Energy-efficient Semantic Communication for Aerial-aided Edge Networks

Guhan Zheng, *Member, IEEE*, Qiang Ni, *Senior Member, IEEE*, Keivan Navaie, *Senior Member, IEEE*,
Haris Pervaiz, *Member, IEEE*, Aryan Kaushik, *Member, IEEE*, and Charilaos Zarakovitis, *Member, IEEE*

Abstract—Semantic communication holds promise for integration into future wireless networks, offering a potential enhancement in network spectrum efficiency. However, implementing semantic communication in aerial-aided edge networks (AENs) introduces unique challenges. Within AENs, semantic communication strategically substitutes part of the communication load with the computation load, aiming to boost spectrum efficiency. This departure from traditional communication paradigms introduces novel challenges, particularly in terms of energy efficiency. Furthermore, by adding complexity, the use of a semantic coder based on machine learning (ML) in AENs encounters real-time updating challenges, further amplifying energy costs in these complex and energy-limited environments. To address these challenges, we propose an energy-efficient semantic communication system tailored for AENs. Our approach includes a mathematical analysis of semantic communication energy consumption within AENs. To enhance energy efficiency, we introduce an energy-efficient game-theoretic incentive mechanism (EGTIM) designed to optimize semantic transmission within AENs. Moreover, considering the accurate and energy-efficient updating of semantic coders in AENs, we present a game-theoretic efficient distributed learning (GEDL) framework, building upon the foundations of the renewed EGTIM. Simulation results validate the effectiveness of our proposed EGTIM in improving energy efficiency. Additionally, the presented GEDL framework exhibits remarkable performance by increasing model training accuracy and concurrently decreasing training energy consumption.

Index Terms—Semantic communication, energy efficiency, game theoretic, distributed learning.

I. INTRODUCTION

THE 6G wireless communication is considered a three-dimensional (3-D) communication network fully assisted by edge cloud facilities [1]. The aerial facilities with edge clouds, i.e., aerial edge clouds (AECs), are anticipated to provide abundant storage and computing resources to subscribers alongside the terrestrial edge clouds (TECs). Subscribers are allowed to access these edge facilities to offload computationally sensitive tasks for rapid processing or acquire massive image/video information etc. [2].

Aerial-aided edge networks (AENs), however, introduce unprecedented spectrum resource and energy challenges [3]. The

deployment of edge networks poses a significant volume of task interactions and hence dramatically increases the volume of communication transmission tasks [4]. It means the network needs to provide more data transmission within the limited spectrum resources to ensure network quality of service (QoS). How to optimize the spectrum efficiency of the AENs therefore becomes an urgent concern.

Semantic communication [5] looks promising in improving spectrum efficiency. It utilizes the semantic coder largely based on machine learning (ML) instead of the conventional communication coder. The ML-based semantic encoder extracts the specific meaning of the input data, thus significantly reducing communication transmission bits [6]. Several studies have investigated the application of semantic communication in transmitting images [7]/text [8]/video [9]/speech [10], etc. These studies have all demonstrated the effectiveness of semantic communication in improving spectrum efficiency and network QoS. Semantic communication is hence also considered to be one of the essential applications of 6G communication [11].

Several studies already investigated the employment of semantic communication for AEC devices. Kang et. al [12] proposed a new aerial semantic image transmission paradigm based on deep reinforcement learning (DRL) to improve the transmission accuracy of unmanned aerial vehicles (UAVs). In [13], semantic communication was integrated into their presented DRL framework for increasing communication reliability and decreasing the latency of air-ground networks. Kang et. al [14] introduced a task-oriented semantic communication framework for UAVs. The UAV sends only the necessary images to the required users rather than all images, thus reducing its energy consumption. However, these existing studies for semantic communication concentrate more on AEC devices and end-to-end semantic coder design. They also neglect to take into account the influence of semantic communication applications for AENs.

There are some outstanding challenges for semantic communication in AENs. First, the utilization of trained semantic coders in AENs raises the sophisticated network energy optimization challenge. Optimizing energy efficiency for AENs becomes a crucial concern as semantic communication redistributes communication load onto computational resources, thereby enhancing spectral efficiency. This shift necessitates addressing energy optimization challenges associated with the transformation of energy utilization patterns. Developing an energy-efficient semantic communication framework is paramount to effectively manage this issue and ensure optimal energy utilization in AENs.
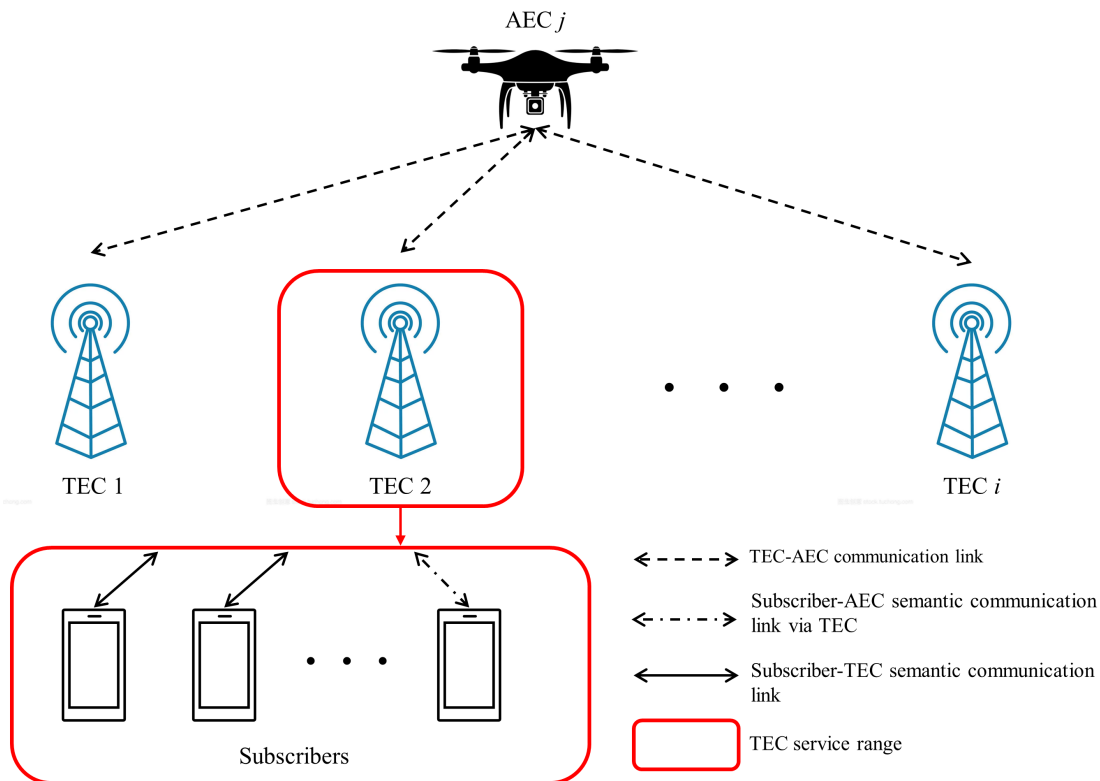
Fig. 1: Proposed system model.

In [15] and [16], different semantic communication frameworks were proposed, they, however, ignore the energy cost of semantic communications. Yang et. al [17] proposed an energy saving semantic coder utilization scheme over wireless networks with rate splitting. Moreover, the optimization of semantic communication energy cost over 2-D edge networks was investigated in [18], [19]. Nonetheless, these approaches did not account for the unique scenario and challenges of AENs. For instance, mobile network operators (MNOs) incur additional monetary and energy costs to implement the AECs. A certain percentage of AEC energy is inevitably consumed in air hover. How to improve the energy efficiency of semantic communication over AENs thus remains a challenge.

Semantic communication also requires real-time updating ML-based semantic coders for various specific content [20]. Designing various distributed learning frameworks for semantic coder updating in different networks is one of the main challenges of semantic communication in networks [21]. The existing studies are limited. Shi et. al [22] proposed a semantic communication framework for general 2-D edge networks and utilized federated learning (FL) to update the ML-based semantic coder. Similarly, Qin et. al [23] investigated the FL framework in semantic communication enabled networks. Furthermore, considering the properties of vehicular networks, e.g., dynamic, an MSFTL [24] framework was designed and tailored for semantic coder updating in vehicular networks. Nevertheless, these 2-D FL frameworks for updating semantic coders are not suitable to be deployed on the AENs directly. Updating semantic coders faces several unique challenges in AENs. For instance, the distributions of training data from

different coder owners are frequently not independent and identically distributed (non-IID) [25]. Furthermore, as the AECs are energy-limited, the energy efficiency of the learning framework has to be considered. How to timely update the semantic coder accurately and energy-efficiently in an AEN with non-IID training data is one of the challenges for semantic communication to apply in AENs. To the best of our knowledge, designing an effective learning framework to update the semantic coders in AENs has not been widely studied.

To address semantic coder updating challenges in AENs while optimizing the energy efficiency of semantic communication, in this paper, we propose a novel energy-efficient semantic communication system for AENs. We then discuss the resource allocation problem during semantic communication usage. A new energy-efficient game theoretic incentive mechanism (EGTIM) based on the proposed semantic communication system is presented to optimize the network energy efficiency in a fair way. In addition, we propose a game theoretic efficient distributed learning (GEDL) framework for semantic coders updating in AENs. It renews the proposed EGTIM and combines EGTIM with a conventional distributed learning approach to update semantic coders accurately and energy efficiently.

The major contributions of this paper are summarized as follows:

- We propose a novel energy-efficient semantic communication system to support AENs. In this system, AECs and TECs provide edge services to users via employed ML-based semantic coders. Moreover, it enables edge devices to schedule the processing locations of computational

tasks due to semantic communication intelligently to improve the energy efficiency of the AEN. The AENs' spectral efficiency and the QoS thus can be improved.

- In particular, we present a new EGTIM in the proposed semantic communication system to further improve the energy efficiency of AENs. The computational and communication workload of the AEC and TECs to perform semantic communication are developed as a Stackelberg game. It is designed to maximise the energy efficiency of the AEN while proportional fairness maximising the service revenue of each edge device in the network.

- A GEDL framework is proposed for semantic coder updating in AENs. It is based on our designed renewed EGTIM for semantic coder updating. Compared to FL, it significantly improves the semantic coder accuracy in IID/non-IID scenarios and improves the training energy efficiency by retraining the model after federated aggregation in the AEC.

The remainder of this paper is organized as follows. We describe the proposed system model in Section II. In Section III, the game problem formulation and the proposed EGTIM are presented. Section IV describes the presented GEDL framework for semantic coder updating in AENs. Simulation results are shown in Section V. Finally, we conclude this paper in Section VI.

II. SYSTEM MODEL

In this paper, we consider a three-dimensional edge network aided by an AEC j (Fig. 1). The TECs provide edge services via semantic coders to subscribers on the terrestrial. An AEC j with semantic coders hovers in the air and assists TECs in providing edge services to subscribers. The semantic communication task processing can be performed in TECs and AEC j. Furthermore, to optimize the allocation of network energy resources, semantic extraction task locations allow for replacement. For instance, in the case of a TEC with insufficient computational resources, a part of the semantic extraction tasks can be provided to the AEC via conventional communication. The semantic extraction tasks are calculated in new locations and the semantic information is then transmitted to the subscribers. In addition, the semantic coders in TECs and AEC j need to be updated in real-time according to different tasks.

We assume that the energy power of AEC j hovers in the air is P_j^l. The free computational capability (free CPU-cycle frequency) of AEC j is f_j. Moreover, there are I TECs within the service range of AEC j that provide edge service to subscribers. We denote the data size of tasks that each TEC i prepares to transmit semantic extraction tasks to AEC j as $m_{i,j}$ bits. The semantic encoder execution latency of TEC i for these tasks can be expressed as:

$$T_i^C = \frac{am_{i,j}}{f_i}, \tag{1}$$

where f_i is the CPU-cycle frequency of TEC i to process these semantic extraction tasks and the unit is cycles/s. Further, a is

the pure number of CPU-cycle consumed to calculate each 1-bit [26]. According to [27], the computing power of the TEC i can be denoted by

$$P_i^C = \kappa f_i^3, \tag{2}$$

where κ is the CPU architecture-related coefficient and is considered to be the same across various devices [27]. We thus have the execution energy consumption of TEC i for these semantic extraction tasks as:

$$E_i^C = \kappa a m_{i,j} f_i^2. \tag{3}$$

Similarly, in the case of the TEC i provides the $m_{i,j}$ bits semantic extraction task to the AEC j, the execution latency and energy consumption of AEC j can be expressed as:

$$T_j^C = \frac{am_{i,j}}{f_{j,i}}, \tag{4}$$

$$E_j^C = \kappa a m_{i,j} f_{j,i}^2, \tag{5}$$

where $f_{j,i}$ is the CPU-cycle frequency that AEC j allocate to the task bits $m_{i,j}$. To ensure the QoS and subscribers' satisfaction, in this paper, we assume $f_{j,i} = f_i$.

In addition, during the semantic extraction task providing process, the data transmission rate of the TEC i to the AEC j can be denoted by

$$r_i^T = B_i \log_2(1 + \frac{p_i g_i}{\sigma^2}), \tag{6}$$

where B_i is the bandwidth of the communication channel between the TEC i and the AEC j. Further, p_i, g_i and σ are the transmission power, channel gain and additive white Gaussian noise (AWGN) power in this channel, respectively. We then can have the transmission delay as:

$$T_i^T = \frac{m_{i,j}}{r_i^T} = \frac{m_{i,j}}{B_i \log_2(1 + \frac{p_i g_i}{\sigma^2})}. \tag{7}$$

Thus, the transmission energy consumption is

$$E_i^T = p_i T_i^T = \frac{p_i m_{i,j}}{B_i \log_2(1 + \frac{p_i g_i}{\sigma^2})}. \tag{8}$$

As the completed semantic extraction task result size is much smaller than the task size, resembling [28], [29], we ignore the transmit delay and energy consumption of transmission tasks after semantic extraction.

For easy reference, the main parameters and their description used throughout this paper are presented in Table I.

III. STACKELBERG GAME THEORETIC INCENTIVE MECHANISM DESIGN

To improve the AEN energy efficiency, the fairness optimizing assignment of the number of semantic compression tasks processed by the TECs and the AEC is essential. Because we found that when AEC edge resources are underutilized, the hovering of airborne devices takes longer for the same amount of energy. This results in a significant amount of energy being wasted for hover rather than performing economically efficient semantic message computing/transmission. Therefore, we construct the TECs and the AEC interaction as a Stackelberg game [30] from the economic perspective. Its objective is to enable

TABLE I: NOTATION DEFINITION

Symbol	Definition
f	CPU-cycle frequency
m	Bits of transmitted data
m^t	Bits of transmitted data during coder updating
T^C	Semantic execution latency
T^T	Transmission delay
E^C	Semantic execution energy consumption
E^T	Transmission energy consumption
E_j	Available hover energy of AEC j
P_j^l	Energy power of AEC j hover in the air
P_j^n	Energy power of AEC j utilizing with no economic benefit
b	Unit bonus price
κ	CPU architecture-related coefficient
α	Net income monetary parameter
β	Cost price monetary parameter
γ	Sale price monetary parameter

energy wasted for AEC hovering to be utilized for semantic extraction task processing to improve the network energy efficiency. It thus incentivises TECs to additionally provide partial semantic extraction tasks to the AEC in fairness, where the AEC is trusted. The Stackelberg game is comprised of a leader and followers, where the followers change their policies according to the policies developed by the leader. Thus, the proposed incentive mechanism consists of the game at the AEC (leader) and the game at TECs (followers), which we elaborate on in detail in the following two subsections.

A. Game at the AEC

We elaborate on the game of the AEC in this subsection. First, AEC j can reaps more processing revenues when more semantic extraction tasks from TECs are processed. The energy used to process these tasks can be thought of as saved hover energy. Correspondingly, the reduced hovering time also reduces the revenue of the AEC j for processing regular tasks, e.g., computing offloading, and computing tasks from TECs. Furthermore, a portion of the total revenue of the AEC j is also required to be paid to the TEC to create incentives. Without loss of generality, we define the monetary utility U_j of the AEC j as:

$$U_j = N_j + R_j - B_j - G_j. \tag{9}$$

where N_j is the net income of AEC j to transmit semantic extraction tasks to subscribers. The R_j is the cost price of AEC j's energy to process semantic tasks. This energy is originally wasted for the hover. Moreover, B_j is the bonus paid to TECs providing the tasks and G_j is the revenue loss of AEC j due to the transfer of some holdup energy to the additional semantic extraction execution resulting in a reduction of the holdup time. We consider the net income N_j and cost price R_j as the energy consumption similar to the previous study [31]. We have

$$N_j(\mathbf{m_{i,j}}) = \alpha \sum_{i=1}^{I} E_j^C, \tag{10}$$

$$R_j(\mathbf{m_{i,j}}) = \beta \sum_{i=1}^{I} E_j^C, \tag{11}$$

where $\alpha > 0$ is the net income monetary parameter and $\beta > 0$ is the cost price monetary parameter of energy. We further set γ is the sale price monetary parameter and $\gamma = \alpha + \beta$.

The revenue loss G_j depends on the aerial hover time and we define it as revenue loss of not performing its regular tasks. To obtain the G_j, we first formula the residence time of AEC j without additional semantic compression tasks as:

$$T_j^0(\mathbf{m_{i,j}}) = \frac{E_j}{P_j^l + P_j^n + \kappa f_{j0}^3}, \tag{12}$$

where E_j is the available hover energy of AEC j and f_{j0} is the CPU-cycle frequency required for the AEC j to perform its regular tasks. Further, P_j^l is the AEC power for hovering in the sky and P_j^n is the AEC utilizing power with no economic benefit. We then have the residence time of AEC j with additional semantic compression tasks as:

$$T_j^1(\mathbf{m_{i,j}}) = \frac{E_j - e_j}{P_j^l + P_j^n + \kappa f_{j0}^3}, \tag{13}$$

where $e_j = \sum_{i=1}^{I} E_j^C$ is the energy consumption of the AEC j to execute the provided tasks. This is due to the energy consumption e_j of processing semantic tasks reducing the total energy E_j of the AEC. Therefore, we can find the G_j as:

$$G_j(\mathbf{m_{i,j}}) = \gamma \kappa f_{j0}^3 (T_j^0 - T_j^1), \tag{14}$$

where γ is the sale price monetary parameter as energy here is not sold and receives zero economic benefit.

In addition, we set the unit bonus price of each task bit being transmitted from the TEC to the AEC to b. The bonus paid B_j to TECs providing the tasks can be expressed by

$$B_j(b, \mathbf{m_{i,j}}) = \sum_{i=1}^{I} b m_{i,j}. \tag{15}$$

Therefore, we have

$$U_j(b, \mathbf{m_{i,j}}) = \gamma \sum_{i=1}^{I} E_j^C - \sum_{i=1}^{I} b m_{i,j} - \gamma \kappa f_{j0}^3 (T_j^0 - T_j^1). \tag{16}$$

Mathematically, the AEC's game problem can be presented as:

Problem 1:

$$\max_{b} \quad \gamma \sum_{i=1}^{I} E_j^C - \sum_{i=1}^{I} b m_{i,j} - \gamma \kappa f_{j0}^3 (T_j^0 - T_j^1) \tag{17a}$$

$$s.t. \quad f_{j0} + \sum_{i=1}^{I} f_{j,i} \le f_j \tag{17b}$$

$$b > 0 \tag{17c}$$

$$E_j > e_j \tag{17d}$$

$$\text{if } m_{i,j} = 0, f_{j,i} = 0 \tag{17e}$$

where constraint (17b) ensures the CPU-cycle used for semantic processing is less than the total AEC computational capacity. Furthermore, constraint (17c) guarantees the semantic task unit price is greater than 0 and constraint (17d) is intended to

ensure the AEC has sufficient energy to process the semantic task. Constraint (17e) shows the relationship between $m_{i,j}$ and $f_{j,i}$.

B. Game at TECs

Similarly, based on energy variation, we can define the utility of a TEC i as:

$$U_i = B_j - N_i - C_i^t - S_i. \tag{18}$$

where B_j is the bonus gain of TEC i from the AEC j and N_i is the net income of processing semantic tasks. These parameters are the same as Eq. (9). Moreover, C_i^t is the transmission cost from the TEC i to the AEC j. In particular, S_i is the potential decrease in subscriber satisfaction due to the change in the location of the semantic transmission service. First, based on Eq. (15), we have B_j as:

$$B_j(\mathbf{b}, m_{i,j}) = b m_{i,j}. \tag{19}$$

The N_i from Eq. (18) is the net income forgone of TEC i to transmit semantic compression tasks to subscribers. The net income is transferred to the AEC. Therefore, similar to Eq. (10), we have the net income forgone of TEC i as:

$$N_i(m_{i,j}) = \alpha E_i^C. \tag{20}$$

In addition, C_i^t is the transmission energy revenue loss from the TEC i to the AEC. As no economic benefit is generated from this energy, we denoted the C_i^t by

$$C_i^t(m_{i,j}) = \gamma E_i^T. \tag{21}$$

In Eq. (17) S_i is set as the satisfaction revenue change of TEC i due to the semantic transmission tasks transfer from the TEC to the AEC. The lower satisfaction results in a lower motivation for subscribers to access the edge services, resulting in lower gains. In this paper, we argue that subscriber satisfaction is related to task processing delay. We hence model the satisfaction revenue as a logarithmic function related to execution delay. Because the logarithmic function based on execution delay precisely expresses the satisfaction of subscribers with the edge services [32], [33]. The S_i can be denoted by

$$S_i(m_{i,j}) = \varphi(\ln(1 + \theta - T_i^C) - \ln(1 + \theta - T_j^C - T_i^T)), \tag{22}$$

where $\varphi \geq 0$ is the monetary parameter and $\theta \geq T_j^C + T_i^T$ to ensure the satisfaction is positive. Therefore, we have

$$U_i(\mathbf{b}, m_{i,j}) = b m_{i,j} - \alpha \kappa a m_{i,j} f_i^2 - \gamma E_i^T \\ - \varphi(\ln(1 + \theta - T_i^C) - \ln(1 + \theta - T_j^C - T_i^T)). \tag{23}$$

In addition, we also need to consider the privacy leakage of TECs. Because even though the AEC is trusted, setting a TEC privacy breach tolerance threshold ζ is necessary to prevent possible attacks. According to [34], we have the relationship between transfer tasks bits and privacy leakage value p_i^r as:

$$p_i^r = \log_2(1 + e^{1 - \frac{m_i + 1}{m_{i,j}}}). \tag{24}$$

Algorithm 1 EGTIM

1: Initialization: semantic transmission tasks m_{i_j}, CPU-cycle frequency f_i, the maximum number of iteration K, the stopping criterion threshold $\xi > 0$, and learning rate ς
2: **for** each $i = 1, 2, ..., I$
3:   Derive optimal $m_{i,j}^*$, i.e., $f_i(b)$ by $\frac{\partial U_i}{\partial m_{i,j}} = 0$
4: **end for**
5: Substitute $f_i(b)$ in $U_j(b)$
6: **while** $k < K$
7:   $b' = b - \varsigma \bigtriangledown U_j(b)$
8:   $b'' = b, b = b'$
9:   until $b'' - b < \xi$
10: **end while**
11: Derive optimal $m_{i,j}$ according to optimal b
12: **return** b and $m_{i,j}$

---

where m_i is the number of training data that TEC i have. Hence, we can have the upper bound of transfer tasks bits $m_{i,j}^{max}$ via $p_i^r < \zeta$.

The TECs' game problem can be expressed as:

Problem 2:

$$\max_{m_{i,j}} \quad b m_{i,j} - \alpha \kappa a m_{i,j} f_i^2 - \gamma \frac{p_i m_{i,j}}{B_i \log_2(1 + \frac{p_i g_i}{\sigma^2})} \\ - \varphi(\ln(1 + \theta - T_i^C) - \ln(1 + \theta - T_j^C - T_i^T)) \tag{25a}$$
$$s.t. \quad 0 \leq m_{i,j} \leq m_{i,j}^{max} \tag{25b}$$

C. Nash equilibrium for the game

The game of TECs and the AEC can model as a Stackelberg game. To guarantee fairness, the objective of the TECs is to maximise their utility by simultaneously selecting the most appropriate $m_{i,j}$ when given the known unit price b. Meanwhile, the AEC's objective is to maximise its utility by varying b, for a known $m_{i,j}$. The game can be expressed by

$$U_i(\mathbf{b}^*, m_{i,j}^*) \geq U_i(\mathbf{b}^*, m_{i,j}), \tag{26}$$

$$U_j(b^*, \mathbf{m_{i,j}^*}) \geq U_j(b, \mathbf{m_{i,j}^*}), \tag{27}$$

where b^* and $m_{i,j}^*$ are solutions in which the parties jointly pursue the optimal strategies, i.e., the Nash equilibrium (NE) point(s). We demonstrate the existence of NE in this game.

Existence of NE:

The second-order partial derivative of $U_i(\mathbf{b}^*, m_{i,j})$ can be denoted by

$$\frac{\partial^2 U_i}{\partial m_{i,j}^2} = \varphi((\frac{\frac{a}{f_i}}{\theta - T_i^C + 1})^2 - (\frac{\frac{a}{f_i} + \frac{1}{r_i^T}}{\theta - T_i^C - T_i^T + 1})^2). \tag{28}$$

Since $\theta - T_i^C + 1 > \theta - T_i^C - T_i^T + 1$ and $\frac{a}{f_i} < \frac{a}{f_i} + \frac{1}{r_i^T}$, we can observe that $\frac{\partial^2 U_i}{\partial m_{i,j}^2} < 0$. Hence, U_i is concave in $m_{i,j}$. As the strategy set of the TEC i is also compact and convex,

based on the Debreu-Glicksberg-Fan theorem [30], the NE of this game exists.

In order to achieve NE, we utilize the backward induction approach in game theory and obtain the optimal strategies of followers (TECs) first. Subsequently, based on these TECs' strategies, the leader's (AEC's) optimal strategy is developed. Thus, we first derive the first-order partial derivative of U_i as:

$$
\frac{\partial U_i}{\partial m_{i,j}} = b - \alpha \kappa a f_i^2 - \gamma \frac{p_i}{r_i^T}
$$
$$
- \frac{\varphi f_i^2(\theta + 1)}{(f_i - am_{i,j} + \theta f_i)(r_i^T f_i - f_i m_{i,j} - r_i^T am_{i,j} + r_i^T \theta f_i)}.
$$
(29)

As U_i is concave in $m_{i,j}$, the maximum of U_i and corresponding $m_{i,j}$ thus can be derived by $\frac{\partial U_i}{\partial m_{i,j}} = 0$. Due to it being hard to express, we simply denoted the optimal $m_{i,j}^* = f_i(b)$. Therefore, the utility function of U_j can be rewritten as:

$$
U_j(b) = \gamma \sum_{i=1}^{I} \kappa a f_i(b) f_{j,i}^2 - \sum_{i=1}^{I} b f_i(b)
$$
$$
- \gamma \kappa f_{j0}^3 (T_j^0 - T_j^1).
$$
(30)

If we can derive the maximum U_j and corresponding b, we therefore can obtain the corresponding $m_{i,j}^*$ in a closed-form based on Eq. (29). However, due to the complexity of the Eq. (30), we cannot derive the NE closed form. Fortunately, b and $m_{i,j}$ both have boundaries. The NE thus can be obtained by performing a gradient descent method [35] over b and $m_{i,j}$. The solution step is shown in Algorithm 1.

IV. EFFICIENT DISTRIBUTED LEARNING DESIGN

The application of semantic communication significantly improves the network QoS. Nevertheless, how to update users' ML-based semantic coders efficiently and accurately in real-time becomes one of the biggest challenges of semantic communication studies. FL is a potential approach to cope with the challenge of semantic coder updates in the network [23]. Nevertheless, the 3-D network environment is sophisticated, and energy limited. In particular, the case where the users' training data are non-IID significantly reduces the semantic communication QoS. To address these challenges, we propose a GEDL framework for AENs (Fig. 2). Specifically, TECs first transmit some semantic communication transmission tasks to the AEC based on our proposed renewed EGTIM for semantic coder updating. The TECs then update the semantic coder based on their training data and transmit the new coder model to the AEC for the federated aggregation. Subsequently, the AEC performs the federated aggregation and retrains the aggregated model utilizing the tasks provided by TECs. This is because AEC is flexible in terms of data collection, it is often used as a federated aggregation node [36]. Finally, the AEC sends back the model to participated TECs and completes one training epoch. The model accuracy thus can be improved while maximising energy efficiency. We will demonstrate these in our simulations.
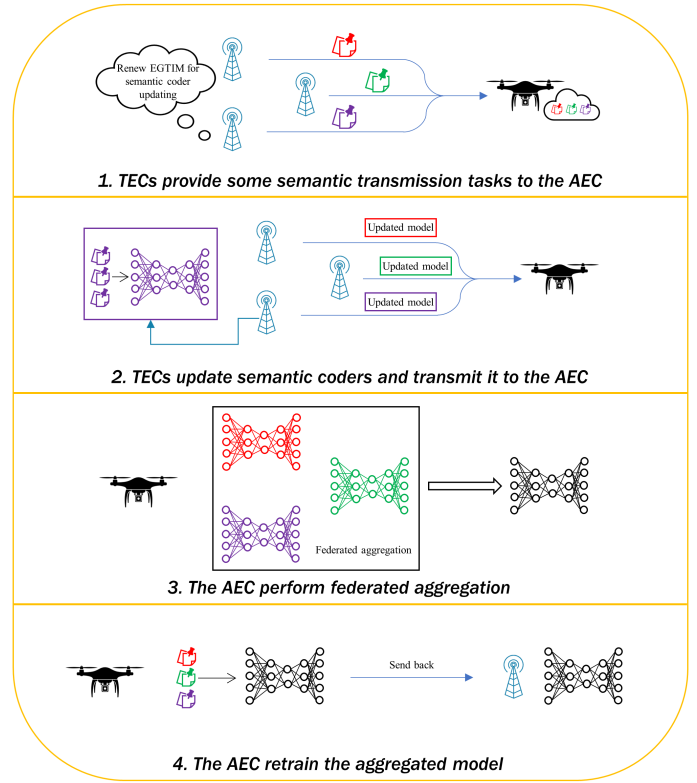

Fig. 2: The process of proposed GEDL.

We first renew the EGTIM for semantic coder updating. As increased semantic coder accuracy can improve the network QoS, it enhances network revenue. Similar to [37], we utilize a logarithmic function to model the relationship between training accuracy and training task size. The revenue of model accuracy improvement thus can be denoted by

$$
A_j^t = \delta(\ln(1 + \sum_{i=1}^{I} m_{i,j}^t) + \eta),
$$
(31)

where $m_{i,j}^t$ is the proving task bits from the TEC i to the AEC j. The difference between $m_{i,j}$ and $m_{i,j}^t$ is that $m_{i,j}$ is the providing tasks during trained semantic coder transmission and $m_{i,j}^t$ is the providing tasks during semantic coder training. Further, δ is the monetary parameter and η is the basic accuracy of FL.

Therefore, we should update the utility function of the AEC j as:

$$
U_j^t = A_j^t + R_j^t - B_j^t - G_j^t.
$$
(32)

Similar to Eq. (9), in Eq. (32), $R_j^t = \beta \kappa a f_j^{t^2} \sum_{i=1}^{I} m_{i,j}^t$ is the energy cost revenue of AEC j gained for additional training $m_{i,j}^t$ data and B_j^t is the bonus paid from the AEC j to TECs providing the tasks. Further, G_j is the gain loss of the AEC j due to the transfer of some holdup energy to additional training.

Therefore, the game problem for AEC j when coder training can be presented as:

Problem 3:

$$\max_b \quad \delta(\ln(1 + \sum_{i=1}^{I} m_{i,j}^t) + \eta) + \beta\kappa a f_j^{t^2} \sum_{i=1}^{I} m_{i,j}^t$$

$$-\sum_{i=1}^{I} b m_{i,j}^t - \gamma\kappa f_{j0}^3 (T_j^0 - T_j^1) \tag{33a}$$

$$s.t. \quad f_{j0} + f_j^t \le f_j \tag{33b}$$

$$b > 0 \tag{33c}$$

$$E_j \ge e_j \tag{33d}$$

$$\text{if } m_{i,j}^t = 0, f_j^t = 0 \tag{33e}$$

where f_j^t is the CPU-cycle frequency of the AEC j to perform the additional training after federated aggregation. Due to the requirement to perform federated aggregation, the power of AEC j for the regular task without economic benefit also needs to be plus the aggregation power. Furthermore, the reduction in training sample size reduces the model accuracy and thus affects the accuracy of the model after federated aggregation [38]. Therefore, TECs still train the number of new tasks they have.

The utility function of proving semantic transmission tasks thus can be changed from Eq. (18) by

$$U_i^t = B_i^t - C_i^{tra} - S_i^t, \tag{34}$$

where B_i^t is the training bonus gain of TEC i from the AEC j and C_i^{tra} is the transmission energy consumption. Further, S_i^t is the revenue change due to the satisfaction change. As satisfaction is associated with training time, we have

$$S_i^t = \varphi(\ln(1 + \theta^t - T_i^t) - \ln(1 + \theta^t - T_j^t - T_i^a)), \tag{35}$$

where T_i^t is the distributed learning training computing time without AEC additional training, i.e., FL training computing time. Further, T_i^a is the AEC additional training time. Since the training time tends to be much greater than the training data transmission time, we ignore the variation in satisfaction due to the transmission time. Hence, we have the game problem for the TEC i during training new coders as:

Problem 4:

$$\max_{m_{i,j}^t} \quad b m_{i,j}^t - \gamma \frac{p_i m_{i,j}^t}{B_i \log_2(1 + \frac{p_i g_i}{\sigma^2})} - \varphi(\ln(1 + \theta^t - T_i^t)$$

$$- \ln(1 + \theta^t - T_i^t - T_i^a)), \tag{36a}$$

$$s.t. \quad 0 \le m_{i,j}^t \le m_{i,j}^{t,max} \tag{36b}$$

where $m_{i,j}^{t,max}$ is the maximum available providing training data considering the risk of privacy leakage arising. Furthermore, m_i^t is the total training task bits of the TEC i. It can be found from Problem 4 that the strategy set of the TEC i is also compact and convex as same as Problem 2. In addition, the second differentiation of U_i^t is similar to U_i and concave in $m_{i,j}^t$. Thus, the NE of this game is still existing and the NE point can be achieved by Algorithm 1.

V. SIMULATION RESULTS

In this section, we provide simulation results to validate the performance of the proposed EGTIM and GEDL. First, we elaborate on the energy efficiency of our EGTIM. The advantage of our GED framework is then assessed by comparing it with baseline distributed learning in image transmission scenarios [22], [23].

A. EGTIM

To the best of our knowledge, there is little previous research on the study of energy-efficient semantic communication in AEN networks. Therefore, in simulations, we demonstrate the effectiveness of EGTIM compared to the straightforward employment of semantic communication in AENs. We first elaborate on the simulation settings in assessing the performance of our proposed EGTIM. We assume there are 5 TECs in the service range of the AEC j. To better demonstrate our proposed mechanism, we assume that all TECs have the same conditions. Similar to [27] and [29], we set $a = 120$; $p_i = 0.2$ w; $\kappa = 10^{(-26)}$; $f_i = 0.5 \times 10^9$ cycles/s; $f_{j0} = 0.5 \times 10^9$ cycles/s. Further, if not mentioned specifically, we assume the monetary parameter $\alpha = 1$, $\beta = 1$ and thus $\gamma = 2$. The hold-up power of the AEC is set as 1 w and by default the constraints are all satisfied.
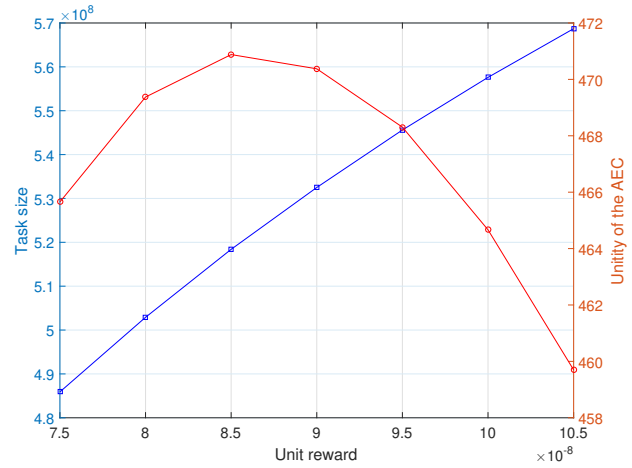

Fig. 3: NE existence under the proposed EGTIM.

In Fig, 3, the existence of NE is demonstrated. It can be observed that as the unit reward value increases, the optimal task size that TECs are willing to provide also increases. This is due to the increased transfer task size allowing TECs to earn greater benefits as the unit rewards increase. However, the utility function of the AEC shows an increasing trend followed by a decreasing trend. There is therefore an NE point that maximises the utility of the AEC while ensuring that the utilities of TECs are maximised (i.e., optimal transfer task size).

Fig. 4 illustrates the energy savings in joules (J) at different amounts of TECs and different hover consumption power. We define energy saving as the reduction in wasted hover consumption minus the lost energy consumption for regular AEC tasks and the power consumption of TECs transmitting.

Mathematically, the energy saving equals $\frac{R_j}{\beta} - \frac{G_j + C_i^t}{\gamma}$. As can be observed, more energy can be saved as the number of TECs increases. This is due to the fact that the increase in the number of TECs decreases the energy consumption in hover and outweighs the resulting loss raise. It is notable that the number of TECs does not grow indefinitely as the AEC has a finite computing capacity. In addition, the higher the hover power, the greater the energy saving, but the magnitude of the increase is decreasing. Because the hover power increase means consuming the same energy for additional semantic transmission tasks, the AEC can be maintained on air for a longer time. The corresponding cost loss thus falls and the magnitude of the increase is decreasing as the percentage of hover energy consumption of the AEC becomes larger.
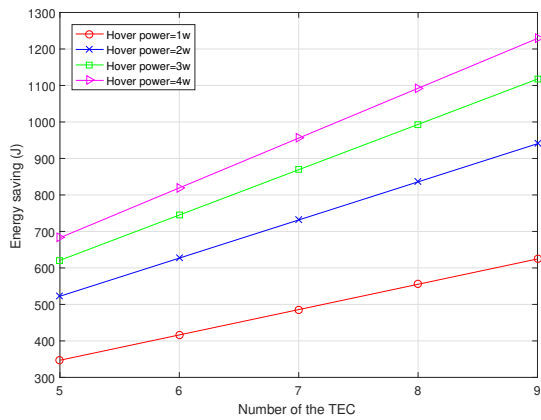

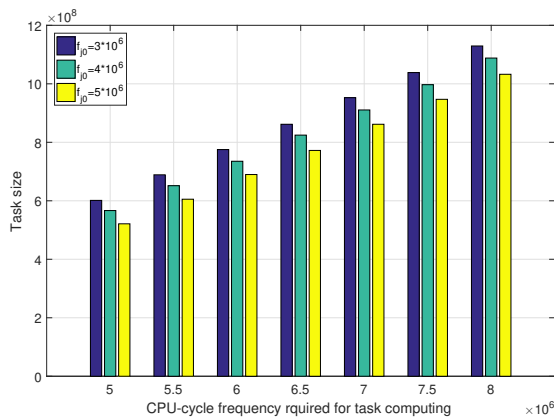Fig. 4: Energy saving of proposed EGTIM in various scenarios.

Fig. 5: Effect of different CPU-cycle on providing task size.

In Fig. 5, we evaluate the influence of different CPU-cycle on providing task size from TECs to the AEC. It is observed that more CPU-cycle frequency required for semantic task transmission makes TECs more inclined to transfer more task bits. However, the increase in CPU-cycle frequency required for regular tasks results in lower providing task sizes. This is because the increased CPU-cycle frequency required for tasks increases the efficiency of AEC hover energy utilization. Therefore, TECs are biased towards providing more tasks for

more revenue. Further, the increased f_{j0} increases the hover time reduction benefit UFT loss and therefore reduces the overall data transfer revenue and hence the unit reward.

B. GEDL

To estimate our GEDL, we employ the convolutional neural network (CNN) as the semantic coder and set the application scenario as an image transmission environment. The semantic coder setting is the same as the previous semantic communication study, i.e., [7]. Further, we train models on the CIFAR-10 [39] dataset with 60000 training data and 10000 test data, which all have 10 class images. As in the same previous subsection, we assume there are 5 TECs involved in the training. To create the non-IID training environment, we enable each TEC in training to have only four classes of the training data in the different 10000 CIFAR-10 data. The transmission accuracy is determined by the PSNR, which is a criterion for the quality of image transmission in semantic communication [7]. We have

$$PSNR = 10lg\frac{MAX^2}{\|x - \hat{x}_j\|^2}, \qquad (37)$$

where MAX is the maximum value for a pixel and x is the input of the image and $\hat{x}_j$ is the output via the semantic coder.
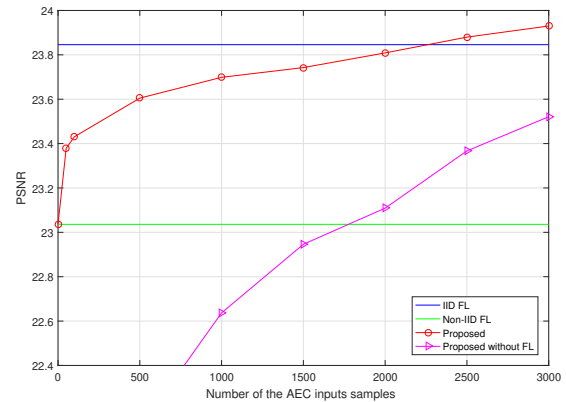

Fig. 6: The accuracy of various training frameworks with the AEC input samples grows.

Fig. 6 demonstrates the comparison of accuracy under different learning frameworks. We compare the different learning frameworks together when the training data is non-IID. Furthermore, we also add the FL model with IID training data as a reference. Since the input data of AEC remains 0 in FL-based frameworks, the PSNR of FL-based frameworks did not change as the AEC input samples grows. It is seen that as the training data obtained by the AEC increases, the coder accuracy also increases. In particular, the trend of the increase exhibits a trend of the logarithmic function, thus verifying our hypothesis in Eq. (31). In addition, with the increase in the volume of data, the accuracy of the proposed GEDL increased and even exceeded the performance of FL trained with the IID model. The accuracy of our proposed GEDL without FL also rapid growth. This is because the greater the amount of data AEC has, the more the training process approaches central

learning. The training data is mixed together for training and therefore the accuracy increases. Nevertheless, it is noteworthy that due to privacy, AEC's available computing resources and energy constraints, the data AEC obtains is limited. However, our proposed GEDL is always more accurate than FL with the non-IID training scenario.
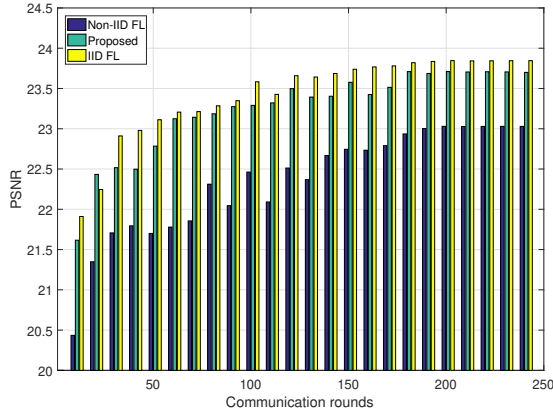

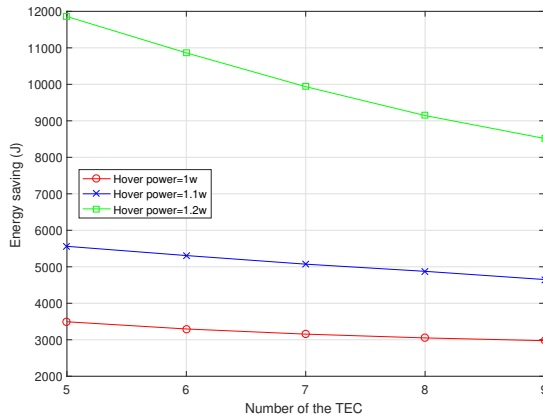Fig. 7: Convergence speed of different training frameworks.

Fig. 8: Energy saving of proposed GEDL in various scenarios.

Fig. 7 shows the comparison of the convergence speed of FL and our proposed distributed learning. We also included FL trained with the IID data as a reference. It can be observed that all learning eventually reaches convergence and the time to reach convergence is almost the same. However, our proposed GEDL is always more accurate than FL after each communication round. This is because our proposed GEDL is based on the FL for accuracy improvement and thus it increases the training accuracy but needs the FL process to reach convergence.

In Fig. 8, the energy savings in joules (J) at different amounts of TECs and different hover consumption power are shown. We set the training epoch is 200. We can see that in contrast to Fig. 4, there is a declining trend in energy savings as the number of TECs increases. This is because accuracy revenue shows a logarithmic function trend. Providing more data when there are more TECs may increase energy savings, but not the corresponding accuracy gains. As a result, the total task size provided by TECs is decreasing and thus

decreases the total energy saving. However, the GEDL we propose can always improve energy efficiency and save energy. Furthermore, the magnitude of the energy saving increase with the hover power increase varies from Fig. 4. It is likewise due to the existence of the trend in the logarithmic function of accuracy revenue. The decrease in regular task revenue due to time reduction makes the task size increase dramatically in order to reach the NE point.
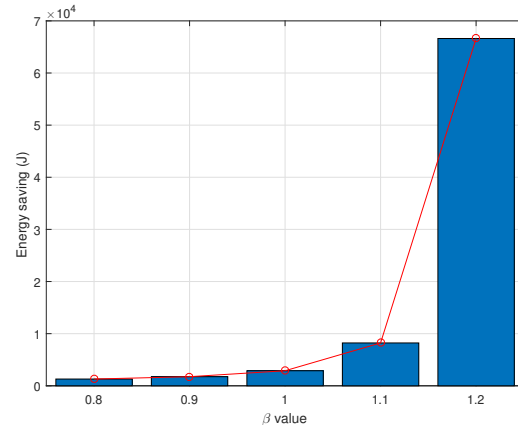

Fig. 9: The impact of β value on energy saving.

Fig. 9 illustrates the impact of changes in β value on energy saving. We evaluate this by adjusting the size of the energy cost monetary factor β. The smaller β means a higher energy cost price. We can observe that as the cost price grows, the overall energy saving of the network also rises exponentially. Due to the reduction in net income, the network members are more inclined to save energy for monetary benefits. Consequently, $m_{i,j}^t$ from the TEC i increases sharply in order to reach the NE point, thus making the energy saving increase.

VI. CONCLUSIONS

In this paper, we first proposed a novel energy-efficient semantic communication system in AENs. We then presented an EGTIM based on the Stackelberg game. In our EGTIM, the edge facilities on the terrestrial are incentivised to transfer part of their semantic transmission tasks to the AEC via the traditional communication encoder. The AEC performs the semantic feature extraction of these tasks and transmits the semantic information to the subscribers. The energy efficiency of the aerial devices thus can be improved. In addition, we further proposed a GEDL framework based on the renewed EGTIM for energy-limited 3-D networks updating semantic coders with non-IID training data. The simulation results demonstrated the effectiveness of our mechanism and learning framework.

REFERENCES

[1] H. Zhang, R. Liu, A. Kaushik and X. Gao, "Satellite Edge Computing with Collaborative Computation Offloading: An Intelligent Deep Deterministic Policy Gradient Approach," *IEEE Internet of Things Journal*, vol. 10, no. 10, pp. 9092-9107, May 2023.

[2] S. S. Shinde, A. Bozorgchenani, D. Tarchi and Q. Ni, "On the Design of Federated Learning in Latency and Energy Constrained Computation Offloading Operations in Vehicular Edge Computing Systems,"*IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2041-2057, Feb. 2022.

[3] Y. Liu, K. Xiong, Q. Ni, P. Fan, and K. B. Letaief, "UAV-assisted wireless powered cooperative mobile edge computing: Joint offloading, CPU control and trajectory optimization," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2777–2790, Apr. 2020.

[4] X. Gao, R. Liu and A. Kaushik, "Hierarchical Multi-Agent Optimization for Resource Allocation in Cloud Computing Systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 3, pp. 692-707, Mar. 2021.

[5] X. Luo, H. -H. Chen and Q. Guo, "Semantic Communications: Overview, Open Issues, and Future Research Directions," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 210-219, February 2022.

[6] Z. Yang, M. Chen, G. Li, Y. Yang and Z. Zhang, " Secure Semantic Communications: Fundamentals and Challenges", arXiv:2301.01421, 2023.

[7] D. B. Kurka and D. Gündüz, "DeepJSCC-f: Deep Joint Source-Channel Coding of Images With Feedback," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 178-193, May 2020.

[8] H. Xie, Z. Qin, G. Y. Li and B. -H. Juang, "Deep Learning Enabled Semantic Communication Systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663-2675, 2021.

[9] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Wireless semantic communications for video conferencing," Feb. 2022, [online] Available: http://arXiv:2204.07790.

[10] Z. Weng and Z. Qin, "Semantic Communication Systems for Speech Transmission," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2434-2444, Aug. 2021.

[11] W. Yang et al., "Semantic Communications for Future Internet: Fundamentals, Applications, and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 213-250, First quarter 2023.

[12] D. Gündüz et al., "Beyond Transmitting Bits: Context, Semantics, and Task-Oriented Communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5-41, Jan. 2023.

[13] W. J. Yun et al., "Attention-based Reinforcement Learning for Real-Time UAV Semantic Communication," *2021 17th International Symposium on Wireless Communication Systems (ISWCS)*, Berlin, Germany, 2021, pp. 1-6.

[14] J. Kang et al., "Personalized Saliency in Task-Oriented Semantic Communications: Image Transmission and Performance Analysis," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 186-201, Jan. 2023.

[15] Y. Xiao, X. Zhang, Y. Li, G. Shi and T. Başar, "Rate-Distortion Theory for Strategic Semantic Communication," *2022 IEEE Information Theory Workshop (ITW)*, Mumbai, India, 2022, pp. 279-284.

[16] B. Güler, A. Yener and A. Swami, "The Semantic Communication Game," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 4, pp. 787-802, Dec. 2018

[17] Z. Yang, M. Chen, Z. Zhang and C. Huang, "Energy Efficient Semantic Communication Over Wireless Networks With Rate Splitting," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1484-1495, May 2023.

[18] Y. Cang et al., "Online Resource Allocation for Semantic-Aware Edge Computing Systems," *IEEE Internet of Things Journal*.

[19] X. Yuan, H. Li, K. Ota and M. Dong, "Building Energy Efficient Semantic Segmentation in Intelligent Edge Computing," *IEEE Transactions on Green Communications and Networking*, vol. 8, no. 1, pp. 572-582, March 2024.

[20] Z. Qin, X. Tao, J. Lu, W. Tong and G. Y. Li, "Semantic communications: Principles and challenges", arXiv:2201.01389, 2021.

[21] G. Zheng, Q. Ni, K. Navaie, H. Pervaiz and C. Zarakovitis, "A Distributed Learning Architecture for Semantic Communication in Autonomous Driving Networks for Task Offloading," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 64-68, November 2023

[22] G. Shi, Y. Xiao, Y. Li and X. Xie, "From Semantic Communication to Semantic-Aware Networking: Model, Architecture, and Open Problems," *IEEE Communications Magazine*, vol. 59, no. 8, pp. 44-50, August 2021.

[23] Z. Qin, G. Y. Li and H. Ye, "Federated Learning and Wireless Communications," *IEEE Wireless Communications*, vol. 28, no. 5, pp. 134-140, October 2021.

[24] G. Zheng et al., "Mobility-Aware Split-Federated With Transfer Learning for Vehicular Semantic Communication Networks," *IEEE Internet of Things Journal*.

[25] H. Zhu, J. Xu, S. Liu, and Y. Jin, "Federated learning on non-iid data: A survey," *Neurocomputing*, vol. 465, pp. 371–390, 2021.

[26] C. You, K. Huang, H. Chae and B. -H. Kim, "energy efficiency Resource Allocation for Mobile-Edge Computation Offloading," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1397-1411, March 2017.

[27] Z. Song, Y. Hao, Y. Liu and X. Sun, "energy efficiency Multiaccess Edge Computing for Terrestrial-Satellite Internet of Things," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 14202-14218, 15 Sept.15, 2021.

[28] Q. Shi, L. Zhao, Y. Zhang, G. Zheng, F. R. Yu and H.-H. Chen, "Energy-efficiency versus delay tradeoff in wireless networks virtualization", *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 837-841, Jan. 2018.

[29] Q. Tang, Z. Fei, B. Li and Z. Han, "Computation Offloading in LEO Satellite Networks With Hybrid Cloud and Edge Computing," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 9164-9176, 1 June1, 2021.

[30] D. Fudenberg and J. Tirole, Game Theory, MIT Press, 1993.

[31] Q. Ni and C. C. Zarakovitis, "Nash Bargaining Game Theoretic Scheduling for Joint Channel and Power Allocation in Cognitive Radio Systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 1, pp. 70-81, January 2012.

[32] Y. Wang et al., "Task Offloading for Post-Disaster Rescue in Unmanned Aerial Vehicles Networks," *IEEE/ACM Transactions on Networking*, vol. 30, no. 4, pp. 1525-1539, Aug. 2022.

[33] H. Zhou, Z. Wang, G. Min and H. Zhang, "UAV-aided Computation Offloading in Mobile Edge Computing Networks: A Stackelberg Game Approach," *IEEE Internet of Things Journal*, 2022.

[34] G. Zheng, Q. Ni, K. Navaie and H. Pervaiz, "Semantic Communication in Satellite-borne Edge Cloud Network for Computation Offloading," *IEEE Journal on Selected Areas in Communications*.

[35] S. Ruder, "An overview of gradient descent optimization algorithms", CoRR, vol. abs/1609.04747, pp. 1-14, Sep. 2016.

[36] H. Yang, J. Zhao, Z. Xiong, K. -Y. Lam, S. Sun and L. Xiao, "Privacy-Preserving Federated Learning for UAV-Enabled Networks: Learning-Based Joint Scheduling and Resource Management," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 10, pp. 3144-3159, Oct. 2021.

[37] Sun, C. et al., Revisiting Unreasonable Effectiveness of Data in Deep Learning Era, arXiv:1707.02968, 2017.

[38] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin and V. Shmatikov, How to backdoor federated learning, arXiv:1807.00459, 2018.

[39] A. Krizhevsky, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep. TR-2009, 2009