Energy-Efficient Resource Allocation in Generative AI-Aided Secure Semantic Mobile Networks

Jie Zheng, Baoxia Du, Hongyang Du, Jiawen Kang, Dusit Niyato, *Fellow, IEEE*, and Haijun Zhang, *Fellow, IEEE*

***Abstract*—The integration of semantic communication with Internet of Things (IoT) technologies has advanced the development of Semantic IoT (SIoT), with edge mobile networks playing an increasingly vital role. This paper presents a framework for SIoT-based image retrieval services, focusing on the application in automotive market analysis. Here, semantic information in the form of textual representations is transmitted to users, such as automotive companies, and stored as knowledge graphs, instead of raw imagery. This approach reduces the amount of data transmitted, thereby lowering communication resource usage, and ensures user privacy. We explore potential adversarial attacks that could disrupt image transmission in SIoT and propose a defense mechanism utilizing Generative Artificial Intelligence (GAI), specifically the Generative Diffusion Models (GDMs). Unlike methods that necessitate adversarial training with specifically crafted adversarial example samples, GDMs adopt a strategy of adding and removing noise to negate adversarial perturbations embedded in images, offering a more universally applicable defense strategy. The GDM-based defense aims to protect image transmission in SIoT. Furthermore, considering mobile devices' resource constraints, we employ GDM to devise resource allocation strategies, optimizing energy use and balancing between image transmission and defense-related energy consumption. Our numerical analysis reveals the efficacy of GDM in reducing energy consumption during adversarial attacks. For instance, in a scenario, GDM-based defense lowers energy consumption by 5.64%, decreasing the number of image retransmissions from 18 to 6, thus underscoring GDM's role in bolstering network security.**

***Index Terms*—Generative AI, resource allocation, semantic communication, energy efficiency.**

I. Introduction

The advent of Sixth-Generation (6G) communications signals a significant shift in information transmission, setting the stage for transformative changes in wireless communication. This new era is characterized by an unprecedented surge in data volume, largely driven by the extensive deployment of Internet of Things (IoT) devices [1], [2]. Additionally, the development of semantic communications has played a crucial role in revolutionizing wireless network operations [3]. Unlike traditional approaches prioritizing bit transmission accuracy, semantic communications emphasize the transmission of meaningful and task-relevant data [4]. This approach has been made viable by developing efficient semantic models, allowing for the integration of semantic encoders and decoders at the network's edge. Such advancements are instrumental in evolving standard IoT systems into Semantic IoT (SIoT) [5], [6], enhancing data exchange efficiency and sustainability. A prime example of this evolution is the emergence of the Semantic Internet of Vehicles (SIoV), where vehicle communication focuses on exchanging semantically rich information, like neural network-derived features, rather than raw imagery. This shift is fundamental in enabling more intelligent and efficient vehicular interactions [7], [8].

The advent of 6G networks, coupled with SIoT's maturation, is set to unlock a spectrum of data-centric services, a development close to realization as indicated in [9], [10]. These networks facilitate various new services, sustained by a continuous data stream. For example, the tripartite service model demonstrates this concept effectively [9]. In this model, a service provider manages a large database on a Publicly Accessible Server (PAS). Network mobile devices, such as car cameras, are integral in gathering and transmitting data to the PAS, keeping it updated with current visual information. Simultaneously, service consumers utilize this extensive data to meet their diverse needs. This model illustrates the interconnected roles of the provider, devices, and consumers within the 6G and SIoT ecosystem, showcasing the potential for data-centric services.

However, the rapid expansion of mobile devices and their integration into networks has led to sophisticated cyber threats,

Manuscript received 20 December 2023; revised 4 April 2024; accepted 29 April 2024. Date of publication 6 May 2024; date of current version 5 November 2024. This research was supported in part by the National Natural Science Foundation of China under Grant 62072362 and Grant 62072373, in part by National Research Foundation, Singapore, and Infocomm Media Development Authority under its Future Communications Research and Development Programme, Defence Science Organisation (DSO) National Laboratories under the AI Singapore Programme AISG under Grant AISG2-RP-2020-019 and under Grant FCP-ASTAR-TG-2022-003, and in part by the Singapore Ministry of Education (MOE) Tier 1 under Grant RG87/22. Recommended for acceptance by F. Wu. *(Corresponding author: Jie Zheng.)*

Jie Zheng is with the State-Province Joint Engineering and Research Center of Advanced Networking and Intelligent Information Services, School of Information Science and Technology, Northwest University, Xi'an 710127, China (e-mail: jzheng@nwu.edu.cn).

Baoxia Du is with the School of Electrical and Information Engineering, JiLin Agricultural Science and Technology University, Jilin 132101, China, and also with the School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin 132022, China (e-mail: dubaoxia@jlict.edu.cn).

Hongyang Du and Dusit Niyato are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: hongyang001@e.ntu.edu.sg; dniyato@ntu.edu.sg).

Jiawen Kang is with the Automation of School, Guangdong University of Technology, Guangzhou 523083, China (e-mail: kjwx886@163.com).

Haijun Zhang is with the Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China (e-mail: haijunzhang@ieee.org).

Digital Object Identifier 10.1109/TMC.2024.3396860

as highlighted in [11], [12], [13]. The growth in IoT technology, while enhancing data exchange and connectivity, also broadens the attack surface for cyber adversaries [14]. A key issue is ensuring the secure transmission of image data in wireless networks. In IoT-based surveillance systems, attackers can manipulate transmitted images using advanced techniques such as adversarial perturbations. These perturbations, as described in [15], can mislead Deep Neural Networks (DNNs) with minimal input changes, remaining virtually undetectable to humans yet significantly altering model predictions. This threat extends beyond image classification to include semantic segmentation [16], object detection [17], and object tracking [18]. Models under attack often misinterpret manipulated images, leading to incorrect predictions and compromising data integrity, as noted in [19]. Such corrupted images can have far-reaching consequences, misguiding decisions, endangering operations, and wasting network resources. The authors in [20] highlight that certain perturbations can affect multiple network classifiers, illustrating the complexity and severity of adversarial attacks in IoT image communications. This situation underscores the imperative for robust security measures to safeguard the accuracy and integrity of image data in transmission.

Various strategies have been investigated in the complex field of protecting deep learning models from adversarial attacks. Initial defenses involved model modifications, with adversarial training as a key method. This technique enhances models by training them with adversarial examples, as detailed in [21]. However, adversarial training has shown weaknesses, including vulnerability to specific types of attacks and questions about its broad applicability, as noted in [22]. Despite these challenges, adversarial training has spurred several refinements, such as misclassification-aware training and margin-maximization techniques [23]. Concurrently, the rise of Generative Diffusion Models (GDMs) presents a novel defensive strategy [24], [25], [26], [27]. These models employ a process of converting data to noise and then reverting it back, effectively purifying adversarial perturbations [28]. Their high-quality generation ensures that the restored images closely resemble the original data. Additionally, the stochastic nature of GDMs enhances their resilience, making them suitable for adversarial purification [26]. Alongside model modification and adversarial training, input transformation methods, such as JPEG compression, act as supplementary defenses, reducing the effectiveness of adversarial inputs. The landscape of defenses is constantly evolving, with each approach presenting its own set of challenges and strengths [29].

While GDM-based defenses are promising for countering adversarial attacks, their application in IoT, particularly in edge mobile networks, requires careful consideration of energy tradeoffs [28]. Edge mobile devices, constrained by limited computational resources and energy, must balance effective defense and energy efficiency. Although GDMs are powerful, their computational demands can significantly increase energy consumption. This becomes more pronounced in the wider context of network communication [28]. Investing more energy in the denoising process can reduce communication energy between the service provider and users by decreasing the likelihood of compromised transmissions, leading to fewer re-transmissions and energy savings [9]. This can result in lower latency, better bandwidth utilization, and improved network efficiency. However, excessive energy use in the denoising process could negate these benefits, quickly depleting edge devices' limited energy and potentially causing operational issues or frequent downtimes. Thus, it is crucial to judiciously manage energy allocation in these systems. The contributions of this paper are summarized as follows:

- We introduce a framework for SIoT that enables efficient image retrieval application services. This framework incorporates a service provider, edge mobile devices, and end-users and utilizes semantic information to facilitate task-oriented data retrieval and transmission.
- We implement adversarial attack schemes during edge mobile networks' image transmission and propose a defense framework using the GDM. We also show that the computing process of the proposed GDM-based defense method can be distributed to various network devices.
- We propose an approach leveraging GDM for energy optimization in the GDM-based defense method. A key feature is determining the optimal number of denoising steps to achieve robust security and energy efficiency.

The remaining sections are organized as follows: In Section II, we provide a brief summary of relevant technologies, including semantic communication, GDM, adversarial attack, and defense to the adversarial attack. Section III introduces the overall system design, which includes the communication service framework applied to SIoT, a case study, and formulas for calculating system energy consumption. Section IV provides a detailed exposition on adversarial attack and GDM-aided defense. Subsequently, in Section V, we elucidate energy-efficient resource allocation schemes. Section VI presents a numerical analysis of the proposed case and Section VII elaborates on the applicability and challenges of the framework presented. Finally, Section VIII draws conclusions.

II. Related Work

In this section, we provide a comprehensive overview of related techniques, including semantic communication, GDMs, adversarial attacks, and defenses to adversarial attacks.

A. Semantic Communications

Unlike conventional communication based on Shannon's Information theory, which only focuses on reliable data transmission without considering the content of information, semantic communication aims to ensure that the transmitted data accurately conveys the intended meaning. Based on this, the receiver can achieve the ultimate goal by using the received data [30], [31]. With the continuous growth in demand for intelligent services, integrating user, application requirements, and the semantics of information into data processing and transmission is poised to become a new core paradigm in 6G [3]. Due to the success of deep learning (DL) in natural language processing (NLP), Nariman Farsad et al. [32] embedded sentences into a semantic space, conveying only the semantic information within the sentences, thereby achieving a lower word error rate. Xie et al. [33] employed the Transformer architecture for textual transmission, thereby enhancing system robustness by restoring the semantic integrity of sentences. In the field of image communication, Hu et al. [34] propose a robust end-to-end semantic communication system framework to address the impact of semantic noise, in which the masked autoencoder (MAE) can reconstruct images based on partial observations, reducing transmission overhead and significantly improving the robustness of the semantic communication system. Lee et al. [35] address energy-constrained IoT devices by introducing a jointly designed transmission-recognition scheme using deep learning. This scheme transmits image features during the recognition process as semantic information and achieves higher recognition accuracy than traditional encoding methods such as JPEG compression. Pan et al. [36] present an image segmentation semantic communication (ISSC) system for vehicular networks. It reduces the transmitted data by extracting features from images using a multi-scale semantic feature extractor based on the Swin Transformer. The receiver can reconstruct segmented images through a semantic feature decoder and reconstructor. Compared to traditional encoding baseline methods, the ISSC system achieves a 75% improvement. Xie et al. [37] consider a task-oriented multi-user semantic communication system for the visual question answering (VQA) task. Receivers can directly generate answers by receiving semantic information from different senders in the form of images and text. This method exhibits better robustness compared to traditional communication systems. Kang et al. [38] reduce communication overhead significantly by matching user query text with the semantic information of images, transmitting only the images that interest users. Furthermore, they allocate resources based on user interests to ensure communication quality. Wei Chong Ng et al. [9] consider a virtual transportation network in the metaverse. They use the YOLO model to extract semantic information from road images collected by edge devices for transmission, thereby reducing power consumption and storage costs during the transmission process. In this paper, we introduce semantic communication technology into IoT applications by leveraging knowledge graphs to store obtained semantic data information, which can significantly reduce overhead during communication processes.

B. Generative Diffusion Model

Denoising GDMs such as denoising diffusion probabilistic models (DDPMs) [39] have emerged as a recent and promising topic in computer vision, showcasing notable outcomes in generative modeling. DDPMs are deep generative models based on two distinct processes: forward diffusion and reverse diffusion. Gaussian noise is progressively added to perturb input data in the forward diffusion process. Subsequently, the model learns the reverse diffusion process to recover the input image in the reverse process. This reverse diffusion process can generate data from noise like a Markov chain. Diffusion models have garnered widespread acclaim due to their ability to generate samples of high quality and diversity. To date, GDMs have found extensive applications across a spectrum of generative tasks. In the realm of computer vision, these applications encompass tasks such as image generation [39], image denoising [40], image super-resolution [41], and image editing [42], among others. Furthermore, it has been found that the latent representations learned from GDMs are of great value in discriminative tasks, including image classification [43], detection [44], and segmentation [45]. Furthermore, it is noteworthy that GDMs can also be employed for generating text [46] and audio [47], among other modalities. This underscores the extensive applicability of denoising GDMs, thereby suggesting that further application avenues await exploration.

C. Adversarial Attack

In recent years, with the remarkable success of AI technologies based on DNNs in various domains, including computer vision, speech recognition, and natural language processing, the integration of AI with edge devices has emerged as a crucial driving force for the advancement of smart cities [48]. An increasing number of DNN models have found widespread application in diverse security-sensitive tasks such as facial recognition, intelligent healthcare, and autonomous driving, significantly expediting the development of smart cities. However, some studies have found that DNNs exhibit a high sensitivity to small noise in the data, which is imperceptible to human observers. By introducing subtle yet carefully crafted modifications to the input data, i.e., adversarial example attacks [15], it is possible to induce incorrect outputs from DNN models [49], [50]. This category of attacks is commonly referred to as adversarial attacks. Adversarial attacks can be categorized into two types: white-box attacks and black-box attacks. White-box attacks have access to detailed information about the model's structure, whereas black-box attacks can only obtain the target model's outputs by feeding raw data into the target model [51].

In computer vision, particularly in common image recognition and object detection tasks [52], [53], when these models are applied to security-sensitive applications, erroneous model outputs can potentially result in severe consequences. For instance, it is possible to deceive facial recognition systems by simply wearing a distinctive pair of eyeglasses [54], [55]; autonomous driving technologies are susceptible to environmental variations, occasionally leading to vehicular loss of control in severe cases [56]. Therefore, research on adversarial attacks contributes to enhancing the robustness of models, enabling them to function more steadfastly and reliably when confronted with adversarial assaults.

D. Defense to the Adversarial Attack

The susceptibility of neural networks to adversarial attacks, notably due to their sensitivity to minor input perturbations, is a critical AI security concern. Since the discovery of the susceptibility of DNNs to adversarial perturbations [15], researchers have been continuously exploring corresponding defensive techniques in various domains [57], [58], [59], [60]. especially in computer vision. In their seminal work, Papernot et al. [61] applied defensive distillation techniques to image classification tasks, effectively enhancing the model's ability against adversarial attacks. Folz et al. [62] propose an architecture based on compressive autoencoders (AEs) with a two-stage training scheme. By leveraging gradient information masking, their approach efficiently defended against gradient-based adversarial examples. Adnan et al. [60] injected trainable Gaussian noise into every layer of the model's activations or weights, incorporating adversarial training, thereby significantly improving the robustness of DNNs against adversarial attacks. These aforementioned works primarily concentrate on model optimization, and some studies also use data optimization to resist adversarial attacks. Xie et al. [63] observed that, due to the continuity of the image feature space, introducing randomization to the test images in certain scenarios can aid in defending against adversarial attacks. Furthermore, some research efforts have explored alternative forms of input data pre-processing to enhance defense capabilities, such as local smoothing [64] and image compression [65]. While these methods have made significant contributions towards advancing the field of adversarial attack defense, they also exhibit certain limitations, such as potential trade-offs with model accuracy, a lack of universality, and inefficacy in handling more sophisticated adaptive attacks.

However, the emergence of Generative Artificial intelligence (GAI), specifically DDPMs, provides a novel approach to address this. Leveraging the adaptability and robustness of DDPMs, a sturdy defense mechanism can be established. Ankile, Midgley, and Weisshaar [24] utilized the reverse diffusion process inherent to DDPMs to enhance system robustness against adversarial attacks by introducing and then strategically removing noise from adversarial examples. Tested on the Patch-Camelyon dataset, their strategy showed notable improvement in classification accuracy, reaching 88% of the accuracy of the original discriminative AI-based model. Kanget al. [25] introduced a defense method named DIFFender for countering adversarial patch attacks in the physical world, DIFFender leverages GDMs in the localization stage to identify the positions of adversarial patches and eliminates the adversarial regions within the image while preserving the visual content intact. This approach consistently outperforms conventional methods in defending against patch attacks on ImageNet and facial recognition datasets. Wang et al. [26] proposed a technique known as the Guided Diffusion Model for Purification (GDMP) to purify attached images. GDMP employs the noise introduced by the DDPM diffusion process to gradually submerge the adversarial perturbations added to the image, thereby eliminating their impact during the denoising process. This method achieves a robust accuracy of 90.1% on the attacked CIFAR10 dataset. However, he limitations of diffusion purification, specifically the substantial resource consumption associated with the operation of DDPMs, are readily apparent, particularly when deployed on resource-constrained edge devices. Consequently, in this paper, we direct our attention toward the resource consumption associated with information transmission and defense against adversarial attacks, with a dedicated focus on maintaining secure communication while simultaneously striving to minimize resource depletion.

III. System Model

In this section, we first provide a comprehensive discussion of the system model and subsequently formulate the energy-efficient resource allocation problem.

A. *Semantic Internet-of-Things*

In SIoT, unlike traditional IoT, which transmits data through bit streams, semantic information is largely task-oriented. In this backdrop, we consider that an image dataset resides in a PAS, and the service provider is responsible for retrieving the requested images from the PAS and sending them to a user. The PAS, however, may contain attack images uploaded by malicious attackers. If the service provider inadvertently sends these attack images to a user, the user would immediately recognize the incorrect content with the human eye, resulting in a retransmission request. This process consumes unnecessary communication resources, as the provider has to re-fetch and retransmit the correct images.

To exemplify the efficacy of our proposed system framework, we delve into a practical case study about automotive market analysis. The scenario, depicted in Fig. 1, involves automotive companies seeking to analyze their market presence and develop competitive strategies. A key aspect of this analysis is understanding the market penetration of their branded vehicles on specific roadways and assessing user driving patterns. A straightforward and effective method for acquiring this information is monitoring various vehicle brands' traffic activities on roads. To facilitate this, service providers install a network of edge devices, such as cameras in mobile devices. These cameras capture and transmit images of passing vehicles to the PAS. The stored images, which include discernible car emblems and models, provide valuable insights into the diversity and frequency of different vehicle brands on these roads. Automotive companies, acting as users in this system, can then access this data by requesting specific image sets from the service provider. This process allows them to gauge their brand's visibility and user preferences in real-time traffic conditions, thereby aiding in strategic decision-making and targeted marketing efforts.

During this process, due to the huge communication costs associated with transmitting all images to the users, one approach is to employ the knowledge graph technique to extract semantic information from each image [38]. The service provider only needs to transmit the semantic information of the images to the users, who then retrieve the desired images based on this semantic information for downloading [38]. However, after one user sends requests, malicious images within PAS may be incorrectly identified and transmitted to this user. Users' dissatisfaction with these results may lead to requests for retransmission, which not only squanders communication resources but also reduces user satisfaction with the service. Specifically, the steps of this process can be summarized as follows:

1) The service provider sends the semantic information of all images, e.g., "Jeep car - parked - road", to the user.
2) The mobile user requests U specific vehicle brand category images, e.g., "Mercedes", from the server provider based on the received semantic information.

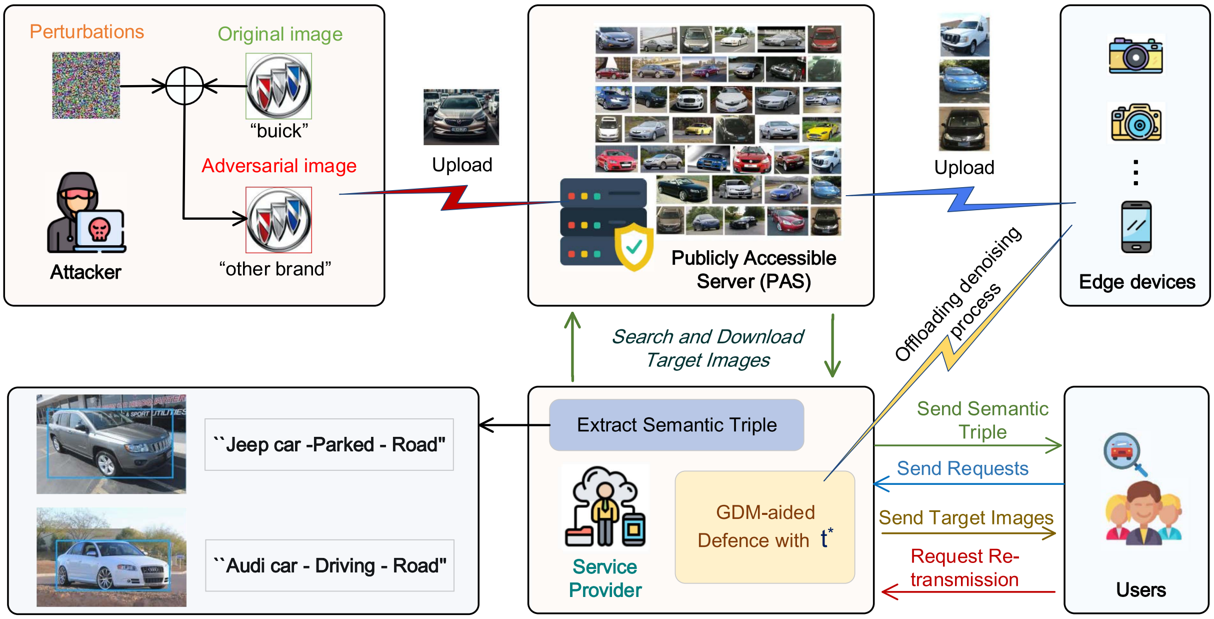


Fig. 1. System model illustrating the process of task-driven selection of specific vehicle brand category images from a service provider's PAS, employing a GDM to counter malicious adversarial attacks and incorporating iterative cycles for erroneous image transmission.

3) The service provider receives the request and selects related images from the PAS.
4) Prior to transmission, the service provider employs a GDM with a time step t^* to counter malicious adversarial attacks.
5) The mobile user receives images from the service provider, which may still include r_i erroneous images, e.g., images of Rolls Royce cars because of the wrong classification. Here, i denotes the round index of requests.
6) The user identifies incorrect images and continues to request r_i images of the same category from the service provider.
7) The service provider repeats the steps 3) and 4).
8) The user repeats steps 5) and 6) until they receive the total of U correct images of the requested vehicle brand category.

B. Problem Formulation

Due to the increased computational load introduced by the utilization of the GDM, its critical parameter t^* needs to strike a balance between computational and communication resources. Increased computational investment typically improves defense effectiveness. To achieve an optimal setting for t^*, we propose an optimization problem aiming at determining the optimal number of denoising steps t^* to be set in defense to minimize the total energy cost. Let $E_d(t^*, r_i)$ denote the energy cost in the GDM-based defense method. As t^* increases, $E_d(t^*, r_i)$ increases. Let E_r represent the resource consumption while transmitting a single image. Thus, the overall energy consumption can be expressed as follows:

$$E\left(t^*\right) = \sum_{i=1}^{k} \left(E_d\left(t^*, r_i\right) + r_i E_r\right), \tag{1}$$

where $r_1 = U$ and $k = i_{\text{end}}$ that makes $r_{i_{\text{end}}} = 0$. Therefore, our objective can be expressed as

$$\begin{aligned} \min_{t^*} \quad & E\left(t^*\right), \\ s.t. \quad & \sum_{i=1}^{k} u_i = U, \end{aligned} \tag{2}$$

where u_i denotes the number of correct images in the r_i-round transmission. Since the optimal defense scheme depends on how the attack is carried out, the extent to which the database is attacked and the database has a certain degree of randomness (the number of attacked images is not the same for each user request), traditional non-convex optimization methods are difficult to be applied. In the following, we first introduce the AI-based attack image generation method, followed by the GDM-aided defense method. Then, the GDM-based solution to the optimization problem (2) is given.

IV. Attack and GDM-Aided Defense

In this section, we provide a comprehensive exposition of the principles underlying adversarial attack and the GDM-aided defense methodology.

A. Adversarial Attack

The image datasets hosted on PAS may include adversarial attack images uploaded by malicious edge devices in SIoT. When users request images of specific categories from the server, there is a possibility that the server may erroneously transmit adversarial attack images to users instead of the desired images. For example, the sales department of *Brand A vehicles* may manipulate the images of their cars to resemble those of *Brand B vehicles* semantically, thereby deceiving classifiers to

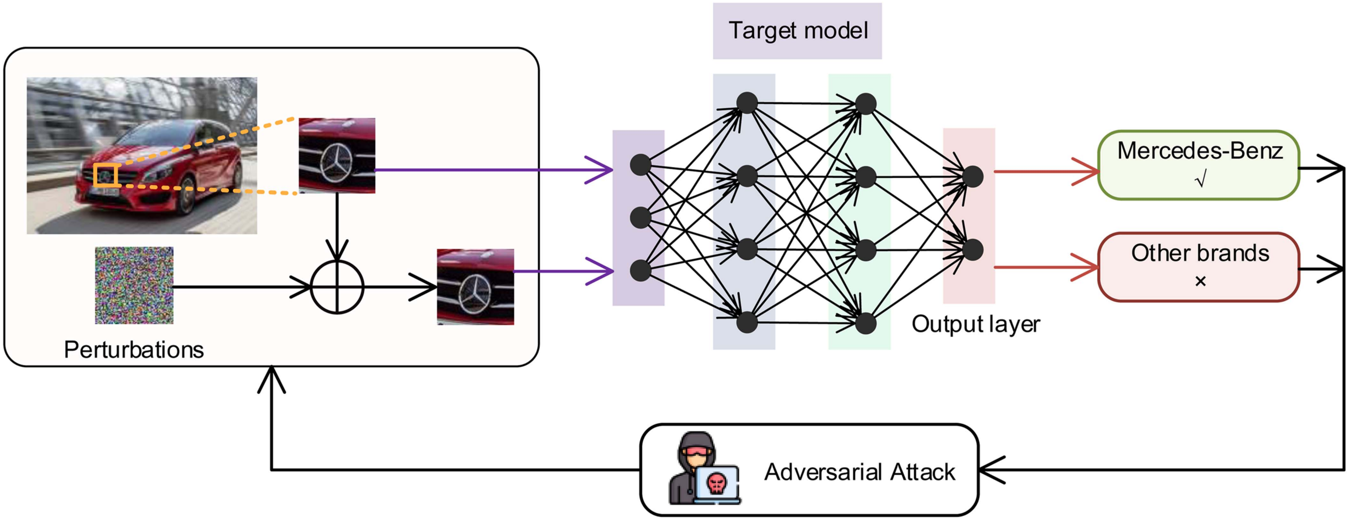


Fig. 2. Adversarial example generation and adversarial attack process. Adversarial perturbations are added to clean images to deceive the classifier into generating incorrect outputs.

achieve the goal of promoting *Brand A*. This attack type can be considered image classification attacks [66].

In the PAS image dataset, we represent the pure RGB images as $\mathbf{x}_t \in \mathbb{R}^{H\times W\times 3}$, where $W \times H \times 3$ denotes the width, height, and number of channels of these images, respectively. Each image is associated with a corresponding label $\mathrm{y}_t = f_c(\mathbf{x}_t)$, where $f_c(\cdot)$ represents the image classifier employed by the service provider. As shown in Fig. 2, when certain images within the PAS are subjected to attacks, these manipulated images are referred to as carrier images $\mathbf{x}_c$. The objective of the attacker is to introduce adversarial perturbations onto the carrier image $\mathbf{x}_c$, thereby generating an adversarial image $\mathbf{x}_a$ that exhibits visual similarity to the carrier image $\mathbf{x}_c$ but is misclassified by $f_c(\cdot)$, i.e., $y_a = f_c(\mathbf{x}_a) \neq y_c$, while maintaining visual resemblances to $\mathbf{x}_c$.

Loss Function: Consider an input image $\mathbf{x}$, the objective of a malevolent attacker is to manipulate the image satisfying $f_c(\mathbf{x}) \neq y_c$, while maintaining a high degree of visual similarity to the original image $\mathbf{x}_c$. Consequently, the loss function can be formulated as

$$L_N\left(\mathbf{x}_c, y_c; \mathbf{x}\right) = -l\left(f_c\left(\mathbf{x}\right), y_c\right) + \lambda\|\mathbf{x} - \mathbf{x}_c\|^2. \tag{3}$$

Function $l(\cdot)$ represents the cross-entropy loss, while $\|\mathbf{x} - \mathbf{x}_c\|^2$ corresponds to the distortion loss between $\mathbf{x}$ and $\mathbf{x}_c$, and$\lambda$is a hyper-parameter to show the impact of the distortion loss. The aforementioned form of adversarial attack can be referred to as non-targeted misclassification. The implementation of Non-targeted misclassification is achieved by increasing the confidence of the classification model towards other categories, i.e., reducing the confidence y_c. In contrast, the objective of Targeted misclassification is to generate an adversarial image that can be classified as y_o. The implementation of targeted misclassification is achieved by minimizing the loss function as

$$L_T\left(\mathbf{x}_c, y_o; \mathbf{x}\right) = l\left(f_c\left(\mathbf{x}\right), y_o\right) + \lambda\|\mathbf{x} - \mathbf{x}_c\|^2. \tag{4}$$

Differing from (3), the cross-entropy loss is minimized for the target class rather than maximized for the carrier class.

Optimization: In our system model, we consider using adaptive attacks designed with full knowledge of the model's defense. The Non-targeted misclassification adversarial examples were created by finding the perturbation from the set of allowable perturbations that minimized the loss given by the following equations:

$$\mathbf{x}_a = \arg\min_{\mathbf{x}} L_N(\mathbf{x}_c, y_c; \mathbf{x}), \tag{5}$$

$$\mathbf{x}_a = \mathbf{x} + \delta, \tag{6}$$

$$\delta \in \{\delta : \|\delta\|_\infty \leq \varepsilon\}. \tag{7}$$

The restriction of the norm of the perturbation δ ensures that the perturbed image is indistinguishable from the original image. Similar to (5) for Non-targeted misclassification, the optimization of (4) generates the adversarial images represented by

$$\mathbf{x}_a = \arg\min_{\mathbf{x}} L_T(\mathbf{x}_c, y_o; \mathbf{x}). \tag{8}$$

Based on the outlined process, malicious edge devices within the SIoT could potentially generate adversarial images and upload them to the PAS. In the absence of a robust defense policy, the service provider may inadvertently transmit irrelevant images in response to the user's request, thereby causing network resource waste. Therefore, we present a cutting-edge defense mechanism based on the GDM to counter such attacks in the following.

B. GDM-Aided Defense

Adversarial purification, which involves using generative models to purify adversarial images before classification, has emerged as a promising approach for defending against adversarial attacks. The DDPM gradually adds noise during the forward process, which allows it to disrupt the adversarial perturbations added to the input image, and hence it cannot perturb the

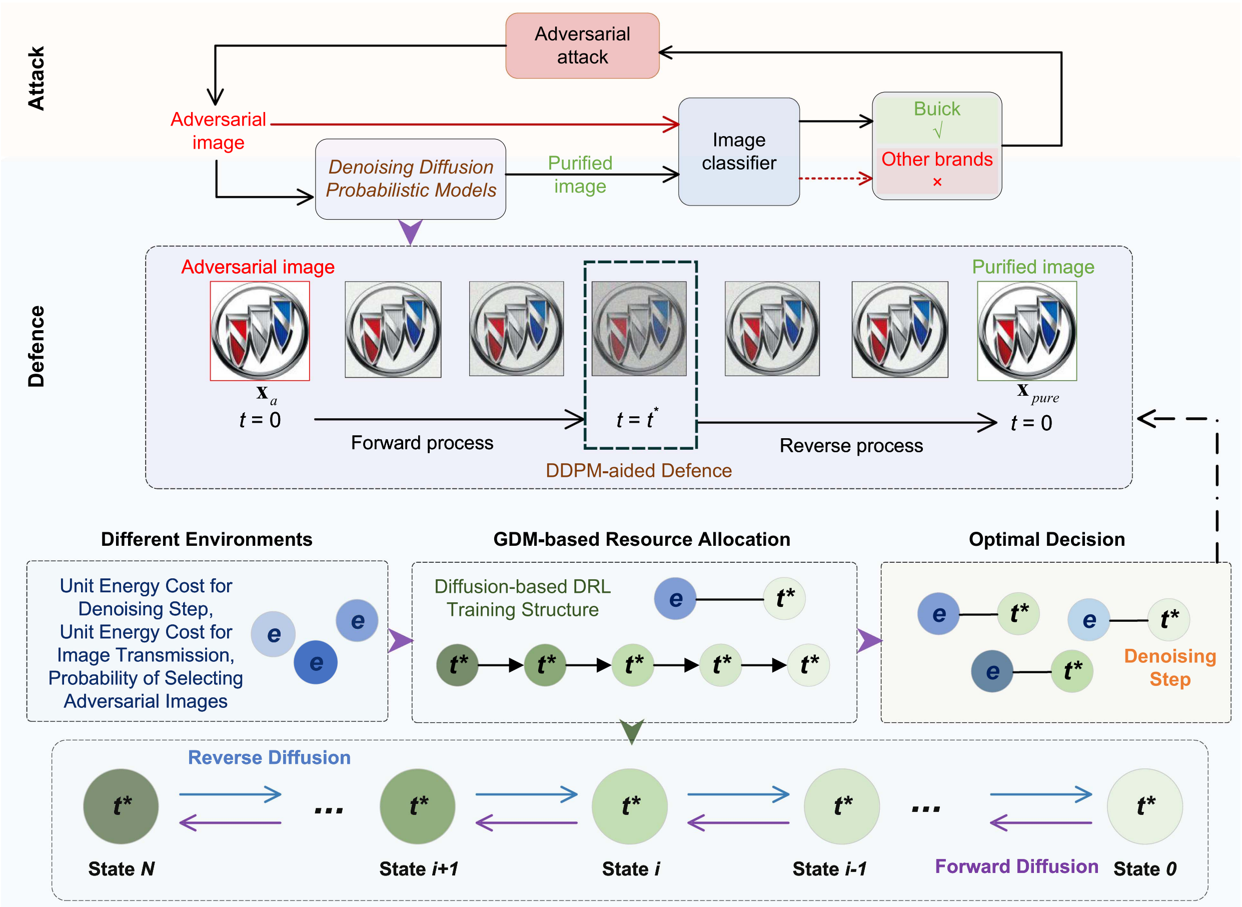


Fig. 3. GDM-based resource allocation for GDM-based defense mechanism. Utilizing a selected denoising step t^*, noise is introduced into the adversarial images, generating diffused variants. These diffused images are then processed by using a pre-trained GDM through an inverse denoising algorithm to yield clean images. For any given environment, the allocation scheme is first randomly generated as Gaussian noise. After multiple steps of denoising through a GDM, the output is a scheme design that maximizes the optimization objective.

classifier [24]. A DDPM consists of two processes: the diffusion process and the reverse process [39]. The detailed principles of these two processes are explained below.

DDPM Forward Process: The diffusion process is defined by a fixed Markov chain from data $\mathbf{x}_0$ to the latent variable $\mathbf{x}_T$

$$q\left(\mathbf{x}_1,\ldots,\mathbf{x}_T|\mathbf{x}_0\right) = \prod_{t=1}^{T} q\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right), \tag{9}$$

where each of $q(\mathbf{x}_t|\mathbf{x}_{t-1})$ can be obtained by $\mathcal{N}(\mathbf{x}_t;\sqrt{1-\beta_t}\mathbf{x}_{t-1},\beta_t\mathbf{I})$, β_t is a small positive constant to controls the progress of noise. The whole forward process gradually adds Gaussian noise converts data $\mathbf{x}_0$ to $\mathbf{x}_T$ according to a variance schedule $\beta_1,\ldots,\beta_T$. We then define constants $\alpha_t = 1-\beta_t$, $\overline{\alpha}_t = \prod_{i=1}^{t}\alpha_i$, the $\mathbf{x}_t$ can be directly sampled through the following equation:

$$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\varepsilon, \tag{10}$$

where ε is a standard Gaussian noise.

DDPM Reverse Process: The denoising process is defined by a Markov chain from $\mathbf{x}_T$ to $\mathbf{x}_0$ parameterized by θ

$$p_\theta\left(\mathbf{x}_{0:T}\right) = p\left(\mathbf{x}_T\right)\prod_{t=1}^{T} p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right), \tag{11}$$

where $p(\mathbf{x}_T) = \mathcal{N}(0,\mathbf{I})$, $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ is parameterized as $\mathcal{N}(\mathbf{x}_{t-1};\mu_\theta(\mathbf{x}_t,t),\sum_\theta(\mathbf{x}_t,t))$ with shared parameter θ. According to [39], the covariance matrix $\sum_\theta(\mathbf{x}_t,t)$ can be set as $\beta_i\mathbf{I}$, and the mean $\mu_\theta(\mathbf{x}_t,t)$ can be represented as $\frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t,t))$, where ϵ_θ represents a function approximator designed to forecast $\mathbf{x}_t$.

The Gaussian noise added in the forward process can be gradually eliminated through $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$. Given the reverse process, the generative procedure involves first sampling an $\mathbf{x}_T \sim \mathcal{N}(0,\mathbf{I})$, followed by sampling $\mathbf{x}_{t-1} \sim p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ for each time step $t = T, T-1,\ldots,1$. The final output, $\mathbf{x}_0$, represents the sample data.

DDPM-aided Defense: As shown in Fig. 3, in the context of GDMs, a critical design parameter is the diffusion time step, denoted as t^*, which represents the amount of noise to be injected during the forward process. t^* necessitates an appropriate selection value, if it is set too large, it can distort the semantic labels of the purified image, while excessively low noise levels would make it challenging to counteract adversarial perturbations [27].

Given an adversarial image, the forward process can start with $\mathbf{x}_0 = \mathbf{x}_\alpha$, i.e., $\mathbf{x}_\alpha$ at denoising step $t = 0$. The forward diffusion process can be computed according to (10) as

$$\mathbf{x}_a^{t^*} = \sqrt{\overline{\alpha}_{t^*}}\mathbf{x}_a + \sqrt{1 - \overline{\alpha}_{t^*}}\varepsilon, \qquad (12)$$

where t^* controls the total steps of the forward diffusion. Subsequently, according to (11), we can use a pre-trained model to obtain the reconstructed purified image as

$$p_{\theta^*}\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right) = \mathcal{N}\left(\mathbf{x}_{t-1}; \mu_{\theta^*}\left(\mathbf{x}_t, t\right), \sum\nolimits_{\theta^*}\left(\mathbf{x}_t, t\right)\right), \qquad (13)$$

$$\mathbf{x}_{pure}^{t^*} := \mathbf{x}_a^{t^*}, \qquad (14)$$

where θ^* represents the pre-trained parameters of the DNN. After this process, the purified image $\mathbf{x}_{pure}^0$ can be obtained through t^* steps of the denoising process.

Unlike methods that train DNNs with specific adversarial examples to defend against particular types of attacks, GDMs employ a unique defense strategy. This strategy involves introducing Gaussian noise into adversarial instances and utilizing the reverse denoising process of the diffusion model to restore the original input, thereby enabling downstream classifiers to accurately recognize the denoised images. Consequently, the defense method based on GDMs can effectively counter a wide range of adversarial attacks, including but not limited to gradient-based attacks (such as Projected Gradient Descent (PGD) [57] and AutoAttack [67]), adversarial patch attacks [25], and Backward Pass Differentiable Approximation (BPDA) [68]. This approach demonstrates greater robustness and more practicality in dealing with complex real-world adversarial scenarios.

Distributed DDPM-aided Defense: To optimize the efficiency and adaptability of our DDPM-aided defense in networked environments, we focus on a distributed approach specifically for the denoising phase of the defense mechanism [69]. Recognizing that the denoising steps are more energy-intensive, we allocate these processes across multiple devices within a network. This strategy taps into the parallel processing potential of GDMs, enabling a more flexible and resource-efficient approach, particularly in terms of computational load and energy consumption.

In our distributed model, the denoising phase, which is the critical component of (13), is partitioned across networked devices. We introduce a parameter, n, representing the number of devices involved in the denoising process. The total denoising steps, denoted by t^*, are evenly distributed among these n devices. Consequently, each device is responsible for a fraction of the total denoising steps, given by t_n^*. Let Denoise$(\mathbf{x}_a^{t^*}, t_n^*, \theta^*)$ denote the processing as (13). The distributed denoising process can be formalized as follows:

$$\text{Device}_i: \quad \mathbf{x}_{\text{denoised},i} = \text{Denoise}\left(\mathbf{x}_a^{t^*}, t_n^*, \theta^*\right), \qquad (15)$$

where $\mathbf{x}_{\text{denoised},i}$ is the partial denoised output from the $i\text{th}$ device. Thus, every denoising chain can be performed by different devices. In networked settings, this distributed approach offers several advantages:

- *Energy Efficiency:* By distributing the computational load, devices can operate at lower energy states, reducing the overall energy consumption of the purification process.
- *Scalability:* The distributed model scales effectively with the number of devices in the network, enabling larger-scale defense implementations without a linear increase in resource demand per device.
- *Resilience:* Distributing the process across multiple devices enhances the system's resilience to individual device failures or targeted attacks, as the process does not rely on a single point of execution.

Algorithm 1: Adversarial Purification With DDPM.

```
Input: an input $\mathbf{x}$, denoising step $t^*$ per each purification run
procedure Purify image ($\mathbf{x}$)
  for $t = 1$ to $t^*$ do
    The diffusion process using (12)
  end for
  for $t = t^*$ to 1 do
    The distributed reverse process using (13)
  end for
  return$\mathbf{x}_{pure}$
end procedure
```

The detailed purification process is outlined in Algorithm 1. Utilizing the GDM-based method, the service provider can markedly decrease the likelihood of transmitting an adversarial image to the user. Moreover, unlike other defense mechanisms, the GDM-based scheme offers the advantage of a tunable parameter, the number of defense denoising steps t^*. This parameter not only has implications for energy consumption but also impacts the effectiveness of the defense strategy. Consequently, achieving an optimal balance between energy efficiency and defense performance becomes a critical consideration.

V. Energy-Efficient Resource Allocation Scheme

In this section, we discuss the significance of denoising steps in the GDM-based defense and provide a GDM-based methodology for identifying the optimal denoising step for defense.

A. Problem Analysis

In the context of our proposed secure semantic method enhanced by GDM, the parameter known as the denoising step, denoted as t^*, holds immense significance to the performance of defense and energy consumption. It is critical to consider this parameter given the inherent resource constraints of IoT devices, such as cameras. When t^* is set to a small value, the effectiveness of defense mechanisms relying on the GDM to thwart adversarial attacks is severely limited. Consequently, this limitation can lead to service providers transmitting a higher number of erroneous images to users, ultimately resulting in users detecting these errors and requesting retransmissions from the service provider. This, in turn, escalates resource consumption.

On the contrary, if t^* is configured with a larger value, users receive notably fewer erroneous images. However, this comes at the cost of increased resource utilization by the GDMs during the diffusion process. Hence, the pursuit of an optimal equilibrium between resource consumption in the realms of adversarial attack defense and communication transmission, essentially identifying the ideal value for t^*, assumes paramount importance in our approach. Balancing these factors can lead to enhanced performance, reduced resource consumption, and ultimately, a more resilient and efficient system. Therefore, the quest for this optimal balance is indispensable for ensuring the efficient and effective operation of the system, addressing both security concerns and resource limitations within the IoT landscape.

B. *Optimization Strategy*

The search for the optimal denoising step t^* can be framed as a decision-making problem. In Reinforcement Learning (RL), two fundamental components come into play: the agent and the environment. Specifically, the agent interacts with the environment to learn the best action policy, aiming to maximize the cumulative reward signal. Deep Reinforcement Learning (DRL) amalgamates deep learning techniques with RL, enabling the agent to acquire highly abstract representations and make decisions in intricate environments.

Among the DRL algorithms, Proximal Policy Optimization (PPO) [70] and Soft Actor-Critic (SAC) [71] stand out as popular choices, exhibiting commendable results in solving complex decision-making problems. PPO is a policy optimization algorithm that leverages the concept of proximal policy optimization by maximizing an objective function augmented with a clipping term to update the policy. On the other hand, SAC is a hybrid approach based on both value functions and policies, primarily suited for continuous action spaces in reinforcement learning. SAC simultaneously learns an actor, responsible for action selection, and a critic, responsible for estimating the state-value function. SAC aims to maximize a weighted sum of expected rewards and policy entropy, striking a balance between exploration and exploitation.

However, in scenarios where the environmental state space becomes excessively complex, DRL-based algorithms may encounter challenges such as training instability, low sample efficiency, difficulty converging to the optimal solution, or achieving suboptimal performance. Consequently, we propose a GDM-based resource allocation method. This approach aims to address these challenges and enhance the efficiency and stability of the learning process in complex environments.

C. *GDM-Based Resource Allocation*

To solve this optimization problem, we use a GDM-based optimization method, namely the GDM-based resource allocation scheme, as proposed in [72]. As shown in Fig. 3, we use vector $\boldsymbol{e}$ to represent the entire system environment, including the probability of selecting adversarial images in the PAS image dataset, the energy cost for image transmission and denoising step. In this given environment, we aim to find the optimal number of diffusion time steps t^*.

TABLE I
GDM-Based Method Training Parameters

Parameter	Value
Batch Size N_b	512
Denoising Step N	5
Soft Target Update Parameter τ	0.005
Exploration Noise ϵ	0.05
The Learning Rate of Network ε_θ	10^{-3}
The Learning Rate of Network Q_v	10^{-3}

The GDM network can be denoted as $\pi_\theta(t^*|\boldsymbol{e})$ with parameters θ. According to (11), the reverse process can be represented as

$$\begin{aligned}\pi_\theta\left(t^*|\boldsymbol{e}\right) &= p_\theta\left(t^{*0:N}|\boldsymbol{e}\right)\\ &= \mathcal{N}\left(t^{*N};0,\mathbf{I}\right)\prod_{i=1}^{N} p_\theta\left(t^{*(i-1)}|t^{*i},\boldsymbol{e}\right).\end{aligned} \tag{16}$$

Thus, $p_\theta(t^{*(i-1)}|t^{*i},\boldsymbol{e})$ can be modeled as a Gaussian distribution $\mathcal{N}(t^{*(i-1)};\mu_\theta(t^{*i},\boldsymbol{e},i),\sum_\theta(t^{*i},\boldsymbol{e},i))$, where

$$\sum\nolimits_\theta\left(t^{*i},\boldsymbol{e},i\right) = \beta_i\mathbf{I}, \tag{17}$$

and

$$\mu_\theta\left(t^{*i},\boldsymbol{e},i\right) = \frac{1}{\sqrt{\alpha_i}}\left(t^{*i} - \frac{\beta_i}{\sqrt{1-\bar{\alpha}_i}}\epsilon_\theta\left(t^{*i},\boldsymbol{e},i\right)\right). \tag{18}$$

To obtain the optimal t^*, we first sample $t^{*N}\sim\mathcal{N}(0,\mathbf{I})$, the reverse diffusion chain can be represented as

$$t^{*(i-1)}|t^{*i} = \frac{t^{*i}}{\sqrt{\alpha_i}} - \frac{\beta_i}{\sqrt{\alpha_i\left(1-\bar{\alpha}_i\right)}}\epsilon_\theta\left(t^{*i},\boldsymbol{e},i\right) + \sqrt{\beta_i}\epsilon. \tag{19}$$

Similar to the Q-function in DRL, we define the Energy Consumption Network Q_v, which represents the future cumulative rewards that an agent can obtain when taking a specific policy action in the environment $\boldsymbol{e}$. Therefore, our objective can be expressed as

$$\pi = \arg\min_{\pi_\theta}\mathcal{L}(\theta) = -\mathbb{E}_{t^{*0}\sim\pi_\theta}\left[Q_v\left(\boldsymbol{e},t^{*0}\right)\right]. \tag{20}$$

According to [72], we use the double Q-learning technique [73] to minimize the Bellman operator.

VI. Numerical Results

This section is primarily dedicated to delineating the configuration of the experimental environment and subsequently presenting the obtained results.

A. *Environment Setup*

Our experimental platform is based on an Ubuntu 20.04 system, powered by two Intel(R) Xeon(R) Silver 4110 CPUs and a GeForce RTX 3060 GPU. The specific training parameters utilized in our experiments are outlined in Table I. We

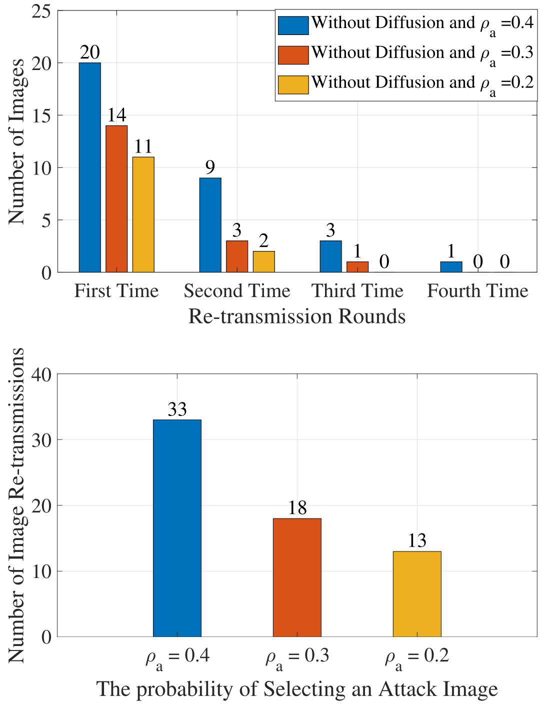


Fig. 4. The image transmission under different attack probabilities, i.e., takes into account the user's request for 50 images, and the probability of selecting an attack image, i.e., ρ_a, in the dataset is 0.2, 0.3, and 0.4, respectively.

quantified energy consumption, finding that each image transmission required 6 W-hours (Wh), and each denoising step in the GDM consumed 0.025 Wh. These energy measurements can be derived by using a calibrated power monitoring setup that tracks the energy used during image transmission and denoising processes [74]. This setup enabled precise energy accounting, ensuring accurate assessment of the energy efficiency of our proposed methods.

B. Results and Analysis

The impact of the probability of selecting attack images ρ_a: In the PAS image dataset, the presence of adversarial attack images determines the probability ρ_a that users receive these attack images during image retrieval requests, which in turn affects the number of user-initiated retransmissions. Considering a user requesting 50 images within a specific category, we analyze the effect of different ρ_a values on retransmission needs. As shown in Fig. 4, there is a direct relationship between ρ_a and the frequency of retransmissions: a lower ρ_a of 0.2 results in just two retransmissions, while an increased ρ_a of 0.4 leads to four retransmissions, with the number of images requiring retransmission jumping from 13 to 33. This increase not only consumes more communication resources but also adversely affects user satisfaction due to the augmented retransmission requirements. These findings emphasize the importance of effectively countering adversarial attacks in the communication service process.

The impact of the Denoising Step t^*: We delve into the effectiveness of GDMs with varying denoising steps t^* in scenarios with different concentrations of adversarial images. As depicted in Fig. 5, for both $\rho_a = 0.2$ and $\rho_a = 0.4$, an increase in t^* results in a notable decrease in both the number of attack images received and the subsequent retransmissions required. Specifically, for $\rho_a = 0.2$, without diffusion defense, a minimum of three retransmissions are necessary, but this reduces to only one retransmission with $t^* = 30$. In scenarios with $\rho_a = 0.4$, the impact of increasing t^* is more pronounced, reducing the initial requirement of four retransmissions to just one at $t^* = 45$. In terms of energy implications, a higher t^* effectively decreases the energy lost due to retransmissions in the image transmission process. However, it's crucial to note that while the energy cost for retransmissions declines, the energy consumption for GDM-based defense increases. For instance, with $t^* = 15$, the total energy cost is 376 Wh. As t^* is raised to 30, although fewer images need retransmission, the increased energy investment in diffusion defense leads to a slight overall increase in energy consumption, reaching 378 Wh. These findings underscore a critical trade-off in the application of GDMs for defending against adversarial attacks: optimizing the denoising step t^* is essential for minimizing retransmissions and enhancing security, but it must be carefully balanced against the associated rise in energy consumption for the defense process.

The effects of GDM-based resource allocation method: Fig. 6 showcases the training progress of the GDM-based resource allocation method, including comparisons with PPO and SAC, under the condition of $\rho_a = 0.3$ for selecting adversarial images. The GDM-based method demonstrates superior stability in learning behavior, outperforming PPO in terms of both convergence speed and stability attainment. In contrast, SAC, despite its initial rapid convergence, struggles to overcome suboptimal solutions in later training stages, as evidenced by its persistently lower training curve. This inferior performance of the GDM-based method is attributed to the GDM's role in policy exploration, which enhances the adaptability of policies and prevents the model from settling into suboptimal states.

Fig. 7 highlights the total energy consumption across different strategies. In scenarios without GDM-based defense, the energy cost stands at 408 Wh. However, after optimizing the denoising steps with PPO, SAC, and the GDM-based methods, all approaches yield a reduction in energy usage. Notably, the GDM-based method achieves the most substantial decrease, lowering the total energy consumption to 385 Wh, which translates to a 23 Wh reduction. These observations reveal that the GDM-based method can maintain a balance between efficient energy usage and robust defense against adversarial attacks, positioning it as a highly viable option for secure mobile networks, where resource constraints are a critical consideration.

VII. Discussion

The applicability of our framework extends beyond the realm of automotive imagery analysis and defense. Given the diverse requirements across various application scenarios, the framework adopts a universal architecture that allows it to adapt to different requests.

- *Broad Applicability of the SIoT System Model [75], [76]:* The first significant aspect of the proposed SIoT model is its broad applicability because of the universal semantic information extraction architecture. For instance, in

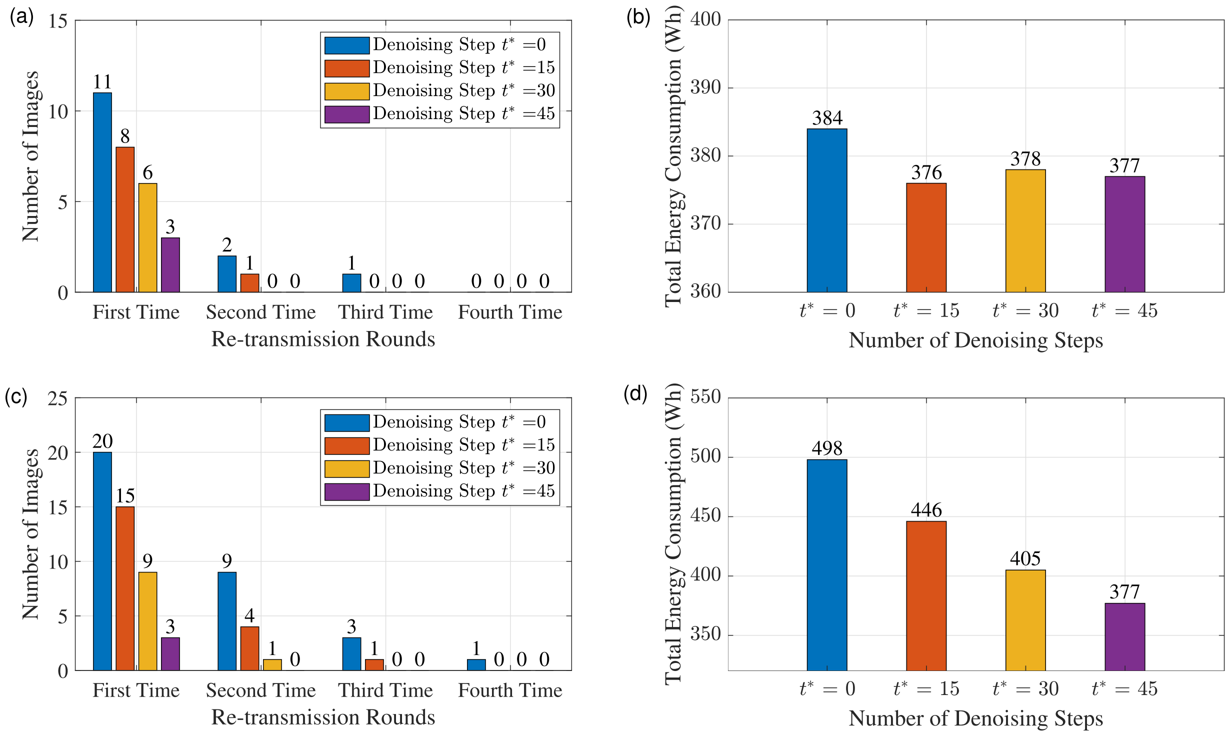


Fig. 5. Image transmission and total energy consumption under different schemes, i.e., the denoising steps for defense are 0, 15, 30, and 45. For subfigures (a) and (b), the probability of containing adversarial images $\rho_a = 0.2$. For subfigures (c) and (d), $\rho_a = 0.4$.

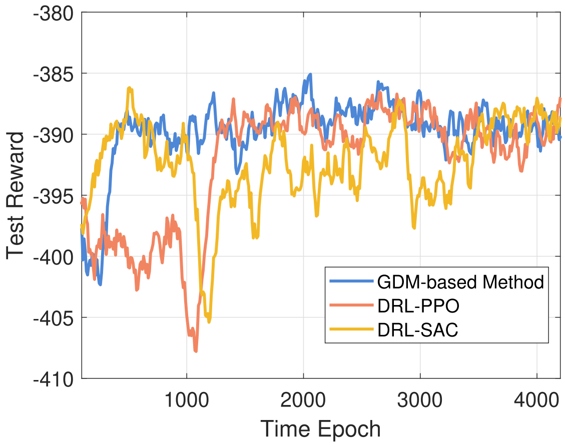


Fig. 6. The training curves for GDM-based resource allocation method, Proximal Policy Optimization (PPO), and Soft Actor-Critic (SAC) optimization. Note that even though the probability of selecting attack images in the database remains constant, the quantity of attack images chosen in each experimental iteration fluctuates, resulting in variable curves. To improve visual clarity, we employ a smoothing function on the training curves.

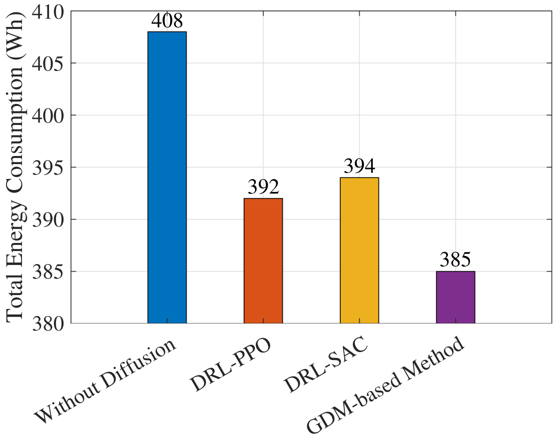


Fig. 7. Energy cost without diffusion defense and other three denoising step optimization methods.

commercial settings or significant public gatherings, our system is instrumental for businesses by facilitating the analysis of market shares for apparel brands, focusing on products such as outerwear and footwear. This versatility underscores the potential of our framework to serve a wide range of practical applications far beyond its initial use case.

- *Generalizability of the GDM-based Defense Mechanism:* By adopting a stochastic approach to introduce and eliminate image noise, the GDM-based method retains the crucial semantic content and effectively counters adversarial alterations. This strategy ensures the framework's resilience against various adversarial strategies, bolstering its applicability in scenarios where security and data integrity are crucial. Our resource allocation algorithm further improves the defense mechanism's flexibility, robustness, and adaptability for real-world application deployment.

Additionally, due to the requirement of the GDM to progressively diffuse and denoise data for adversarial defense, the computational resources needed for this process remain a significant challenge. Some research has made progress in this area, such as optimizing sampling steps [77], or quantitatively compressing the model [78], and so forth. Integrating these improvements with the adversarial defense mechanism can lead to more efficient and energy-saving defense strategies.

VIII. Conclusion

In this paper, we explored the application of GDM in enhancing security and optimizing resource allocation in the SIoT. We utilized an advanced GDM framework to effectively counter adversarial threats in mobile networks. Additionally, recognizing the resource limitations inherent in mobile devices, we developed a GDM-based resource allocation strategy, aiming to strike a balance between efficient image transmission and robust adversarial defense mechanisms. Our numerical analysis revealed that this GDM-based method significantly improves resource utilization. Notably, it achieved a 5.64% reduction in energy consumption, alongside a drastic decrease in the number of retransmissions, which fell from 18 to just 6. This highlights the efficacy of GDM not only as a defense tool against adversarial attacks but also as a means to enhance energy efficiency in mobile networks, offering valuable insights for future research and development.

References

[1] D. C. Nguyen et al., "6G Internet of Things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2022.

[2] H. Zhang, D. Wang, S. Wu, W. Guan, and X. Liu, "USTB 6G: Key technologies and metaverse applications," *IEEE Wireless Commun.*, vol. 30, no. 5, pp. 112–119, Oct. 2023.

[3] W. Yang et al., "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Commun. Survey Tut.*, vol. 25, no. 1, pp. 213–250, First Quarter 2023.

[4] H. Zhang, H. Wang, Y. Li, K. Long, and A. Nallanathan, "DRL-driven dynamic resource allocation for task-oriented semantic communication," *IEEE Trans. Commun.*, vol. 71, no. 7, pp. 3992–4004, Jul. 2023.

[5] H. Du et al., "Rethinking wireless communication security in semantic Internet of Things," *IEEE Wireless Commun.*, vol. 30, no. 3, pp. 36–43, Jun. 2023.

[6] H. Zhang, H. Wang, Y. Li, K. Long, and V. C. Leung, "Toward intelligent resource allocation on task-oriented semantic communication," *IEEE Wireless Commun.*, vol. 30, no. 3, pp. 70–77, Jun. 2023.

[7] H. Zhang, L. Feng, X. Liu, K. Long, and G. K. Karagiannidis, "User scheduling and task offloading in multi-tier computing 6G vehicular network," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 446–456, Feb. 2023.

[8] W. Du, T. Wang, H. Zhang, Y. Dong, and Y. Li, "Joint resource allocation and trajectory optimization for completion time minimization for energy-constrained UAV communications," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 4568–4579, Apr. 2023.

[9] W. C. Ng, H. Du, W. Y. B. Lim, Z. Xiong, D. T. Niyato, and C. Miao, "Stochastic resource allocation for semantic communication-aided virtual transportation networks in the metaverse," 2022, *arXiv:2208.14661*. [Online]. Available: https://api.semanticscholar.org/CorpusID:251953606

[10] Y. Liu, K. A. Hassan, M. Karlsson, Z. Pang, and S. Gong, "A data-centric Internet of Things framework based on azure cloud," *IEEE Access*, vol. 7, pp. 53839–53858, 2019.

[11] S. Verma, Y. Kawamoto, and N. Kato, "A smart internet-wide port scan approach for improving IoT security under dynamic WLAN environments," *IEEE Internet Things J.*, vol. 9, no. 14, pp. 11951–11961, Jul. 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:244894332

[12] K. Eykholt et al., "Physical adversarial examples for object detectors," 2018, *arXiv: 1807.07769*. [Online]. Available: https://api.semanticscholar.org/CorpusID:49904930

[13] S. Verma, Y. Kawamoto, Z. M. Fadlullah, H. Nishiyama, and N. Kato, "A survey on network methodologies for real-time analytics of massive IoT data and open research issues," *IEEE Commun. Survey Tut.*, vol. 19, no. 3, pp. 1457–1477, Third Quarter 2017.

[14] J. Zheng, H. Zhang, J. Kang, L. Gao, J. Ren, and D. Niyato, "Covert federated learning via intelligent reflecting surfaces," *IEEE Trans. Commun.*, vol. 71, no. 8, pp. 4591–4604, Aug. 2023.

[15] C. Szegedy et al., "Intriguing properties of neural networks," 2013, *arXiv:1312.6199*.

[16] A. Arnab, O. Miksik, and P. H. Torr, "On the robustness of semantic segmentation models to adversarial attacks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 888–897.

[17] H. Zhang and J. Wang, "Towards adversarially robust object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 421–430.

[18] Y. J. Jia et al., "Fooling detection alone is not enough: Adversarial attack against multiple object tracking," in *Proc. Int. Conf. Learn. Representations*, 2020, pp. 1–15, [Online]. Available: https://openreview.net/pdf?id=rJl31TNYPr

[19] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access*, vol. 6, pp. 14410–14430, 2018.

[20] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, "Universal adversarial perturbations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1765–1773.

[21] T. Bai, J. Luo, J. Zhao, B. Wen, and Q. Wang, "Recent advances in adversarial training for adversarial robustness," 2021, *arXiv:2102.01356*.

[22] H. Zhang, H. Chen, Z. Song, D. Boning, I. S. Dhillon, and C.-J. Hsieh, "The limitations of adversarial training and the blind-spot attack," 2019, *arXiv: 1901.04684*.

[23] Y. Wang, D. Zou, J. Yi, J. Bailey, X. Ma, and Q. Gu, "Improving adversarial robustness requires revisiting misclassified examples," in *Proc. Int. Conf. Learn. Representations*, 2019, pp. 1–14. [Online]. Available: https://openreview.net/pdf?id=rklOg6EFwS

[24] L. L. Ankile, A. Midgley, and S. Weisshaar, "Denoising diffusion probabilistic models as a defense against adversarial attacks," 2023, *arXiv:2301.06871*.

[25] C. Kang, Y. Dong, Z. Wang, S. Ruan, H. Su, and X. Wei, "DIFFender: Diffusion-based adversarial defense against patch attacks in the physical world," 2023, *arXiv:2306.09124*. [Online]. Available: https://api.semanticscholar.org/CorpusID:259165314

[26] J. Wang, Z. Lyu, D. Lin, B. Dai, and H. Fu, "Guided diffusion model for adversarial purification," 2022, *arXiv:2205.14969*. [Online]. Available: https://api.semanticscholar.org/CorpusID:249192338

[27] W. Nie, B. Guo, Y. Huang, C. Xiao, A. Vahdat, and A. Anandkumar, "Diffusion models for adversarial purification," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 16805–16827. [Online]. Available: https://api.semanticscholar.org/CorpusID:248811081

[28] H. Cao et al., "A survey on generative diffusion model," 2022, *arXiv:2209.02646*. [Online]. Available: https://api.semanticscholar.org/CorpusID:252090040

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

[30] Q. Lan et al., "What is semantic communication? A view on conveying meaning in the era of machine intelligence," 2021, *arXiv:2110.00196*. [Online]. Available: https://api.semanticscholar.org/CorpusID:238253071

[31] A. Cavagn, N. Li, A. Iosifidis, and Q. Zhang, "Semantic communication enabling robust edge intelligence for time-critical IoT applications," 2022, *arXiv:2211.13787*. [Online]. Available: https://api.semanticscholar.org/CorpusID:254017713

[32] N. Farsad, M. Rao, and A. J. Goldsmith, "Deep learning for joint source-channel coding of text," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2018, pp. 2326–2330. [Online]. Available: https://api.semanticscholar.org/CorpusID:3400480

[33] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:219792180

[34] Q. Hu, G. Zhang, Z. Qin, Y. Cai, and G. Yu, "Robust semantic communications against semantic noise," in *Proc. IEEE 96th Veh. Technol. Conf.*, 2022, pp. 1–6. [Online]. Available: https://api.semanticscholar.org/CorpusID:246634062

[35] C. Han Lee, J. W. Lin, P.-H. Chen, and Y.-C. Chang, "Deep learning-constructed joint transmission-recognition for Internet of Things," *IEEE Access*, vol. 7, pp. 76547–76561, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:195428320

[36] Q. Pan et al., "Image segmentation semantic communication over Internet of Vehicles," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2022, pp. 1–6. [Online]. Available: https://api.semanticscholar.org/CorpusID: 252815484

[37] H. Xie, Z. Qin, and G. Y. Li, "Task-oriented multi-user semantic communications for VQA task," 2021, *arXiv:2108.07357*.

[38] J. Kang et al., “Personalized saliency in task-oriented semantic communications: Image transmission and performance analysis,” *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 186–201, Jan. 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:252531406

[39] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” 2020, *arXiv: 2006.11239*. [Online]. Available: https://api.semanticscholar.org/CorpusID:219955663

[40] K. Gong, K. A. Johnson, G. E. Fakhri, Q. Li, and T. Pan, “PET image denoising based on denoising diffusion probabilistic models,” 2022, *arXiv:2209.06167*. [Online]. Available: https://api.semanticscholar.org/CorpusID:252211768

[41] B. B. Moser, F. Raue, S. Frolov, J. Hees, S. M. Palacio, and A. R. Dengel, “Hitchhiker’s guide to super-resolution: Introduction and recent advances,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 9862–9882, Aug. 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:252545235

[42] B. Kawar et al., “Imagic: Text-based real image editing with diffusion models,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 6007–6017. [Online]. Available: https://api.semanticscholar.org/CorpusID:252918469

[43] A. C. Li, M. Prabhudesai, S. Duggal, E. L. Brown, and D. Pathak, “Your diffusion model is secretly a zero-shot classifier,” 2023, *arXiv:2303.16203*. [Online]. Available: https://api.semanticscholar.org/CorpusID:257771787

[44] S. Chen, P. Sun, Y. Song, and P. Luo, “DiffusionDet: Diffusion model for object detection,” 2022, *arXiv:2211.09788*. [Online]. Available: https://api.semanticscholar.org/CorpusID:253581633

[45] T. Amit, E. Nachmani, T. Shaharbany, and L. Wolf, “SegDiff: Image segmentation with diffusion probabilistic models,” 2021, *arXiv:2112.00390*. [Online]. Available: https://api.semanticscholar.org/CorpusID:244773420

[46] X. L. Li, J. Thickstun, I. Gulrajani, P. Liang, and T. Hashimoto, “Diffusion-LM improves controllable text generation,” 2022, *arXiv:2205.14217*. [Online]. Available: https://api.semanticscholar.org/CorpusID:249192356

[47] R. Huang et al., “FastDiff: A fast conditional diffusion model for high-quality speech synthesis,” in *Proc. Int. Joint Conf. Artif. Intell.*, 2022, pp. 4157–4163. [Online]. Available: https://api.semanticscholar.org/CorpusID:248300058

[48] A. Singh and B. K. Sikdar, “Adversarial attack and defence strategies for deep-learning-based IoT device classification techniques,” *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2602–2613, Feb. 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:245520127

[49] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” 2014, *arXiv:1412.6572*. [Online]. Available: https://api.semanticscholar.org/CorpusID:6706414

[50] Y. Zhong and W. Deng, “Towards transferable adversarial attack against deep face recognition,” *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1452–1466, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:215745571

[51] H. Liang, E. He, Y. Zhao, Z. Jia, and H. Li, “Adversarial attack and defense: A survey,” *Electronics*, vol. 11, pp. 1–19, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/8/1283

[52] D. Wang, W. Yao, T. Jiang, G. Tang, and X. Chen, “A survey on physical adversarial attack in computer vision,” 2022, *arXiv:2209.14262*. [Online]. Available: https://api.semanticscholar.org/CorpusID:252567921

[53] H. Zhang and J. Wang, “Towards adversarially robust object detection,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 421–430. [Online]. Available: https://api.semanticscholar.org/CorpusID:198229380

[54] M. Sharif, S. Bhagavatula, L. Bauer, and M. K. Reiter, “Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition,” in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2016, pp. 1528–1540. [Online]. Available: https://api.semanticscholar.org/CorpusID:207241700

[55] M. Sharif, S. Bhagavatula, L. Bauer, and M. K. Reiter, “A general framework for adversarial examples with objectives,” *ACM Trans. Privacy Secur.*, vol. 22, pp. 1–30, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:132058467

[56] J. Hu et al., “Potential auto-driving threat: Universal rain-removal attack,” *iScience*, vol. 26, no. 9, 2023, Art. no. 107393. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2589004223014700

[57] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, “Towards deep learning models resistant to adversarial attacks,” 2017, *arXiv: 1706.06083*. [Online]. Available: https://api.semanticscholar.org/CorpusID:3488815

[58] Y. Dong et al., “Boosting adversarial attacks with momentum,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 9185–9193. [Online]. Available: https://api.semanticscholar.org/CorpusID:4119221

[59] Z. Huang and T. Zhang, “Black-box adversarial attack with transferable model-based embedding,” 2019, *arXiv: 1911.07140*. [Online]. Available: https://api.semanticscholar.org/CorpusID:208139568

[60] A. S. Rakin, Z. He, and D. Fan, “Parametric noise injection: Trainable randomness to improve deep neural network robustness against adversarial attack,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 588–597. [Online]. Available: https://api.semanticscholar.org/CorpusID:53716271

[61] N. Papernot, P. Mcdaniel, X. Wu, S. Jha, and A. Swami, “Distillation as a defense to adversarial perturbations against deep neural networks,” in *Proc. IEEE Symp. Secur. Privacy*, 2015, pp. 582–597. [Online]. Available: https://api.semanticscholar.org/CorpusID:2672720

[62] J. Folz, S. M. Palacio, J. Hees, D. Borth, and A. R. Dengel, “Adversarial defense based on structure-to-signal autoencoders,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2018, pp. 3568–3577. [Online]. Available: https://api.semanticscholar.org/CorpusID:4021558

[63] C. Xie, J. Wang, Z. Zhang, Z. Ren, and A. L. Yuille, “Mitigating adversarial effects through randomization,” 2017, *arXiv: 1711.01991*. [Online]. Available: https://api.semanticscholar.org/CorpusID:3526769

[64] W. Xu, D. Evans, and Y. Qi, “Feature squeezing: Detecting adversarial examples in deep neural networks,” 2017, *arXiv: 1704.01155*. [Online]. Available: https://api.semanticscholar.org/CorpusID:3851184

[65] U. Shaham et al., “Defending against adversarial images using basis functions transformations,” 2018, *arXiv: 1803.10840*. [Online]. Available: https://api.semanticscholar.org/CorpusID:4549456

[66] G. Tolias, F. Radenovic, and O. Chum, “Targeted mismatch adversarial attack: Query with a flower to retrieve the tower,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 5036–5045. [Online]. Available: https://api.semanticscholar.org/CorpusID:201657307

[67] F. Croce and M. Hein, “Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks,” in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 2206–2216. [Online]. Available: https://api.semanticscholar.org/CorpusID:211818320

[68] A. Athalye, N. Carlini, and D. A. Wagner, “Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 274–283. [Online]. Available: https://api.semanticscholar.org/CorpusID:3310672

[69] H. Du et al., “User-centric interactive AI for distributed diffusion model-based AI-generated content,” 2023, *arXiv:2311.11094*.

[70] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” 2017, *arXiv: 1707.06347*. [Online]. Available: https://api.semanticscholar.org/CorpusID:28695052

[71] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” 2018, *arXiv: 1801.01290*. [Online]. Available: https://api.semanticscholar.org/CorpusID:28202810

[72] H. Du, J. Wang, D. Niyato, J. Kang, Z. Xiong, and D. I. Kim, “AI-generated incentive mechanism and full-duplex semantic communications for information sharing,” *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2981–2997, Sep. 2023.

[73] H. V. Hasselt, “Double Q-learning,” in *Proc. 23rd Int. Conf. Neural Inf. Process. Syst.*, 2010, pp. 2613–2621. [Online]. Available: https://api.semanticscholar.org/CorpusID:5155799

[74] H. Du et al., “Diffusion-based reinforcement learning for edge-enabled AI-generated content services,” *IEEE Trans. Mobile Comput.*, early access, Jan. 19, 2024, doi: 10.1109/TMC.2024.3356178.

[75] K. Li, B. P. L. Lau, X. Yuan, W. Ni, M. Guizani, and C. Yuen, “Toward ubiquitous semantic metaverse: Challenges, approaches, and opportunities,” *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21855–21872, Dec. 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259847638

[76] C. Liang et al., “Generative AI-driven semantic communication networks: Architecture, technologies and applications,” 2024, *arXiv:2401.00124*. [Online]. Available: https://api.semanticscholar.org/CorpusID:266693775

[77] Z. Duan, C. Wang, C. Chen, J. Huang, and W. Qian, “Optimal linear subspace search: Learning to construct fast and high-quality schedulers for diffusion models,” in *Proc. 32nd ACM Int. Conf. Inf. Knowl. Manage.*, 2023, pp. 463–472. [Online]. Available: https://api.semanticscholar.org/CorpusID:258866130

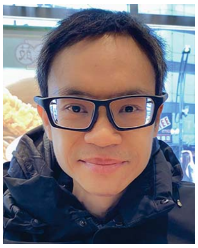
[78] Y. Shang, Z. Yuan, B. Xie, B. Wu, and Y. Yan, “Post-training quantization on diffusion models,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1972–1981. [Online]. Available: https://api.semanticscholar.org/CorpusID:254069829

Jie Zheng received the BSc degree in communications engineering from Nanchang University, in 2008, and the PhD degree from the Department of Telecommunications Engineering, Xidian University, China, in 2014. He is currently an associate professor with the School of Information Science and Technology, Northwest University, Xi'an, China. His research interests include energy-efficient transmission, wireless resource allocation, and edge intelligence.

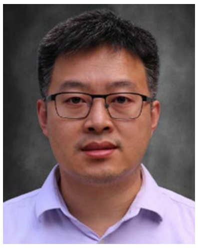
Baoxia Du received the BS degree from the Shandong University of Science and Technology, Qingdao, China, in 2020. He is currently working toward the MS degree with the School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin, China. His current research interests include deep learning, computer vision, and semantic communication.

Hongyang Du received the BS degree from Beijing Jiaotong University, Beijing, China, in 2021. He is currently working toward the PhD degree with the School of Computer Science and Engineering, Energy Research Institute at NTU, Nanyang Technological University, Singapore, under the Interdisciplinary Graduate Program. He was recognized as an exemplary reviewer of *IEEE Transactions on Communications*, in 2021. He was the recipient of IEEE Daniel E. Noble Fellowship Award, in 2022. His research interests include semantic communications, resource allocation, and communication theory.

Jiawen Kang received the PhD degree from the Guangdong University of Technology, China, in 2018. He was 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. His research interests include blockchain, security, and privacy protection in wireless communications and networking.



Dusit Niyato (Fellow, IEEE) received the BEng degree from the King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1999, and the PhD degree in electrical and computer engineering from the University of Manitoba, Canada, in 2008. He is currently a professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests are in the areas of the Internet of Things (IoT), machine learning, and incentive mechanism design.



Haijun Zhang (Fellow, IEEE) was a post-doctoral research fellow with the Department of Electrical and Computer Engineering, The University of British Columbia (UBC), Canada. He is currently a full professor and an associate dean of the University of Science and Technology Beijing, China. He received the IEEE ComSoc Young Author Best Paper Award, in 2017, the IEEE CSIM Technical Committee Best Journal Paper Award, in 2018, and the IEEE ComSoc Asia–Pacific Best Young Researcher Award, in 2019. He serves/served as the track co-chair for VTC Fall 2022 and WCNC 2020/2021; the Symposium chair for Globecom'19; the TPC co-chair for INFOCOM 2018 Workshop on Integrating Edge Computing, Caching, and Offloading in Next Generation Networks; and the general co-chair for GameNets'16. He serves/served as an editor for *IEEE Transactions on Information Forensics and Security*, *IEEE Transactions on Communications*, and *IEEE Transactions on Network Science and Engineering*. He is a distinguished lecturer of IEEE.