telecom

MDPI

Article

Semantic-Aware Resource Allocation for Infrared Small-Target Detection in UAV Communication Systems

Weicheng Qiu, Jiujiu Chen *, Bangshu Xiong and Wei Li

Jiangxi Key Laboratory of Image Processing and Pattern Recognition, Nanchang Hangkong University, Nanchang 330063, China; 2504082500001@stu.nchu.edu.cn (W.Q.); xiongbs@nchu.edu.cn (B.X.); weili@nchu.edu.cn (W.L.)

* Correspondence: chenjiujiu@nchu.edu.cn

Abstract

In resource-constrained unmanned aerial vehicle (UAV) infrared image transmission, infrared small targets usually occupy only a small number of pixels, while conventional uniform resource allocation strategies fail to consider the semantic differences among image regions, resulting in inefficient resource utilization and loss of target-related information. To address this issue, this paper proposes a semantic-aware resource allocation method for infrared small-target detection. First, infrared images are divided into multiple grid cells, and the semantic importance of each cell is estimated based on preliminary detection results. Then, the cells are classified into different semantic levels, followed by differentiated bit allocation and region-wise reconstruction. Finally, a joint optimization problem of grid partitioning and resource allocation parameters is formulated and solved by the covariance matrix adaptation evolution strategy (CMA-ES) to optimize both detection performance and image reconstruction quality. Experimental results demonstrate that the proposed method achieves mIoU values of 0.6106 and 0.7445 at available bit rates of 12.5% and 50%, respectively, showing significant improvements over comparison methods. Moreover, CMA-ES obtains stable optimization results with fewer evaluations and provides a favorable balance between detection accuracy and reconstruction quality. These results indicate that the proposed method can effectively preserve target-related semantic information under limited communication resources, providing an effective solution for efficient UAV infrared image transmission and reliable small-target detection.

Keywords: UAV semantic communication; small target detection; semantic importance assessment; region-adaptive coding; task-oriented communication

Check for updates

Academic Editor: Barbara M. Masini

Received: 30 July 2026
Revised: 3 September 2026
Accepted: 4 September 2026
Published: 8 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

UAVs have become important platforms for wide-area visual sensing owing to their flexible deployment, high-resolution imaging capabilities, and ability to collect observations from top and oblique viewpoints. UAV-based remote sensing has been increasingly applied to environmental monitoring, infrastructure inspection, emergency response, and other large-area observation scenarios [1]. These applications generate large volumes of high-resolution imagery that must be efficiently represented and transmitted for subsequent analysis. Improving the efficiency of visual information representation and transmission is therefore important for reliable UAV-based perception.

Small-target detection in UAV imagery remains challenging because distant targets generally occupy only a few pixels, while their low-level spatial details can be progressively

weakened during repeated downsampling and feature extraction [2]. Infrared small-target detection presents additional difficulties because the targets are typically dim, low-contrast, and weakly structured, with limited texture, contour, and shape information. Their responses can also be easily confused with background clutter, sensor noise, and other local interference. Recent studies have therefore emphasized the joint preservation and enhancement of weak target cues in both the spatial and frequency domains [3]. Preserving small-target responses and fine-grained spatial information during image processing and transmission is consequently essential for maintaining reliable detection performance.

Conventional image transmission and compression methods commonly evaluate reconstruction performance using global distortion or perceptual-quality criteria. However, these image-level criteria do not explicitly account for the unequal information value of different spatial regions. In UAV infrared imagery, small targets occupy only a limited portion of the image, whereas background content covers most of the spatial area. Uniform representation or compression may therefore devote substantial resources to background regions without producing a proportional improvement in small-target detection performance. Recent importance-aware image transmission research has demonstrated that image patches can be assigned different quantization precisions according to their relative importance, providing a feasible basis for spatially differentiated representation [4].

Semantic communication provides a possible means of addressing this mismatch by emphasizing the effective preservation of semantic information rather than the exact recovery of all transmitted bits or pixels. Recent research has investigated semantic communication through multiple technical paradigms, including deep joint source–channel coding, generative semantic communication, and semantic knowledge modeling, with the common objective of reducing communication overhead while retaining useful information [5]. These developments indicate that communication resources can be configured according to the relative value of transmitted information instead of being distributed uniformly across the entire image.

Against this background, this study addresses the preservation of infrared small-target information in resource-constrained semantic transmission of UAV imagery. The objective is to reduce the resources consumed by extensive background content while retaining the visual information required for reliable infrared small-target detection.

1.1. Related Work

Recent studies on small-object detection in UAV imagery have mainly focused on preserving fine-grained features, enhancing multiscale representations, and suppressing interference from complex backgrounds. Li et al. [6] proposed SME-DETR, which combines a scale-aware attentional backbone with multidomain feature enhancement to integrate spatial, frequency, and channel information. Hu et al. [7] developed CSFPR-RTDETR by jointly modeling spatial–frequency features and positional relationships, thereby strengthening small-object cues while maintaining relatively efficient detection. Related studies have further improved UAV small-object detection through real-time Transformer architectures, scale-invariant feature disentanglement, and semantic-guided global–local feature fusion [8–10]. Visible–infrared cross-modal fusion has also been introduced to exploit complementary information from different imaging modalities [11]. Although these methods improve small-object representation at the detector level, they do not address the spatial allocation of limited transmission resources before detection.

Infrared small-target detection places greater emphasis on preserving weak target responses and separating targets from complex backgrounds. Guo et al. [12] proposed a diverse feature harmonization framework that coordinates complementary feature representations to enhance small-target information and suppress background interference.

Xiong et al. [13] developed DRPCA-Net by embedding the sparsity prior of infrared small targets and the low-rank property of backgrounds into a dynamic deep-unfolding network. Dense nested attention and interpretation-weighted sparse modeling have also been used to strengthen weak target responses while retaining background-suppression capability [14,15]. More recent methods introduce cross-layer relative-position embedding or dual-path edge and semantic modeling to improve target localization and feature discrimination [16,17]. Nevertheless, these approaches primarily optimize the detector itself and do not explicitly consider how weak target information should be preferentially preserved during resource-constrained image transmission.

In semantic communication, recent studies have incorporated semantic importance into feature transmission and resource allocation. Zhou et al. [18] proposed a feature-allocation framework with space–time importance awareness, which evaluates the relative importance of encoded features and assigns them to suitable temporal or spatial channels. Guo et al. [19] quantified the importance of semantic content and configured adaptive modulation and coding according to different importance levels. For UAV collaborative networks, semantic-aware resource allocation has also been investigated for multimodal information transmission [20]. In addition, task-adaptive visual semantic communication selectively transmits information required by downstream tasks, whereas multiscale Vision Transformer architectures improve the robustness of image semantic transmission [21,22]. These studies demonstrate that semantic importance provides a useful basis for differentiated transmission, although the allocation decisions are generally performed at the feature, channel, or overall semantic-content level.

At the image level, Hu et al. [23] proposed a resource-efficient multibranch semantic communication architecture based on saliency detection. Image regions with different levels of complexity are processed by semantic branches with different capacities, while communication and computation resources are jointly optimized. Adaptive semantic-and-bit communication has further been studied to select appropriate transmission modes and allocate resources for multiuser image transmission [24]. Wang et al. [25] jointly optimized semantic compression, transmission, and communication-resource allocation for multiuser image-classification tasks. These methods establish important connections between image characteristics, semantic relevance, and resource allocation. Nevertheless, existing studies do not specifically evaluate semantic importance from infrared small-target responses at the grid-cell level, nor do they jointly optimize grid partitioning and graded regional representation parameters. These limitations motivate the semantic-aware regional resource allocation method developed in this paper. For clarity, a comparison between representative related works and the proposed method is summarized in Table 1.

Table 1. Comparison between representative related works and our work.

Ref.	UAV-Assisted Scenario	Semantic Relevance Awareness	Region-Level Processing	Adaptive Transmission	Image Reconstruction	Joint Optimization	Detection Performance
[11]	✓						✓
[18]		✓		✓			
[20]	✓	✓		✓		✓	
[21]		✓		✓			✓
[23]			✓	✓	✓	✓	
[24]				✓	✓	✓	
[25]			✓	✓		✓	
Our Work	✓	✓	✓	✓	✓	✓	✓

✓ indicates that the corresponding characteristic is explicitly considered in the related work, while a blank entry indicates otherwise.

1.2. Contributions

The main contributions of this paper are summarized as follows:

- A semantic-aware resource allocation framework is developed for infrared small-target detection. Different from conventional ROI-based schemes that mainly distinguish target and background regions using predefined allocation rules, the proposed framework evaluates the semantic importance of individual grid cells and performs differentiated resource allocation under limited communication resources. This enables a more flexible allocation of communication resources according to the semantic importance of different image regions.
- A semantic-importance-based resource allocation method is proposed. The infrared image is first divided into multiple grid cells, and the semantic weight of each cell is calculated from the initial detection response. The semantic weights are then quantized into different levels, according to which corresponding bit resources and representation scales are assigned. This allows small-target-related regions to be represented with higher precision while reducing the resources allocated to background regions.
- A joint parameter optimization method based on CMA-ES is developed. The method jointly searches the grid partition and graded resource allocation parameters under a fixed communication budget, and determines a suitable grid configuration and its corresponding allocation parameters through semantic importance estimation, resource allocation, image reconstruction, and performance evaluation.

Experimental comparisons with existing schemes show that the proposed method effectively preserves small-target detection performance while reducing resource consumption, thereby achieving a more favorable trade-off between the two and validating the effectiveness of semantic-weight-based regional resource allocation.

2. System Model and Problem Formulation

2.1. System Model

We consider a UAV-assisted semantic communication system for infrared small-target detection, as illustrated in Figure 1. The system consists mainly of a UAV equipped with an infrared imaging sensor and a mobile edge computing (MEC) server. The UAV acquires infrared images of the ground scene through the sensing link and performs a region-wise semantic representation and differentiated compression according to the spatial distribution of task-relevant information. The processed visual information is subsequently transmitted to the MEC server through an air-to-ground semantic communication link. At the receiver, the infrared image is reconstructed and then used to perform the final small-target detection task. It should be noted that the preliminary detector at the UAV and the detector at the MEC server use the same detection model. The UAV-side detector is applied to the original infrared image to generate the detection response used for semantic importance estimation and subsequent resource allocation. In contrast, the detector at the MEC server is applied to the reconstructed image to evaluate whether the information related to small-target detection is effectively preserved after resource-constrained representation and transmission.

Infrared small targets generally occupy only a small number of pixels, whereas most image regions correspond to background content with relatively limited relevance to the downstream detection task. Applying uniform representation precision and compression strength to the entire image would distribute the limited communication resources equally between target-related and background regions, thereby reducing the resource efficiency of task-relevant information transmission. To address this issue, the preliminary detection result obtained at the UAV is employed as a spatial prior for evaluating the semantic importance of different image regions.

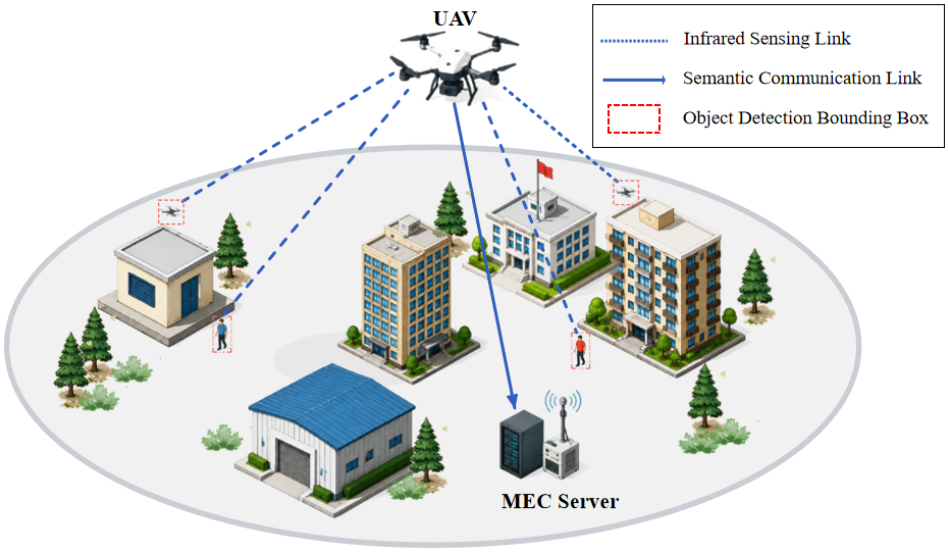


Figure 1. System model of UAV-assisted infrared sensing semantic communication.

Let $X_i \in \mathbb{R}^{H\times W}$ denote the i-th infrared image acquired by the infrared imaging sensor mounted on the UAV, where H and W denote the image height and width, respectively. A preliminary infrared small-target detector is deployed on the UAV to process the acquired image and generate a preliminary prediction map:

$$P_i = \mathbb{I}\big(\mathcal{F}_{\text{pre}}(X_i) \geq \tau\big), \tag{1}$$

where $\mathcal{F}_{\text{pre}}(\cdot)$ denotes the preliminary detector at the UAV, τ denotes the binarization threshold, and $P_i \in \{0,1\}^{H\times W}$.

For a given grid configuration $g = (r, c)$, the infrared image is divided into $r \times c$ non-overlapping local cells, denoted by $\Omega_{u,v}(g)$, where $u = 1, \ldots, r$ and $v = 1, \ldots, c$.

The semantic importance of cell $\Omega_{u,v}(g)$ in image X_i is defined as

$$w_i^{u,v}(g) = \frac{\sum_{(x,y)\in\Omega_{u,v}(g)} P_i(x,y)}{|\Omega_{u,v}(g)|}, \tag{2}$$

where $P_i(x,y)$ denotes the value of the preliminary prediction map at pixel location (x,y), and $|\Omega_{u,v}(g)|$ denotes the number of pixels contained in the corresponding cell. The regional semantic importance therefore represents the proportion of pixels identified as target-related within the corresponding grid cell. The semantic importance values of all cells form the regional semantic importance matrix:

$$\mathbf{W}_i(g) = \big[w_i^{u,v}(g)\big]_{r\times c}. \tag{3}$$

During the transmission of image X_i, let $\mathbf{q}_i$ and $\mathbf{q}_{\text{M}}$ denote the horizontal locations of the UAV and the MEC server, respectively, and let H_{U} denote the UAV altitude. The communication distance between the UAV and the MEC server is given by

$$d_i = \sqrt{H_{\text{U}}^2 + \|\mathbf{q}_i - \mathbf{q}_{\text{M}}\|_2^2}. \tag{4}$$

Considering a quasi-static air-to-ground channel dominated by line-of-sight propagation, the channel power gain is modeled as

$$h_i = \beta_0 \left(\frac{d_0}{d_i}\right)^2, \tag{5}$$

where d_0 denotes the reference distance, and β_0 denotes the channel power gain at d_0.

The corresponding air-to-ground channel capacity is expressed as

$$C_i = W_c \log_2\left(1 + \frac{P_t h_i}{\sigma^2}\right), \tag{6}$$

where W_c denotes the channel bandwidth, P_t denotes the UAV transmit power, and σ^2 denotes the receiver noise power.

Let T denote the predefined transmission-slot duration allocated to each infrared image. The maximum number of bits that can be transmitted over the air-to-ground channel during this slot is therefore TC_i. Accordingly, the image-level total bit budget is defined as

$$B_i^{\text{tot}} = \min\{B_{\max}, \lfloor TC_i \rfloor, \rho HW\}, \tag{7}$$

where $B_{\max}$ denotes the maximum allowable bit budget for a single image and ρ denotes the number of bits required to represent each pixel in the original image.

For a given grid configuration $g = (r, c)$ and image-level total bit budget B_i^{tot}, the regional bit-allocation result of image X_i is represented by

$$\mathbf{B}_i(g) = \left[b_i^{u,v}(g)\right]_{r\times c}, \tag{8}$$

where $b_i^{u,v}(g)$ denotes the number of bits allocated to cell $\Omega_{u,v}(g)$.

The regional bit allocation further determines the number of spatial samples retained in each cell and the corresponding integer compression dimensions. The height and width of cell $\Omega_{u,v}(g)$ are denoted by $H_{u,v}(g)$ and $W_{u,v}(g)$, respectively. The compressed height and width of cell $\Omega_{u,v}(g)$ are denoted by $\widetilde{H}_i^{u,v}(g)$ and $\widetilde{W}_i^{u,v}(g)$, respectively, and are determined jointly by the allocated regional bits and the original regional dimensions:

$$\left(\widetilde{H}_i^{u,v}(g), \widetilde{W}_i^{u,v}(g)\right) = \Phi\left(b_i^{u,v}(g); H_{u,v}(g), W_{u,v}(g)\right), \tag{9}$$

where $\Phi(\cdot)$ denotes the integer compression-dimension determination rule under the regional bit constraint. The compressed heights and widths of all cells are expressed as

$$\widetilde{\mathbf{H}}_i(g) = \left[\widetilde{H}_i^{u,v}(g)\right]_{r\times c}, \tag{10}$$

$$\widetilde{\mathbf{W}}_i(g) = \left[\widetilde{W}_i^{u,v}(g)\right]_{r\times c}. \tag{11}$$

Based on the compressed dimensions of all cells, the reconstructed image is expressed as

$$\hat{X}_i(g) = \mathcal{R}\left(X_i; g, \widetilde{\mathbf{H}}_i(g), \widetilde{\mathbf{W}}_i(g)\right), \tag{12}$$

where $\mathcal{R}(\cdot)$ denotes the overall image reconstruction process, including regional compression, regional recovery, and spatial recombination.

The reconstructed image is subsequently processed by the infrared small-target detector deployed at the MEC server:

$$\hat{P}_i(g) = \mathcal{F}_{\text{det}}\left(\hat{X}_i(g)\right), \tag{13}$$

where $\mathcal{F}_{\text{det}}(\cdot)$ denotes the downstream infrared small-target detector.

Accordingly, the grid configuration determines the spatial granularity of regional semantic modeling, while the regional semantic importance affects the spatial distribution of the available bit resources. The bits allocated to each cell further determine its com-

pressed spatial dimensions, thereby influencing the reconstructed image quality and the downstream small-target detection performance.

2.2. Problem Formulation

Based on the system model described above, the performance of regional semantic resource allocation is affected not only by the grid granularity, but also by the resource proportions assigned to cells with different semantic importance. Allocating excessive resources to low-importance background regions may reduce resource efficiency, whereas an excessively small allocation may cause severe background degradation and consequently affect image reconstruction and downstream small-target detection. Therefore, the resource-allocation weight should be considered together with the grid configuration.

Let $\boldsymbol{\eta} \in \mathcal{E}$ denote the resource-allocation scale vector associated with cells of different semantic importance, where $\mathcal{E}$ denotes the corresponding candidate set. The parameter $\boldsymbol{\eta}$ controls the relative distribution of the image-level bit budget among cells with different semantic importance. Thus, the number of bits allocated to cell $\Omega_{u,v}(g)$ can be described as $b_i^{u,v}(g, \boldsymbol{\eta})$.

For a given grid configuration g and resource-allocation weight $\boldsymbol{\eta}$, the comprehensive system performance is defined as

$$J(g, \boldsymbol{\eta}) = \mathcal{J}(M_{\text{task}}(g, \boldsymbol{\eta}), Q_{\text{rec}}(g, \boldsymbol{\eta})), \tag{14}$$

where $\mathcal{J}(\cdot)$ aggregates the downstream detection performance and image reconstruction quality into a scalar performance score. $M_{\text{task}}(g, \boldsymbol{\eta})$ denotes the downstream small-target detection performance, and $Q_{\text{rec}}(g, \boldsymbol{\eta})$ denotes the image reconstruction quality.

Based on the above definitions, the considered optimization problem is formulated as

$$(\text{P1}): \quad \max_{g \in \mathcal{G}, \boldsymbol{\eta} \in \mathcal{E}} J(g, \boldsymbol{\eta}) \tag{15}$$

$$\text{s.t.} \quad \sum_{u=1}^{r} \sum_{v=1}^{c} b_i^{u,v}(g, \boldsymbol{\eta}) = B_i^{\text{tot}}, \qquad \forall i, \tag{16}$$

$$\rho \le b_i^{u,v}(g, \boldsymbol{\eta}) \le \rho H_{u,v}(g) W_{u,v}(g), \qquad \forall i, u, v, \tag{17}$$

$$b_i^{u,v}(g, \boldsymbol{\eta}) \in \mathbb{Z}, \qquad \forall i, u, v, \tag{18}$$

where $\mathcal{G}$ denotes the set of candidate grid configurations. Constraint (16) requires that the image-level bit budget be completely allocated among all cells. Constraint (17) ensures that each cell is allocated at least ρ bits, corresponding to one pixel sample, while its allocated bits do not exceed the uncompressed data volume $\rho H_{u,v}(g, \boldsymbol{\eta}) W_{u,v}(g, \boldsymbol{\eta})$ of the corresponding cell. Constraint (18) requires the regional bit allocations to be integers.

3. Proposed Solution

Based on the system model and optimization problem formulated above, we propose a task-oriented region-wise resource allocation method. First, the regional semantic importance is discretized into ordered semantic levels, which are then mapped to allocation weights for regional bit allocation. Subsequently, the regional compression dimensions are determined according to the allocated bits, followed by region-wise compression and image reconstruction. Finally, the overall performance $J(g, \boldsymbol{\eta})$ is evaluated in terms of downstream task performance and image reconstruction quality. The overall framework of the proposed method is illustrated in Figure 2.

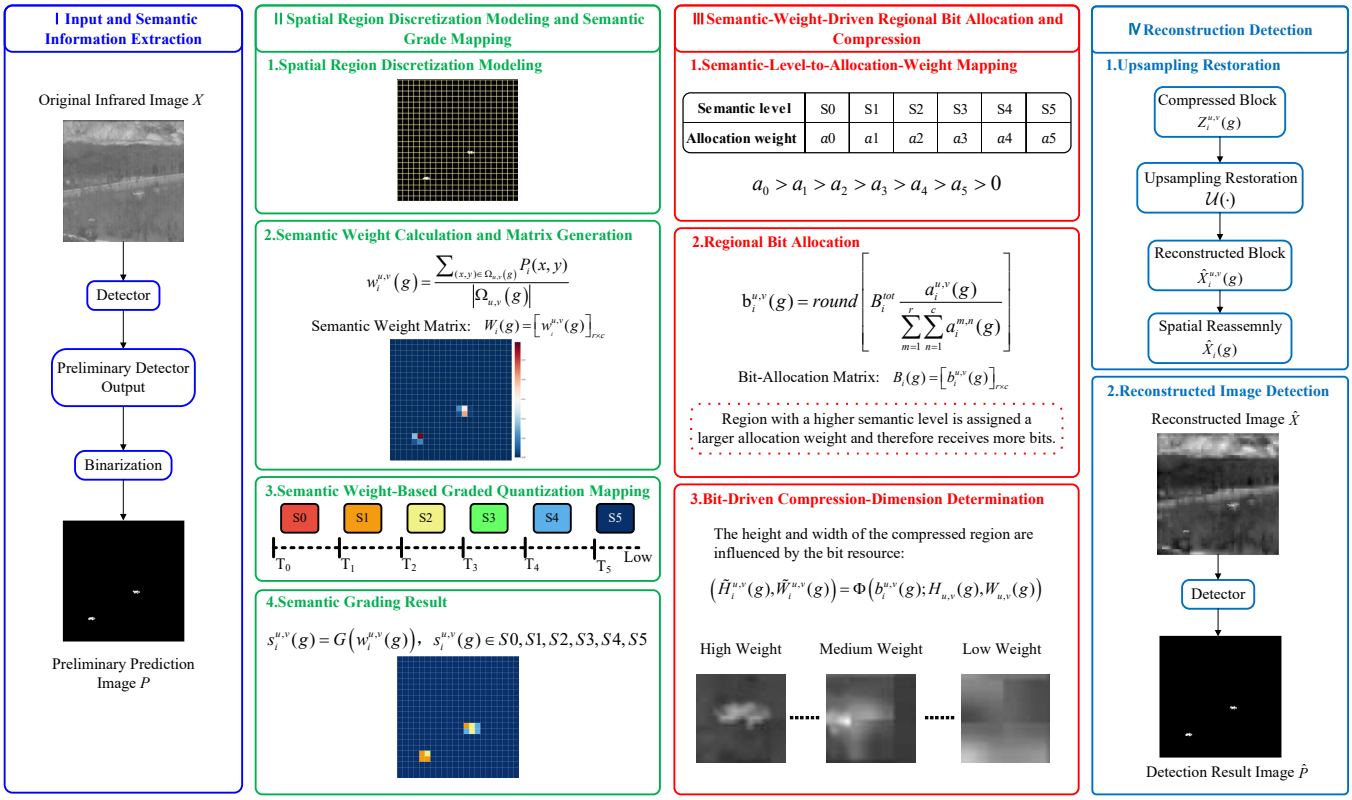


Figure 2. Overall framework of the proposed semantic-importance-driven regional resource allocation method.

3.1. Semantic-Weight-Driven Regional Resource Allocation

3.1.1. Regional Semantic-Level Classification

For the regional semantic importance $w_i^{u,v}(g)$, the continuous semantic importance values are discretized into six ordered semantic levels:

$$\mathcal{S}=\{S_0,S_1,S_2,S_3,S_4,S_5\}, \tag{19}$$

where S_0 represents the highest semantic level, and S_5 represents the lowest semantic level.

Let the semantic-level thresholds satisfy

$$0\le T_5<T_4<T_3<T_2<T_1<T_0\le 1. \tag{20}$$

The semantic level of cell $\Omega_{u,v}(g)$ is defined as

$$s_i^{u,v}(g)=\begin{cases}S_0, & T_1\le w_i^{u,v}(g)\le T_0,\\ S_k, & T_{k+1}\le w_i^{u,v}(g)<T_k, \quad k=1,2,3,\\ S_4, & T_5<w_i^{u,v}(g)<T_4,\\ S_5, & w_i^{u,v}(g)=T_5.\end{cases} \tag{21}$$

which realizes the mapping from continuous semantic importance to discrete semantic levels, enabling different cells to receive differentiated resource-allocation strategies.

The semantic levels of all cells form the semantic-level matrix

$$\mathbf{S}_i(g)=\left[s_i^{u,v}(g)\right]_{r\times c}. \tag{22}$$

Higher regional semantic importance indicates that the corresponding cell contains more task-relevant information and is therefore assigned a semantic level closer to S_0. Conversely, a cell dominated by background information is assigned a semantic level closer to S_5. This discretization converts continuous regional semantic importance values into a finite set of resource-allocation states.

3.1.2. Semantic-Level Mapping and Regional Bit Resource Allocation

Based on the resource-allocation scale vector $\boldsymbol{\eta}$, the allocation weight associated with semantic level S_k is defined as

$$a_k(\boldsymbol{\eta}) = \frac{1}{\eta_k}, \qquad k = 0, \ldots, 5. \tag{23}$$

The allocation scale coefficients satisfy

$$1 \leq \eta_0 < \eta_1 < \cdots < \eta_5. \tag{24}$$

The allocation weight of cell $\Omega_{u,v}(g)$ is determined by its semantic level:

$$a_i^{u,v}(g, \boldsymbol{\eta}) = a_k(\boldsymbol{\eta}), \qquad s_i^{u,v}(g) = S_k. \tag{25}$$

Under the image-level total bit budget B_i^{tot}, the number of bits allocated to cell $\Omega_{u,v}(g)$ is determined according to the proportion of its allocation weight in the total weight of all cells:

$$b_i^{u,v}(g, \boldsymbol{\eta}) = \text{round}\left[B_i^{\text{tot}} \frac{a_i^{u,v}(g, \boldsymbol{\eta})}{\sum_{m=1}^{r} \sum_{n=1}^{c} a_i^{m,n}(g, \boldsymbol{\eta})} \right], \tag{26}$$

where $\text{round}(\cdot)$ denotes the rounding operation used to obtain an integer number of allocated bits. After proportional allocation, an integer correction is performed to ensure that the regional bit allocations satisfy the prescribed lower and upper bounds and that their sum is exactly equal to B_i^{tot}.

According to this allocation rule, a cell with a higher semantic level is assigned a larger allocation weight and therefore receives more bits. In contrast, a cell with a lower semantic level receives fewer bits. The regional bit-allocation results form the bit-allocation matrix

$$\mathbf{B}_i(g, \boldsymbol{\eta}) = \left[b_i^{u,v}(g, \boldsymbol{\eta}) \right]_{r \times c}. \tag{27}$$

3.1.3. Bit Resource-Driven Regional Compression

The number of bits allocated to a cell directly determines the number of spatial samples that can be retained after compression. Let $H_{u,v}(g)$ and $W_{u,v}(g)$ denote the original height and width of cell $\Omega_{u,v}(g)$, respectively, and let $b_i^{u,v}(g, \boldsymbol{\eta})$ denote the number of bits allocated to this cell. Since each grayscale sample requires ρ bits, the maximum number of spatial samples that can be retained is defined as

$$N_i^{u,v}(g, \boldsymbol{\eta}) = \min\left\{ H_{u,v}(g) W_{u,v}(g), \left\lfloor \frac{b_i^{u,v}(g, \boldsymbol{\eta})}{\rho} \right\rfloor \right\}. \tag{28}$$

where $H_{u,v}(g) W_{u,v}(g)$ denotes the original number of pixels in the cell, and $\lfloor b_i^{u,v}(g, \boldsymbol{\eta}) / \rho \rfloor$ denotes the number of grayscale samples supported by the allocated regional bits. Taking the smaller value ensures that the number of retained samples exceeds neither the original

number of grid-cell pixels nor the number permitted by the regional bit budget. Since $\rho \le b_i^{u,v}(g,\boldsymbol{\eta})$, at least one grayscale sample is retained in each grid cell.

To determine the specific compressed height and width, all feasible candidate heights are sequentially examined. The candidate height h is selected from

$$h \in \left\{1, 2, \ldots, \min\left(H_{u,v}(g), N_i^{u,v}(g,\boldsymbol{\eta})\right)\right\}. \tag{29}$$

For a given candidate height h, the corresponding maximum candidate width is defined as

$$w_i^{u,v}(h,g,\boldsymbol{\eta}) = \min\left\{W_{u,v}(g), \left\lfloor \frac{N_i^{u,v}(g,\boldsymbol{\eta})}{h} \right\rfloor\right\}. \tag{30}$$

where $\left\lfloor N_i^{u,v}(g,\boldsymbol{\eta})/h \right\rfloor$ denotes the maximum integer width permitted by the retained-sample constraint when the candidate height is h. Taking the minimum with the original regional width $W_{u,v}(g)$ ensures that the candidate width does not exceed the original regional boundary.

To utilize the allocated regional bits as fully as possible, the set of candidate heights corresponding to the maximum retained area is first selected as

$$\mathcal{H}_i^{u,v}(g,\boldsymbol{\eta}) = \underset{1 \le h \le \min\left\{H_{u,v}(g), N_i^{u,v}(g,\boldsymbol{\eta})\right\}}{\arg\max} h\, w_i^{u,v}(h,g,\boldsymbol{\eta}). \tag{31}$$

where $h\, w_i^{u,v}(h,g,\boldsymbol{\eta})$ represents the retained area.

Different height–width combinations may retain the same number of spatial samples. Therefore, maximizing the retained area alone may not uniquely determine the compression dimensions. To reduce geometric distortion caused by dimensional changes, the aspect-ratio error between a candidate dimension pair and the original grid cell is defined as

$$E_i^{u,v}(h,g,\boldsymbol{\eta}) = \left| \ln\left(\frac{w_i^{u,v}(h,g,\boldsymbol{\eta})/h}{W_{u,v}(g)/H_{u,v}(g)} \right) \right|. \tag{32}$$

When the two aspect ratios are identical, $E_i^{u,v}(h,g,\boldsymbol{\eta}) = 0$. A smaller error indicates that the shape of the compressed region is closer to that of the original region, thereby reducing geometric distortion caused by dimensional changes.

After maximizing the number of retained spatial samples, the final compressed height is determined as

$$\widetilde{H}_i^{u,v}(g,\boldsymbol{\eta}) = \underset{h \in \mathcal{H}_i^{u,v}(g,\boldsymbol{\eta})}{\arg\min} E_i^{u,v}(h,g,\boldsymbol{\eta}). \tag{33}$$

The corresponding compressed width is given by

$$\widetilde{W}_i^{u,v}(g,\boldsymbol{\eta}) = w_i^{u,v}\left(\widetilde{H}_i^{u,v}(g,\boldsymbol{\eta}), g, \boldsymbol{\eta}\right). \tag{34}$$

Accordingly, the regional compression dimensions are determined according to two sequential criteria. First, the number of retained spatial samples is maximized under the regional bit constraint. Second, when multiple dimension pairs retain the same number of samples, the pair whose aspect ratio is closest to that of the original region is selected. This approach ensures that the compressed data volume does not exceed the allocated regional bits while reducing geometric distortion caused by changes in the regional shape.

3.1.4. Region Compression and Image Reconstruction

The local image block of X_i within grid cell $\Omega_{u,v}(g)$ is denoted by $X_i^{u,v}(g) = X_i|_{\Omega_{u,v}(g)}$. According to the compression dimensions determined in the preceding subsection, the compressed regional representation is defined as

$$Z_i^{u,v}(g,\boldsymbol{\eta}) = \mathcal{D}\Big(X_i^{u,v}(g); \widetilde{H}_i^{u,v}(g,\boldsymbol{\eta}), \widetilde{W}_i^{u,v}(g,\boldsymbol{\eta})\Big). \tag{35}$$

where $\mathcal{D}(\cdot)$ denotes the regional downsampling operation. Area-based interpolation is employed to reduce aliasing introduced by spatial compression. If the compression dimensions are identical to the original dimensions, the regional image block remains unchanged.

At the receiver, the compressed regional representation is restored to its original dimensions to obtain an image with the same spatial resolution as the original image. The restored regional image block is expressed as

$$\hat{X}_i^{u,v}(g,\boldsymbol{\eta}) = \mathcal{U}\big(Z_i^{u,v}(g,\boldsymbol{\eta}); H_{u,v}(g), W_{u,v}(g)\big), \tag{36}$$

where $\mathcal{U}(\cdot)$ denotes the regional upsampling operation. Bicubic interpolation is employed to obtain a smooth reconstructed regional block.

After all regions have been individually downsampled and restored, the restored regional blocks are reassembled according to their original spatial positions to obtain the complete reconstructed image:

$$\hat{X}_i(g,\boldsymbol{\eta}) = \mathcal{M}\left(\{\hat{X}_i^{u,v}(g,\boldsymbol{\eta})\}_{\substack{u=1,\ldots,r \\ v=1,\ldots,c}}\right). \tag{37}$$

where $\mathcal{M}(\cdot)$ denotes the spatial reassembly operation. Since the non-overlapping grid cells jointly cover the entire image, $\hat{X}_i(g,\boldsymbol{\eta})$ has the same spatial dimensions as the original image X_i.

Finally, the reconstructed image is fed into the downstream infrared small-target detector to obtain $\hat{P}_i(g,\boldsymbol{\eta})$, which is used for the subsequent evaluation of detection performance and reconstruction quality.

3.2. Performance Evaluation

To evaluate the overall system performance $J(g,\boldsymbol{\eta})$ under grid configuration g and resource-allocation scale vector $\boldsymbol{\eta}$, mIoU, global SSIM, target-region local SSIM, and PSNR are employed to assess the downstream small-target detection performance and image reconstruction quality. To ensure a consistent sample basis, all metrics are calculated over the common valid sample set $\mathcal{I}$, in which each sample has an available original image, reconstructed image, ground-truth annotation, detection result, and valid local target region. The number of valid samples is denoted by $N_{\mathrm{v}} = |\mathcal{I}|$. For each $i \in \mathcal{I}$, $Y_i \in \{0,1\}^{H\times W}$ denotes the corresponding ground-truth binary mask, where target and background pixels are represented by 1 and 0, respectively.

The mean intersection over union between the detection result $\hat{P}_i(g)$ and the ground-truth binary mask Y_i is defined as

$$\mathrm{mIoU}(g,\boldsymbol{\eta}) = \frac{1}{N_{\mathrm{v}}} \sum_{i\in\mathcal{I}} \frac{\big|\hat{P}_i(g,\boldsymbol{\eta}) \cap Y_i\big|}{\big|\hat{P}_i(g,\boldsymbol{\eta}) \cup Y_i\big|}. \tag{38}$$

This metric measures the average overlap between the predicted target regions obtained from the reconstructed images and the corresponding ground-truth target regions.

Global structural similarity evaluates the consistency between the original and reconstructed images in terms of luminance, contrast, and structural information. The average global SSIM is calculated as

$$\mathrm{SSIM}(g,\boldsymbol{\eta}) = \frac{1}{N_{\mathrm{v}}} \sum_{i\in\mathcal{I}} \frac{\Big(2\mu_{X_i}\mu_{\hat{X}_i(g,\boldsymbol{\eta})} + C_1\Big)\Big(2\sigma_{X_i,\hat{X}_i(g,\boldsymbol{\eta})} + C_2\Big)}{\Big(\mu_{X_i}^2 + \mu_{\hat{X}_i(g,\boldsymbol{\eta})}^2 + C_1\Big)\Big(\sigma_{X_i}^2 + \sigma_{\hat{X}_i(g,\boldsymbol{\eta})}^2 + C_2\Big)}. \tag{39}$$

where μ_{X_i} and $\mu_{\hat{X}_i(g,\boldsymbol{\eta})}$ denote the mean intensities of the original and reconstructed images, respectively; $\sigma^2_{X_i}$ and $\sigma^2_{\hat{X}_i(g,\boldsymbol{\eta})}$ denote their variances; $\sigma_{X_i,\hat{X}_i(g,\boldsymbol{\eta})}$ denotes their covariance; and C_1 and C_2 are stability constants. A larger SSIM indicates higher global structural consistency.

Since infrared small targets occupy only a small proportion of the entire image, global SSIM may be dominated by large background regions. To further evaluate information preservation in the target region, the minimum bounding rectangle covering all target pixels in Y_i is first obtained and expanded outward by d_{pad} pixels. Boundaries exceeding the image range are clipped to the corresponding image boundaries, resulting in the local target region

$$\mathcal{T}_i = \text{ROI}\Big(Y_i, d_{\text{pad}}\Big). \tag{40}$$

The average local structural similarity in the target region is calculated as

$$\text{SSIM}_{\text{local}}(g,\boldsymbol{\eta}) = \frac{1}{N_{\text{v}}} \sum_{i\in\mathcal{I}} \text{SSIM}\Big(X_i^{\text{local}}, \hat{X}_i^{\text{local}}(g,\boldsymbol{\eta})\Big). \tag{41}$$

where $X_i^{\text{local}} = X_i|_{\mathcal{T}_i}$ and $\hat{X}_i^{\text{local}}(g,\boldsymbol{\eta}) = \hat{X}_i(g,\boldsymbol{\eta})|_{\mathcal{T}_i}$ denote the local image blocks extracted from the target region $\mathcal{T}_i$ of the original and reconstructed images, respectively. The boundary expansion preserves a certain amount of surrounding background information and avoids unstable structural-similarity evaluation caused by an excessively small target region.

Peak signal-to-noise ratio is employed to evaluate the pixel-level distortion between the original and reconstructed images. Its average value is calculated as

$$\text{PSNR}(g,\boldsymbol{\eta}) = \frac{1}{N_{\text{v}}} \sum_{i\in\mathcal{I}} 10\log_{10}\left[\frac{L^2}{\frac{1}{HW}\sum_{x=1}^{H}\sum_{y=1}^{W}\left[X_i(x,y) - \hat{X}_i(g,\boldsymbol{\eta};x,y)\right]^2}\right]. \tag{42}$$

where L denotes the maximum possible pixel value. The denominator represents the pixel-wise mean squared error between the original image X_i and the reconstructed image $\hat{X}_i(g,\boldsymbol{\eta})$, which measures their average pixel deviation. A larger PSNR indicates lower pixel-level reconstruction distortion.

Since the numerical range of PSNR differs from those of the other evaluation metrics, its normalized form is used in the overall system performance $J(g,\boldsymbol{\eta})$:

$$\overline{\text{PSNR}}(g,\boldsymbol{\eta}) = \min\left\{\frac{\text{PSNR}(g,\boldsymbol{\eta})}{P_{\text{ref}}}, 1\right\}, \tag{43}$$

where P_{ref} denotes a predefined PSNR reference value. This operation limits the normalized PSNR to a maximum of one, thereby reducing the numerical-scale difference between PSNR and the other evaluation metrics.

Based on the above evaluation metrics, the overall performance function in (14) is specified as

$$\begin{aligned} J(g,\boldsymbol{\eta}) = {} & \lambda_1\,\text{mIoU}(g,\boldsymbol{\eta}) + \lambda_2\,\text{SSIM}_{\text{local}}(g,\boldsymbol{\eta}) \\ & + \lambda_3\,\text{SSIM}(g,\boldsymbol{\eta}) + \lambda_4\overline{\text{PSNR}}(g,\boldsymbol{\eta}), \end{aligned} \tag{44}$$

where λ_1, λ_2, λ_3 and λ_4 are the weighting coefficients of the evaluation metrics. It should be noted that the weighting coefficients in the objective function represent the relative importance of different evaluation metrics and are used to combine the detection and reconstruction performance into a unified scalar objective, rather than serving as optimizable

parameters of the proposed resource allocation method. The individual mIoU, Local SSIM, SSIM, and PSNR values are also reported separately in the performance evaluation. Therefore, the aggregated Score is mainly used for candidate evaluation during optimization and for overall comparison among different optimization methods.

3.3. Joint Parameter Optimization Based on CMA-ES

The grid configuration g is a discrete variable, whereas the resource-allocation scale vector η consists of multiple continuous variables. Moreover, the regional bit-allocation procedure involves integer rounding and constraint correction. The subsequent compression dimension determination, regional reconstruction, and downstream detection operations also do not establish an explicitly differentiable relationship between the decision variables and the comprehensive evaluation score. Therefore, the objective function $J(g, \eta)$ can be regarded as a computationally expensive black-box function that is non-smooth and non-differentiable.

To solve the resulting mixed discrete–continuous optimization problem, a covariance matrix adaptation evolution strategy (CMA-ES) is employed to jointly search for the grid configuration and resource-allocation parameters within a predefined evaluation budget. CMA-ES generates candidate solutions from a multivariate Gaussian distribution and adaptively updates the mean, covariance matrix, and global step size according to their evaluated performance. The mean determines the current search center, the covariance matrix learns the correlations among optimization variables and favorable search directions, and the global step size controls the search range. Since CMA-ES does not require the analytical form or gradient information of the objective function, it is suitable for the black-box optimization problem considered in this work.

Since standard CMA-ES is primarily designed for continuous variables, the discrete grid configuration and continuous decay parameters are represented in a unified form as

$$\boldsymbol{\xi} = (\gamma, \boldsymbol{\beta}), \qquad \gamma \in [0,1], \quad \boldsymbol{\beta} \in (0,1)^5. \tag{45}$$

where γ denotes the continuous encoding of the grid configuration, and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_5]^{\mathrm{T}}$ represents the resource decay parameters between adjacent semantic levels. This representation combines one grid-encoding variable and five continuous resource parameters into a unified search variable, allowing CMA-ES to jointly adjust both types of variables under the same search distribution.

For the candidate grid set $\mathcal{G}$, the grid configurations are first arranged in a fixed order, and the interval $[0,1]$ is divided into subintervals corresponding to the number of candidate grids. Each subinterval is associated with one candidate grid. Once γ is generated by CMA-ES, the grid configuration g corresponding to the location of γ within $[0,1]$ is selected. The decay vector $\boldsymbol{\beta}$ is then converted into the resource-allocation scale vector $\boldsymbol{\eta}(\boldsymbol{\beta})$ according to the cumulative mapping introduced previously. Since all decay parameters are restricted to $(0,1)$, the prescribed monotonic relationship among the resource-allocation scales can be maintained.

At generation t, CMA-ES constructs a multivariate Gaussian search distribution using the current mean vector $\boldsymbol{m}_t$, covariance matrix $\boldsymbol{C}_t$, and global step size σ_t. Candidate parameter combinations are sampled as

$$\boldsymbol{\xi} \sim \mathcal{N}\left(\boldsymbol{m}_t, \sigma_t^2 \boldsymbol{C}_t\right). \tag{46}$$

where $\mathcal{N}(\cdot)$ denotes a multivariate Gaussian distribution and $\boldsymbol{m}_t$ denotes the current search center and determines the parameter region in which the candidates are mainly distributed. The covariance matrix $\boldsymbol{C}_t$ describes the correlations among optimization variables and the

favorable search directions, while σ_t controls the overall search range. A population of N_{pop} candidates is generated from this distribution at each generation. If any candidate component exceeds its feasible range, it is projected onto the corresponding boundary to ensure that the grid encoding and decay parameters remain feasible.

For each candidate vector $\boldsymbol{\xi}$, the component γ is first decoded into a candidate grid configuration g, and the decay vector $\boldsymbol{\beta}$ is converted into the resource-allocation scale vector $\boldsymbol{\eta}(\boldsymbol{\beta})$. The complete procedures of regional bit-resource allocation, local image compression, image reconstruction, and downstream small-target detection are subsequently performed to calculate the overall system performance $J(g, \boldsymbol{\eta}(\boldsymbol{\beta}))$.

Standard CMA-ES is generally implemented for minimization, whereas the objective considered in this work is to maximize the overall system performance. Accordingly, the fitness function is defined as

$$f(\boldsymbol{\xi}) = -J(g, \boldsymbol{\eta}). \tag{47}$$

After all candidates in the current generation have been evaluated, CMA-ES ranks them according to their fitness values and selects the best μ candidates to update the search distribution, where μ is automatically determined by the standard CMA-ES implementation according to the population size N_{pop}. Specifically, the mean vector moves toward parameter regions with higher performance, the covariance matrix learns the correlations among variables and promising search directions from the selected candidates, and the global step size is adaptively adjusted according to the search progress. These quantities follow the standard CMA-ES update rules to balance global exploration and local refinement.

Let J_n denote the overall system performance obtained at the n-th complete evaluation. The best-observed performance after n evaluations is defined as

$$J_{\text{best}}(n) = \max_{1 \le l \le n} J_l. \tag{48}$$

The search terminates when the number of complete evaluations reaches the predefined upper limit $\nu_{\max}$. In addition, if the improvement in the best-observed performance over the most recent ν_s evaluations is smaller than the minimum effective improvement threshold ε, further searching is considered unlikely to produce a meaningful improvement:

$$J_{\text{best}}(n) - J_{\text{best}}(n - \nu_s) < \varepsilon. \tag{49}$$

where ν_s denotes the convergence-checking window, and ε is used to exclude ineffective improvements caused by minor numerical fluctuations. The optimization process may also terminate when the search step size, covariance matrix, or other internal states satisfy the standard stopping conditions of CMA-ES.

After termination, the parameter combination with the highest overall system performance among all completely evaluated candidates is selected as

$$(g^\star, \boldsymbol{\beta}^\star) = \arg\max_{(g,\boldsymbol{\beta}) \in \mathcal{V}} J(g, \boldsymbol{\eta}(\boldsymbol{\beta})), \tag{50}$$

where $\mathcal{V}$ denotes the set of all candidate parameter combinations that have undergone complete system evaluation during the CMA-ES search. The variables $g^\star$ and $\boldsymbol{\beta}^\star$ denote the best-observed grid configuration and decay vector, respectively. The final resource-allocation scale vector is obtained as

$$\boldsymbol{\eta}^\star = \boldsymbol{\eta}(\boldsymbol{\beta}^\star). \tag{51}$$

The joint optimization algorithm to solve the primal problem (P1) is summarized in Algorithm 1.

Algorithm 1 Joint Optimization of Grid Configuration and Resource-Allocation Parameters Based on CMA-ES

Input: Candidate grid set $\mathcal{G}$, optimization dataset, population size N_{pop}, maximum number of evaluations $\nu_{\max}$, convergence window ν_s, and minimum effective improvement threshold ε
Output: Best grid configuration $g^\star$ and best resource-allocation scale vector $\boldsymbol{\eta}^\star$
Initialize: the mean vector $\boldsymbol{m}_0$, covariance matrix $\boldsymbol{C}_0$, global step size σ_0, evaluation count, and best-observed performance
while $n < \nu_{\max}$ and the stopping criterion is not satisfied **do**
    Sample candidate vectors from the Gaussian distribution defined in (46)
    for each candidate vector $\boldsymbol{\xi}$ **do**
        Decode $\boldsymbol{\xi}$ into g and $\boldsymbol{\beta}$, obtain $\boldsymbol{\eta}(\boldsymbol{\beta})$ and evaluate $J(g, \boldsymbol{\eta}(\boldsymbol{\beta}))$
        Compute the fitness according to (47)
        Update $g^\star$, $\boldsymbol{\beta}^\star$, and $J^\star$ when a better candidate is obtained
        $n \leftarrow n + 1$
    end for
    Rank the candidates by fitness and select the best μ candidates and update $\boldsymbol{m}_t$, $\boldsymbol{C}_t$, and σ_t according to the standard CMA-ES rules
    $t \leftarrow t + 1$
end while

4. Results

4.1. Dataset Description

The NUDT-SIRST dataset developed by Li et al. [14] was employed for the experiments in this study. The dataset contains 1327 infrared small-target images synthesized using real infrared backgrounds, together with accurate pixel-level ground-truth masks. It covers five representative background categories, including city, field, highlight, sea, and cloud scenes. The target categories include point, spot, and extended targets, with the extended targets involving different objects such as unmanned aerial vehicles, aircraft, and ships. These targets exhibit diverse sizes, poses, and signal-to-clutter ratios. To improve the physical rationality and visual realism of the synthesized images, the target type and size were adaptively selected according to the background scene, while intensity adjustment and Gaussian blurring were applied to improve the target intensity distribution and boundary smoothness. Approximately 37% of the images contain at least two targets; 27% of the targets occupy no more than 0.01% of the entire image area; and 96% satisfy the SPIE definition of an infrared small target, in which the target area is smaller than 0.15% of the image area. Moreover, approximately 32% of the targets are outside the top 10% of the image brightness range, indicating relatively low visual saliency.

For all optimization methods, the NUDT-SIRST dataset is divided into an optimization set and an independent test set at a ratio of 8:2. The optimization set is used to search for the grid configuration and the corresponding parameters, while the detection network remains fixed throughout the optimization process. After the search is completed, the optimal grid configuration and parameter values obtained by each optimization method are fixed and evaluated only once on the independent test set that has never participated in the search process. The same data partition and evaluation protocol are adopted for all compared optimization algorithms to ensure a fair comparison and avoid information leakage from the test set.

4.2. Parameters Settings

The semantic-level thresholds and the weighting coefficients in Table 2 are selected according to the characteristics and performance priorities of the considered infrared small-target detection task. Since infrared small targets generally occupy only a small number of pixels, the semantic importance values of most grid cells tend to be relatively low.

Therefore, finer threshold intervals are adopted in the low-semantic-importance range to better distinguish background-dominated regions with different levels of target relevance, while relatively wider intervals are used for regions with higher semantic importance.

For the weighting coefficients, preserving small-target detection performance is given the highest priority, followed by the structural preservation of target-related regions. Therefore, mIoU and Local SSIM are assigned weights of 0.40 and 0.30, respectively, while Global SSIM and normalized PSNR are each assigned a weight of 0.15 to account for overall structural and pixel-level reconstruction quality. The four weights sum to one.

All experiments are conducted on a computing platform equipped with a 12th Gen Intel Core i7-12700F CPU at 2.10 GHz, an NVIDIA GeForce RTX 3060 GPU with 12 GB memory, and 16 GB RAM. CMA-ES is performed during the offline optimization stage to determine the optimal grid configuration g^* and the corresponding resource allocation parameters $\boldsymbol{\eta}^*$. After optimization, these parameters are fixed and directly applied during actual UAV–MEC operation.

Table 2. Simulation parameter settings.

Notation	Definition	Parameter Value
H_U	UAV altitude	100 m
β_0	Reference channel power gain	-40 dB
W_c	Channel bandwidth	100 MHz
P_t	UAV transmit power	10 dBm
σ^2	Receiver noise power	-84 dBm
T	Transmission-slot duration per image	1 ms
d_0	Reference distance	1 m
$H \times W$	Image resolution	256×256 pixels
ρ	Bits per grayscale pixel	8 bits/pixel
$B_{\max} = \rho HW$	Maximum image budget	524,288 bits
τ	Prediction threshold	127
$(T_0, T_1, T_2, T_3, T_4, T_5)$	Semantic thresholds	$(1, 0.65, 0.40, 0.12, 0.02, 0)$
d_{pad}	Bounding-box padding	5 pixels
L	Maximum pixel value	255
P_{ref}	PSNR reference	40 dB
λ_1	mIoU weight	0.40
λ_2	Local SSIM weight	0.30
λ_3	Global SSIM weight	0.15
λ_4	Normalized PSNR weight	0.15
$\nu_{\max}$	Maximum evaluations	300
N_{pop}	Population size	10
ν_s	Convergence window	20
ε	Improvement threshold	10^{-4}

4.3. Experimental Results

4.3.1. Optimization Algorithm Selection Analysis

Three representative black-box optimization algorithms are selected for comparison: Tree-structured Parzen Estimator (TPE) constructs probabilistic models of the objective function and selects promising candidates based on the estimated distributions, enabling efficient parameter exploration. Sequential Model-based Algorithm Configuration (SMAC) builds a surrogate model using random forests and performs sequential sampling to improve the optimization process. Genetic Algorithm (GA) imitates natural evolutionary mechanisms and iteratively updates candidate solutions through selection, crossover, and mutation operations. All these algorithms are gradient-free optimization methods and are suitable for solving the complex optimization problem in this work, which involves discrete grid configurations and continuous resource allocation parameters.

For clarity, an evaluation refers to one complete performance assessment of a candidate parameter configuration, including resource allocation, image reconstruction, and metric calculation. An iteration refers to one optimizer update cycle, whose specific implemen-

tation depends on the optimization algorithm. For population-based methods such as CMA-ES and GA, a generation consists of evaluating a population of multiple candidate solutions followed by one population update. Since the internal update mechanisms and the number of candidate solutions evaluated per iteration or generation differ across optimization algorithms, the number of iterations or generations is not directly comparable. Therefore, all methods are evaluated under the same maximum budget of 300 candidate evaluations. To reduce the influence of randomness in the stochastic optimization process, each optimization algorithm is repeated 10 times using the same data partition, maximum evaluation budget, optimization variables, and objective function. Table 3 reports the mean and standard deviation of the resulting performance metrics to characterize the stability of different optimization methods.

As shown in Table 3, CMA-ES effectively solves the formulated resource allocation optimization problem and obtains a stable optimization result with fewer evaluations. At the 113th evaluation, CMA-ES achieves an optimal Score of 0.8275, which is very close to the results obtained by SMAC and TPE, being approximately 0.08% lower than SMAC and 0.19% higher than TPE, while improving by about 2.20% compared with GA. This demonstrates its favorable search efficiency and optimization stability. In terms of downstream detection performance, CMA-ES achieves an mIoU of 0.6106, which is comparable to the best result obtained by SMAC and higher than those of TPE and GA, indicating its effectiveness in preserving semantic information related to infrared small-target detection.

Table 3. Comparison of optimal solutions obtained by different optimization algorithms.

Method	Best Evaluation	Score	mIoU	Local SSIM	SSIM	PSNR
CMA-ES	**113**	$\underline{0.8275} \pm 0.0022$	$\underline{0.6106} \pm 0.0051$	$\underline{0.9840} \pm 0.0008$	$\mathbf{0.9631} \pm 0.0006$	$\mathbf{38.2888} \pm 0.0305$
TPE	160	0.8259 ± 0.0077	0.6079 ± 0.0159	$\mathbf{0.9841} \pm 0.0013$	$\underline{0.9613} \pm 0.0033$	38.2061 ± 0.2863
SMAC	193	$\mathbf{0.8282} \pm 0.0000$	$\mathbf{0.6138} \pm 0.0001$	$\underline{0.9840} \pm 0.0000$	$\underline{0.9613} \pm 0.0000$	38.2056 ± 0.0000
GA	154	0.8097 ± 0.0017	0.5873 ± 0.0082	0.9729 ± 0.0016	0.9575 ± 0.0031	37.1524 ± 0.2338

Bold and underlined values indicate the best and second-best performances in each column, respectively.

Regarding image reconstruction quality, CMA-ES exhibits favorable performance, obtaining the highest SSIM and PSNR values among the four optimization algorithms, with values of 0.9631 and 38.2888, respectively. Compared with SMAC, CMA-ES improves SSIM and PSNR by approximately 0.19% and 0.22%, respectively, indicating that the resource allocation strategy obtained by CMA-ES can better preserve the global image structure and visual quality. Meanwhile, CMA-ES achieves a Local SSIM of 0.9840, which is comparable to those of TPE and SMAC. The corresponding standard deviations of these evaluation metrics remain relatively small, indicating limited performance fluctuations and good stability of the obtained solutions. Although SMAC achieves a slightly higher Score and mIoU, CMA-ES reaches a competitive solution with fewer evaluations while maintaining favorable reconstruction quality. Therefore, considering optimization efficiency, detection performance, reconstruction quality, and result stability, CMA-ES provides a favorable trade-off among these aspects. The computational costs of the main processing stages are summarized in Table 4.

Table 4. Computational cost of the main processing stages.

Stage	Computational Time
Preliminary Detection	44.24 ms/image
Semantic Importance Estimation + Resource Allocation	6.42 ms/image
Regional Compression/Reconstruction	26.18 ms/image
CMA-ES Optimization	5.71 s/evaluation

4.3.2. Overall Comparison with Different Methods

To evaluate the effectiveness of the proposed method, six representative baseline methods are considered. Specifically, Average Allocation uniformly distributes the total bit budget among all image regions, whereas Random Allocation randomly assigns regional resources subject to the total budget constraint. Huffman Coding performs lossless variable-length coding according to the occurrence frequencies of pixel symbols, while JPEG Coding represents a conventional lossy image compression scheme with an adjustable compression quality. ROI Coding adopts differentiated coding strategies for target and background regions to preferentially preserve regions of interest. DeepJSCC employs neural-network-based encoders and decoders to realize end-to-end joint source–channel coding.

To investigate the performance under different communication conditions, the UAV altitude, transmit power, channel bandwidth, receiver noise power and transmission-slot duration are fixed according to Table 2, while the UAV–MEC communication distance is varied. For each communication distance, the corresponding channel power gain and channel capacity are calculated according to (5) and (6), respectively, and the available image-level bit budget is subsequently determined by (7). Therefore, different communication distances correspond to different available resource budgets for infrared image transmission. For convenient comparison of communication resource availability, the available bit rate is further defined as the percentage of the image-level bit budget relative to the uncompressed bit volume of the original image:

$$\gamma_i = \frac{B_i^{\mathrm{tot}}}{\rho HW} \times 100\%. \tag{52}$$

Accordingly, the available bit rate is determined by the available bit budget obtained from the channel model and provides a normalized representation of the communication resources available at different UAV–MEC distances. For example, when the UAV–MEC communication distance is 221 m, the corresponding SNR is approximately 7.12 dB, resulting in an available bit budget of 262,144 bits and an available bit rate of 50%. As the communication distance increases, the available communication resources gradually decrease. At a distance of 661 m, the available bit rate decreases to approximately 12.5%, while at 2953 m, it further decreases to approximately 0.78%.

Under the same UAV–MEC communication condition, the proposed method, JPEG, ROI coding, and Huffman coding are assigned the same nominal image-level resource budget determined by the channel model. For the proposed method, JPEG, and ROI coding, the budget mainly constrains the retained 8-bit spatial samples, with JPEG and ROI using their respective fixed coding settings after the sampling dimensions are determined. For Huffman coding, the encoded pixel payload is constrained within the same nominal budget. Method-specific side information, such as ROI coordinates, Huffman tables, and protocol overhead, is not uniformly included in the current resource accounting. Therefore, the comparisons in this study are conducted under the same nominal equivalent resource budget rather than under strictly identical end-to-end communication costs.

For the DeepJSCC baseline, a convolutional encoder–decoder architecture is employed. The encoder produces a latent representation of size $C \times 16 \times 16$ from each 256×256 input image, which is power-normalized and transmitted through an AWGN channel before image reconstruction by the decoder. The network is trained under an AWGN channel with an SNR of 10 dB using Adam for 80 epochs with a batch size of 8 and a learning rate of 10^{-3}, using the reconstruction loss $L = L_1 + 0.2L_{\mathrm{MSE}}$. Since DeepJSCC does not generate a conventional discrete bitstream, its communication resource is characterized by the number of real-valued channel symbols and the channel bandwidth ratio (CBR). The CBR is defined as the ratio between the number of transmitted latent channel symbols and the number

of source-image pixels. To align its resource levels with the nominal image-level budgets used by the other methods, each real-valued channel symbol is mapped to 8 equivalent bits, and the latent-channel dimension C is adjusted accordingly, with an independently trained model used for each resource level. These equivalent bits are used only for resource normalization and do not represent physically transmitted digital bits.

To provide a consistent basis for comparison, all methods use the same input images, downstream detector, evaluation protocol, and the UAV–MEC communication distance of 661 m, corresponding to an available bit rate of 12.5%. The overall performance comparison under this setting is presented in Table 5.

Table 5. Overall performance comparison of different methods under the available bit rate of 12.5%.

Method	mIoU	Local SSIM	SSIM	PSNR (dB)
Proposed Method	**0.6106**	**0.9840**	**0.9631**	**38.2888**
Average Allocation	0.1937	0.8717	0.9585	37.8376
Random Allocation	0.2323	0.8888	0.9624	38.1326
Huffman Coding	0.3493	0.8679	0.9417	35.9884
JPEG Coding	0.3335	0.8902	0.9586	38.0444
ROI	0.2682	0.9705	0.7524	27.5831
DeepJSCC	0.0350	0.7606	0.8746	31.9552

Bold values indicate the best performance.

To further analyze the influence of preliminary detection errors on the proposed resource allocation method, we examined the actual preliminary detection results over all 1327 images. Among them, 14 images, accounting for approximately 1.06% of the dataset, exhibited no overlap between the preliminary detection and the ground-truth target, including 13 images with empty predictions. After the subsequent semantic-importance estimation, resource allocation, and image reconstruction, the detection IoU of these 14 images at the MEC server remained zero. These results indicate that preliminary detection errors can propagate through the subsequent semantic-importance estimation and resource allocation process. In particular, complete target misses may assign insufficient semantic priority to the corresponding target regions, resulting in reduced resource allocation and further degradation of the detection performance at the MEC server. Improving the robustness of the proposed method against preliminary missed detections therefore remains an important direction for future work.

4.3.3. Performance Under Different Resource Budgets

As shown in Figure 3, the mIoU values of all methods generally increase as the available bit rate rises from 0.78% to 50%, indicating that additional transmission resources help preserve task-relevant information in the reconstructed images. The proposed method consistently achieves the highest mIoU under all evaluated resource settings, demonstrating that semantic-importance-based allocation can assign more bits and retain larger spatial dimensions to target-related regions, thereby reducing the loss of small-target features during compression. At an available bit rate of 12.5%, the proposed method achieves an mIoU of 0.6106, representing an improvement of approximately 74.8% over Huffman coding, the best-performing baseline under this setting. When the available bit rate reaches 50%, its mIoU increases to 0.7445, outperforming JPEG coding by approximately 11.8%. The performance gaps gradually narrow as more resources become available because both target and background regions can be reconstructed with higher quality. Nevertheless, the proposed method maintains a clear advantage, indicating that it is effective not only under extremely constrained resources but also under relatively high resource budgets. In contrast, average and random allocation do not distinguish regional importance; conventional coding methods mainly optimize overall compression quality; and the global reconstruction objective of DeepJSCC provides limited protection for small targets occu-

pying only a small portion of the image. Figure 4 provides a direct reference between the UAV–MEC communication distance and the corresponding available bit rates used in the mIoU evaluation.

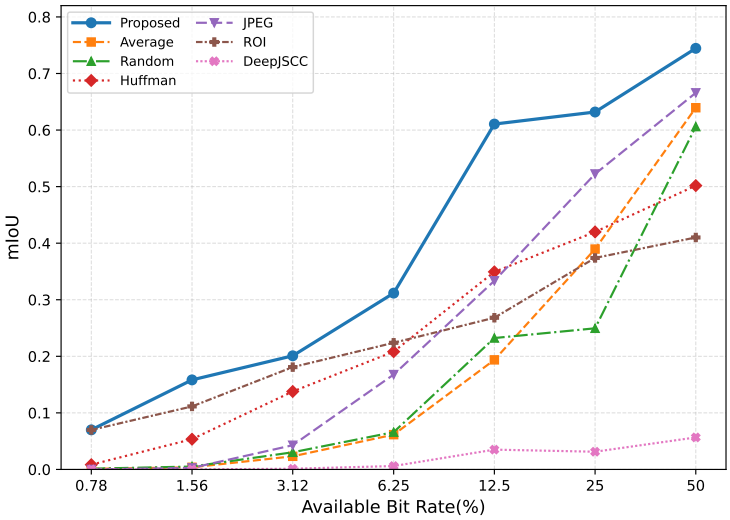


Figure 3. mIoU under different available bit rates.

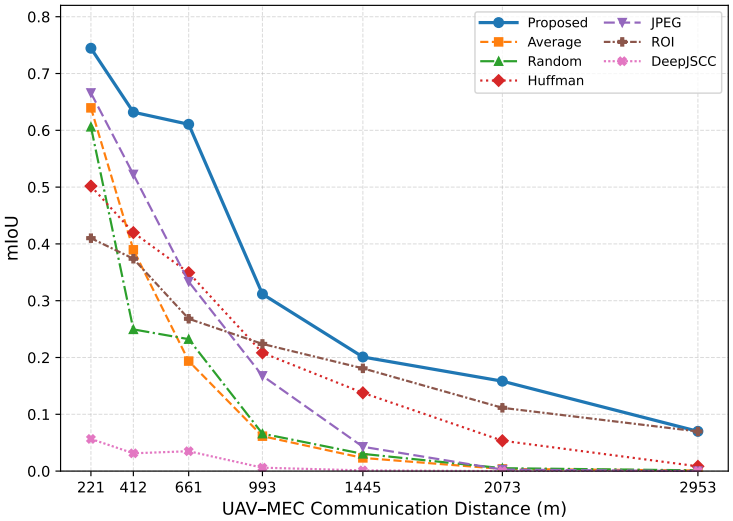


Figure 4. mIoU under different UAV–MEC communication distances.

As shown in Figure 5, the $SSIM_{local}$ values of all methods gradually improve with increasing resource availability, demonstrating that additional resources enhance structural preservation in target-related regions. The proposed method achieves the highest local structural similarity under most resource settings and reaches 0.9975 at an available bit rate of 50%, indicating that the allocated resources are effectively converted into improved reconstruction quality around the targets. Average and random allocation exhibit relatively limited target-region preservation under low-resource conditions because they do not distinguish regional semantic importance. The local reconstruction quality of Huffman and JPEG coding gradually improves as more resources become available, with JPEG demonstrating strong reconstruction capability under relatively high available bit rates. In contrast, DeepJSCC consistently achieves comparatively low $SSIM_{local}$, suggesting that its global reconstruction objective provides insufficient protection for infrared targets occupying only a small image area. At available bit rates of 3.125% and 6.25%, ROI coding slightly outperforms the proposed method in terms of $SSIM_{local}$, since it directly concentrates resources on predefined ROI regions and therefore provides high local pixel

similarity. However, at an available bit rate of 6.25%, the proposed method still improves the mIoU by approximately 39.2% over ROI coding. This result indicates that local structural similarity alone cannot fully characterize the usefulness of reconstructed images for the detection task. In addition to target-region fidelity, the surrounding background, local contrast, and boundary continuity also affect downstream detection. Therefore, the two curves jointly demonstrate that the proposed method preserves both local reconstruction quality and task-relevant semantic information more effectively. Figure 6 provides a direct reference between the UAV–MEC communication distance and the corresponding available bit rates used in the local SSIM evaluation.

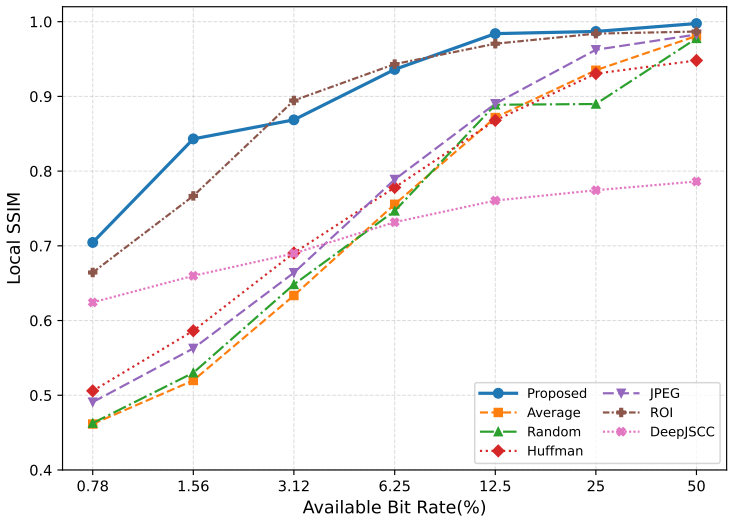


Figure 5. Local SSIM under different available bit rates.

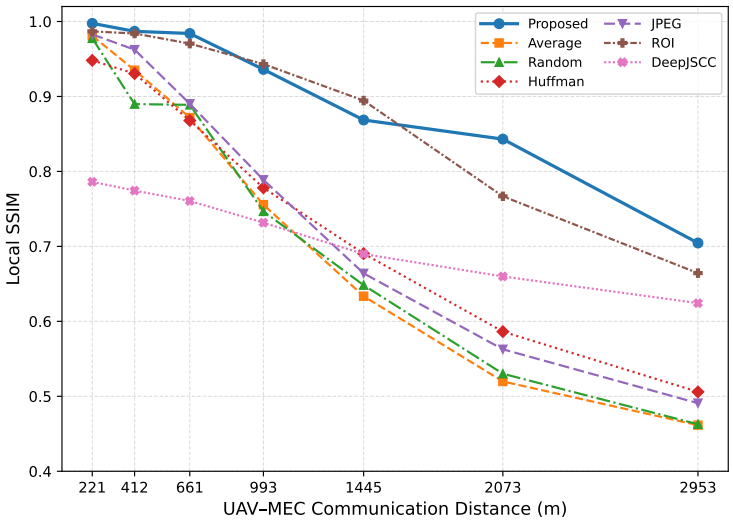


Figure 6. Local SSIM under different UAV–MEC communication distances.

4.3.4. Ablation Study

To analyze the contributions of the main components of the proposed method, ablation experiments are conducted under the same resource budget and evaluation conditions, as summarized in Table 6. The proposed method employs semantic-importance-based differentiated resource allocation and jointly optimizes the grid configuration g and resource-allocation scale vector $\boldsymbol{\eta}$ using CMA-ES. For "w/o Semantic-aware Allocation", average allocation is used instead of semantic-aware differentiated allocation. For "Grid-only Optimization", $\boldsymbol{\eta}$ is fixed and CMA-ES is used only to optimize the grid configuration. For "Allocation-only Optimization", the grid configuration is fixed and CMA-ES is used

only to optimize $\boldsymbol{\eta}$. For "w/o CMA-ES-based Optimization", the CMA-ES-based joint parameter search is removed. The objective function defines the evaluation criterion for candidate configurations during optimization, and replacing it would directly redefine the optimization problem.

Table 6. Ablation study of the main components of the proposed method.

Configuration	Score	mIoU	Local SSIM	SSIM	PSNR (dB)
Proposed Method	**0.8275**	**0.6106**	**0.9840**	0.9631	**38.2888**
w/o Semantic-aware Allocation	0.6247	0.1937	0.8717	0.9585	37.8376
Grid-only Optimization	0.7933	0.5329	0.9771	0.9660	37.7831
Allocation-only Optimization	0.8076	0.5673	0.9782	**0.9673**	37.8891
w/o CMA-ES-based Optimization	0.8009	0.5495	0.9774	0.9628	38.2573

Bold values indicate the best performance.

As shown in Table 6, replacing semantic-aware differentiated allocation with average allocation substantially decreases the Score from 0.8275 to 0.6247 and the mIoU from 0.6106 to 0.1937, together with a noticeable reduction in Local SSIM. This demonstrates the importance of allocating limited resources according to the semantic importance of different image regions for preserving small-target-related information. When only the resource allocation parameters are optimized, the Score and mIoU reach 0.8076 and 0.5673, respectively, whereas optimizing only the grid configuration results in values of 0.7933 and 0.5329. In contrast, jointly optimizing the grid configuration and resource allocation parameters yields the highest Score of 0.8275 and mIoU of 0.6106, indicating that optimizing either of these two groups of variables independently is insufficient to achieve the overall performance of the joint optimization.

Although the Allocation-only Optimization configuration achieves a slightly higher SSIM, its Score, mIoU, Local SSIM, and PSNR remain lower than those of the proposed method, indicating that an improvement in a single reconstruction metric does not necessarily correspond to better overall performance. Without CMA-ES-based joint optimization, the Score and mIoU decrease to 0.8009 and 0.5495, respectively. Overall, semantic-aware differentiated allocation provides the most pronounced contribution to the performance improvement, while the CMA-ES-based joint optimization of the grid configuration and resource allocation parameters further improves the detection performance and overall evaluation result, allowing the proposed method to achieve a favorable balance between small-target information preservation and image reconstruction quality.

4.3.5. Sensitivity Analysis of Semantic-Level Thresholds

To evaluate the influence of the semantic-level partition thresholds on the proposed method, the adopted configuration $(0.65, 0.40, 0.12, 0.020)$ is used as the baseline, and four perturbed configurations are constructed in its neighborhood. Specifically, the lower configuration decreases all thresholds to $(0.60, 0.36, 0.10, 0.015)$, whereas the upper configuration increases them to $(0.70, 0.44, 0.14, 0.025)$. In addition, two mixed perturbation settings, denoted as Mixed-A and Mixed-B, are introduced to avoid restricting the sensitivity analysis to uniform threshold shifts, with values of $(0.68, 0.37, 0.14, 0.018)$ and $(0.62, 0.43, 0.10, 0.022)$, respectively. The number of semantic levels and all other experimental settings remain unchanged, and only the partition thresholds are varied.

Figure 7 presents the variations in mIoU, Local SSIM, SSIM, and PSNR relative to the baseline configuration. The overall fluctuations of all metrics remain limited under different resource conditions. Specifically, the variations in mIoU and Local SSIM are within approximately ± 0.008 and ± 0.004, respectively, while the variation in SSIM is on the order of 10^{-4} and the maximum PSNR variation is approximately 0.04 dB. Moreover, none

of the perturbed configurations consistently exhibits a clear performance advantage or degradation across all resource conditions. These results indicate that moderate variations in the semantic-level thresholds have limited influence on the overall performance, demonstrating that the adopted semantic-level partition has good parameter stability within a reasonable neighborhood.

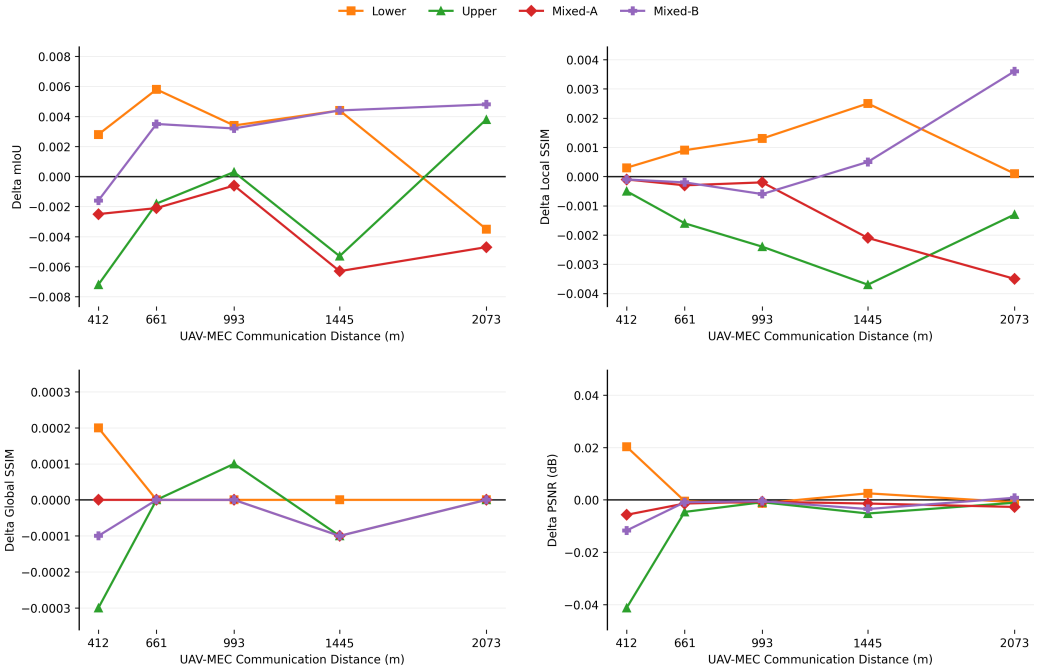


Figure 7. Performance variations under different semantic-level threshold configurations.

5. Conclusions

This paper proposed a semantic-aware image resource allocation method for UAV-based infrared small-target detection under constrained communication resources. By allocating transmission resources according to the task relevance of different image regions, the proposed method prioritizes the preservation of target-related semantic information while satisfying the total bit budget. Experimental results demonstrate that the proposed method achieves a favorable balance among small-target detection performance, local structural preservation, and overall reconstruction quality under different resource budgets. Its superior overall performance over several comparison methods further verifies the effectiveness of the resource allocation strategy based on regional semantic importance.

The current experimental validation is conducted only on the NUDT-SIRST dataset. Therefore, the reported results mainly demonstrate the effectiveness of the proposed method under the data distribution and infrared small-target scenarios represented by this dataset, while its generalization to different imaging conditions, target characteristics, and data distributions remains to be further investigated. Future work will extend the evaluation to additional infrared small-target datasets and more diverse scenarios to further assess the generalization and applicability of the proposed resource allocation strategy.

Future work may investigate more flexible target-aware region partitioning strategies to reduce the influence of fixed grid boundaries on semantic integrity. An end-to-end joint optimization framework may also be developed to integrate semantic importance estimation, resource allocation, and target detection. Moreover, practical communication factors, such as channel fading, bandwidth variation, and UAV resources, will be considered. Further evaluations on real UAV communication platforms will be conducted to improve the adaptability, real-time performance, and engineering applicability of the proposed method.

Author Contributions: Conceptualization and methodology, W.Q. and J.C.; investigation, B.X. and W.L.; formal analysis, data curation and writing—original draft, W.Q.; writing—review and editing, W.Q., J.C. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Open Foundation of Jiangxi Provincial Key Laboratory of Image Processing and Pattern Recognition (ET202404438), Early-Career Young Scientists and Technologists Project of Jiangxi Province (20244BCE52091), Science and Technology Project of Jiangxi Provincial Department of Education (GJJ2401015), Doctoral Start-up Fund of Nanchang Hangkong University (EA202404174), the National Natural Science Foundation of China (62473187) and the High-Tech Field Key R&D Program of Jiangxi Province (20261BCE310029).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cheng, B.; Deng, Q.; Gan, W.; Li, Y.; Wu, C.; Wang, Q. A review of remote sensing image object detection: Advances and challenges in data, techniques, and applications. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2026**, *19*, 10438–10472. [CrossRef]
2. Wang, R.; Lin, C.; Li, Y. RPLFDet: A lightweight small object detection network for UAV aerial images with rational preservation of low-level features. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 5013514. [CrossRef]
3. Zhu, Y.; Ma, Y.; Fan, F.; Huang, J.; Yao, Y.; Zhou, X.; Huang, R. Toward robust infrared small target detection via frequency and spatial feature fusion. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 2001115. [CrossRef]
4. Park, J.; Oh, Y.; Kim, Y.; Jeon, Y.-S. Vision Transformer-based semantic communications with importance-aware quantization. *IEEE Internet Things J.* **2025**, *12*, 35662–35677. [CrossRef]
5. Nguyen, L.X.; Raha, A.D.; Aung, P.S.; Niyato, D.; Han, Z.; Hong, C.S. A contemporary survey on semantic communications: Theory of mind, generative AI, and deep joint source–channel coding. *IEEE Commun. Surv. Tutor.* **2026**, *28*, 2377–2417. [CrossRef]
6. Li, J.; Shi, Y.; Hong, Q.; Jia, Y. A scale-aware multidomain DETR for small object detection in UAV remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4421520. [CrossRef]
7. Hu, L.; Yuan, J.; Cheng, B.; Xu, Q. CSFPR-RTDETR: Real-time small object detection network for UAV images based on cross-spatial-frequency domain and position relation. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5638219. [CrossRef]
8. Ma, S.; Zhang, Y.; Peng, L.; Sun, C.; Ding, L.; Ding, B.; Zhu, Y. OWRT-DETR: A novel real-time Transformer network for small-object detection in open-water search and rescue from UAV aerial imagery. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4205313. [CrossRef]
9. Liu, F.; Yao, L.; Zhang, C.; Wu, T.; Zhang, X.; Jiang, X.; Zhou, J. Boost UAV-based object detection via scale-invariant feature disentanglement and adversarial learning. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5622113. [CrossRef]
10. Chen, Y.; Ye, Z.; Sun, H.; Gong, T.; Xiong, S.; Lu, X. Global–local fusion with semantic information guidance for accurate small object detection in UAV aerial images. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4701115. [CrossRef]
11. Luo, M.; Zhao, R.; Zhang, S.; Chen, L.; Shao, F.; Meng, X. IM-CMDet: An intramodal enhancement and cross-modal fusion network for small object detection in UAV aerial visible-infrared imagery. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5008316. [CrossRef]
12. Guo, T.; Zhou, B.; Luo, F.; Zhang, L. Infrared small target detection via diverse feature harmonization. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5004513. [CrossRef]
13. Xiong, Z.; Zhou, F.; Wu, F.; Yuan, S.; Fu, M.; Peng, Z.; Yang, J.; Dai, Y. DRPCA-Net: Make robust PCA great again for infrared small target detection. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5005516. [CrossRef]
14. Li, B.; Xiao, C.; Wang, L.; Wang, Y.; Lin, Z.; Li, M.; An, W.; Guo, Y. Dense nested attention network for infrared small target detection. *IEEE Trans. Image Process.* **2023**, *32*, 1745–1758. [CrossRef] [PubMed]
15. Zhang, Y.; Li, Z.; Siddique, A.; Azeem, A.; Chen, W.; Cao, D. Infrared small target detection based on interpretation weighted sparse method. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5001415. [CrossRef]
16. Jiang, Z.; Chen, S.; Tian, C.; Li, X.; Yue, Y. CRPE-Net: Infrared small target detection Transformer with cross-layer relative-position embedding. *Pattern Recognit.* **2026**, *172*, 112640. [CrossRef]
17. Zuo, J.; Pei, S.; Li, Q.; Huang, Y.; Wang, S. DENet: Dual-path edge network with global–local attention for infrared small target detection. *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 5605615. [CrossRef]
18. Zhou, K.; Zhang, G.; Cai, Y.; Hu, Q.; Yu, G.; Swindlehurst, A.L. Feature allocation for semantic communication with space–time importance awareness. *IEEE Trans. Wirel. Commun.* **2025**, *24*, 8847–8862. [CrossRef]

19. Guo, S.; Wang, Y.; Ye, J.; Zhang, A.; Zhang, P.; Xu, K. Semantic importance-aware communications with semantic correction using large language models. *IEEE Trans. Mach. Learn. Commun. Netw.* **2025**, *3*, 232–245. [CrossRef]
20. Hu, H.; Zhu, X.; Zhou, F.; Wu, W.; Hu, R.Q.; Zhu, H. Resource allocation for multi-modal semantic communication in UAV collaborative networks. *IEEE Trans. Commun.* **2025**, *73*, 7599–7616. [CrossRef]
21. Park, J.; Yoon, S.W. Transmit what you need: Task-adaptive semantic communications for visual information. *IEEE J. Sel. Areas Commun.* **2025**, *43*, 4182–4197. [CrossRef]
22. Peng, X.; Qin, Z.; Tao, X.; Lu, J.; Letaief, K.B. A robust image semantic communication system with multi-scale Vision Transformer. *IEEE J. Sel. Areas Commun.* **2025**, *43*, 1278–1291. [CrossRef]
23. Hu, H.; Song, K.; Fan, R.; Zhan, C.; Xu, J.; Yang, J. Energy-efficient image semantic communication: Architecture design and optimal joint allocation of communication and computation resources. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 6466–6480. [CrossRef]
24. Li, P.; Wang, Y.; Tang, X.; Liu, M.; Liu, H.; Cheng, J. Resource allocation for image transmission using adaptive semantic and bit communication. *IEEE Trans. Wirel. Commun.* **2026**, *25*, 3526–3540. [CrossRef]
25. Wang, Q.; Ye, J.; Qian, L.P.; Jiang, W.; Yang, Q.; Liang, Y.-C.; Kam, P.-Y. Adaptive semantic compression and transmission with joint resource allocation optimization for multi-user image classification. *IEEE Trans. Mob. Comput.* **2026**, *25*, 11694–11709. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.