Robust Unsupervised Multifeature Representation for Infrared Small Target Detection

Liqiong Chen, Tong Wu, Shuyuan Zheng, Zhaobing Qiu, and Feng Huang

***Abstract*—Infrared small target detection is critical to infrared search and tracking systems. However, accurate and robust detection remains challenging due to the scarcity of target information and the complexity of clutter interference. Existing methods have some limitations in feature representation, leading to poor detection performance in complex scenes. Especially when there are sharp edges near the target or in cluster multitarget detection, the "target suppression" phenomenon tends to occur. To address this issue, we propose a robust unsupervised multifeature representation (RUMFR) method for infrared small target detection. On the one hand, robust unsupervised spatial clustering (RUSC) is designed to improve the accuracy of feature extraction; on the other hand, pixel-level multiple feature representation is proposed to fully utilize the target detail information. Specifically, we first propose the center-weighted interclass difference measure (CWIDM) with a trilayer design for fast candidate target extraction. Note that CWIDM also guides the parameter settings of RUSC. Then, the RUSC-based model is constructed to accurately extract target features in complex scenes. By designing the parameter adaptive strategy and iterative clustering strategy, RUSC can robustly segment cluster multitargets from complex backgrounds. Finally, RUMFR that fuses pixel-level contrast, distribution, and directional gradient features is proposed for better target representation and clutter suppression. Extensive experimental results show that our method has stronger feature representation capability and achieves better detection performance than several state-of-the-art methods.**

***Index Terms*—Infrared small target detection, pixel-level multifeature representation, robust unsupervised spatial clustering (RUSC), "target suppression" phenomenon.**

I. Introduction

As A key technology of infrared search and tracking (IRST) systems, infrared (IR) small target detection plays an important role in many applications such as missile interception and battlefield reconnaissance [1], [2], [3], [4]. For early warning at a long distance, IR small targets usually occupy few pixels and lack texture detail [5], [6], [7]. In addition, IR small targets are easily submerged in complex backgrounds, leading to the "target suppression" phenomenon and lower detection performance. As shown in Fig. 1, real targets are significantly suppressed by existing methods when they are near sharp edges or in cluster multitarget detection. Therefore, it remains challenging to design an excellent detection method that can robustly cope with the above difficulties.

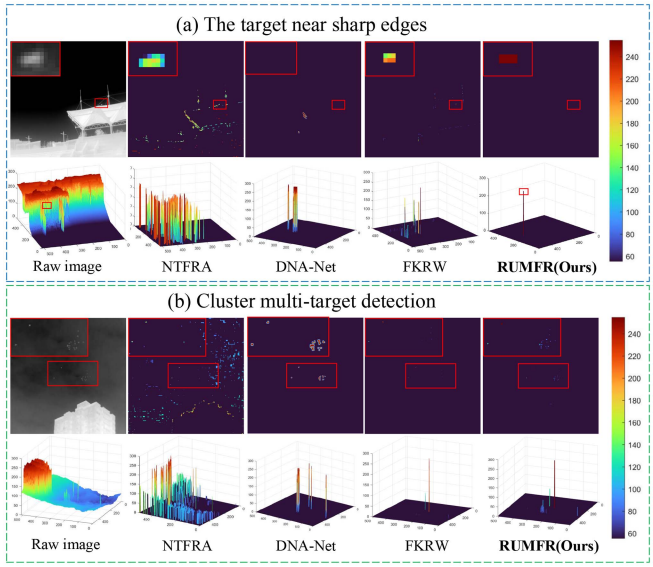


Fig. 1. Phenomenon of "target suppression": (a) the target near sharp edges, (b) cluster multitarget detection. The red rectangle represents real targets.

Over the past decades, many methods have been researched to detect small targets from infrared images. In general, they can be roughly divided into four categories, namely conventional filtering (CF)-based methods, sparse and low-rank representation (SLR)-based methods, neural network (NN)-based methods, and local feature representation (LFR)-based methods. CF-based methods design image filtering operations to suppress the background and achieve good detection performance in simple scenes. However, filters such as Max-Mean [8] have limited feature extraction capability to capture target and background differences, leading to poor performance in suppressing complex interference such as high-intensity clutter. For better background suppression, SLR-based methods model the target and background as sparse and low-rank matrices, respectively, and solves the final detection results using component analysis. However, low-rank matrices fail to accurately represent sparse interference such as pixel-sized noise with high brightness (PNHB [9]). As shown in Fig. 1, one of the latest SLR-based methods (NTFRA [10]) shows many false detections with sparse interference.

Manuscript received 9 February 2024; revised 9 April 2024; accepted 24 April 2024. Date of publication 8 May 2024; date of current version 30 May 2024. This work was supported in part by the Nature Science Foundation of Fujian Province under Grant 2022J05113, and in part by the Educational Research Program for Young and Middle-aged Teachers of Fujian Province under Grant JAT210035. *(Corresponding authors: Zhaobing Qiu; Feng Huang.)*

The authors are with the School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China (e-mail: liqiongchen@fzu.edu.cn; 230227130@fzu.edu.cn; 220220067@fzu.edu.cn; qiuzhaobing@fzu.edu.cn; huangf@fzu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3398361

Moreover, the solution process contains large-scale matrix operations and multiple iterative optimizations, leading to high complexity and poor real-time performance. With massive data and optimized models, NN-based methods have achieved impressive representation capabilities and detection performance [11]. However, they are not robust enough to detection scenarios that lack real data, limiting their wider application. For example, one state-of-the-art NN-based method (DNA-Net [12]) still shows many missed and misdetected targets in Fig. 1. Recently, LFR-based methods have attracted a lot of attention due to their higher efficiency and robustness. They achieve detection by measuring the difference between the target and local backgrounds (LBs), which improves feature representation by considering both target and background information comprehensively. Moreover, operations on local areas can achieve higher efficiency and wider applications. However existing methods still have some limitations in feature representation, which makes it difficult to cope with complex detection scenarios. For example, the recent LFR-based method (FKRW [13]) achieves higher feature extraction accuracy using supervised clustering. However, as a supervised method, FKRW requires a preset number and class of clusters, which makes it difficult to extract target features from the complex background shown in Fig. 1, leading to the phenomenon of "target suppression."

To address the above issues, we propose a robust unsupervised multifeature representation (RUMFR) method for infrared small target detection. As shown in Fig. 1, RUMFR can effectively solve the "target suppression" problem and remove almost all the clutter interference. Its main motivation is to design a robust unsupervised spatial clustering (RUSC)-based model to improve the feature extraction accuracy, while fusing multiple pixel-level features to fully utilize the target detail information. The proposed method consists of four main steps. First, the center-weighted interclass difference measure (CWIDM) with a trilayer design is introduced for fast candidate target extraction. Note that CWIDM also provides reference for parameter settings of RUSC. Then, we propose RUSC-based model to accurately extract target features in complex scenes. By designing parameter adaptive strategy and iterative clustering strategy, the proposed RUSC can robustly segment an arbitrary number and distribution of cluster multitargets from complex backgrounds. After that, RUMFR that fuses pixel-level contrast, distribution, and directional gradient features is proposed to further enhance the target and suppress the background. Thanks to RUSC, RUMFR can effectively capture the detailed features of targets with high accuracy, and thus is more robust to complex scenarios (e.g., cluster multitarget detection). Finally, we detect real targets from the RUMFR enhanced map via threshold segmentation. Extensive experiments demonstrate that our method effectively solves the "target suppression" problem and achieves better detection performance than several state-of-the-art methods. In this article, our main contributions can be summarized as follows:

1) In order to improve the accuracy of feature extraction in complex scenes, we design the robust clustering method (RUSC) and propose a RUSC-based model.
2) For better balance of efficiency and effectiveness, the CWIDM with a trilayer design is proposed to quickly extract candidate targets.
3) In order to effectively enhance cluster multitargets and suppress clutter interference, we propose RUMFR that fuses pixel-level multifeatures.

The rest of this article is organized into five sections. Section II briefly describes the related work, including existing detection methods, unsupervised clustering, and density-based spatial clustering of applications with noise (DBSCAN [14]), which is the basis of RUSC. In Section III, we describe the robust unsupervised clustering method RUSC in detail. Section IV presents the implementation details of the RUMFR detection method. In Section V, a large number of experiments are implemented to verify the effectiveness of our method. Finally, Section VI concludes the article.

II. Related Work

In this section, we first introduce the research status of IR small target detection, then present several classical unsupervised clustering methods, and finally describe the principle of DBSCAN, which is the basis of RUSC.

A. Infrared Small Target Detection

As mentioned above, existing detection methods can be broadly categorized into four groups based on the detection principle: CF-based methods, SLR-based methods, NN-based methods, and LFR-based methods. As the earliest detection method, CF-based methods design image filtering operations to achieve target enhancement and background suppression. CF-based methods contain two main branches: spatial domain filtering and frequency domain filtering. The former exploits the spatial continuity of the background and designs a spatial filter to suppress the background by directly processing the raw image. Many classical algorithms such as max-mean/max-median filter [8], bilateral filter [15], and guided filter [16] belong to this category. They are simple and effective, but difficult to represent target and high-intensity background differences, and thus not suitable for complex scenes. The latter transforms the image to the frequency domain and designs a high-pass filter to suppress low-frequency background signals. Due to the limited feature representation of traditional Fourier transforms, Qi et al. [17] utilized the phase spectrum of quaternion Fourier transform to enhance Gaussian-like shape targets. In order to obtain more spectral information, many methods based on wavelet transform have been developed [18], [19], [20], [21], [22]. For example, Xin et al. [21] used the Gabor wavelet transform to extract features from different scales and angles for better background suppression. These methods assume that the target and background have high-frequency and low-frequency characteristics, respectively, and thus have more false alarms when subjected to heavy noise and complex background interference.

SLR-based methods believe that the infrared image can be decomposed into a sparse matrix representing the target and a low-rank matrix representing the background, and thus the

final target image can be obtained by component analysis. Gao et al. [23] first proposed an infrared patch image (IPI) model to obtain the target image by principle component pursuit, which achieves better background suppression than CF-based methods. Subsequently, many researchers develop improved models based on IPI [24], [25], [26]. For example, Dai et al. [24] exploited total variation regularization to suppress strong edges; Zhang et al. [26] utilized $l_{2,1}$-norm to further reduce the clutter residuals.; Zhang et al. [27] applied l_p-norm constraint to construct sparser target images. To reduce the computational complexity, the tensor is used to extend the data dimensionality [28], [29], [30]. For example, Dai et al. [31] proposed infrared patch-tensor model based on spatial correlation to suppress remaining edges. Subsequently Zhang et al. [32] proposed partial sum of the tensor nuclear norm (PSTNN) to better protect small targets. To improve the flexibility of the model and enhance noise suppression, Kong et al. [10] proposed nonconvex tensor fibered rank approximation (NTFRA). In general, SLR-based methods achieve significant improvement in background suppression. However, the low-rank matrix struggles to represent sparse interferences such as PNHB, leading to a number of false alarms. In addition large-scale matrix decomposition and reconstruction has large computational complexity, which limits their practical application.

NN-based methods achieves target detection in a data-driven manner. For better feature representation, some researchers design customized network frameworks for IR small targets. For example, denoising autoencoder network [33] achieve good background suppression by treating small targets as noise. In order to improve the detection performance, Wang et al. [34] decomposed the infrared target detection problem into two opposing subproblems of missed detection and false alarm and proposed conditional generative adversarial network (CGAN). In order to preserve the target features, Li et al. [12] proposed dense nested attention network (DNA-Net) for progressive interaction between high-level and low-level features. To capture more target features, Sun et al. [35] proposed receptive-field and direction-induced attention network (RDIAN). To alleviate data scarcity, some researchers attempt to augment data [36], [37] or fuse traditional feature representations [38], [39], [40], [41]. For the space-based infrared tiny ship detection, Wu et al. [42] proposed multilevel TransUNet (MTU-Net) that achieves better multilevel feature extraction. In order to fully utilize local semantics and contextual relevance, Zhang et al. [3] proposed the attention-guided pyramid context network (AGPCNet) to perceive pixel correlations within and between patches at a specific scale. In order to balance the model accuracy and computational efficiency, Kou et al. [43] proposed the lightweight IR small target segmentation network (LW-IRSTNet), which optimizes the feature extraction structure. NN-based methods are developing rapidly and achieving impressive results in specific scenarios. However, they are not robust enough to detection scenarios that lack real data, limiting their wider application.

Recently, LFR-based methods have received a lot of attention for their balance of effectiveness and efficiency [5], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53]. Their main motivation is to achieve detection by representing the difference in features between IR small targets and their LBs. Chen et al. [54] proposed local contrast measure (LCM) based on a 3×3 nested structure to represent the difference in saliency between the target and LBs. Since LCM is sensitive to PNHB, Han et al. [9] proposed to represent the contrast feature using mean operation. In order to better represent the unknown scale target, Wei et al. [55] proposed a multiscale nested structure. Since the multiscale model is difficult to accurately extract arbitrary rectangular target features, an adaptive scale model [56] is proposed in our earlier work. To better extract features from arbitrarily shaped targets, Huang et al. [57] proposed density peaks searching and maximum-gray region growing (DPSMRG). Since the region growth algorithm is sensitive to noise, Qin et al. [13] applied random walker (RW) [58] and proposed facet kernel and random walker (FKRW). As a supervised clustering, RW relies on the initial label setting, which reduces the detection accuracy in complex backgrounds. Therefore, Chen et al. [59] proposed improved fuzzy C-means (IFCM) for better feature extraction from chaotic clutter. However, IFCM still needs to initialize the number of clusters and thus cannot effectively handle complex scenarios such as cluster multitarget detection. In our previous work, improved DBSCAN (IDBSCAN) [44] is proposed to effectively enhance cluster multitargets. However, the customized feature extraction model is difficult to adapt to different detection scenarios, while the limited feature representation capability limits further improvement of detection performance. Some other researchers have attempted to introduce more features to enhance the feature representation capability, such as scale [52], gradient [45], and weighted information [60]. However, limited by the accuracy of feature extraction, they are prone to cause more missed detections while reducing false detections. In summary, designing a robust and accurate feature representation is the key to further improve the detection performance.

B. *Unsupervised Clustering*

The task of unsupervised clustering is to classify a set of samples into several different clusters without labeled samples, where any sample is more similar to samples in the same cluster than to samples in different clusters. Depending on the principle, common unsupervised clustering algorithms mainly include centroid-based methods (e.g., fuzzy C-means [61]), connectivity-based methods (e.g., hierarchical clustering [62]), and distribution-based methods (e.g., Gaussian mixture model clustering [63]). However, these methods are sensitive to noise interference in IR images. For example, they usually misclassify PNHBs as target or background clusters, leading to large feature extraction errors. Moreover, they typically require specific prior knowledge or consume a large amount of computation to determine the number of clusters in the database, which limits their wider application in complex scenarios such as cluster multitarget detection.

Another representative unsupervised clustering method is density-based spatial clustering (e.g., DBSCAN). It has better robustness by classifying the noise as outliers. In addition, it can efficiently adapt to the database with unknown number of clusters. Therefore, DBSCAN is a good candidate for solving the

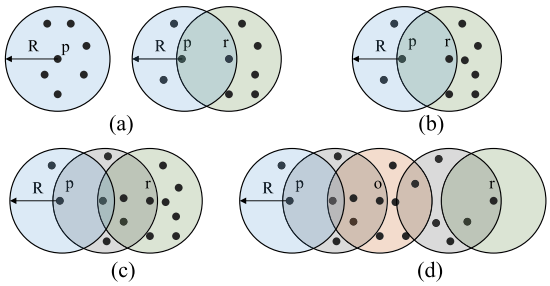


Fig. 2. Schematic diagram of DBSCAN definition ($\varepsilon = R$, MinPts=5). (a) Core sample and border sample. (b) Directly density-reachable. (c) Density-reachable. (d) Density-connected.

difficulty of feature extraction in infrared small target detection. Nevertheless, DBSCAN has the disadvantage of being quite sensitive to parameter settings, which leads to poor robustness for complex detection scenarios. To address this challenge, Section III of this article designs a robust unsupervised clustering method called RUSC based on DBSCAN.

C. DBSCAN

The main motivation of DBSCAN is to achieve clustering by setting parameters ε and *MinPts* to describe the closeness of sample distribution. Specifically, DBSCAN believes that for each sample point of a cluster, the neighborhood of a given radius ε should contain at least *MinPts* sample points. The DBSCAN is implemented based on the following definition.

Definition 1 (ε-Neighborhood): For a sample $\mathbf{p}$, the ε-neighborhood denotes the set of all samples within the hypersphere with its center and ε as its radius, defined as follows:

$$N_\varepsilon(\mathbf{p}) = \{\mathbf{r} \in \mathrm{S} \mid d(\mathbf{p}, \mathbf{r}) \leq \varepsilon\} \tag{1}$$

where S denotes the sample set and d denotes the distance function, such as Euclidean distance [64].

Definition 2 (Core Sample and Border Sample): As shown in Fig. 2(a), a sample $\mathbf{p}$ is a core sample if $|N_\varepsilon(\mathbf{p})| \geq$ *MinPts*; a sample $\mathbf{p}$ is a boundary sample if $|N_\varepsilon(\mathbf{p})| <$ *MinPts* and $\mathbf{p} \in N_\varepsilon(\mathbf{r})$, where $\mathbf{r}$ is a core sample.

Definition 3 (Directly Density-Reachable)): As shown in Fig. 2(b), a sample $\mathbf{p}$ is directly density-reachable from a sample $\mathbf{r}$ with respect to ε and *MinPts*, if $\mathbf{p} \in N_\varepsilon(\mathbf{r})$ and $\mathbf{r}$ is a core sample.

Definition 4 (Density-Reachable): As shown in Fig. 2(c), a sample $\mathbf{p}$ is density-reachable from a sample $\mathbf{r}$ with respect to ε and *MinPts*, if there is a chain of samples $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ with $\mathbf{p}_1 = \mathbf{r}$ and $\mathbf{p}_n = \mathbf{p}$ such that $\mathbf{p}_{i+1}$ is density-reachable from $\mathbf{p}_i$.

Definition 5 (Density-Connected): As shown in Fig. 2(d), a sample $\mathbf{p}$ is density-connected to a sample $\mathbf{r}$ with respect to ε and *MinPts*, if there is a sample $\mathbf{o}$ such that both $\mathbf{p}$ and $\mathbf{r}$ is density-reachable from $\mathbf{o}$.

Definition 6 (Cluster): A cluster C_i with respect to ε and *MinPts* is a nonempty subset of S which satisfies the following conditions:

Algorithm 1: The DBSCAN Algorithm.

Input: the sample set S, parameters ε and *MinPts*
Output: *ClassID* of each sample in S
1: Construct the distance matrix $\mathbb{D}$ of all samples from S;
2: Solve ε-Neighborhood of each sample base on Def. 1 with $\mathbb{D}$;
3: Select and label core samples according to Def. 2;
4: Identify density-connected samples using Defs. 3-5;
5: Label *ClassID* of each sample base on Defs. 6 and 7.

1) $\forall \mathbf{p}, \mathbf{r}$: if $\mathbf{p} \in C_i$ and $\mathbf{r}$ is density-reachable from $\mathbf{p}$ with respect to ε and *MinPts*, then $\mathbf{r} \in C_i$ (Maximality).
2) $\forall \mathbf{p}, \mathbf{r} \in C_i$: $\mathbf{p}$ is density-connected to $\mathbf{r}$ with respect to ε and *MinPts* (Connectivity).

Definition 7 (Outlier Sample): A sample $\mathbf{p}$ is a outlier sample if it is neither a core sample nor a border sample. This implies that outlier sample does not belong to any cluster.

Based on the above definition, the implementation process of DBSCAN can be summarized as Algorithm 1. Specifically, it first chooses an arbitrary sample from the sample set S that satisfies the core sample condition. Then, it retrieves all samples density-reachable from the core samples and determines all samples density connected to each other as one cluster.

III. Robust Unsupervised Spatial Clustering

Although the classical DBSCAN algorithm can adapt to arbitrarily distributed data and has good noise immunity, it is parameter-sensitive and difficult to be applied to complex detection scenarios. Therefore, in order to accurately extract clustered multitargets from complex backgrounds, the RUSC is designed in this section. Specifically, we first set the range of clustering region according to the IR small target detection task. Then, the distance function is redefined according to the target characteristics. Finally, to improve the accuracy and robustness of clustering, we propose parameter adaptive strategy and iterative clustering strategy.

A. Clustering Region Setting

Since DBSCAN is sensitive to parameter settings, it is difficult to get satisfactory clustering results by processing the whole infrared image with the same parameters. In addition, clustering all pixels in an image simultaneously faces a large computational burden, limiting wider applications. Thus, in this article, we propose to perform clustering in a local region of $\mathrm{L_p} \times \mathrm{L_p}$ pixels centered on each candidate target.

B. Distance Function Redefinition

The distance function of a single metric has a limited ability to characterize the differences between the target and the background. Consider that IR small targets are heterogeneous and compact, i.e., their own pixels are spatially connected to each other and have similar intensities, while there is usually a large intensity difference with the surrounding background. Thus, by combining spatial and intensity features, we redefine

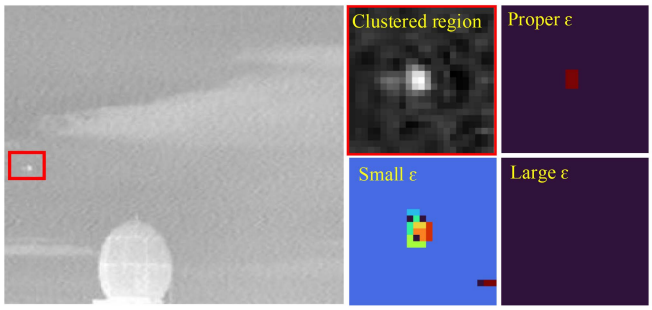


Fig. 3. Cluster results with different ε values set in the target local area. (a) Raw image. (b) Clustering results with different ε.

the distance function of pixels $\mathbf{p}_i$ and $\mathbf{p}_j$ as

$$d\left(\mathbf{p}_i, \mathbf{p}_j\right) = Th(\boldsymbol{\phi}\left(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j\right)) \times \left(\boldsymbol{\phi}\left(I_i, I_j\right) + 1\right) \tag{2}$$

where $\boldsymbol{\theta}_i$ and I_i denote the spatial position vector and intensity of pixel i, respectively. $\boldsymbol{\phi}$ denotes the squared Euclidean distance function, and Th denotes the spatial threshold function, which is defined as follows:

$$Th(\boldsymbol{\phi}\left(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j\right)) = \begin{cases} 1, & \boldsymbol{\phi}\left(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j\right) \leq d_L \\ \infty, & \text{otherwise} \end{cases} \tag{3}$$

where d_L controls the distance between associated pixels. Since infrared small targets have compactness [13], we set d_L to 2, i.e., pixels in the eight neighborhoods may belong to the same category. To prevent the spatial property of small targets from being ignored, the intensity feature between pixels is represented as $(\boldsymbol{\phi}(I_i) + 1)$.

With the above distance function, one target pixel has a small intraclass distance from pixels of the same target while having a large interclass distance from pixels of the background or different targets.

C. Parameter Adaptive Strategy

The parameter *MinPts* in DBSCAN determines the minimum number of samples contained in a single cluster. Since an IR small target may contain as few as 2×1 pixels [55], too large a *MinPts* will lose the real target. Thus, we propose the strategy of fixing parameter *MinPts* to 2 and adaptively adjusting parameter ε.

Parameter ε determines the neighborhood range of clustering. As shown in Fig. 3, too large ε tends to classify the target and the background into the same category, while too small ε may misclassify a single target into multiple categories, both of which will lead to poor detection performance. The complex and diverse detection backgrounds make the optimal ε parameter settings highly variable. Thus, we adaptively set the initial clustering parameter $\varepsilon^{(0)}$ as

$$\varepsilon^{(0)} = \delta_1 \times \text{CWIDM} \tag{4}$$

where δ_1 is an empirical parameter and CWIDM denotes the interclass distance estimate, which will be introduced in Section IV-A. Since the value of parameter ε should be smaller than the interclass distance between the target and the background, δ_1 is recommended to be set in the interval [0.8, 1]. In the experiments, we set δ_1 to 0.8.

D. Iterative Clustering Strategy

Although setting ε with the adaptive initial parameter $\varepsilon^{(0)}$ can handle most scenarios, it is difficult to cope with complex backgrounds due to the estimation error of CWIDM and the setting error of the empirical parameter δ_1. To address this issue, this section proposes an iterative clustering strategy to further fine-tune the parameter ε. Specifically, we set the parameter ε as follows:

$$\varepsilon = \varepsilon^{(n)} \tag{5}$$

where n denotes the number of iterative clustering and $\varepsilon^{(n)}$ denotes the result of the nth iteration, which is iteratively updated according to (6) until the clustering completion condition is satisfied:

$$\varepsilon^{(n)} = \delta_2 \times \varepsilon^{(n-1)} \tag{6}$$

where δ_2 controls the iteration speed. The larger δ_2 is, the faster the algorithm converges and the easier it is to miss the optimal solution. Therefore, under the requirement of real-time performance, the smaller δ_2 is, the better detection accuracy can be obtained. In the experiment, we set δ_2 to 0.3.

The key to iterative clustering is the appropriate completion condition setting. We define C to denote the clustering result, v_i to denote the ith pixel, and $C(i)$ to denote the category of pixel v_i after clustering. In the ideal clustering result, the number of pixels in small target clusters should be less than 81 [55], thus **Completion Condition 1** of iterative clustering is defined as

$$\text{NUM}\left(\{v_i \mid C(i) == C(t), i \in [1, N]\}\right) \leq 81 \tag{7}$$

where the tth pixel corresponds to the target pixel in the cluster region center, and NUM denotes the number statistics function. When the clustered region is background clutter, the number of target pixels may always be greater than 81 and the clustering will iterate infinitely. Thus, we define the **Completion Condition 2** of iterative clustering as follows:

$$\text{NUM}\left(\textit{unique}\left(\{C(i) \mid i \in [1, N]\}\right)\right) \geq N_c \tag{8}$$

where *unique* denotes the function to remove duplicate elements and N_c denotes the expected maximum number of clustering categories, which is set to 5. When the above conditions are satisfied, most of the clustered regions can be effectively segmented. However, the uniform background region itself has only one category, thus **Completion Condition 3** is defined as

$$\varepsilon \leq N_\varepsilon \tag{9}$$

where N_ε controls the minimum interclass distance. Since small infrared targets are usually brighter than the surrounding background, N_ε is recommended to be set in the interval [20, 50]. In the experiments, we set N_ε to 30.

In summary, by iteratively updating ε according to (6) until stopping when either of the completion conditions in (7)–(9) is satisfied, we can obtain accurate and robust clustering results

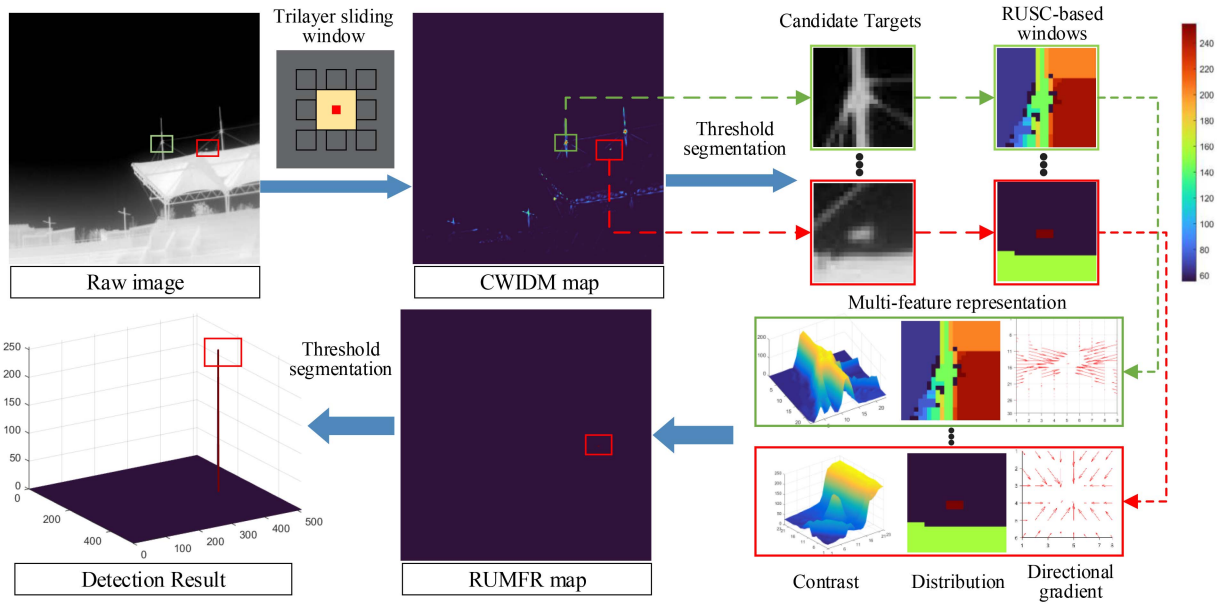


Fig. 4. Framework of proposed RUMFR method. The red rectangle represents the real target and the green rectangle represents typical clutter.

C. Assuming that C contains $c+1$ categories (i.e., $0, 1, \ldots, c$), then the cluster of category 0 denotes the isolated boundary sample.

IV. Proposed RUMFR Detection Method

In this section, we present a concrete implementation of the RUMFR detection method. As shown in Fig. 4, we first construct the CWIDM map to extract candidate targets and estimate the initial parameters of RUSC. Then, the RUSC-based model is constructed to accurately extract target features in complex scenes. After that, RUMFR fusing multiple features is computed to enhance the target and suppress clutter interference. Finally, threshold segmentation is utilized to detect real targets from the RUMFR map.

A. Candidate Target Extraction and Parameter Estimation

In order to efficiently extract candidate targets and guide the parameter settings of RUSC, we design the following CWIDM for target enhancement and background suppression.

First, to better represent the target and LBs, we design a trilayer sliding window structure as shown in Fig. 5. Specifically, the window consists of three parts: LB layer, TA layer, and TC layer, where LB layer consists of eight patches (i.e., B1, B2 $\cdots$ B8). In order to mitigate interference from sharp edges around the target, we fix the patch Bi as 3×3 pixels and fix the TC layer as one pixel. For better adaptation to scale variations, the TA layer is set to multiple scales. Note that as with most existing methods [54], [55], [65], this article focuses on targets with sizes between 2×1 and 9×9 pixels. Therefore, the TA layer is set to four scales of 3×3, 5×5, 7×7, and 9×9.

Then, we calculate the CWIDM for each pixel to enhance the target and suppress the background. Specifically, construct a trilayer window centered on each pixel, one by one, and represent

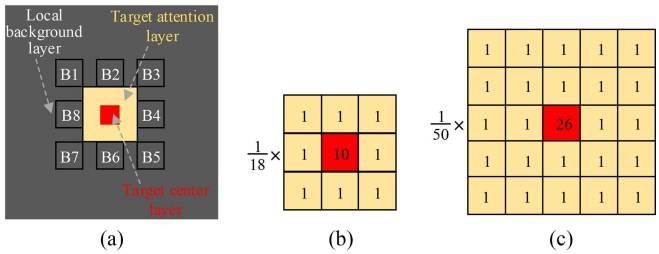


Fig. 5. Structure schematic of CWIDM. (a) Trilayer sliding window. (b) 3×3 mask for m_{T}. (c) 5×5 mask for m_{T}.

the target intensity with the following equation:

$$m_{\mathrm{T}} = \frac{1}{2}\left(m_{\mathrm{TA}} + m_{\mathrm{TC}}\right) \tag{10}$$

where m_{TA} and m_{TC} denote the average intensity of TA and TC layer. Compared to existing methods, m_{T} can better preserve target intensity features while suppressing noise. In addition, m_{T} can be efficiently calculated by convolving the image using masks as shown in Fig. 5. The intensity difference between the target and LBs can be expressed as

$$\mathrm{ID}_i = m_{\mathrm{T}} - m_{\mathrm{Bi}},\ i \in [1, 8] \tag{11}$$

where m_{Bi} denotes the average intensity of Bi. Thus, the interclass intensity difference between the target and surrounding background for scale L can be estimated as

$$\mathrm{CWIDM_L} = \min_{i=1,2,3,4}\left\{\mathrm{ID}_i \times \mathrm{ID}_{i+4}\right\}. \tag{12}$$

For better scale invariance, we define the final interclass distance estimation as follows:

$$\mathrm{CWIDM} = \max_{L=3,5,7,9}\left\{\mathrm{CWIDM_L}\right\}. \tag{13}$$

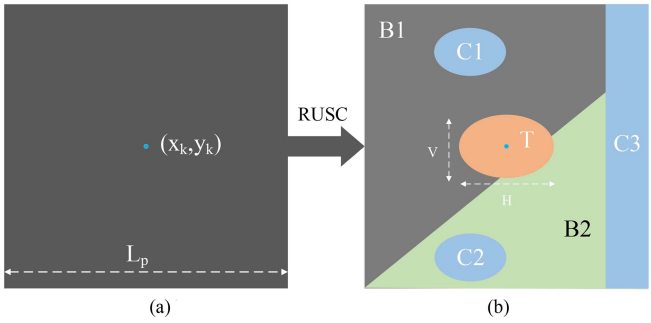


Fig. 6. Schematic diagram of RUSC-based feature extraction model. T denotes the center target, B1 and B2 denote the surrounding background, and C1-C3 denote the nearby clutter. (a) Initial stage. (b) Completion stage.

In the CWIDM map, the target is effectively enhanced with a higher intensity than the background. In addition, CWIDM can guide the parameter settings of RUSC in (4).

Finally, candidate targets can be extracted from the CWIDM map via the following threshold segmentation:

$$\mathrm{S_{CT}} = \{(x, y, \varepsilon) \mid \mathrm{CWIDM}(x, y) \geq \mathrm{Th_{CT}}\} \tag{14}$$

where $\mathrm{S_{CT}}$ denotes the candidate target set including coordinates (x, y) and clustering parameter ε, and $\mathrm{Th_{CT}}$ is the adaptive threshold, which is defined as follows:

$$\mathrm{Th_{CT}} = \min\{\sigma \times \mathrm{M}_1, \mathrm{M}_k\} \tag{15}$$

where M_1 and M_k denote the grayscale value of the brightest and kth brightest pixel in the CWIDM map, respectively. $\mathrm{Th_{CT}}$ can preserve at least k candidate targets, and thus is more robust to multitarget detection and complex background interference. The empirical parameters σ and k are set to 0.6 and 30 in the experiment, respectively.

B. RUSC-Based Feature Extraction Model

As shown in Fig. 6, the RUSC-based model is constructed to accurately extract target features from complex backgrounds. At the initial stage, a $\mathrm{L_p} \times \mathrm{L_p}$ local region is constructed centered on one candidate target (x_k, y_k). $\mathrm{L_p}$ controls the clustering region. Since the size of IR small targets usually ranges from 2×1 pixels to 9×9 pixels [9], $\mathrm{L_p}$ needs to be set larger than 9. If $\mathrm{L_p}$ is set too small, it may categorize individual small targets as background clutter, reducing the detection rate; if $\mathrm{L_p}$ is set too large, it will increase the amount of computation and reduce real-time performance. Experimental results show that the algorithm maintains good detection performance when $\mathrm{L_p}$ is set in a wide range from 19 to 29. Thus, we set $\mathrm{L_p}$ to 23 in the experiments.

When RUSC has been performed in the local region, the segmentation result can be obtained as shown in Fig. 6(b). Where T denotes the center target, Bk denotes the surrounding background, and Ck denotes the nearby clutter.

As described in Section I, nearby clutter (e.g., sharp edges and cluster targets) degrades the accuracy of feature extraction and leads to the "target suppression" problem. Therefore, it is important to distinguish between the surrounding background and nearby clutter for RUSC-based models. Based on observations and studies, they are found to exhibit large differences in spatial distribution. Specifically, the surrounding background Bk is usually adjacent to the center target and thus satisfy

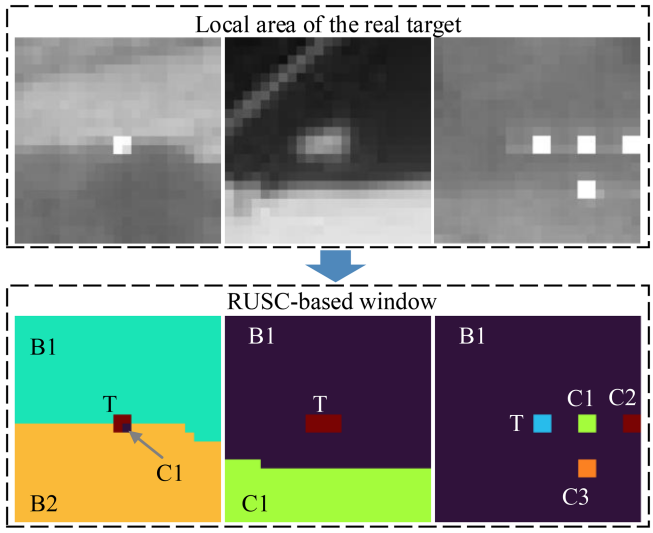


Fig. 7. RUSC-based window structure for real IR targets. T denotes the center target, B1 and B2 denote the surrounding background, and C1-C3 denote the nearby clutter.

$$\min_{\mathbf{p}_i \in \mathrm{Bk}, \mathbf{p}_j \in \mathrm{T}} \{d(\mathbf{p}_i, \mathbf{p}_j)\} \leq 1 \tag{16}$$

where d denotes the Euclidean distance. In addition, the intensity of the background usually changes slowly and has a large distribution area (e.g., greater than 9×9) [9]. Therefore, Bk should also satisfy the condition

$$\mathrm{H_{Bk}} > 9 \parallel \mathrm{W_{Bk}} > 9 \tag{17}$$

where $\mathrm{H_{Bk}}$ and $\mathrm{W_{Bk}}$ denote the maximum pixel distance of Bk in the vertical and horizontal directions, respectively.

As shown in Fig. 6(b), clusters (B1 and B2) that satisfy the above two conditions are categorized as the surrounding background, while cluster targets (C1 and C2) as well as sharp edges (C3) are identified as nearby clutter. To verify the effectiveness of RUSC-based model, Fig. 7 shows the actual feature extraction for different detection scenarios. It can be seen that both the surrounding background, as well as sharp edges and clustered multitargets are correctly recognized, providing robust and accurate feature extraction.

C. Pixel-Level Multifeature Representation

Limited by feature extraction accuracy and single feature representation, existing methods are difficult to effectively suppress complex backgrounds. To this end, we propose the pixel-level multifeature representation RUMFR, which fuses high-precision contrast, distribution, and directional gradients features.

LCM has proven to be an effective method for IR small target detection [55]. However, when the target is near high-intensity clutter, the contrast feature measured by existing methods have

large errors, leading to "target suppression." Since RUSC-based model can distinguish between target and nearby clutter at pixel level, we design the following more accurate representation of contrast feature:

$$\mathrm{ACF} = \min_{i=1,2,\ldots,N_B} (\mathrm{m}_T - \mathrm{m}_{Bi}) \tag{18}$$

where N_B denotes the number of surrounding background areas, m_T and m_{Bi} denote the average intensity of areas T and B_i, respectively. It can be seen that ACF excludes the interference of nearby clutter and achieves a more accurate feature representation. As a result, ACF can better enhance real targets and suppresses low-contrast clutter.

However, the single contrast feature fails to remove high-intensity backgrounds, including PNHB and LABHB. Therefore, we fuse target distribution feature to further suppress the background interference. For one real target, it is usually located within a local area from 2×1 and 9×9 [66]. Thus, the target distribution feature is represented as follows:

$$\mathrm{TDF} = \begin{cases} 1, & V \times H > 1 \text{ and } V, H \leq 9 \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

where V and H denote the distribution distances in the vertical and horizontal directions of area T. With TDF, background interference such as PNHB that does not satisfy the target distribution characteristics can be effectively removed.

In complex detection scenarios, it is still challenging to suppress high-contrast background (HCB) that has similar contrast and distribution characteristics as the target. Recently, the gradient feature has been shown effective in suppressing complex backgrounds [67]. However, existing methods usually represent gradient features with fixed templates, which are not robust to targets with unknown distributions. In order to characterize the target gradient more accurately, we propose a pixel-level directional gradient representation. Specifically, we first construct a gradient feature extraction region [see Fig. 8(b)] centered on the target (x_0, y_0). As shown in Fig. 8(c), the region is divided into eight parts (i.e., $\mathrm{D1}, \mathrm{D2} \cdots, \mathrm{D8}$), depending on their orientation. Their direction vectors are denoted by $\boldsymbol{\tau_\alpha} = (\sin\alpha, \cos\alpha)$, where α is $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$, and 315° respectively. For one pixel (x, y) in the region, its gradient orientation angle can be obtained as follows:

$$\theta = \underset{\alpha=0^\circ,45^\circ,\ldots,315^\circ}{\operatorname{argmax}} |\boldsymbol{\tau_0} \cdot \boldsymbol{\tau_\alpha}| \tag{20}$$

where vector $\boldsymbol{\tau_0} = (x - x_0, y - y_0)$. Then, we can calculate the directional gradient of this pixel

$$\mathrm{DG}(x, y) = \frac{\partial f}{\partial l} = \frac{I(x + \Delta x, y + \Delta y) - I(x, y)}{|\boldsymbol{\tau_\theta}|} \tag{21}$$

where Δx and Δy denote the offsets, which are defined as follows:

$$\begin{cases} \Delta x = \mathrm{Round}(\cos\theta) \\ \Delta y = \mathrm{Round}(\sin\theta) \end{cases} \tag{22}$$

where Round indicates the rounding function. When all pixels are processed, the directional gradient vector map can be constructed as shown in Fig. 8(d). In order to effectively suppress interference such as HCB, we denote the directional gradient feature as follows:

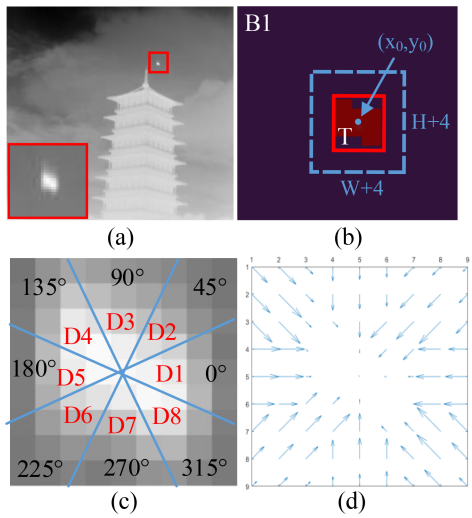


Fig. 8. Schematic diagram of the robust gradient feature representation. (a) Raw IR image with a small target. (b) RUSC-based window. (c) Feature extraction region. (d) Directional gradient vector map.

$$\mathrm{DGF} = \underset{i=1,2,3,4}{\mathrm{mean}} \{\mathrm{MDG}_i \times \mathrm{MDG}_{i+4}\} \tag{23}$$

where MRG_i denotes the maximum directional gradient of all pixels in area Di, which is defined as

$$\mathrm{MDG}_i = \max_{(x,y)\in \mathrm{Di}} \{\mathrm{DG}(x, y)\}. \tag{24}$$

Combining the above pixel-level multifeatures, we define the final RUMFR as

$$\mathrm{RUMFR} = \mathrm{ACF} \times \mathrm{TDF} \times \mathrm{DGF}. \tag{25}$$

D. RUMFR-Based Target Detection

As mentioned above, RUMFR improves performance in both feature extraction and feature representation. To illustrate the effectiveness of RUMFR, we provide a brief analysis below. When different candidate pixels are processed by RUMFR, the following situations may occur:

1) For a real target, it is accurately segmented from cluster multitargets or sharp edges. Thanks to the accurate feature extraction, its TDF = 1, ACF and DGF are all large. As a result, the target is effectively enhanced in the RUMFR map.
2) For low-contrast clutter, its ACF is small. As a result, the RUMFR is small, and the clutter is effectively suppressed.
3) For high-intensity clutter, such as LABHB, its $\mathrm{W} > 9$ or $\mathrm{H} > 9$. Thus, TDF = 0 and RUMFR = 0. As a result, the clutter is effectively removed. Similarly, the bright PNHB can be removed from the RUMFR map.
4) For high-contrast clutter, it usually has a large gradient difference only in a particular direction. Therefore, its DGF and RUMFR are small, and it is effectively suppressed.

Algorithm 2: RUMFR-Based Small Target Detection Algorithm.

Input: An infrared image
Output: Small target coordinates (x, y)
Calculate CWIDM for each pixel according to (13) and construct the CWIDM map;
Estimate parameter ε according to (4), and construct the candidate target set S_{CT} using (14);
repeat
    Select an unprocessed candidate target (x, y);
    Construct the RUSC-based window centered on (x, y)
    Calculate RUMFR(x, y) according to (25);
until all candidate pixels have been processed
Output the target coordinates (x, y) that satisfy RUMFR$(x, y) >$ Th;

It can be concluded that after the RUMFR calculation, the real targets are effectively enhanced and various clutters are removed or suppressed. In the RUMFR map, real targets usually have a higher intensity. Therefore, the final detection result can be obtained via threshold segmentation:

$$\text{Th} = \lambda \times R_{\max} + (1 - \lambda) \times R_{\text{mean}} \tag{26}$$

where $R_{\max}$ and R_{mean} denote the maximum and average grayscale of RUMFR map, respectively, and λ is an empirical parameter that controls the background suppression level. When λ is set too small, more background clutter will remain in the detection result, while when λ is set too large, the real target pixels may be removed as background, reducing the detection rate. Thus, λ is recommended to be set between [0.1, 0.7].

With the above preparations, we design the RUMFR-based detection method shown in Algorithm 2. Specifically, our method contains the following four main steps:

1) Construct CWIDM map for candidate target extraction and parameter estimation;
2) Construct RUSC-based windows for each candidate target to accurately extract features;
3) Construct RUMFR map to enhance real targets and suppress clutter interference;
4) Detect the targets using threshold segmentation in RUMFR map.

V. Experimental Results

To illustrate the effectiveness and robustness of the proposed method, this section conducts a large number of experiments in different detection scenarios. Specifically, we first describe the experimental setup, including the dataset, evaluation metrics, and baseline methods in Section V-A. Then, to verify the contribution of each component of the proposed method, Section V-B provides ablation experiments. Finally, Section V-C demonstrates the effectiveness of our method through comparative experiments with several state-of-the-art methods. All our experiments are accomplished with MATLAB or Pytorch on a computer with 16 GB RAM and dual-core 2.7 GHz processor. The configuration of each comparison algorithm is consistent with their literature.

A. Experimental Setup

1) Dataset: The experimental dataset consists of more than 1600 frames real IR images of different scenes. For convenience of analysis, we divide it into six sequences. As shown in Fig. 9, Seq. 1 [9], [13], [56] contains a large number of single-frame images with different backgrounds. Seq. 4 and Seq. 5 are derived from publicly available datasets IRSTD-1k [68] and SIRST [69], respectively. Meanwhile, Seq. 2 [70], Seq. 3 [70], and Seq. 6 [71] are publicly available video sequences with challenges. Table I summarize the detection difficulties faced by individual sequences, including sharp edges, cluster multitarget, complex backgrounds, and heavy noise. In addition, details such as number of frames, spatial resolution, and target scale are also provided in Table I.

2) Evaluation Metrics: In order to better evaluate the detection performance, two widely used metrics, background suppression factor (BSF), and signal to clutter ratio gain (SCRG) [13] are chosen for comparison. BSF evaluates the background suppression ability and is expressed as

$$\text{BSF} = \frac{\sigma_r}{\sigma_p} \tag{27}$$

where σ_r and σ_p denote the standard deviation (STD) of the raw image and algorithmically processed image, respectively. In order to fairly evaluate the background suppression performance, BSF is set to 1 when there is no real target in the processed image. SCRG evaluates the target enhancement ability and is expressed as

$$\text{SCRG} = \frac{\text{SCR}_{\text{out}}}{\text{SCR}_{\text{in}}} \tag{28}$$

where SCR denotes signal to clutter ratio and is defined as follows:

$$\text{SCR} = |\mu_{ct} - \mu_{nb}| / \sigma_{nb} \tag{29}$$

where μ_{ct} denotes the average intensity of the central target area, and μ_{nb} and σ_{nb} denote the average intensity and STD of the neighboring background area [52], respectively.

For a more comprehensive comparison of detection performance, receiver operating characteristic (ROC) is also chosen as an evaluation metric. ROC is presented by a curve, where the horizontal and vertical axes are the true positive rate (TPR) and false positive rate (FPR), respectively:

$$\text{TPR} = \frac{\text{True Detections}}{\text{Total Targets}} \tag{30}$$

$$\text{FPR} = \frac{\text{False Detection Pixel Number}}{\text{Total Number of Pixels}}. \tag{31}$$

Considering the close proximity between cluster multitargets, the maximum distance error for true detection is defined as three pixels. For quantitative comparison, the area under the curve (AUC) [72] is also calculated and presented in the lower right corner of the curve.

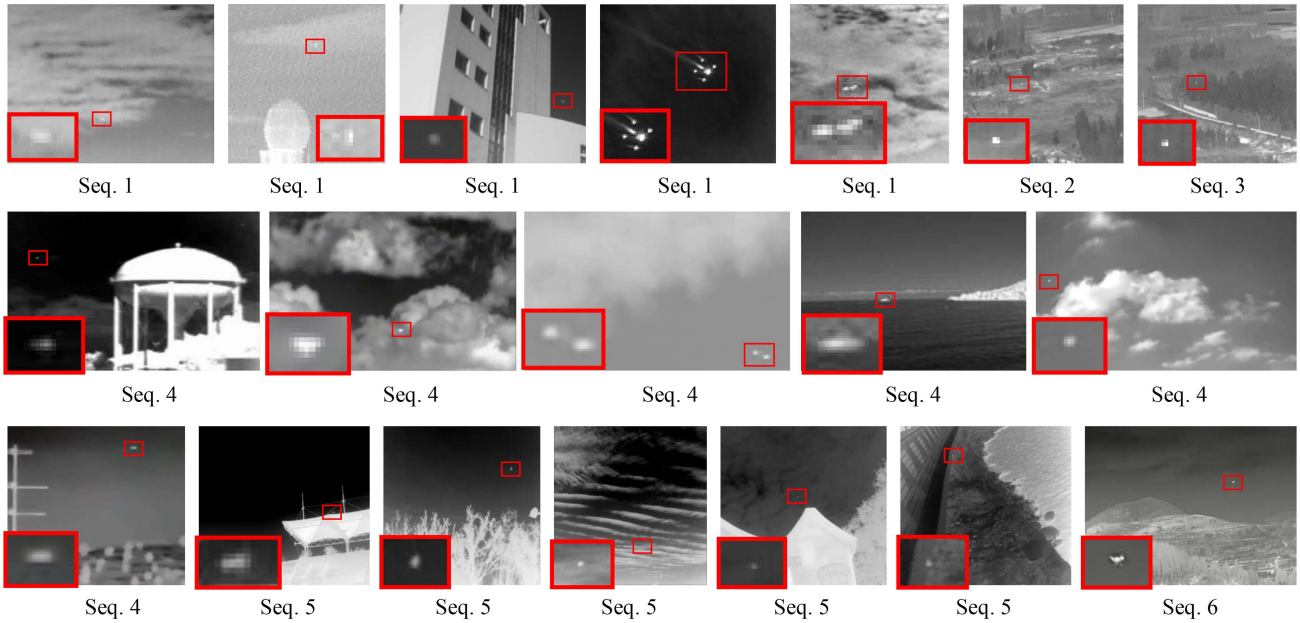


Fig. 9. Evaluation dataset consisting of six sequences containing multiple scenes. Where Seq. 1, Seq. 4, and Seq. 5 contain a large number of single-frame images with different backgrounds, while Seq. 2, Seq. 3, and Seq. 6 are video sequences with challenges. Note that the red rectangle represents the real target, and a large-scale view of the target is provided below the image.

TABLE I
DETAILS OF THE EXPERIMENTAL DATASET

Sequence	Frames	Spatial Resolution	Target Size	Detection Challenges			
				Sharp Edges	Cluster Multitarget	Complex Backgrounds	Heavy Noise
1	82	$256 \times 200 \sim 512 \times 512$	$2 \times 1 \sim 9 \times 9$	✓	✓		✓
2	200	256×256	$2 \times 1 : 3 \times 3$	✓		✓	
3	400	256×256	$2 \times 1 : 3 \times 3$	✓		✓	
4	334	$261 \times 171 \sim 375 \times 319$	$2 \times 1 : 9 \times 9$	✓	✓		✓
5	294	512×512	$2 \times 1 : 9 \times 9$	✓		✓	
6	353	640×512	$6 \times 8 : 7 \times 9$	✓		✓	

3) Baseline Methods: In this article, we select several representative methods for comparison, including the classic CF-based method (Max-Mean [8]), two state-of-the-art SLR-based methods (PSTNN [32] and NTFRA [10]), four recent LFR-based methods (FKRW [13], IFCM [59], ASPCM [56] and IDB-SCAN [44]), and three latest NN-based methods (DNA-Net [12], MTU-Net [42], and LW-IRSTNet [43]). Max-Mean is a classical filtering algorithm. PSTNN and NTFRA are the latest improved algorithms based on IPI and are widely recognized. ASPCM, FKRW, IFCM, and IDBSCAN are representative LFR-based methods. ASPCM optimizes the feature extraction model to an adaptive scale and achieves a better improvement in detection accuracy. FKRW, IFCM, and IDBSCAN are recent algorithms that apply clustering techniques for detection. DNA-Net, MTU-Net, and LW-IRSTNet are the latest state-of-the-art NN-based methods. More details of the parameter settings of comparison algorithms are presented in Table II.

TABLE II
DETAILS OF PARAMETER SETTINGS FOR BASELINE METHODS

No.	Methods	Parameter settings
1	Max-Mean [8]	Local window size: 5×5
2	PSTNN [32]	Patch size: 40×40, sliding step: 40
3	FKRW [13]	β=200, p=6, K=4, window size: 11×11
4	NTFRA [10]	Patch size: 40×40, sliding step: 40
5	IFCM [59]	β=43, λ=0.5, window size: 11×11
6	ASPCM [56]	Adaptive scale window size: [2,1] to [9,9]
7	IDBSCAN [44]	λ=0.1,λ_E=0.6,window size: 23×23
8	DNA-Net [12]	Learning rate: 0.05, batch size: 16, epoch=1500, $D_{\text{thresh}} = 3$
9	MTU-Net [42]	Learning rate: 0.05, batch size: 16, epoch=1500, $D_{\text{thresh}} = 3$
10	LW-IRSTNet [43]	Learning rate: 0.05, batch size: 64, epoch=100, kernel Sizes: 7×7, number of channels: 8, 32, 64

B. Ablation Analysis

The improvements of RUMFR consist of three main aspects: the candidate target extraction method CWIDM that balances

TABLE III
ABLATION ANALYSIS ON EXPERIMENTAL DATASETS, WHERE MAX-MEAN, LCM, DNGM, AND TLLCM ARE REPRESENTATIVE FILTER-BASED, SINGLE-LAYER-WINDOW-BASED, DUAL-LAYER-WINDOW-BASED, AND TRILAYER-WINDOW-BASED CANDIDATE-TARGET-EXTRACTION METHODS, RESPECTIVELY

	Methods	Dataset					
		Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5	Seq. 6
Pd	Max-Mean	0.2653	0.425	0.3484	0.5106	0.2790	0.5788
	LCM	0.8980	0.965	0.9298	0.8822	0.7127	0.9914
	DNGM	0.8367	0.985	0.7243	0.9849	0.9254	0.8453
	TLLCM	0.6531	0.855	0.4486	0.9033	0.7182	0.8481
	CWIDM	**0.9898**	**1**	**1**	**0.9997**	**1**	**1**
Times/s	Max-Mean	0.6742	0.6888	0.7034	0.7091	3.0033	3.7611
	LCM	0.0466	0.04571	0.0442	0.0426	0.1350	0.1575
	DNGM	0.0241	0.02839	0.0295	0.0259	0.0885	0.1013
	TLLCM	2.1202	2.6017	2.593	1.9475	8.2761	9.5385
	CWIDM	0.0486	0.0533	0.0593	0.0530	0.1667	0.1950

Bold values indicate the best method.

TABLE IV
ABLATION ANALYSIS OF THE PROPOSED METHOD ON THE EXPERIMENTAL DATASET, WHERE MF_{RW} AND MF_{IFCM} DENOTE MULTIFEATURE REPRESENTATIONS BASED ON DIFFERENT FEATURE EXTRACTION MODELS, AND RU_{NC}, RU_{ND}, AND RU_{NG} DENOTE RUMFR WITHOUT CONTRAST, DISTRIBUTION AND GRADIENT FEATURES, RESPECTIVELY

Component	MF_{RW}	MF_{IFCM}	RU_{nG}	RU_{nD}	RU_{nC}	RUMFR
ACF	✓	✓	✓	✓		✓
TDF	✓	✓	✓		✓	✓
DGF	✓	✓		✓	✓	✓
Average BSF	37.0322	26.834	25.6635	38.8942	33.4434	**39.0568**
Average SCRG	3.7084	1.647	3.9697	4.2417	4.1667	**4.2424**

Bold values indicate the best method.

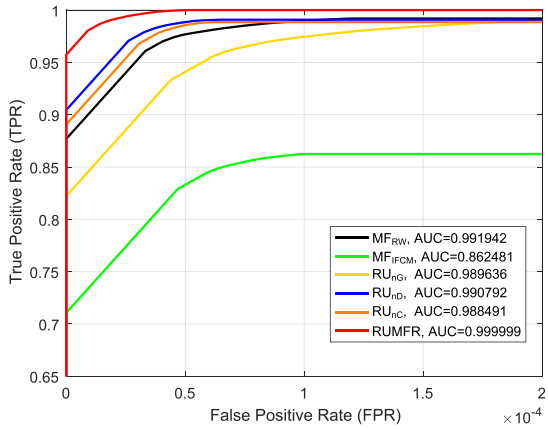


Fig. 10. Average ROC curves for ablation experiments of our method.

efficiency and effectiveness, an accurate feature extraction, and a robust multifeature representation. To demonstrate the contribution of each component in RUMFR, we carry out the following ablation experiments:

First, in order to demonstrate the improvement of CWIDM, we compare several candidate target extraction methods, including Max-Mean [8], LCM [9], DNGM [67], and TLLCM [51]. Where Max-Mean, LCM, DNGM, and TLLCM are representative filter-based, single-layer-window-based, dual-layer-window-based, and trilayer-window-based methods, respectively. Then, to illustrate the importance of feature extraction models, we design different models for multifeature representation. MF_{RW} and MF_{IFCM} adopt RW-based [13] and IFCM-based [59] feature extraction models, respectively. For fair comparison, they fuse the same multifeatures as RUMFR. Finally, to demonstrate the contribution of each feature to the RUMFR, we remove the corresponding feature separately for comparison. Specifically, RU_{nC}, RU_{nD}, and RU_{nG} denote the two-feature representations of RUMFR without contrast, distribution, and gradient features, respectively.

Table III compares the detection performance when using different candidate target extraction methods. It can be seen that CWIDM achieves the highest detection rate in each sequence. This indicates that CWIDM can effectively suppress the background interference and extract real targets as candidates. In addition, CWIDM utilizes the sliding window operation to optimize the computation and achieves better detection efficiency than Max-Mean and TLLCM. Therefore, CWIDM is a better candidate target extraction method that balances effectiveness and efficiency

Table IV shows the results of ablation experiments on the dataset, where comparison methods are all variants of RUMFR. MF_{RW} and MF_{IFCM} have different feature extraction models, while RU_{nC}, RU_{nD}, and RU_{nG} contain only two features. Note that the experimental dataset consists of all the sequences presented in Table I. It can be clearly seen that each component of the proposed method contributes positively to the detection performance. First, the RUSC-based methods achieve significantly higher SCRG values than MF_{RW} and MF_{IFCM}, demonstrating that our feature extraction model can effectively improve target enhancement ability. Furthermore, one can see that both BSF and SCRG values of RU_{nC}, RU_{nD}, and RU_{nG} decrease simultaneously. It suggests that each feature of RUMFR contributes to the performance improvement. Finally, it can be seen that proposed RUMFR achieves the highest BSF and SCRG values, which suggests that each component plays a positive role in target enhancement and background suppression.

For visual comparison, Fig. 10 shows average ROC curves of different methods. It can be seen that RUMFR is consistently in the upper left of the other curves and has the largest AUC value. It suggests that either changing the model or reducing the feature results in more missed and false detections. In other words, all components of the proposed method have positive effects, resulting in an optimal detection performance of RUMFR.

C. Performance Comparison

To further illustrate the important progress of our method in terms of detection performance, we compare it with ten representative algorithms presented in Section V-A. For comprehensive and objective evaluation, this section conducts experiments in both qualitative and quantitative comparisons.

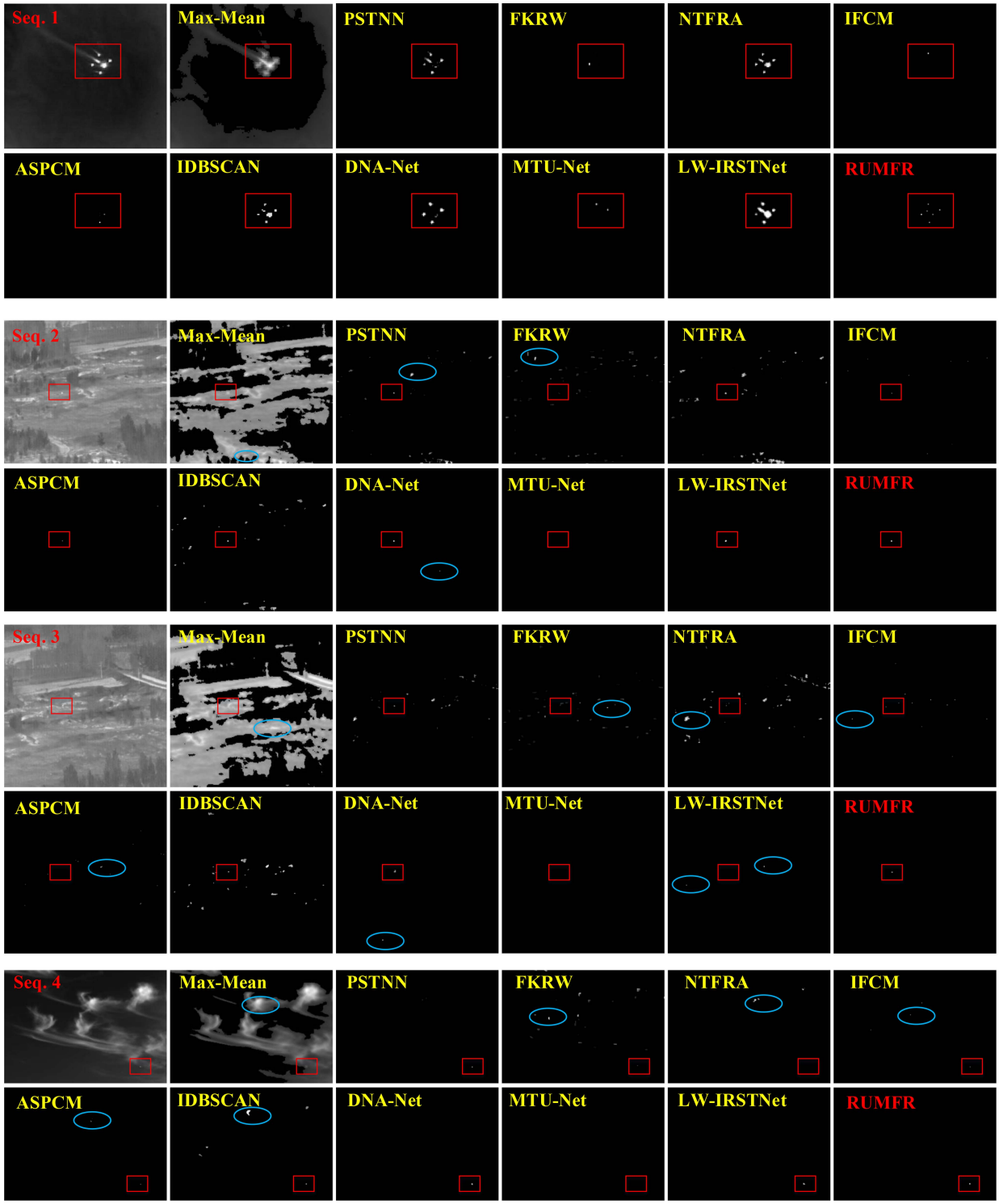


Fig. 11. IR raw images and detection results of different algorithms on the six sequences. The red rectangle represents real targets and the blue ellipse represents false detection.

For qualitative comparison, Fig. 11 shows the detection results of different methods on the six sequences. It can be clearly seen that RUMFR achieves better performance in terms of target enhancement and background suppression. Specifically, there is almost no residual clutter in the RUMFR maps, while real targets are salient. The CF-based method Max-Mean enhances real targets, but its poor background suppression leads to many false alarms. As representative SLR-based methods (PSTNN and NTFRA), they perform well in clustered multitarget detection for simple scenes such as Seq. 1. However, their low-rank

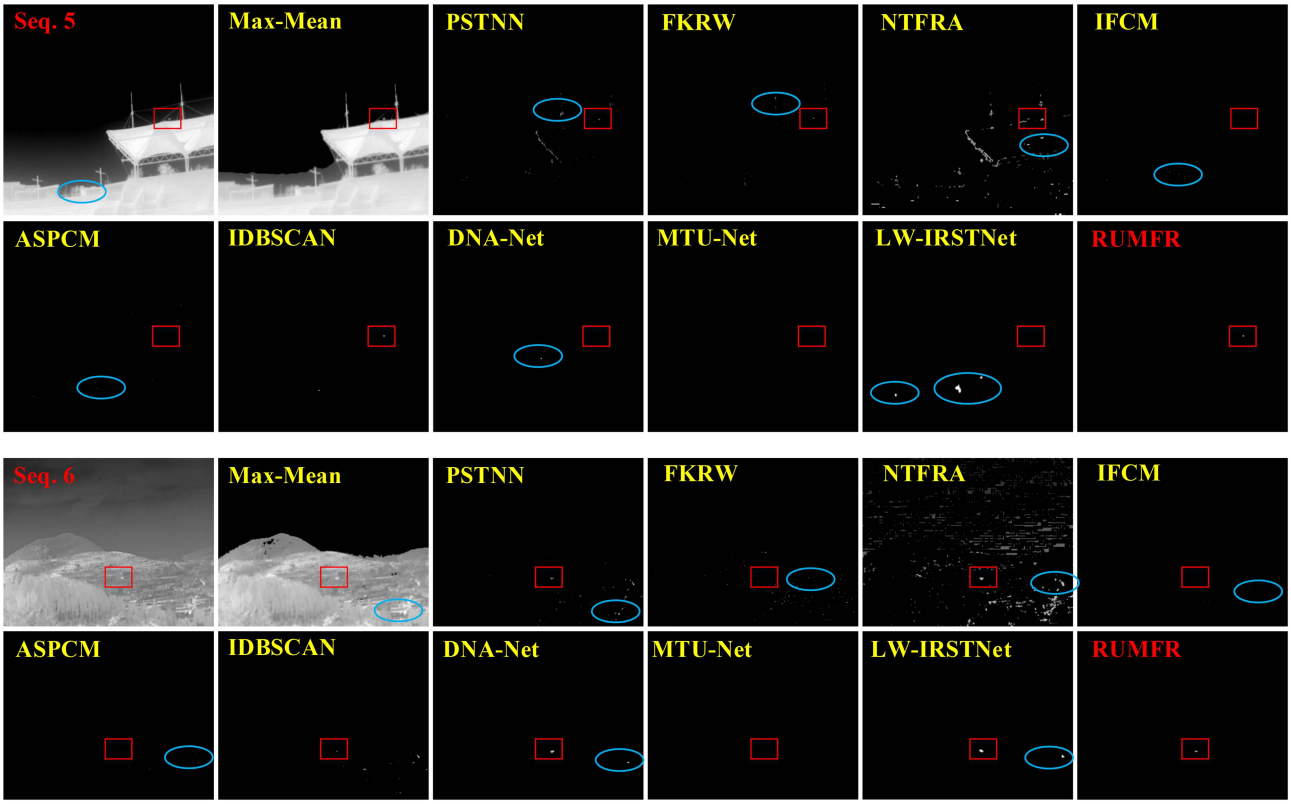


Fig. 11. *(Continued.)* IR raw images and detection results of different algorithms on the six sequences. The red rectangle represents real targets and the blue ellipse represents false detection.

matrices are difficult to represent complex interferences such as PNHB, making more false detections in other sequences. By designing adaptive models to extract features, three representative LFR-based methods (FKRW, IFCM, and ASPCM) have made significant improvements in clutter suppression. However, they struggle to cope with the complex scenarios in Seq. 1 and Seq. 5, leading to many false and missed detections. The latest method (IDBSCAN) accurately detects each target in Seq. 1. However, due to the limited feature representation capability, IDBSCAN is prone to misdetect the background clutter in Seq. 4, resulting in a high false alarm rate. Three latest NN-based methods (DNA-Net, MTU-Net, and LW-IRSTNet) remove almost all background clutter. However, as data-driven methods, they struggle to cope with scenarios where real data is scarce. For example, MTU-Net for tiny ship detection misses many real targets in most scenarios, DNA-Net misses a real target in Seq. 1, and LW-IRSTNet misidentifies two targets in Seq. 1 as one.

For quantitative comparison, Tables V and VI show the average BSF and SCRG values for all methods on different sequences, respectively. Note that larger BSF and SCRG values imply better local target enhancement and global background suppression performance, respectively. It can be seen that Max-Mean has the lowest BSF value, implying poor background suppression. DNA-Net and LW-IRSTNet achieve the top two high BSF values, implying the best background suppression performance. Compared to other model-driven methods, RUMFR has high BSF values and good background suppression. In terms of target enhancement, RUMFR achieves the highest SCRG value, and IFCM and ASPCM achieve the top three highest values. In general, RUMFR achieves better comprehensive performance in terms of target enhancement and background suppression.

To further evaluate the detection performance, Fig. 12 shows the ROC curves of different algorithms on the six sequences. Note that curves near the upper left corner have better detection performance. One can see that the curves of RUMFR are almost always in the upper leftmost part and achieve the best AUC values in all sequences. It suggests that our method can achieve a higher detection rate with fewer false alarms. Thanks to the local structure tensor, the latest method NTFRA achieves the highest TPR at FPR=0 in Seq. 2. However, the limited robustness makes its performance decrease significantly in other sequences such as Seq. 5. Advanced methods (FKRW, IFCM, and ASPCM) effectively minimize false detection of complex backgrounds. However, feature extraction models with limited robustness results in their low detection rates and AUC values, especially when dealing with cluster target detection in Seq. 1 and sharp edge interference in Seq. 5. Although the latest method (IDBSCAN) shows the second best detection accuracy in Seq. 1 and Seq. 5, the false alarm rate increases significantly when there is complex background interference. DNA-Net, MTU-Net, and LW-IRSTNet are sensitive to training data. For example, MTU-Net, which targets tiny ship detection, performs poorly in most sequences. Thanks to sufficient data and model

TABLE V
AVERAGE BSF OF 11 METHODS ON THE SIX SEQUENCES

Methods	Dataset					
	Seq. 1	**Seq. 2**	**Seq. 3**	**Seq. 4**	**Seq. 5**	**Seq. 6**
Max-Mean	0.33271	0.29582	0.25986	0.35679	0.64147	0.52184
PSTNN	10.9307	3.5162	3.5974	9.6504	23.0106	18.3526
FKRW	8.8612	5.1371	5.8333	12.093	37.799	13.8467
NTFRA	4.4328	1.7085	1.9544	3.4188	9.9921	2.6435
IFCM	13.4905	10.0925	8.0499	20.6335	51.501	43.7284
ASPCM	19.0679	15.7224	12.8067	20.856	57.4645	54.9315
IDBSCAN	18.5595	2.5998	2.793	22.1097	25.0395	12.604
DNA-Net	**973.357**	**2387.96**	**2453.17**	**1525.12**	**4062.14**	**3796.468**
MTU-Net	**394.322**	1	1	**68.7096**	**1280.76**	**548.425**
LW-IRSTNet	**1133.20**	**1724.62**	**1028.03**	**1519.40**	**3105.62**	**2846.67**
RUMFR	21.7979	**17.9806**	**17.0834**	25.2381	77.2709	60.9504

The red, blue, and green fonts indicate the best, second best, and third best results, respectively.

TABLE VI
AVERAGE SCRG OF 11 METHODS ON THE SIX SEQUENCES

Methods	Dataset					
	Seq. 1	**Seq. 2**	**Seq. 3**	**Seq. 4**	**Seq. 5**	**Seq. 6**
Max-Mean	0.63505	0.23916	0.28414	0.62682	0.71663	0.65095
PSTNN	2.9789	2.5758	3.4171	1.6081	1.4753	1.1937
FKRW	2.9187	1.828	1.9562	1.8788	1.7625	1.2816
NTFRA	2.0738	1.8666	2.3771	1.3197	1.2177	0.76663
IFCM	4.6799	**4.3643**	**4.5341**	**3.4849**	**2.703**	**2.1692**
ASPCM	**5.0957**	**4.4243**	**5.3555**	3.2039	**2.6601**	**1.8931**
IDBSCAN	**6.1245**	2.3819	2.5951	**3.8306**	1.6463	1.0073
DNA-Net	1.0094	2.1559	3.0709	1.0109	0.8113	0.6820
MTU-Net	0.1395	0	0	0.0232	0.1321	0.0270
LW-IRSTNet	1.909	1.5975	1.4184	1.0178	0.5890	0.6493
RUMFR	**6.8666**	**4.5752**	**5.6878**	**3.9329**	**3.6295**	**2.6498**

The red, blue, and green fonts indicate the best, second best, and third best results, respectively.

optimization, LW-IRSTNet achieves the second highest AUC value in Seq. 4. However, it still misses the real target in Seq. 5 and misdetects the clutter interference in Seq. 6. In comparison, RUMFR achieves the best detection performance by reducing false detections while maintaining a high detection rate.

For efficiency comparison, Table VII presents the complexity and processing time of different methods. The main computational process of Max-Mean is to compute the mean value in different directions centered on individual pixels and solve for the maximum value. Therefore its complexity can be approximated as $\mathcal{O}(\mathrm{MN})$, where M and N denote the height and width of the infrared image. For LFR-based methods, computing the local feature differences is the main time-consuming step. For one single pixel in the image, it requires $\mathcal{O}(s^2)$, where $s \times s$ denotes the local window size. Since recent methods utilize local filtering operations instead of pixel-by-pixel computation, the computation of a single pixel is optimized to $\mathcal{O}(1)$. Therefore, the complexity of LFR-based methods is approximated as $\mathcal{O}(\mathrm{LMN})$, where L denotes the number of scales. For single-scale IFCM and ASPCM, their complexity is $\mathcal{O}(\mathrm{MN})$. Since FKRW and IDBSCAN first extract candidate targets using multiscale methods and then further operate on some candidate targets, their complexity can also be approximated as $\mathcal{O}(\mathrm{LMN})$. The computational complexity of SLR-based methods mainly comes from the singular value decomposition process, thus the complexity is $\mathcal{O}(\mathrm{mn}^2)$, where m and n denote the rows and columns of the patch image. PSTNN and NTFRA reduce the complexity to $\mathcal{O}(\mathrm{n_1n_2n_3}log(\mathrm{n_1n_2}) + \mathrm{n_1n_2^2}[(\mathrm{n_3}+1)/2])$ and $\mathcal{O}(\mathrm{n_1n_2n_3}log(\mathrm{n_1n_2}) + \sum_{i=1}^{3}\min(\mathrm{n}_i, \mathrm{n}_{i+1}))$, where $\mathrm{n_1} \times \mathrm{n_2} \times \mathrm{n_3}$ represents the size of the patch tensor. The main computation of NN-based methods comes in the convolutional layer operations, thus their complexity can be approximated as $\mathcal{O}(\mathrm{L_N C_i C_o K^2 W^2})$, where $\mathrm{L_N}$ denotes the number of network convolutional layers, $\mathrm{C_i}$ and $\mathrm{C_o}$ denote the number of channels of the input and output feature maps, respectively, $\mathrm{K^2}$ is the convolution kernel size, and $\mathrm{W^2}$ denotes the size of the output feature map. Since the size of the output feature map is related to the original image size, their complexity can be approximated as $\mathcal{O}(\mathrm{L_N C_i C_o K^2 MN})$. In terms of computation time, IFCM and LW-IRSTNet consume the least time and achieve the best real-time performance. It can be seen that the proposed method

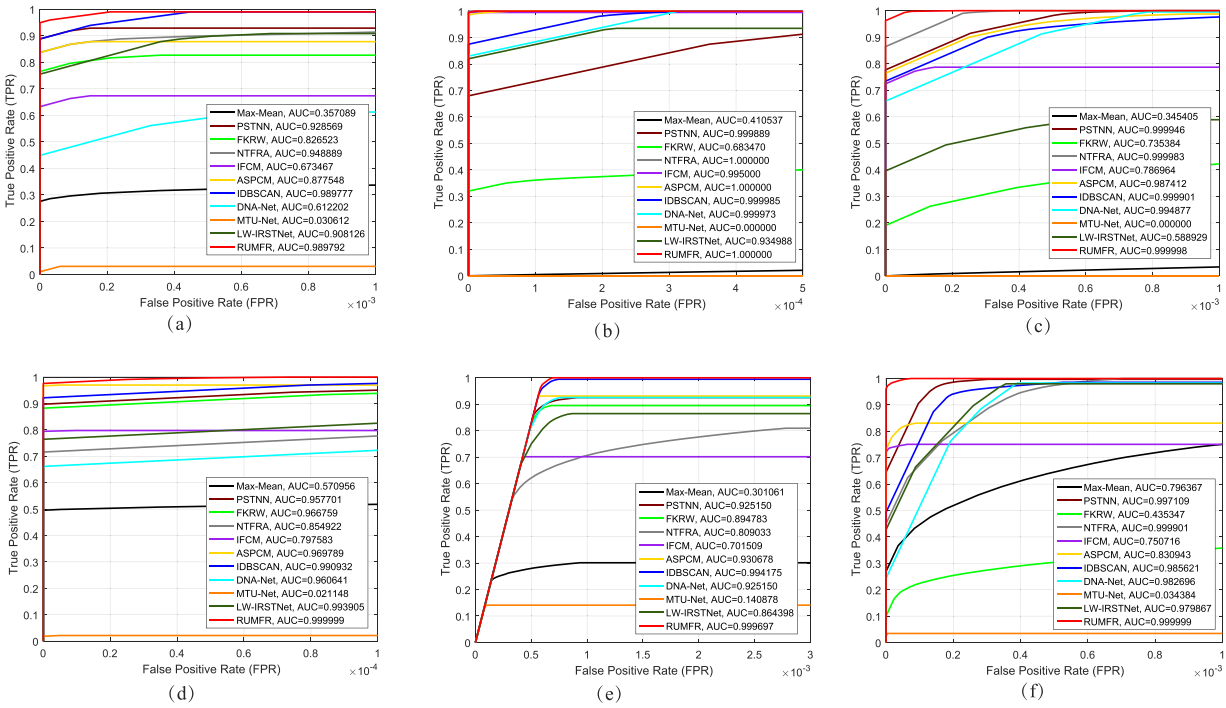


Fig. 12. ROC curves of different methods on the six sequences. (a) Seq. 1. (b) Seq. 2. (c) Seq. 3. (d) Seq. 4. (e) Seq. 5. (f) Seq. 6.

TABLE VII
COMPARISON OF COMPLEXITY AND AVERAGE PROCESSING TIME (IN SECONDS) OF DIFFERENT ALGORITHMS

Methods	Complexity	Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5	Seq. 6
Max-Mean	$\mathcal{O}(MN)$	0.6742	0.6888	0.7034	0.7091	3.0033	3.7611
PSTNN	$\mathcal{O}(n_1n_2n_3\log(n_1n_2) + n_1n_2^2[(n_3+1)/2])$	0.09240	0.1633	0.1635	0.0658	0.4473	0.6547
FKRW	$\mathcal{O}(MN)$	0.1946	0.2345	0.2491	0.1774	0.6860	1.1942
NTFRA	$\mathcal{O}(n_1n_2n_3\log(n_1n_2) + \sum_{i=1}^{3}\min(n_i, n_{i+1}))$	1.083	1.2255	1.2164	1.1811	5.471	6.7272
IFCM	$\mathcal{O}(MN)$	0.0202	0.0195	0.0183	0.0152	0.0497	0.0610
ASPCM	$\mathcal{O}(MN)$	0.3723	0.4161	0.4138	0.4151	1.7491	2.1724
IDBSCAN	$\mathcal{O}(LMN)$	0.4150	0.5713	0.5392	0.2621	0.5645	0.6829
DNA-Net	$\mathcal{O}(L_NC_iC_oK^2MN)$	0.36	0.382	0.38	0.376	0.361	0.391
MTU-Net	$\mathcal{O}(L_NC_iC_oK^2MN)$	0.412	0.374	0.404	0.406	0.413	0.414
LW-IRSTNet	$\mathcal{O}(L_NC_iC_oK^2MN)$	0.039	0.041	0.039	0.038	0.04	0.04
RUMFR	$\mathcal{O}(LMN)$	0.2295	0.2377	0.2394	0.2098	0.3410	0.3835

Where M×N denotes the image size, L denotes the number of different scales, and n1 × n2 × n3 denotes the tensor size that is positively correlated with M× N.

is also efficient and runs faster than most of recent methods such as IDBSCAN and DNA-Net. In general, the proposed method provides a better balance between detection performance and efficiency.

D. Discussion

1) Limitations and Further Research: The proposed RUMFR achieves better detection performance in complex scenarios by designing robust unsupervised models. However, the detection process is not efficient enough for scenes with higher real-time requirements. In order to improve the detection speed and extend wider applications, we will carry out further research work as follows: on one hand, we will investigate parallel computing techniques to characterize multiple candidate targets simultaneously; on the other hand, we will investigate more efficient unsupervised clustering methods, such as adaptive step-size parameter tuning strategies.

2) Other Applications: Thanks to the robust feature representation, the proposed method has the potential to be applied for other types of data. For this purpose, we have collected some synthetic aperture radar (SAR) and visible light data containing small targets for testing. Fig. 13 presents the detection results of our method on different data. It can be seen that our method

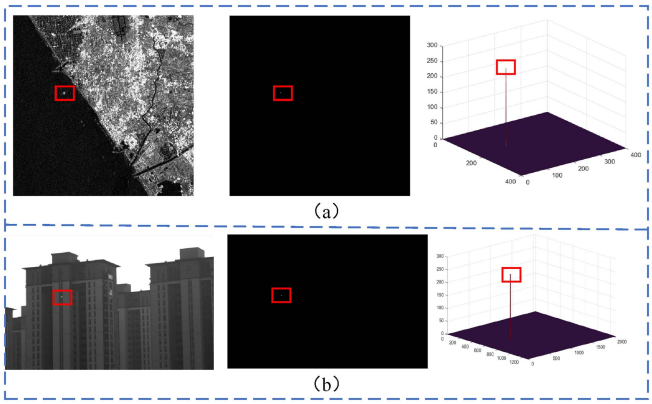


Fig. 13. Application of the proposed method for other types of data. (a) SAR image. (b) Visible light image.

has good background suppression performance in SAR and visible scenes with complex background interference (e.g., bright buildings, cluttered clouds). Therefore, if the targets in other types of data have similar characteristics to IR small targets, our method can still achieve accurate detection.

VI. Conclusion

In this article, a robust RUMFR-based infrared small target detection method is proposed. In order to solve the "target suppression" problem in complex detection scenarios, we make improvements in both feature extraction and feature representation. On the one hand, we design the robust clustering method RUSC and propose the RUSC-based model to accurately extract target features. In particular, by designing the parameter adaptive strategy and iterative clustering strategy, RUSC can robustly segment an arbitrary number and distribution of cluster multitargets from complex backgrounds. On the other hand, to further suppress clutter interference, we propose RUMFR that fuses pixel-level multifeatures with stronger characterization capability. Extensive experimental results show that our method has better target enhancement and background suppression capabilities, especially in sharp edge interference and cluster multitarget detection. Compared to several state-of-the-art methods, our method achieves a higher detection rate with fewer false detections.

References

[1] P. Yang, L. Dong, and W. Xu, "Small maritime target detection using gradient vector field characterization of infrared image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1827–1841, 2023.

[2] H. Yi et al., "Spatial-temporal tensor ring norm regularization for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[3] T. Zhang, L. Li, S. Cao, T. Pu, and Z. Peng, "Attention-guided pyramid context networks for detecting infrared small target under complex background," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, pp. 4250–4261, 2023.

[4] D. Zhou and X. Wang, "Robust infrared small target detection using a novel four-leaf model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1462–1469, 2024.

[5] X. Wang, F. Xie, W. Liu, S. Tang, and J. Yan, "Robust small infrared target detection using multi-scale contrast fuzzy discriminant segmentation," *Expert Syst. Appl.*, vol. 212, 2023, Art. no. 118813. .

[6] J. He, J. Chen, H. Xu, and M. S. Ayub, "Small target detection method based on low-rank sparse matrix factorization for side-scan sonar images," *Remote Sens.*, vol. 15, no. 8, 2023, Art. no. 2054.

[7] H. Gao, Y. Zhang, Z. Chen, S. Xu, D. Hong, and B. Zhang, "A multidepth and multibranch network for hyperspectral target detection based on band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–18, 2023.

[8] S. D. Deshpande, M. H. Er, R. Venkateswarlu, and P. Chan, "Max-mean and max-median filters for detection of small targets," in *Proc. Int. Soc. Opt. Photon. Signal Data Process. Small Targets*, vol. 3809, 1999, pp. 74–83.

[9] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2168–2172, Dec. 2014.

[10] X. Kong, C. Yang, S. Cao, C. Li, and Z. Peng, "Infrared small target detection via nonconvex tensor fibered rank approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–21, 2022.

[11] R. Kou et al., "Infrared small target segmentation networks: A survey," *Pattern Recognit.*, vol. 143, 2023, Art. no. 109788.

[12] B. Li et al., "Dense nested attention network for infrared small target detection," *IEEE Trans. Image Process.*, vol. 32, pp. 1745–1758, 2022.

[13] Y. Qin, L. Bruzzone, C. Gao, and B. Li, "Infrared small target detection based on facet kernel and random walker," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7104–7118, Sep. 2019.

[14] M. Ester et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. Kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[15] T. W. Bae, "Small target detection using bilateral filter and temporal cross product in infrared images," *Infrared Phys. Technol.*, vol. 54, no. 5, pp. 403–411, 2011.

[16] H. Zhu, J. Zhang, G. Xu, and L. Deng, "Balanced ring top-hat transformation for infrared small-target detection with guided filter kernel," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 5, pp. 3892–3903, May 2020.

[17] S. Qi, J. Ma, H. Li, S. Zhang, and J. Tian, "Infrared small target enhancement via phase spectrum of quaternion Fourier transform," *Infrared Phys. Technol.*, vol. 62, pp. 50–58, 2014.

[18] Z. Song, J. Yang, D. Zhang, S. Wang, and Z. Li, "Semi-supervised dim and small infrared ship detection network based on haar wavelet," *IEEE Access*, vol. 9, pp. 29686–29695, 2021.

[19] G. Boccignone, A. Chianese, and A. Picariello, "Small target detection using wavelets," in *Proc. IEEE 14th Int. Conf. Pattern Recognit.*, vol. 2, 1998, pp. 1776–1778.

[20] Y. Zhang, J. Zhang, D. Wang, and C. Chen, "Infrared small target detection based on morphology and wavelet transform," in *Proc. IEEE 2nd Int. Conf. Artif. Intell., Manage. Sci. Electron. Commerce*, 2011, pp. 4033–4036.

[21] Y. Xin, J. Zhou, and Y. Chen, "Dual multi-scale filter with SSS and GW for infrared small target detection," *Infrared Phys. Technol.*, vol. 81, pp. 97–108, 2017.

[22] H. Wang and Y. Xin, "Wavelet-based contourlet transform and Kurtosis map for infrared small target detection in complex background," *Sensors*, vol. 20, no. 3, p. 755, 2020. .

[23] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4996–5009, Dec. 2013.

[24] Y. Dai, Y. Wu, Y. Song, and J. Guo, "Non-negative infrared patch-image model: Robust target-background separation via partial sum minimization of singular values," *Infrared Phys. Technol.*, vol. 81, pp. 182–194, 2017.

[25] X. Wang, Z. Peng, P. Zhang, and Y. He, "Infrared small target detection via nonnegativity-constrained variational mode decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1700–1704, Oct. 2017.

[26] L. Zhang, L. Peng, T. Zhang, S. Cao, and Z. Peng, "Infrared small target detection via non-convex rank approximation minimization joint $l_{2,1}$-norm," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1821.

[27] T. Zhang, H. Wu, Y. Liu, L. Peng, C. Yang, and Z. Peng, "Infrared small target detection based on non-convex optimization with l_p-norm constraint," *Remote Sens.*, vol. 11, no. 5, p. 559, 2019.

[28] T. Liu, J. Yang, B. Li, Y. Wang, and W. An, "Infrared small target detection via nonconvex tensor tucker decomposition with factor prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, p. 1–17, 2023.

[29] Z. Zhang, C. Ding, Z. Gao, and C. Xie, "Anlpt: Self-adaptive and nonlocal patch-tensor model for infrared small target detection," *Remote Sens.*, vol. 15, no. 4, 2023, Art. no. 1021.

[30] D. Pang, P. Ma, Y. Feng, T. Shan, R. Tao, and Q. Jin, "Tensor spectral k-support norm minimization for detecting infrared dim and small target against urban backgrounds," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5002513.
[31] Y. Dai and Y. Wu, "Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3752–3767, Aug. 2017.
[32] L. Zhang and Z. Peng, "Infrared small target detection based on partial sum of the tensor nuclear norm," *Remote Sens.*, vol. 11, no. 4, p. 382, 2019.
[33] M. Shi and H. Wang, "Infrared dim and small target detection based on denoising autoencoder network," *Mobile Netw. Appl.*, vol. 25, no. 4, pp. 1469–1483, 2020.
[34] H. Wang, L. Zhou, and L. Wang, "Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 8509–8518.
[35] H. Sun, J. Bai, F. Yang, and X. Bai, "Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, 2023.
[36] J.-H. Kim and Y. Hwang, "Gan-based synthetic data augmentation for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.
[37] A. Wang, W. Li, Z. Huang, X. Wu, F. Jie, and R. Tao, "Prior-guided data augmentation for infrared small target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 10027–10040, 2022.
[38] T. Liu, Q. Yin, J. Yang, Y. Wang, and W. An, "Combining deep denoiser and low-rank priors for infrared small target detection," *Pattern Recognit.*, vol. 135, 2023, Art. no. 109184.
[39] Y. Dai, X. Li, F. Zhou, Y. Qian, Y. Chen, and J. Yang, "One-stage cascade refinement networks for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–17, 2023.
[40] X. Zhou, P. Li, Y. Zhang, X. Lu, and Y. Hu, "Deep low-rank and sparse patch-image network for infrared dim and small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, 2023.
[41] Y. Zhu, Y. Ma, F. Fan, J. Huang, K. Wu, and G. Wang, "Towards accurate infrared small target detection via edge-aware gated transformer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 8779–8793, 2024.
[42] T. Wu et al., "MTU-Net: Multilevel transunet for space-based infrared tiny ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–15, 2023.
[43] R. Kou et al., "LW-IRSTNet: Lightweight infrared small target segmentation network and application deployment," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5621313.
[44] Z. Qiu, Y. Ma, F. Fan, J. Huang, L. Wu, and Y. Du, "Improved DBSCAN for infrared cluster small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
[45] X. Zhang, J. Ru, and C. Wu, "Infrared small target detection based on gradient correlation filtering and contrast measurement," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, 2023.
[46] J. Han, Y. Ma, J. Huang, X. Mei, and J. Ma, "An infrared small target detecting algorithm based on human visual system," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 452–456, Mar. 2016.
[47] H. Fu, Y. Long, R. Zhu, and W. An, "Infrared small target detection based on multiscale center-surround contrast measure," in *Proc. 9th Int. Conf. Graphic Image Process.*, vol. 10615, 2018, Art. no. 106150I.
[48] Y. Shi, Y. Wei, H. Yao, D. Pan, and G. Xiao, "High-boost-based multiscale local contrast measure for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 1, pp. 33–37, Jan. 2017.
[49] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Infrared small-target detection using multiscale gray difference weighted image entropy," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 1, pp. 60–72, Jan. 2016.
[50] J. Liu, Z. He, Z. Chen, and L. Shao, "Tiny and dim infrared target detection based on weighted local contrast," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 11, pp. 1780–1784, Nov. 2018.
[51] J. Han, K. Liang, B. Zhou, X. Zhu, J. Zhao, and L. Zhao, "Infrared small target detection utilizing the multiscale relative local contrast measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 612–616, Apr. 2018.
[52] Z. Qiu, Y. Ma, F. Fan, J. Huang, M. Wu, and X. Mei, "A pixel-level local contrast measure for infrared small target detection," *Defence Technol.*, vol. 18, no. 9, pp. 1589–1601, 2022.
[53] Z. Qiu, Y. Ma, F. Fan, J. Huang, and L. Wu, "Global sparsity-weighted local contrast measure for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[54] C. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 574–581, Jan. 2013.
[55] Y. Wei, X. You, and H. Li, "Multiscale patch-based contrast measure for small infrared target detection," *Pattern Recognit.*, vol. 58, pp. 216–226, 2016.
[56] Z. Qiu, Y. Ma, F. Fan, J. Huang, and M. Wu, "Adaptive scale patch-based contrast measure for dim and small infrared target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[57] S. Huang, Z. Peng, Z. Wang, X. Wang, and M. Li, "Infrared small target detection by density peaks searching and maximum-gray region growing," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1919–1923, Dec. 2019.
[58] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
[59] L. Chen and L. Lin, "Improved fuzzy c-means for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.
[60] Y. Xu et al., "Infrared small target detection based on local contrast-weighted multidirectional derivative," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, 2023.
[61] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 21, no. 3, pp. 193–199, Mar. 2002.
[62] L. Rokach and O. Maimon, "Clustering methods," in *Data Mining Knowledge Discovery Handbook*. Berlin, Germany: Springer, 2005, pp. 321–352.
[63] J. D. Banfield and A. E. Raftery, "Model-based Gaussian and non-Gaussian clustering," *Biometrics*, pp. 803–821, 1993. .
[64] M. M. Deza and E. Deza, *Encyclopedia of Distances*. New York, NY, USA: Springer, 2009.
[65] J. Han, S. Moradi, I. Faramarzi, C. Liu, H. Zhang, and Q. Zhao, "A local contrast method for infrared small-target detection utilizing a tri-layer window," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1822–1826, Oct. 2019.
[66] X. Guan, L. Zhang, S. Huang, and Z. Peng, "Infrared small target detection via non-convex tensor rank surrogate joint local contrast energy," *Remote Sens.*, vol. 12, no. 9, 2020, Art. no. 1520.
[67] L. Wu, Y. Ma, F. Fan, M. Wu, and J. Huang, "A double-neighborhood gradient method for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 8, pp. 1476–1480, Aug. 2021.
[68] M. Zhang, R. Zhang, Y. Yang, H. Bai, J. Zhang, and J. Guo, "ISNet: Shape matters for infrared small target detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 877–886.
[69] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Asymmetric contextual modulation for infrared small target detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2021, pp. 950–959.
[70] B. Hui et al., "A dataset for infrared image dim-small aircraft target detection and tracking under ground/air background," Oct. 2019. [Online]. Available: https://doi.org/10.11922/sciencedb.902
[71] J. Zhao et al., "The 3rd anti-UAV workshop & challenge: Methods and results," 2023, *arXiv:2305.07290*.
[72] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 299–310, Mar. 2005.

Liqiong Chen received the B.S. and Ph.D. degrees from the School of Electronic Information, Wuhan University, Wuhan, China, in 2016 and 2021, respectively.

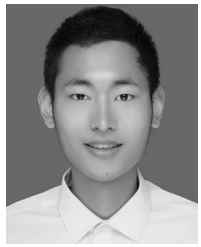
She is currently an Associate Professor with the School of Mechanical Engineering and Automation, Fuzhou University. Her current research interests include object detection, scene recognition and deep learning.

Tong Wu received the B.S. degree in energy and power engineering from the Wuhan University of Technology, Wuhan, China, in 2022. He is currently working toward the M.S. degree in mechanical design and theory at Fuzhou University, Fuzhou, China.

His main research interests include image processing, computer vision, and deep learning.

Shuyuan Zheng received the B.S. degree in mechanical engineering and automation from the Huaqiao University, Xiamen, China, in 2023. She is currently working toward the M.S. degree in mechanical engineering with Fuzhou University.

Her main research interests include infrared small target detection and deep learning.



Zhaobing Qiu received the B.S., M.S., and Ph.D. degrees from the School of Electronic Information, Wuhan University, Wuhan, China, in 2016, 2019, and 2023.

He is currently a Lecturer with the School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou, China. His research interests include computer vision, pattern recognition, and optomechanical system control.

Feng Huang was born in Fuzhou, China, in 1979.

He is currently with the School of Mechanical Engineering and Automation, Fuzhou University, China. His research interest includes optoelectronic imaging.