Image Steganography For Securing Intellicise Wireless Networks: "Invisible Encryption" Against Eavesdroppers

Bizhu Wang, Song Gao, Rui Meng, *Member, IEEE,* Haixiao Gao, Xiaodong Xu, *Senior Member, IEEE,* Mengying Sun, *Member, IEEE,* Chen Dong, Ping Zhang, *Fellow, IEEE,* and Dusit Niyato, *Fellow, IEEE*

Abstract—As one of the most promising technologies for *intellicise (intelligent and consice)* wireless networks, Semantic Communication (SemCom) significantly improves communication efficiency by extracting, transmitting, and recovering semantic information, while reducing transmission delay. However, an integration of communication and artificial intelligence (AI) also exposes SemCom to security and privacy threats posed by intelligent eavesdroppers. To address this challenge, image steganography in SemCom embeds secret semantic features within cover semantic features, allowing intelligent eavesdroppers to decode only the cover image. This technique offers a form of "invisible encryption" for SemCom. Motivated by these advancements, this paper conducts a comprehensive exploration of integrating image steganography into SemCom. Firstly, we review existing encryption techniques in SemCom and assess the potential of image steganography in enhancing its security. Secondly, we delve into various image steganographic paradigms designed to secure SemCom, encompassing three categories of joint source-channel coding (JSCC) models tailored for image steganography SemCom, along with multiple training strategies. Thirdly, we present a case study to illustrate the effectiveness of coverless steganography SemCom. Finally, we propose future research directions for image steganography SemCom.

Index Terms—Intellicise wireless networks, semantic communications, image steganography, anti-eavesdropping.

I. INTRODUCTION

The *intellicise (intelligent and concise)* wireless network is an innovative paradigm designed to address complex challenges in modern communication systems by harmonizing intelligence and simplicity [1]. Rooted in the integration of multidisciplinary theories, including information theory, artificial intelligence (AI), complex science, and system theory, it employs advanced methodologies such as basic probability,

Bizhu Wang, Song Gao, Rui Meng, Haixiao Gao, Mengying Sun, Chen Dong, and Ping Zhang are with Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: wangbizhu_7@bupt.edu.cn; wkd251292@bupt.edu.cn; buptmengrui@bupt.edu.cn; haixiao@bupt.edu.cn; smy_bupt@bpt.edu.cn; dongchen@bupt.edu.cn; pzhang@bupt.edu.cn).

Xiaodong Xu is with Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with Peng Cheng Laboratory, Shenzhen 518066, Guangdong, China (e-mail: xuxiaodong@bupt.edu.cn).

Dusit Niyato is with College of Computing and Data Science, Nanyang Technological University, Singapore (email: dniyato@ntu.edu.sg).

fuzzy measurement, and logical reasoning to explore fundamental principles of information and networks. As one of the most promising technologies for intellicise wireless networks, Semantic Communication (SemCom) is an emerging communication paradigm that focuses on the semantics of the data itself during information transmission rather than the precise transmission of bit streams. By comprehending, compressing, and conveying the inherent meaning of information, it reduces redundant data transmission, thus enhancing communication efficiency and supporting more intelligent interactions [2], [3].

However, with the development of SemCom, a series of complex security and privacy threats that traditional communication systems have never encountered have emerged. Compared with the security challenges faced by traditional wireless communication, such as eavesdropping attacks and jamming attacks, SemCom is confronted with more severe security threats due to the deep integration of AI and communication technologies [2], such as semantic knowledge base poisoning [4], gradient leakage [5], and intelligent eavesdropping [2].

To effectively address the security and privacy threats in SemCom, researchers have explored numerous defense technologies at the levels of model training, model transfer, and semantic information transmission [3]. Data cleansing ensures the purity of model training by detecting and removing malicious samples from the semantic knowledge base [4]. During the model training process, adversarial samples are proactively introduced to enhance the model's robustness against semantic adversarial attacks [5]. Differential privacy-based noises are added to semantic data to prevent attackers from inferring individual information from the output results [5]. Cryptography converts plaintext semantic data into ciphertext semantic data to safeguard confidentiality and integrity [6]. Blockchain achieves the immutability and traceability of semantic data through distributed ledgers and smart contracts [3]. Model compression simplifies the model structure through techniques such as pruning, quantization, and knowledge distillation, indirectly improving security [4]. Physical layer security leverages the physical characteristics of wireless channels to provide endogenous semantic security guarantees [6].

In addition to the aforementioned defense methods, image steganography [7]–[9] offers a promising approach for securing SemCom by embedding secret images into cover images, effectively concealing encryption traces. This technique, often referred to as "invisible encryption," holds significant potential for enhancing the security of SemCom systems. By integrating

steganography, the proposed secure SemCom approach by Tang et al. [10] effectively safeguards private image data during transmission, leveraging Invertible Neural Networks (INNs) to embed private images within host images. The method ensures high reconstruction quality for authorized receivers while denying eavesdroppers access to sensitive content. Building on this, Tang et al. [11] introduced novel intelligent eavesdroppers and demonstrated that, the INN-based steganography module can reduce eavesdropping success rates to 0%, while maintaining data quality for legitimate semantic receivers. Additionally, Huo et al. [12] employed Generative Adversarial Networks (GANs) to embed secret semantic features into cover semantic features, ensuring reliable extraction and maintaining covertness throughout SemCom. This approach not only preserves the integrity of SemCom but also strengthens security by addressing distortions at both pixel and semantic levels. Despite these advancements, research on image steganography-based SemCom is still in its nascent stages. The full potential of leveraging image steganography techniques to secure SemCom has yet to be thoroughly explored. Several key questions remain unanswered:

- Q1: What makes image steganography particularly suitable for securing SemCom?
- Q2: How can image steganography be effectively integrated with SemCom to defend against eavesdroppers?
- Q3: What challenges arise when applying image steganography to SemCom?

In light of these questions, we conduct a review of image SemCom encryption approaches, evaluating the feasibility of applying image steganography to SemCom. The main contributions of this work are as follows:

- We revisit existing image SemCom encryption schemes and present an overview of image steganography techniques, subsequently analyzing the potential of image steganography in securing SemCom.
- We detail image steganography paradigms for securing SemCom, comparing Convolutional Neural Network (CNN)-based, GAN-based, and INN-based image steganography joint source-channel coding (JSCC) models, and listing image semantic steganography strategies. Moreover, we present a case study to show its effectiveness.
- We envision future research directions for image steganography SemCom.

II. OVERVIEW OF IMAGE SEMCOM ENCRYPTION SCHEMES AND IMAGE STEGANOGRAPHY TECHNIQUES

A. Image SemCom Encryption

1) Why Encrypt SemCom?: Encryption is crucial for ensuring the security of SemComs, with its necessity as follows:

- **Protection Against Eavesdropping and Privacy Leakage:** Wireless channels possess broadcasting characteristics, allowing any device within the coverage area to intercept signals. Compressed semantic feature vectors may expose critical information, and eavesdroppers can reconstruct original sensitive data using feature inversion algorithms. A more insidious risk lies in the implicit associations within semantics: for instance, unencrypted semantic reports transmitted by medical IoT devices may be linked to metadata such as timestamps and location information, enabling precise inference of private information like individual identities and health statuses [3].
- **Prevention of Malicious Attacks and Intent Tampering:** SemCom directly conveys the logical meaning of information. If unencrypted, attackers can launch precise semantic-layer attacks. For example, in smart home scenarios, tampering with the semantics of device control commands, such as replacing "turn off the air conditioner" with "continue cooling," leading to energy waste. Besides, in autonomous driving systems, falsified environmental semantic information, such as altering "road collapse ahead" to "lanes clear," attracts vehicles into dangerous areas. Furthermore, attackers could interfer with the semantic information decoding at the legitimate receiver by fine-tuning unencrypted semantic vectors.

2) How to Encrypt SemCom?: Table I lists existing encryption techniques for image SemCom and representative research papers, including cryptography-based SemCom, covert SemCom, physical-layer encrypted SemCom, and application-layer encrypted SemCom.

B. Image Steganography

1) What is Image Steganography?: In today's digital era, the demand for data security has driven the development of security systems: cryptography systems and information hiding systems. Cryptography systems focus on the unreadability of information, while information hiding systems emphasize concealing data to create the appearance of no secret data existing. Data hiding systems encompass watermarking and steganography techniques. Watermarking aims to prove ownership or integrity of carriers, or enables tamper-proof authentication. In contrast, steganography prioritizes hiding the existence of secret information, ensuring that third parties remain unaware of the presence of secret data [8]. The basic architecture of image steganography is illustrated in Figure 1a.

2) How to Image Steganography?: Image steganography is realized through embedding processes, which are mainly divided into two categories:

- **Traditional Image Steganography:** According to processing domains, it can be categorized into spatial domain and transform domain. Spatial domain methods utilize manually designed hiding rules and distortion functions to perform operations in the pixel domain, enabling the embedding and extraction of secret information. Transform domain methods transform images from the pixel domain to the frequency domain and embed secret information into high-frequency regions to minimize interference with the original cover image. Transform methods involve Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), integer transforms, and so on [7].
- **Deep Learning (DL)-Based Image Steganography:** With advancements in DL-based steganalysis techniques, traditional image steganography methods face increasing

TABLE I: Image SemCom encryption methods and representative research papers, where AES: Advanced Encryption Standard; PLK: Physical-Layer Key; FDD: Frequency Division Duplexing; RIS: Reconfigurable Intelligence Surface; SSR: Secure Semantic Rate; SSSE: Secure Semantic Spectrum Efficiency. **Cryptography-based SemCom:** The compatibility of encryption in complex systems where bits and semantics coexist is worth considering. Since cryptography is used in traditional communication systems, expanding cryptography-based SemCom methods is an important research direction. This includes traditional cryptography, homomorphic encryption, and quantum cryptography. **Covert SemCom:** Compared to information encryption methods, covert communication can hide the SemCom behavior between legitimate transmitters and receivers, achieving encryption from another perspective. The combination of covert communication and SemCom has broad applications, such as in military communications. **Physical-Layer Encrypted SemCom:** Physical-layer encryption mainly utilizes the characteristics of wireless channels and designs appropriate technologies from the physical-layer to enhance the security of SemCom systems. This includes information-theoretic security, RIS-aided security, PLK-aided security, and beamforming-aided security. **Application-Layer Encrypted SemCom:** Compared to physical-layer encryption, application-layer encryption can customize and optimize encryption schemes according to specific application requirements to ensure the confidentiality and integrity of semantic information. This includes cross-layer encryption and neural network-based encryption.

Encryption Methods		Descriptions
Cryptography-based SemCom	Traditional Cryptography	• *Deep joint source-channel and encryption coding: Secure semantic communications:* Based on learning with errors, and using the affine property of encryption method to realize privacy security. • *Lightweight and robust wireless semantic communications:* AES encryption algorithm, adversarial identification, and SemCom are combined to resist eavesdropping attacks and adversarial attacks.
	Homomorphic Encryption	• *Secure semantic communication with homomorphic encryption:* The frequency of key update can be adjusted according to service requirements without affecting transmission performance.
	Quantum Cryptography	• *Quantum semantic communications for metaverse: Principles and challenges:* The quantum semantic layer can ensure the privacy of semantic information through public keys generated by quantum key distribution. • *Cooperative resource management in quantum key distribution (QKD) networks for semantic communication:* Analysis of the cooperation of quantum key distribution service providers. • *Integrating sustainable big AI: Quantum anonymous semantic broad-cast:* Semantic extraction is realized by large models, and semantic information is broadcast by quantum anonymous communication protocol.
Covert SemCom		• *Multi-agent reinforcement learning for covert semantic communications over wireless networks:* Jointly maximizing semantic information transmission and power control to ensure the quality and security. • *Unveiling covert semantics: Joint source-channel coding under a covertness constraint:* Verifing the feasibility of training neural network to realize covert SemCom for classification tasks.
Physical-Layer Encrypted SemCom	Information Theory Security	• *A nearly information theoretically secure approach for semantic communications over wiretap channel:* Ensuring that the eavesdropper's symbol error probability is equal to random guess.
	RIS-aided Security	• *Star-RIS-assisted privacy protection in semantic communication system:* Designing the reflection coefficient vector to form the interference at task level and signal-to-noise ratio level in eavesdroppers.
	PLK-aided Security	• *Securing semantic communications with physical-layer semantic encryption and obfuscation:* A semantic ambiguity mechanism to provide physical-layer protection and promote key generation in static environments.
	Beamforming-aided Security	• *Secure design for integrated sensing and semantic communication system:* Maximizing the cumulative semantic secrecy rate of all users, while ensuring the minimum quality of service for each user.
Application-Layer Encrypted SemCom	Cross Layer Encryption	• *IRS-enhanced secure semantic communication networks: Cross-layer and context-aware resource allocation:* Definitions of SSR and SSSE to quantify the semantic security performance of task level. • *Eavesdropping on semantic communication: Timing attacks and countermeasures:* This study found that the privacy protection of SemCom is time-sensitive, and provides theoretical guidance based on Markov.
	Neural Network-based Encryption	• *Deep joint source-channel coding for image transmission with visual protection:* Transforming the original image into a visually protected image. • *The model inversion eavesdropping attack in semantic communication systems:* Introductions of the model inversion eavesdropping attack to reveal the risk of privacy leakage in SemCom system. • *Wireless image transmission with semantic and security awareness:* An autoencoder based on residual module and convolution level is proposed to flexibly adjust the trade-off between efficiency and privacy. • *Disentangled information bottleneck guided privacy-protective joint source and channel coding for image transmission:* A privacy-preserving SemCom method based on decoupling information bottleneck.

security risks. To address this challenge, DL-based image steganography has emerged. This approach primarily adopts an encoder-decoder architecture, where secret information is embedded into images through encoding, and the original secret information and cover image are subsequently recovered using the decoder [8].

C. What Can SemCom Benefit from Image Steganography?

Compared with existing SemCom encryption techniques, image steganography offers the following advantages:

- **Concealing Encryption Traces:** Data processed by traditional encryption methods becomes unreadable, directly revealing the act of semantic encryption. In contrast, steganography can integrate secret information into the semantic feature space, enabling the stego semantic features to convey both normal semantic information and additional secret semantic information, thus achieving the concealment of encryption [12].

- **End-to-End Implementation:** Traditional encryption requires separate encryption and decryption modules, which not only increase system complexity but also introduce additional latency. In contrast, steganography can be deeply integrated with the encoding and decoding architecture of SemCom, allowing the receiver to extract secret semantic information without the need for additional processing steps, thereby improving communication efficiency [10], [11].

- **Defense Against Intelligent Eavesdroppers:** Traditional encryption methods generate ciphertext that, even if it cannot be directly decrypted, allows intelligent eavesdroppers to infer sensitive data through model inversion attacks (MIA) and Generative AI (GAI) techniques. In contrast, steganography embeds secret semantic features
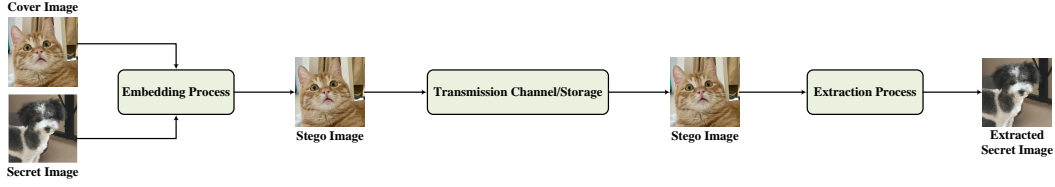
(a) Basic architecture of the image steganography system, where secret images are confidential images that need to be protected or transmitted; cover images appear to be ordinary images on the surface but possess high complexity (rich textures, diverse colors, etc.) to conceal information without causing visual anomalies; and stego images are the cover images in which secret information is embedded [7].
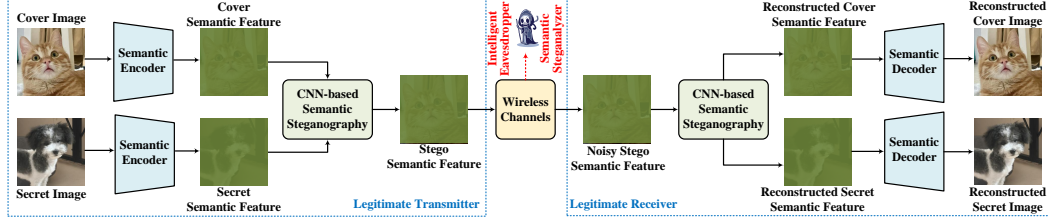
(b) CNN-based image steganography JSCC model, where the semantic steganography modules deployed at the transmitter and receiver are usually the encoder and decoder of variational autoencoders, respectively.
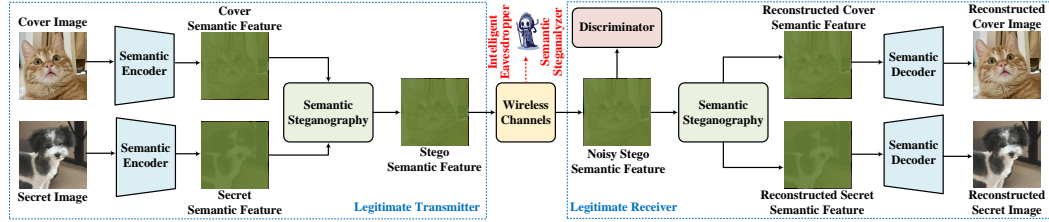
(c) GAN-based image steganography JSCC model, where the adversarial training process forces the semantic steganography to produce undetectable embeddings.
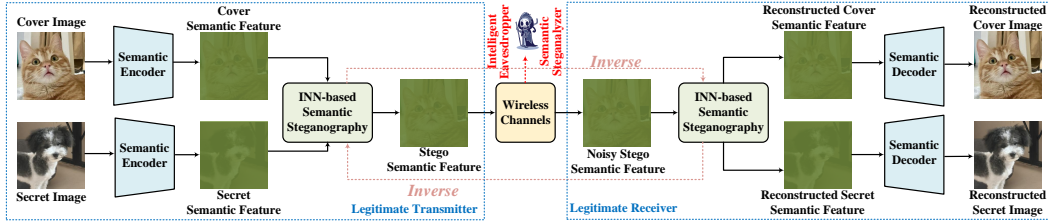
(d) INN-based image steganography JSCC model, where the semantic steganography modules deployed at the transmitter and receiver are inverse to improve recover performance.

Fig. 1: Illustration of image steganography SemCom systems, where (a) presents the basic architecture of the image steganography system, (b) CNN-based image steganography JSCC model has lowest training time and model size but worst secret semantic capacity and transmission quality, (c) GAN-based image steganography JSCC model has the best anti-eavesdropping ability through adversarial training but the most parameters, and (d) INN-based image steganography JSCC model has the highest secret semantic capacity and transmission quality but longest training time.

into cover semantic features without altering the cover image's visual semantics. This makes it difficult for intelligent eavesdroppers to detect the presence of hidden information, let alone extract or infer sensitive content, thereby defending against them [11].

- **High Flexibility and Compatibility:** Steganography exhibits excellent flexibility and can adaptively adjust to the requirements of different scenarios. In bandwidth-constrained situations, steganography can adjust the steganographic capacity to prioritize the quality of semantic transmission. When facing severe security threats, more robust steganography methods can be employed to enhance wireless security. Additionally, steganogra-

phy can complement traditional encryption technologies, forming a "dual protection" mechanism. Even if traditional encryption methods are compromised, attackers still need to perform complex steganalysis to access the secret content, providing more reliable security for SemCom.

III. IMAGE STEGANOGRAPHY PARADIGMS FOR SECURING SEMCOM

A. Image Steganography Joint Source-Channel Coding (JSCC) Models

SemCom is typically implemented via JSCC, which can be integrated with DL-based image steganography models

TABLE II: Comparison of CNN-based, GAN-based, and INN-based image steganography JSCC models.

Image Steganography JSCC Model	Training Time	Model Size	Secret Semantic Capacity	Semantic Transmission Quality	Anti Intelligent Eavesdroppers
CNN-based	**Low**	**Low** (about 0.5 million parameters)	Low	Low	Low
GAN-based	Medium	High (more than 1 million parameters)	Medium	Medium	**High**
INN-based	High	Medium (about 0.6 million parameters)	**High**	**High**	Medium

to achieve secure SemCom. Figures 1b, 1c, and 1d show CNN-based, GAN-based, and INN-based image steganography JSCC models, respectively. Table II compares the above three models, and the detailed descriptions are as follows.

1) Convolutional Neural Network (CNN)-based Image Steganography JSCC: In traditional fully connected neural networks, each neuron connects to all inputs, leading to an explosion in the number of parameters. In contrast, neurons in CNNs only connect to local regions of the input data, imitating the local perception characteristics of human eyes, which greatly reduces the parameter count. Additionally, the same convolutional kernel slides across the entire input data to extract consistent types of features, such as edges and textures. This parameter-sharing mechanism further minimizes parameters while enhancing the model's generalization capability. CNN-based image steganography JSCC typically include the following modules:

- The *Semantic Encoder* extracts high-level semantic information and removes redundant data, thereby converting the cover image and secret image into cover semantic features and secret semantic features, respectively.
- At the transmitter, the *CNN-based Semantic Steganography* module covertly embeds the secret semantic features into the cover semantic features, leveraging CNN's dynamic perception capabilities to select optimal embedding locations, resulting in stego semantic features. Thus, stego semantic features are the cover semantic features where secret semantic information is embedded. At the receiver, the *CNN-based Semantic Steganography* module extracts the reconstructed cover semantic features and reconstructed secret semantic features from the noisy stego semantic features.
- During model training, a noise layer can be introduced to simulate wireless channel effects such as noise, fading, and interference. The stego semantic features pass through the *Wireless Channel* to become noisy stego semantic features.
- The *Semantic Decoder* restores the reconstructed cover semantic features and reconstructed secret semantic features into the reconstructed cover image and reconstructed secret image, respectively.

2) Generative Adversarial Network (GAN)-based Image Steganography JSCC: GAN is a deep learning model that employs adversarial training to pit two neural networks, the generator and the discriminator, against each other. Its objective is to generate synthetic data that closely resembles the distribution of real data, a minimax game $\min_G \max_D V(D, G)$. In this framework: the generator learns to create realistic data from random noise, such as Gaussian distributions, aiming to "deceive" the discriminator ($D(G(z)) \rightarrow 1$), through

minimizing $V(D, G)$; the discriminator acts as a classifier, striving to accurately determine whether input data is a real sample ($D(x) \rightarrow 1$) or a fake sample synthesized by the generator ($D(G(z)) \rightarrow 0$), through maximizing $V(D, G)$. This adversarial interplay drives continuous improvement in both components, enabling the generator to produce highly convincing synthetic outputs. Therefore, in GAN-based image steganography JSCC, the *generator* is designed to produce stego semantic features, while the *discriminator* distinguishes between the cover semantic features encoded by the semantic encoder and the noisy stego semantic features. This adversarial process forces the *generator* to produce undetectable embeddings. Additionally, by modeling the presence of an intelligent eavesdropper via the *discriminator's* detection attempts, this method enhances resistance to semantic steganalysis.

For instance, Huo et al. [12] used two discriminators to enhance both security and semantic extraction accuracy. The first discriminator focuses on distinguishing between cover and stego semantic features, prompting the generator to produce stego features that closely resemble the original cover features. The second discriminator operates after the semantic reconstruction stage, ensuring that even after image reconstruction, the embedded secret information remains secure and the system is resistant to steganalysis. In short, the first discriminator ensures that the stego features are undetectable during semantic transmission, while the second discriminator guarantees that secret information remains hidden even in the reconstructed image, thus enhancing the overall robustness and security of SemCom.

3) Invertible Neural Network (INN)-based Image Steganography JSCC: In INNs, the input to each layer can be fully recovered from its output, meaning there exists a well-defined mathematical inverse operation for every transformation. If the forward calculation is $y = f(x)$, there must exist $x = f^{-1}(y)$. Additionally, during backpropagation, inputs can be recalculated from outputs without storing intermediate activations, making memory consumption independent of network depth, which is ideal for training ultra-deep architectures. Therefore, in INN-based image steganography JSCC, the invertible design of *INN-based semantic steganography* ensures no loss of input information during forward propagation, realizing precise semantic information reconstruction.

For instance, Tang et al. [10] introduced an INN-based signal steganography framework to enhance security in SemCom. The method leverages multiple invertible blocks and additive affine transformations to embed secret signals while preserving reversibility. The loss function incorporates both forward and backward propagation processes, along with a privacy protection term to safeguard sensitive information. Notably, the training process assumes no prior knowledge of Eves, i.e.,
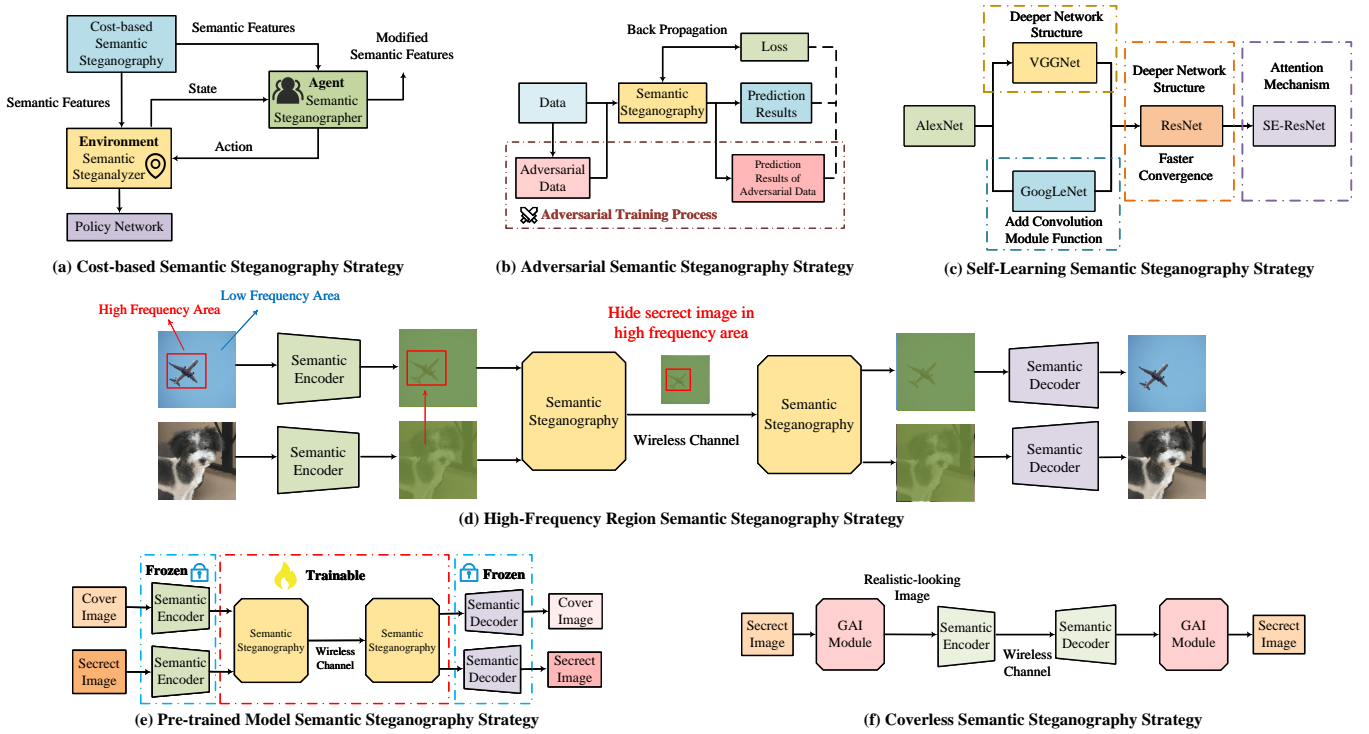
Fig. 2: Illustration of six image semantic steganography strategies. (a) Cost-based strategy leverages reinforcement learning to optimize embedding decisions with minimal cost, making it ideal for scenarios with limited steganographic capacity. (b) Adversarial strategy integrates adversarial examples during training to enhance resilience against detection by semantic steganalyzers. (c) Self-learning strategy utilizes end-to-end DL models to enable autonomous semantic steganography, eliminating the need for manual preprocessing or post-processing steps. (d) High-frequency region strategy embeds secret semantic information into high-frequency domains, leveraging regions with rich cover semantic features to evade semantic steganalysis. (e) Pre-trained model strategy avoids time-consuming model retraining, thus enhancing steganographic efficiency. (f) Coverless strategy generates stego images directly through GAI models, bypassing the requirement for cover images.

an eavesdropper. Experimental results demonstrate that eavesdroppers employing naive decoding or model inversion attacks can only reconstruct cover images, validating the framework's effectiveness against intelligent eavesdroppers. Tang et al. [11] conducted further assessments of the framework's security performance against various intelligent eavesdropping scenarios, including glass-box, closed-box, GAI+Glass-box, and GAI+closed-box eavesdropping. Simulation results reveal that the proposed scheme effectively mitigates the risks associated with these eavesdropping threats. Additionally, to degrade the performance of eavesdroppers, Li et al. [13] proposed a hiding and deception method based on INNs. This method deceives eavesdroppers by generating virtual semantic information for a certain modality while simultaneously hiding the original semantic information of that modality into another modality. Legitimate users can extract the hidden semantic information and accomplish their tasks normally through INNs.

B. Image Semantic Steganography Strategies

As illustrated in Figure 2, we present six strategies to support the training of image steganography JSCC models, which are discussed in detail as follows.

1) Cost-based Semantic Steganography: Cost-based semantic steganography can be designed using a semantic distortion minimization strategy to modify stego semantic features with least detectable artifacts. By introducing a reinforcement learning framework, the optimal trade-off between semantic steganographic security and semantic distortion can be achieved through dynamic game interactions and intelligent decision-making [14]. Specifically, the agent acts as a semantic steganographer, taking the semantic distortion levels as actions, aiming to learn the optimal steganography strategy associated with semantic steganography costs. Meanwhile, the environment simulates a semantic steganalyzer, providing feedback to the agent's semantic embedding actions: negative rewards for modifications easily detectable and positive rewards for imperceptible ones.

2) Adversarial Semantic Steganography: Adversarial examples are generated by adding imperceptible perturbations to original images, which can cause neural network-based classifiers to produce erroneous judgments. Inspired by this, adversarial semantic steganography embeds secret semantic features into cover semantic features while leveraging adversarial perturbations to enhance the security of stego semantic features, enabling them to resist detection by semantic steganalyzers. Specifically, we can leverage the generator in GANs to learn universal adversarial perturbation patterns, enabling it to automatically generate perturbations tailored to diverse semantic features, thereby enhancing the deceptive nature of stego semantic features.

3) Self Learning-based Semantic Steganography: The self learning-based semantic steganography strategy leverages the end-to-end optimization capability of deep learning models

to automatically learn the optimal way of embedding secret semantic features into cover semantic features without relying on manually predefined rules. Some typical models are as follows. AlexNet has a relatively small number of parameters and a simple structure, making it suitable for low-capacity semantic steganography in low-resolution images. VGGNet employs deep convolutional layers to capture more complex semantic features, enhancing the concealment of semantic feature embedding. GoogleNet utilizes Inception modules to simultaneously capture local details and global context, providing adaptive semantic region embedding. ResNet introduces residual structures, which extend the depth of networks and reduce the reconstruction error of secret semantic features. The attention mechanism can dynamically learn importance weights of different channels, thereby optimizing the deep semantic representation of images and improving the quality of stego semantic features as well as the accuracy of secret semantic feature extraction.

4) High-frequency Region Semantic Steganography: The low-frequency components of images are associated with smooth regions where the human visual system is particularly sensitive. Consequently, even minor alterations in these areas are easily noticeable by steganalyzers. In contrast, high-frequency components correspond to rapidly changing areas, such as textures and edges, which are often rich in semantic information but less perceptible to human vision [9]. This characteristic allows subtle modifications in these regions to go undetected by semantic steganalyzers. As a result, high-frequency region semantic steganography presents a practical strategy. For example, we can apply wavelet transforms to secret images to preserve high-frequency semantic features while eliminating low-frequency components.

5) Pre-trained Model-based Semantic Steganography: Compared to single-stage end-to-end training, the pre-trained model-based semantic steganography strategy adopts a two-stage separable training approach. The semantic encoder and decoder focus on efficiently compressing and reconstructing semantic information, while the semantic steganography module concentrates on secret semantic feature embedding and extraction. This separable training strategy avoids direct conflicts between the objectives of the two types of tasks, preventing the image steganography JSCC model from falling into local optima. For example, Tang et al. [11] employed the DeepJSCC architecture[1] as the trained semantic encoder and decoder, then froze their parameters, and further trained the parameters of INN-based signal steganography modules.

6) Coverless Semantic Steganography: Cover-based semantic steganography may disrupt statistical properties, leaving traces of rewriting. In contrast, coverless semantic steganography leverages GAI models to directly produce stego semantic features without requiring any cover images [15]. For example, we can construct specialized dictionaries and semantic knowledge bases using GAI models. At the transmitter, secret images are mapped onto semantic labels through these dictionaries, which are then utilized by GAI models to generate realistic-looking semantic features to confuse intelligent

---

[1]https://arxiv.org/abs/1809.01733

eavesdroppers. The receiver can identify these semantic labels to retrieve the hidden information.

C. Lessons Learned

We conducted a comparative analysis of three representative image steganography JSCC models, including CNN-based, GAN-based [12], and INN-based models [10], [11], [13]. Notably, INN-based models stand out due to their high capacity for secret semantic information and superior semantic transmission quality, which has led to them being the most extensively studied. Furthermore, we systematically categorized and elaborated on six strategic approaches for training semantic steganography modules. These approaches encompass the cost-based strategy, adversarial strategy, self learning-based strategy, high-frequency region semantic steganography, pre-trained model-based strategy, and coverless strategy. These strategies can be adopted either independently or in combination to meet the needs of secure SemCom.

IV. CASE STUDY: A COVERLESS STEGANOGRAPHY SEMCOM

A. Proposed Conditional Diffusion-based Coverless Steganography SemCom Scheme

Existing image steganography SemCom schemes [10]–[13] have the following limitations. Firstly, they are heavily constrained by the inherent characteristics of cover images. There is a close positive correlation between the capacity of secret images and the size and quality of cover images. This means that to transmit a high-dimensional secret image, it is necessary to simultaneously obtain a cover image with highly accurate matching. This requirement undoubtedly poses new challenges to resource scheduling and management in SemCom systems. Secondly, the selection and update mechanisms for cover images need further research. If there are insufficient cover images to meet update requirements, or if the update method appears abnormal, it can easily arouse the vigilance and suspicion of intelligent attackers. Thirdly, the embedding of secret semantic features may trigger statistical anomalies. Once intelligent attackers employ advanced steganalysis techniques, these potential statistical anomalies may become the breakthrough point for them to detect semantic steganography behavior.

Against this background, we present a coverless steganography SemCom scheme, as shown in Figure 3. Specifically, we employ conditional diffusion models to generate stego images to confuse eavesdroppers. In this regard, if the attackers do not have the private key, they cannot recover secret images. This is because stego images do not contain semantic features related to secret images through conditional generation. The private keys are generated based on a pre-trained semantic encoder, and they are distributed by a key agreement center. The interaction between legitimate users and the key management center is assumed to be conducted through secure channels.
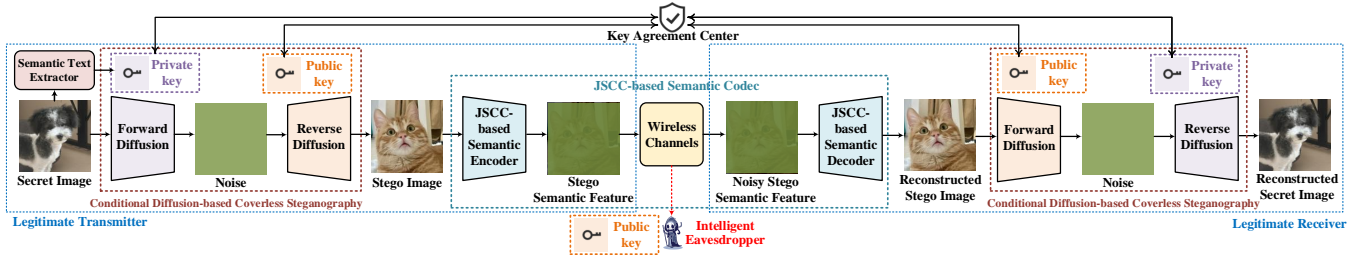
Fig. 3: Illustration of the proposed coverless steganography SemCom scheme, where the conditional diffusion-based coverless steganography module and JSCC-based semantic codec are trained separately, and keys are generated based on semantic features. The public keys are open access to everyone, including intelligent eavesdroppers, and they are semantic information related to steganographic images but not to secret images.
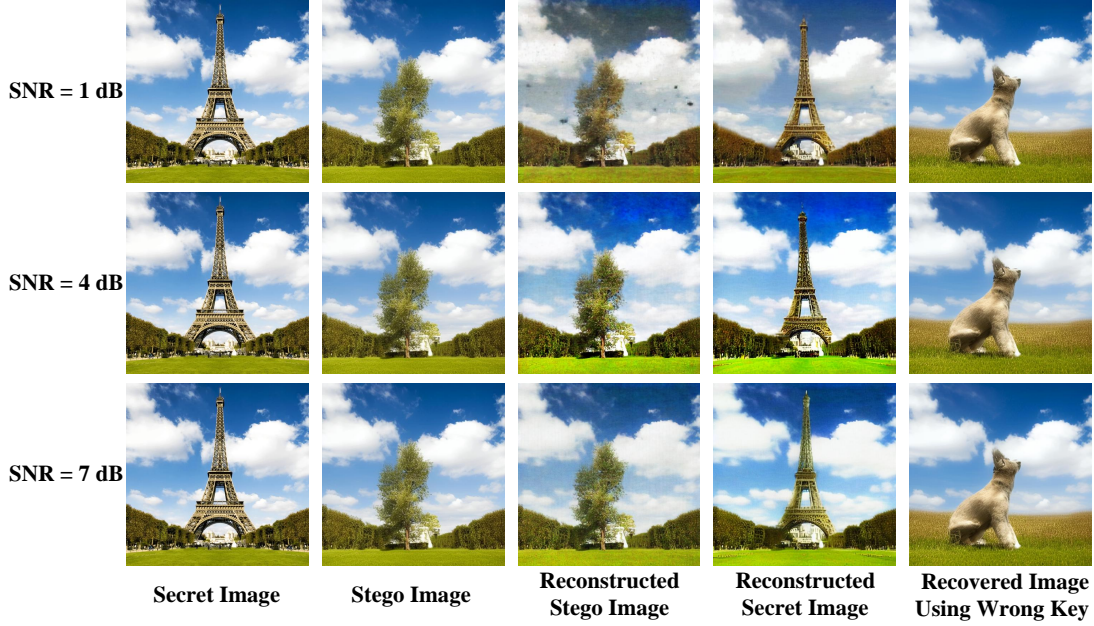

Fig. 4: Simulation results of the proposed conditional diffusion-based coverless steganography SemCom scheme. We select the Stable Diffusion version 1.5 as the conditional diffusion model and employ the deterministic DDIM as the sampling algorithm. Both the forward and reverse processes are configured to include 50 steps. The DeepJSCC architecture is employed as the trained semantic encoder and decoder.

B. Simulation Results

In our experiments, the datasets are sourced from two publicly available datasets[2][3] and Google search engines. Figure 4 provides the simulation results of the proposed scheme. Specifically, the Eiffel Tower serves as the confidential component requiring protection. Consequently, the prompt "*Eiffel Tower under the blue sky*" functions as the private key, while "*a tree under the blue sky*" acts as the public key. A legitimate receiver applying the correct private key successfully reconstructs the secret image, whereas an incorrect key fails to produce the intended result.

C. Lessons Learned

When compared to current image steganography SemCom methods, our proposed coverless steganography SemCom approach utilizes conditional diffusion models to produce stego images. This method eliminates the reliance on cover images

[2]https://github.com/aisegmentcn/matting_human_datasets

[3]https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals

and offers greater adaptability. Nevertheless, there are several ways in which this scheme could be enhanced: (1) Is it feasible to utilize the forward diffusion module to manage the noise addition process for generating stego images, thereby dispensing with the need for the reverse diffusion module? (2) In situations where multiple semantic features require protection, can we refine the joint conditional input process to derive the joint private keys?

V. FUTURE RESEARCH DIRECTIONS

A. Theoretical Analysis of Semantic Steganography Capacity

According to the mathematical theory of SemCom[4], we can know that, for band-limited Gaussian channels, the channel capacity can be represented as $C_S = B \log \left[S^4 \left(1 + \frac{P}{N_0 B} \right) \right]$, where S is the average synonymous length. In steganographic SemCom, when we embed secret semantic features into cover semantic features, the effective channel capacity will decrease compared to SemCom without steganography. This

[4]https://arxiv.org/abs/2401.13387

is because, to achieve the steganographic goal and deceive semantic steganalyzers, we need to transmit cover information as well as the correlation information between cover and secret semantic features. To maximally confuse attackers, reducing the transmission rate of secret semantic features is a direct and effective method. However, doing so will inevitably degrade communication efficiency. Therefore, how to provide a theoretical analysis of semantic steganographic capacity based on channel capacity and steganographic capacity is crucial for guiding further research in the future.

B. Multimodal Steganography SemCom

Compared to single-image modality SemCom, multimodal SemCom leverages the semantic correlations between multiple modalities to significantly enhance the completeness and precision of semantic understanding. However, the diverse data structures and varying semantic representation methods across different modalities necessitate that steganographic techniques be tailored to the unique characteristics of each modality. Furthermore, given the consistency of semantics across modalities, the stego semantic features must maintain a high degree of uniformity, because any deviation could easily arouse the suspicion of semantic steganalyzers. To effectively address these challenges, we can employ large multimodal AI models to map data from different modalities into a unified semantic space, within which semantic steganography strategies can be elaborately designed. Concurrently, leveraging the correlations between modalities to embed secret information can further enhance its concealment.

C. GAI-based Semantic Steganalysis

Traditional discriminative AI achieves steganalysis by learning the classification boundaries between data. However, it is sensitive to local features and prone to missing complex semantic steganography patterns. In contrast, GAI possesses powerful capabilities for modeling complex distributions and can detect traces left by semantic steganography through latent space analysis. From an intelligent attacker's perspective, diffusion models model data distributions through a gradual denoising process, which is expected to capture the latent space shifts caused by semantic steganography. In addition, GAI can automatically synthesize a rich variety of semantic steganography samples or enable adversarial training to effectively improve the generalization performance of detection, making it more advantageous when dealing with various complex semantic steganography scenarios.

D. Combination of Encryption and Steganography for Securing SemCom

Semantic steganography and semantic steganalysis are always in a state of confrontation, continuously evolving through this dynamic interplay. To provide more reliable privacy protection for SemCom, integrating encryption technology into semantic steganography is a promising solution. Homomorphic encryption allows direct homomorphic operations on encrypted data, ensuring that the privacy of the plaintext is not compromised during processing and providing security protection during semantic extraction and recovery[5]. Based on this, we can first perform semantic extraction on the ciphertext processed by homomorphic encryption. Then, we can design a semantic steganography module that supports homomorphic encryption to embed the secret semantic ciphertext into the cover semantic features. In this way, even if intelligent attackers obtain the secret semantic ciphertext through advanced semantic steganalysis techniques, they still need to crack the homomorphic encryption key to access the plaintext content.

VI. CONCLUSIONS

In this paper, we have first reviewed image SemCom encryption techniques, including cryptography-based SemCom, covert SemCom, physical-layer encrypted SemCom, and application-layer encrypted SemCom. Then, we have analyzed the potential of image steganography in securing SemCom. Furthermore, we have provided CNN-based, GAN-based, and INN-based image steganography JSCC models. Additionally, we have presented six image semantic steganography strategies, including cost-based semantic steganography, adversarial semantic steganography, self learning-based semantic steganography, high-frequency region semantic steganography, pre-trained model-based semantic steganography, and coverless semantic steganography. Moreover, we have given a case study to show the potential of coverless steganography in securing SemCom. Lastly, we have provided future research directions, including theoretical analysis of semantic steganography capacity, multimodal steganography SemCom, GAI-based semantic steganalysis, and combination of encryption and steganography for securing SemCom.

REFERENCES

[1] P. Zhang, W. Xu, Y. Liu, X. Qin, K. Niu, S. Cui, G. Shi, Z. Qin, X. Xu, F. Wang, Y. Meng, C. Dong, J. Dai, Q. Yang, Y. Sun, D. Gao, H. Gao, S. Han, and X. Song, "Intellicise wireless networks from semantic communications: A survey, research issues, and challenges," *IEEE Communications Surveys & Tutorials*, 2024.

[2] S. Guo, Y. Wang, N. Zhang, Z. Su, T. H. Luan, Z. Tian, and X. Shen, "A survey on semantic communication networks: Architecture, security, and privacy," *IEEE Communications Surveys & Tutorials*, 2024.

[3] R. Meng, S. Gao, D. Fan, H. Gao, Y. Wang, X. Xu, B. Wang, S. Lv, Z. Zhang, M. Sun *et al.*, "A survey of secure semantic communications," *Journal of Network and Computer Applications*, p. 104181, 2025.

[4] M. Shen, J. Wang, H. Du, D. Niyato, X. Tang, J. Kang, Y. Ding, and L. Zhu, "Secure semantic communications: Challenges, approaches, and opportunities," *IEEE Network*, vol. 38, no. 4, pp. 197–206, 2024.

[5] Z. Yang, M. Chen, G. Li, Y. Yang, and Z. Zhang, "Secure semantic communications: Fundamentals and challenges," *IEEE Network*, vol. 38, no. 6, pp. 513–520, 2024.

[6] H. Du, J. Wang, D. Niyato, J. Kang, Z. Xiong, M. Guizani, and D. I. Kim, "Rethinking wireless communication security in semantic internet of things," *IEEE Wireless Communications*, vol. 30, no. 3, pp. 36–43, 2023.

[7] S. Rustad, P. N. Andono, G. F. Shidik *et al.*, "Digital image steganography survey and investigation (goal, assessment, method, development, and dataset)," *Signal processing*, vol. 206, p. 108908, 2023.

[8] K. Hu, M. Wang, X. Ma, J. Chen, X. Wang, and X. Wang, "Learning-based image steganography and watermarking: A survey," *Expert Systems with Applications*, p. 123715, 2024.

[9] B. Song, P. Wei, S. Wu, Y. Lin, and W. Zhou, "A survey on deep-learning-based image steganography," *Expert Systems with Applications*, p. 124390, 2024.

[5]https://arxiv.org/abs/2501.10182

[10] S. Tang, C. Liu, Q. Yang, S. He, and D. Niyato, "Secure semantic communication for image transmission in the presence of eavesdroppers," *arXiv preprint arXiv:2404.12170*, 2024.

[11] S. Tang, Y. Chen, Q. Yang, R. Zhang, D. Niyato, and Z. Shi, "Towards secure semantic communications in the presence of intelligent eavesdroppers," *arXiv preprint arXiv:2503.23103*, 2025.

[12] Y. Huo, S. Xiang, X. Luo, and X. Zhang, "Image semantic steganography: A way to hide information in semantic communication," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 2, pp. 1951–1960, 2025.

[13] Z. Li, O. Huan, W. Zhang, and T. Luo, "Multi-modal task-oriented secure semantic communication: A hide-and-deceive approach," in *2024 10th International Conference on Computer and Communications (ICCC)*. IEEE, 2024, pp. 1477–1482.

[14] W. Tang, B. Li, M. Barni, J. Li, and J. Huang, "An automatic cost learning framework for image steganography using deep reinforcement learning," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 952–967, 2021.

[15] J. Yu, X. Zhang, Y. Xu, and J. Zhang, "Cross: Diffusion model makes controllable, robust and secure image steganography," *Advances in Neural Information Processing Systems*, vol. 36, pp. 80 730–80 743, 2023.