangyuan Liu , Hongyang Du , Dusit Niyato , Jiawen Kang , Zehui Xiong , Dong In Kim , and Xuemin Shen 

ABSTRACT

Artificial Intelligence Generated Content (AIGC) Services have significant potential in digital content creation. The distinctive abilities of AIGC, such as content generation based on minimal input, hold huge potential, especially when integrating with semantic communication (SemCom). In this paper, a novel comprehensive conceptual model for the integration of AIGC and SemCom is developed. Particularly, a content generation level is introduced on top of the semantic level that provides a clear outline of how AIGC and SemCom interact with each other to produce meaningful and effective content. Moreover, a novel framework that employs AIGC technology is proposed as an encoder and decoder for semantic information, considering the joint optimization of semantic extraction and evaluation metrics tailored to AIGC services. The framework can adapt to different types of content generated, the required quality and the semantic information utilized. By employing a Deep Q Network (DQN), a case study is presented that provides useful insights into the feasibility of the optimization problem and its convergence characteristics.

INTRODUCTION

Artificial Intelligence Generated Content services (AIGC services) have gained significant attention due to their ability to enhance creativity, accelerate design processes, provide personalized content and promote accessibility in various domains, such as digital marketing, video game design and filmmaking [1]. AIGC employs advanced Generative AI (GAI) techniques to produce content corresponding to human instructions by deciphering the intent and generating appropriate content in response.

The considerable advancements in AIGC can be attributed to the development of increasingly sophisticated generative models, expanded foundation model architectures and the accessibility of vast computational resources. AIGC services hold substantial importance for a variety of reasons, including:

- **Enhancing Creativity:** GAI models support users in generating unique and original images from textual prompts, thereby unveiling new avenues for artistic and creative expression.¹
- **Streamlining Design Processes:** AIGC services facilitate rapid generation of visual content, optimizing workflows for designers and artists which subsequently reducing time and effort required for manual design tasks.
- **Tailoring Content:** These services enable the creation of customized content that is specifically designed to cater to individual user preferences or distinct target audiences, ultimately resulting in heightened user engagement.
- **Promoting Accessibility:** AIGC services enable the users with limited design abilities to generate professional-quality visuals, thereby fostering inclusivity and democratizing the design process.

The efficient functioning and widespread adoption of AIGC services are contingent on robust communication infrastructure and technologies, such as wireless networks and semantic communication (SemCom). The technologies offer many benefits to AIGC services including [2]:

- **Ubiquitous access:** Wireless networks facilitate widespread access to AIGC services, allowing users to benefit from these services regardless of their location or device.
- **Real-time processing:** Wireless networks enable rapid data transfer, promoting real-time interactions between users and AI models for content generation, thus improving the overall user experience.
- **Scalability:** Wireless networks can efficiently support a large number of users simultaneously, enhancing the availability and accessibility of AIGC services.

On the other hand, SemCom can focus on the intended meaning or context of the transmitted data, rather than on its raw form, which leads to a more efficient and intelligent exchange of information. When SemCom is applied to AIGC services, this results in optimized transmission and interpretation of generated content.

¹ <https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work>, accessed July 06, 2023.

Guangyuan Liu and Hongyang Du are with the School of Computer Science and Engineering and the Energy Research Institute @ NTU, Interdisciplinary Graduate Program, Nanyang Technological University, Singapore 639798; Dusit Niyato is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798; Jiawen Kang (corresponding author) is with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China; Zehui Xiong is with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 487372; Dong In Kim is with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon 16419, South Korea; Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Digital Object Identifier:
10.1109/MNET.2024.3352917
Date of Current Version:
16 September 2024
Date of Publication:
11 January 2025

Specifically, the benefits of integrating SemCom into AIGC services include [3]:

- **Bandwidth Efficiency:** SemCom makes possible more efficient use of bandwidth, allowing generated contents to be transmitted with lower latency and less resource consumption.
- **Enhanced Privacy:** SemCom also contributes to increased privacy, ensuring secure exchanges of information between users and AIGC services due to shorter transmission time and less amount of transferred data.
- **Energy Efficiency:** By focusing on the meaningful components of data, SemCom can reduce the energy consumption of data transmission and processing in AIGC services.

Evidently, the development of efficient and secure wireless and SemCom technologies is essential for the widespread adoption and effective implementation of AIGC services. As such, further clear understanding of their integration is necessary to guide the design and implementation of AIGC services to meet the diverse needs of different applications and domains.

In this paper, we introduce a conceptual model for integrating and designing AIGC services with SemCom. Specifically, we extend three-level of SemCom, i.e., physical level, semantic level and effectiveness level, by having a generation level that allows AIGC services at the receiver to generate meaningful contents from input provided by the transmitter over a communication link. To realize the proposed model, we then explore the effect of extracting semantic information from AIGC inputs on the AIGC output quality. We propose an innovative framework that quantitatively evaluates the impact of various factors, such as image resolution and compression levels. The contributions of this paper are summarized as follows:

- We present the model that serves as a foundation to develop AIGC services by leveraging the capability of SemCom. The model can represent building blocks of the SemCom-enabled AIGC services and their interactions to achieve effective and efficient content generation by generative AI over wireless networks.
- We realize the model by conducting a thorough analysis of how alterations in the extraction and reduction of input information, such as changes in image resolution and compression, influence AIGC output quality.
- We introduce a novel framework that describes the vital role of AIGC in semantic information transmission. This strategy aims to optimize resource allocation to maximize user satisfaction with AIGC services.
- We present an in-depth experimental study of joint resource allocation to understand the complexities of accommodating various services within a single transmission framework. This study offers crucial insights into optimization results and their convergence characteristics. Then, we provide validation of our proposed framework through a Deep Q-Network (DQN), demonstrating its feasibility and promising implications for enhancing AIGC services efficiency.

SEMANTIC COMMUNICATIONS-ENABLED AI-GENERATED CONTENT MODEL

Traditional SemCom model prioritizes accurate transmission and reception of semantic content over relying solely on average information related to source data probabilities [4]. However, the advent of AIGC offers an opportunity to augment and benefit from SemCom model. In this section, we propose two integral components, the **Effective Information Creation Level** and the **Generation Level**. By implementing these components, AIGC can fulfill its dual objectives of information generation and communication more effectively.

INTEGRATION OF SEMCOM WITH AIGC

In traditional SemCom model, as shown in Fig. 1, physical and semantic levels form the two-pronged structure. **The physical level** encapsulates transmission operations within the physical layer, including modules responsible for channel encoding and decoding. This level necessitates symbol transmission over a physical channel susceptible to noise. **The semantic level** is composed of semantic representation and encoding modules at the transmitter end. These components work together to extract and represent semantic information. Furthermore, semantic interpretation and decoding modules situated at the receiver end are for reconstructing and interpreting the received semantic information.

To extend the capabilities of the SemCom model, the introduction of **effective information creation level** and **the generation level** is essential [4]. Specifically, the **effective information creation level** in the AIGC framework transforms raw user data into semantic representations, which are then used to guide the generation process. It involves feature extraction and the creation of model fine-tuning parameters based on context-specific requirements. Meanwhile, the **generation level** provides an engine of content creation. It utilizes the semantically-encoded information and the generation parameters to run controlled diffusion models, generating the final content. Together, these levels augment the traditional SemCom model by adding control and personalization, thereby making AIGC a more adaptable and effective tool for content generation and transmission.

The generation level has the following components:

- **Background Knowledge:** This component represents the foundation of knowledge and expertise necessary for content generation. On the transmitter side, it encompasses generation parameters that are finely tuned for specific generation tasks. On the receiver side, it includes the knowledge of how to fine-tune models with provided parameters and how to infer the generation process using the given generation parameters.
- **Shared Knowledge:** As a shared knowledge base, this component allows both the transmitter and receiver to access relevant information during content generation and interpretation. It consists of extensive data containing Pretrained Foundation Models (PFMs) that can be utilized for additional fine-tuning, thereby enhancing the

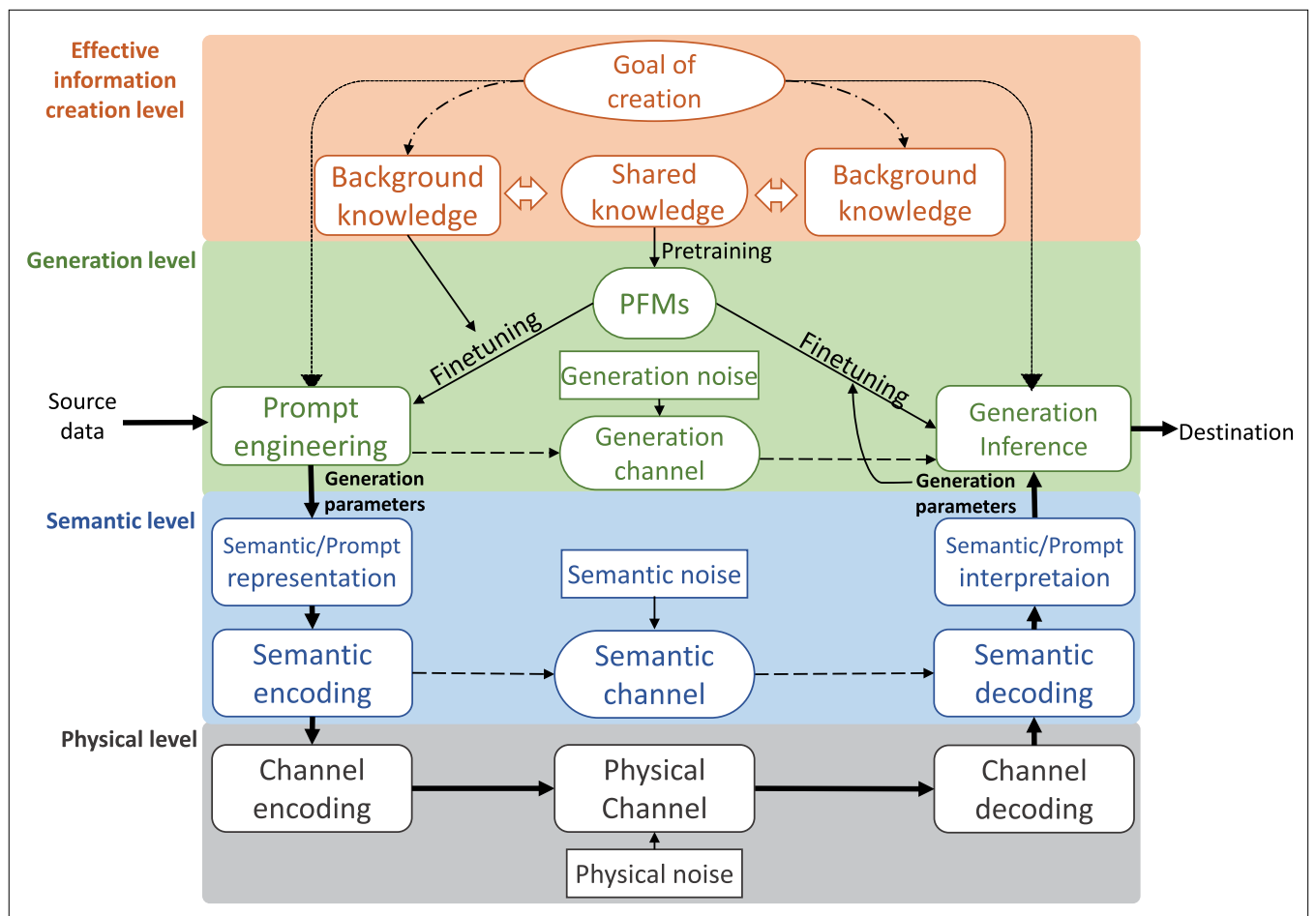


FIGURE 1. SemCom-enabled AIGC model, which is divided into physical level and semantic level for SemCom to support generation level and effective information creation level.

effectiveness and relevance of content generation [5].

- **Goal of Creation:** This component represents the specific objective or intention behind content creation. It guides the generative process to ensure that the output aligns with the desired outcome. The generation level has several integral components, including:

- **PFMs:** As the primary source of content generation, PFMs are pre-trained on extensive datasets [5]. This pretraining process endows them with considerable knowledge and capabilities, allowing them to be fine-tuned for specific tasks. Thus, they are capable of generating diverse, high-quality content such as images, stories and code.

Elaborating further on PFMs, it is important to understand that these models assimilate the underlying patterns from the vast amounts of training data. This wide-ranging and extensive training allows them to generate a rich variety of content, providing the basis for generating content in a variety of contexts.

- **Prompt Engineering:** The process of creating prompts to guide AI model generation forms the crux of prompt engineering. By optimizing and customizing these prompts, the desired content output can be achieved. Prompt engineering plays a pivotal role in steering the generative process, resulting

in more accurate and effective content generation.

The effectiveness of Prompt Engineering is determined by how well it aligns with the information generation task and the desired outcome. It necessitates a comprehensive understanding of the generative model and the task at hand to produce precise prompts that yield the most desirable output.

- **Generation Inference:** This process involves content recreation at the receiver end, integrating and applying the generated content in the pertinent context.
- **Generation Channel:** A virtual medium that facilitates content generation. This channel enables the transfer of creativity from the transmitter to the receiver.
- **Generation Noise:** Representing potential inconsistencies or errors arising during the generative process [5], this noise can lead to mismatches in the quality, values and utility of the generated content between the transmitter and receiver.

PRACTICAL APPLICATIONS AND INSIGHTS

The AIGC framework offers a significant advantage in terms of personalization, a feature that sets it apart from traditional SemCom. While traditional SemCom often remains constrained to transmitting generic semantic information, the SemCom-enabled AIGC framework exhibits

the ability to tailor the generation of content to specific scenarios or user preferences such as personal driving assistant and smart healthcare [2].

To illustrate, consider a news organization employing the AIGC framework to generate digital content. Traditional SemCom model may merely transmit a generic scene of a car accident. However, the AIGC framework empowers the news organization (the receiver) to modify this scene according to the context of their report. The organization can integrate specific elements such as a particular street scene or additional prompts such as “crowded street,” thereby rendering the content more relevant and impactful for their audience, e.g., through personal driving assistant services. The framework thereby facilitates greater receiver control over the content, enhancing the value and applicability of the generative process.

Transitioning to data security considerations, the AIGC framework’s feature of customization at the receiver’s end becomes particularly advantageous. As transmitters only send semantic information and prompts, the receiver’s specific adjustments remain undisclosed, significantly reducing the potential for inadvertent exposure of sensitive information. Thus, the AIGC framework proves especially suitable for applications requiring a high degree of privacy and data security, e.g., smart healthcare. Furthermore, the AIGC framework’s potential to introduce generation noise actually contributes to its overall robustness. Although generation noise originates from inconsistencies within the AIGC model, service providers employing precise prompts and meticulous model fine-tuning techniques can mitigate these inconsistencies, yielding more reliable and high-quality generation results. Therefore, despite the inherent complexities of generation-based communication, the AIGC framework represents a robust mechanism, proficient in ensuring secure and personalized digital content generation and transmission.

CONTROLLABLE AI GENERATED CONTENT SERVICES

In this section, we explore the latest AIGC advancements, focusing on image generation from text prompts. We address how models such as ControlNet increases user control in the content generation process. Furthermore, we propose the application of controlled diffusion as generation encoder/decoder in the aforementioned generation framework, outlining the role of prompt engineering, service validation, receiver generation inference, and highlight how this shift improves data security and user experience.

AIGC IN IMAGE CONTENT GENERATION

Our research primarily focuses on image content generation from text prompts in AIGC services, a recent development with transformative potential.

- **DALL-E:** Developed by OpenAI, DALL-E is a state-of-the-art AI model that demonstrates the capability to generate a wide array of images from text prompts [6]. This model represents a major step forward in translating abstract textual concepts into concrete visual imagery.
- **VQGAN+CLIP:** In subsequent research, CLIP, a joint text-image encoder from OpenAI, was combined with a variant of

Generative Adversarial Network (GAN) known as Vector Quantized GAN (VQGAN).² This integration results in the capability of generating complex and creative images from text prompts.

- **ControlNet+Stable Diffusion:** Similar to DALL-E, Stable diffusion generates images from textual prompts. However, ControlNet enhances the user experience by providing more control over the generated output [7]. By enabling modifications to the descriptive text during the generative process, ControlNet allows for more precise and interactive image creation.

These models, with their emphasis on high-resolution image synthesis, are revolutionizing the industry, lessening the reliance on individual skills and years of experience traditionally considered essential in this field. Therefore, they continue to progress and already significantly impact various domains, including advertising, game development and film effects. This transformative potential is paving the way for new standards and workflows.

ENHANCED CONTROL IN GENERATIVE AI

One of the most significant challenges faced by generative models has been their limited controllability. Traditional methods have primarily relied on brute force techniques, involving linear combinations of prompts and the mass production of images. While this approach provides users with a broad scope of outputs, it also faces many issues, primarily rooted in inefficiency and a lack of precision.

However, the advent of ControlNet [7] signals a transformative shift in AI-driven creation. ControlNet is adeptly integrated into the Stable Diffusion model, which is a large text-to-image diffusion model based on a U-Net architecture.

By using a unique dual-parameter mechanism involving a locked and trainable copy of parameters, ControlNet ensures efficient learning while preventing overfitting. Moreover, it utilizes zero convolution layers as a key technique, providing an efficient connection between different network blocks. Even though these layers start with zero weights, they adapt during the training phase, contributing to the network’s output and enhancing control. This structure allows for layer-wise manipulation of the Stable Diffusion model, adding controllability to the image generation process.

ControlNet allow users to interactively guide the image generation process. By adding control over the neural network’s behavior by manipulating the input conditions of network blocks, ControlNet provides another modality of control in AI-based content creation. In essence, ControlNet improved control and precision capabilities. For example, a generated dancing scene can be controlled with the dedicated pose extracted from a guiding image.

CONTROLLED DIFFUSION AS THE GENERATION ENCODER/DECODER

Interestingly, the shift towards controlled generation aligns closely with the core principles of the proposed framework. SemCom inherently involves the extraction and construction of information, paralleling the controlled generative process as exemplified by controlled diffusion technologies. This characteristic alignment underscores the

² <https://github.com/nerdyrodent/VQGAN-CLIP>, accessed July 06, 2023.

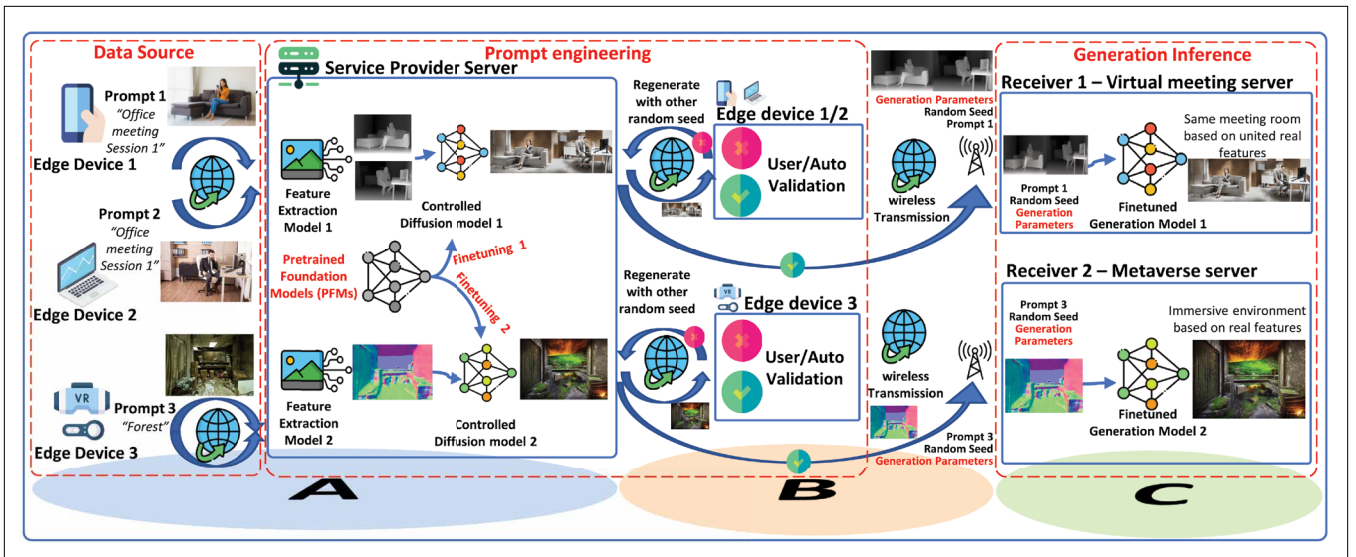


FIGURE 2. AIGC model as generation encoder and decoder. In this AIGC framework, edge devices 1 and 2 represent users in a professional meeting. This step fuses their **Semantic/prompt representation** through **Prompt engineering**, guiding AI models to generate a unified modern meeting room theme in **Generation inference**. On the other hand, edge device 3 applied this framework in VR to generate immersive environments and to synchronize with metaverse server, addressing VR bounding issues by automatically adapting to the user's surroundings environments.

potential synergy between generative AI models and SemCom, indicating the vast opportunities present for their integration. Given the idea, The data flow across the proposed information generation and communication can be described through the following steps:

Step 1. Prompt Engineering: As illustrated in Fig. 2 Part A, the transmission process commences with data collection by the edge devices, which can encompass various user interfaces such as sensors and smartphones. These devices gather raw user inputs, ranging from images and text prompts to userspecific data like the service type that ultimately defines the generation prompt, and transmit this data to a server. This server is a specialized unit designed for feature extraction which transforms the raw data into an abstract semantic representation. The server then employs controlled diffusion models guided by a collection of generation parameters including the random seed and generation steps, among other settings related to the generative process. Subsequently, these models will generate content derived from the given prompts and extracted semantic information. Within this framework, Pretrained Foundation Models (PFMs) play a critical role as a shared knowledge base, accessible to both the transmitter and receiver, thus instead of transmitting entire large pretrained models, only the transmission of the fine-tuning parameters is required. This strategic approach significantly reduces communication overhead and enhances the overall creation efficiency. Furthermore, the procedure is designed to be shareable across multiple edge devices, an attribute that optimizes resource allocation and maximizes system performance. The outcome of this stage is a semantic/prompt representation—an encoded version of the user's input that encapsulates both the extracted features and the generation parameters.

Step 2. Service Validation: As shown in Figure 2 Part B, upon generating the content, the server

conveys it back to the transmitter for validation. This step can either be supervised by the user or automated mechanism through predefined criteria or algorithms such as convolutional neural networks for image recognition. Though it may create more communication overheads, this validation process is integral to the framework, as it not only enhances communication but also enables creative possibilities. The validation process fortifies the reliability and quality of the service, ensuring that the produced images meet the established standards and user requirements. If validation criteria is met, the corresponding semantic/prompt representation is then dispatched for transmission. Conversely, if the content does not meet validation standards, regeneration is initiated with adjustments made to the generation variables. Traditional SemCom model views Parts A and B together as a semantic encoder, transforming user inputs into semantically encoded information. However, within the context of the AIGC framework, this process is interpreted as prompt engineering. This process not only includes the encoding of semantic content but also encompasses the formulation and refinement of generation prompts, thus ensuring a more precise, efficient and user-specific content generation. In addition, the validation process could be adopted in determining the minimal prompt/semantic representation necessary for recreating content. Aligned with Semantic Entropy(SE) [8], such a concept can be termed as "generation entropy". This concept concurs with the process of prompt engineering and ensures maximum overhead reduction during transmission.

Step 3. Receiver Generation Inference: Following successful validation, the semantic/prompt representation is sent to the receiving entity, which can be another edge device or a server, as depicted in Fig. 2 Part C. In contrast to the traditional SemCom model, where this step serves as a semantic decoder, the AIGC framework interprets it as an embodiment of generation

inference, thus underlining the creative aspect of content interpretation from the received information. The versatility of this process is demonstrated across various domains. For example, in gaming, it facilitates the creation of user-specific avatars using the received semantic information. In the realm of virtual reality, this technology addresses the common issue of obstructed vision within VR bounding areas.³ Rather than disrupting the user experience with the necessity of manually drawing a bounding box, the system can regenerate an environment reflecting the user's actual surroundings, thereby enhancing the immersive experience.

In conventional content generation methods, transmission of sensitive user data, such as images or videos, over the Internet potentially risks user privacy. However, the AIGC model operates as an encoder-decoder system for SemCom, transmitting only semantic information and generation variables. This method maintains the transmitted data encrypted and inaccessible to unauthorized entities equipped with different AIGC models and generation parameters, thereby amplifying data privacy and security. For example, eavesdroppers may not have the correct AIGC inference engine, unable to generate the same or similar content as the legitimate receiver. Moreover, this approach transitions power to users, offering a versatile tool in place of a predetermined outcome.

CASE STUDY: JOINT OPTIMIZATION OF SEMCOM AND AIGC

This section presents a case study on the joint optimization of SemCom and AIGC services based on ControlNet [7]. We investigate most suitable semantic extraction methods, the role of evaluation metrics in content quality maintenance and the challenges of joint resource allocation across multiple components, from physical, semantic and generation levels. Through empirical studies and the use of DQN, this section offers insights and solutions for more efficient resource allocation in the multifarious realm of AIGC services.

SEMANTIC EXTRACTION METHODS

Semantic extraction is a crucial component of the transmission framework of AIGC services. The method of semantic extraction depends primarily on the type of service and can vary considerably. As depicted in Fig. 3, several common semantic information such as depth maps, Multi-Instance Depth via Attention Sampling (MiDaS), segmentation, pose estimation, Holistically-Nested Edge Detection (HED) and Canny edge detection have been employed in our proposed transmission framework. Depth maps typically measure the distance between the imaging sensor and each pixel's corresponding real-world point, resulting in a grayscale image representation. Meanwhile, MIDAS uses a novel attention mechanism to produce high-quality depth maps from 2D images [9]. On the other hand, segmentation separates an image into regions or objects for further analysis. For pose estimation, which interprets the human form in an image or video, recognizing human posture and outputting skeleton position information. Moreover, HED leverages the power of deep

learning to predict the presence and location of edges in an image, providing a structural layout. In contrast, Canny edge detection is a multi-stage algorithm used to detect a wide range of image edges, offering a binary output that emphasizes the boundaries of objects within an image [10]. These examples represent a fraction of the broad, flexible range of methods to obtain semantic representation (As discussed in Section II) that services can deploy.

A case in point is when a service pertains to human figures, where pose estimation semantic information have the potential to retain substantial information. This allows for satisfactory reconstruction results at the receiver's end. In our experimental studies, we found that the final output only diverged significantly from the original text prompt, "a man sitting in a modern meeting room" when the semantic information is subjected to a downscaling factor of 10. This observation suggests that even under significant compression, the chosen semantic (pose) preserved essential information effectively.

However, the selection of semantic information may only sometimes align with the service type or user requirements. For instance, another scenario outlined in Fig. 3 used a depth map as the semantic information for a different service. In this case, the maximum permissible downscaling factor without noticeable loss of information is only 4, corresponding to image compression of 16 times. Therefore, the degree of compression semantic information can be adapted without impacting quality depends heavily on the chosen semantic information. This highlights the importance of carefully selecting the most suitable semantic extraction method based on service type and user requirements.

EVALUATION METRICS

The proposed methodology for identifying the most suitable semantic extraction method and corresponding evaluation metrics for a particular service is illustrated in Fig. 4, using four service examples.

For Services 1 and 3, the process begins by selecting an evaluation metric aligning with the provider's quality assessment needs. This step is followed by identifying a semantic extraction method that pairs effectively with the chosen metric, particularly the semantic extraction method and evaluation metric pair that is providing a predictable and stable linear response to various levels of image degradation. This approach ensures predictable and maintainable content quality, even under diverse degrees of downscaling. On the contrary, Service 4 employs a reversed process where the service provider selects the most suitable semantic extraction method that aligns with the service type or user requirements and subsequently identifies an evaluation metric offering a predictable linear response to the chosen extraction method.

Service 2 presents a different scenario, where the service provider simultaneously defines both the extraction method and evaluation metric. This approach is advantageous when service requirements are well-understood and quality standards are clearly established.

This study utilizes a variety of established evaluation metrics, including Mean Squared

³ <https://github.com/Zetaphor/webxr-environment-mapper>, accessed July 28, 2023.

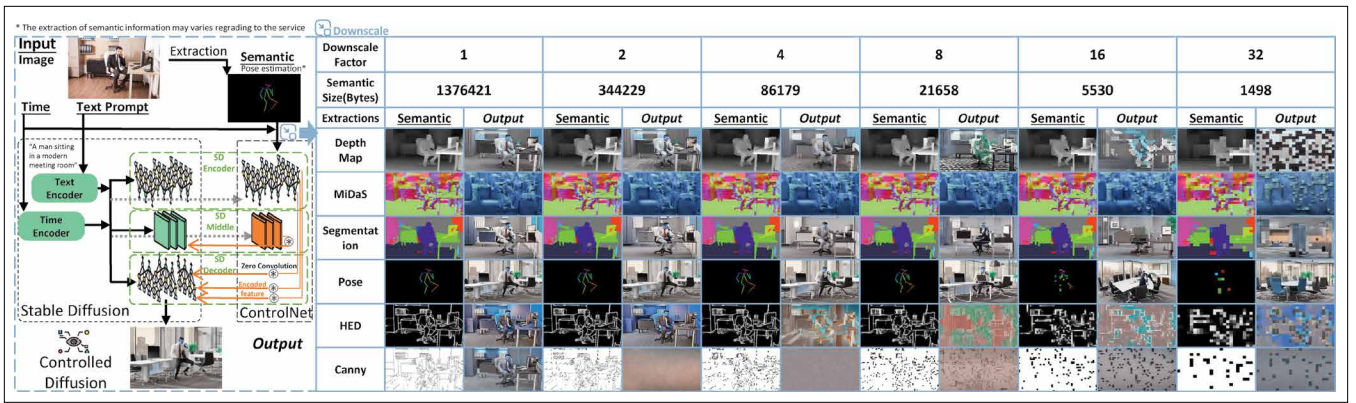


FIGURE 3. Semantic variation and downscaling impact in generative task. The figure illustrates ControlNet's structure and where downscaling in semantic information could be deployed. Different extraction models show varied scaling levels, reducing transmitted data. Each method exhibits different sensitivity to downscaling, highlighting the need for flexible resource allocation. Tailoring the downscale factor to the specific model ensures efficient communication by prioritizing less sensitive to more sensitive methods.

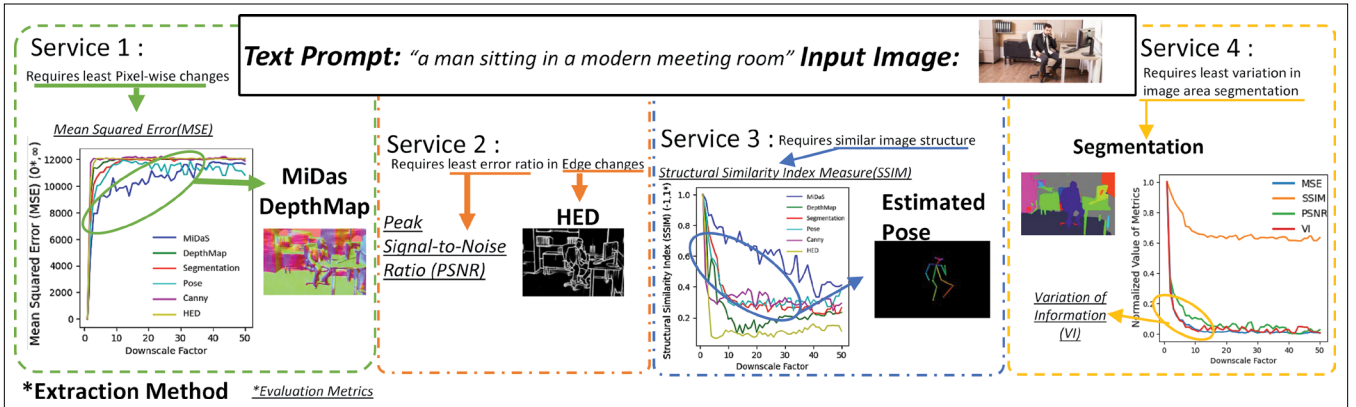


FIGURE 4. Workflow for defining predictable pairs of semantic extraction methods and evaluation metrics. Service 2 presents a traditional case where a service provider defines both the extraction method and evaluation metric. Services 1 and 3 illustrate the process of selecting the most suitable semantic extraction method when the service provider defines the evaluation metrics. Conversely, Service 4 demonstrates the process of choosing the most appropriate evaluation metrics when the semantic extraction method is defined. This iterative workflow ensures the identification of most predictable combinations for effective semantic extraction and evaluation in the given context.

Error (MSE), Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) and Variation of Information (VI) [11]. MSE measures the average pixel-wise squared difference between the generated images with and without semantic downscaling. PSNR compares the maximum possible power of the original signal to the power of the noise that affects its quality. SSIM assesses structural similarity which provides insight into perceived changes in structural information and VI quantifies the change in shared information, reflecting the effectiveness of semantic content transmission. Although other metrics, such as Inception Score (IS), Fréchet Inception Distance (FID) [12] and Content Loss [13] could be considered, this study concentrates on a selected group of common metrics for simplicity and clarity.

In summary, selecting the most predictable pair of semantic extraction methods and evaluation metrics can be complex and service-specific. However, a systematic approach that considers the provider's needs and the pair's linear response under compression can yield predictable image quality.

MULTI SERVICE JOINT RESOURCE ALLOCATION

Following the selection of distinct semantic extraction methods and evaluation metrics tailored to individual service, a complex dynamic emerges when accommodating multiple services' transmission requirements concurrently. To provide a comprehensive understanding of our framework's deployable environment, we detail the specifications of the equipment specification used in our experiments:

- CPU: AMD Ryzen Threadripper PRO 3975WX 32-Cores
- GPU: NVIDIA RTX A5000

The scenario, as represented in Fig. 4, involves four services, each possessing unique evaluation metrics and semantic extraction methods. For the purpose of maintaining consistency across varying metrics, all evaluation metrics were normalized to fall within a range of zero to one. Here, a value of one denotes an exact match between the original and regenerated image, indicating no loss of information or quality. In this study, we employed Deep Q-Learning Network (DQN) as a solution technique to assess the feasibility of joint resource

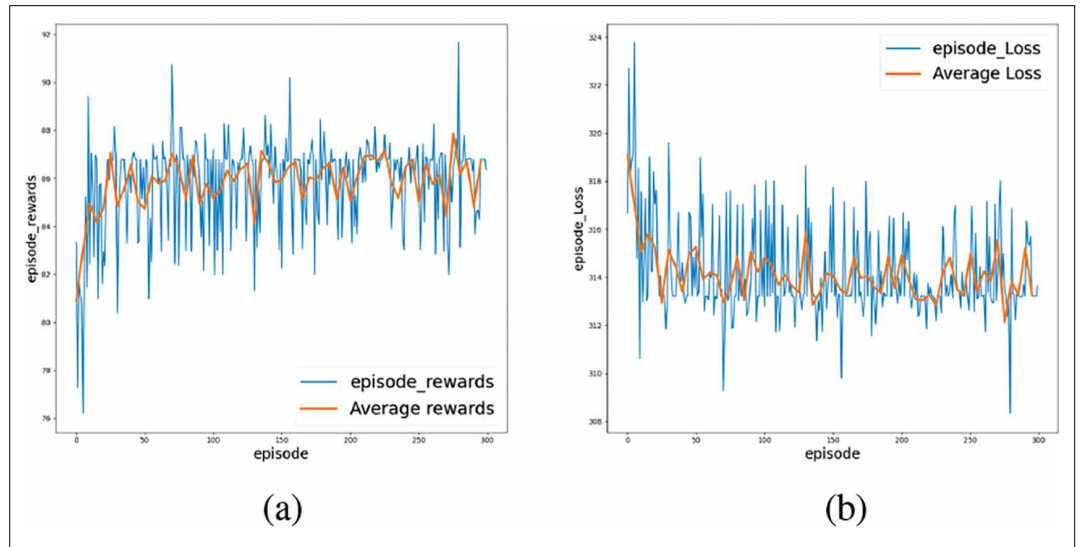


FIGURE 5. Rewards and down-sampling losses over episodes. a) Reward analysis. b) Loss analysis.

allocation. In the DQN model, the action space is defined by the downscaling factor for each service, while the reward signal is measured based on the evaluation results.

The outcomes of the DQN optimization are depicted in Fig. 5. As the results indicate, the optimization process converges after approximately 50 epochs. This convergence is significant as it indicates the attainment of the most efficient resource allocation scheme, which is essential for maximizing the efficiency and effectiveness of our AIGC framework. The loss is calculated by comparing the quality of images generated with original and down-scaled semantics, reflecting how resource allocation schemes affects image quality. To enhance system performance, the reward is inversely modeled related to loss, guiding the DQN to efficiently allocate resources while preserving image quality by reducing loss. The substantial fluctuation observed within this process is a likely consequence of the escalated complexity of image restoration when the downscaling factor increments. Specifically, an increased downscaling factor signifies a greater division of the image's width and height. This substantial partitioning invariably complicates restoring the original semantic information, thereby influencing the quality of the final generated image.

Nonetheless, despite these challenges, the results validate the feasibility of the proposed framework for joint resource allocation. The approach demonstrates adaptability to the diverse range of requirements inherent in contemporary and future AIGC services. Therefore, this framework holds significant potential for widespread deployment, facilitating more efficient and optimized resource allocation within the diverse landscape of AIGC services.

In Section I, we highlighted AIGC + SemCom's capability to manage complex content generation, which is later substantiated by our experiments in Section IV-B. Our model adeptly managed resource allocation for four distinct services, showcasing its effectiveness in complex scenarios that align with our theoretical assertions about

the AIGC + SemCom framework's practicality in real-world applications. Additionally, using the DQN is pivotal in determining the most suitable allocation schemes. This strategy efficiently down-scales semantics for transmission with the least effect on generated image quality, demonstrating efficient and high-quality content creation. Such results affirm the framework's efficiency and quality standards.

FUTURE RESEARCH DIRECTIONS

In this section, we present future research directions related to the integration of SemCom and AIGC services:

- **Development of Universal Evaluation**

Metrics: Current AIGC services present a challenge in creating universal evaluation metrics, as different services have varying requirements [2]. A comprehensive metric that quantifies performance across various AIGC services can be developed to provide a consistent performance measure independent of the services. This would enable easier comparison across different services and provide a better understanding of performance trade-offs in various AIGC applications.

- **Exploration of Low-rank Adaptation (LoRa):**

Investigate the application of LoRa, a method that condenses large language models into smaller, efficient versions without significant performance loss [14]. By leveraging LoRa's capabilities to optimize image creation and finetuning, it is possible to enhance the transmission and decoding of semantic information in AIGC services. This approach can improve the accessibility and usability of AIGC services in resource-limited environments by providing streamlined and efficient fine-tuning of PFMs, making the process more standardized and effective.

- **Development of Lightweight AIGC Services:**

Adapting AIGC services to be more lightweight and efficient for mobile edge computing can be a potential direction. This direction focuses on optimizing AIGC

algorithms to ensure they are resource-efficient and can operate effectively in the constrained environments typical of edge devices for a better quality of the service [15]. Such a method aims to enhance real-time data processing and content generation capabilities in edge computing scenarios, which could contribute to advancements in smart transportation and connected vehicle technologies.

CONCLUSION

This article has presented a conceptual model for integrating SemCom into AIGC services, featuring an additional generation level for content creation. Our analysis has revealed how modifications in input information extraction, like image resolution and compression, can impact AIGC output. The introduced evaluation framework highlighted AIGC's role in semantic information transmission and facilitated optimized resource allocation. Validation through a Deep Q-Network confirmed the feasibility of our framework, promising significant efficiency improvements for AIGC services.

ACKNOWLEDGMENT

This work was supported in part by the National Research Foundation, Singapore, and Infocomm Media Development Authority through the Future Communications Research and Development Programme, DSO National Laboratories, with in the AI Singapore Programme (AISG), under Award AISG2-RP-2020-019 and Award FCP-AS-TAR-TG-2022-003; 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 RG8722; in part by NSFC under Grant 62102099 U22A2054; in part by the Pearl River Talent Recruitment Program under Grant 2021QN02S643; in part by the Guangzhou Basic Research Program under Grant 2023A04J 1699; in part by the Ministry of Science and ICT (MSIT), Korea, through the ICT Creative Consilience Program supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP) under Grant IITP-2020-001821; in part by the National Research Foundation (NRF) and Infocomm Media Development Authority through the Future Communications Research Development Programme (FCP); in part by the SUTD under Grant SRG-ISTD-2021165; and in part by the Ministry of Education, Singapore, under SMU-SUTD Joint Grant 22-SIS-SMU048.

REFERENCES

- [1] Y. Cao et al., "A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT," 2023, *arXiv:2303.04226*.
- [2] H. Du et al., "Enabling AI-generated content (AIGC) services in wireless edge networks," 2023, *arXiv:2301.03220*.
- [3] Y. Lin et al., "A unified framework for integrating semantic communication and AI-generated content in metaverse," 2023, *arXiv:2305.11911*.
- [4] W. Yang et al., "Semantic communication meets edge intelligence," *IEEE Wireless Commun.*, vol. 29, no. 5, pp. 28–35, Oct. 2022.
- [5] C. Zhou et al., "A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT," 2023, *arXiv:2302.09419*.

- [6] A. Ramesh et al., "Zero-shot text-to-image generation," in *Proc. 38th Int. Conf. Mach. Learn.*, vol. 139, M. Meila and T. Zhang, Eds., Jul. 2021, pp. 8821–8831. [Online]. Available: <https://proceedings.mlr.press/v139/ramesh21a.html>
- [7] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Feb. 2023, pp. 3836–3847.
- [8] L. Yan et al., "QoE-aware resource allocation for semantic communication networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2022, pp. 3272–3277.
- [9] S. Fuster, T. Eftestøl, and K. Engan, "Nested multiple instance learning with attention mechanisms," 2021, *arXiv:2111.00947*.
- [10] S. Xie and Z. Tu, "Holistically-nested edge detection," 2015, *arXiv:1504.06375*.
- [11] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
- [12] Y. Benny et al., "Evaluation metrics for conditional image generation," *Int. J. Comput. Vis.*, vol. 129, no. 5, pp. 1712–1731, May 2021.
- [13] J. Chen et al., "Learning to evaluate the artness of AI-generated images," 2023, *arXiv:2305.04923*.
- [14] E. J. Hu et al., "LoRA: Low-rank adaptation of large language models," 2021, *arXiv:2106.09685*.
- [15] S. Verma, Y. Kawamoto, and N. Kato, "Energy-efficient group paging mechanism for QoS constrained mobile IoT devices over LTE-A pro networks under 5G," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 9187–9199, Oct. 2019.

BIOGRAPHIES

GUANGYUAN LIU (liug0022@e.ntu.edu.sg) is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Energy Research Institute @ NTU, Interdisciplinary Graduate Program, Nanyang Technological University, Singapore. His research interests include generative AI, semantic communications, and Metaverse.

HONGYANG DU (hongyang001@e.ntu.edu.sg) is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Energy Research Institute @ NTU, Interdisciplinary Graduate Program, Nanyang Technological University, Singapore. His research interests include generative AI, semantic communications, and communication theory. He was the recipient of IEEE Daniel E. Noble Fellowship Award in 2022.

DUSIT NIYATO (Fellow, IEEE) (dniyato@ntu.edu.sg) received the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada, in 2008. He is a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests are in the areas of sustainability, edge intelligence, decentralized machine learning, and incentive mechanism design.

JIAWEN KANG (kavinkang@gdut.edu.cn) received the Ph.D. degree from the Guangdong University of Technology, China, in 2018. He has been a Post-Doctoral Researcher with Nanyang Technological University, Singapore, from 2018 to 2021. He is currently a Full Professor with the Guangdong University of Technology, China. His research interests focus on blockchain, security, and privacy protection.

ZEHUI XIONG (zehui_xiong@sutd.edu.sg) received the Ph.D. degree from Nanyang Technological University (NTU), Singapore. He is an Assistant Professor with the Singapore University of Technology and Design, and also an Honorary Adjunct Senior Research Scientist with the Alibaba-NTU Singapore Joint Research Institute, Singapore. His research interests include wireless communications, blockchain, edge intelligence, and Metaverse.

DONG IN KIM (Fellow, IEEE) (dikim@skku.ac.kr) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1990. He is a Professor with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea. His research interests include the Internet of Things, wireless power transfer, and connected intelligence.

XUEMIN (SHERMAN) SHEN (Fellow, IEEE) (sshenn@uwaterloo.ca) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on network resource management, network security, the Internet of Things, 5G, and beyond.