



Article

Improved Weighted Nuclear Norm Minimization for Infrared Dim and Small-Target Detection Based on Low-Rank Clustering

Yizhi Deng ¹, Xiangsuo Fan ^{1,2,3,*}, Qingxiao Ma ¹ and Jiajun Liang ¹¹ School of Automation, Guangxi University of Science and Technology, No.19, Guantang Avenue, Yufeng District, Liuzhou 545006, China² Guangxi Key Laboratory of Logistics Unmanned Aircraft Technology for Transportation Industry, No.19, Guantang Avenue, Yufeng District, Liuzhou 545616, China³ Guangxi Low-Altitude Unmanned Aircraft Key Technologies Engineering Research Center, No.19, Guantang Avenue, Yufeng District, Liuzhou 545616, China

* Correspondence: 100002085@gxust.edu.cn

Highlights

What are the main findings?

- A low-rank clustering mechanism is proposed to adaptively group image patches based on local variance and Euclidean distance, which significantly enhances the low-rank property of the background matrix in infrared images.
- A class-adaptive weighted nuclear norm penalty is designed for different categories of image patches, where the singular value penalty strength is dynamically adjusted according to the statistical characteristics of each patch group.
- An energy shrinkage mechanism is introduced to effectively alleviate the problem of target energy dilution during aggregation, improving the contrast of dim targets.

What is the implication of the main finding?

- This study demonstrates that incorporating low-rank clustering priors into the Weighted Nuclear Norm Minimization framework can effectively suppress background clutter and enhance target saliency in complex infrared scenes. The proposed method achieves superior performance in six challenging scenarios, with BSF and AUC ranking the highest among several state-of-the-art methods. The main limitation is the relatively high computational cost, which future work can address through parallel computing or model acceleration.

Abstract

In complex infrared scenarios, where prominent edge contours and strong noise interference coexist, effectively suppressing these interfering factors has become the core challenge faced by algorithms for detecting weak and small targets. Current mainstream algorithms generally suffer from a high false alarm rate, mainly due to the interference effect of background clutter. To this end, this paper proposes a Weighted Nuclear Norm Minimization method based on low-rank clustering idea. This method first clusters the image patches adaptively by a similarity measure combining local statistical features and Euclidean distance, which enhances the low-rank property of the background matrix. On this basis, a differentiated weighted nuclear norm penalty is designed for image patches of different categories, which further improves the ability of the model to separate the background from the target. An energy shrinkage mechanism is also introduced to alleviate target energy dilution. Experimental results on six different scenes, compared with six mainstream detection methods and two deep learning methods, using BSF, SNR, and three AUC metrics, demonstrate that the proposed method achieves superior background suppression



Academic Editor: Haipeng Wang

Received: 22 June 2026

Revised: 6 July 2026

Accepted: 16 July 2026

Published: 23 July 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

and detection performance. Although the computational cost is high (80 to 245 s per frame) and currently not suitable for real-time or near-real-time applications, its advantages in background suppression and false alarm control make it particularly suitable for offline or batch processing scenarios, such as post-mission analysis and surveillance data review, where accuracy takes precedence over real-time requirements. This innovative modeling approach provides a new technical solution for weak and small target detection in complex scenarios.

Keywords: infrared small target detection; low-rank clustering; weighted nuclear norm; adaptive; energy shrinkage

1. Introduction

Infrared detection offers advantages such as all-weather operation, long-range detection [1], low background redundancy information, wide night-vision range, strong penetration capability, and strong anti-interference ability [2]. Due to these advantages, it has found wide applications in both military and civilian fields [3], including missile defense [4], intelligent traffic management [5], radar detection technology [6], and forest fire early warning [7]. However, several challenges still limit the detection of dim and small targets. On the one hand, due to the very long imaging distance, the target size in the image is extremely small, possibly occupying only a few pixels. This means the target lacks sufficient spatial information and textural features, making it difficult to extract effective features for detection [8], and the target shape varies significantly [9]. On the other hand, the background of infrared images is often complex, containing various interfering factors such as clouds, trees, and buildings, and the target signal strength is very low, making it easily overwhelmed by complex background noise [10]. These issues not only pose challenges for infrared dim and small target detection [11] but also limit the effectiveness of advanced computer vision techniques and manual interpretation [12].

Current mainstream infrared dim and small target detection methods can be broadly classified into traditional model-driven methods and data-driven (deep learning) methods. Among traditional methods, the Infrared Patch-Image (IPI) model and its derivative algorithms, which separate background and target through low-rank sparse decomposition, are representative. However, they are prone to target energy loss or background residue under complex noise and structured background interference. Weighted Nuclear Norm Minimization (WNNM) methods improve the accuracy of low-rank matrix recovery by applying differentiated weights to different singular values. However, they still do not fully account for the statistical heterogeneity of local image regions, leading to suboptimal suppression performance in varying backgrounds.

To overcome the above limitations, this paper proposes an improved WNNM model incorporating low-rank clustering. The core innovations are:

1. Introduction of a low-rank clustering mechanism and its optimized model. A matching mechanism combining local statistical features (grayscale variance) and spatial Euclidean distance with weighted fusion is used for adaptive grouping of image patches, ensuring higher similarity and low-rankness within each group. On this basis, an energy shrinkage mechanism is designed. During the solution process for each group, the energy proportion of the sparse part (target image) is monitored. When it falls below 20% of the original energy, it is forcibly increased to 30%, effectively alleviating the problem of target energy dilution during aggregation in traditional low-rank clustering methods.

2. Design of category-adaptive weighted nuclear norm weights for different groups of image patches. The penalty strength for singular values is dynamically adjusted according to the statistical characteristics of each class, thereby better preserving target information while suppressing the background.

Experiments are conducted on six scenes with different background complexities and target motion characteristics. Comparisons are made with six mainstream detection methods (including single-frame and multi-frame algorithms) and two deep learning methods. Quantitative evaluation employs BSF, SNR, and three AUC metrics. Experimental results demonstrate that the proposed algorithm achieves optimal performance in background suppression capability ($AUC_{(p_f, \tau)}$) and leads in overall detection performance. Ablation studies verify the effectiveness of each module.

The structure of this paper is as follows: Section 2 reviews related work on infrared dim and small target detection. Section 3 details the proposed low-rank clustering improved WNNM model and its solution method. Section 4 validates the method's effectiveness through multiple experiments, including objective metric comparisons and discussion. Section 5 concludes the paper and outlines future research directions.

2. Related Work

Currently, infrared small target detection techniques are broadly categorized into data-driven deep learning methods and traditional model-driven methods, each with various subcategories.

2.1. Traditional Methods for Dim and Small Target Detection

Traditional infrared dim and small target detection methods can be further divided into single-frame algorithms and multi-frame algorithms [13]. For single-frame detection algorithms, the Infrared Patch-Image (IPI) model proposed by Gao et al. [14] is a representative algorithm. This model can simultaneously achieve dim and small target detection and background prediction without requiring prior target size and background information. Building on this theory, Dai et al. [15] improved it by introducing reweighting and structure tensor concepts, further enhancing background suppression. However, these algorithms generally have high computational complexity, leading to long detection times and poor real-time performance. Researchers have also improved the traditional Top-Hat algorithm from various perspectives. Bai et al. [16] optimized its structural window, using the central area to extract the target while utilizing surrounding areas to obtain the local background. Zhang et al. [17] proposed an improved algorithm based on grayscale values adopting differentiated strategies, effectively enhancing the target and suppressing the background. Other single-frame methods include the Weighted Strengthened Local Contrast Measure (WSLCM) proposed by Han et al. [18] and the Local Contrast Measure (LCM) method by Chen et al. [19].

Multi-frame detection algorithms primarily rely on the spatio-temporal correlation between consecutive frames in an image sequence, utilizing the difference in motion characteristics between dim and small targets and the randomness of background noise for detection [20]. Classical algorithms include the 3D-matched filtering algorithm proposed by Zhang et al. [21], which sets up 3D-matched filters for all possible target trajectories and calculates the Signal-to-Noise Ratio (SNR) for each, selecting the trajectory with the maximum SNR as the target. Liu et al. [22] proposed a variable weighted pipeline filtering algorithm, which uses an adaptive learning scheme to modify the pipeline center based on target position, effectively suppressing edge noise interference. Pang et al. [23] formulated the target detection problem as a low-rank sparse tensor optimization problem based on spatio-temporal prior information of the target and background. They defined a

multi-directional edge-aware map based on facet derivatives and integrated it as sparse prior information into the Spatial–Temporal Tensor Model (STTM), significantly improving detection performance for dim and small targets. Liu et al. [24] proposed a Fast Graph Laplacian Regularization with Minimax Concave Penalty (FGLR-MCP) method, which constructs a graph structure using each frame of the sequence as a vertex, significantly reducing computational complexity. It replaces the traditional nuclear norm with fast graph Laplacian regularization to describe low-rankness, avoiding Singular Value Decomposition (SVD), and uses a non-convex MCP function to enhance sparse modeling capability, achieving a good balance between detection speed and accuracy on multiple sequences. Liu et al. [25] further proposed an efficient detection method based on Representative Coefficient Total Variation Regularization (RCTVW). This method imposes total variation constraints on the representative coefficient matrix to describe local smooth priors, based on the theory that the representative coefficient matrix inherits the spatial structure of the data matrix. It maintains low-rankness by controlling the number of columns of the coefficient matrix, avoiding complex regularization terms and SVD computations. Simultaneously, a weighted l_1 norm based on the log-sum function is used to adaptively enhance target sparsity. This method significantly outperforms most existing low-rank methods in detection efficiency and possesses good noise robustness. The various algorithms introduced above are relatively classic detection algorithms among traditional methods.

2.2. Deep Learning-Based Dim and Small Target Detection

Traditional infrared small target detection methods have revealed inherent limitations over time [26]. On one hand, they are susceptible to complex backgrounds and noise, exhibiting insufficient detection capability for low-SNR targets and poor adaptability. On the other hand, their recognition speed is slow and generalization ability is weak, making it difficult to effectively unify real-time and accuracy requirements.

Deep learning-based dim and small target detection methods have received extensive research in recent years. For example, Yao et al. [27] introduced a multi-dimensional attention mechanism into ODConv, enhancing small target detection capability through adaptive feature learning and multi-dimensional information capture. Ma et al. [28] proposed an SSD network-based detection algorithm, constructing a dedicated SSD framework for dim and small targets to fully leverage the detail description capability of convolutional features, thereby increasing the feature difference between target and background for accurate extraction. Chen et al. [29] proposed DynamicConv, which improves model adaptability to targets of different scales and shapes with a small increase in computation, enabling better handling of the multi-scale and multi-directional characteristics of infrared small targets. Kou et al. [30] proposed a detection and tracking algorithm combining segmentation networks with multi-strategy fusion. It first usesIRSTNet and a pixel-clustering strategy for candidate target segmentation and extraction, then further filters candidates through a threshold adjustment strategy, ultimately achieving real-time detection and tracking of infrared dim and small targets by combining large-range search, local area tracking, and gate tracking strategies. Pang et al. [31] proposed a novel Spatial–Temporal Tensor Model with Saliency Filter Regularization (STTM-SFR) for infrared small target detection. First, using spatio-temporal priors of background and target, the small target detection task is transformed into a sparse low-rank tensor optimization problem; the overall STTM construction preserves the complete spatio-temporal information of the original infrared image sequence. Then, an SFR term is defined between background and foreground to promote target saliency learning. As deep learning models achieve significant success in infrared small target detection, issues of large parameter sizes, high computational costs, and difficulty in deployment on resource-constrained devices are becoming prominent.

To address this, Li et al. [32] introduced network quantization technology into infrared small target segmentation for the first time, proposing a Symmetric Progressive Decreasing Mixed-Precision Quantization method (SPMix-Q). By analyzing the sensitivity differences of different network layers to quantization, they designed three quantization principles: (1) the first convolution layer maintains full precision (FP32); (2) the encoder sub-network is assigned higher bit-widths than the decoder sub-network; (3) shallower layers are assigned higher bit-widths than deeper layers. Based on these principles, SPMix-Q builds symmetric progressively decreasing modules for weights and activations, maintaining performance close to the full-precision model even at extremely low bit-widths (average weights 2.42 bits, activations 3.82 bits), while compressing model size, memory usage, and computational cost to approximately 1/13, 1/4.6, and 1/29, respectively. It has been successfully deployed on mobile chips, achieving high-precision, high-efficiency real-time infrared small target segmentation. To further balance efficiency and detection performance, lightweight network architectures have recently emerged. Liu et al. [33] proposed the Recurrent Reusable-Convolution Attention Network (RRCANet). The core innovation of this network is the design of a Reusable Convolution Block (RuCB), which iteratively reuses the same set of convolution kernels for feature extraction and refinement, achieving preservation and enhancement of high-level semantic features with almost no extra parameters. The network also introduces a Dual-Interactive Attention Aggregation Module (DIAAM) to promote contextual information interaction and fusion between adjacent layers. Additionally, the authors designed a target-inspired composite loss function (DpT-k Loss) combining the advantages of Dice Loss, Poly Loss, and Top-k Loss, aiming to emphasize hard samples and improve convergence stability. Experiments show that RRCANet achieves performance comparable to state-of-the-art methods on multiple benchmark datasets like NUAA-SIRST and IRSTD-1k, while maintaining a very small parameter count and serving as a plug-and-play module that brings consistent performance improvements to various mainstream infrared small target detection methods. The algorithms described above represent relatively classic ones among deep learning-based detection methods.

Why does this paper not adopt deep learning methods? There are three main reasons: (1) The dataset used in this paper contains diverse scenes (ground-air mixed, target maneuvering, etc.), but each scene has only 80 frames, which is insufficient to train a deep learning model with strong generalization capability from scratch. (2) The goal of this paper is to propose an interpretable and stable detection method without relying on large-scale annotated data, facilitating engineering deployment and parameter tuning. (3) Traditional low-rank models have natural advantages in background suppression, while false alarm control in deep learning methods often requires post-processing. Therefore, this paper adheres to the model-driven route and strives to improve its adaptability in complex backgrounds and target energy preservation capability. Of course, combining the low-rank prior of this paper with deep networks (e.g., as a regularization module for networks) is a valuable future direction.

2.3. Weighted Nuclear Norm Minimization and Its Improvements

Weighted Nuclear Norm Minimization (WNNM) overcomes the over-shrinkage problem caused by traditional nuclear norm's equal penalization of all singular values by assigning different weights to different singular values. It is widely used in image restoration and target separation tasks. However, traditional WNNM methods still lack adaptive modeling capability for the statistical characteristics of local regions when processing infrared images with spatial heterogeneity. In infrared small target detection, WNNM has been directly used for the low-rank term in the IPI framework, achieving better background suppression than the standard nuclear norm. Nevertheless, traditional WNNM

still suffers from the following limitations: (1) Global unified processing: it uses the same weight settings for all image patches, ignoring statistical heterogeneity among different regions (e.g., flat vs. texture regions). (2) Fixed similar patch grouping: practical WNNM typically uses non-local self-similar patch grouping, but the search window size and the number of similar patches are fixed, leading to computational redundancy in flat regions and insufficient structural search in texture regions. (3) Target energy dilution: during the overlapping patch reconstruction process, target energy is diluted by multiple averaging. To address this, this paper proposes integrating the low-rank clustering concept into the WNNM framework. Through a strategy of clustering first and then applying differentiated weighting, more refined background–target separation is achieved. Furthermore, we introduce an optimization model that improves upon the low-rank clustering idea, including an adaptive mechanism in the similar patch matching stage and an energy shrinkage mechanism, significantly alleviating the target signal energy dilution problem during the clustering process.

3. Proposed Method

3.1. Weighted Nuclear Norm Minimization (WNNM)

When using low-rank sparse inversion models to solve dim and small target detection problems, the IPI model is a common algorithm. The traditional infrared image model can be expressed as:

$$f_D(x, y) = f_B(x, y) + f_T(x, y) + f_N(x, y) \quad (1)$$

where f_D, f_T, f_B, f_N and (x, y) represent the original image, target image, background image, random noise image, and pixel position, respectively. Using a fixed-size sliding window and step size, the original image can be divided into several image patches. Each patch is vectorized into a column vector, and these column vectors are combined in their original order to form a new matrix [34], yielding a patch-image model representation for infrared images:

$$D = B + T + N \quad (2)$$

where D, T, B , and N represent the original patch image, target patch image, background patch image, and random noise patch image, respectively. Equation (2) is known as the Infrared Patch-Image model, or IPI model. Clearly, when the selected sliding window size equals the original image size, the IPI model degenerates into the traditional infrared image model.

Its specific constrained model is as follows:

$$\min_{B, T} \|B\|_* + \lambda \|T\|_1 \quad s.t. \quad \|D - B - T\|_F \leq \sigma \quad (3)$$

where $\|\cdot\|_*$ is the matrix nuclear norm; $\|\cdot\|_1$ is the matrix L1 norm; $\|\cdot\|_F$ is the Frobenius norm; and σ represents the strength of the noise.

In most algorithms, the low-rank sparse model often uses the nuclear norm of the low-rank matrix to replace its rank. It is defined as the sum of all singular values after Singular Value Decomposition (SVD). Research by Recht [35] has proven that Nuclear Norm Minimization can accurately reconstruct the low-rank component of a matrix under specific conditions. This study also introduced the soft-thresholding operation, providing an efficient solution for nuclear norm minimization problems and promoting the widespread application of this method across multiple fields. However, in the presence of measurement noise, the reconstruction accuracy of this convex relaxation algorithm significantly decreases, leading to noticeable deviation between its solution and the true solution of the

original rank minimization problem, accompanied by an over-shrinkage phenomenon in rank estimation.

To address the problems arising from the above phenomenon, some scholars have proposed treating different rank components differently to improve the accuracy of low-rank recovery, rather than using a single weight as in traditional nuclear norm minimization. For example, methods like Truncated Nuclear Norm Regularization (TTNR) [36] and Partial Sum Minimization (PSM) [37] retain the first r dominant singular values unchanged and only constrain the smallest $N - r$ singular values. However, these methods have limitations: the rank r of the matrix is difficult to estimate accurately in practice and varies dynamically with data. To more rationally utilize the prior information embedded in different singular values, Gu [38] further proposed the weighted nuclear norm (WNN), as shown below:

$$\|A\|_{\omega,*} = \sum_i \omega_i \sigma_i(A) \quad (4)$$

where $\omega = [\omega_1, \omega_2, \dots, \omega_n]$, $\omega_i > 0$ represents the weight assigned to the singular values.

Substituting the above WNN into the low-rank term of the IPI algorithm's constrained model yields:

$$\min_{B,T} \|B\|_{\omega,*} + \lambda \|T\|_1 \quad s.t. \quad \|D - B - T\|_F \leq \sigma \quad (5)$$

The constrained model above is the Weighted Nuclear Norm Minimization (WNNM) method. WNNM effectively implements the function of assigning different penalties to singular values of different magnitudes in the low-rank matrix. Larger singular values are assigned larger penalties, while smaller singular values receive smaller penalties. In contrast, the standard nuclear norm commonly used in algorithms assigns equal penalties to all singular values of the low-rank matrix. Therefore, WNNM is more conducive to improving image recovery accuracy.

3.2. Low-Rank Clustering Concept and Its Optimization

3.2.1. Basic Low-Rank Clustering Concept

The original infrared images we process contain various characteristics, such as local statistical properties (grayscale mean, variance, etc.) and textural properties. These properties are prevalent throughout the infrared original image. Based on the construction concept of the IPI model, we use a fixed-size sliding window and step size to divide the original infrared image into several image patches. Clearly, these image patches also possess the various characteristics mentioned above. Since the size of an image patch is generally much smaller than the entire original infrared image, it can typically be assumed that the properties of each image patch are relatively constant. Let the image patches obtained by dividing the original infrared image f_D as described above be denoted as $A_1, A_2, \dots, A_k$. Based on a certain characteristic mentioned earlier, these k image patches are classified according to their similarity regarding that characteristic. Taking variance as the selected characteristic for illustration: first, we calculate the grayscale variance for each of the k image patches. Then, an interval is set based on the maximum and minimum variances among these patches. This interval is further divided into r mutually exclusive sub-intervals. Each image patch is then assigned to a sub-interval based on its grayscale variance magnitude. This embodies the "clustering" concept. Clearly, patches assigned to the same sub-interval exhibit a certain similarity in grayscale variance, which embodies the "low-rank" concept.

The above describes the basic concept of low-rank clustering. In the proposed algorithm, the k image patches obtained from the original infrared image f_D are reasonably assigned to specific intervals based on a chosen characteristic. First, a search window W ,

larger than an image patch but smaller than the original image f_D , is set. The sliding step size for this search window and the number of patch images within each interval are also determined. When using the sliding window to partition the original image f_D into patches, the currently segmented patch is designated as the reference patch $A_i (i = 1, 2, \dots, k)$. After determining the reference patch A_i , the search window W begins to slide according to the set step size. For the image patches within the area occupied by W at each slide, their similarity to the reference patch A_i is judged based on quantitative values of a chosen characteristic. A threshold is set. If the difference in the quantitative characteristic value between the reference patch A_i and a candidate patch is less than this threshold, they are considered similar; otherwise, they are dissimilar. After the search window traverses the entire image, the characteristic interval for reference patch A_i is obtained. Subsequently, the sliding window moves to the next reference patch A_{i+1} , and the process continues, traversing the entire image, as shown in Figure 1. This yields several different groups based on the chosen characteristic.

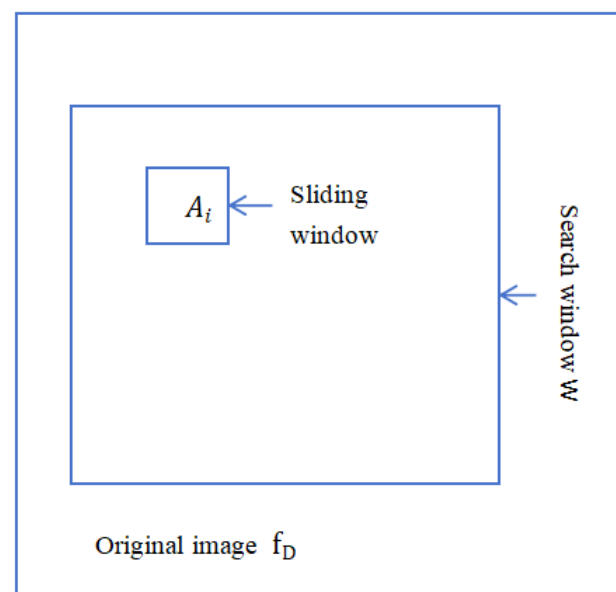


Figure 1. Low-rank clustering classification concept.

However, in practical applications, there is a well-established principle in infrared image processing: adjacent pixels exhibit high similarity. That is, image patches with small Euclidean distances are highly similar. Therefore, the influence of Euclidean distance must be considered when judging the similarity between image patches. Moreover, studies have found that Euclidean distance is a primary factor affecting similarity. Consequently, we ultimately adopt a weighted combination of Euclidean distance and a chosen characteristic to determine similarity between image patches.

3.2.2. Optimization Improvements of the Low-Rank Clustering Concept

Adaptive Similar Patch Matching: In infrared dim and small target detection using the low-rank clustering concept, the similar patch search stage for classifying image patches is a core part of this concept. The main principle is to search for patches similar to a given one based on textural properties, primarily including grayscale variance, eigenvalues of the structure tensor, etc. These properties can describe the texture complexity of a patch. In particular, the eigenvalues of the structure tensor can clearly capture structural similarity between patches, significantly enhancing algorithm robustness. However, using structure tensor eigenvalues imposes a heavier computational burden. As a compromise, using grayscale variance has become a mainstream choice. While grayscale variance may not

be as effective as structure tensor eigenvalues in the similar patch search stage, it offers significantly better performance in terms of algorithm time complexity. In this paper, we employ a similarity matching mechanism based on a weighted fusion of the Euclidean distance between patches and the difference in their grayscale variances, as mentioned in the previous section. For traditional low-rank clustering, the search window size and the number of patches per group are typically fixed values, set before the algorithm runs. This means the same search window size and the same number of similar patches per group are used for patches with different textural characteristics. For patches in flat regions (with simpler texture), a large search window only introduces redundant patches, increasing computation without aiding denoising. For patches in complex regions (with more intricate texture), a small search window limits the ability to find other patches with similar texture, potentially leading to insufficient fitting of the low-rank component and resulting in many false alarms. Based on this analysis of the search window's impact, we decided to dynamically adjust the similar patch search radius for each patch based on the local variance reflecting regional structural complexity, i.e., adapt the search window size based on local variance. The specific method is as follows: Let the grayscale variance of a patch be σ^2 ; the low threshold for grayscale variance be V_{low} , and the high threshold be V_{high} ; the minimum search window radius be r_{min} , and the maximum be r_{max} . Then, the search window radius r for this patch is:

$$r = \begin{cases} r_{min}, & \sigma^2 < V_{low} \\ r_{min} + (r_{max} - r_{min}) \frac{\sigma^2 - V_{low}}{V_{high} - V_{low}}, & V_{low} \leq \sigma^2 \leq V_{high} \\ r_{max}, & \sigma^2 > V_{high} \end{cases} \quad (6)$$

A linear model is used to make the search window radius vary smoothly with patch texture complexity. This allows small windows in flat regions to prevent mismatching with distant noise patches, avoiding the introduction of false textures or over-smoothing; large windows in texture-complex regions capture more true structure, improving low-rank decomposition accuracy and preserving target details in the sparse part. It also avoids extensive exhaustive searches in flat regions, significantly reducing unnecessary patch matching computations, which is particularly important for large images or real-time applications.

The number of patches in each similar patch group is another important parameter in the similar patch matching stage. For flat regions, taking extra patches introduces redundancy, increases computation, and may lead to noise residue in the sparse part. For complex regions, too few patches prevent the low-rank decomposition from capturing sufficient structure, causing details to be mistakenly filtered out as noise. We adopt the core observation that local variance reflects regional structural complexity to dynamically determine the number of similar patches to match for each reference patch, i.e., adapt the number of patches per group based on local variance. The specific method is: Let the grayscale variance of a patch be σ^2 ; the low threshold for grayscale variance be H_{low} , and the high threshold be H_{high} ; the minimum number of patches per group be n_{min} , and the maximum be n_{max} . Then, the number of patches n for this group is:

$$n = \begin{cases} n_{min}, & \sigma^2 < H_{low} \\ n_{min} + (n_{max} - n_{min}) \frac{\sigma^2 - H_{low}}{H_{high} - H_{low}}, & H_{low} \leq \sigma^2 \leq H_{high} \\ n_{max}, & \sigma^2 > H_{high} \end{cases} \quad (7)$$

Again, a linear model ensures the number of similar patches transitions smoothly with complexity. For flat regions, only the minimum number of similar patches is used (set to 5

in this paper), as all patches are highly similar, and a small number suffices to represent the low-rank structure; taking more only adds redundancy and computational cost, potentially causing noise residue. For complex regions, the maximum number of similar patches is used (set to 30 in this paper). Non-local self-similarity is weaker in complex regions, requiring more patches from a wider search area to build a robust low-rank matrix, ensuring that low-rank sparse decomposition effectively separates target details (placed in the sparse part) from background noise (placed in the low-rank part).

Summarizing the relationships between these adaptively optimized parameters: The original image is divided into K patches using a fixed-size sliding window ($p \times p$) and fixed step s . The number of patches K is approximately:

$$K \approx \left(\frac{M-p}{s} + 1 \right)^2 \quad (8)$$

where $M \times M$ is the original image size. Each patch serves as a reference patch, forming a similar patch group through similarity matching. Therefore, the number of similar patch groups is $R = K$. In other words, the number of groups R equals the total number of image patches, determined by the sliding window size and step.

Energy Shrinkage Mechanism:

After the similar patch matching stage, the low-rank clustering method proceeds to solve the constrained model for each group to obtain the corresponding low-rank and sparse components. This is another core part of the proposed algorithm. Traditional low-rank clustering often uses the simplest RPCA method to solve each group. This is mainly because each group contains relatively few patches, and many groups need to be solved. Using overly complex methods risks overfitting the low-rank component, forcing target information from the sparse variable into the low-rank variable, leading to missed detections. Simpler methods also help manage time complexity effectively. However, precisely because low-rank clustering solves multiple groups and reconstructs the final background and target images by aggregating results from all groups, a given patch may appear in different groups, yielding different solutions. During the final aggregation reconstruction, the energy of important target information becomes diluted. This is a persistent problem in traditional low-rank clustering. To mitigate this, we should aim to preserve the energy of the sparse variables for each group as much as possible during the model solution. Although this might also preserve noise, the noise becomes diluted during the final reconstruction, leaving mainly the target information. Analyzing the solution of the traditional simple RPCA method, its constrained model is:

$$\min_{B, T} \|B\|_* + \lambda \|T\|_1 \quad s.t. \quad D = B + T \quad (9)$$

where D is the original variable; B is the low-rank variable; T is the sparse variable; $\|\cdot\|_*$ is the matrix nuclear norm; $\|\cdot\|_1$ is the matrix L1 norm; $\|\cdot\|_F$ is the Frobenius norm. The augmented Lagrangian function is:

$$L = \|B\|_* + \lambda \|T\|_1 + \langle Y, D - B - T \rangle + \frac{\rho}{2} \|D - B - T\|_F^2 \quad (10)$$

where Y is the Lagrangian multiplier; $\langle \cdot, \cdot \rangle$ denotes inner product; $\rho > 0$ is the penalty factor.

For the low-rank variable B , the closed-form solution for its iterative update is:

$$B^{k+1} = \mathcal{D}_{\frac{\lambda}{\rho}} \left(D - T^k + \frac{Y^k}{\rho} \right) \quad (11)$$

where $\mathcal{D}_\tau(X) = U \cdot \text{diag}(\max(\sigma_i - \tau, 0)) \cdot V^T$, $X = U \Sigma V^T$ is the SVD.

For the sparse variable T , the closed-form solution for its iterative update is:

$$T^{k+1} = \mathcal{S}_{\frac{\lambda}{\rho}}(D - B^{k+1} + \frac{Y^k}{\rho}) \quad (12)$$

where $\mathcal{S}_{\tau}(x) = \text{sign}(x) \cdot \max(|x| - \tau, 0)$.

The energy shrinkage mechanism introduced in this paper involves monitoring the energy indicator of the current sparse variable during its iterative update. Here, the energy indicator is defined as the square of the Frobenius norm of the matrix. Calculate the original signal energy $E_{orig} = \|D_j\|_F^2$, where D_j is the input similar patch group matrix for the current group. Calculate the signal energy of the current sparse variable T^k : $E_{T^k} = \|T^k\|_F^2$. When $E_{T^k} < a \cdot E_{orig}$ (with $a = 0.20$ in this paper), indicating insufficient sparse variable energy, execute:

$$T^k \leftarrow T^k \cdot \text{scale}, \quad \text{scale} = \sqrt{\frac{b \cdot E_{orig}}{E_{T^k} + \varepsilon}} \quad (13)$$

where $\varepsilon > 0$ prevents division by zero. This boosts the energy of the insufficient sparse variable to b times the original energy (with $b = 1.3$ in this paper), while also adjusting the low-rank variable to maintain consistency: $B^k \leftarrow D_j - T^k$. A mild singular value shrinkage (coefficient 0.02) is applied to the new low-rank variable to ensure its low-rank property is not compromised.

This ensures that if the sparse variable's energy proportion is too low, indicating target details are not fully separated and may be incorrectly assigned to the low-rank variable, the energy boost forces the target to remain in the sparse output.

3.3. Improved Weighted Nuclear Norm Algorithm Based on Low-Rank Clustering

In the WNNM algorithm mentioned in Section 3.1, larger singular values should be assigned larger penalties, while smaller singular values receive smaller penalties. Based on this idea, Sun [39] provided a specific expression for the weights in the weighted nuclear norm (WNN):

$$\omega_i = \frac{C}{\sigma_i(B) + \varepsilon} \quad (14)$$

where $\sigma_i(B)$ is the i -th singular value of the low-rank matrix B ; C is a positive constant; $\varepsilon > 0$ prevents division by zero. This achieves the desired functionality.

We introduce the low-rank clustering concept from Section 3.2 into the proposed algorithm's model. First, we reconstruct a matrix for each of the R groups classified by low-rank clustering. This involves vectorizing the image patches in each group into column vectors and rearranging them into a matrix, as shown in Figure 2.

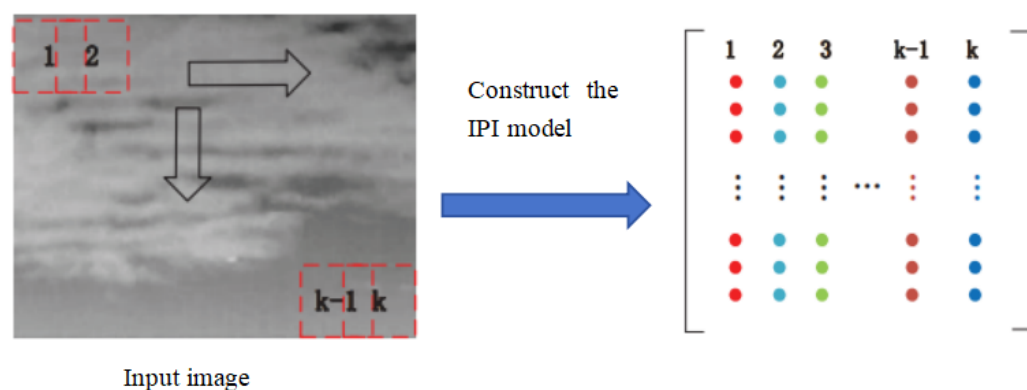


Figure 2. Schematic diagram of IPI model construction.

This yields R original patch images D_j ($j = 1, 2, \dots, R$). We use the local statistical characteristic—variance—as the characteristic in the low-rank clustering concept. Therefore, the proposed algorithm ultimately uses a weighted combination of variance and Euclidean distance to judge patch similarity. As mentioned, patches within each group exhibit a degree of similarity; hence, the original patch images D_j ($j = 1, 2, \dots, R$) generated from each group possess potential low-rankness. Moreover, there are significant characteristic differences among these D_j . Based on this, we can set different constrained models for each original patch image D_j according to its characteristic differences. Thus, our innovation is to improve upon Formula (6) and provide the weighted nuclear norm formula set by the proposed algorithm for different original patch images D_j :

$$\omega_{ji} = \frac{C_j}{\sigma_i(B_j) + \varepsilon} \quad (15)$$

where B_j is the low-rank matrix of the j -th group; ω_{ji} is the weight for the i -th singular value of the low-rank matrix B_j of the j -th group; C_j is the characteristic constant for the j -th group; $\varepsilon > 0$ prevents division by zero.

The characteristic constant C_j ($j = 1, 2, \dots, R$) for each group is set based on the variance characteristic of that group. We define it using a linear method:

$$C_j = \begin{cases} C_{\max}, & V_j < V_{\text{low}} \\ C_{\max} - (C_{\max} - C_{\min}) \frac{V_j - V_{\text{low}}}{V_{\text{high}} - V_{\text{low}}}, & V_{\text{low}} \leq V_j \leq V_{\text{high}} \\ C_{\min}, & V_j > V_{\text{high}} \end{cases} \quad (16)$$

where V_j is the average variance of the j -th group; V_{high} and V_{low} are the upper and lower bounds of the average variances of the R groups; $C_{\max}$ and $C_{\min}$ are the upper and lower bounds of the characteristic constants for the R groups.

Thus, we obtain the low-rank term constraint for each group:

$$\|A_j\|_{\omega_j, *} = \sum_i |\omega_{ji} \sigma_{ji}(A_j)|_1 \quad (17)$$

where $\omega_{ji} = [\omega_{j1}, \omega_{j2}, \dots, \omega_{jn}]$, $\omega_{ji} > 0$ represents the weight assigned to the singular value.

Next, we set the constrained model for each group. The model for the j -th group ($j = 1, 2, \dots, R$) is:

$$\min_{B_j, T_j} \|B_j\|_{\omega_j, *} + \lambda \|T_j\|_1 \quad \text{s.t.} \quad \|D_j - B_j - T_j\|_F \leq \sigma \quad (18)$$

where D_j, B_j, T_j ($j = 1, 2, \dots, R$) are the original patch image, background patch image, and target patch image for the j -th group, respectively, and $\sigma > 0$ represents the noise strength for the j -th group.

By solving each model, the background tensor B_j and target tensor T_j for each group are obtained. The background and target patches from these tensors are then arranged in their original order to form the final background image B and target image T . This constitutes the core of the proposed algorithm.

3.4. Image Patch Clustering

The background patches and target patches from the reconstructed background tensor B_j and target tensor T_j of each group are arranged in their original order to form the final background image B and target image T . However, an overlapping patch issue exists: the same image patch may be assigned to different groups, resulting in potentially different

grayscale values for certain pixels from the solutions of different group models. To address this overlapping patch issue, the proposed algorithm uses a redundant information fusion method. That is, for each pixel point that may have multiple grayscale values from different patches, the final value is taken as the average of all these grayscale values. Imagine each pixel as a “voter,” and each similar patch group as a “voting meeting.” The same pixel may participate in several different voting meetings, each yielding a denoised result. Ultimately, the pixel’s value is determined by the average of all meeting votes. Let pixel (x, y) be covered by N different similar patch groups, with each group providing a denoised value $I_j(x, y)$ ($j = 1, \dots, N$). The final result is:

$$I(x, y) = \frac{1}{N} \sum_{j=1}^N I_j(x, y) \quad (19)$$

$I(x, y)$ is the final grayscale value of pixel (x, y) .

The flowchart of the proposed algorithm is shown in Figure 3.

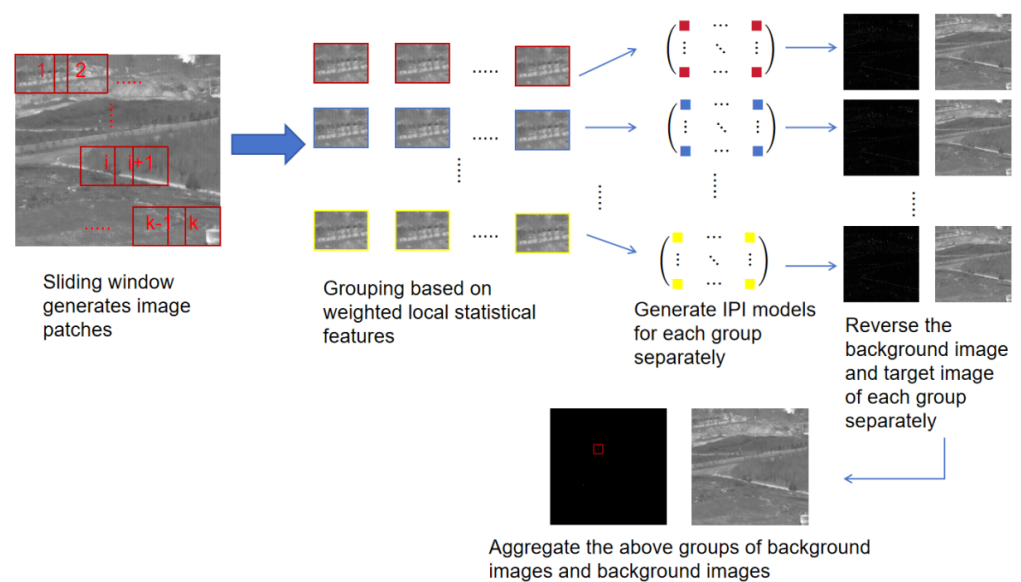


Figure 3. Flowchart of the proposed algorithm.

3.5. Algorithm Solution

For all classified groups, the constrained models have the same form; we can solve each group’s model using the same method. As mentioned, the constrained model for the j -th group is shown in Formula (18).

To solve this constrained model, we adopt the ADMM method based on the solution concept in reference [40], converting the background–target separation problem into a reweighted RPCA problem. First, we introduce the Lagrangian multiplier Y_j and penalty term to construct the Lagrangian function:

$$L = \|B_j\|_{\omega_j, *} + \lambda \|T_j\|_1 + \langle Y_j, D_j - B_j - T_j \rangle + \frac{\mu_j}{2} \|D_j - B_j - T_j\|_F^2 \quad (20)$$

where Y is the Lagrangian multiplier matrix; $\langle \cdot, \cdot \rangle$ denotes inner product; $\mu_j > 0$ is the penalty parameter.

Next, we solve for each iterative variable alternately. The alternating solution for (B_j, T_j) is:

$$\begin{aligned} B_j^{k+1} &= \arg \min_{B_j} L(B_j, T_j^k, Y_j^k, \mu_j^k) \\ &= \arg \min_{B_j} \|B_j\|_{\omega_j, *} + \frac{\mu_j^k}{2} \|D_j - B_j - T_j^k + (\mu_j^k)^{-1} Y_j^k\|_F^2 \end{aligned} \quad (21)$$

$$\begin{aligned} T_j^{k+1} &= \arg \min_{T_j} L(B_j^{k+1}, T_j, Y_j^k, \mu_j^k) \\ &= \arg \min_{T_j} \lambda \|T_j\|_1 + \frac{\mu_j^k}{2} \|D_j - T_j - B_j^{k+1} + (\mu_j^k)^{-1} Y_j^k\|_F^2 \end{aligned} \quad (22)$$

The matrix Y is updated via:

$$Y_j^{k+1} = Y_j^k + \mu_j^k (D_j - B_j^{k+1} - T_j^{k+1}) \quad (23)$$

If $E_{T_j^k} \geq 0.2E_{orig}$, Equations (21) and (22) can be solved via ADMM as:

$$B_j^{k+1} = \mathcal{D}_{(\mu_j^k)^{-1}\omega_j} [D_j - T_j^k - (\mu_j^k)^{-1} Y_j^k] \quad (24)$$

$$T_j^{k+1} = \mathcal{S}_{(\lambda\mu_j^k)^{-1}} [D_j - B_j^{k+1} - (\mu_j^k)^{-1} Y_j^k] \quad (25)$$

where $\mathcal{D}_{(\mu_j^k)^{-1}\omega_j}(\cdot)$ and $\mathcal{S}_{(\lambda\mu_j^k)^{-1}}(\cdot)$ are soft-thresholding operators.

If $E_{T_j^k} < 0.2E_{orig}$, T_j^k is reassigned via (13) to boost energy, and B_j^k is set to $B_j^k \leftarrow D_j - T_j^k$. After this reassignment, the sparse variable meets the energy threshold, and the low-rank and sparse variables continue updating via (24) and (25).

The specific pseudo-code for the above algorithm is shown in Algorithm 1.

Algorithm 1: Pseudo-code of the algorithm in this study.

Input: Block image matrix D_j , parameters $\lambda, \mu, \varepsilon$

Output: B_j, T_j

Step 1: Initialization: $B_j^0, T_j^0 = \mathbf{0}, Y_j^0 = \mathbf{0}, \mu_0 > 0, \rho > 1$

Step 2: Solve using IALM:

while not converge do

Update B_j^{k+1} via: $B_j^{k+1} = \mathcal{D}_{(\mu_j^k)^{-1}\omega_j} [D_j - T_j^k - (\mu_j^k)^{-1} Y_j^k]$

Update T_j^{k+1} via: $T_j^{k+1} = \mathcal{S}_{(\lambda\mu_j^k)^{-1}} [D_j - B_j^{k+1} - (\mu_j^k)^{-1} Y_j^k]$

Update Y_j^{k+1} via: $Y_j^{k+1} = Y_j^k + \mu_j^k (D_j - B_j^{k+1} - T_j^{k+1})$

Update μ_j^{k+1} via: $\mu_j^{k+1} = \rho \mu_j^k$

end while

Step 3: Update weights: $\omega_{ji}^{k+1} = \frac{C_j}{\sigma_{ji}^k (B_j) + \varepsilon}$

Step 4: Stop iteration when $k > k_{\max}$

The above details the specific solution process for the constrained model of the j -th group ($j = 1, 2, \dots, R$). Subsequently, the image blocks derived from the constraint models of all groups are clustered to obtain the final result image.

4. Experimental Results and Analysis

This section presents experiments using 6 different scenes and two evaluation metrics specifically for traditional methods, BSF and SNR, as well as three AUC metrics for

quantitative analysis. 3D-ROC curves are also used to test the detection performance of different algorithms.

4.1. Original Image Information

The raw image data used in this chapter is sourced from Hui et al. at the National University of Defense Technology [41]. Collected and publicly released between 2017 and 2019, these images were captured using midwave infrared imaging equipment (wavelength range 3–5 μm), with each infrared image maintaining a fixed resolution of 256×256 pixels. This dataset captures various scenes against sky and ground backgrounds. Six scenarios were selected for experimental validation. The selected image sequences are shown in Figure 4, with detailed sequence information provided in Table 1.

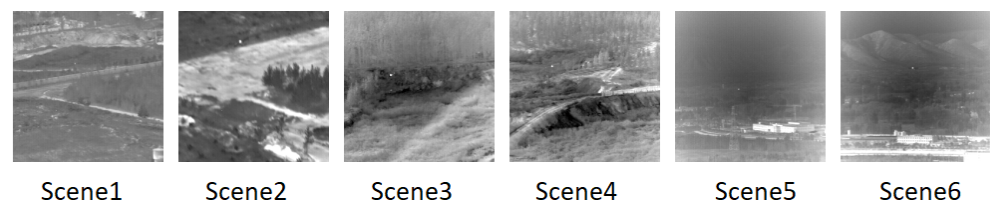


Figure 4. Six scenes of the original image.

Table 1. Scene Information.

Series	Number of Frames	Size	Background Description
Scene 1	80	256×256	Target approaches from far to near, single target, ground background
Scene 2	80	256×256	Target moves from far to near, single target, target maneuvers midway, ground background
Scene 3	80	256×256	Target approaches from far to near, single target, extended target, target maneuvers, ground background
Scene 4	80	256×256	Target approaches from far to near, single target, ground background
Scene 5	80	256×256	Single target, target maneuvers, ground background
Scene 6	80	256×256	Single target, target maneuvers, air-ground background

4.2. Parameter Analysis and Setting

Certain important parameters in the proposed low-rank clustering improved WNNM model significantly affect model performance. In this section, we analyze the influence of parameters on the model and validate the results through experiments. During the solution process, we used grid search for the energy shrinkage parameters (a, b) and the control variable method for parameters (step, size, max_iter).

First, we determine the two hyperparameters a and b of the energy shrinkage mechanism. In natural images, the energy of the sparse component (target) typically accounts for 20–25% of the total energy. Below 20% usually means the target signal is indistinguishable from the noise level and is easily misassigned to the low-rank background. Therefore, we set the range of a to (0.15, 0.25). A b that is too small cannot significantly improve the target energy, while a b that is too large will excessively boost noise energy. Thus, the range of b is set to (1.2, 1.4). We select three representative scenarios to determine the optimal parameters: Scene 1 (complex background texture, weak target radiation), Scene 3 (complex

background texture, strong target radiation), and Scene 5 (simple background texture, weak target radiation). The quantitative analysis is shown in Tables 2–4.

Table 2. Energy shrinkage mechanism parameter optimization for Scene 1 (SNR unit: dB) (The optimal item is in bold).

<i>a</i>	<i>b</i>	BSF	SNR
0.15	1.2	10.2423	6.83
0.15	1.3	11.5654	7.42
0.15	1.4	11.0246	7.05
0.20	1.2	12.8774	8.21
0.20	1.3	13.6286	8.65
0.20	1.4	11.3985	8.33
0.25	1.2	12.9452	7.68
0.25	1.3	10.8571	7.19
0.25	1.4	10.0314	6.72

Table 3. Energy shrinkage mechanism parameter optimization for Scene 3 (SNR unit: dB) (The optimal item is in bold).

<i>a</i>	<i>b</i>	BSF	SNR
0.15	1.2	42.3782	15.26
0.15	1.3	45.8512	16.03
0.15	1.4	44.1252	15.71
0.20	1.2	48.5648	16.95
0.20	1.3	51.1905	17.64
0.20	1.4	49.0316	17.02
0.25	1.2	46.7702	16.44
0.25	1.3	44.2594	15.89
0.25	1.4	41.8572	15.23

Table 4. Energy shrinkage mechanism parameter optimization for Scene 5 (SNR unit: dB) (The optimal item is in bold).

<i>a</i>	<i>b</i>	BSF	SNR
0.15	1.2	15.2384	11.86
0.15	1.3	16.8772	12.54
0.15	1.4	16.1298	12.17
0.20	1.2	18.3428	13.42
0.20	1.3	19.2520	14.05
0.20	1.4	18.6792	13.63
0.25	1.2	17.0851	12.91
0.25	1.3	16.2393	12.38
0.25	1.4	15.4689	11.75

Based on the grid search results, the optimal parameter combination is uniformly $a = 0.20$, $b = 1.3$ across all three scenarios. This set of parameters achieves optimal detection performance in all three representative scenarios, verifying its effectiveness and robustness across different scenarios. In the subsequent parameter analysis, all experiments use this fixed set of energy shrinkage parameters.

Next, we determine the other parameters in the model: sliding window size (size), window sliding step (step), and number of iterations (max_iter). Based on the analysis, we selected optimal parameters for each scene as summarized below.

(a) Sliding Window Size: To construct a patch image matrix, the entire image is traversed by a sliding window with a fixed step. Data within the window forms a column vector, and these vectors construct the patch matrix. Therefore, the sliding window size significantly affects matrix construction.

(b) Window Sliding Step: The sliding step directly affects the overlap between image patches, thus influencing their correlation. A smaller step increases overlap and correlation between patches.

(c) Number of Iterations: The number of iterations plays a crucial role in model solution. Too few iterations may lead to insufficient convergence, poor denoising, and residual information containing not just noise but useful data. Too many iterations can cause long runtimes with negligible denoising improvement, potentially leading to overfitting and loss of image details, while dramatically increasing computation time.

Based on the qualitative analysis of these three parameters, they all significantly impact the proposed algorithm. To further determine optimal parameters, we conducted a quantitative analysis. We selected three representative scenes for parameter selection: Scene 1 (complex background texture, weak target radiation), Scene 3 (complex background texture, strong target radiation), and Scene 5 (simple background texture, weak target radiation), as shown in Tables 5, 6, and 7, respectively.

Table 5. Optimal parameter selection for Scene 1 (SNR unit: dB) (The optimal item is in bold).

Step	Size	Max_iter	BSF	SNR
3	6	3	5.2458	4.17
4	6	3	4.9409	5.09
5	6	3	4.9537	5.47
6	6	3	4.1094	−15.50
5	5	3	3.3119	5.81
5	7	3	5.0045	−15.67
5	8	3	5.3409	−15.86
5	9	3	7.1762	−15.50
5	6	1	13.6286	8.65
5	6	2	8.2244	6.86
5	6	4	3.6007	4.74
5	6	5	3.0632	4.15

From Table 5, regarding step selection, although no single value optimizes both metrics, considering BSF and SNR together, and considering that smaller steps increase runtime, step = 5 was chosen. For size selection, many values resulted in missed detections (negative SNR), so size = 6 was chosen based on BSF and SNR. For Max_iter, because the algorithm achieves sufficient separation of sparse and low-rank components within one iteration through energy redistribution and adaptive thresholds, more iterations can disrupt target energy balance, introduce artifacts, and reduce efficiency; hence, Max_iter = 1 was chosen. In summary, step = 5, size = 6, max_iter = 1 were selected as optimal for the scene with complex texture and weak target radiation.

From Table 6, regarding step selection, although step = 3 gives maximum BSF and SNR, the runtime is too long. Since the results for step = 4 are close to step = 3, step = 4 was chosen considering these factors. Size = 8 clearly outperforms others, so it was chosen. Max_iter = 1 was chosen for reasons similar to Table 5. In summary, step = 4, size = 8, max_iter = 1 were selected as optimal for the scene with complex texture and strong target radiation.

From Table 7, for step selection, step = 3 gives maximum BSF and SNR with a notable gap from other values. Although runtime is longer, BSF and SNR are primary considerations, so step = 3 was chosen. Size = 9 clearly outperforms others, so it was chosen. Max_iter = 1 was chosen for reasons similar to Table 5. In summary, step = 3, size = 9, max_iter = 1 were selected as optimal for the scene with simple texture and weak target radiation.

Table 6. Optimal parameter selection for Scene 3 (SNR unit: dB) (The optimal item is in bold).

Step	Size	Max_iter	BSF	SNR
3	6	3	12.3322	14.92
4	6	3	12.3032	14.22
5	6	3	12.5643	12.34
6	6	3	10.4648	14.16
4	7	3	13.6615	12.91
4	8	3	16.5588	15.02
4	9	3	15.5251	14.10
4	5	3	9.0142	12.74
4	8	1	51.1905	17.64
4	8	2	38.8538	17.30
4	8	4	12.7411	14.99
4	8	5	10.6246	14.45

Table 7. Optimal parameter selection for Scene 5 (SNR unit: dB) (The optimal item is in bold).

Step	Size	Max_iter	BSF	SNR
3	6	3	7.8892	9.48
4	6	3	7.6508	6.15
5	6	3	7.2461	−16.77
6	6	3	6.3544	6.04
3	5	3	7.2787	4.90
3	7	3	9.3284	9.37
3	8	3	9.7193	8.02
3	9	3	9.5532	11.27
3	9	1	19.2520	14.05
3	9	2	12.3842	8.83
3	9	4	9.1646	7.72
3	9	5	8.1032	7.45

Based on the above analysis, Scene 1 uses parameters step = 5, size = 6, max_iter = 1; Scenes 2, 3, and 4 use step = 4, size = 8, max_iter = 1; Scenes 5 and 6 use step = 3, size = 9, max_iter = 1.

4.3. Experimental Analysis

Detecting dim and small targets in complex backgrounds is challenging. To effectively improve detection performance, this paper proposes an infrared dim and small target detection method based on low-rank clustering improved WNNM. The method uses low-rank clustering to classify image patches based on a weighted combination of local statistical features and Euclidean distance. Low-rank sparse models are constructed and solved for the patch matrices formed from each group. Finally, the resulting patches from each group are aggregated to obtain the final target and background images. Experiments were conducted on sequential images, focusing on evaluating the proposed method's noise suppression capability in complex space backgrounds and its target detection accuracy in the spatio-temporal domain. Experimental results are shown in Figures 5 and 6. In these

figures, (a–f) represent scenes 1–6; columns 1, 2, and 3 show the original image, the binary result image from the algorithm, and the SNR image, respectively. The red box indicates the labeled target neighborhood.

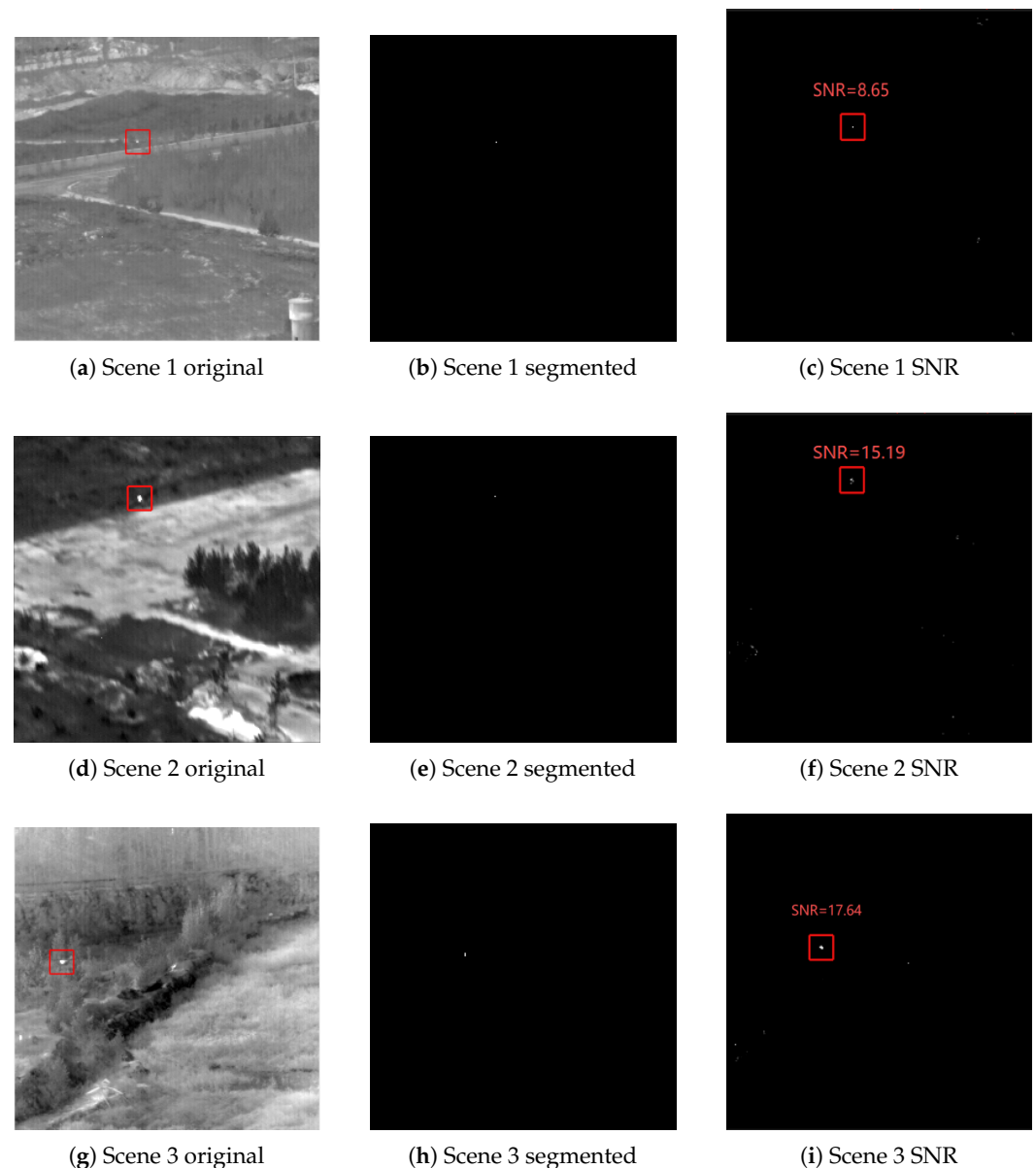


Figure 5. Result images for Scenes 1–3.

From the experimental results, in terms of background suppression, the target images obtained after solving the model improved by low-rank clustering show significantly weakened background noise. Comparing the original images (column 1) with the background-modeled target images (column 2), most background clutter has been removed, with only slight residual noise. In conclusion, the experimental results demonstrate that the proposed method effectively suppresses noise and achieves effective detection of dim and small targets in complex space backgrounds.

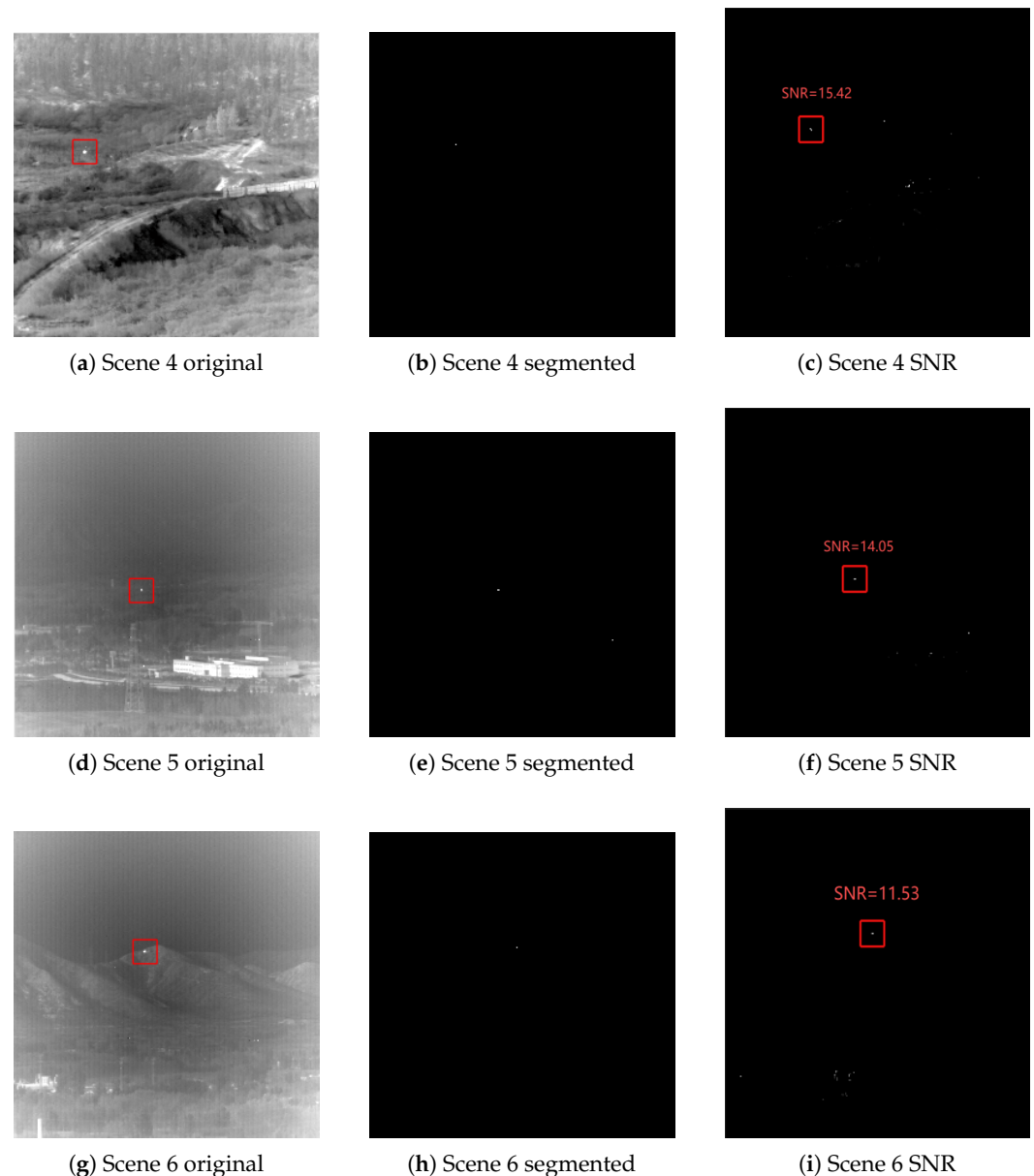


Figure 6. Result images for Scenes 4–6.

To better demonstrate the performance of the proposed algorithm, it was compared with six different target detection methods across 6 scenes. These include 4DST-BTMD [42], ASTTV-NTLA [43], PSTNN [44], GDR-PCD [45], STWRT [46], and IPI [14]. Among them, 4DST-BTMD, ASTTV-NTLA, and STWRT are multi-frame detection algorithms, while the proposed algorithm is single-frame. The purpose of comparing it with multi-frame algorithms is to demonstrate the superiority of the proposed algorithm in macroscopic detection performance (such as the three AUC metrics) compared to typical multi-frame methods. PSTNN and GDR-PCD are single-frame algorithms; comparing with them aims to show the superiority of the proposed algorithm in microscopic detection performance (such as BSF, SNR). The results for the seven algorithms across Scenes 1 to 6 are shown in Figures 7–12. The lower right corner of each image is a partial enlarged view of the target area. In each figure, (a) is the original image, (b) is the proposed algorithm, (c) is 4DST-BTMD, (d) is ASTTV-NTLA, (e) is PSTNN, (f) is GDR-PCD, (g) is STWRT, and (h) is IPI. The red solid box indicates the labeled target neighborhood, while the red dashed box represents a zoomed-in view of the target neighborhood.

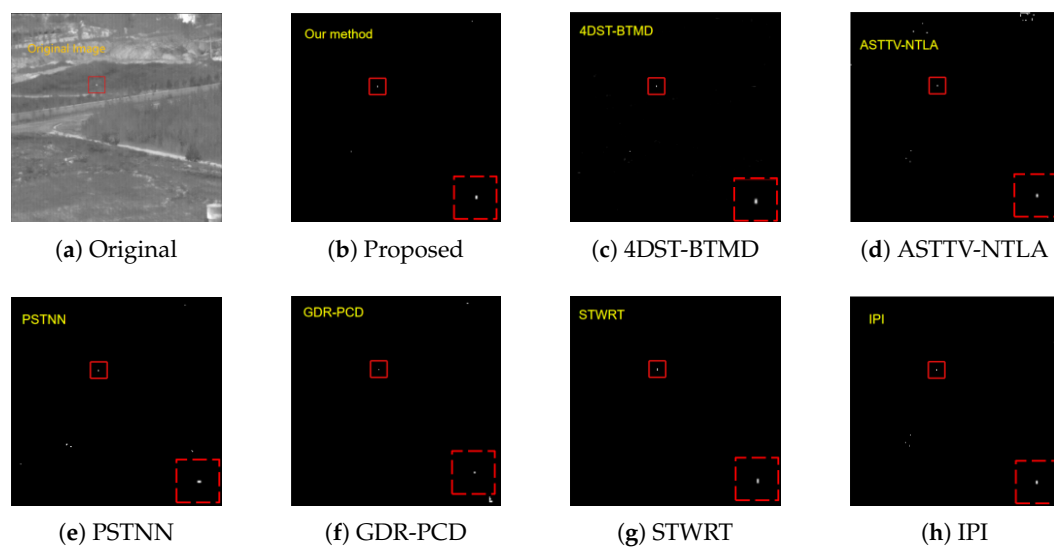


Figure 7. Scene 1 comparison results.

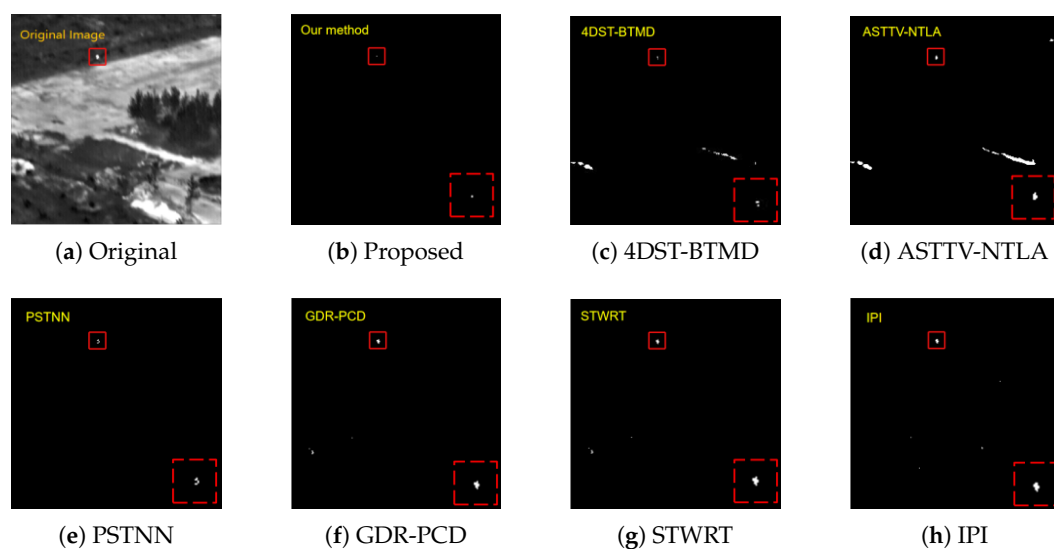


Figure 8. Scene 2 comparison results.

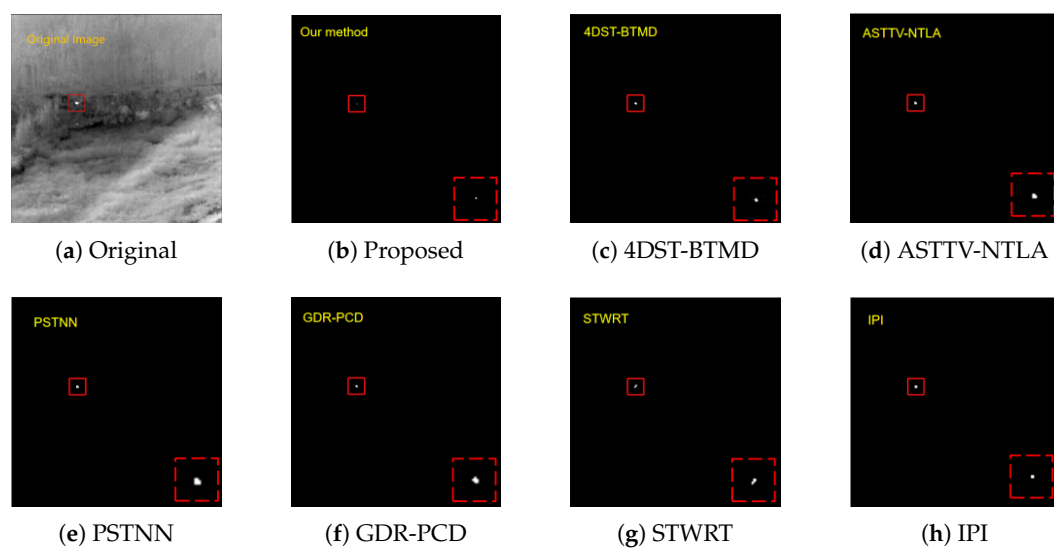


Figure 9. Scene 3 comparison results.

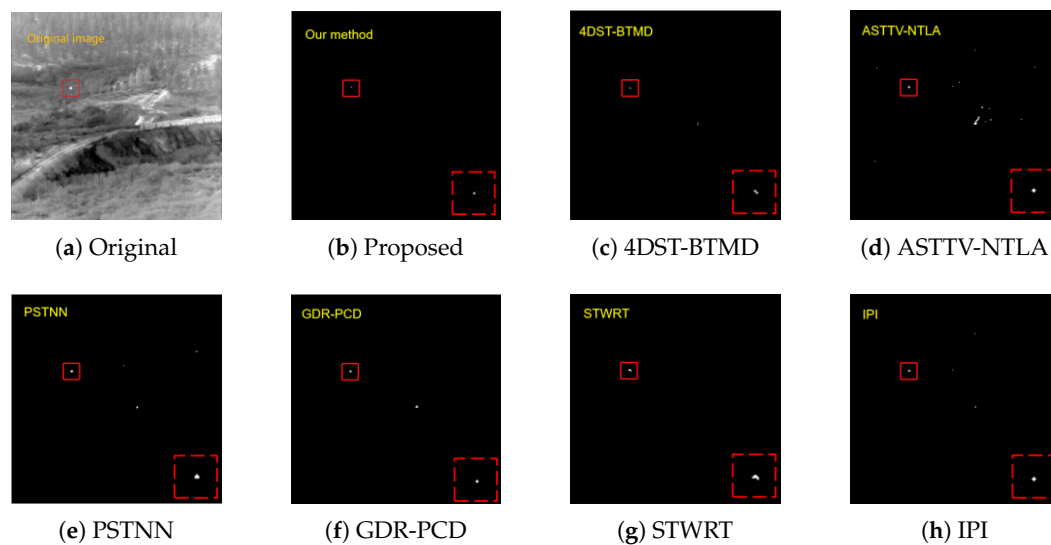


Figure 10. Scene 4 comparison results.

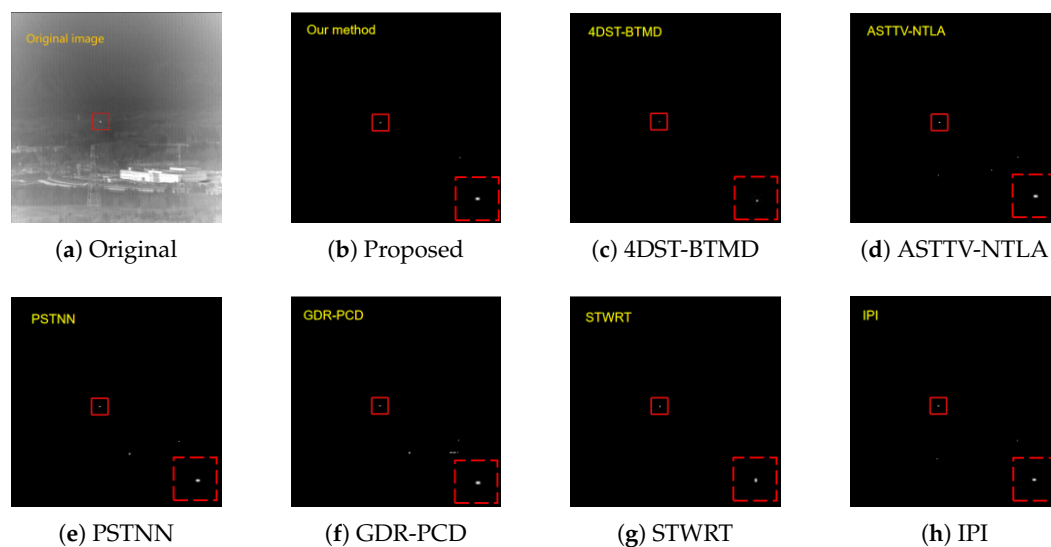


Figure 11. Scene 5 comparison results.

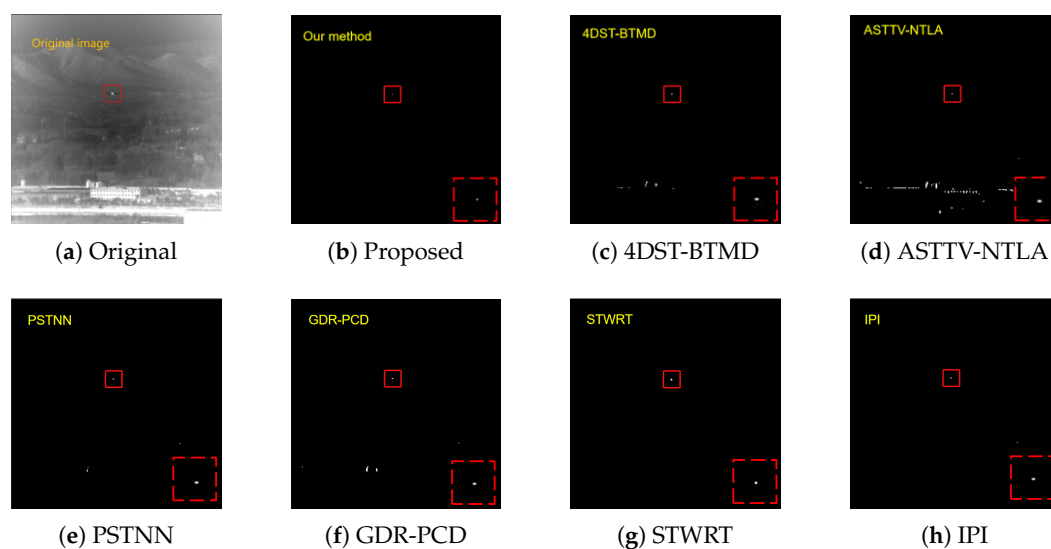


Figure 12. Scene 6 comparison results.

From the experimental results, all seven algorithms possess certain detection capabilities, effectively detecting targets and suppressing background noise to some extent. However, performance differences persist. In terms of target localization accuracy, the proposed algorithm can identify the target region in all scenes, with localization results basically matching the true target position. In terms of background clutter suppression, the non-target region residual grayscale values in the output image of the proposed algorithm are relatively low, resulting in relatively clean background suppression. Although most comparison algorithms can detect targets, there are still obvious background residues or false alarm responses around targets in some scenes, among which IPI and ASTTV-NTLA have more significant residual clutter, GDR-PCD and PSTNN have moderate false alarm levels, while 4DST-BTMD and STWRT perform relatively close to the proposed algorithm in some scenes. In addition, in scenes with weak target radiation energy such as Scene 1 and Scene 5, the target signal intensity detected by the proposed algorithm is relatively low, and the contrast needs further improvement; this problem also exists to varying degrees in other comparison algorithms. Combining target localization accuracy and clutter suppression effects, the comprehensive performance of the proposed algorithm in the 6 scenes has certain advantages over other comparison algorithms. Particularly in scenes with complex background textures, the proposed algorithm can detect targets more accurately, effectively reducing noise interference in detection results, providing a more reliable and efficient solution for target detection tasks.

4.4. Evaluation Metrics Analysis

To further compare the performance of the proposed algorithm with others, we employed several evaluation metrics to assess background suppression, target energy enhancement, and noise elimination. Three metrics were introduced: Background Suppression Factor (BSF), Signal-to-Noise Ratio (SNR), and AUC. BSF evaluates the effectiveness of the algorithm's predicted background image using the difference image and the original image. The calculation formula is:

$$BSF = \sigma_{in} / \sigma_{out} \quad (26)$$

where σ_{in} and σ_{out} are the standard deviations of the input image and the difference image, respectively.

SNR assesses the contrast between target and noise energy obtained by the algorithm, validating detection and background estimation effects. In this paper, SNR refers to the local SNR, typically taking a background region three times the size of the target region. Its calculation involves the mean of the target region, the mean of the background region, and the standard deviation of the background region. The SNR value clearly indicates the prominence of target information against background noise, thereby judging the method's ability to extract target information. The formula is:

$$SNR = 10 \times \log_{10} \left(\frac{A_E - A_B}{\sigma_B} \right) \quad (27)$$

where A_E is the mean of the target region; A_B is the mean of the background region; σ_B is the standard deviation of the background region.

The calculated results for the proposed algorithm and the other six comparison algorithms across 6 scenes for these two metrics are shown in Tables 8 and 9.

According to the BSF comparison data in Table 8, the proposed algorithm demonstrates a clear advantage in background suppression capability. Across the 6 scenes, the proposed algorithm achieves the highest BSF values in Scenes 1, 3, 5, and 6, at 13.6286, 51.1905, 19.2520, and 21.6940, respectively, significantly outperforming other algorithms. In Scene 2, the proposed algorithm's BSF is 46.7646, slightly lower than STWRT's 50.8724, but still substantially higher than others (e.g., GDR-PCD's 35.8935, IPI's 30.9752). In Scene 4, the proposed algorithm's BSF is 38.316, close to the optimal value of 41.7321 (STWRT), and higher than all other comparison algorithms except STWRT.

Table 8. Comparison of BSF metric (The optimal item is in bold).

BSF	Proposed	4DST-BTMD	ASTTV-NTLA	PSTNN	GDR-PCD	STWRT	IPI
Scene 1	13.6286	10.4661	8.2102	9.1322	6.9284	11.4989	6.1985
Scene 2	46.7646	50.8724	39.7392	44.6573	35.8935	49.7892	30.9752
Scene 3	51.1905	39.6733	36.3674	48.0464	35.9705	49.5950	37.6812
Scene 4	38.316	38.0279	27.6896	38.6434	28.8356	41.7321	26.9525
Scene 5	19.2520	15.7894	9.7856	18.4348	12.8746	17.3278	8.6974
Scene 6	21.6940	18.2028	15.7569	17.9307	12.7969	19.8954	10.6281

Overall, the proposed algorithm's BSF is at a leading level in all scenes, particularly excelling in complex scenes with strong background clutter and weak target energy (e.g., Scenes 1 and 5). This indicates that the proposed low-rank clustering improved WNNM method effectively suppresses background noise and enhances background–target separation. Compared to the multi-frame algorithm STWRT, the proposed algorithm achieves better BSF in most scenes, demonstrating that a single-frame framework with adaptive clustering and energy shrinkage can achieve or even surpass the background suppression performance of multi-frame methods. Compared to IPI, the BSF improvement is more significant, validating the effectiveness of incorporating the low-rank clustering concept and optimization model. Therefore, the proposed algorithm shows excellent comprehensive performance in the key metric of Background Suppression Factor.

Table 9. Comparison of SNR metric (SNR unit: dB) (The optimal item is in bold).

SNR	Proposed	4DST-BTMD	ASTTV-NTLA	PSTNN	GDR-PCD	STWRT	IPI
Scene 1	8.65	8.73	7.53	9.28	8.25	9.18	6.53
Scene 2	15.19	12.84	13.26	12.64	10.51	14.67	11.54
Scene 3	17.64	15.73	14.82	16.84	12.53	17.21	10.96
Scene 4	15.42	14.89	13.54	13.97	10.67	15.21	9.77
Scene 5	14.05	14.76	12.54	16.38	11.86	15.43	10.89
Scene 6	11.53	12.47	9.36	10.95	9.74	11.38	8.74

According to the SNR comparison data in Table 9, the proposed algorithm also demonstrates strong competitiveness in target energy enhancement. Across the 6 scenes, the proposed algorithm achieves the highest SNR values in Scenes 2, 3, and 4, at 15.19, 17.64, and 15.42, respectively, significantly outperforming other algorithms. In Scene 1, its SNR of 8.65 is slightly lower than STWRT's 9.18 and PSTNN's 9.28, but higher than other algorithms. In Scene 5, its SNR is 14.05, slightly lower than PSTNN's 16.38 and STWRT's 15.43, ranking third. In Scene 6, its SNR is 11.53, slightly lower than 4DST-BTMD's 12.47, but higher than other comparison algorithms.

Overall, the proposed algorithm's SNR is at a leading or forefront level in most scenes, indicating it not only effectively suppresses the background but also adequately preserves target energy in the sparse part. Compared to IPI, the SNR improvement is significant (e.g., Scene 3: 17.64 vs. 10.96), validating the effectiveness of the energy shrinkage mechanism

and adaptive clustering strategy in enhancing target contrast. Compared to the multi-frame algorithm STWRT, the proposed algorithm performs better in Scenes 2, 3, and 4, suggesting that an optimized single-frame model can achieve excellent target enhancement. It is noted that in Scenes 1 and 5, the proposed algorithm's SNR is slightly lower than the optimal algorithm, indicating room for further optimization in target energy preservation under extremely low SNR or very weak target energy conditions. Overall, the proposed algorithm demonstrates robust and excellent comprehensive performance in the SNR metric.

4.5. Detection Rate and False Alarm Rate Analysis

For a comprehensive evaluation of the algorithms in this chapter and to further highlight the effectiveness of the proposed algorithm, we compare three AUC metrics and plot detection rate vs. false alarm rate vs. threshold (3D-ROC) curves [46] to test performance. The detection rate and false alarm rate are calculated as shown in Formula (28):

$$\begin{cases} Pd = \frac{\text{Number of correctly detected targets}}{\text{Number of detected images}} \\ Pf = \frac{\text{Area occupied by false alarms}}{\text{Image area}} \end{cases} \quad (28)$$

Here, Pd represents the detection rate; Pf represents the false alarm rate. These two metrics are significant for evaluating the accuracy of detection methods. Pd reflects the proportion of correctly detected targets, while Pf reflects the proportion of non-targets incorrectly identified as targets. Together, they provide a comprehensive assessment of detection method performance. The 3D-ROC curves for the six scenes are shown in Figures 13–18.

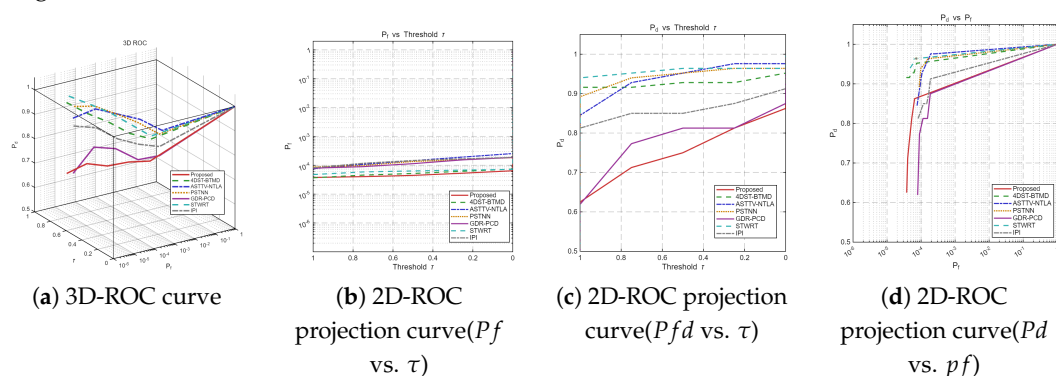


Figure 13. 3D-ROC curves and projected curves for Scene 1.

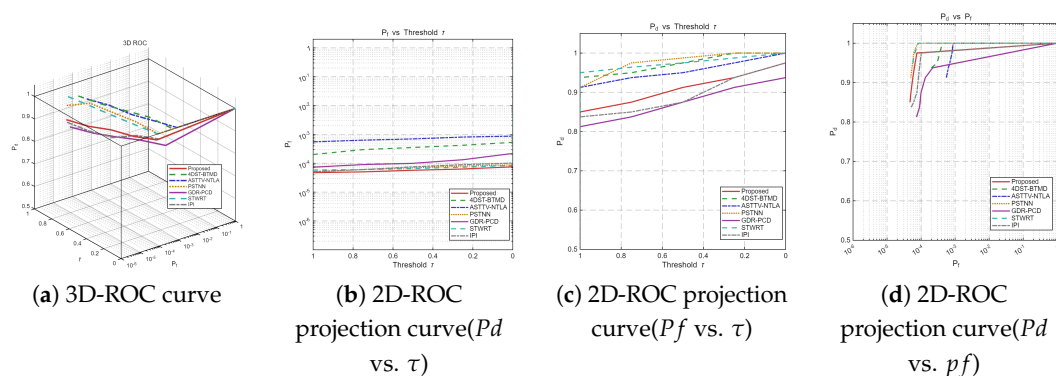


Figure 14. 3D-ROC curves and projected curves for Scene 2.

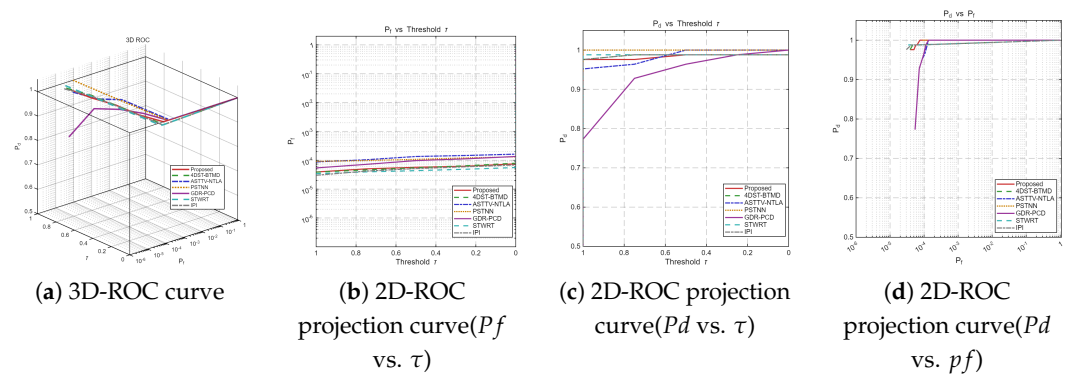


Figure 15. 3D-ROC curves and projected curves for Scene 3.

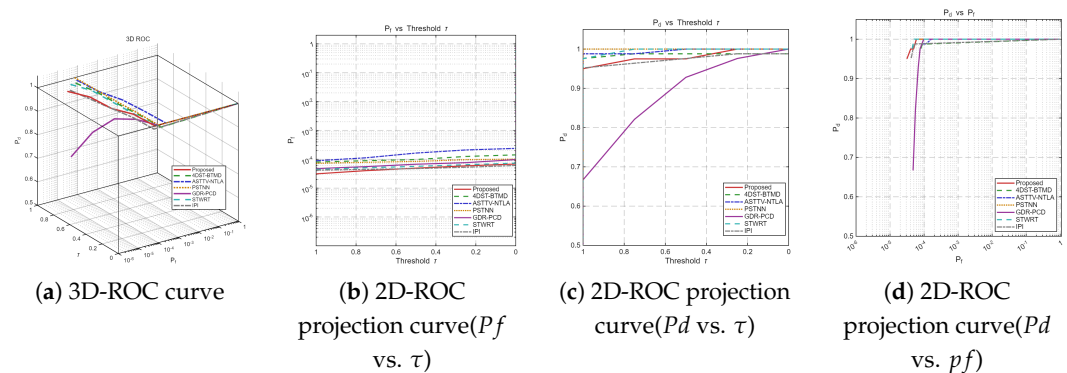


Figure 16. 3D-ROC curves and projected curves for Scene 4.

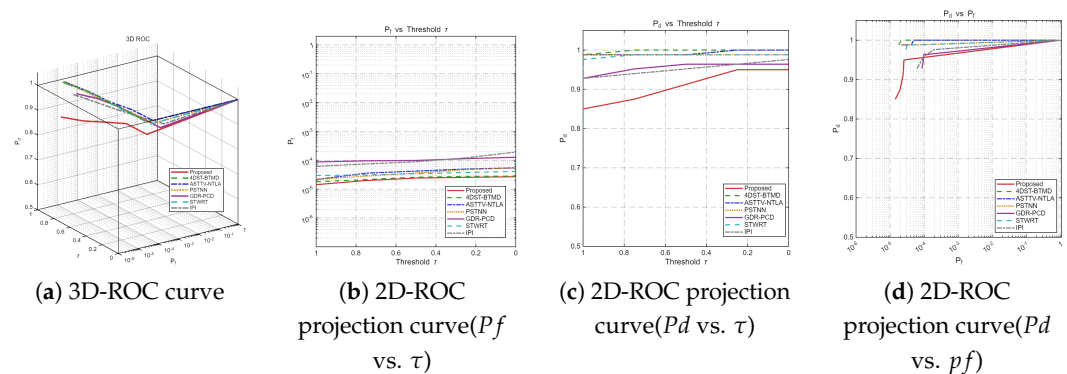


Figure 17. 3D-ROC curves and projected curves for Scene 5.

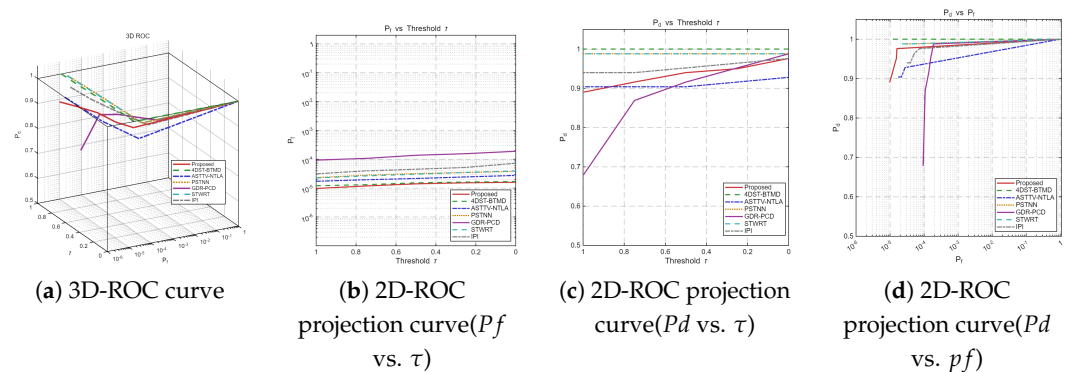


Figure 18. 3D-ROC curves and projected curves for Scene 6.

From the ROC curves for the 6 different scenes, we can calculate the Area Under the Curve (AUC) for each projection plane, namely $AUC_{(P_d, P_f)}$, $AUC_{(P_d, \tau)}$, and $AUC_{(P_f, \tau)} \subset$

(0, 1). Higher values of $AUC_{(P_d, P_f)}$ and $AUC_{(P_d, \tau)}$ indicate better detection performance, while lower values of $AUC_{(P_f, \tau)}$ indicate better background suppression performance. $AUC_{(P_d, \tau)}$ measures the algorithm's ability to detect true targets at different thresholds; $AUC_{(P_f, \tau)}$ measures how fast the false alarm rate decreases as the threshold increases; $AUC_{(P_d, P_f)}$ reflects the overall trade-off between detection rate and false alarm rate.

Table 10 shows the AUC values for the 6 algorithms across the 6 different scenes. The optimal values are bolded.

Table 10. Comparison of three AUC metrics across 6 scenes (The optimal item is in bold).

Scene	Index	Proposed	4DST-BTMD	ASTTV-NTLA	PSTNN	GDR-PCD	STWRT	IPI
Scene 1	$AUC_{(P_d, \tau)}$	0.754687	0.926500	0.941625	0.946000	0.786250	0.958000	0.859375
	$AUC_{(P_f, \tau)}$	0.000047	0.000054	0.000152	0.000132	0.000123	0.000063	0.000142
	$AUC_{(P_d, P_f)}$	0.931223	0.975979	0.987950	0.981950	0.937436	0.981975	0.956194
Scene 2	$AUC_{(P_d, \tau)}$	0.909375	0.973437	0.954687	0.979688	0.875000	0.999969	0.987466
	$AUC_{(P_f, \tau)}$	0.000058	0.000359	0.000714	0.000070	0.000117	0.000068	0.000075
	$AUC_{(P_d, P_f)}$	0.987472	0.999885	0.999685	0.999972	0.968699	0.999969	0.987466
Scene 3	$AUC_{(P_d, \tau)}$	0.985000	0.986500	0.985000	1.000000	0.941625	0.988000	0.986500
	$AUC_{(P_f, \tau)}$	0.000057	0.000056	0.000128	0.000110	0.000094	0.000044	0.000051
	$AUC_{(P_d, P_f)}$	0.999979	0.993980	0.999953	0.999954	0.999962	0.993983	0.993984
Scene 4	$AUC_{(P_d, \tau)}$	0.981250	0.986500	0.995500	1.000000	0.889625	0.997000	0.974500
	$AUC_{(P_f, \tau)}$	0.000050	0.000106	0.000163	0.000087	0.000068	0.000058	0.000049
	$AUC_{(P_d, P_f)}$	0.999983	0.993959	0.999954	0.999964	0.999964	0.999977	0.993978
Scene 5	$AUC_{(P_d, \tau)}$	0.909375	0.998500	0.992500	0.988000	0.956500	0.986500	0.952000
	$AUC_{(P_f, \tau)}$	0.000023	0.000024	0.000041	0.000037	0.000104	0.000035	0.000103
	$AUC_{(P_d, P_f)}$	0.974991	0.999991	0.999989	0.993989	0.981954	0.993985	0.987964
Scene 6	$AUC_{(P_d, \tau)}$	0.934750	1.000000	0.910000	0.988000	0.892500	0.988000	0.953500
	$AUC_{(P_f, \tau)}$	0.000014	0.000015	0.000022	0.000031	0.000035	0.000031	0.000047
	$AUC_{(P_d, P_f)}$	0.987994	0.999994	0.963991	0.993988	0.993931	0.993989	0.987983

According to the comparison data for the three AUC metrics in Table 10, the proposed algorithm demonstrates good balance and robustness in target detection capability ($AUC_{(P_d, \tau)}$), background suppression capability ($AUC_{(P_f, \tau)}$), and overall detection performance ($AUC_{(P_d, P_f)}$).

In terms of target detection capability ($AUC_{(P_d, \tau)}$), the proposed algorithm achieves values of 0.985000 in Scene 3, 0.981250 in Scene 4, 0.909375 in Scene 2, and 0.934750 in Scene 6, all at relatively high levels. Although slightly lower than the optimal values in some scenes (e.g., Scene 1: 0.754687 vs. PSTNN's 0.946000; Scene 5: 0.909375 vs. 4DST-BTMD's 0.998500), overall it maintains a stable detection rate, mostly above 90%, proving its effective target capture capability.

In terms of background suppression capability ($AUC_{(P_f, \tau)}$), the proposed algorithm achieves the minimum (i.e., optimal) values in all 6 scenes: Scene 1: 0.000047, Scene 2: 0.000058, Scene 3: 0.000057, Scene 4: 0.000050, Scene 5: 0.000023, Scene 6: 0.000014. Lower values indicate faster decrease in false alarm rate with increasing threshold, meaning cleaner background suppression. The proposed algorithm significantly outperforms or matches other algorithms in all scenes, especially in Scenes 5 and 6 where $AUC_{(P_f, \tau)}$ is as low as 2.3×10^{-5} and 1.4×10^{-5} , indicating absolute superiority in suppressing background clutter and noise. This benefits from the effective combination of the low-rank clustering mechanism and energy shrinkage strategy, ensuring almost no background residue in the sparse part (target image).

In terms of overall detection performance ($AUC_{(P_d, P_f)}$), the proposed algorithm achieves optimal values in Scenes 3 and 4 (0.999979 and 0.999983), near-optimal values in

Scenes 2 and 6 (0.987472 and 0.987994), and maintains relatively high levels in Scenes 1 and 5 (0.931223 and 0.974991). $AUC_{(P_d, P_f)}$, reflecting the area under the ROC curve, is a comprehensive measure of the trade-off between detection rate and false alarm rate. The proposed algorithm achieves or approaches optimality in most scenes, indicating that it maintains high detection rates while keeping extremely low false alarm rates, demonstrating excellent overall detection performance.

Combining the three AUC metrics, the proposed algorithm shows overwhelming advantage in background suppression capability ($AUC_{(P_f, \tau)}$), robust performance in target detection capability ($AUC_{(P_d, \tau)}$), and leading performance in comprehensive evaluation ($AUC_{(P_d, P_f)}$). Compared to the multi-frame algorithm STWRT, the proposed algorithm is superior in false alarm control. Compared to traditional low-rank methods IPI and WNNM, the proposed algorithm shows significant improvements across all metrics. This fully validates the effectiveness and advancement of the proposed low-rank clustering improved WNNM method for infrared dim and small target detection in complex scenes.

4.6. Ablation Study

To analyze the effectiveness of the innovations in the proposed algorithm, this section compares the traditional low-rank clustering model, the traditional model with only WNNM added, the traditional model with only the optimization model added, and the proposed algorithm's constrained model, using BSF and SNR metrics. We use three representative scenes (Scene 1, Scene 3, and Scene 5) as input images. The comparison of these four models across the three scenes is shown in Tables 11–13. The checkmark in the table indicates that the model adopts the corresponding innovation.

Table 11. Ablation study results (Scene 1) (SNR unit: dB) (The optimal item is in bold).

Model	WNNM	Optimization	BSF	SNR
Traditional low-rank clustering			5.45144	4.325
Traditional + WNNM	✓		8.1772	5.19
Traditional + Optimization		✓	9.7538	5.42
Proposed (WNNM+Optimization)	✓	✓	13.6286	8.65

Table 12. Ablation study results (Scene 3) (SNR unit: dB) (The optimal item is in bold).

Model	WNNM	Optimization	BSF	SNR
Traditional low-rank clustering			18.2347	12.58
Traditional + WNNM	✓		36.4512	15.83
Traditional + Optimization		✓	38.9275	16.21
Proposed (WNNM+Optimization)	✓	✓	51.1906	17.64

Table 13. Ablation study results (Scene 5) (SNR unit: dB) (The optimal item is in bold).

Model	WNNM	Optimization	BSF	SNR
Traditional low-rank clustering			7.8623	9.76
Traditional + WNNM	✓		13.4587	11.92
Traditional + Optimization		✓	14.8732	12.48
Proposed (WNNM+Optimization)	✓	✓	19.2520	14.05

From Tables 11–13, in all three scenes, adding only WNNM brings significant performance improvements compared to the traditional model: in Scene 1, BSF improves by approximately 50% (5.45 → 8.18); in Scene 3, by approximately 100% (18.23 → 36.45); in Scene 5, by approximately 71% (7.86 → 13.46). The varying gains of WNNM across

different scenes are mainly because Scene 3 has stronger target radiation energy, making the distinction between low-rank background and sparse target more pronounced, allowing the adaptive singular value penalty strategy to play a more substantial role. Adding only the optimization model also brings significant improvements: BSF improves by approximately 79%, 114%, and 89% in the three scenes, and SNR improves by approximately 25%, 29%, and 28%, respectively. The optimization model effectively improves target energy preservation capability across different backgrounds through adaptive similar patch matching and energy shrinkage mechanisms.

The complete algorithm achieves BSF of 13.63, 51.19, and 19.25 in the three scenes, representing improvements of approximately 150%, 181%, and 145% over the traditional model; SNR reaches 8.65, 17.64, and 14.05, with improvements of approximately 100%, 40%, and 44%, respectively. The SNR improvement in Scene 1 is the largest (approximately 100%), mainly because the target energy in this scene is extremely weak, and the energy shrinkage mechanism's effect on preserving weak targets is most significant. The BSF improvement in Scene 3 is the largest (approximately 181%), because the separation difficulty between strong targets and complex textured backgrounds is moderate, and WNNM's adaptive penalty and the optimization model's structural search form the strongest complementary effect. The improvements in Scene 5 remain stable across all metrics, verifying the effectiveness of the proposed algorithm under simple backgrounds.

The above data further demonstrate that WNNM focuses on differentiated shrinkage of singular values in the low-rank decomposition process, while the optimization model focuses on data-driven adaptation in patch matching and active preservation of sparse energy; their mechanisms are orthogonal. In summary, the ablation study quantitatively verifies the indispensable contribution of each innovation in the proposed algorithm to improving background suppression and target SNR across three representative scenes, demonstrating the universal effectiveness of each module under different scenarios and the best performance when combined.

4.7. Algorithm Complexity Analysis

The proposed algorithm mainly consists of two stages: the low-rank clustering stage (similar patch search and grouping) and the WNNM optimization solution stage. The time complexity for each stage is analyzed below.

4.7.1. Low-Rank Clustering Stage

This stage divides the original infrared image into multiple patches and computes similarity with other patches within an adaptive search window for each patch, completing grouping. Assuming the original image size is $M \times M$, sliding window size $p \times p$, step s , the number of patches K is given by Formula (8).

For each patch, the adaptive search window radius r is dynamically adjusted based on local variance (upper and lower bounds $r_{\max}, r_{\min}$), with average radius $\bar{r} = (r_{\max} + r_{\min})/2$. Similarity measures (weighted sum of Euclidean distance and grayscale variance difference) need to be computed with all candidate patches within the window. The computational cost for each candidate patch is $O(p^2)$. The number of candidate patches in the window is approximately $(2\bar{r} + 1)^2$. Thus, the clustering complexity for a single patch is $O(p^2 \cdot \bar{r}^2)$. The total clustering complexity for all patches is $O(K \cdot p^2 \cdot \bar{r}^2)$.

4.7.2. Group Optimization Solution Stage

After clustering, K groups are obtained (one per reference patch). Each group j contains n_j patches (adaptive), forming the data matrix $D_j \in \mathbb{R}^{p^2 \times n_j}$. Each group's WNNM model is solved independently using the ADMM framework. The main computational cost is the Singular Value Decomposition (SVD) required for low-rank matrix updates. According

to the solution process, each group completes solving within one iteration ($\text{max_iter} = 1$). The time complexity of SVD for a matrix of size $m \times n$ (assuming $m \geq n$) is $O(mn^2)$. Here, $m = p^2$ and $n = n_j$. Typically, $p^2 > n_j$, so the complexity is $O(p^2 \cdot n_j^2)$. The total solution complexity for all groups is:

$$O\left(\sum_{j=1}^K p^2 \cdot n_j^2\right) \quad (29)$$

Since each patch corresponds to a group and patches overlap, the distribution of n_j relates to local variance. Let the average of n_j be n_{avg} . The total complexity for this stage is approximately:

$$O(K \cdot p^2 \cdot n_{avg}^2) \quad (30)$$

4.7.3. Overall Complexity

Combining both stages, the total time complexity of the proposed algorithm is approximately:

$$O(K \cdot p^2 \cdot (\bar{r}^2 + n_{avg}^2)) \quad (31)$$

Below is a horizontal comparison of the time complexity of the proposed algorithm with other comparison algorithms. The input image size for all algorithms is uniformly set to $M \times M$, and the asymptotic time complexities of each algorithm are re-derived. Table 14 summarizes the theoretical complexities of the proposed algorithm and the six comparison algorithms, all using a unified parameter system.

For proposed, IPI, and GDR-PCD: the sliding window size is $p \times p$, the step size is s , and the number of image patches is $K \approx (\frac{M-p}{s} + 1)^2 \approx \frac{M^2}{s^2}$. The proposed algorithm additionally needs to solve the WNNM model for R groups ($R = K$) separately, with an average of n_{avg} similar patches per group and an average search radius of $\bar{r}$.

For PSTNN, STWRT, and ASTTV-NTLA: the $M \times M$ image is divided into $n_1 \times n_2$ patches, and n_3 patches are stacked to form a third-order tensor (typically $n_1 n_2 \ll M^2$, and n_3 the number of stacked frames or the transform-domain dimension).

For 4DST-BTMD: $n_1 \times n_2$ are the spatial dimensions, n_3 and n_4 are the spatio-temporal dimensions, and the block-term decomposition rank is R' .

The number of iterations for model solving is set to t for all the above algorithms.

Table 14. Time complexity comparison of various algorithms.

Algorithm	Time Complexity
Proposed	$O(\frac{M^2}{s^2} \cdot p^2 \cdot (\bar{r}^2 + n_{avg}^2) \cdot t)$
4DST-BTMD	$O(t \cdot n_1 n_2 n_3 n_4 R')$
ASTTV-NTLA	$O(t \cdot n_1 n_2 \cdot \min(n_1, n_2) n_3)$
PSTNN	$O(t \cdot n_1 n_2 \cdot \min(n_1, n_2) n_3)$
GDR-PCD	$O(t \cdot p^2 \cdot \frac{M^2}{s^2} \cdot \min(p^2, \frac{M^2}{s^2}))$
STWRT	$O(t \cdot n_1 n_2 \cdot \min(n_1, n_2) n_3)$
IPI	$O(t \cdot p^2 \cdot \frac{M^2}{s^2} \cdot \min(p^2, \frac{M^2}{s^2}))$

Since the proposed algorithm's solution model sets $\text{max_iter} = 1$, the iteration count $t = 1$ is directly used. However, for fairness in comparison, the iteration count t is included in the complexity expression.

5. Algorithm Limitations and Discussion

Although the proposed infrared dim and small target detection algorithm based on low-rank clustering improved WNNM demonstrates good detection performance across

multiple complex scenes, it still has some limitations, which are reflected in specific experimental scenes and metric comparisons. The two main drawbacks of this paper are the relatively low detection rate for scenes with weak target energy and the long algorithm runtime.

5.1. In-Depth Analysis of Low Detection Rate

As shown in Table 10, the proposed algorithm's $AUC_{(P_d, \tau)}$ in Scene 1 is 0.754687, while that in Scene 5 is 0.909375, both lower than comparison algorithms such as PSTNN and 4DST-BTMD. This section provides an in-depth analysis of the causes.

5.1.1. Failure Mechanism of Similar Patch Matching in Complex Background and Weak-Target Scenarios

The proposed algorithm employs a weighted fusion strategy of grayscale variance and Euclidean distance to measure similarity. The core assumption is that patches with similar texture features have similar grayscale variances, and patches containing targets can be distinguished in the variance dimension. However, this assumption fails when weak targets coexist with complex textures. Let the grayscale variance of a patch be $\sigma^2 = \frac{1}{N} \sum_{(x,y) \in \Omega} (I(x,y) - \mu)^2$, where Ω is the set of pixels in the patch, N is the total number of pixels, and μ is the grayscale mean. For a patch of size $p \times p$ ($p = 6-9$), if the target occupies $3 \times 3 - 5 \times 5$ pixels, the proportion of target pixels to total patch pixels ρ is approximately 25–69% (when $p = 6$) or 11–31% (when $p = 9$). When target radiation is weak, the grayscale difference $\Delta = \mu_{target} - \mu_{background}$ is small. The target's contribution to patch variance $\Delta\sigma^2 = \rho(1 - \rho)\Delta^2$ is masked by background variance fluctuations. In other words, the distinguishability $d = |\sigma^2_{with\ target} - \sigma^2_{without\ target}|$ approaches 0 under weak-target conditions. Complex textures inherently have high-variance fluctuations, causing overlap between target-containing and pure background patches. Since the algorithm relies only on grayscale variance and Euclidean distance, without introducing structural features (e.g., gradient histograms, local binary patterns, or structure tensor eigenvalues), target patches may be incorrectly assigned to pure background groups. This triggers a cascading failure: (1) The low-rank matrix B_j is composed entirely of pure background patches, (2) the target signal is suppressed as an anomaly, (3) the sparse variable T_j contains no target information, and (4) the target energy is lost during aggregation, leading to missed detections.

5.1.2. Limitations of the Energy Shrinkage Mechanism Under Incorrect Grouping

The energy shrinkage mechanism boosts sparse energy when it falls below a times the original energy (with $a = 0.20, b = 1.3$). Under correct grouping, it effectively enhances target contrast. However, under incorrect grouping, its effect becomes complex. Suppose a pure background group is incorrectly assigned a target-containing patch. Since the target signal is weak, the sparse energy proportion $\frac{E_{Tk}}{E_{orig}}$ may still trigger the condition (<0.20). The mechanism does not distinguish whether the group truly contains a target—it uniformly boosts energy. This may amplify noise rather than the target signal. More fundamentally, the effectiveness of the energy shrinkage mechanism depends on whether the energy threshold can distinguish target groups from noise groups. In weak-target scenarios, the sparse energy distributions of target and noise groups overlap. The single fixed threshold $a = 0.20$ cannot achieve optimal discrimination.

5.2. Analysis of Long Algorithm Runtime

Another drawback is the long runtime. Table 15 shows the runtime for processing one frame from each scene for the proposed algorithm and the comparison algorithms.

The table shows that the proposed algorithm's runtime is significantly higher than other algorithms. The primary reason is the need to match multiple groups of similar patches and solve multiple constrained models.

Table 15. Runtime of various algorithms across 6 scenes (unit: seconds).

Scene	Proposed	4DST-BTMD	ASTTV-NTLA	PSTNN	GDR-PCD	STWRT	IPI
Scene 1	80.6492	0.0594	0.0671	1.5389	96.2656	0.0209	4.1561
Scene 2	217.8245	0.0601	0.0437	1.6040	106.5998	0.0276	4.5314
Scene 3	137.2814	0.0578	0.0472	1.2322	97.7692	0.0187	4.4071
Scene 4	245.4941	0.0568	0.0517	1.2425	97.1730	0.0196	3.9631
Scene 5	166.6625	0.0566	0.0405	1.6768	91.5446	0.0198	4.5305
Scene 6	217.4874	0.0568	0.0383	1.7137	98.9944	0.0207	4.6852

5.3. Future Outlook

The above analysis indicates that the core limitations of the proposed algorithm stem from three aspects: (1) insufficient separability between weak targets and complex textures in the similar patch matching stage; (2) “blind” triggering of the energy shrinkage mechanism under incorrect grouping conditions; (3) long runtime. To address these three issues, future research can pursue systematic improvements from the following directions.

5.3.1. Multi-Feature Fusion Similar Patch Matching Mechanism

The current algorithm relies only on grayscale variance and Euclidean distance in the similar patch matching stage, failing to fully utilize the structural information of patch images. Future research can be improved from the following two levels.

(1) Structural feature supplementation: Introduce descriptors more sensitive to local structures, such as image gradient orientation histograms or local binary patterns, into the similarity measure. These descriptors can capture the spatial arrangement information of pixels within patches, effectively distinguishing texture blocks and target blocks with the same grayscale variance. Specifically, let the gradient orientation histograms of two image patches be h_i and h_j . Then, the structural similarity can be defined as:

$$S_{struct}(i, j) = \cos(h_i, h_j) = \frac{h_i \cdot h_j}{\|h_i\| \cdot \|h_j\|} \quad (32)$$

This is combined with the grayscale variance difference $\Delta\sigma^2$ and Euclidean distance d_{euclid} through joint weighting to form a multi-dimensional similarity criterion, reducing the probability of misgrouping in weak-target scenarios caused by reliance on a single statistical feature.

(2) Multi-scale variance estimation: To address the inherent problem of single-scale variance being sensitive to local noise, a multi-scale analysis strategy can be introduced. Grayscale variances are computed at windows of different sizes (e.g., 3×3 , 5×5 , 7×7), and their weighted average is taken as the texture complexity measure of the patch. Multi-scale averaging can effectively smooth variance fluctuations caused by single pixels or tiny local regions, making the variance dimension difference between target-containing patches and pure-texture patches more stable and distinguishable.

5.3.2. Target Confidence-Aware Energy Shrinkage Strategy

The current energy shrinkage mechanism uses a fixed threshold $a = 0.20$ for binary judgment of the trigger condition, and performs uniform energy boosting on the sparse variable after triggering. This unified processing strategy is effective when grouping

is reliable, but may amplify noise when grouping is incorrect. Future research can be improved from the following two directions:

(1) Confidence-weighted adaptive threshold: Introduce a target confidence score $q_j \in [0, 1]$ for each group ($j = 1, 2, \dots, R$), which can be computed based on the structural similarity of patches in that group to a typical target template or local contrast. The trigger threshold a_j and boost multiplier b_j of the energy shrinkage should be dynamically adjusted according to q_j : high-confidence groups use larger a and b to ensure sufficient preservation of target energy; low-confidence groups use smaller a and b to suppress noise amplification. This can be designed as a linear mapping of the form $a_j = a_{\min} + q_j(a_{\max} - a_{\min})$, where $a_{\min}$ and $a_{\max}$ are the minimum and maximum values of a_j .

(2) Asymmetric energy enhancement strategy: Consider directional enhancement of components in the sparse variable that exhibit target morphological characteristics, rather than uniformly scaling all non-zero elements. Connected component analysis can be performed on the sparse variable T to identify connected components with compact spatial distribution (i.e., typical target morphology), and only these components are subjected to energy boosting, while scattered noise points maintain their original shrinkage intensity. This strategy can amplify target energy while suppressing noise energy, improving signal–noise separation.

5.3.3. Computational Efficiency Optimization Strategies

In addition to detection accuracy, computational efficiency is also a direction that urgently needs improvement for the proposed algorithm (as shown in Table 15). Future research can pursue acceleration from the following two levels:

(1) Fast approximate Singular Value Decomposition: Since the core time consumption of WNNM solution for each group is the SVD, random SVD algorithms can be used to approximately replace traditional SVD. Random SVD can reduce the computational complexity from $O(p^2 n_j)$ to $O(p^2 n_j \log n_j)$ while maintaining sufficient accuracy. For groups with large n_j , this acceleration strategy will be particularly effective.

(2) Parallel computing: The WNNM solutions for each group are independent of each other and possess natural parallelism. In the future, inter-group parallel computation can be implemented based on GPU or multi-core CPUs, reducing the total runtime from the sum of each group's solution time to the maximum solution time of a single group. Considering that the number of groups R may reach hundreds to thousands, the theoretical upper limit of parallel acceleration is very considerable.

The above content presents specific strategies for addressing the problems existing in the proposed algorithm. Future work can incorporate these strategies to achieve better results.

6. Conclusions

Aiming at the challenges of difficult background clutter suppression, easy loss of target energy, and high false alarm rates in infrared dim and small target detection under complex backgrounds, this paper proposes a method based on low-rank clustering improved Weighted Nuclear Norm Minimization (WNNM). Built upon the traditional WNNM framework, the method introduces a low-rank clustering mechanism. By fusing local statistical features (grayscale variance) of patches with spatial Euclidean distance, it achieves adaptive grouping of image patches, ensuring higher low-rankness within each group. Based on this, differentiated WNNM penalization strategies are designed for groups with different characteristics. Furthermore, an energy shrinkage mechanism is proposed to effectively alleviate the problem of target energy dilution during aggregation in tradi-

tional low-rank clustering methods. The main contributions of this paper are summarized as follows:

(1) An adaptive similar patch matching strategy is proposed, dynamically adjusting the search window radius and the number of similar patches based on the local variance of the reference patch. For flat regions, small windows and fewer patches are used to avoid redundant computation and noise introduction. For texture-complex regions, larger windows and more patches ensure the low-rank decomposition captures sufficient structural information. This strategy improves background suppression accuracy while reducing computational burden.

(2) An energy shrinkage mechanism is introduced. During the solution process for each group's WNNM, the energy proportion of the sparse part (target image) is monitored. When it falls below 20% of the original energy, it is forcibly boosted to 30%, with corresponding adjustments to the low-rank part to maintain consistency. This mechanism significantly improves the problem of target energy attenuation during multiple groupings and aggregations, enhancing the contrast of dim and small targets in the output image.

(3) Based on the low-rank clustering concept, the WNNM model is extended to a group-wise differentiated weighting form. Each group independently sets the constant factor for its WNNM weights based on its average variance, achieving adaptive constraint strength for different texture regions and further improving background–target separation capability.

Experiments were conducted on a public dataset, selecting 6 scenes with different background complexities and target motion characteristics. Comparisons were made with 6 mainstream detection methods (including single-frame and multi-frame algorithms) and two deep learning methods. Quantitative evaluation employed BSF, SNR, and three AUC metrics ($AUC_{(P_d, \tau)}$, $AUC_{(P_f, \tau)}$, $AUC_{(P_d, P_f)}$). Experimental results show the following: In terms of BSF, the proposed algorithm achieves optimal values in 4 scenes and ranks among the top in others. In terms of SNR, it achieves optimal values in 3 scenes and is close to the optimal value in others. In terms of $AUC_{(P_f, \tau)}$ (background suppression capability), the proposed algorithm achieves optimal values in all 6 scenes, with false alarm control significantly superior to comparison algorithms. In terms of overall detection performance $AUC_{(P_d, P_f)}$, the proposed algorithm achieves optimal or near-optimal values in multiple scenes. Ablation studies further validate the individual contributions of WNNM and the optimization model (adaptive matching + energy shrinkage), as well as the synergistic gains from their combination.

Although the proposed algorithm achieves superior performance in background suppression and target detection, it still has some limitations: First, the running time is long. Because similar patch search and group optimization solutions are performed independently for each reference patch, the computational load is significantly higher than traditional single-model methods (e.g., IPI), requiring further optimization for practical engineering applications. Second, in some scenes with dense textures or periodic noise (e.g., Scene 1), the algorithm exhibits over-suppression of corner structures in the background, leading to reduced contrast in the output target image. Future research could consider introducing multi-scale analysis or more robust structure-preserving priors to enhance adaptability in extremely complex scenes. Additionally, exploring fast approximate SVD or parallel computing strategies could reduce computational time and improve real-time performance. The computational time of the proposed algorithm is significantly higher than that of single-model methods like IPI (as shown in Table 15), mainly due to two factors: (1) adaptive similar patch search and grouping for each reference patch; (2) independent WNNM optimization solutions (including multiple SVDs) for R groups. Therefore, the proposed method is currently suitable for offline or batch processing scenarios where

detection accuracy, background suppression capability, and false alarm control take precedence over real-time requirements. Typical applications include post-mission intelligence analysis, surveillance data review, and offline processing of scientific experimental data. For real-time or near-real-time deployment needs, further acceleration is required through fast SVD approximations, parallel computing, or group sparsification strategies. This will be a key focus of future work.

In summary, the proposed low-rank clustering improved WNNM method provides a new and effective technical approach for infrared dim and small target detection in complex backgrounds, demonstrating significant advantages in background suppression, false alarm control, and target energy preservation, with good application prospects and research value.

Author Contributions: Conceptualization, Y.D. and X.F.; methodology, Y.D.; software, Y.D.; validation, Y.D. and Q.M., X.F. and J.L.; formal analysis, Y.D.; investigation, Y.D. and Q.M.; resources, X.F.; data curation, Y.D.; writing—original draft preparation, Y.D.; writing—review and editing, X.F.; visualization, Y.D.; supervision, X.F. and J.L.; project administration, X.F.; funding acquisition, X.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Guangxi Outstanding Youth Science Foundation Project (No. 2025GXNSFFA069014), National Natural Science Foundation of China under Grant (Grant No. 62261004).

Data Availability Statement: The data presented in this study, as well as the detailed derivation of the optimization algorithm and additional experimental results, are available upon request from the corresponding author.

Acknowledgments: The authors would like to thank Hui et al. at the National University of Defense Technology for providing the infrared image dataset.

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

Abbreviations

The following abbreviations are used in this manuscript:

WNNM	Weighted Nuclear Norm Minimization
IPI	Infrared Patch Image
SNR	Signal-to-Noise Ratio
BSF	Background Suppression Factor
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
SVD	Singular Value Decomposition
ADMM	Alternating Direction Method of Multipliers
RPCA	Robust Principal Component Analysis

References

1. Hou, F.; Zhang, Y.; Zhou, Y.; Zhang, M.; Lv, B.; Wu, J. Review on Infrared Imaging Technology. *Sustainability* **2022**, *14*, 11161. [\[CrossRef\]](#)
2. Yan, W.; Hongbo, H.; Zhenming, P. Low-Rank Tensor Subspace Learning for Infrared Small Target Detection. *Data Acquis. Process.* **2025**, *40*, 349–364.
3. Pi, Y.; Kong, L.; Yang, B.; Chang, R.; Pu, H.; Zhou, M.; Luo, J. Irtransunet: Efficient transformer embedding unet for infrared small target detection. *Infrared Phys. Technol.* **2025**, *151*, 106061. [\[CrossRef\]](#)
4. Zhou, F.; Fu, M.; Qian, Y.; Yang, J.; Dai, Y. Sparse prior is not all you need: When differential directionality meets saliency coherence for infrared small target detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5039818. [\[CrossRef\]](#)
5. Li, S.; Ye, M.; Ji, L.; Tang, S.; Gan, Y.; Zhu, X. Illumination distribution-aware thermal pedestrian detection. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 18688–18700. [\[CrossRef\]](#)

6. Ahmeda, S.S.; Keche, M.; Haarison, I. Adaptive for tracking multiple targets in cluttered joint probabilistic environment. *Radar Sonar Navig.* **2014**, *144*, 309–314.
7. Yao, R.; Li, W.; Zhou, Y.; Sun, J.; Yin, Z.; Zhao, J. Dual-stream edge-target learning network for infrared small target detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5007314. [[CrossRef](#)]
8. Gao, C.; Meng, D.; Yang, Y.; Wang, Y.; Zhou, X.; Hauptmann, A.G. Infrared Patch-Image Model for Small Target Detection in a Single Image. *IEEE Trans. Image Process.* **2013**, *22*, 4996–5009. [[CrossRef](#)] [[PubMed](#)]
9. Wang, Q.; Wu, L.; Wang, Y.; Wang, H.; Yang, W. Infrared Small Target Detection Based on Keypoint Detection. *Acta Aeronaut. Astronaut. Sin.* **2023**, *44*, 284–294.
10. Han, J.; Ma, Y.; Zhou, B.; Fan, F.; Liang, K.; Fang, Y. A robust infrared small target detection algorithm based on human visual system. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 2168–2172. [[CrossRef](#)]
11. Luo, K. Space-based infrared sensor scheduling with high uncertainty: Issues and challenges. *Syst. Eng.* **2014**, *18*, 102–113. [[CrossRef](#)]
12. Luo, F.; Li, Y.; Zeng, G.; Peng, P.; Wang, G.; Li, Y. Thermal infrared image colorization for nighttime driving scenes with top-down guided attention. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 15808–15823. [[CrossRef](#)]
13. Pan, S.; Zhang, S.; Zhao, M.; An, B. Maritime infrared weak target detection based on weighted scene prior. *J. Infrared Millim. Waves* **2019**, *38*, 9.
14. Dai, Y.; Wu, Y.; Song, Y.; Guo, J. Non-negative infrared patch-image model: Robust target-background separation via partial sum minimization of singular values. *Infrared Phys. Technol.* **2017**, *81*, 182–194. [[CrossRef](#)]
15. Bai, X.; Zhou, F. Analysis of new top-hat transformation and the application for infrared dim small target detection. *Pattern Recognit.* **2010**, *43*, 2145–2156. [[CrossRef](#)]
16. Zhang, J.; Cao, S.; Cui, W.; Zhang, T. Infrared small target detection based on improved top-hat transformation. *J. Electron. Inf. Technol.* **2024**, *46*, 267–276.
17. Han, J.; Moradi, S.; Faramarzi, I.; Zhang, H.; Zhao, Q.; Zhang, X.; Li, N. Infrared small target detection based on the weighted strengthened local contrast measure. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 806–810. [[CrossRef](#)]
18. Chen, C.L.P.; Li, H.; Wei, Y.; Xia, T.; Tang, Y.Y. A local contrast method for small infrared target detection. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 574–581. [[CrossRef](#)]
19. Sun, Y.; Yang, J.; An, W. Infrared dim and small target detection via multiple subspace learning and spatial-temporal patch tensor model. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 3737–3752. [[CrossRef](#)]
20. Zhang, T.; Li, M.; Zuo, Z.; Yang, W.; Sun, X. Moving dim point target detection with three-dimensional wide-to-exact search directional filtering. *Pattern Recognit. Lett.* **2007**, *28*, 246–253. [[CrossRef](#)]
21. Liu, J.; Ji, H. Detection method for small targets in the IR image based on the variable weighted pipeline filter. *J. Infrared Millim. Waves* **2007**, *34*, 743–747.
22. Pang, D.; Shan, T.; Li, W.; Ma, P.; Tao, R.; Ma, Y. Facet Derivative-Based Multidirectional Edge Awareness and Spatial-Temporal Tensor Model for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5001015. [[CrossRef](#)]
23. Liu, T.; Liu, Y.; Yang, J.; Li, B.; Wang, Y.; An, W. Graph Laplacian regularization for fast infrared small target detection. *Pattern Recognit.* **2025**, *158*, 111077. [[CrossRef](#)]
24. Liu, T.; Yang, J.; Li, B.; Zhang, L.; Peng, Z.; Xiao, X. Infrared small target detection via nonconvex tensor tucker decomposition with factor prior. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–16. [[CrossRef](#)]
25. Ran, Q.; Wang, Q.; Zhao, B.; Wu, Y.; Pu, S.; Li, Z. Lightweight oriented object detection using multiscale context and enhanced channel attention in remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5786–5795. [[CrossRef](#)]
26. Li, C.; Zhou, A.; Yao, A. Omni-dimensional dynamic convolution. *arXiv* **2022**, arXiv:2209.07947.
27. Ma, Y.; Huang, C.; Huang, J.; Wang, W. Research of infrared dim and small target detection based on SSD framework. *Laser Infrared* **2021**, *51*, 1342–1347.
28. Chen, Y.; Dai, X.; Liu, M.; Chen, D.; Yuan, L.; Liu, Z. Dynamic convolution: Attention over convolution kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2020; pp. 11030–11039.
29. Kou, R.; Wang, C.; Yu, Y.; Peng, Z.; Huang, F.; Fu, Q. Infrared Small Target Tracking Algorithm via Segmentation Network and Multi-strategy Fusion. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–12. [[CrossRef](#)]
30. Pang, D.; Ma, P.; Shan, T.; Li, W.; Tao, R.; Ma, Y.; Wang, T. STTM-SFR: Spatial-Temporal Tensor Modeling with Saliency Filter Regularization for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [[CrossRef](#)]
31. Li, B.; Wang, L.; Wang, Y.; Wu, T.; Lin, Z.; Li, M.; An, W.; Guo, Y. Mixed-Precision Network Quantization for Infrared Small Target Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5000812. [[CrossRef](#)]
32. Zhang, X.; Wang, L.; Peng, Z.; Wang, C.; Liu, T. RRCANet: Recurrent Reusable-Convolution Attention Network for Infrared Small Target Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 24632–24646. [[CrossRef](#)]
33. Chunbo, C. Research on Infrared Weak and Small Target Detection Algorithm Based on Local Prior and Tensor Recovery. Ph.D. Dissertation, Huazhong University of Science and Technology, Hubei, China, 2022.

34. Candes, E.; Recht, B. Exact matrix completion via convex optimization. *Commun. ACM* **2012**, *55*, 111. [[CrossRef](#)]
35. Zhang, D.; Hu, Y.; Ye, J.; Li, X.; He, X. Matrix completion by Truncated Nuclear Norm Regularization. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2012; pp. 2192–2199.
36. Oh, T.-H.; Kim, H.; Tai, Y.-W.; Bazin, J.-C.; Kweon, I.-S. Partial Sum Minimization of Singular Values in RPCA for Low-Level Vision. In *Proceedings of the 2013 IEEE International Conference on Computer Vision*; IEEE: New York, NY, USA, 2013; pp. 145–152.
37. Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Weighted Nuclear Norm Minimization with Application to Image Denoising. In *Proceedings of the Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2014; pp. 2862–2869.
38. Yang, S. Research on Infrared Weak Target Detection Technology in Complex Background. Ph.D. Dissertation, National University of Defense Technology, Changsha, China, 2020.
39. Hui, B.; Song, Z.; Fan, H.; Zhong, P.; Hu, W.; Zhang, X.; Ling, J.; Su, H.; Jin, W.; Zhang, Y.; et al. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background. *China Sci. Data* **2020**, *5*, 2096–2223. [[CrossRef](#)]
40. Luo, Y.; Li, X.; Chen, S.; Xia, C. 4DST-BTMD: An infrared small target detection method based on 4-D data-sphered space. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–20. [[CrossRef](#)]
41. Liu, T.; Yang, J.; Li, B.; Xiao, C.; Sun, Y.; Wang, Y.; An, W. Nonconvex Tensor Low-Rank Approximation for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–18. [[CrossRef](#)]
42. Zhang, L.; Peng, Z. Infrared Small Target Detection Based on Partial Sum of the Tensor Nuclear Norm. *Remote. Sens.* **2019**, *11*, 382. [[CrossRef](#)]
43. Wu, A.; Fan, X.; Chen, H.; Min, L.; Xu, Z. Infrared Dim and Small Target Detection Algorithm Combining Multiway Gradient Regularized Principal Component Decomposition Model. *IEEE Access* **2022**, *10*, 36057–36072. [[CrossRef](#)]
44. YYin, J.-J.; Li, H.-C.; Zheng, Y.-B.; Gao, G.; Hu, Y.; Tao, R. Spatial-Temporal Weighted and Regularized Tensor Model for Infrared Dim and Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–17. [[CrossRef](#)]
45. Wu, F.; Yu, H.; Liu, A.; Luo, J.; Peng, Z. Infrared Small Target Detection Using Spatio-Temporal 4D Tensor Train and Ring Unfolding. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5002922. [[CrossRef](#)]
46. Chang, C.-I. An Effective Evaluation Tool for Hyperspectral Target Detection: 3D Receiver Operating Characteristic Curve Analysis. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5131–5153. [[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.