

UAV-Assisted Integrated Sensing, Communication and Computation Edge Inference: Task-Driven Joint Design With DAE-VAE

Guofang Wu, Yejun He, *Senior Member, IEEE*, Xiaowen Cao, and Mohsen Guizani, *Fellow, IEEE*

Abstract—Integrated Sensing, Communication, and Computing (ISCC) has been recognized as a key enabler for realizing the intelligent connectivity vision of 6G. Conventional pipelines struggle to extract high-quality, task-relevant sensing features, degrading the overall performance of the ISCC system. To address this, we propose a UAV-assisted ISCC edge-inference architecture with intelligent feature extraction. A denoising autoencoder (DAE) is employed to achieve intelligent noise suppression during the sensing stage, while a controlled latent-space variational autoencoder (CL-VAE) with a scenario-specific loss function is utilized to perform nonlinear feature extraction and dimensionality reduction of high-dimensional sensing data, thereby learning compact semantic representations. Based on this intelligent sensing architecture, a tightly coupled joint modeling framework among sensing, communication, and computing is further established, and a system resource optimization problem is formulated to maximize the overall discriminant gain. To tackle the non-convexity of the formulated problem, a hierarchical optimization framework is developed, where the outer layer handles the fractional programming via quadratic transformation and enumeration, and the inner layer solves subproblems through second-order cone programming (SOCP) and a Karush-Kuhn-Tucker (KKT)-based water-filling algorithm for fine-grained resource allocation. Simulation results show that the proposed scheme improves inference accuracy by 0.5% to 7.75% over the strongest baseline across the tested bandwidth, energy, and multi-device configurations, reaching up to 97.5% accuracy and demonstrating its feasibility and effectiveness as a promising design paradigm for future 6G systems.

Index Terms—Integrated sensing, communication, and computation, artificial intelligence, edge inference, denoising autoencoder, variational autoencoder.

I. INTRODUCTION

Manuscript received 2 December 2025; revised 30 April 2026; revised 10 August 2026; revised 16 September 2026. This work is supported in part by National Key Research and Development Program of China under Grant 2023YFE0107900, in part by the National Natural Science Foundation of China (NSFC) under Grant U2541208 and 62501407, in part by the Key Program of Shenzhen Natural Science Foundation under Grant JCYJ20241202124219023, and in part by the Program of Shenzhen Key Laboratory Evaluation under Grant SYSPG20241211173908022.

Guofang Wu, Yejun He and Xiaowen Cao are with the State Key Laboratory of Radio Frequency Heterogeneous Integration, Sino-British Antennas and Propagation Joint Laboratory of MOST, Guangdong Engineering Research Center of Base Station Antennas and Propagation, Shenzhen Key Laboratory of Antennas and Propagation, College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: wuguo-fangwy@163.com; heyejun@ieee.org; caoxwen@szu.edu.cn).

Mohsen Guizani is with the Machine Learning Department, Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE (e-mail: mguizani@ieee.org).

TO realize the vision of Intelligent Network of Everything, future 6G networks will impose higher demands on transmission rates, connection numbers, and task latency [1]. This requires larger bandwidth, more intelligent sensing capabilities, and more efficient computing capabilities. How to effectively utilize the limited spectrum resources and how to achieve low-latency intelligent applications for massive edge devices have become key research topics. In response, integrated sensing and communication (ISAC) [2] and integrated communication and computation (ICAC) technologies have emerged, each focusing on different aspects. ISAC focuses on the integration of sensing and communication functions to enhance spectrum efficiency [3], reduce hardware costs, and achieve efficient signal processing, while ICAC focuses on task and resource scheduling based on network and computational resources, maximizing resource utilization to meet the real-time computational demands [4]. The primary enabling technologies in this domain are mobile edge computing (MEC) [5] and over-the-air computation (AirComp) [6]. It is evident that communication, sensing, and computing are tightly interconnected. From a global perspective, intelligently integrating these three capabilities, referred to as integrated sensing, communication, and computation (ISCC), will enable wireless networks to possess sensing capabilities anywhere and anytime while efficiently processing large amounts of sensing data, thus achieving seamless mapping from the physical world to the digital world [7].

With the continuous development of technology, an increasing number of intelligent devices are being deployed at the network edge. However, the limited computational power and resources of edge devices are insufficient to meet the growing demands of intelligent tasks [8]. At the same time, the coverage of future high-frequency communication base stations is expected to shrink, making it increasingly difficult to achieve ultra-high precision, reliability, and low-latency sensing, transmission, and processing. In this context, unmanned aerial vehicles (UAVs), with their unique mobility, offer a promising solution. UAVs can provide flexible, on-demand communication and computing capabilities and offer complementary wide-area sensing perspectives for ground devices. Additionally, UAVs can serve as efficient relay nodes, connecting ground edge devices with remote base stations, and help establish a seamless, highly integrated low-altitude edge intelligence infrastructure. This makes UAVs a key enabler for applications in areas such as smart cities and emergency response, with great potential [9].

Data quality control and task-relevant feature extraction are crucial within the ISCC framework. On one hand, noise and interference can significantly degrade the accuracy of downstream inference tasks. On the other hand, fixed or empirically set compression ratios often fail to adapt to the semantic granularity required for optimal task discriminability. This not only consumes communication resources with redundant data but also increases the computational burden. To address these issues, this paper introduces a denoising autoencoder (DAE) [10] to enhance raw data quality and employs a variational autoencoder (VAE) [11] with a scenario-tailored controlled latent-space loss to learn compact task-relevant semantic representations, while enabling adaptive latent-dimension selection according to system conditions. This approach reduces communication and computation overhead while maintaining inference accuracy.

In this paper, we design a task-oriented UAV-assisted ISCC edge inference framework with a two-stage semantic processing pipeline and adaptive latent-dimension control. Unlike traditional schemes, this architecture is designed for intelligent transformation from raw sensor data to task-oriented semantic features, performing resource coordination and location optimization centered around final task effectiveness. The discriminant gain serves as a bridge linking physical resources and inference performance. The main contributions are summarized as follows:

- We develop a task-oriented two-stage data processing pipeline that cascades a denoising autoencoder (DAE) with a controlled latent-space variational autoencoder (CL-VAE). The DAE suppresses sensing noise and interference before feature extraction, and its residual effect is characterized by a single measurable factor, allowing the trained denoiser to be represented analytically within the resource-allocation model. Rather than adopting a conventional VAE, the CL-VAE augments the reconstruction and KL objectives with class-center and classification constraints, which encourage compact intra-class representations and enhanced inter-class separability in the latent space. The resulting latent features are therefore compact for transmission while remaining task-discriminative, and the latent dimension is further treated as an adjustable decision variable that can be selected according to the available system resources.
- We derive a closed-form expression for the discriminant gain that jointly accounts for the quality of the processed sensing data and the physical resources of the system. This expression shows how the achieved sensing quality, reflecting both the allocated sensing power and the residual effect of the denoising stage, combines with the communication distortion, the latter shaped by the UAV's position and the allocated transmit power, to determine the separability of the features received at the edge server. Based on this expression, we formulate an optimization problem that maximizes the overall discriminant gain jointly over the sensing and communication power, the communication time, the latent dimension, the bandwidth, and the UAV's position.

- To solve the formulated mixed-integer non-convex optimization problem, we design a two-layer iterative algorithm. The outer layer handles the fractional objective and the integer latent dimension via quadratic transformation together with a problem-specific bounded enumeration, in which explicit constraint-derived upper bounds and a dynamic feasibility test sharply reduce the search set. The inner layer employs a block coordinate descent approach, solving the subproblems through closed-form updates, second-order cone programming for the coupled power-and-position block, and a Karush-Kuhn-Tucker (KKT)-based water-filling algorithm for bandwidth.
- We conduct extensive simulations comparing the proposed scheme against several baselines. The results validate the effectiveness of the individual components (DAE, CL-VAE) and the overall optimization algorithm, demonstrating the superiority of the proposed framework and method.

The remainder of this paper is organized as follows. Section II reviews the related work. Section III presents the system model for the ISCC framework. The problem formulation is provided in Section IV. Section V proposes the joint optimization algorithm. Simulation results and corresponding analyses are discussed in Section VI. Finally, Section VII concludes the paper.

II. RELATED WORK

The related work can be categorized into two parts. The first part reviews studies on ISCC systems with or without the involvement of UAVs. The second part classifies existing ISCC studies according to their research objectives.

In ISCC research without UAV participation, the typical architecture consists of edge devices and base stations equipped with edge servers. Depending on the entity that performs the sensing function, these systems can be divided into device-sensing [12]–[14] and base-station-sensing [15], [16] architectures. In device-sensing scenarios, edge devices are equipped with both communication and sensing capabilities. They sense environmental targets using the sensing function and offload computation tasks to the base station via communication links. The base station, serving as an edge server, assists in processing these tasks. In base-station-sensing scenarios, the base station itself is equipped with ISAC capabilities, enabling it to both process computation tasks and sense targets within its service region. Furthermore, [16] introduces a cloud computing layer to assist base stations in handling computational workloads. These studies primarily focus on terrestrial cellular networks, which are suitable for relatively static user environments. However, with the rapid increase in mobile users and smart devices, UAVs have been introduced as flexible aerial platforms to serve in scenarios where ground base station coverage is insufficient, such as remote areas or large-scale events. In UAV-assisted ISCC research, two main directions have emerged. The first replaces the ground ISAC device in the aforementioned architecture with UAVs [17], [18], enabling them to flexibly sense larger areas, while base stations assist UAVs in performing computation tasks.

The second approach involves employing UAVs as aerial platforms [19]–[21], primarily assisting ground-based ISAC devices in completing communication and computation tasks. This approach effectively leverages the sensing capabilities of ground devices to obtain perception data from multiple perspectives, while the mobility and flexibility of UAVs ensure that the communication and computation requirements of the ground devices are satisfied.

According to research objectives, ISCC studies can be categorized into joint optimization-oriented and task-oriented approaches. In the joint optimization category, the goal is to simultaneously optimize multiple functions such as communication, sensing, and computation. For example, [17] formulates a joint optimization of radar beam pattern design and UAV energy consumption; [22] minimizes both radar beam pattern error and computational offloading energy; [18] aims to minimize sensing data collection time and UAV energy consumption; and [12] introduces a utility function combining three functional metrics to balance system performance. In Aircomp systems, another important metric is the Aircomp error [23], [24]. These studies mainly focus on optimization targets such as beam pattern error, energy consumption, and latency, as summarized in tables of [25]. Overall, these references optimize one or more system metrics depending on the specific application scenario. In contrast, task-oriented ISCC research [26]–[29] focuses directly on the task performance itself, regarding sensing, communication, and computation as integrated processes contributing to the final task outcome. For edge inference tasks, the objective is to achieve the best inference accuracy under limited resources, which requires coupling task accuracy with sensing, communication, and computation processes. [26]–[29] introduce the concept of discriminant gain as an intermediary metric. Specifically, [26] proves the relationship between discriminant gain and inference accuracy, and subsequent studies adopt the maximization of discriminant gain as the system optimization objective. This task-oriented line of research elegantly links the three domains to achieve overall system optimality. Within this framework, data acquisition, processing, and transmission all influence final inference performance, where sensing data quality is a key factor. Clean and informative features not only improve inference accuracy but also reduce communication overhead. To the best of our knowledge, representative prior task-oriented ISCC studies [26], [27] mainly optimize sensing, communication, and computation resources under fixed-dimensional linear feature representations. The feature-extraction method and feature dimension are therefore generally determined before resource optimization, while DAE-based denoising is not commonly incorporated into the ISCC pipeline. Moreover, when VAE-type models are considered in related ISAC or edge-inference settings, they are typically used in a standard form or as standalone feature-processing modules, rather than being specifically designed to provide task-discriminative latent representations that can be analytically incorporated into subsequent resource optimization. Motivated by these limitations, we develop a two-stage DAE-CL-VAE semantic-processing framework with task-oriented latent representation learning and adaptive latent-dimension selection for UAV-assisted task-

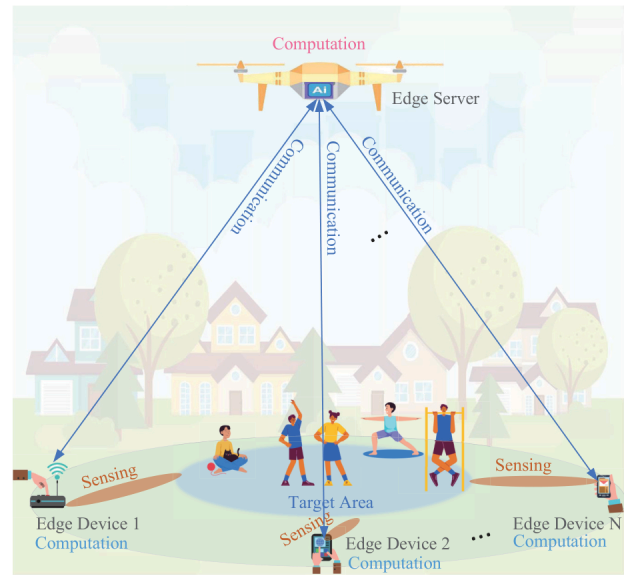


Fig. 1. UAV-assisted ISCC edge-inference network.

oriented ISCC. The denoising and feature-extraction stages are designed so that their outputs can be analytically incorporated into resource-aware optimization, while discriminant gain provides the analytical bridge connecting the resulting data-processing pipeline to the joint optimization of sensing, communication, and computation resources, together with the UAV position.

Based on the above analysis, we consider a UAV-assisted ISCC scenario, where UAVs act as aerial platforms to assist ground ISAC devices in communication and computation tasks. The ground ISAC devices cooperatively sense targets from multiple perspectives, and the sensing data are processed through a DAE and a CL-VAE for denoising and dimensionality reduction. The goal is to obtain clean and effective semantic representations that maximize discriminant gain.

III. SYSTEM MODEL

We consider an edge-inference network consisting of N single antenna ISAC devices and a UAV equipped with an edge server, as shown in Fig. 1. Each ISAC device implements sensing and communication functions in a time-division manner. The communication between the edge server and the devices employs orthogonal frequency-division multiplexing, allocating different bandwidth resources to each device to enable simultaneous transmission. The preprocessed sensing data from all devices are aggregated at the edge server and then used as the input for the edge inference task. The overall workflow proceeds as follows:

- **Task Announcement:** The edge server broadcasts a sensing-inference task to all ISAC devices.
- **Multi-View Sensing:** Each ISAC device acquires environmental observations from its own viewpoint, producing multi-view sensing data.
- **Local Preprocessing:** Onboard each device, the raw sensing measurements are sampled, denoised, and dimensionally reduced. The DAE and VAE models used

for preprocessing are trained offline on the edge server with sufficient computational resources, while only the trained models are deployed at the devices for real-time inference.

- **Data Transmission:** The dimension-reduced sensing data are uploaded from each ISAC device to the edge server via the wireless uplink.
- **Edge-Server Aggregation and Inference:** The edge server aggregates the received data streams from all devices and performs further inference on the fused multi-view data.

A. Sensing Model

When a sensing task arrives, each ISAC device uses its sensing function to transmit a sensing signal. We adopt the commonly used frequency-modulated continuous-wave (FMCW) waveform for sensing. After the sensing signal probes the target region, the transmitted FMCW signal returns as an echo containing the target's state information. Upon reception of the echo, each device performs data preprocessing to facilitate subsequent communication and inference.

1) *Sensing Signal:* The transmitted signal consists of M chirps, each of duration $T_0 = \frac{T_{s,n}}{M}$, where $T_{s,n}$ is the sensing period of device n . The sensing signal is then given by

$$s_n(t) = \sqrt{P_{s,n}} \sum_{m=0}^{M-1} \text{rect}\left(\frac{t - mT_0}{T_0}\right) \cos\left[2\pi f_n^c(t - mT_0) + \pi \frac{B_{s,n}}{T_0}(t - mT_0)^2\right], \quad (1)$$

where f_n^c is the carrier frequency of the sensing signal, $B_{s,n}$ is the bandwidth of the sensing signal, $\text{rect}(\cdot)$ denotes the rectangular window function, defined as

$$\text{rect}\left(\frac{t - mT_0}{T_0}\right) = \begin{cases} 1, & mT_0 < t < (m+1)T_0, \\ 0, & \text{otherwise.} \end{cases}$$

2) *Echo Signal:* The received echo signal at device n contains both the direct-path and the multi-path components, and is given by

$$y_n(t) = \sqrt{P_{s,n}} H_{s,n}(t) s_n(t - \tau) + \sqrt{P_{s,n}} \sum_{l=1}^L G_{n,l}(t) s_n(t - \tau_l) + w_{s,n}(t), \quad (2)$$

where $H_{s,n}(t)$ is the direct-path channel response, $G_{n,l}$ denotes the complex gain of the l -th indirect multi-path component, τ and τ_l are the propagation delays of the direct and l -th indirect paths, respectively, $w_{s,n}(t)$ denotes the additive noise in the sensing process, modeled as a zero mean Gaussian random process.

3) *Normalization:* The echo signal is normalized as

$$y_{\text{NOR},n}(t) = a_n(t) + b_n(t) + \frac{w_{s,n}(t)}{\sqrt{P_{s,n}}}, \quad (3)$$

where $a_n(t) = H_{s,n}(t) s_n(t - \tau)$ is the target signal, and $b_n(t) = \sum_{l=1}^L G_{n,l}(t) s_n(t - \tau_l)$ is the clutter from non-direct paths. We assume that $b_n(t)$ follows a zero-mean Gaussian distribution.

B. Sensing Signal Processing

1) *Intermediate Frequency Signal:* To extract target information such as range and velocity from the echo signal, the normalized signal is mixed with the original transmit signal $s_n(t)$ to obtain the intermediate frequency signal:

$$y_{\text{IF},n}(t) = y_{\text{NOR},n}(t) \cdot s_n(t) = a_{\text{IF},n}(t) + b_{\text{IF},n}(t) + \frac{w_{\text{IF},n}(t)}{\sqrt{P_{s,n}}}. \quad (4)$$

Since the transmitted signal $s_n(t)$ is known and the mixing process is linear, both $b_{\text{IF},n}(t)$ and $w_{\text{IF},n}(t)$ remain zero-mean Gaussian random processes. Specifically, the interference component $b_{\text{IF},n}(t) \sim \mathcal{N}(0, \sigma_{n,b}^2)$, and the noise term $w_{\text{IF},n}(t) \sim \mathcal{N}(0, \sigma_w^2)$.

2) *Sampling:* Assume the sampling rate is $f_{n,s}$. The number of samples per chirp is given by $O = f_{n,s}T_0$. After sampling, the signal is first reshaped from a $1 \times OM$ complex-valued vector $\mathbf{y}_n$ into a matrix $\mathbf{Y}_n \in \mathbb{C}^{O \times M}$, where $\mathbf{Y}_n = [y_n(o, m)]$, with O representing the fast-time dimension (containing range information), and M representing the slow-time dimension (containing velocity information). The sampled signal can be expressed as

$$y_{\text{SAM},n}(o, m) = a_{\text{SAM},n}(o, m) + b_{\text{SAM},n}(o, m) + \frac{w_{\text{SAM},n}(o, m)}{\sqrt{P_{s,n}}}. \quad (5)$$

3) *Sensing Data Denoising:* To suppress noise and interference, a DAE is employed. Since the sampled signal is complex-valued, it is decomposed into its real and imaginary parts before being fed into the neural network. Specifically, the complex matrix $\mathbf{Y}_n \in \mathbb{C}^{O \times M}$ is split into two real-valued matrices: $\mathbf{Y}_n^{\text{real}} \in \mathbb{R}^{O \times M}$, $\mathbf{Y}_n^{\text{imag}} \in \mathbb{R}^{O \times M}$, which are then combined as a two-channel input. The output of the DAE is the denoised signal:

$$\mathbf{Y}_n^{\text{denoised}} = f_{\text{DAE}}(\mathbf{Y}_n), \quad \mathbf{Y}_n^{\text{denoised}} \in \mathbb{C}^{O \times M}. \quad (6)$$

The denoised signal at position (o, m) is expressed as

$$y_n^{\text{denoised}}(o, m) = a_n^{\text{denoised}}(o, m) + \sqrt{\alpha_n} \left(b_n^{\text{denoised}}(o, m) + \frac{w_n^{\text{denoised}}(o, m)}{\sqrt{P_{s,n}}} \right), \quad (7)$$

where the residual interference and noise after denoising are assumed to have a reduced variance of $\alpha_n \left(\sigma_{n,b}^2 + \frac{\sigma_w^2}{P_{s,n}} \right)$, $\alpha_n \in [0, 1]$, with $\alpha_n = 1$ indicating no denoising, and $\alpha_n = 0$ representing complete suppression of interference and noise. The mean of the target signal component is preserved after denoising. After training the DAE for each device, α_n is estimated offline on a held-out validation set using paired noisy/clean samples for that device. It quantifies the average fraction of residual interference-and-noise energy remaining after denoising under the nominal dataset operating condition. The estimator is defined as

$$\hat{\alpha}_n = \frac{\sum_s \|\mathbf{Y}_{n,s}^{\text{denoised}} - \mathbf{Y}_{n,s}^{\text{clean}}\|_F^2}{\sum_s \|\mathbf{Y}_{n,s}^{\text{noisy}} - \mathbf{Y}_{n,s}^{\text{clean}}\|_F^2},$$

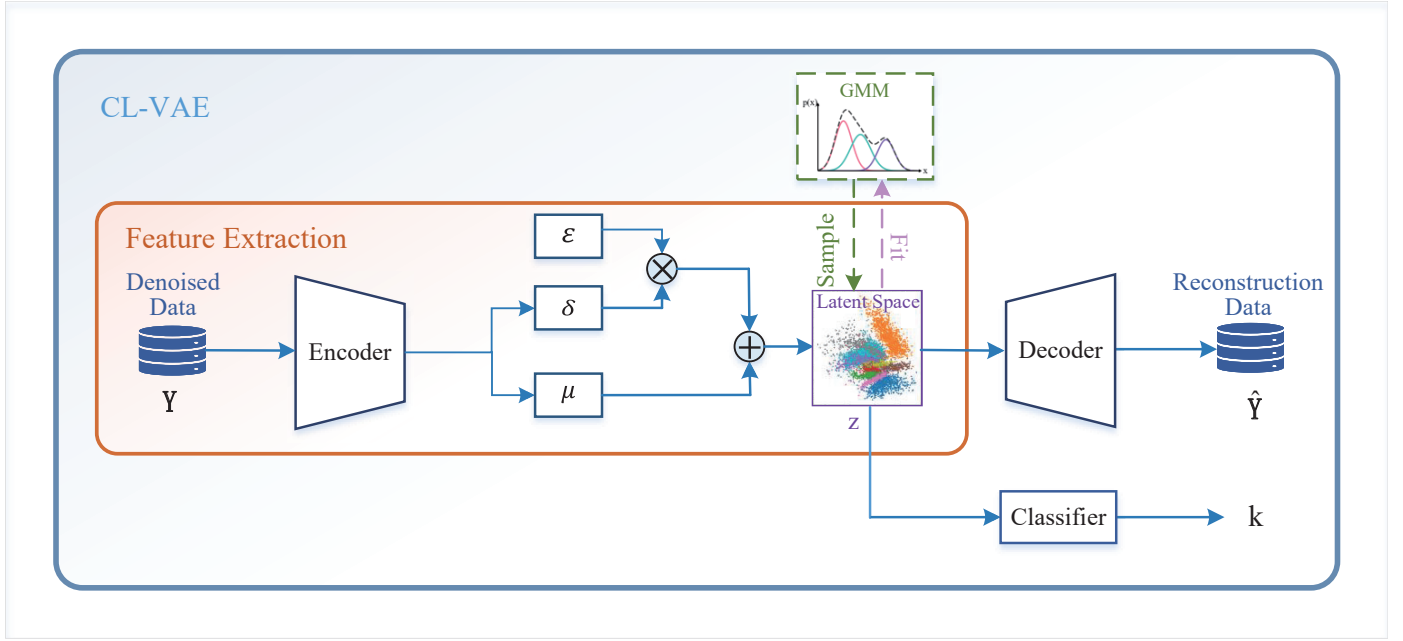


Fig. 2. Proposed CL-VAE architecture with data flow and latent space design.

where $\mathbf{Y}_{n,s}^{\text{noisy}}$ is the noisy input sample, $\mathbf{Y}_{n,s}^{\text{denoised}}$ is the denoised output, and $\mathbf{Y}_{n,s}^{\text{clean}}$ is the paired clean reference. This quantity characterizes the fraction of interference-and-noise energy that remains after denoising under the nominal operating conditions.

4) *VAE for Dimensionality Reduction*: To reduce the communication overhead, it is advantageous to perform dimensionality reduction of data at the edge devices. VAE can better capture the nonlinear characteristics of the raw data. Moreover, by modeling a probabilistic distribution in the latent space, a VAE facilitates quantification and evaluation of the uncertainty of the reduced-dimension data during subsequent inference tasks.

a) *Controlled Latent-space VAE Model*: We extend the standard VAE by introducing classification and clustering losses, and by fitting a Gaussian mixture distribution in the latent space [30], [31]. This ensures that the latent space not only preserves reconstruction quality and prior consistency, but also provides class-discriminative features. We denote the improved model as the *Controlled Latent-space VAE* (CL-VAE). The structure of CL-VAE is illustrated in Figure 2.

The total loss is defined as

$$L_{\text{loss}} = \frac{1}{S_{\text{total}}} \sum_{n=1}^N \sum_{i=1}^{S_n} \left(\lambda_1 L_{\text{recon}}^{(n,i)} + \lambda_2 L_{\text{KL}}^{(n,i)} + \lambda_3 L_{\text{center}}^{(n,i)} + \lambda_4 L_{\text{class}}^{(n,i)} \right) \quad (8)$$

where

$$\begin{aligned} L_{\text{recon}}^{(n,i)} &= -\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{Y}_{n,i}^{\text{denoised}})} \left[\log p_{\theta}(\mathbf{Y}_{n,i}^{\text{denoised}} | \mathbf{z}) \right], \\ L_{\text{KL}}^{(n,i)} &= D_{\text{KL}} \left(q_{\phi}(\mathbf{z} | \mathbf{Y}_{n,i}^{\text{denoised}}) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I}) \right), \\ L_{\text{center}}^{(n,i)} &= \left\| \mu_{n,i} - \mathbf{c}_{y_{n,i}} \right\|_2^2, \\ L_{\text{class}}^{(n,i)} &= -\sum_{k=1}^K y_{n,i,k} \log p_{n,i}^{(k)}, \end{aligned}$$

where N is the total number of devices; S_n is the number of training samples collected by device n ; $S_{\text{total}} = \sum_{n=1}^N S_n$ is the total number of training samples. $\mathbf{Y}_{n,i}^{\text{denoised}} \in \mathbb{C}^{O \times M}$ denotes the DAE-filtered spectrogram of the i -th sample from device n . $q_{\phi}(\mathbf{z} | \cdot)$ is the encoder posterior with mean vector $\mu_{n,i}$ and diagonal standard deviation $\sigma_{n,i}$; $p_{\theta}(\cdot | \mathbf{z})$ is the decoder likelihood, assumed complex Gaussian; $\mathcal{N}(\mathbf{0}, \mathbf{I})$ is the standard normal prior. $\mathbf{c}_{y_{n,i}}$ is the centre of class $y_{n,i}$; $y_{n,i,k}$ is the one-hot indicator for class k ; $p_{n,i}^{(k)}$ is the predicted probability of class k . Finally, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are the weighting coefficients for the four loss terms.

b) *Encoder and Decoder*: The denoised complex-valued matrix $\mathbf{Y}_n^{\text{denoised}} \in \mathbb{C}^{O \times M}$ is split into its real part $\mathbf{Y}_n^{\text{dereal}} \in \mathbb{R}^{O \times M}$ and imaginary part $\mathbf{Y}_n^{\text{deimag}} \in \mathbb{R}^{O \times M}$, then combined into a two-channel real tensor:

$$[\mathbf{Y}_n^{\text{dereal}}, \mathbf{Y}_n^{\text{deimag}}] \in \mathbb{R}^{O \times M \times 2}.$$

The VAE encoder maps the denoised complex signal into a latent variable with a specified distribution:

$$\mathbf{z}_n = f_{\text{VAE}}(\mathbf{Y}_n^{\text{denoised}}), \quad \mathbf{z}_n \in \mathbb{R}^D.$$

The decoder reconstructs the real and imaginary parts, which are then recombined into the complex signal $\hat{\mathbf{Y}}_n \in \mathbb{C}^{O \times M}$.

c) *Statistical Analysis of Latent Variables*: For the d -th element of the feature vector of device n , we decompose it as

$$z_n(d) = a_n^{\text{target}}(d) + z_n^{\text{residual}}(d), \quad (9)$$

where $a_n^{\text{target}}(d)$ is the true target signal component, and $z_n^{\text{residual}}(d)$ is the residual component.

We assume the following probability density functions:

$$f(a_n^{\text{target}}(d)) = \frac{1}{K} \sum_{k=1}^K \mathcal{N}(\mu_{n,d,k}, \sigma_{n,d}^2), \quad (10)$$

$$f(z_n^{\text{residual}}(d)) = \mathcal{N}\left(0, s_{n,d}^2 \alpha_n (\sigma_{n,b}^2 + \frac{\sigma_w^2}{P_{s,n}})\right), \quad (11)$$

where

- K is the total number of classes.
- $\mu_{n,d,k}$ is the mean of class k on latent dimension d at device n .
- $s_{n,d}$ is the singular value corresponding to dimension d of the encoder's Jacobian matrix.

Remark 1: The class-conditional Gaussian mixture assumption on $a_n^{\text{target}}(d)$ and the equal-variance structure across classes, i.e., the shared variance $\sigma_{n,d}^2$ for different k , are adopted to obtain a tractable closed-form discriminant-gain expression in Section III-D. These assumptions are consistent with the design objective of the proposed CL-VAE. Specifically, the KL-divergence loss regularizes the latent distribution, while the centre loss L_{center} and the classification loss L_{class} encourage intra-class compactness and inter-class separability, thereby promoting an approximately class-conditional Gaussian mixture structure in the latent space. Empirical analysis of the trained CL-VAE shows that the class-conditional latent distributions closely follow Gaussian profiles. The coefficient of variation of the per-class variance traces is 0.164, indicating that the equal-variance assumption provides a reasonable approximation for the learned latent space. Moreover, the mean Spearman rank correlation coefficient between the homoscedastic and heteroscedastic discriminant-gain rankings is 0.991 across all devices. This indicates that, under the observed level of variance heterogeneity, the dimension-selection ordering is robust and is not materially affected by the equal-variance approximation.

d) Computation Energy for Dimensionality Reduction:

After the CL-VAE encoder is trained once with the maximum latent dimension D_{max} , it is deployed on the edge device to perform dimensionality reduction. The latent dimensions are indexed in decreasing order of their discriminant-gain contribution, established once from the trained encoder. At runtime, when the system optimization selects a target dimension D , only the first D latent outputs of the last layer are activated. Therefore, the preceding encoder layers remain unchanged, and only the computation associated with the final latent projection stage scales with D . The computation time is given by

$$T_n^{\text{comp}} = T_0 + \frac{\rho_0 N_{\text{MAC}}}{f_n}, \quad (12)$$

where T_0 is the fixed time of the preceding layers, ρ_0 (cycles/MAC) is the number of CPU cycles required to complete a multiply-accumulate (MAC) operation, $N_{\text{MAC}} = H_2 \cdot 2D$ is the number of MAC operations in the last layer, H_2 is the output dimension of the penultimate layer, and f_n (cycles/s) is the CPU clock frequency.

Given the average computation power consumption P_n^{comp} of the device, the computation energy can be expressed as

$$E_n^{\text{comp}} = P_n^{\text{comp}} T_n^{\text{comp}}. \quad (13)$$

This indicates that both the computation time and the computation energy vary linearly with the latent dimension D .

C. UAV-Assisted Edge Inference Model

The UAV provides computing resources for ground edge devices to perform edge inference tasks. Let the UAV fly at a fixed altitude h , with initial position $\mathbf{p}_u^0 = (x_u^0, y_u^0, h)$, and optimal hovering position $\mathbf{p}_u = (x_u, y_u, h)$. After flying to the destination, the UAV executes the edge inference task. Denote the UAV flying speed as v . The flying distance is given by

$$\Delta p = \|\mathbf{p}_u - \mathbf{p}_u^0\|_2 = \sqrt{(x_u - x_u^0)^2 + (y_u - y_u^0)^2}. \quad (14)$$

The flying time is

$$T_{\text{move}} = \frac{\Delta p}{v}. \quad (15)$$

The energy consumption during flying is given by

$$E_u^{\text{fly}} = \left(\eta_1 v^3 + \frac{\eta_2}{v}\right) T_{\text{move}}, \quad (16)$$

where η_1 and η_2 are constants related to the UAV's weight and wing area.

The UAV hovers for a duration of T_{hover} , and the hovering energy consumption is

$$E_u^{\text{hover}} = P_u^{\text{hover}} T_{\text{hover}}, \quad (17)$$

where P_u^{hover} is the hovering power of the UAV. Ground edge devices are assumed to be stationary during each time slot t , and the position of the n -th device is $\mathbf{p}_n = [x_n, y_n, 0]$.

1) Communication Model: Similar to [20], [32], [33], the link between the UAV and the ground devices is assumed to follow a line-of-sight (LoS) channel model, as UAVs typically operate at altitudes where the probability of LoS conditions is significantly higher than in terrestrial networks. The channel power gain therefore depends only on their Euclidean distance. We define the 3-D distance

$$p_{u,n} = \|\mathbf{p}_u - \mathbf{p}_n\|_2 = \sqrt{(x_u - x_n)^2 + (y_u - y_n)^2 + h^2}. \quad (18)$$

With the free-space path-loss exponent $\beta = 2$, the corresponding channel power gain is

$$H_{u,n} = \frac{H_0}{p_{u,n}^2} = \frac{H_0}{(x_u - x_n)^2 + (y_u - y_n)^2 + h^2}, \quad (19)$$

where H_0 is the channel power gain at the reference distance of 1 m.

After dimensionality reduction (to a common size D), device n enters the communication phase and uploads its latent vector $\mathbf{z}_n = [z_n(1), \dots, z_n(D)]$. The d -th element received at the UAV is

$$z'_n(d) = \sqrt{P_{c,n} H_{u,n}} z_n(d) + c_n(d), \quad (20)$$

where $P_{c,n}$ is the transmit power of device n , $H_{u,n}$ is the channel gain in (19), and $c_n(d) \sim \mathcal{CN}(0, \delta_{n,c}^2)$ is complex Gaussian noise.

The achievable uplink rate is

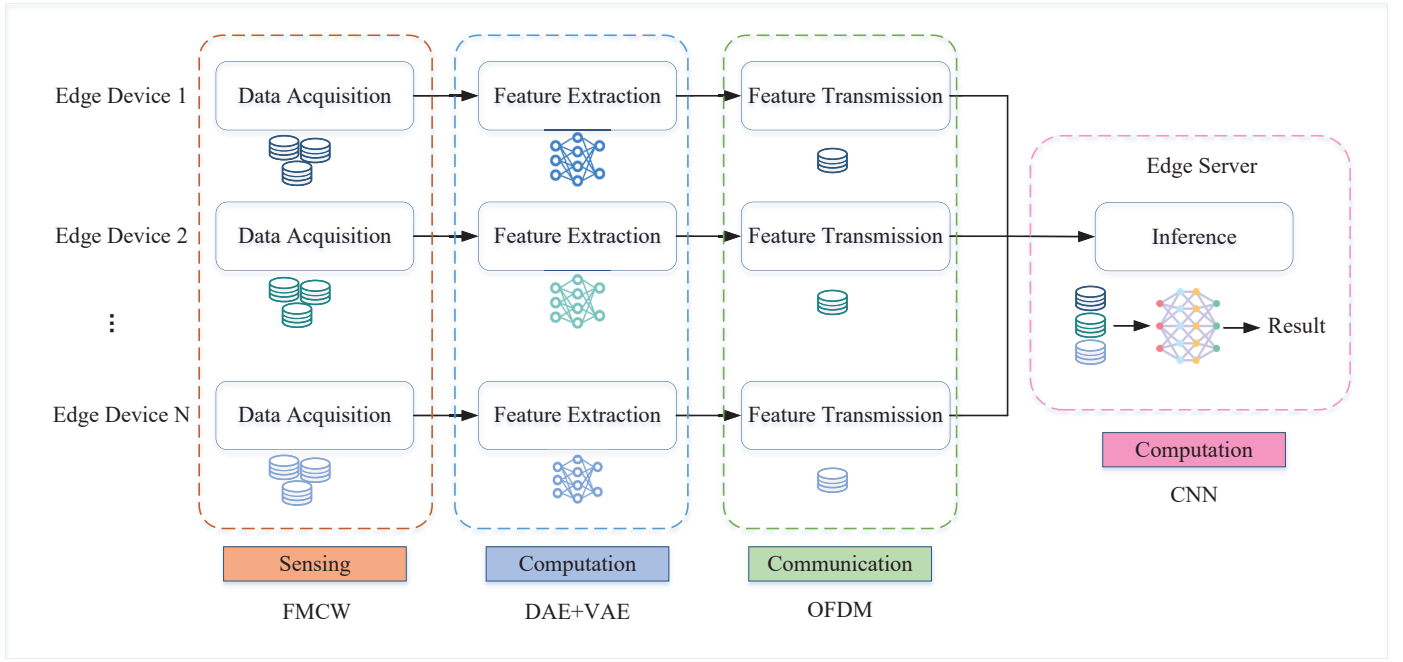


Fig. 3. The data processing workflow of ISCC.

$$R_{u,n} = B_{c,n} \log_2 \left(1 + \frac{P_{c,n} H_{u,n}}{\delta_{n,c}^2} \right), \quad (21)$$

with bandwidth $B_{c,n}$.

The received signal is normalized as follows:

$$\tilde{z}_n(d) = \frac{z'_n(d)}{\sqrt{P_{c,n} H_{u,n}}} = z_n(d) + \frac{c_n(d)}{\sqrt{P_{c,n} H_{u,n}}}. \quad (22)$$

Substituting the decomposition $z_n(d) = a_n^{\text{target}}(d) + z_n^{\text{residual}}(d)$ yields

$$\tilde{z}_n(d) = a_n^{\text{target}}(d) + z_n^{\text{residual}}(d) + \frac{c_n(d)}{\sqrt{P_{c,n} H_{u,n}}}. \quad (23)$$

The probability density function (PDF) of $\tilde{z}_n(d)$ is given by

$$f_k(\tilde{z}_n(d)) = \mathcal{N} \left(\mu_{n,d,k}, \sigma_{n,d}^2 + s_{n,d}^2 \alpha_n \left(\sigma_{n,b}^2 + \frac{\sigma_w^2}{P_{s,n}} \right) + \frac{\delta_{n,c}^2}{H_{u,n} P_{c,n}} \right). \quad (24)$$

2) *Inference Model*: The volume of data to be transmitted after dimensionality reduction is $Q = \rho D$, where ρ is the number of bits per dimension. The UAV aggregates the data from all devices to obtain multi-view sensing information. The total amount of data received by the UAV is $Q_u = NQ$. The inference time at the UAV is calculated as

$$T_u = \frac{Q_u f_0}{f_u}, \quad (25)$$

where f_0 is the number of CPU cycles required to process one bit of data, and f_u is the UAV's computing capability.

The energy consumption for inference is given by

$$E_u^{\text{inf}} = \kappa Q_u f_0 f_u^2, \quad (26)$$

where κ is a hardware-dependent constant of the UAV.

D. Discriminant Gain

To evaluate the inference performance of the sensing task at the edge server, we adopt the discriminant gain as the evaluation metric. Intuitively, the discriminant gain measures the separability between class-conditional feature distributions. A larger value indicates more distinguishable features and, as motivated by a geometric interpretation of class separability and confirmed empirically in [26] (see Fig. 5 therein), a lower inference error. Maximizing the discriminant gain therefore provides an analytically tractable surrogate for maximizing the downstream inference accuracy, which is precisely why it is adopted as the objective of our optimization. For any two classes k and k' , the discriminant gain of the normalized feature element $\tilde{z}(d_n)$ is defined as

$$\begin{aligned} G_{k,k'}(\tilde{z}_n(d)) &= D_{\text{KL}}(f_k(\tilde{z}_n(d)) \| f_{k'}(\tilde{z}_n(d))) \\ &\quad + D_{\text{KL}}(f_{k'}(\tilde{z}_n(d)) \| f_k(\tilde{z}_n(d))) \\ &= \frac{(\mu_{n,d,k} - \mu_{n,d,k'})^2}{\sigma_{n,d}^2 + s_{n,d}^2 \alpha_n \left(\sigma_{n,b}^2 + \frac{\sigma_w^2}{P_{s,n}} \right) + \frac{\delta_{n,c}^2}{H_{u,n} P_{c,n}}}, \forall (k, k'). \end{aligned} \quad (27)$$

Summing over all devices $n = 1, \dots, N$ and feature dimensions $d = 1, \dots, D$ gives the per-class-pair discriminant gain

$$G_{k,k'}(\tilde{\mathbf{z}}) = \sum_{n=1}^N \sum_{d=1}^D G_{k,k'}(\tilde{z}_n(d)). \quad (28)$$

Finally, the global discriminant gain is defined as the average over all $\frac{K(K-1)}{2}$ unordered class pairs:

$$G = \frac{2}{K(K-1)} \sum_{n=1}^N \sum_{d=1}^D \sum_{k=1}^{K-1} \sum_{k'=k+1}^K G_{k,k'}(\tilde{z}_n(d)). \quad (29)$$

The entire data processing workflow is illustrated in Fig. 3.

IV. PROBLEM FORMULATION

A. Constraints

1) *Energy Constraints*: Each edge device must satisfy an individual energy budget constraint given by:

$$P_{s,n}T_{s,n} + P_n^{\text{comp}}T_n^{\text{comp}} + P_{c,n}T_{c,n} \leq E_n^{\text{max}}, \quad 1 \leq n \leq N, \quad (\text{C1})$$

where $P_{s,n}$ and $T_{s,n}$ denote the sensing power and sensing time of device n , P_n^{comp} and T_n^{comp} represent the local computation power and time, $P_{c,n}$ and $T_{c,n}$ denote the communication power and time, and E_n^{max} is the maximum energy budget of device n .

The UAV must also operate within a total energy constraint, which includes inference, flight, and hovering energy:

$$E_u^{\text{inf}} + E_u^{\text{fly}} + E_u^{\text{hover}} \leq E_u^{\text{max}}. \quad (\text{C2})$$

2) *Latency Constraints*: Each inference task is subject to a latency constraint, which includes the sensing stage, local computation stage, communication stage, UAV flight time, and inference time. It is assumed that the UAV starts flying as soon as the edge devices begin sensing and processing the data. The latency constraint for each device is given by:

$$\max(T_{s,n} + T_n^{\text{comp}}, T_{\text{move}}) + T_{c,n} + T_u \leq T^{\text{max}}, \quad (\text{C3})$$

where $T_{s,n}$ is the sensing time, T_n^{comp} is the local computation time, T_{move} is the UAV flight time, $T_{c,n}$ is the transmission time, and T_u is the inference time at the UAV. T^{max} denotes the maximum tolerable delay for each task.

Since the UAV must remain hovering during the communication and inference stages, the hovering time must be no less than the sum of the transmission time and inference time:

$$T_{c,n} + T_u \leq T_{\text{hover}}. \quad (\text{C4})$$

3) *Bandwidth Constraint*: The total communication bandwidth at the edge server is B_{total} , thus

$$\sum_{n=1}^N B_{c,n} \leq B_{\text{total}} \quad (\text{C5})$$

4) *Communications Transmission Constraint*: The amount of data to be transmitted (after dimensionality reduction) must not exceed the achievable capacity under the given latency $T_{c,n}$, bandwidth $B_{c,n}$, and channel conditions:

$$\rho D \leq T_{c,n} B_{c,n} \log_2 \left(1 + \frac{P_{c,n} H_{u,n}}{\delta_{n,c}^2} \right) \quad (\text{C6})$$

5) *Variable Value Range*: The variable D is discrete, whereas all other variables are continuous, subject to the following constraints:

$$T_{c,n}, P_{s,n}, P_{c,n}, B_{c,n} \in \mathbb{R}^+, \quad \forall n \quad (\text{C7})$$

$$D_{\min} \leq D \leq D_{\max}, \quad D \in \mathbb{Z}^+ \quad (\text{C8})$$

$$P_{s,n}, P_{c,n} \leq P_{\max} \quad (\text{C9})$$

B. Optimization Problems

$$\begin{aligned} \text{P1:} \quad & \max_{\{T_{c,n}, P_{s,n}, P_{c,n}, D, B_{c,n}, \mathbf{p}_u\}} G \\ \text{s.t.} \quad & (\text{C1})\text{--}(\text{C9}). \end{aligned} \quad (30)$$

Substituting the definition of G into P1 yields

$$\begin{aligned} \text{P2:} \quad & \max_{\{T_{c,n}, P_{s,n}, P_{c,n}, D, B_{c,n}, \mathbf{p}_u\}} \frac{2}{K(K-1)} \sum_{n=1}^N \sum_{d=1}^D \sum_{k=1}^{K-1} \sum_{k'=k+1}^K G_{k,k'}(\tilde{z}_n(d)) \\ \text{s.t.} \quad & (\text{C1})\text{--}(\text{C9}), \end{aligned} \quad (31)$$

where

$$G_{k,k'}(\tilde{z}_n(d)) = \frac{(\mu_{n,d,k} - \mu_{n,d,k'})^2}{\sigma_{n,d}^2 + s_{n,d}^2 \alpha_n \left(\sigma_{n,b}^2 + \frac{\sigma_w^2}{P_{s,n}} \right) + \frac{\delta_{n,c}^2}{H_{u,n} P_{c,n}}}.$$

The formulated optimization problem involves both continuous and discrete variables. The objective function consists of multiple fractional summation terms containing nonlinear components such as $1/P_{s,n}$ and $1/P_{c,n}$. The denominators include non-convex terms, and the coupling of variables in constraints C1 and C6 further introduces non-convexity. Therefore, the overall problem is classified as a mixed-integer nonlinear programming (MINLP) problem and is inherently non-convex.

C. Problem Transformation

To facilitate the transformation from problem P2 to problem P3, we first define

$$\zeta_{n,d,k,k'} = \frac{2(\mu_{n,d,k} - \mu_{n,d,k'})^2}{K(K-1)}, \quad (32)$$

and

$$\begin{aligned} g_{n,d}(P_{s,n}, P_{c,n}, \mathbf{p}_u) &= \sigma_{n,d}^2 + s_{n,d}^2 \alpha_n \left(\sigma_{n,b}^2 + \frac{\sigma_w^2}{P_{s,n}} \right) \\ &\quad + \frac{\delta_{n,c}^2}{H_{u,n} P_{c,n}}, \end{aligned} \quad (33)$$

where $\zeta_{n,d,k,k'}$ denotes the normalized squared difference between $\mu_{n,d,k}$ and $\mu_{n,d,k'}$, and $g_{n,d}(\cdot)$ characterizes the effective noise-plus-interference component as a function of $P_{s,n}$, $P_{c,n}$, and $\mathbf{p}_u$.

Using these definitions, the function $G_{k,k'}(\tilde{z}_n(d))$ in P2 can be rewritten as

$$G_{k,k'}(\tilde{z}_n(d)) = \frac{\zeta_{n,d,k,k'}}{g_{n,d}(P_{s,n}, P_{c,n}, \mathbf{p}_u)}. \quad (34)$$

Applying the quadratic transform to each fractional term and introducing nonnegative auxiliary variables $y_{n,d,k,k'} \geq 0$ yields the equivalent problem formulation:

$$\begin{aligned}
 \text{P3: } \max_{\substack{\{T_{c,n}, P_{s,n}, P_{c,n}, \\ D, B_{c,n}, \mathbf{p}_u, \\ y_{n,d,k,k'}\}}} & \sum_{n=1}^N \sum_{d=1}^D \sum_{k=1}^{K-1} \sum_{k'=k+1}^K \left(2y_{n,d,k,k'} \sqrt{\zeta_{n,d,k,k'}} \right. \\
 & \left. - y_{n,d,k,k'}^2 g_{n,d}(P_{s,n}, P_{c,n}, \mathbf{p}_u) \right) \\
 \text{s.t. } & \text{(C1)–(C9),} \\
 & \text{(C10) } y_{n,d,k,k'} \geq 0, \forall n, d, k, k'.
 \end{aligned} \quad (35)$$

For conciseness, P3 is rewritten in the following form as P4.

$$\begin{aligned}
 \text{P4: } \max_{\substack{\{T_{c,n}, P_{s,n}, P_{c,n}, \\ D, B_{c,n}, \mathbf{p}_u, y_\gamma\}}} & \sum_{\gamma \in \mathcal{I}} \left(2y_\gamma \sqrt{\zeta_\gamma} - y_\gamma^2 g_\gamma \right) \\
 \text{s.t. } & \text{(C1)–(C10),}
 \end{aligned} \quad (36)$$

where $\gamma \in \mathcal{I}$, $\mathcal{I} = \{(n, d, k, k') : 1 \leq n \leq N, 1 \leq d \leq D, 1 \leq k < k' \leq K\}$, and for $\gamma = (n, d, k, k')$ we define $\zeta_\gamma = \zeta_{n,d,k,k'}$ and $g_\gamma = g_{n,d}(P_{s,n}, P_{c,n}, \mathbf{p}_u)$.

V. OPTIMIZATION PROBLEM SOLVING

To address the aforementioned non-convex optimization problem, a three-layer iterative framework combined with a block-coordinate strategy is employed. Specifically, the quadratic transform is first applied to reformulate P2 as P4. The feasible values of D are then enumerated, and for each candidate D , the resulting AO procedure is repeated until convergence. After all candidates have been evaluated, the solution yielding the largest objective value is selected as the final output. The overall solution framework is illustrated in Fig. 4.

A. Closed-Form Auxiliary-Variable Update

For a fixed candidate D , the auxiliary variables $\{y_\gamma\}$ are first updated in closed form according to (37) using the current values of $(P_{s,n}, P_{c,n}, \mathbf{p}_u)$:

$$y_\gamma^* = \frac{\sqrt{\zeta_\gamma}}{g_\gamma}, \quad \forall \gamma \in \mathcal{I}. \quad (37)$$

With $\{y_\gamma\}$ fixed, the four variable blocks are subsequently optimized in sequence. This update cycle is repeated until convergence.

B. Dimension Enumeration

To reduce the number of enumeration points and avoid ineffective computations, explicit filtering and feasibility detection are performed according to the constraint conditions, enabling rapid iterations. The detailed procedure is as follows.

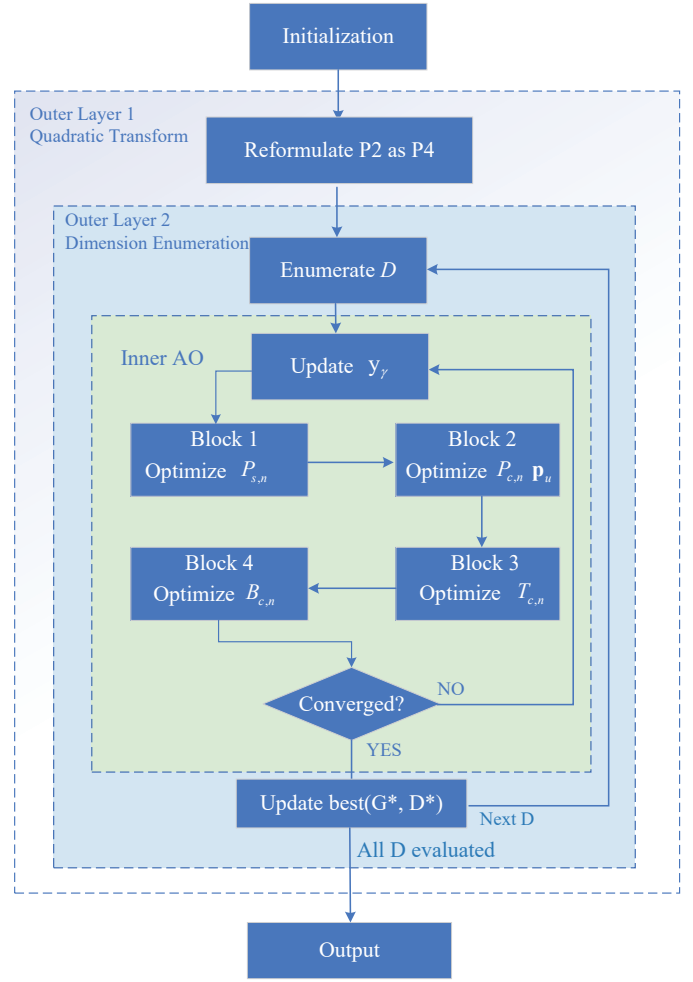


Fig. 4. The framework for solving the problem.

1) *Explicit Filtering*: To obtain the most relaxed condition, we set $P_{s,n} = 0$, $P_{c,n} = 0$, and $T_{c,n} = 0$. Under this setting, constraint (C1) can be transformed into an upper bound on D as

$$D \leq D_{\max}^{(C1)} = \frac{(E_n^{\max} - P_n^{\text{comp}} T_0) f_n}{2\rho_0 H_2 P_n^{\text{comp}}}. \quad (38)$$

By setting the UAV displacement term in (C2) to zero, i.e., letting $E_u^{\text{fly}} = 0$, the relaxed upper bound of D is obtained as

$$D \leq D_{\max}^{(C2)} = \frac{E_u^{\max} - P_u^{\text{hover}} T_{\text{hover}}}{\rho \kappa N f_0 f_u^2}. \quad (39)$$

Since (C3) contains a max term, it can be decomposed into two branches, namely the computation branch and the displacement branch. For notational simplicity, let

$$a_t = \frac{2\rho_0 H_2}{f_n}, \quad b_t = \frac{N\rho f_0}{f_u}.$$

The relaxed constraints are then expressed as

$$T_{s,n} + T_0 + a_t D + b_t D \leq T^{\max} \\ \Rightarrow D \leq D_{\max}^{(C3a)} = \frac{T^{\max} - T_{s,n} - T_0}{a_t + b_t}, \quad (40)$$

$$\frac{\Delta p}{v} + b_t D \leq T^{\max} \\ \Rightarrow D \leq D_{\max}^{(C3b)} = \frac{T^{\max} - \frac{\Delta p}{v}}{b_t}. \quad (41)$$

To make the explicit upper bound independent of the variables, the effect of the UAV displacement is ignored. In this case, $D_{\max}^{(C3b)} = \frac{T^{\max}}{b_t}$ is larger than $D_{\max}^{(C3a)}$, and therefore, constraint (C3) is relaxed as (40).

Constraint (C4) is relaxed as

$$D \leq D_{\max}^{(C4)} = \frac{T_{\text{hover}}}{b_t}. \quad (42)$$

Combined Explicit Upper Bound: The minimum among the previously obtained upper bounds of D is taken to obtain the explicitly filtered upper bound:

$$D \leq D_{\max}^{\text{exp}} = \min\{D_{\max}^{(C1)}, D_{\max}^{(C2)}, D_{\max}^{(C3a)}, D_{\max}^{(C4)}\}. \quad (43)$$

2) *Dynamic Tightening:* After the explicit filtering, a relatively small range of feasible values for D is obtained. Within this range, enumeration of D is carried out. For each enumerated D , a secondary filtering is conducted based on the constraints before entering the alternating optimization, so as to exclude infeasible values of D from the optimization loop. The mechanism of this secondary filtering is described as follows.

Step 1: Positive Available Communication Time for Each Device. Based on the current enumerated value of D and the constants, the maximum available communication time for each device is computed as

$$T_{c,n}^{\max}(D) = \min\left\{T_{\text{hover}} - b_t D, T^{\max} - (T_{s,n} + T_0 + a_t D) - b_t D\right\}. \quad (44)$$

A necessary condition is $T_{c,n}^{\max}(D) > 0, \forall n$. If this fails for any n , the candidate D is impossible.

Step 2: Total bandwidth check. The main idea is that, for a given D , under the optimal transmission rate condition, we check whether the total bandwidth is sufficient to complete the data transmission within $T_{c,n}^{\max}$. First, the available flight time is computed. Since $T_{c,n} > 0$, from constraints (C2) and (C3), we obtain

$$T_{\text{move}}^{\max}(D) = \min\left\{\frac{E_u^{\max} - E_u^{\text{hover}} - \kappa N \rho f_0 f_u^2 D}{\eta_1 v^3 + \eta_2 / v}, T^{\max} - b_t D\right\}. \quad (45)$$

Within this duration, the UAV can travel a maximum radius of

$$R_{\text{reach}}(D) = v T_{\text{move}}^{\max}(D). \quad (46)$$

For each device, starting from the initial position $\mathbf{p}_u^0$, the minimum possible distance between the UAV and device n after the maximum flight is given by

$$r_n(D) = \max\{0, \|\mathbf{p}_u^0 - \mathbf{p}_n\|_2 - R_{\text{reach}}(D)\}. \quad (47)$$

Since a smaller distance yields a stronger channel gain, the upper bound of the channel gain can be expressed as

$$H_{u,n}^{\uparrow}(D) = \frac{H_0}{r_n(D)^2 + h^2}. \quad (48)$$

By taking the maximum transmit power $P_{\max}$, the achievable rate upper bound is

$$R_n^{\uparrow}(D) = \log_2\left(1 + \frac{P_{\max} H_{u,n}^{\uparrow}(D)}{\delta_{n,c}^2}\right). \quad (49)$$

From constraint (C6), the minimum bandwidth required for each device can then be derived by considering the maximum communication time $T_{c,n}^{\max}$ and the achievable rate upper bound:

$$B_{c,n}^{\text{req}}(D) = \frac{\rho D}{T_{c,n}^{\max}(D) R_n^{\uparrow}(D)}. \quad (50)$$

This represents the minimum bandwidth required to ensure that device n can transmit all its data within the longest communication window and the best achievable rate.

Finally, from constraint (C5), the feasibility criterion is given by

$$\sum_{n=1}^N B_{c,n}^{\text{req}}(D) \leq B_{\text{total}}. \quad (51)$$

If this condition is not satisfied, it indicates that the sum of the minimum bandwidth requirements of all devices exceeds the total available bandwidth. In this case, the corresponding D is discarded directly without entering the AO process.

C. Inner Alternating Optimization

1) *Block 1: Optimization of Sensing Powers $P_{s,n}$:* After the quadratic transform, in each inner iteration, given γ , D , and $(P_{c,n}, \mathbf{p}_u, B_{c,n}, T_{c,n})$, problem P4 can be simplified as

$$\text{P4.1: } \max_{\{P_{s,n}\}} \sum_{\gamma \in \mathcal{I}} \left(A_{\gamma} - \frac{B_{\gamma}}{P_{s,n}}\right) \quad (52) \\ \text{s.t. (C1), (C9),}$$

where

$$A_{\gamma} = 2y_{\gamma} \sqrt{\zeta_{\gamma}} - y_{\gamma}^2 \left(\sigma_{n,d}^2 + s_{n,d}^2 \alpha_n \sigma_{n,b}^2 + \frac{\delta_{n,c}^2}{H_{u,n} P_{c,n}} \right),$$

$$B_{\gamma} = y_{\gamma}^2 s_{n,d}^2 \alpha_n \sigma_w^2.$$

Constraint (C1) can be equivalently expressed as

$$P_{s,n} \leq \frac{E_n^{\max} - P_n^{\text{comp}} T_n^{\text{comp}} - P_{c,n} T_{c,n}}{T_{s,n}}. \quad (53)$$

Since A_{γ} and B_{γ} are constants with $B_{\gamma} \geq 0$, the term $-\frac{B_{\gamma}}{P_{s,n}}$ is concave for $P_{s,n} > 0$. Moreover, constraint (C1) is linear, and thus problem P4.1 is convex. In addition, since $-\frac{B_{\gamma}}{P_{s,n}}$ is monotonically increasing within the feasible range

of $P_{s,n}$, the optimal solution is attained at the upper bound of $P_{s,n}$, given by

$$P_{s,n}^* = \min \left\{ P_{\max}, \frac{E_n^{\max} - P_n^{\text{comp}} T_n^{\text{comp}} - P_{c,n} T_{c,n}}{T_{s,n}} \right\}. \quad (54)$$

2) *Block 2: Joint Optimization of $(P_{c,n}, \mathbf{p}_u)$* : By fixing y_γ , D , $P_{s,n}$, $B_{c,n}$, and $T_{c,n}$, problem P4 reduces to

$$\begin{aligned} \text{P4.2: } \quad & \max_{\{P_{c,n}, \mathbf{p}_u\}} \sum_{\gamma \in \mathcal{I}} \left(C_\gamma - \frac{D_\gamma \|\mathbf{p}_u - \mathbf{p}_n\|_2^2}{P_{c,n}} \right) \\ \text{s.t. } \quad & \text{(C1), (C2), (C3), (C6), (C9),} \end{aligned} \quad (55)$$

where

$$C_\gamma = 2y_\gamma \sqrt{\zeta_\gamma} - y_\gamma^2 s_{n,d}^2 \alpha_n \left(\sigma_{n,b}^2 + \frac{\sigma_u^2}{P_{s,n}} \right), \quad (56)$$

$$D_\gamma = \frac{y_\gamma^2 \delta_{n,c}^2}{H_0}. \quad (57)$$

In the objective function (55), C_γ and D_γ are constants. The part involving $P_{c,n}$ and $\mathbf{p}_u$ is a sum of negative quadratic-over-linear terms, which is therefore convex.

- Constraint (C1) is linear with respect to $P_{c,n}$.
- Constraint (C2) is a norm inequality(second-order cone) with respect to the variable $\mathbf{p}_u$, and it is convex:

$$E_u^{\text{inf}} + \underbrace{\left(\eta_1 v^3 + \frac{\eta_2}{v} \right) \frac{\|\mathbf{p}_u - \mathbf{p}_u^0\|_2}{v}}_{E_u^{\text{fly}}(\mathbf{p}_u)} + E_u^{\text{hover}} \leq E_u^{\text{max}}.$$

- The max operation in constraint (C3) follows the same idea as in (40) and (41). Since the variable $\mathbf{p}_u$ is related to the displacement branch, the part involving $\mathbf{p}_u$ is equivalent to the following form:

$$\|\mathbf{p}_u - \mathbf{p}_u^0\|_2 \leq v(T^{\max} - T_{c,n} - T_u), \quad \forall n, \quad (58)$$

which is a convex constraint.

- Constraint (C6) can be equivalently rewritten as a lower bound on the transmit power:

$$P_{c,n} \geq \frac{\delta_{n,c}^2}{H_0} \left(2^{\frac{\rho D}{T_{c,n} B_{c,n}}} - 1 \right) \|\mathbf{p}_u - \mathbf{p}_n\|_2^2, \quad \forall n, \quad (59)$$

which takes the form of a convex quadratic constraint.

Problem P4.2 is reformulated as a standard second-order cone programming (SOCP) problem as follows. Let $\mathbf{v}_n = [x_u - x_n, y_u - y_n, h]^\top$ and introduce auxiliary variables $t_\gamma \geq 0$ for each $\gamma \in \mathcal{I}$. The objective function is equivalently written as

$$\min \sum_{\gamma \in \mathcal{I}} D_\gamma t_\gamma, \quad (60)$$

where t_γ serves as an upper bound on the quadratic-over-linear term, i.e., $t_\gamma \geq \|\mathbf{v}_n\|_2^2 / P_{c,n}$. This constraint is equivalently enforced by the following second-order cone (SOC) constraint:

$$\| [2\mathbf{v}_n; t_\gamma - P_{c,n}] \|_2 \leq t_\gamma + P_{c,n}, \quad t_\gamma \geq 0, P_{c,n} \geq 0. \quad (61)$$

Together with the convex constraints discussed above, problem P4.2 is thus reformulated as a standard SOCP, which can be efficiently solved using off-the-shelf conic solvers.

3) *Block 3: Optimization of Transmission Durations $T_{c,n}$* : Since the objective function does not involve $T_{c,n}$, by fixing the other variables, the constraints in problem (36) with respect to $T_{c,n}$ are convex (affine in $T_{c,n}$). Therefore, the optimal strategy for selecting $T_{c,n}$ is given as follows.

The lower bound is determined by constraint (C6):

$$L_n = \frac{\rho D}{B_{c,n} \log_2 \left(1 + \frac{P_{c,n} H_{u,n}}{\delta_{n,c}^2} \right)}. \quad (62)$$

The upper bounds are derived from constraints (C1), (C3), and (C4):

$$U_{\text{hover},n} = T_{\text{hover}} - T_u, \quad (63)$$

$$U_{\text{delay},n} = T^{\max} - T_u - \max\{T_{s,n} + T_n^{\text{comp}}, T_{\text{move}}\}, \quad (64)$$

$$U_{\text{energy},n} = \frac{E_n^{\max} - E_n^{\text{comp}} - P_{s,n} T_{s,n}}{P_{c,n}}, \quad (65)$$

$$U_n = \min\{U_{\text{hover},n}, U_{\text{delay},n}, U_{\text{energy},n}\}. \quad (66)$$

Thus, the optimal solution is obtained as

$$T_{c,n}^* = \begin{cases} L_n, & \text{if } L_n \leq U_n, \\ \text{infeasible}, & \text{if } L_n > U_n. \end{cases} \quad (67)$$

4) *Block 4: Bandwidth Allocation*: With all other variables fixed, the objective function does not involve $B_{c,n}$. Therefore, during bandwidth allocation, the strategy is to first assign the minimum bandwidth required to satisfy constraints (C5) and (C6). If additional bandwidth resources remain, they are redistributed to accelerate the optimization of Block 2 in the next iteration.

The procedure is as follows:

a) *Minimum bandwidth computation and allocation*:

$$B_n^{\min} = \frac{\rho D}{T_{c,n} R_n} = \frac{\rho D}{T_{c,n} \log_2 \left(1 + \frac{P_{c,n} H_{u,n}}{\delta_{n,c}^2} \right)}, \quad (68)$$

and each user is allocated its minimum bandwidth:

$$B_{c,n} \leftarrow B_n^{\min}, \quad n = 1, \dots, N. \quad (69)$$

b) *Total bandwidth check*:

$$S_{\min} = \sum_{n=1}^N B_n^{\min}. \quad (70)$$

If $S_{\min} > B_{\text{total}}$, the current solution is infeasible; if $S_{\min} = B_{\text{total}}$, the current block terminates; otherwise, if $S_{\min} < B_{\text{total}}$, the surplus bandwidth is redistributed among users according to the maximum marginal gain principle¹.

¹The maximum marginal gain principle refers to the greatest power savings achieved per unit of bandwidth. The system allocates bandwidth preferentially to users with higher marginal gain, thereby minimizing the total communication power. This approach balances the bandwidth allocation for each user via the water-level equation, ensuring an optimal bandwidth distribution.

Algorithm 1 Optimization Scheme via Quadratic Transformation, Enumeration, and Block-Coordinate Descent (QTE-BCD)

- 1: **Input:** System parameters $B_{\text{total}}, E_n^{\max}, E_u^{\max}, T^{\max}, T_{\text{hover}}, P_{\max}, \mathbf{p}_u^0, D_{\min}, D_{\max}$; convergence tolerance ϵ ; maximum number of inner iterations $I_{\max}$.
- 2: **Output:** D^* and $\{P_{s,n}^*, P_{c,n}^*, \mathbf{p}_u^*, T_{c,n}^*, B_{c,n}^*\}$.
- 3: Transform problem (31) into problem (36); set $G^* \leftarrow -\infty$.
- 4: Compute $D_{\max}^{\text{exp}}$ according to (43).
- 5: **for** each integer $D \in [D_{\min}, \min\{D_{\max}, D_{\max}^{\text{exp}}\}]$ **do**
- 6: Apply dynamic tightening according to (44)–(51); **if** infeasible, discard D and **continue**.
- 7: Initialize $\{P_{s,n}, P_{c,n}, \mathbf{p}_u, T_{c,n}, B_{c,n}\}$ to a feasible point; set $i \leftarrow 0$ and $G^{(0)} \leftarrow -\infty$.
- 8: **repeat**
- 9: $i \leftarrow i + 1$.
- 10: Update y_γ according to (37).
- 11: **Block 1:** Update $P_{s,n}$ according to (54).
- 12: **Block 2:** Update $(P_{c,n}, \mathbf{p}_u)$ via the SOCP (60)–(61).
- 13: **Block 3:** Update $T_{c,n}$ according to (67).
- 14: **Block 4:** Update $B_{c,n}$ via the KKT-based water-filling (78)–(82).
- 15: Compute $G^{(i)}$ according to (31).
- 16: **until** $i \geq 2$ and $|G^{(i)} - G^{(i-1)}| \leq \epsilon$, or $i = I_{\max}$
- 17: Check feasibility; **if** feasible and $G^{(i)} > G^*$, update $G^* \leftarrow G^{(i)}$, $D^* \leftarrow D$, and store the current solution.
- 18: **end for**
- 19: **return** D^* and $\{P_{s,n}^*, P_{c,n}^*, \mathbf{p}_u^*, T_{c,n}^*, B_{c,n}^*\}$.

c) *KKT-based water-filling:* Consider the following convex optimization problem:

$$\begin{aligned} \min_{\{B_{c,n}\}} \quad & \sum_{n=1}^N g_n(B_{c,n}) \\ \text{s.t.} \quad & B_{c,n} \geq B_n^{\min}, \quad n = 1, \dots, N, \\ & \sum_{n=1}^N B_{c,n} = B_{\text{total}}, \end{aligned} \quad (71)$$

where

$$g_n(B) = C_n \left(2^{F_n/B} - 1 \right), \quad B > 0, \quad (72)$$

with

$$C_n = \frac{\delta_{n,c}^2}{H_0} \|\mathbf{p}_u - \mathbf{p}_n\|_2^2, \quad F_n = \frac{\rho D}{T_{c,n}},$$

and $g_n(B)$ is a decreasing convex function.

The Lagrangian function is given by

$$\begin{aligned} \mathcal{L} = & \sum_{n=1}^N g_n(B_{c,n}) + \lambda \left(\sum_{n=1}^N B_{c,n} - B_{\text{total}} \right) \\ & - \sum_{n=1}^N \eta_n (B_{c,n} - B_n^{\min}), \end{aligned} \quad (73)$$

where λ and $\eta_n \geq 0$ are the dual variables.

The KKT conditions imply that, for any user n ,

$$\begin{cases} g'_n(B_{c,n}^*) + \lambda = 0, & \text{if } B_{c,n}^* > B_n^{\min}, \\ g'_n(B_n^{\min}) + \lambda \geq 0, & \text{if } B_{c,n}^* = B_n^{\min}, \end{cases} \quad (74)$$

where

$$g'_n(B) = -C_n (\ln 2) F_n \frac{2^{F_n/B}}{B^2} < 0, \quad (75)$$

with constants $C_n = \frac{\delta_{n,c}^2}{H_0} \|\mathbf{p}_u - \mathbf{p}_n\|_2^2$ and $F_n = \frac{\rho D}{T_{c,n}}$. For users with $B_{c,n}^* > B_n^{\min}$ (inactive lower bound), the KKT stationarity $g'_n(B_{c,n}^*) + \lambda = 0$ yields

$$C_n (\ln 2) F_n \frac{2^{F_n/B_{c,n}^*}}{(B_{c,n}^*)^2} = \lambda. \quad (76)$$

Let $I_n = F_n/B_{c,n}^*$. Then the stationarity becomes

$$\left(\frac{C_n \ln 2}{F_n} \right) I_n^2 e^{(\ln 2) I_n} = \lambda. \quad (77)$$

Define $\vartheta = \frac{\ln 2}{2}$ and note that $I_n^2 e^{(\ln 2) I_n} = (I_n e^{\vartheta I_n})^2$. Thus $I_n e^{\vartheta I_n} = \sqrt{\frac{\lambda F_n}{C_n \ln 2}}$, which gives the closed form via Lambert-W:

$$I_n(\lambda) = \frac{2}{\ln 2} W_0 \left(\frac{\ln 2}{2} \sqrt{\frac{\lambda F_n}{C_n \ln 2}} \right), \quad (78)$$

$$B_{c,n}^{\text{wf}}(\lambda) = \frac{F_n}{I_n(\lambda)}, \quad (79)$$

where $W_0(\cdot)$ denotes the principal branch of the Lambert-W function. Therefore, the water-filling solution is

$$B_{c,n}^*(\lambda) = \max \left\{ B_n^{\min}, B_{c,n}^{\text{wf}}(\lambda) \right\}, \quad (80)$$

where the multiplier $\lambda \geq 0$ must satisfy

$$\sum_{n=1}^N B_{c,n}^*(\lambda) = B_{\text{total}}, \quad (81)$$

otherwise the problem is infeasible. Since $B_{c,n}^{\text{wf}}(\lambda)$ is strictly decreasing with respect to λ , the aggregate $\sum_{n=1}^N B_{c,n}^*(\lambda)$ is also monotonically decreasing. Thus, in practice, λ can be efficiently determined by a bisection search over the interval $[0, \lambda_{\max}]$, until all the bandwidth resources are allocated, where

$$\lambda_{\max} = \max_n \lambda_n^{\max} = \max_n C_n (\ln 2) F_n \frac{2^{F_n/B_{c,n}^{\min}}}{(B_{c,n}^{\min})^2}. \quad (82)$$

According to the preceding analysis, the solution procedure of the original optimization problem (30) is outlined in Algorithm 1.

D. Complexity Analysis

Let $M = NDK^2$ denote the number of transformed fractional terms for a given D , and let Q denote the number of feasible enumerated values of D after explicit filtering and dynamic tightening. For each feasible D , each inner iteration first updates y_γ , which has complexity $\mathcal{O}(M)$. Blocks 1 and 3 admit closed-form solutions and thus each requires $\mathcal{O}(N)$. Block 4 solves a one-dimensional dual problem by bisection;

TABLE I
SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
Number of ISAC devices, N	3	Weighting coefficients ($\lambda_1, \lambda_2, \lambda_3, \lambda_4$)	(1, 0.1, 0.5, 1.0)
Number of classes, K	5	MAC cycles per operation, ρ_0	1
Sensing time, $T_{s,n}$	0.1 s	Device CPU frequency, f_n	0.6 GHz
Sensing carrier frequency, f_n^c	77 GHz	Device computation power, P_{comp}	0.3 W
Sensing bandwidth, $B_{s,n}$	150 MHz	UAV energy coefficients, (η_1, η_2)	$(10^{-3}, 2.5 \times 10^2)$
Permitted latency, T_{max}	2 s	UAV speed, v	15 m/s
Clutter variance, $\sigma_{n,b}^2$	{0.15, 0.2, 0.25}	Channel power gain at 1 m, H_0	-50 dB
Noise variance, σ_w^2	1	Bits per latent dimension, ρ	12
Communication channel noise power, $\delta_{n,c}^2$	10^{-12} W	UAV CPU frequency, f_u	3 GHz

each bisection step evaluates all N users, yielding complexity $\mathcal{O}(NI_{bi})$, where I_{bi} is the number of bisection iterations. The dominant computational cost lies in Block 2, where $(P_{c,n}, \mathbf{p}_u)$ are optimized via SOCP. After standard epigraph reformulation, the SOCP size grows linearly with N , and the corresponding complexity is $\mathcal{O}(I_{soc}n_{soc}^3)$, where n_{soc} is the SOCP problem size and I_{soc} is the number of solver iterations. Hence, the overall complexity of the proposed algorithm is $\mathcal{O}(QI_{AO}[M + I_{soc}n_{soc}^3 + NI_{bi}])$, where I_{AO} is the number of inner iterations for each feasible D , each consisting of one update of y_γ and one sweep over the four blocks.

VI. SIMULATION RESULTS

This section evaluates the proposed ISCC framework and optimization algorithm through simulations on a human motion recognition task. First, the simulation environment and parameter settings are described. Next, five baseline schemes are introduced for comparison to validate the effectiveness of the proposed method. We then systematically compare the classification accuracy of our scheme with those of the baselines under various bandwidth and energy consumption constraints. To further demonstrate the advantage of multi-device data fusion, the system performance is analyzed under single-device, two-device, and three-device configurations. Moreover, we further visualize the VAE latent space and compare it with PCA. Finally, a heatmap analysis is conducted to reveal the correlation between UAV position and the system's average accuracy.

A. Simulation Setup

The simulated system comprises three ISAC devices and one UAV. The UAV is initially positioned randomly within a [-300 m, 300 m] square area at a fixed altitude of 80 m. The devices are randomly located within a circular area of a 50 m radius centered at [0, 0], while the target is placed randomly within an annular region centered at [0, 0] with an inner radius of 50 m and an outer radius of 100 m. All reported inference accuracy results are averaged over 10 independent runs. Following the high-fidelity wireless sensing simulation methodology in [34], we construct a dataset for a five-class human activity recognition task, including "standing," "walking," "running," "waving," and "sitting." The resulting dataset contains 12,000 samples from three edge devices. Each motion

realization is observed by three devices at fixed positions around the target, producing three view-specific frames that are subsequently fused through feature concatenation. The dataset is partitioned at the motion-realization level using a stratified 64/16/20 train-validation-test split, which is consistently adopted for the DAE, CL-VAE, and multi-view CNN. For each device, the DAE is trained separately on paired clean and noisy frames, where the noisy frames are generated by adding multipath clutter and Gaussian noise to the corresponding line-of-sight frames. The DAE uses a learning rate of 10^{-4} , a batch size of 8, and 10 training epochs. The CL-VAE is trained on the pooled denoised data with a learning rate of 10^{-4} , a batch size of 32, and 200 epochs. The multi-view CNN classifier is trained on the concatenated latent features with a learning rate of 10^{-3} , a batch size of 64, and up to 50 epochs. Other key parameters are summarized in Table I. All simulations are performed on a desktop equipped with an Intel Core i7-13700 CPU, 16 GB RAM, and Windows 11, using Python 3.10.

B. Model Architectures

The DAE architecture consists of an encoder with three convolutional layers followed by three pooling layers, and a decoder with three transposed convolutional layers. A final 1×1 convolution layer is used to reconstruct the input. For the VAE, the encoder comprises four convolutional and pooling layers. The latent mean and log-variance are parameterized by fully connected layers, with the decoder consisting of transposed convolutions for reconstruction. The CNN classifier is constructed from a stack of four convolutional and pooling layers, followed by three fully connected layers.

C. Baselines and Proposed Scheme

The following baselines are used to evaluate the effects of denoising, feature compression, and resource optimization.

Baseline 1 (No-DAE+VAE): The raw sensing data are directly compressed by the CL-VAE without DAE-based denoising.

Baseline 2 (DAE+PCA): The sensing data are denoised by the DAE and compressed using PCA with a fixed feature dimension. This baseline represents the conventional fixed-dimensional PCA-based design adopted in existing task-oriented ISCC studies.

Baseline 3 (Equal+DAE+VAE): The system employs equal resource allocation. The sensing data are first denoised using

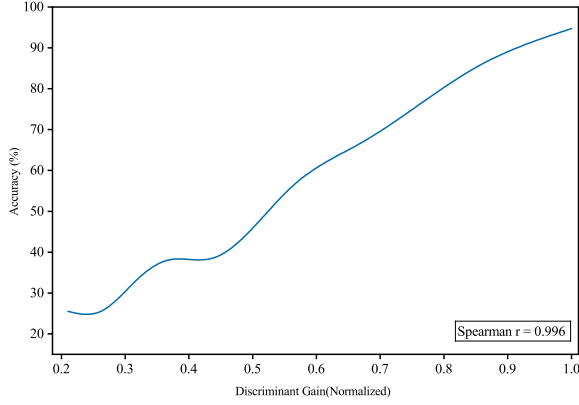


Fig. 5. Inference accuracy versus discriminant gain.

the DAE and then compressed through dimensionality reduction by the CL-VAE.

Baseline 4 (DAE+SupCon): The denoised sensing data are encoded using supervised contrastive learning [35], which encourages samples from the same class to have similar representations while separating samples from different classes. This scheme provides a non-generative learning-based feature-compression baseline.

Baseline 5 (DAE+SAE): The denoised data are processed by a supervised autoencoder with the same encoder-decoder capacity as the proposed CL-VAE. The SAE uses a deterministic latent representation and is trained with reconstruction and classification losses, without KL-divergence regularization or center loss. This scheme evaluates whether reconstruction and classification supervision alone can achieve performance comparable to the complete CL-VAE.

Proposed Scheme: The data are processed by the DAE for noise suppression and the CL-VAE for latent feature extraction. System resources are then optimally allocated according to the proposed optimization algorithm.

D. Simulation Results

To comprehensively evaluate the effectiveness of the proposed scheme, we first show the correlation between discriminant gain and inference accuracy. Its classification accuracy is then compared with the baseline methods under varying bandwidth, energy consumption, and multi-view configurations. In addition, to further demonstrate the VAE's contribution to system performance, we visualize its learned latent space. Finally, a heatmap analysis is conducted to reveal the correlation between UAV position and the resulting system accuracy.

1) Correlation Between Discriminant Gain and Inference Accuracy: The relationship between discriminant gain and inference accuracy is shown in Fig. 5. As the discriminant gain increases, the measured accuracy rises correspondingly, from 25.5% to 94.75% across the tested conditions. This strong relationship is confirmed by a Spearman rank correlation coefficient of 0.996, indicating a strong correlation between discriminant gain and inference accuracy.

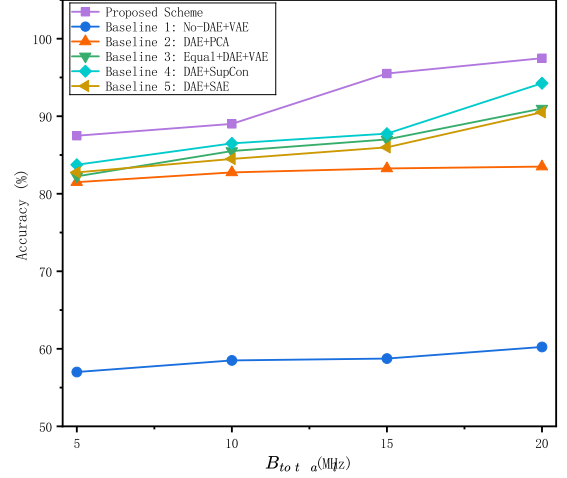


Fig. 6. Accuracy under different bandwidth thresholds.

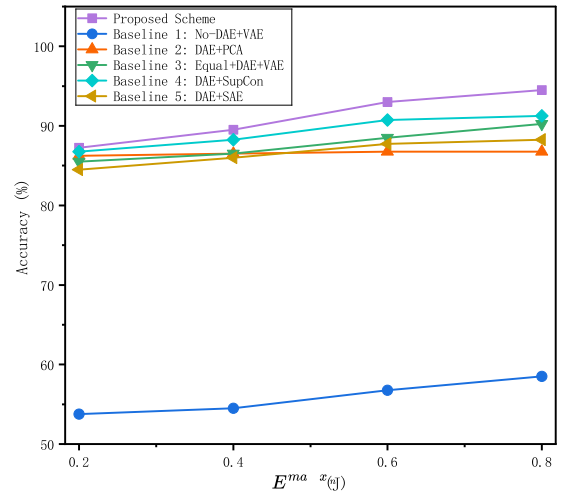


Fig. 7. Accuracy under different energy thresholds.

2) Accuracy vs. Bandwidth Threshold: The total bandwidth B_{total} is set to $\{5 \text{ MHz}, 10 \text{ MHz}, 15 \text{ MHz}, 20 \text{ MHz}\}$. The classification accuracy achieved by different schemes under these bandwidth settings is illustrated in Fig. 6. The proposed scheme consistently outperforms all five baselines across all bandwidth values. Among these, Baseline 1, which operates without the DAE module, exhibits the lowest accuracy. The superior performance of the DAE-equipped schemes highlights the critical role of the DAE module in data preprocessing. Additionally, Baseline 2 with PCA-based linear compression performs worse than the schemes using nonlinear or task-supervised latent representations, underscoring the necessity of the proposed CL-VAE module in our system. A comparison between the proposed scheme and Baseline 3 further demonstrates the performance gain attributable to the proposed optimization algorithm. Compared with DAE+SupCon and DAE+SAE, it provides accuracy gains of 2.50%–7.75% and 4.50%–9.50%, respectively. These results show that the pro-

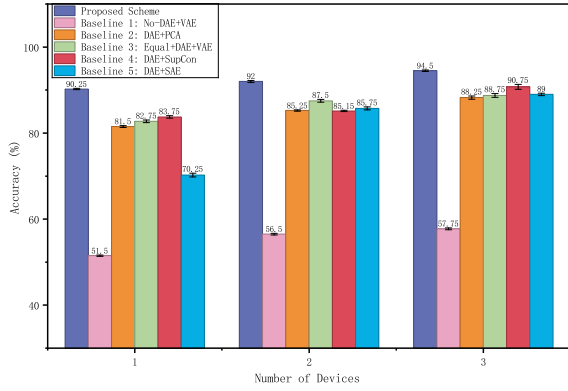
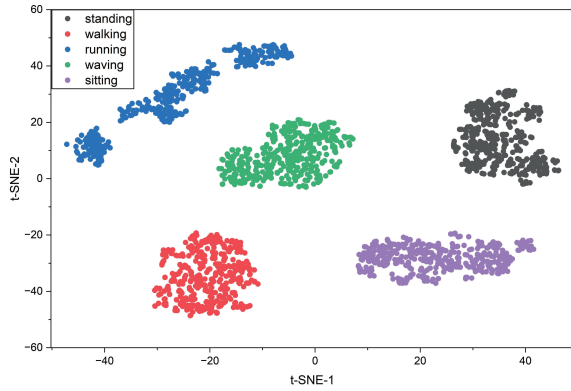
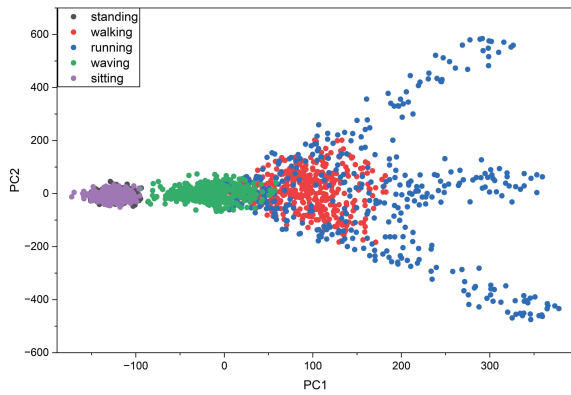


Fig. 8. Accuracy under different views. Error bars denote one standard deviation over 10 independent runs.



(a)



(b)

Fig. 9. VAE vs PCA latent space. (a) VAE. (b) PCA.

posed CL-VAE remains effective against both supervised contrastive learning and deterministic supervised autoencoding. A general trend observed across all schemes is the monotonic improvement in system accuracy with increasing bandwidth, indicating that greater resource availability directly enhances performance.

3) *Accuracy vs. Energy Threshold*: The relationship between energy thresholds and system accuracy is illustrated

TABLE II
ABLATION STUDY RESULTS.

Model	λ_1	λ_2	λ_3	λ_4	Acc. (%)
Full CL-VAE	1.0	0.1	0.5	1.0	88.75
Std VAE	1.0	0.1	0	0	20.00
w/o L_{recon}	0	0.1	0.5	1.0	82.58
w/o L_{KL}	1.0	0	0.5	1.0	87.67
w/o L_{center}	1.0	0.1	0	1.0	85.50
w/o L_{class}	1.0	0.1	0.5	0	20.00

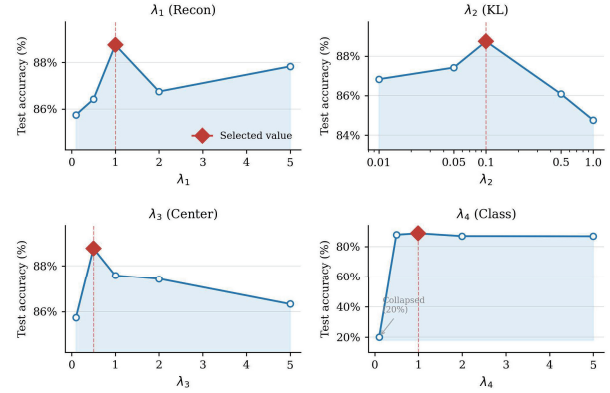


Fig. 10. Sensitivity analysis of loss weights.

in Fig. 7. Overall, system accuracy is positively correlated with the energy threshold. As observed in previous studies, the baseline without the DAE module performs the worst. However, the proposed scheme deviates from the other DAE-equipped baselines as energy becomes more abundant, achieving significant performance gains. In particular, compared with the supervised feature-compression baselines DAE+SupCon and DAE+SAE, the proposed scheme achieves gains of 0.50%–3.25% and 2.75%–6.25%, respectively. The advantage becomes more evident as the available energy increases, indicating that the proposed latent representation can more effectively utilize the allocated system resources. This result underscores the effectiveness of our resource allocation algorithm in leveraging available energy to maximize system performance.

4) *Accuracy under Single-View and Multi-View Configurations*: Fig. 8 illustrates the impact of the number of views on system performance. The recognition accuracy consistently improves as more views are incorporated, owing to the supplementary information provided by additional perspectives for the inference task. Under each view configuration, the proposed scheme achieves the highest accuracy among all five baselines, clearly demonstrating the superior effectiveness of our approach. Notably, even in the single-view scenario, the proposed scheme remains comparable to the strongest baseline in the three-view configuration and outperforms the remaining baselines in the three-view configuration, showing that the proposed design can maintain high recognition accuracy even with limited view information. Since the effects of the DAE, CL-VAE, and optimization algorithm on performance improvement have been thoroughly analyzed in the bandwidth-accuracy subsection of this section, a similar trend is observed

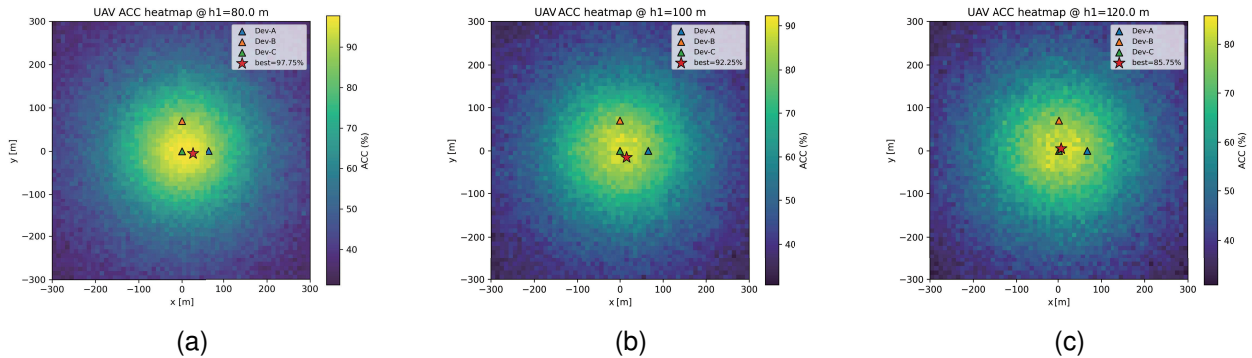


Fig. 11. Heatmaps of recognition accuracy under different UAV altitudes. (a) $h=80\text{m}$. (b) $h=100\text{m}$. (c) $h=120\text{m}$.

here and will not be further discussed. This range reflects the small-team cooperative sensing scenario targeted in this work. More devices can be added if higher accuracy is required, but since three devices already achieve 94.5% accuracy, the number of devices used in practice remains bounded, keeping the solving complexity manageable.

The run-to-run variation in inference accuracy is small, with the standard deviation remaining below 0.48 percentage points for all schemes and settings. In addition, paired t -tests are performed between the proposed scheme and DAE+SupCon, the most competitive baseline, using matched random realizations, yielding $p < 0.05$ for all considered settings. These results confirm that the observed accuracy improvements are statistically significant.

5) *Latent Space Visualization via VAE*: To provide an intuitive comparison of dimensionality reduction capabilities, the latent spaces produced by both the VAE and PCA are visualized in Fig. 9. The PCA projection shows considerable overlap among several classes, whereas the VAE latent space effectively separates the categories into distinct clusters.

6) *Ablation Study and Sensitivity Analysis*: To validate the design of the CL-VAE loss function, we conduct an ablation study and a sensitivity analysis of the loss weights.

Table II reports the test accuracy of six model variants, each trained under identical conditions with one loss term removed. Full_CL_VAE denotes the proposed model with all loss terms included, while Std_VAE refers to the standard VAE without the center loss and classification loss. The prefix “w/o” indicates that the corresponding term is excluded. All hyperparameters are fixed at the baseline values ($\lambda_1=1.0$, $\lambda_2=0.1$, $\lambda_3=0.5$, $\lambda_4=1.0$). The full CL-VAE achieves the highest accuracy of 88.75%, and every ablated variant performs worse, confirming that each term makes an independent contribution to the final performance. The substantial accuracy reduction after removing L_{class} shows that explicit classification supervision is essential for adapting the latent representation to the downstream inference task. This is consistent with the design objective of the CL-VAE, since the compressed latent features are directly used for inference rather than for reconstructing the original sensing data. Without L_{center} , the model retains the reconstruction, KL-divergence, and classification losses and achieves 85.50%. Adding the center loss further increases the accuracy to 88.75%, demonstrating its additional contribution

to intra-class compactness. PCA and the standard VAE are therefore interpreted as task-agnostic compression baselines, whereas the variant without L_{center} provides a task-supervised comparison.

To verify that the baseline weights are well-justified rather than arbitrarily chosen, we sweep each λ_i independently over one to two orders of magnitude while fixing the remaining three parameters at their baseline values. The results are shown in Fig. 10. The selected values achieve the best accuracy in their respective sweeps, confirming that the weight configuration is optimal within the searched range. Furthermore, aside from the performance collapse observed at very small values of λ_4 , where the classification signal is too weak to drive discriminative learning, accuracy variations across all other sweeps remain within 3%, demonstrating that the model is robust to moderate perturbations of the loss weights.

7) *Heatmap Analysis: UAV Position vs. Accuracy*: Fig. 11 illustrates the relationship between UAV position and system accuracy at different flight altitudes. The overall trend indicates that the system accuracy improves as the UAV moves closer to the ground devices. The proposed optimization algorithm effectively identifies the optimal UAV position that maximizes performance. Furthermore, a gradual decline in overall accuracy is observed as the UAV’s flight altitude increases.

VII. CONCLUSION

This paper proposes a UAV-assisted ISCC edge inference framework, featuring a DAE+VAE-based sensing data processing module. Following the direction of the data flow, we model the sensing, communication, and computation processes and formulate an optimization problem aimed at maximizing the discriminant gain. The optimization jointly considers communication and sensing power allocation, communication time, dimensionality reduction level, and UAV positioning. To solve the formulated problem, a QTE-BCD algorithm is developed. Simulation results verify the effectiveness of the proposed framework and algorithm. Future work will further investigate UAV-assisted ISCC edge inference, including multi-UAV cooperative scenarios, mobile device settings such as vehicular networks, as well as user privacy and security protection.

REFERENCES

- [1] C.-X. Wang *et al.*, "On the road to 6g: Visions, requirements, key technologies, and testbeds," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 905–974, Feb. 2023.
- [2] S. Lu *et al.*, "Integrated sensing and communications: Recent advances and ten open challenges," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 19 094–19 120, Jun. 2024.
- [3] M. Lyu, H. Chen, D. Wang, G. Feng, C. Qiu, and X. Xu, "Reference signal-based waveform design for integrated sensing and communications system," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2025, pp. 6621–6626.
- [4] W. Chu, X. Jia, Z. Yu, J. C. Lui, and Y. Lin, "Joint service caching, resource allocation and task offloading for mec-based networks: A multi-layer optimization approach," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, pp. 2958–2975, Apr. 2024.
- [5] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, Aug. 2017.
- [6] Z. Wang, Y. Zhao, Y. Zhou, Y. Shi, C. Jiang, and K. B. Letaief, "Over-the-air computation for 6g: Foundations, technologies, and applications," *IEEE Internet Things J.*, vol. 11, no. 14, pp. 24 634–24 658, Jul. 2024.
- [7] D. Wen, Y. Zhou, X. Li, Y. Shi, K. Huang, and K. B. Letaief, "A survey on integrated sensing, communication, and computation," *IEEE Commun. Surveys Tuts.*, vol. 27, no. 5, pp. 3058–3098, Oct. 2025.
- [8] W. Jiang *et al.*, "Terahertz communications and sensing for 6g and beyond: A comprehensive review," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 4, pp. 2326–2381, Apr. 2024.
- [9] H. Hao, C. Xu, W. Zhang, S. Yang, and G.-M. Muntean, "Joint task offloading, resource allocation, and trajectory design for multi-uav cooperative edge computing with task priority," *IEEE Trans. Mobile Comput.*, vol. 23, no. 9, pp. 8649–8663, Sep. 2024.
- [10] H. Raeis, M. Kazemi, and S. Shirmohammadi, "Cae-mas: Convolutional autoencoder interference cancellation for multiperson activity sensing with fmcw microwave radar," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–10, Feb. 2024.
- [11] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Found. Trends Mach. Learn.*, vol. 12, no. 4, pp. 307–392, Nov. 2019.
- [12] L. Zhao, D. Wu, L. Zhou, and Y. Qian, "Radio resource allocation for integrated sensing, communication, and computation networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8675–8687, Oct. 2022.
- [13] Y. Chen, Z. Chang, G. Min, S. Mao, and T. Hämmäläinen, "Joint optimization of sensing and computation for status update in mobile edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 8230–8243, Mar. 2023.
- [14] X. Li, G. Feng, Y. Liu, S. Qin, and Z. Zhang, "Joint sensing, communication, and computation in mobile crowdsensing enabled edge networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2818–2832, Apr. 2023.
- [15] Y. He, G. Yu, Y. Cai, and H. Luo, "Integrated sensing, computation, and communication: System framework and performance optimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 1114–1128, Feb. 2024.
- [16] Z. Wang, X. Mu, Y. Liu, X. Xu, and P. Zhang, "Noma-aided joint communication, sensing, and multi-tier computing systems," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 574–588, Mar. 2023.
- [17] Y. Liu, B. Zhang, D. Guo, H. Wang, and G. Ding, "Joint precoding design and location optimization in joint communication, sensing and computing of uav systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 10, no. 2, pp. 541–552, Apr. 2024.
- [18] N. Huang, C. Dou, Y. Wu, L. Qian, B. Lin, and H. Zhou, "Unmanned-aerial-vehicle-aided integrated sensing and computation with mobile-edge computing," *IEEE Internet Things J.*, vol. 10, no. 19, pp. 16 830–16 844, Oct. 2023.
- [19] B. Li, W. Liu, W. Xie, N. Zhang, and Y. Zhang, "Adaptive digital twin for uav-assisted integrated sensing, communication, and computation networks," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 4, pp. 1996–2009, Dec. 2023.
- [20] Y. Liu *et al.*, "Joint communication and computation resource scheduling of a uav-assisted mobile edge computing system for platooning vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8435–8450, Jul. 2022.
- [21] Y. Tang, G. Zhu, W. Xu, M. H. Cheung, T.-M. Lok, and S. Cui, "Integrating sensing, communication, and computation in the sky," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2024, pp. 12 886–12 890.
- [22] C. Ding, J.-B. Wang, H. Zhang, M. Lin, and G. Y. Li, "Joint mimo precoding and computation resource allocation for dual-function radar and communication systems with mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2085–2102, Jul. 2022.
- [23] S. Wang, X. Li, F. Liu, and Y. Gong, "Integrated sensing, communication, and computation over-the-air: Beam pattern design," in *Proc. IEEE Int. Conf. Commun. Workshops*, May 2023, pp. 458–463.
- [24] X. Li *et al.*, "Integrated sensing, communication, and computation over-the-air: Mimo beamforming design," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5383–5398, Aug. 2023.
- [25] G. Wu, Y. He, X. Cao, and C. Yuen, "Ai-enabled integrated sensing, communication, and computation survey: Techniques, status, and perspectives," *IEEE Internet Things J.*, vol. 12, no. 16, pp. 32 676–32 700, Aug. 2025.
- [26] D. Wen *et al.*, "Task-oriented sensing, computation, and communication integration for multi-device edge ai," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 2486–2502, Mar. 2024.
- [27] Z. Zhuang, D. Wen, Y. Shi, G. Zhu, S. Wu, and D. Niyato, "Integrated sensing-communication-computation for over-the-air edge ai inference," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3205–3220, Apr. 2024.
- [28] D. Wen, X. Jiao, P. Liu, G. Zhu, Y. Shi, and K. Huang, "Task-oriented over-the-air computation for multi-device edge ai," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 2039–2053, Mar. 2024.
- [29] X. Jiao, D. Wen, G. Zhu, W. Jiang, W. Luo, and Y. Shi, "Task-oriented over-the-air computation for edge-device co-inference with balanced classification accuracy," *IEEE Trans. Veh. Technol.*, vol. 73, no. 11, pp. 17 818–17 823, Nov. 2024.
- [30] N. Dilokthanakul *et al.*, "Deep unsupervised clustering with gaussian mixture variational autoencoders," Jan. 2017, *arXiv:1611.02648*.
- [31] L. D. Chamain, S. Qi, and Z. Ding, "End-to-end image classification and compression with variational autoencoders," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21 916–21 931, Nov. 2022.
- [32] Y. Zeng and R. Zhang, "Energy-efficient uav communication with trajectory optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3747–3760, Jun. 2017.
- [33] F. Zhou, Y. Wu, R. Q. Hu, and Y. Qian, "Computation rate maximization in uav-enabled wireless-powered mobile-edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 1927–1941, Sep. 2018.
- [34] G. Li, S. Wang, J. Li, R. Wang, X. Peng, and T. X. Han, "Wireless sensing with deep spectrogram network and primitive based autoregressive hybrid channel model," in *Proc. IEEE 22nd Int. Workshop Signal Process. Adv. Wireless Commun.*, Sep. 2021, pp. 481–485.
- [35] P. Khosla *et al.*, "Supervised contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 18 661–18 673.



Guofang Wu received the M.S. degree in communication engineering from Henan University of Technology, China, in 2021. She is currently pursuing the Ph.D. degree in Electronic Information with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China. Her current research interests focus on integrated sensing, communication, and computation, edge inference, and low-altitude intelligence.



Yejun He (Senior Member, IEEE) received the Ph.D. degree in Information and Communication Engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2005. From 2005 to 2006, he was a Research Associate with the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, China. From 2006 to 2007, he was a Research Associate with the Department of Electronic Engineering, Faculty of Engineering, The Chinese University of Hong Kong, Hong Kong,

China. In 2012, he joined the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, as a Visiting Professor. From 2013 to 2015, he was an Advanced Visiting Scholar (Visiting Professor) with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. From 2023 to 2024, he was an Advanced Research Scholar (Visiting Professor) with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore.

Since 2006, he has been a faculty of Shenzhen University, where he is currently a Distinguished Professor with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China, the Director of Sino-British Antennas and Propagation Joint Laboratory of Ministry of Science and Technology of the People's Republic of China (MOST), the Director of the Guangdong Engineering Research Center of Base Station Antennas and Propagation, and the Director of the Shenzhen Key Laboratory of Antennas and Propagation. He was selected as an Expert with Special Government Allowance from the State Council in China, and a Leading Talent in the "Guangdong Special Support Program" in 2024. He was promoted to the Shenzhen "Pengcheng Scholar" Distinguished Professor in 2020. He has authored or coauthored more than 370 refereed journal and conference papers and seven books. He holds more than 30 patents. His research interests include wireless communications, antennas, and radio frequency.

Dr. He was also a recipient of the Shenzhen Overseas High-Caliber Personnel Level B (Peacock Plan Award B) and Shenzhen High-Level Professional Talent (Local Leading Talent). He received the Second Prize of Shenzhen Science and Technology Progress Award in 2017, the Three Prize of Guangdong Provincial Science and Technology Progress Award in 2018, the Second Prize of Guangdong Provincial Science and Technology Progress Award in 2023, and the 10th Guangdong Provincial Patent Excellence Award in 2023. He is currently the Chair of IEEE Antennas and Propagation Society-Shenzhen Chapter and obtained the 2022 IEEE APS Outstanding Chapter Award. He has served as a Technical Program Committee Member or a Session Chair for various conferences, including the IEEE Global Telecommunications Conference (GLOBECOM), the IEEE International Conference on Communications (ICC), the IEEE Wireless Communication Networking Conference (WCNC), and the IEEE Vehicular Technology Conference (VTC). He served as the TPC Chair for IEEE ComComAp 2021 and the General Chair for IEEE ComComAp 2019. He was selected as a Board Member of the IEEE Wireless and Optical Communications Conference (WOCC). He served as the TPC Co-Chair for WOCC 2023/2022/2019/2015, APCAP 2023, UCMMT 2023, ACES-China2023, NEMO 2020 and so on. He acted as the Publicity Chair of several international conferences such as the IEEE PIMRC 2012. He has served as an Executive Chair of 2024 IEEE International Workshop of Radio Frequency and Antenna Technologies (RFAT 2024), 2025 IEEE International Workshop of Radio Frequency and Antenna Technologies (RFAT 2025), as well as 2026 IEEE International Conference of Radio Frequency and Antenna Technologies (RFAT 2026). He is acting as an Executive Chair of 2027 IEEE International Conference of Radio Frequency and Antenna Technologies (RFAT2027). He is the Principal Investigator for over 40 current or finished research projects, including the National Natural Science Foundation of China, the Science and Technology Program of Guangdong Province, and the Science and Technology Program of Shenzhen City. He has served as a Reviewer for various journals, such as the IEEE Transactions on Wireless Communications, the IEEE Transactions on Vehicular Technology, the IEEE Transactions on Communications, the IEEE Transactions on Industrial Electronics, the IEEE Transactions on Antennas and Propagation, the IEEE Wireless Communications, the IEEE Communications Letters, the International Journal of Communication Systems.

Dr. He is a Fellow of IET, and a Fellow of China Institute of Communications (CIC). He is serving as an Associate Editor for IEEE Transactions on Vehicular Technology, IEEE Transactions on Antennas and Propagation, IEEE Antennas and Wireless Propagation Letters, IEEE Antennas and Propagation Magazine, International Journal of Communication Systems, China Communications, and ZTE Communications.



Xiaowen Cao (Member, IEEE) received the B.Eng. and Ph.D. degrees from the Guangdong University of Technology in 2017 and 2022, respectively. She is now an assistant professor in College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China. Her research interests include edge learning, over-the-air computation, as well as integrated sensing, communication, and computation.



Mohsen Guizani (Fellow, IEEE) received the BS (with distinction), MS and PhD degrees in Electrical and Computer engineering from Syracuse University, Syracuse, NY, USA in 1985, 1987 and 1990, respectively. He is currently a Professor of Machine Learning and the Associate Provost at Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, UAE. Previously, he worked in different institutions in the USA. His research interests include applied machine learning and artificial intelligence, Internet of Things (IoT),

intelligent autonomous systems, smart city, and cybersecurity. He was elevated to the IEEE Fellow in 2009 and was listed as a Clarivate Analytics Highly Cited Researcher in Computer Science in 2019, 2020 and 2021. Dr. Guizani has won several research awards including the "2015 IEEE Communications Society Best Survey Paper Award", the Best ComSoc Journal Paper Award in 2021 as well five Best Paper Awards from ICC and Globecom Conferences. He is the author of ten books and more than 800 publications. He is also the recipient of the 2017 IEEE Communications Society Wireless Technical Committee (WTC) Recognition Award, the 2018 AdHoc Technical Committee Recognition Award, and the 2019 IEEE Communications and Information Security Technical Recognition (CISTC) Award. He served as the Editor-in-Chief of IEEE Network and is currently serving on the Editorial Boards of many IEEE Transactions and Magazines. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the Chair of the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker and is currently the IEEE ComSoc Distinguished Lecturer.