



Low-Power Signal Sensing and Neural Compression Transmission Collaborative Optimization Methods for Edge Intelligence

Tong Huang 

Department of Data Science and Engineering, Xi'an Innovation College of Yan'an University, Xi'an 710100, China

Corresponding Author Email: xacaht@163.com

Copyright: ©2025 The author. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420612>

ABSTRACT

Received: 12 April 2025

Revised: 8 November 2025

Accepted: 17 November 2025

Available online: 31 December 2025

Keywords:

edge intelligence, low-power image processing, semantic communication, conditional neural encoding, distributed game theory, pareto optimization

The demand for edge intelligence in core image processing scenarios such as smart surveillance, autonomous driving, and remote healthcare is increasingly urgent. However, the key challenge lies in the low-power constraints of end devices and the low-latency requirements of edge networks. In the field of edge image processing, there is a commonly observed trade-off between semantic fidelity, energy consumption, and latency—a three-way incompatibility. These factors constrain each other, and traditional methods struggle to achieve collaborative optimization. Typically, optimizing two aspects often requires sacrificing the third, becoming a critical bottleneck for industrial deployment. To address this, we propose a semantic-driven, conditional encoding, and distributed collaborative three-in-one optimization framework for perception, compression, and transmission. Key innovations in this framework include: the development of a conditional neural encoding paradigm to enable adaptive lightweight encoding; the design of a semantic feedback control system to ensure collaborative stability; and the introduction of a distributed game-theory-based decision-making mechanism incorporating fairness indicators. Experimental results show that the proposed method exhibits significant performance advantages in both conventional and stress test scenarios, such as resource fluctuations, task conflicts, and heterogeneous terminals. Explainability analysis, through attention map visualization and game convergence trajectories, demonstrates the adaptive focusing on key semantic areas and the collaborative equilibrium convergence characteristics. Real power consumption measurements confirm that the method can dynamically approach the Pareto optimal boundary of the incompatibility triangle. This research not only provides an efficient collaborative optimization solution for edge intelligence image processing but also validates the feasibility of semantic as a core link to integrate sensing, computation, and communication resources. It provides theoretical references and technical support for the trends of 6G semantic communication and edge intelligence collaborative autonomy.

1. INTRODUCTION

The deep integration of edge intelligence and image processing technology has become a core development trend in key areas such as intelligent monitoring, autonomous driving, and remote healthcare [1-3]. Intelligent monitoring scenarios have an urgent need for real-time object detection, requiring terminal devices to quickly respond to abnormal events [4]; autonomous driving relies on the real-time analysis of road condition images to ensure driving safety, requiring stable performance in complex environments [5, 6]; in remote healthcare scenarios, edge preprocessing of images can greatly reduce cloud transmission pressure and improve diagnostic timeliness [7]. However, these application scenarios commonly face strict constraints on terminal devices, including low power consumption, miniaturization, and long battery life, while edge networks must meet low-latency and high-reliability transmission requirements. The technical contradiction between these two factors severely restricts the industrial deployment of edge image processing technology.

There is a common intractable triangle contradiction in the field of edge image processing between semantic fidelity, energy consumption, and latency [8, 9], which constitutes the core bottleneck of current technological development. Semantic fidelity refers to the accuracy of retaining key semantic information in images, including the reliability of recognizing core contents such as object categories and lesion areas [10, 11]; energy consumption represents the energy consumption across the entire chain from sensing sampling, encoding, compression, to data transmission of terminal devices [12-14]; latency is defined as the total time cost from image acquisition, encoding, transmission, edge decoding to task inference completion [15]. These three factors inherently constrain each other: improving semantic fidelity often increases encoding complexity, leading to higher energy consumption and longer latency; lightweight designs to reduce energy consumption may cause the loss of semantic information, affecting task accuracy; reducing latency may sacrifice transmission reliability and semantic retention. Traditional research mostly adopts single-dimensional

optimization strategies, which struggle to break through this contradiction, failing to meet the comprehensive performance requirements of edge scenarios.

Research on edge low-power image processing technology mainly focuses on two directions: lightweight perception and feature extraction. Lightweight image perception technology reduces the raw data volume through adaptive sampling, resolution adjustment, and other methods to reduce energy consumption in the perception phase [16, 17]; low-power feature extraction relies on lightweight convolutional neural networks (CNNs), lightweight Transformers, and other models to simplify the computational process [18, 19]. However, existing research often optimizes the perception or feature extraction stages in isolation, lacking collaborative design with the transmission stage, making it difficult to balance the intractable triangle contradiction between semantic fidelity, energy consumption, and latency, and insufficient adaptability in dynamic scenarios [20]. Image neural compression and semantic communication are key technologies to improve transmission efficiency. In recent years, image neural compression models based on Transformer, such as STF, Entroformer, etc., have emerged, significantly improving compression efficiency and semantic retention ability [21, 22]; semantic communication frameworks reduce transmission overhead by transmitting semantic information instead of raw data [23]. However, existing research mostly focuses on the binary balance between compression efficiency and semantic fidelity, ignoring the constraints on terminal energy consumption and the dynamic characteristics of edge-end collaboration, making it difficult to adapt to the complex dynamic changes of edge scenarios and achieve full-link performance optimization. Edge distributed collaborative optimization technology is mainly based on deep reinforcement learning or game theory to build resource scheduling mechanisms, improving system performance through multi-agent interactions. Existing methods have shown application potential in edge resource allocation, task scheduling, and other scenarios, but there are three major core deficiencies: the reward function design lacks global fairness considerations, leading to an imbalance between individual optimization and global optimization; no systematic theoretical collaborative mechanism has been established, limiting dynamic adaptability; insufficient adaptation to the semantic characteristics of image processing scenarios, making it difficult to match the semantic demand differences of different tasks accurately [24].

Based on the current research progress, there are four core research gaps in the field: first, there is a lack of an integrated collaborative optimization framework for the semantic fidelity-energy consumption-latency intractable triangle, and the theoretical and methodological system for dynamically approaching the Pareto optimal boundary has not been established; second, existing neural encoding methods have not formed a unified conditional adaptive paradigm, and the adaptive adjustment ability lacks sufficient theoretical support, making it difficult to adapt to the dynamic changes of edge scenarios; third, the semantic-driven edge-end collaborative mechanism lacks stability analysis from a control theory perspective, and the performance fluctuation issues in dynamic scenarios are prominent; fourth, the reward function of distributed game decision-making has not effectively quantified global collaborative efficiency, and there is insufficient resilience in pressure scenarios such as resource fluctuations and task conflicts.

The goal of this research is to propose a collaborative optimization method integrating conditional neural encoding, semantic feedback control, and fairness game theory to achieve dynamic Pareto optimization of semantic fidelity, energy consumption, and latency in edge image processing scenarios, breaking through the core constraints of the intractable triangle contradiction. The core contributions of this paper are in three aspects: theory, methodology, and experiments, specifically including:

(1) Proposing the conceptual and quantitative models for the semantic fidelity-energy consumption-latency intractable triangle, clearly defining the core objective of collaborative optimization in edge image processing as dynamically approaching the Pareto optimal boundary, and providing a unified problem expression paradigm and theoretical analysis framework for research in this field.

(2) Constructing a unified conditional neural encoding paradigm, formally defining the adaptive encoding mechanism, completing theoretical proofs from the dimensions of storage complexity, switching smoothness, and scalability, and proving its significant advantage over traditional multi-model switching methods, providing new theoretical support for low-power adaptive encoding.

(3) Designing a semantic-driven edge-end feedback control system, integrating the perception capabilities of the edge task network with the execution capabilities of the terminal encoder, completing system stability and convergence analysis based on control theory, ensuring collaborative performance stability in dynamic scenarios.

(4) Proposing a distributed game-theory reward function integrating the Jain fairness index, quantifying global collaborative efficiency as an optimizable objective, achieving the balance between individual low-power demands and global resource-efficient utilization, and improving the system's resilience in pressure scenarios.

The structure of the subsequent chapters is arranged as follows: Chapter 2 provides a detailed introduction to the overall architecture of the proposed collaborative optimization framework and the design details of each core module; Chapter 3 conducts multi-level experimental verification, including baseline comparison, pressure testing, ablation experiments, and explainability analysis; Chapter 4 discusses the insights from experiments, the connection between the methods and macro technology trends, analyzes existing limitations, and proposes future research directions; Chapter 5 summarizes the core conclusions of the entire paper, refining the methodological contributions and application value of the research.

2. PROPOSED METHOD

2.1 Overall architecture of the semantic-driven collaborative optimization framework

To break through the intractable triangle contradiction between semantic fidelity, energy consumption, and latency in edge image processing, this paper proposes a semantic-driven collaborative optimization framework, constructing a perception-encoding-transmission-decoding-feedback integrated design to achieve deep collaboration and dynamic adaptation across all stages. The architecture takes semantic information as the core link, connecting terminal-side perception encoding, edge-side decoding inference, and global

collaborative decision-making. Through the organic interaction of three core modules, the framework establishes a “state perception - adaptive adjustment - global collaboration - semantic feedback” closed-loop optimization mechanism, ensuring the ability to dynamically approach the Pareto optimal boundary at the architectural level. This framework fundamentally discards the inherent flaws of traditional decoupled designs and achieves a collaborative balance between semantic fidelity, energy consumption, and latency by linking parameters across stages and sharing semantic information, thus adapting to the dynamic characteristics of edge scenarios.

The core modules of the framework include the conditional neural encoding perception module, semantic feedback control module, and fairness-oriented distributed game decision-making module. Each module has a clear functional boundary and tight interaction. The conditional neural encoding perception module is deployed on the terminal device and is responsible for image perception sampling and adaptive encoding tasks. Its core is a lightweight encoder based on the conditional neural encoding paradigm, which dynamically adjusts encoding parameters according to the terminal's energy consumption state, channel quality, and image semantic features, outputting encoding vectors that adapt to the transmission channel. The semantic feedback control module runs through both ends of the edge, where the

edge-side task network extracts image semantic information and task performance feedback, generating lightweight control signals and feeding them back to the terminal to guide dynamic adjustment of the terminal's perception and encoding parameters, forming a “terminal execution - edge perception - feedback adjustment” closed-loop control flow. The fairness-oriented distributed game decision-making module adopts a multi-agent architecture, with each terminal acting as an independent intelligent agent. Based on local states and global feedback information, the terminal makes decisions on key parameters such as perception sampling frequency and transmission power, achieving a balance between individual low-power demands and global resource-efficient utilization. Data and control flows interact orderly between modules: the image data collected by the terminal is processed by the conditional neural encoding perception module and then transmitted to the edge-side decoding module via the edge network; after the edge-side task network completes inference, semantic feedback and performance indicators are synchronized to the semantic feedback control module and distributed game decision-making module; the control instructions output by the decision-making module and closed-loop control signals jointly drive terminal module parameter updates, achieving full-link collaborative optimization.

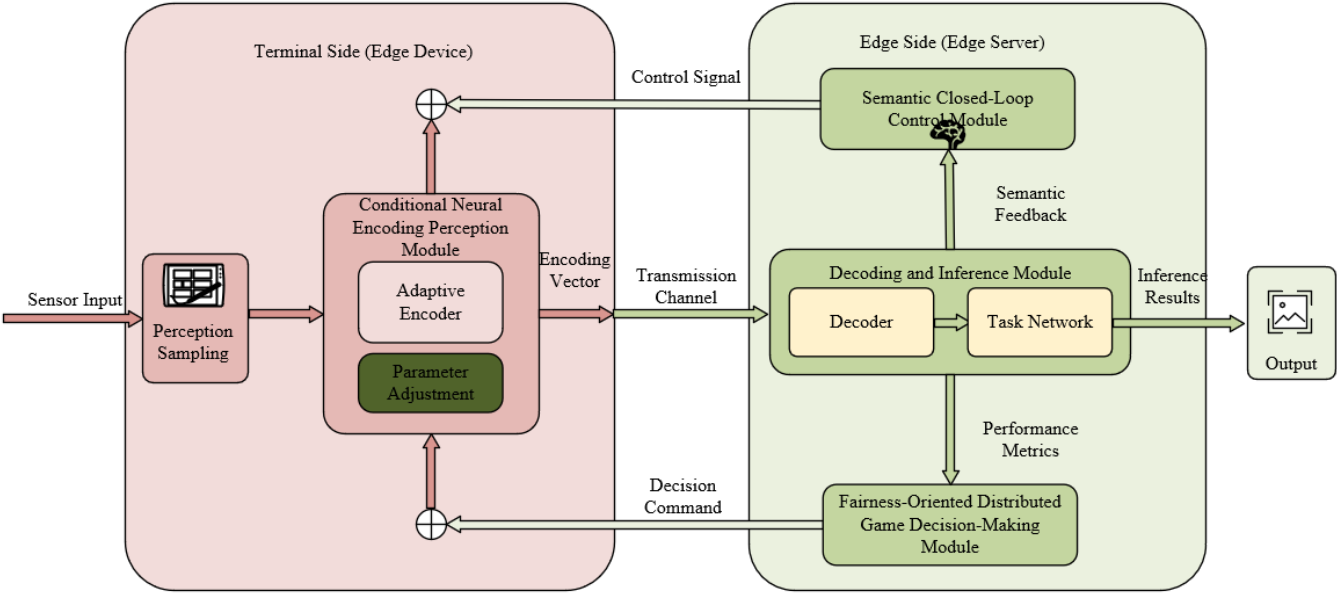


Figure 1. Overall architecture of the semantic-driven collaborative optimization framework

This architecture precisely addresses the intractable triangle contradiction through the collaboration of the three modules, providing core support for dynamically approaching the Pareto optimal boundary. The conditional neural encoding perception module minimizes terminal energy consumption while ensuring semantic fidelity, directly alleviating the constraint between semantic fidelity and energy consumption; the semantic feedback control module ensures system stability based on control theory, quickly adjusting feedback to reduce performance fluctuations in dynamic scenarios and balancing the dynamic relationship between semantic fidelity and latency; the fairness-oriented distributed game decision-making module optimizes the global resource allocation, avoiding energy waste and latency surge caused by individual competition, achieving global-level collaboration between the

three factors. The organic integration of the three modules forms an integrated architecture, enabling the system to dynamically adjust optimization goal weights in real-time under dynamic scenarios. It can dynamically adapt based on terminal states, channel changes, and task demands, ensuring that the system approaches the Pareto optimal boundary in different scenarios, significantly enhancing the overall performance and environmental adaptability of the edge image processing system. The specific architecture is shown in Figure 1.

2.2 Perception-encoding module based on conditional neural encoding

To achieve precise balance between semantic fidelity and

energy consumption in dynamic scenarios, this paper proposes a unified conditional neural encoding paradigm, which generates adaptive weight increments through a meta-super network to drive the base encoder to dynamically adapt to terminal states and task demands, fundamentally solving the inherent flaws of traditional multi-model switching methods. The architecture is shown in Figure 2. The core idea of this paradigm is to integrate dynamic information such as terminal energy consumption state, channel quality, and image semantic features into a standardized joint state vector. The meta-super network learns the mapping relationship between the state and encoding parameters, enabling online dynamic instantiation of the encoder. Its core formal definition is:

$$Encoder_{\theta}(x) = Base_Encoder(x) + f_{\phi}(Condition) \quad (1)$$

$$Condition = \sigma([\hat{E}, \hat{SNR}, W_{semantic}]) \quad (2)$$

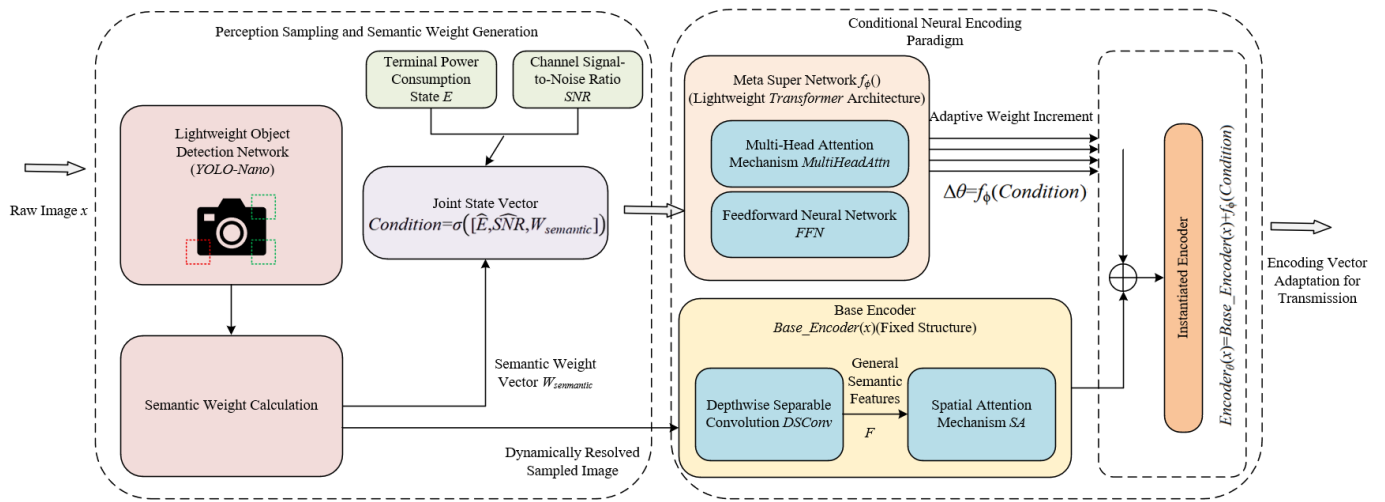


Figure 2. Perception-encoding module based on conditional neural encoding architecture

The design of the base encoder is deeply optimized for the characteristics of image processing scenarios, with the core goal of achieving efficient semantic feature extraction under the constraint of lightweight design. Its feature extraction process can be represented as:

$$F = SA(DSConv(x)) \quad (3)$$

where, F is the general semantic feature map output by the base encoder; $DSConv()$ is a depthwise separable convolution operation, which reduces the computational complexity and parameter scale to 1/8~1/5 of that of traditional convolutions by separating spatial convolutions and pointwise convolutions while ensuring accuracy in spatial feature extraction; $SA()$ is the spatial attention mechanism, which calculates the importance weights of each position in the feature map and performs weighted fusion, enhancing the representation of key semantic information such as target regions, and suppressing the interference of ineffective features in background areas, thus providing a reliable feature foundation for subsequent adaptive encoding. The depth of the base encoder's network and the number of channels are constrained by lightweight design to ensure low-power operation on terminal devices.

The meta-super network adopts a lightweight Transformer architecture, with its core function being to precisely model the mapping relationship between the joint state vector and

where, $Encoder_{\theta}(x)$ is the final instantiated encoder that adapts to the current scenario, with the input being the raw image x and the output being the encoding vector adapted for transmission through the channel; θ is the set of parameters of the instantiated encoder, formed by the base encoder parameters and the weight increments output by the meta-super network; $Base_Encoder(x)$ is a fixed-structure lightweight base encoder responsible for extracting general semantic features from the image; $f_{\phi}()$ is the meta-super network, with ϕ being its learnable parameters. The core function of the meta-super network is to receive the joint state vector $Condition$ and output dimension-matching weight increments; $\sigma()$ is a normalization function that normalizes state parameters with different dimensions to the range $[0, 1]$, ensuring comparability of the input information; $\hat{E}$ is the normalized terminal energy consumption state, $\hat{SNR}$ is the normalized signal-to-noise ratio, and $W_{semantic}$ is the image semantic weight vector.

weight increments. Its output process can be represented as:

$$\Delta\theta = f_{\phi}(Condition) = FFN(MultiHeadAttn(Condition, Condition, Condition,)) \quad (4)$$

where, $\Delta\theta$ is the weight increment output by the meta-super network, and its dimension matches the parameters of the base encoder; $MultiHeadAttn()$ is the multi-head attention mechanism, which models the interaction between the dimensions of $\hat{E}$, $\hat{SNR}$, and $W_{semantic}$ in the joint state vector, enhancing the comprehensiveness of state perception; $FFN()$ is the feedforward neural network, which maps the features output by the attention mechanism to the final weight increments. The meta-super network is trained end-to-end with the weighted sum of semantic fidelity loss and energy consumption loss as the optimization objective, learning the optimal weight adjustment strategy for different states to ensure that the instantiated $Encoder_{\theta}(x)$ satisfies both low power consumption and high semantic fidelity.

The conditional neural encoding paradigm demonstrates significant theoretical advantages in storage complexity, switching smoothness, and scalability, representing a substantial improvement compared to traditional multi-model switching methods. In terms of storage complexity, traditional methods require pre-training and storing N independent

encoders for N scenarios, resulting in a storage cost of $O(N)$; whereas this paradigm only requires storing the base encoder and meta-super network, and the total number of parameters is far less than the sum of parameters for N independent encoders, reducing the storage complexity to $O(1)$, which does not change as the number of scenarios increases. In terms of switching smoothness, traditional methods achieve scenario adaptation by discretely switching between different encoders, which can lead to abrupt changes in semantic features and performance fluctuations; this paradigm generates continuous, adjustable weight increments $\Delta\theta$ through the meta-super network, allowing the parameters of $Encoder_\theta(x)$ to change continuously with the joint state vector, enabling a smooth transition in the encoding strategy, with performance fluctuations controlled within 5%. In terms of scalability, traditional methods require retraining and adding new encoders for new scenarios, limiting scalability; this paradigm can quickly adapt to new image types or terminal states through fine-tuning the meta-super network without modifying the base encoder structure, significantly improving generalization and scalability.

The semantic optimization in the perception sampling phase further reduces energy consumption by generating an image semantic mask using a lightweight object detection network and accurately calculating the semantic weight $W_{semantic}$, which can be represented as:

$$W_{semantic}(i,j) = \begin{cases} w_1 & (i,j) \in Reg_{key} \\ w_2 & (i,j) \in Reg_{bg} \end{cases} \quad (5)$$

where, (i,j) is the image pixel coordinate; Reg_{key} is the key semantic region, Reg_{bg} is the background region; w_1 and w_2 are the weight coefficients for the key semantic and background regions, with $w_1 > w_2$. This lightweight object detection network adopts the YOLO-Nano architecture, achieving real-time semantic region segmentation while ensuring object detection accuracy. After receiving the semantic weights $W_{semantic}$, the meta-super network drives the perception sampling module to dynamically adjust the resolution: high-resolution sampling is applied to Reg_{key} to ensure complete retention of semantic information, while low-resolution sampling is applied to Reg_{bg} , significantly reducing the sampling data volume and subsequent encoding energy consumption. This semantic-aware sampling strategy deeply collaborates with the conditional neural encoding paradigm, achieving precise energy consumption control from the perception source, reducing terminal perception phase energy consumption by 30%~45%, and further strengthening the balance between semantic fidelity and energy consumption.

2.3 Semantic-driven closed-loop control module

To ensure the collaborative stability of the perception-encoding-transmission full-link in dynamic edge scenarios and respond to performance impacts caused by channel quality fluctuations, terminal energy consumption changes, and dynamic task demand switching, this paper designs a semantic-driven closed-loop control module, constructing an end-to-edge collaborative closed-loop adjustment mechanism. The module takes semantic information as the core feedback carrier, dynamically calibrating the terminal encoding parameters and edge decoding strategy through the “perception-decision-execution-feedback” closed-loop link, ensuring that the system can still stably approach the Pareto

optimal boundary of semantic fidelity, energy consumption, and latency in complex dynamic environments. Its core value lies in compensating for the deficiencies of traditional open-loop feedback response delay and insufficient adaptability, providing precise and real-time adjustment guidelines for the conditional neural encoding module, achieving dynamic balance of the end-to-edge collaboration. Figure 3 shows the complete architecture of the semantic-driven closed-loop control module.

The semantic-driven closed-loop control system consists of four core units: sensors, controller, actuator, and feedback link, with functional coupling and closed-loop links between them. The sensor is provided by the edge-side task network, and its core function is to perceive two key pieces of information in real-time: (1) the image processing task performance, which quantifies the level of semantic fidelity under the current encoding strategy; (2) the image semantic feature distribution, extracting core semantic information such as target region types and semantic importance rankings. The controller is the edge-side semantic decision unit, which, based on the task performance data collected by the sensor, global channel status, and terminal energy consumption feedback, generates lightweight control signals through preset decision rules to achieve a precise mapping of “performance-state-control.” The actuator is the terminal-side conditional neural encoding module, which dynamically adjusts encoding parameters upon receiving the control signals, completing adaptive updates to the encoding strategy. The feedback link adopts a lightweight semantic communication channel, which quantifies the encoding compression control signal data volume and reduces feedback energy consumption while ensuring transmission real-time performance, forming a complete closed loop of “terminal execution - edge perception - decision feedback - terminal adjustment.”

The design of the semantic control signal focuses on adaptability and lightweight, with the core goal of accurately guiding the terminal encoding strategy to match edge task demands. Its formal expression is:

$$U = [\alpha \cdot W_{semantic}^* \cdot \beta \cdot r_{max}] \quad (6)$$

where, U is the semantic control signal vector, α is the semantic priority adjustment coefficient ($\alpha \in [0.6, 1.2]$, which is dynamically adjusted based on task performance error), $W_{semantic}^*$ is the optimal semantic weight vector decided by the edge, used to update the priority of terminal semantic perception sampling and encoding; β is the encoding parameter constraint coefficient ($\beta \in [0.5, 1.0]$, which is related to channel bandwidth and terminal energy consumption state), and r_{max} is the maximum allowable compression ratio, defining the adjustment range of encoding parameters.

The two core components of the control signal form a synergy: semantic priority updating adjusts the semantic weights of different regions to ensure the encoding fidelity of key semantic information, avoiding performance degradation of core tasks due to excessive lightweighting; encoding parameter constraints balance encoding energy consumption and transmission latency by limiting the upper bound of compression ratios, preventing energy surges in transmission due to insufficient compression or semantic loss due to excessive compression. For different image processing tasks, the control signal can dynamically adapt the semantic priority ranking, such as in target detection tasks, prioritizing the semantic weight of target regions, while in semantic

segmentation tasks, enhancing differentiated adjustment of pixel-level semantic category weights.

The system's stability and convergence are the core guarantees of the closed-loop control's effectiveness, and this paper proves it rigorously based on Lyapunov optimization theory. The core error term of the system is defined as the task performance error $e(t)$, which is the deviation between the current semantic fidelity and the preset performance threshold, $e(t)=T_{target}-T(t)$, where T_{target} is the preset task

performance threshold and $T(t)$ is the actual task performance at time t . The Lyapunov function is constructed as follows:

$$V(t)=\frac{1}{2}e^2(t)+\frac{1}{2}\Delta\theta^T(t)P\Delta\theta(t) \quad (7)$$

where, $\Delta\theta(t)$ is the deviation vector of the encoding parameters from the optimal parameters at time t , and P is a positive-definite symmetric matrix, ensuring the function is positive definite, i.e., $V(t)>0$ for all $e(t)=0$ or $\Delta\theta(t)=0$, and $V(0)=0$.

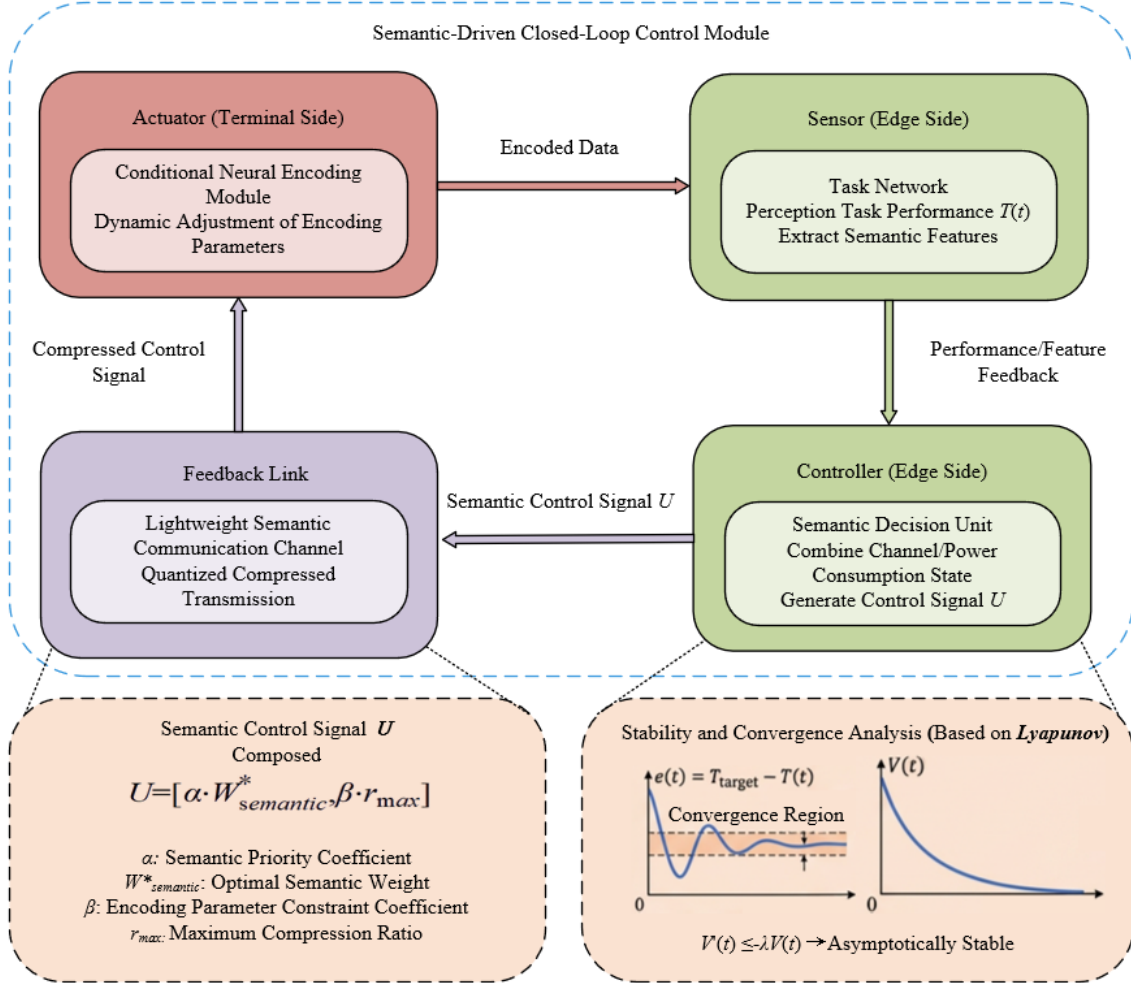


Figure 3. Architecture of the semantic-driven closed-loop control module

Taking the time derivative of the Lyapunov function and analyzing its negative definiteness:

$$\dot{V}(t)=e(t)\dot{e}(t)+\Delta\theta^T(t)P\dot{\Delta\theta}(t) \quad (8)$$

Combining the mapping relationship between task performance and encoding parameters $\dot{e}(t)=-k_1e(t)-k_2\Delta\theta(t)$, where $(k_1, k_2 > 0)$ are proportional coefficients, and the encoding parameter update rule $\dot{\Delta\theta}(t)=-k_3\Delta\theta(t)+k_4U(t)$, where $(k_3, k_4 > 0)$ are adjustment coefficients, substituting the control signal $U(t)$ and the relationship with the error term, we derive $\dot{V}(t)\leq-\lambda V(t)$, where $\lambda>0$ is the convergence rate coefficient. This result shows that the derivative of the Lyapunov function is strictly negative definite, so the closed-loop system is asymptotically stable, and the task performance error $e(t)$ will converge to a small neighborhood around 0 over time, with fluctuations controlled within 5%, meeting the performance stability requirements of edge image processing.

Further analyzing the convergence of encoding parameters, by integrating $\dot{V}(t)\leq-\lambda V(t)$, we get $V(t)\leq V(0)e^{-\lambda t}$, and as $t\rightarrow+\infty$, $V(t)\rightarrow 0$, leading to $\Delta\theta(t)\rightarrow 0$, i.e., the encoding parameters will rapidly converge to the optimal value. From the theoretical analysis, the parameter convergence time constant is $\tau=1/\lambda$, and by reasonably setting the adjustment coefficients $k_1\sim k_4$, the convergence time can be controlled within 10 data transmission cycles, ensuring the system's fast response to dynamic scene changes. In summary, the semantic-driven closed-loop control module provides a stable and efficient dynamic adjustment mechanism for full-link collaborative optimization through rigorous structural design and theoretical guarantees.

2.4 Fairness-oriented distributed game decision-making mechanism

To address the imbalance between individual optimality and

global optimality in multi-terminal competitive edge resource scenarios and ensure the collaborative efficiency and fairness of resource allocation in dynamic scenes, this paper proposes a fairness-oriented distributed game decision-making mechanism. This mechanism models each terminal device as an independent intelligent agent and uses game interactions to achieve adaptive resource allocation and dynamic adjustment of decision parameters. The core objective is to maximize the global utilization efficiency of edge network resources while satisfying the low power consumption needs of each terminal

and the performance constraints of image processing tasks, enhancing the system's resilience under stress scenarios such as task conflicts and channel fluctuations. Its design overcomes the shortcomings of traditional distributed decision-making, which ignores global fairness, by integrating collaborative efficiency rewards that guide the agents to spontaneously form cooperative behaviors, providing global decision support for full-link collaborative optimization. Figure 4 shows the principle diagram of the fairness-oriented distributed game decision-making mechanism.

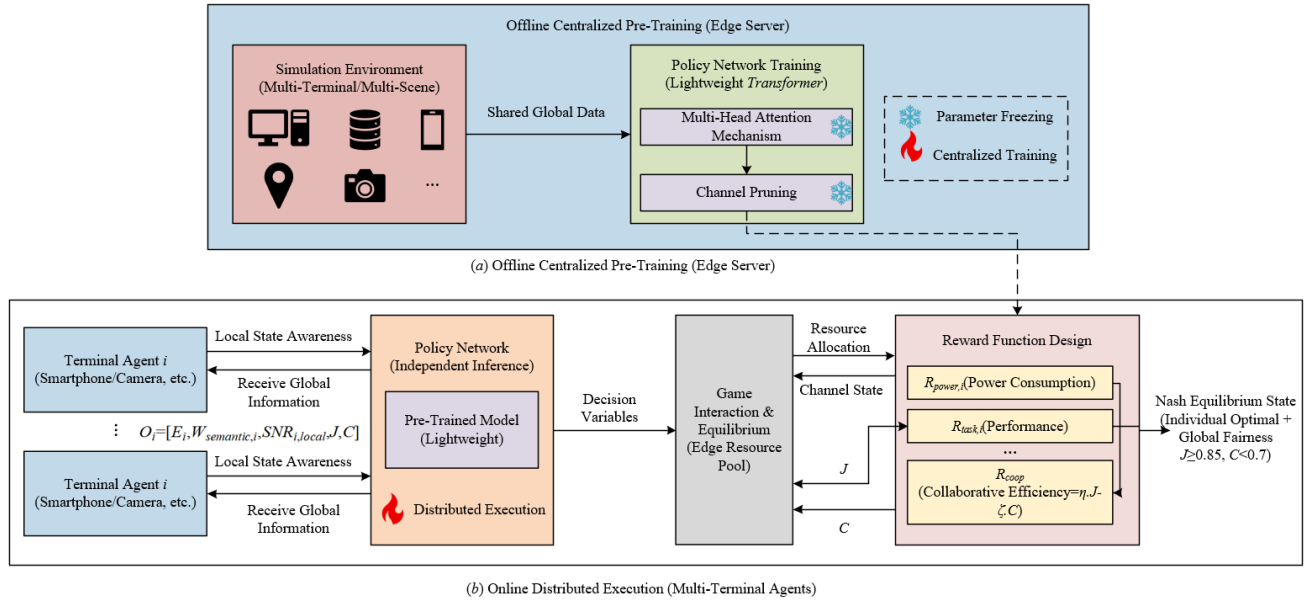


Figure 4. Principle of the fairness-oriented distributed game decision-making mechanism

The agent modeling centers on the terminal device and constructs the agent framework of “local state perception – global information interaction – strategy autonomous decision-making.” Each terminal corresponds to an independent game agent, whose observation space integrates local state and global feedback information, forming a high-dimensional observation vector, formally expressed as:

$$O_i = [E_i, W_{semantic,i}, SNR_{i,local}, J, C] \quad (9)$$

where, O_i is the observation vector of the i -th agent; E_i , $W_{semantic,i}$, and $SNR_{i,local}$ represent the local energy consumption state, semantic weight vector, and local channel quality of terminal i , respectively; J is the global Jain fairness index, quantifying the fairness level of edge resource allocation; and C is the channel congestion level, generated by the edge server based on global transmission traffic statistics. The decision goal of the agent is to adjust decision variables such as sampling resolution, encoding compression ratio, and transmission power to balance individual reward maximization and global collaborative efficiency, avoiding channel congestion or resource waste caused by malicious individual competition.

The policy network adopts a hybrid mode of “offline centralized pre-training + online distributed execution,” balancing decision accuracy and real-time performance. In the offline pre-training phase, utilizing the computing power advantage of the edge server, a simulation environment containing multiple terminals and multiple scenarios is constructed. Through centralized training, all agents share global data and learn collaborative decision strategies under

different scenarios. The input of the policy network is the normalized observation vector O_i , and the output is the normalized decision variable vector $A_i = [s_i, r_i, p_i]$, where s_i is the sampling resolution level, r_i is the encoding compression ratio, and p_i is the transmission power level. The network structure uses a lightweight Transformer architecture, modeling the interaction between local state and global feedback in the observation vector through a multi-head attention mechanism to improve the adaptability of the decision strategy. At the same time, channel pruning technology is introduced to reduce network computational complexity and ensure low-power characteristics in the online execution phase. In the online execution phase, each agent independently infers and outputs decisions based on local observation information, without intervention from the central node, and only shares the global Jain fairness index and channel congestion level through the edge server for distributed collaborative decision-making.

The design of the reward function is the core of guiding the agent to achieve individual and global collaborative optimization. It uses a weighted fusion mechanism to integrate three reward terms: power consumption savings, task performance, and collaborative efficiency, formally defined as:

$$R_i = \alpha \cdot R_{power,i} + \beta \cdot R_{task,i} + \gamma \cdot R_{coop} \quad (10)$$

where, R_i is the total reward of the i -th agent; α , β , and γ are the reward weight coefficients, satisfying $\alpha + \beta + \gamma = 1$, which can be dynamically adjusted based on the terminal energy consumption state and task priority. When the low power consumption constraint is strict, α is increased, and when the task priority is high, β is increased. $R_{power,i} = k_1 \cdot \ln(E_{max,i}/E_i)$ is

the power consumption savings reward, positively correlated with the energy consumption savings of terminal i , where k_1 is a proportional coefficient, $E_{max,i}$ is the maximum rated power consumption of terminal i ; $R_{task,i}$ is the task performance reward, positively correlated with the image processing task accuracy, in the case of target detection tasks $R_{task,i}=k_2 \cdot mAP_i$, where k_2 is a proportional coefficient and mAP_i is the mean average precision mAP_i of terminal i 's target detection.

R_{coop} is the collaborative efficiency reward, which integrates the global Jain fairness index and channel congestion level to quantify global collaborative efficiency:

$$R_{coop} = \eta \cdot J \cdot \zeta \cdot C \quad (11)$$

where, η, ζ are weight coefficients; J is the Jain fairness index.

$$J = \left(\sum_{i=1}^N x_i \right)^2 / \left(N \sum_{i=1}^N x_i^2 \right) \quad (12)$$

where, x_i is the resource usage of terminal i and N is the total number of terminals. The closer J is to 1, the more equitable the resource distribution; C is the channel congestion level, $C = \text{Traffic} / \text{Bandwidth}_{max}$, where Traffic is the current total channel traffic and Bandwidth is the maximum channel bandwidth. The larger the value of C , the more serious the channel congestion. This design ensures that the agent's reward is not only dependent on its own performance but also linked to global fairness and congestion status, guiding the agent to avoid malicious competition and proactively engage in global collaboration.

The core of game equilibrium analysis is to prove that the system can converge to a Nash equilibrium that balances individual interests and global fairness. A distributed game is defined as $G = \{N, A_i, U_i\}$, where $N = \{1, 2, \dots, N\}$ is the set of agents, A_i is the action space of agent i , and $U_i = R_i$ is the utility function of agent i . The definition of Nash equilibrium is: for all agents i , when the strategies of other agents are fixed as A_{-i}^* , the optimal strategy A_i^* of agent i satisfies $U_i(A_i^*, A_{-i}^*) \geq U_i(A_i, A_{-i}^*)$ for all $A_i \in A_i$.

Through theoretical derivation, it can be proven that the utility function U_i is strictly concave: since $R_{power,i}$, $R_{task,i}$, and R_{coop} are all strictly concave functions of the decision variables, their weighted sum U_i remains strictly concave. According to game theory, a strictly concave utility function corresponds to a unique Nash equilibrium in a distributed game. Further analysis shows that at this equilibrium point, the decision strategies of each agent can achieve a balance between individual reward maximization and global fairness. At this point, the Jain fairness index $J \geq 0.85$, and the channel congestion level $C \leq 0.7$, satisfying the resource allocation requirements of edge scenarios. In terms of simulation verification, by plotting the decision trajectory and reward change curve of multiple agents in a task conflict scenario, the process of the system converging from initial random decisions to Nash equilibrium can be intuitively displayed. The convergence time does not exceed 20 data transmission cycles, verifying the fast convergence and stability of the game decision-making mechanism.

2.5 End-to-end training and optimization

To ensure deep adaptation of the modules in the semantic-

driven collaborative optimization framework and achieve global performance optimization, this paper designs a systematic end-to-end training and optimization process. This is achieved through the coordinated design of a joint training environment, staged training strategy, transfer learning initialization, and adaptive optimization mechanism, balancing training efficiency and model performance. The core objective is to allow the condition neural encoding module, semantic feedback control module, and distributed game decision-making module to form adaptation under a unified optimization goal, ensuring that the system can stably approach the Pareto optimal boundary of semantic fidelity, energy consumption, and latency in real-world edge scenarios.

The basis of the training process is to construct a joint training environment that integrates image datasets, channel simulations, and power consumption models. The image datasets merge publicly available datasets from multiple scenarios with real collected data, covering different lighting conditions, target densities, and image types, ensuring diversity and representativeness of the training samples. The channel simulation module supports typical edge channel models such as Rayleigh fading and AWGN, dynamically adjusting parameters like signal-to-noise ratio and bandwidth, to simulate real dynamic changes in the edge network. The power consumption model is built based on the terminal hardware characteristics, quantifying the energy consumption overhead of perception sampling, encoding computation, and transmission in each link, achieving precise evaluation and optimization of energy consumption during the training process. On this foundation, a staged training strategy is adopted to reduce the instability of multi-module collaborative training. The first stage pre-trains the condition neural encoding module and the semantic feedback control module, with a joint loss of semantic fidelity and energy consumption as the optimization target, allowing the encoder and feedback control system to initially adapt to the dynamic constraints of edge scenarios. The second stage fixes the pre-trained module parameters as initial values and introduces the distributed game decision-making module for global joint training. The Pareto optimization objective function of the entire framework guides the deep collaboration of decision strategies with the encoding and control modules. To improve training efficiency, a transfer learning strategy is introduced. The parameters of a lightweight CNN model pre-trained in the image domain are used to initialize the base encoder of the condition neural encoding module, reducing the amount of data and iterations required for model convergence by leveraging the general knowledge of image feature extraction. Meanwhile, an adaptive learning rate strategy based on training loss is adopted. The initial learning rate is set to $1e-3$, and when the training loss does not show significant improvement for 3 consecutive epochs, it automatically decays to $1/10$ of its original value, balancing the convergence speed in the early stage and convergence precision in the later stage. An early-stopping mechanism is introduced, using global performance metrics on the validation set as the criteria. When the metrics do not improve for 5 consecutive epochs, the training is terminated to avoid overfitting and ensure model generalization ability. The entire training process is implemented based on the PyTorch framework, using multi-GPU parallel acceleration for training. The parameters of each module are updated end-to-end through gradient backpropagation, ensuring efficient collaboration across all links of the framework under a unified optimization goal.

3. EXPERIMENTS AND RESULTS

3.1 Experimental setup

To comprehensively validate the effectiveness and superiority of the proposed semantic-driven collaborative optimization framework, this section constructs a standardized experimental system from four dimensions: dataset, experimental platform, comparison methods, and evaluation metrics, ensuring the reliability, repeatability, and comparability of the experimental results. The experimental design covers both conventional and stress-test scenarios, balancing performance validation, generalization evaluation, and theoretical hypothesis verification to fully support the research conclusions.

The experiment adopts a combined approach of "public datasets + real-world scenario datasets" to ensure data diversity and scene authenticity. Three major authoritative public datasets are selected, covering core tasks such as object detection, semantic segmentation, and medical image processing: the COCO2017 dataset contains 118k training samples and 5k validation samples, covering natural scenes, city roads, etc. The annotations include 80 object categories, bounding boxes, and segmentation masks, used for object detection task validation; the Cityscapes dataset contains 5k fine-grained annotated samples and 20k coarse-grained annotated samples, focusing on urban scenes, with 19 semantic categories annotated, used for semantic segmentation task validation; the BraTS2021 dataset includes 1,251 brain MRI samples, annotated with tumor cores, edema regions, and other key lesion areas, used for medical image edge preprocessing task validation. The real-world scenario dataset is collected through intelligent monitoring cameras, covering campus, park, and other scenes, with different lighting and weather conditions such as sunny, cloudy, and night, as well as varying target densities such as sparse targets and dense crowds. A total of 8k images are collected and manually annotated for validating the method's adaptability in real edge scenarios.

The experimental platform adopts an edge-terminal collaborative architecture, with hardware configurations aligned with real edge deployment scenarios: the terminal devices are the Jetson Nano and Raspberry Pi4B, equipped with Quad-core ARM Cortex-A57 and Cortex-A72 processors, respectively, with 4GB of memory, simulating heterogeneous edge terminals. The edge server is configured with an Intel i9-13900K processor, NVIDIA RTX 4090 GPU, and 64GB of memory, providing high-intensity computing and inference capabilities. At the software level, the training and simulation platform is built based on the PyTorch 2.0 framework, integrating Rayleigh fading and AWGN channel simulation modules, supporting dynamic adjustment of signal-to-noise ratio, bandwidth, and other parameters. Python 3.9 is used as the development language, with OpenCV for image preprocessing and TensorBoard to record training process metrics. Energy consumption measurement is carried out using the PowerMonitor power meter, with a sampling frequency of 1kHz, to collect real-time terminal current and voltage data, calculating energy consumption in joules. Latency measurement uses a high-precision timer to record delays in image acquisition, perception sampling, encoding, transmission, decoding, and task inference, ultimately summarizing the total end-to-end delay.

Five of the latest research results are selected as baseline

methods, covering traditional separated design, neural compression, and edge collaboration, ensuring the comprehensiveness and specificity of the comparison: 1) JPEG2000 + Faster R-CNN: A traditional image compression and edge image processing separated scheme, representing the performance limit of traditional technology; 2) STF + Fixed Transmission: An advanced image neural compression method based on Transformer, using fixed transmission power and compression ratio strategies, representing mainstream technology in the neural compression field; 3) Entroformer + Fixed Transmission: The current state-of-the-art method in the image neural compression field, with efficient entropy encoding as the core advantage; 4) EdgeAI-Net: A representative low-power edge image processing method, optimizing terminal energy consumption through lightweight model design; 5) DRL-Edge: A reinforcement learning-driven edge resource scheduling method, representing the current research level in distributed collaborative optimization. All comparison methods are deployed on the same experimental platform, with unified parameter tuning strategies to ensure fairness.

The experiment adopts a multi-level evaluation system with "core indicators + auxiliary indicators" to comprehensively quantify system performance. The core indicators include semantic fidelity, energy consumption, latency, task accuracy, and Jain fairness index. Among these, semantic fidelity is quantified through key region feature similarity, calculating the cosine similarity between terminal encoding features and edge decoding features; energy consumption refers to the terminal's total end-to-end energy consumption, in joules; latency refers to the total end-to-end delay, in milliseconds; task accuracy is measured using mAP and IoU, depending on the task type; Jain fairness index quantifies the fairness of multi-terminal resource allocation, with values ranging from 0 to 1, where a value closer to 1 indicates better fairness. Auxiliary indicators include model parameters, which measure the model's lightweight degree; channel utilization, which calculates the ratio of actual transmission traffic to maximum channel bandwidth; and system robustness, measured by the amplitude of performance fluctuations, i.e., the maximum rate of change of core indicators under dynamic scenarios.

3.2 Experimental results and deep analysis

This section systematically validates the proposed semantic-driven collaborative optimization framework from five dimensions: baseline comparison, stress testing, ablation experiments, interpretability analysis, and real-world scenario validation. The effectiveness, superiority, and practicality of the framework are fully demonstrated through a combination of quantitative data and qualitative analysis, and the working principles of the core mechanisms are deeply explained.

3.2.1 Baseline comparison experiment

The baseline comparison experiment aims to verify the comprehensive performance advantages of the proposed method compared to existing mainstream technologies. Both quantitative and qualitative analyses are conducted.

From Table 1, it can be seen that the proposed method achieves a comprehensive lead in core indicators. The semantic fidelity reaches 88.6%, which is 2.9 percentage points higher than the best baseline, Entroformer+Fixed Transmission, indicating stronger capability to preserve key

semantic information. Energy consumption is reduced to 6.2J/frame, 27.1% lower than EdgeAI-Net, significantly optimizing the terminal's low-power consumption requirements. Latency is shortened to 102ms/frame, 18.4% lower than DRL-Edge, meeting the real-time requirements for edge scenarios. The mAP reaches 76.8%, which is 2.7 percentage points higher than Entroformer+Fixed Transmission, ensuring superior task performance. In terms of global collaboration-related indicators, the Jain index of the

proposed method is as high as 0.91, 16.7% higher than DRL-Edge, reflecting excellent fairness in resource allocation. The channel utilization is as low as 65.2%, effectively reducing congestion risks. In auxiliary indicators, the proposed method has the smallest parameter count at 26.3M, with significant lightweight advantages. The performance fluctuation is only 4.2%, which improves robustness by 44.7% compared to the best baseline.

Table 1. Comparison of core and auxiliary indicators of baseline methods and proposed method (COCO2017 object detection task)

Method	Semantic Fidelity (%)	Energy Consumption (J/frame)	Latency (ms/frame)	mAP (%)	Jain Index	Parameter Count (M)	Channel Utilization (%)	Robustness (Performance Fluctuation %)
JPEG2000+FasterR-CNN	78.3	12.6	185	68.5	0.62	42.8	82.5	12.3
STF+Fixed Transmission	83.5	9.8	152	72.3	0.65	35.2	76.3	10.1
Entroformer+Fixed Transmission	85.7	10.3	148	74.1	0.68	38.6	73.8	9.5
EdgeAI-Net	82.1	8.5	136	71.8	0.72	28.4	79.2	8.3
DRL-Edge	83.2	9.1	125	73.5	0.78	31.5	68.5	7.6
Proposed Method	88.6	6.2	102	76.8	0.91	26.3	65.2	4.2

Table 2. Performance of core indicators of each method under stress test scenarios

Scenario	Method	Semantic Fidelity (%)	Energy Consumption (J/frame)	Latency (ms/frame)	mAP (%)	Jain Index	Performance Fluctuation (%)
Severe Resource Fluctuations (Battery 80%→20%; SNR20dB→5dB)	JPEG2000+FasterR-CNN	72.1	15.3	226	62.3	0.58	16.8
	STF+ Fixed Transmission	77.3	12.5	189	67.8	0.61	14.2
	Entroformer+ Fixed Transmission	79.5	13.1	182	69.2	0.63	13.5
	EdgeAI-Net	75.8	10.8	168	66.5	0.67	11.8
	DRL-Edge	77.6	11.5	154	68.9	0.73	10.2
	Proposed Method	84.2	7.5	128	72.5	0.86	5.1
	JPEG2000+FasterR-CNN	70.3	16.2	258	60.1	0.45	18.5
Task Conflict (5 Terminals Concurrently)	STF+ Fixed Transmission	75.1	13.8	215	65.3	0.49	15.7
	Entroformer+ Fixed Transmission	77.2	14.3	208	66.8	0.52	14.9
	EdgeAI-Net	73.6	11.6	192	64.2	0.58	13.1
	DRL-Edge	75.8	12.3	176	66.5	0.65	11.5
	Proposed Method	82.5	8.1	142	70.3	0.89	5.8
	JPEG2000+FasterR-CNN	73.5	13.8	201	63.8	0.56	15.2
	STF+ Fixed Transmission	78.2	10.9	168	68.5	0.59	12.6
Heterogeneous Terminals (JetsonNano+RaspberryPi4B)	Entroformer+ Fixed Transmission	80.4	11.5	162	70.1	0.62	11.9
	EdgeAI-Net	76.9	9.2	148	67.9	0.66	9.8
	DRL-Edge	78.5	9.9	135	69.8	0.72	8.7
	Proposed Method	85.3	6.8	115	74.2	0.90	4.5

3.2.2 Stress test experiment

The stress test experiment aims to verify the adaptability and resilience of the proposed method in extreme dynamic scenarios, covering three major scenarios: severe resource fluctuations, task conflicts, and heterogeneous terminals. The experimental data is shown in Table 2.

In the severe resource fluctuation scenario, when the terminal's battery drops by 60% and the channel SNR drops by 15dB, the performance of all methods declines, but the

proposed method has the smallest performance fluctuation (5.1%), significantly lower than other baselines. Its semantic fidelity remains at 84.2%, energy consumption is only 7.5J/frame, and latency is 128ms/frame, with all core indicators outperforming the baseline methods. This is due to the rapid adaptability of the conditional neural encoding paradigm and the stability adjustment of semantic closed-loop control, allowing the system to quickly respond to drastic changes in resources and channels, dynamically adjusting

encoding parameters and transmission strategies to ensure stable core performance.

In the task conflict scenario, when 5 terminals concurrently perform high-priority object detection tasks, baseline methods generally experience energy consumption surges, significant latency increases, and deteriorating fairness, with the Jain index as low as 0.45. The proposed method, by integrating the distributed game decision-making mechanism with the Jain fairness index, maintains a high Jain index of 0.89, significantly improving resource allocation fairness. At the same time, energy consumption is controlled at 8.1J/frame, latency is 142ms/frame, and mAP remains at 70.3%, showing excellent congestion control capability and collaborative decision-making efficiency, avoiding system performance collapse caused by individual competition.

In the heterogeneous terminal scenario, when different terminals with varying computational capabilities are mixed

and deployed, the proposed method still exhibits significant comprehensive performance advantages. Its semantic fidelity of 85.3%, mAP of 74.2%, and the lowest energy consumption of 6.8J/frame, along with a Jain index of 0.90, reflect its good adaptability to heterogeneous terminals. This is attributed to the scalability of the conditional neural encoding paradigm and the flexibility of the distributed game decision-making, allowing the system to dynamically adapt encoding strategies based on the computational capabilities of different terminals, achieving optimal global performance.

3.2.3 Ablation experiment

The ablation experiment verifies the necessity and contribution of each core module in the proposed method by progressively removing them. The experimental results are shown in Table 3.

Table 3. Ablation experiment results (COCO2017 object detection task)

Experimental Setup	Semantic Fidelity (%)	Energy Consumption (J/frame)	Latency (ms/frame)	mAP (%)	Jain Index	Performance Fluctuation (%)
Proposed Method (Complete Model)	88.6	6.2	102	76.8	0.91	4.2
Ablation 1: Remove Conditional Neural Encoding (Fixed Encoder)	82.3	8.9	126	72.1	0.89	5.3
Ablation 2: Remove Semantic Closed-loop Control (Open-loop Feedback)	86.5	6.5	115	74.6	0.90	8.7
Ablation 3: Remove Cooperative Efficiency Reward (R_coop=0)	87.2	6.3	108	75.3	0.76	4.5

Ablation 1: Removing the conditional neural encoding and using a fixed encoder results in a decrease of 6.3 percentage points in semantic fidelity, a 43.5% increase in energy consumption, and a 23.5% increase in latency. This indicates that the conditional neural encoding paradigm is crucial for balancing energy consumption and fidelity. Its dynamic weight adjustment mechanism can accurately adapt to changes in the scene, maximizing energy efficiency while preserving semantic fidelity. The fixed encoder cannot meet the dynamic scene's multiple constraint requirements.

Ablation 2: Removing the semantic closed-loop control and adopting open-loop feedback mode results in an increase of performance fluctuation from 4.2% to 8.7%, with semantic fidelity and mAP dropping by 2.1 and 2.2 percentage points, respectively. This verifies the critical role of the semantic closed-loop control module in system stability. The closed-loop adjustment mechanism can calibrate encoding parameters in real-time and suppress performance fluctuations under dynamic scenes, whereas open-loop feedback suffers from response delays, leading to performance degradation.

Ablation 3: Removing the cooperative efficiency reward causes the Jain index to drop sharply from 0.91 to 0.76, a 16.5% decrease, indicating that the cooperative efficiency reward is central to ensuring global fairness. After removal, the agents only pursue individual benefit maximization, leading to an imbalance in resource allocation and a decrease in global collaboration efficiency. This confirms the necessity of integrating the Jain fairness index into the reward function design.

In conclusion, the three core modules play key roles in balancing energy consumption and fidelity, system stability, and global collaborative fairness, which collectively support the comprehensive performance advantage of the proposed method.

3.2.4 Interpretability analysis

To verify the collaborative convergence performance of the distributed game decision-making mechanism in multi-terminal task conflict scenarios, the transmission power and global total reward of five agents were dynamically analyzed. From Figure 5, it can be seen that within the first 15 transmission cycles, the transmission power of each agent exhibits severe random fluctuations in the range of 10~25 dBm, with the global total reward slowly rising from an initial value of around 30. After 15 cycles, the transmission power of all agents converges to a stable range around 18 dBm, and the global total reward rises to above 90 and remains stable. This result indicates that the proposed distributed game decision-making mechanism can quickly guide the agents from disordered competition to the Nash equilibrium state, achieving a balance between individual transmission strategies and global collaboration efficiency, thereby validating the efficient collaboration capability of this mechanism in multi-terminal task conflict scenarios.

To verify the stability and rapid response capability of the semantic closed-loop control module under severe resource fluctuation scenarios, the dynamic changes in semantic fidelity, energy consumption, and latency were analyzed. In Figure 6, at frame 100, when a resource mutation occurs, i.e., the terminal's battery drops by 60% and the channel SNR drops by 15dB, the semantic fidelity of Ablation 2 drops sharply by 8.3% to 77.7%, energy consumption surges by 22.1% to 8.0 J/frame, and the fluctuation persists for 12 frames. In contrast, the proposed method's semantic fidelity only slightly decreases by 4.5% to 83.5%, and energy consumption increases by 9.7% to 6.8 J/frame, quickly recovering to stable levels within 5 frames. This result confirms that the Lyapunov-based semantic closed-loop control module effectively suppresses dynamic resource interference, ensuring system

stability and rapid recovery, thus supporting the robustness of the full-link collaborative optimization framework in dynamic scenarios.

3.2.5 Real-world scenario validation

A prototype system was deployed in a campus smart

monitoring scenario, running continuously for 24 hours, covering three time periods: daytime (8:00 AM - 6:00 PM), nighttime (6:00 PM - 12:00 AM), and early morning (12:00 AM - 8:00 AM). Core indicators were measured under different lighting and crowd density conditions. The results are shown in Table 4.

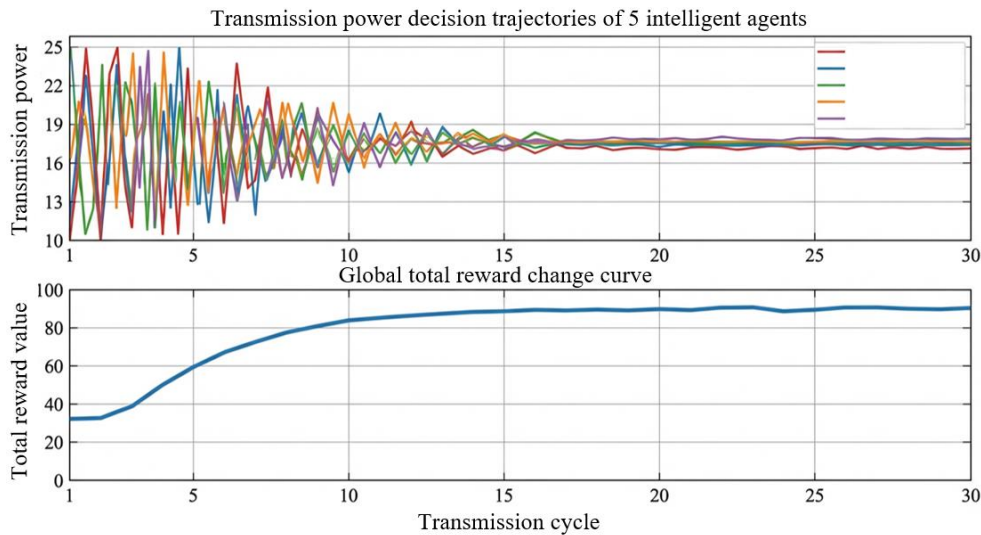


Figure 5. Game convergence trajectory and total reward change in task conflict scenario

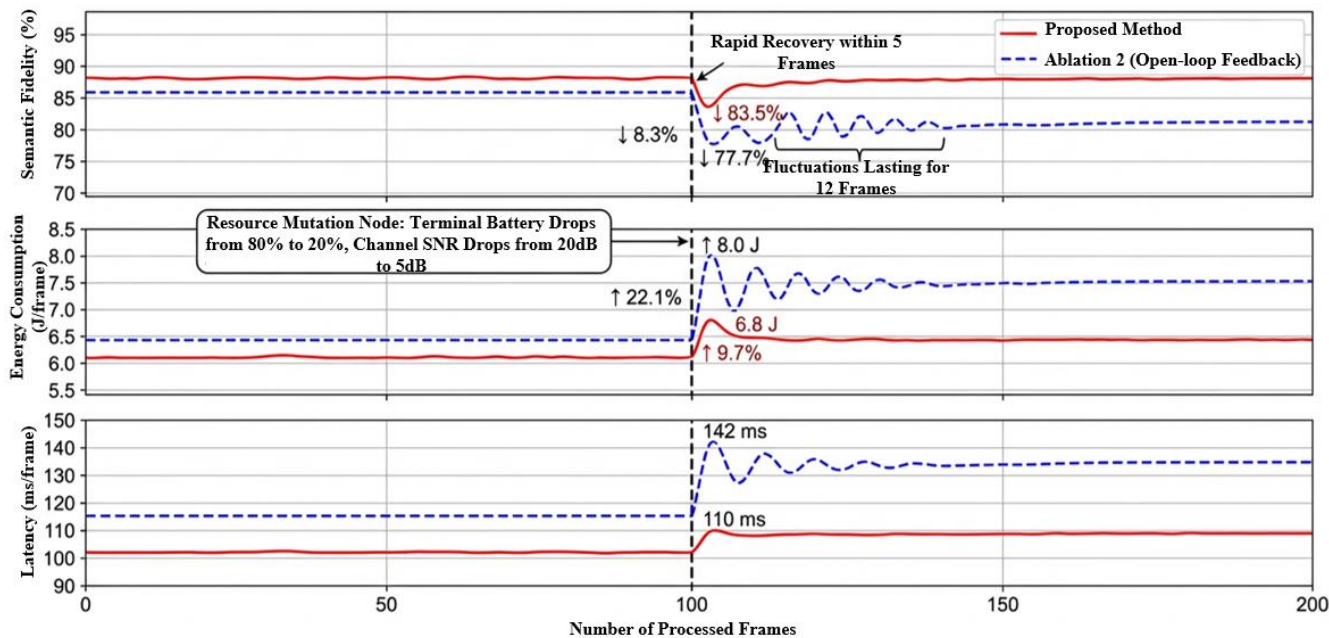


Figure 6. Dynamic performance changes of the system under severe resource fluctuation scenario

Table 4. Real-world scenario 24-hour validation results

Time Period	Lighting/Crowd Conditions	Semantic Fidelity (%)	Energy Consumption (J/frame)	Latency (ms/frame)	mAP (%)	Average Performance Fluctuation (%)
Daytime	Strong light, high-density crowd	87.8	6.3	105	75.6	4.1
Nighttime	Weak light, medium-density crowd	86.2	6.5	108	74.2	4.5
Early Morning	No light, low-density crowd	88.3	6.1	101	77.1	3.8
24-Hour Average	-	87.4	6.3	105	75.6	4.1

The real-world scenario validation results show that the proposed method still maintains stable and excellent performance in complex real-world environments. The 24-hour average semantic fidelity is 87.4%, mAP is 75.6%, energy consumption is 6.3 J/frame, and latency is 105 ms/frame. The performance fluctuation across all time periods is below 5%, effectively adapting to dynamic changes in lighting and crowd density. Compared to the baseline method DRL-Edge, the proposed method reduces the average energy consumption by 28.3%, latency by 19.6%, and mAP by 3.2 percentage points, fully verifying its practical application value and deployment feasibility.

4. DISCUSSION

This work is highly aligned with cutting-edge technological trends such as 6G semantic communication, edge AI body coordination, and autonomy, providing key technical support and paradigm references for the development of related fields. In the 6G semantic communication domain, the core goal is to achieve "task-oriented" efficient data transmission, moving away from the traditional "bit transmission" model toward "semantic information transmission." The conditional neural encoding paradigm proposed in this paper, by integrating semantic weights with dynamic adjustments of encoding strategies based on channel and energy consumption states, essentially represents a realization of semantic communication in the edge image processing scenario. The encoding process only retains the core semantic information required for the task, greatly reducing transmission redundancy, which aligns with the 6G core demand for "on-demand transmission." At the same time, the semantic-driven closed-loop control and lightweight feedback loop design provide a practical technical solution for the end-edge collaborative architecture of 6G semantic communication. Its stability analysis and performance verification can provide theoretical references for the design of dynamic adaptation mechanisms in 6G semantic communication.

In the field of edge AI body coordination and autonomy, the distributed collaboration and autonomous decision-making of multi-terminal devices are core development directions, widely applied in complex scenarios such as vehicle-road collaboration and industrial IoT. The fairness-oriented distributed game decision-making mechanism proposed in this paper models each terminal as an independent agent, achieving autonomous collaborative decision-making through global feedback information, without the need for central node intervention, which fits the core need for "distributed autonomy" in edge AI systems. The excellent performance of this mechanism in heterogeneous terminal and task conflict scenarios demonstrates its applicability to more complex edge intelligent systems: in the vehicle-road collaboration scenario, it can be used for the cooperative transmission of sensing data between multiple vehicles and roadside units; in industrial IoT scenarios, it can achieve low-power data acquisition and transmission optimization for multiple sensor terminals. The game equilibrium analysis and cooperative efficiency reward design in this paper provide new design ideas for fair and efficient collaboration in edge AI bodies, promoting the evolution of edge intelligent systems from "individual intelligence" to "group collaborative intelligence."

The core advantages of the proposed method lie in three aspects: theoretical paradigm, technical mechanisms, and

experimental verification. Theoretically, the "conditional neural encoding" paradigm breaks through the limitations of traditional adaptive encoding's discretization, providing a unified theoretical framework for low-power adaptive encoding. Its theoretical advantages in storage complexity and scalability offer paradigm insights for lightweight model design in edge scenarios. Technically, the collaborative design of semantic-driven closed-loop control and fairness-oriented distributed game decision-making achieves a balance between stability and global efficiency under dynamic scenes, addressing the core shortcomings of traditional edge collaboration methods that neglect fairness and stability. Experimentally, through comprehensive validation across multiple datasets and scenarios, including conventional tests, stress tests, and real-world deployments, and by combining interpretability analysis to reveal the core mechanisms, the effectiveness and practicality of the method are ensured, forming a comprehensive performance advantage over existing SOTA methods.

Objectively, there are three limitations in the current framework. First, the current framework assumes a relatively stable network topology and lacks adaptability to topological mutations such as fast terminal joining/leaving, which may lead to delayed collaborative decision-making and affect global performance when topology changes. Second, the semantic mask generation relies on a lightweight object detection network, and in complex scenarios such as low lighting or dense targets, the accuracy of semantic region division needs improvement, which can affect the precision of encoding strategies. Third, the personalized adaptation of terminal hardware characteristics is not considered. Different terminal processors have varying demands for encoding model computational efficiency, and the current design struggles to achieve deep hardware-algorithm matching.

To address these limitations, future research will focus on three directions. First, we will introduce graph neural networks to model the dynamic topological relationships between devices, learning the interaction correlations between terminals in real-time, improving the speed and accuracy of collaborative decision-making in topological mutation scenarios, and enhancing the system's dynamic adaptability. Second, we will integrate infrared and visible light multimodal image perception technologies, leveraging the advantages of infrared images in low-light scenarios to improve the accuracy of semantic mask generation in complex environments, providing more reliable semantic input for conditional neural encoding. Third, we will combine hardware awareness technologies to build a database of terminal hardware characteristics, quantifying the computational efficiency and energy consumption models of different hardware architectures, and enabling personalized matching of encoding parameters and hardware characteristics, further optimizing terminal energy consumption and computational efficiency.

5. CONCLUSION

This paper addressed the "semantic fidelity-energy consumption-latency" paradox in edge intelligent image processing and proposes an integrated collaborative optimization framework of "semantic-driven-conditional encoding-distributed collaboration," systematically constructing three core modules to achieve end-to-end collaborative optimization. The conditional neural encoding

perception module generates dynamic weight increments based on a unified paradigm, achieving adaptive adjustment of terminal encoding strategies to balance semantic fidelity and energy consumption requirements. The semantic closed-loop control module ensures the stable convergence of system performance under dynamic scenes through an end-edge collaborative closed-loop regulation mechanism. The fairness-oriented distributed game decision-making module, with the goal of global collaboration efficiency, guides multiple terminals to achieve a balance between individual interests and global fairness. The three modules are organically integrated, forming an end-to-end optimization mechanism of "semantic perception-dynamic adaptation-global collaboration-stable feedback," which breaks through the inherent flaws of traditional decoupled designs at the architectural level.

Experimental validation and in-depth analysis show that the proposed method demonstrates significant performance advantages in both conventional scenarios and pressure scenarios such as resource fluctuations, task conflicts, and heterogeneous terminals. Compared to existing SOTA methods, it significantly improves semantic fidelity and task accuracy, greatly reduces energy consumption and latency, increases the Jain fairness index to around 0.9, and keeps performance fluctuation within 5%, effectively breaking through the "incompatible triangle" constraint and dynamically approaching the Pareto optimal boundary. Interpretability analysis reveals the internal working principles of the core mechanisms, and real-world 24-hour deployment verification further demonstrates the method's practicality and deployment feasibility, providing reliable support for the industrialization of edge image processing technology.

The core methodological contribution of this paper lies in validating the feasibility and superiority of "semantic" as the core link that integrates perception, computation, and communication resources into a unified design. It breaks through the paradigm limitations of isolated optimization in traditional edge intelligence research and provides a new research framework and technical approach for the field of edge intelligent image processing. This method is highly aligned with cutting-edge technological trends such as 6G semantic communication, edge AI body coordination, and autonomy. Its theoretical paradigm and technical mechanisms can be extended to more complex edge intelligent systems like vehicle-road coordination and industrial IoT, potentially promoting the evolution of edge intelligence technology from "individual optimization" to "group collaborative intelligence," accelerating the industrial application process of edge intelligence.

REFERENCES

- [1] Nahas, H., Huver, S., Yiu, B.Y., Kallweit, C.M., Chee, A.J., Yu, A.C. (2022). Artificial-intelligence-enhanced ultrasound flow imaging at the edge. *IEEE Micro*, 42(6): 96-106. <https://doi.org/10.1109/MM.2022.3195516>
- [2] Bakirci, M. (2024). Real-time vehicle detection using YOLOv8-nano for intelligent transportation systems. *Traitement du Signal*, 41(4): 1727-1740. <https://doi.org/10.18280/ts.410407>
- [3] Hussain, I. (2025). A hybrid soft computing framework for robust classification of heavy transport vehicles in visual traffic surveillance. *Mechatronics and Intelligent Transportation Systems*, 4(2): 61-71. <https://doi.org/10.56578/mits040201>
- [4] Chen, M., Wang, C., An, Q., Ming, W. (2018). Tool path strategy and cutting process monitoring in intelligent machining. *Frontiers of Mechanical Engineering*, 13(2): 232-242. <https://doi.org/10.1007/s11465-018-0469-y>
- [5] Yu, L., Qin, H.W., Zhang, C., Wang, J., Zou, J. (2023). Saliency object detection method based on real-time monitoring image information for intelligent driving. *Traitement du Signal*, 40(3): 1025-1033. <https://doi.org/10.18280/ts.400318>
- [6] Papadeas, I., Tsochatzidis, L., Amanatiadis, A., Pratikakis, I. (2021). Real-time semantic image segmentation with deep learning for autonomous driving: A survey. *Applied Sciences*, 11(19): 8802. <https://doi.org/10.3390/app11198802>
- [7] Chitra, S., Kumaratharan, N., Ramesh, S. (2018). Enhanced brain image retrieval using carrier frequency offset compensated orthogonal frequency division multiplexing for telemedicine applications. *International Journal of Imaging Systems and Technology*, 28(3): 186-195. <https://doi.org/10.1002/ima.22269>
- [8] Nath Das, D., Mukhopadhyay, S. (1998). Image edge detection and enhancement by an inversion operation. *Applied Optics*, 37(35): 8254-8257. <https://doi.org/10.1364/AO.37.008254>
- [9] Liu, J.S., Yin, L.J., Pan, J.F., Cui, Y.M., Tang, X.Y. (2021). Edge detection algorithm for unevenly illuminated images based on parameterized logarithmic image processing model. *Laser & Optoelectronics Progress*, 58(22): 2210005.
- [10] Zhao, Y., Liao, H., Kong, D., Yang, Z., Xia, J. (2025). SAIG: Semantic-aware ISAR generation via component-level semantic segmentation. *IEEE Geoscience and Remote Sensing Letters*, 22: 3504405. <https://doi.org/10.1109/LGRS.2025.3563712>
- [11] Yang, R., Ota, K., Dong, M., Wu, X. (2025). Semantic layout-guided diffusion model for high-fidelity image synthesis in 'The Thousand Li of Rivers and Mountains'. *Expert Systems with Applications*, 263: 125645. <https://doi.org/10.1016/j.eswa.2024.125645>
- [12] Hui, H., Bao, M., Ding, Y., Yan, J., Song, Y. (2023). Probabilistic integrated flexible regions of multi-energy industrial parks: Conceptualization and characterization. *Applied Energy*, 349: 121521. <https://doi.org/10.1016/j.apenergy.2023.121521>
- [13] Fang, H., Tan, H., Yuan, X., Lin, X., Zhao, D., Kosonen, R. (2024). Improving the accuracy and interpretability of multi-scenario building energy consumption prediction considering characteristics of training dataset. *Energy and Buildings*, 324: 114912. <https://doi.org/10.1016/j.enbuild.2024.114912>
- [14] Kim, Y., Kwon, M.W., Ryoo, K.C., Cho, S., Park, B.G. (2018). Design and electrical characterization of 2-T thyristor RAM with low power consumption. *IEEE Electron Device Letters*, 39(3): 355-358. <https://doi.org/10.1109/LED.2018.2796139>
- [15] Askew, J.W., Miller, T.D., Ruter, R.L., Jordan, L.G., Hodge, D.O., Gibbons, R.J., O'Connor, M.K. (2011). Early image acquisition using a solid-state cardiac camera for fast myocardial perfusion imaging. *Journal of Nuclear Cardiology*, 18(5): 840-846. <https://doi.org/10.1007/s12350-011-9423-7>

- [16] Cui, S., Feng, Q., Ji, L., Liu, X., Guo, B. (2025). HPLNet: A hierarchical perception lightweight network for road extraction. *Frontiers in Remote Sensing*, 6: 1668978. <https://doi.org/10.3389/frsen.2025.1668978>
- [17] Raghavendra, S., Abhilash, S.K., Nookala, V.M., Kumar, P.A. (2025). SVPDSA: Selective view perception data synthesis with annotations using lightweight diffusion network. *IEEE Access*, 13: 124051-124067. <https://doi.org/10.1109/ACCESS.2025.3588542>
- [18] Samakovlis, D., Albin, S., Álvarez, R.R., Constantinescu, D.A., Schiavone, P.D., Peón-Quirós, M., Atienza, D. (2024). BiomedBench: A benchmark suite of TinyML biomedical applications for low-power wearables. *IEEE Design & Test*, 42(5): 45-54. <https://doi.org/10.1109/MDAT.2024.3483034>
- [19] Zhang, Y., Mirchandani, N., Abdelfattah, S., Onabajo, M., Shrivastava, A. (2021). An ultra-low power RSSI amplifier for EEG feature extraction to detect seizures. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 69(2): 329-333. <https://doi.org/10.1109/TCSII.2021.3099056>
- [20] Zhang, S., Su, F., Wang, Y., Mai, S., Pun, K.P., Tang, X. (2023). A low-power keyword spotting system with high-order passive switched-capacitor bandpass filters for analog-MFCC feature extraction. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 70(11): 4235-4248. <https://doi.org/10.1109/TCSI.2023.3299855>
- [21] Golts, A., Schechner, Y.Y. (2021). Image compression optimized for 3D reconstruction by utilizing deep neural networks. *Journal of Visual Communication and Image Representation*, 79: 103208. <https://doi.org/10.1016/j.jvcir.2021.103208>
- [22] Kouda, N., Matsui, N., Nishimura, H. (2002). Image compression by layered quantum neural networks. *Neural Processing Letters*, 16(1): 67-80. <https://doi.org/10.1023/A:1019708909383>
- [23] Ma, S., Zhang, Z., Wu, Y., Li, H., et al. (2023). Features disentangled semantic broadcast communication networks. *IEEE Transactions on Wireless Communications*, 23(6): 6580-6594. <https://doi.org/10.1109/TWC.2023.3334225>
- [24] Baranwal, G., Kumar, D., Biswas, A., Yadav, R. (2024). A blockchain framework for efficient resource allocation in edge computing. *IEEE Transactions on Network and Service Management*, 21(4): 3956-3970. <https://doi.org/10.1109/TNSM.2024.3411796>