Article

Robust Infrared Small Target Detection via Jointly Sparse Constraint of $l_{1/2}$-Metric and Dual-Graph Regularization

Fei Zhou, Yiquan Wu *, Yimian Dai and Kang Ni

College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; F.zhouip@nuaa.edu.cn (F.Z.); dym@nuaa.edu.cn (Y.D.); baiyaruibigbang@nuaa.edu.cn (K.N.)

* Correspondence: imagestrong@nuaa.edu.cn; Tel.: +86-137-7666-7415

check for updates

Abstract: Small target detection is a critical step in remotely infrared searching and guiding applications. However, previously proposed algorithms would exhibit performance deterioration in the presence of complex background. It is attributed to two main reasons. First, some common background interferences are difficult to eliminate effectively by using conventional sparse measure. Second, most methods only exploit the spatial information typically, but ignore the structural priors across feature space. To address these issues, this paper gives a novel model combining the spatial-feature graph regularization and $l_{1/2}$-norm sparse constraint. In this model, the spatial and feature regularizations are imposed on the sparse component in the form of graph Laplacians, where the sparse component is enforced as the eigenvectors of their graph Laplacian matrices. Such an approach is to explore the geometric information in both data and feature space simultaneously. Moreover, $l_{1/2}$-norm acts as a substitute of the traditional l_1-norm to constrain the sparse component, further reducing the fake targets. Finally, an efficient optimization algorithm equipped with linearized alternating direction method with adaptive penalty (LADMAP) is carefully designed for model solution. Comprehensive experiments on different infrared scenes substantiate the superiority of the proposed method beyond 11 competitive algorithms in subjective and objective evaluation.

Keywords: infrared small target detection; spatial and feature graph regularization; $l_{1/2}$-norm constraint; LADMAP

1. Introduction

Small target detection is a pivotal technique in infrared search and tracking applications, such as precise guidance and antimissile systems, maritime target search equipment, and small unmanned aerial vehicle surveillance systems [1–3]. The main purpose of small target detection is to search and locate potentially suspicious targets in the distance as early as possible, which facilitates people to take adequate preparations for emergencies. In long-range infrared scenes, projected targets only possess one or few pixels in size let alone other concrete discriminating features, i.e., texture, edge, contour, and might even be buried in diverse interferences or heavy noise. Additionally, their visibility varies greatly depending on target type and background environment. Although great advances have been made for detecting small target in recent decades, it remains a formidable task due to the above challenges.

Generally, relying on whether the prior information like target velocity and trajectory is applied or not, existing detection algorithms can be roughly viewed as sequential detection or single-frame detection. Traditional sequential detecting schemes present acceptable performance under given conditions [4–6]. However, they degrade performance possibly in the absence of the priors. Small target

detection in single frame, in comparison, has been attracting wide attention in view of the advantages of easy implementation and few prior requirements. Current single-frame detection models are mostly designed from background characteristics, target features or both of them in spatial domain. For instance, methods based on background characteristics are driven by the assumption of local spatial consistency, and essentially dominated by background prediction results [7–10]. They perform well for uniform scenes, but are fragile to clutter and random noise in jumbled background. It mainly lies in that the manner of background prediction based on spatial statistical distribution over pixel appearance is incompetent to encode the mutated and irregular components, e.g., clutter, glints. On the other hand, methods based on target regional saliency depict the difference between a target and its local area as a contrast enhancement factor to protrude the target while suppressing background [11–14]. However, the sensitivity to strong edge and sporadic glint limits the robustness and scalability of these algorithms for practical applications. It is because the saliency of a dim yet real target would be covered by high brightness interferences, or be easily mixed by the saliency of random glitters.

Different from the methods derived individually on background or target characteristics, there occurs a novel trend that is to integrate both of them to overcome the limitations of single characteristics for improving detection performance [15–18]. Especially, a series of representative approaches are based on the theory of low-rank recovery through using both background nonlocal correlation and target sparsity [19]. For example, Gao et al. [18] initially proposed the infrared patch-image (IPI) model in the fashion of patch vectorization, converting small target detection into a low-rank and sparse matrices separation problem. Wang et al. [20] presented a multi-subspace learning algorithm based on low-rank representation to overcome the drawbacks caused by the hypothesis of single subspace. Dai et al. [21] generalized the patch-image model to an infrared patch-tensor (IPT) model to mine more information from the nonlocal correlation in patch space. The low-rank recovery-based methods present promising performance in slowly transitional and homogeneous background, and show great potential for improving detection robustness in complex scenes. They, however, still suffer from two deficiencies. First, the existing methods based on low-rank recovery only consider the prior knowledge within background patch space, namely nonlocal correlation. It essentially assumes that all patches can be expressed by a single subspace or a mixture of low-rank subspace clusters. However, some rare components, such as sun flash, cloudy edges, and heavy sea waves, deviate from the low-rank subspace and show analogous sparsity to small targets under insufficient sample patches. Therefore, when the feature implied in patch space is only exploited, the performance of the methods will be greatly limited. Second, l_1-norm as a sparse measure may cause a dilemma where either a dim small target may be missed due to the excessive shrinking or the sparse glitter points remain in the target image due to inadequate constraints. Some recent ameliorations [22–24] indirectly attribute the unsatisfactory phenomenon to the constant trade-off parameter, and apply a weighted penalty to adjust the parameter invariance. However, the intrinsic reason is that l_1-norm as a loose approximation of l_0-norm commonly introduces extra bias into the sparse constraint when using minimum measurement [25].

In common, an observed matrix owns two spaces, namely column space and row space, separately corresponding to patch space and feature space in our paper. The two spaces coincide in the matter of the rank of a matrix. However, our key observation is that when estimating the rank of a corrupted patch-image, the rank in one space, say in patch space, may not be as precise as that of another. It is because the rank of the corrupted image will be relaxed in one space more than another one. Therefore, another space becomes particularly useful in constraining the rank mutually, as illustrated in Figure 1. Moreover, some researchers [26] have identified that the high dimensional data generally reside on a nonlinear low dimensional manifold in data space, also in feature space. The manifold in feature space is viewed as the feature manifold. Some published studies [27–30] have reported some models that integrate all two spaces in the form of dual-graph regularization. Nevertheless, it is worth noting that the dual-graph regularized pattern has not been designed for detecting infrared small targets. Moreover, $l_{1/2}$ regularization can provide a representative sparse solution among all l_q regularization

with q in (0,1), which has been verified by a phase diagram study, and its fast solution has also been designed ingeniously in [25]. Therefore, using $l_{1/2}$ regularization can better wipe out fake targets, distinguishing a real small target precisely. Additionally, the sparse negative components obtained in an iterative process is not related to a small target in actual physical sense. Given the above analysis, we propose a novel model based on $l_{1/2}$-norm combining dual-graph regularization for infrared small target detection, in which the two graphs are constructed from the patch and feature space via k-nearest neighboring manner.
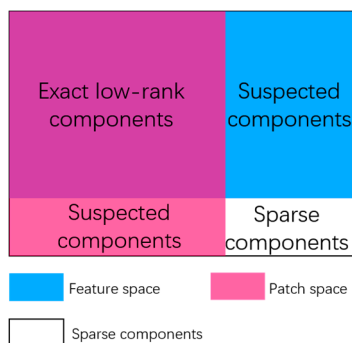

Figure 1. Explanation of our motivation to integrate patch space with feature space.

The main contributions of this paper are summarized in the following:

(1) We propose a novel model based on dual-graph regularization for infrared small target detection, which simultaneously incorporates both the data and feature manifold in the form of graph Laplacian.

(2) To eliminate fake targets effectively, $l_{1/2}$-norm instead of l_1-norm in traditional methods is used to constrain the sparse part. Additionally, a non-negative constraint in sparse component is appended to cater to the fact that targets have higher intensity.

(3) To accelerate the efficiency of the proposed algorithm, we skillfully design an efficient optimization algorithm based on the linearized alternating direction method with adaptive penalty (LADMAP) [31], which uses fewer auxiliary variables with convergence guarantee. Extensive experiments on various scenes demonstrate the superiority of the proposed model compared with 11 competitive baselines.

The remainder of this paper is organized as follows. We have a brief review about graph representation for data as well as about the methods related to infrared small target detection in Section 2. The proposed dual-graph regularized method is given detailed in Section 3. We propose a simple and feasible optimization algorithm to solve the proposed model in Section 4. The performance of the proposed method is evaluated by extensive experiments in Section 5. Finally, our discussion and conclusions are presented in Sections 6 and 7, respectively.

2. Preliminaries and Related Algorithms

2.1. Graph Laplacian

Suppose that $X \in \mathbb{R}^{d \times n}$ resides on a potential manifold $\mathcal{M}$, an undirected weighted graph $G(X, E, W)$ with n vertices can be constructed via k-nearest neighboring manner, as shown in Figure 2, where $E = \{e_{ij}\}$ is the edge set with each edge e_{ij} connecting vertices x_i and x_j. $W = \{w_{ij}\}$ denotes edge weight set measuring the correlation between vertices, which can be calculated by binary method, heat kernel, or correlation distance. If vertices x_i and x_j are not in a cluster set, then $w_{ij}(x_i, x_j) = 0$. The unnormalized graph Laplacian matrix $L \in \mathbb{R}^{n \times n}$ corresponding to graph G is presented as

$L = H - W$, where H is the degree matrix with each entry $H_{ii} = \sum_j W_{ij}$. It can be expressed mathematically as follows:

$$L_{ij}\left(x_i, x_j\right) = \begin{cases} -w_{ij}(x_i, x_j), & if \ i \neq j \\ \sum_{j, j \neq i} w_{ij}(x_i, x_j), & otherwise \end{cases} \tag{1}$$
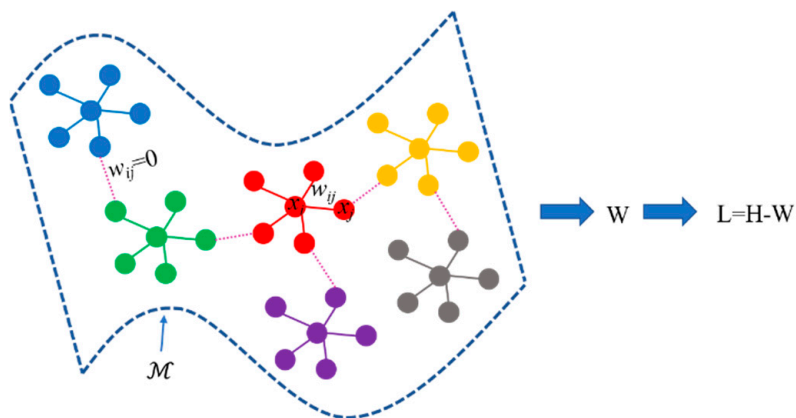

Figure 2. Graph construction.

2.2. Related Algorithms

Significant advances in single-frame infrared small target detection have been made in recent decades. Existing methods can be roughly classified into background prediction-based, local saliency-based, transform domain-based, dictionary learning, multiple features integration, deep learning, and subspace learning.

Background prediction-based methods model the appearance of each pixel by numerical statistics, such as a filtering manner [8,9,32] or using a parametric probability density function, such as Gaussian mixture function [7,33], and non-parametric algorithms, such as kernel density estimation functions [10]. These methods can locate small targets precisely when the background presents uniformity visually but may fail to deal with abrupt structures in heterogeneous scenes. Local saliency-based methods delineate the difference based on the local region around suspicious targets to enhance the target and suppress background, such as local contrast measure (LCM) [11], novel local contrast method (NLCM) [34], multiscale patch-based measure (MPCM) [12], derivative entropy-based contrast measure (DECM) [35], weighted local difference measure (WLDM) [13], dual-window local contrast method [14]. Such types of methods successfully enhance the dim small target and neglect the smooth areas of background, increasing the detection rate. However, they are less robust to sun glints and high-contrast edges in intricate sea backgrounds, causing high false alarm rates. Transform domain-based models explore more useful features in different domains, such as Fourier domain [1], fuzzy space [36], gradient vector field [37], to discriminate real target components from such target-like ones. The kind of methods are computational friendly and can complete the task of target extraction in a clean scene with high contrast. However, they are incapable of making a distinction between a dim small target and heavy natural noise. Dictionary learning methods such as those proposed by Li et al. [38] and Wang et al. [39], recognize the real target from several candidates derived from a given dictionary. These methods can deal with different types of small targets well, yet rely heavily on the quality of the given target dictionary. They would obtain unsatisfactory performance when actual targets cannot be represented by a dictionary atom or their combination. Methods based on multiple features integration overcome the drawbacks of the simple feature depicted by raw pixels. For example, Qin et al. [15] and Yao et al. [16] gradually remove clutter interferences and highlight target by combining background consistency and target singularity. Additionally, a multiscale adaptive difference and variance difference are jointly

used to enhance small targets and alleviate the impact of background fluctuation [40]. Methods of this type effectively eliminate the edge clutters and pixelwise noise with high brightness in heterogeneous scenes. However, they become less effective when encountering dim small target scenes, resulting in missing targets with high probability. Recently, deep convolutional neural networks have been employed for the community of small target detection [41–44]. Lin et al. [42] designed a seven-layer conventional neural network in an end-to-end way to automatically extract small target features and eliminate clutter. With the help of massive training samples generation, Zhao et al. [43] suggested a simple conventional network for modeling the background patches. Such methods show good robustness even in some complex situations with heavy clutter but they require a great quantity of labeled training data, which may not always be available in practice. In contrast, the proposed method is unsupervised and does not need any labeled training data.

The proposed method falls into the category of a subspace learning model, so we review the precious subspace learning methods based on robust principal component analysis (RPCA) for small target detection. In [19], RPCA is initially applied to separate the outliers in data, which also is employed for infrared small detection in [18], called infrared patch image model (IPI). Many researchers have put forward effective optimization, enhancements, extensions, and ameliorations of the original IPI model. For the model, one of the limitations is long consuming time due to the slow convergence of the optimization based on accelerated proximal gradient (APG). The alternating direction method of multiplier (ADMM) optimization scheme is used in the models proposed recently since it can get same optimal solution under faster convergence [21,22], [45,46]. To mine more nonlocal self-correlation information from patch models, some extended versions of the IPI model have been proposed, including infrared patch-tensor model (IPT) [21], spatial-temporal patch-image model (STPI) [7], spatial-temporal tensor model (STTM) [47], multi-subspace learning model (SMSL) [20]. Among them, IPT and SMSL perform well in some complex scenes, but they may obtain high false alarms in sea backgrounds with heavy waves and sun glint. The models in [7,47] take spatial-temporal information into account and increase the detection probability of dim small targets in slowly changing backgrounds. Some ameliorated versions are proposed to further improve the robustness of the initial IPI model. For instance, Dai et al. [22] proposed a weighted IPI model (WIPI), which used the target likelihood coefficient based on steering kernel instead of the constant weight. In [23,24], the nonconvex and tighter rank surrogate acts as a substitute for the original nuclear norm to achieve better background suppression. Besides, in the enhancement type of the IPI model, Wang et al. [45] used the total variation regularization (TVPCP) to depict the background feature, which aimed to obtain good target-background separation for some mild situations. In [46], we proposed a combination of nonconvex rank approximation and graph regularization (GRLA) to take full use of the intrinsic structure between patch images.

Difference to related existing subspace learning based methods: As a subspace learning method, our proposed model differs from the aforementioned ones in several aspects. (1) Our method incorporates prior information in both spatial and feature spaces of patch images simultaneously, whereas other methods [21–24] only take the priors within the patch space into account, ignoring the feature space. (2) Our method employs the sparser regularizer instead of commonly used reweighting manner [22–24] to enhance the sparsity of small targets, so as to suppress target-like outlier better. (3) Our method uses LADMAP to give the model solution, while other methods apply the traditional ADMM [45,46], which will introduce too many multipliers, leading to increasing time consumption. The detailed characteristics of the existing methods are summarized in Table 1.

Table 1. Detailed characteristics of several related subspace learning based methods.

Methods	Advantages	Disadvantages
IPI [18]	Perform well in uniform background	Over-shrink leading to missing detection or remaining residuals, time consuming
IPT [21]	Perform well in relative complex scenes, computational friendly	Losing dim target, fails to eliminate target-like point
STPI [7]	Achieve good performance for slowly changing background	Sensitive to strong edges and clutters, difficult to address non-Gaussian noise
STTM [47]	Perform well for homogeneous and slowly changing scenes	Difficult to address highly dynamic scenes, easily leaving residuals
SMSL [20]	Perform well for salient target scenes, computational friendly	Hard to suppress strong edges, easily missing weak target
WIPI [22]	Works well for high contrast scenes	Incapability to address the sparse noise, time consuming
Reference [23]	Eliminate sparse edges and noise, computational friendly	Difficult to suppress the interferences with similar appearance to targets
Reference [24]	Preserve target structure, suppress non-target residuals	Cannot completely suppress significant edge structure
TVPCP [45]	Recover homogeneous background well	Sensitive to the ground disturbance with high thermal, takes a long time
GRLA [46]	Perform better in background suppression	Weaken target energy, unable to maintain target structure

3. Algorithm Description

In this section, we describe the major steps of the proposed model in detail and give its system diagram in Figure 3. An infrared small target image is transformed into a patch image by using a sliding window as the input matrix. The patch and feature graphs are constructed along the columns and rows of the input matrix, and then, these Laplacian matrices of corresponding graphs are imposed on the sparse component to preserve sparse structures. To better suppress sparse outliers, we employ $l_{1/2}$-norm as the surrogate of l_1-norm. The novel objective function is formulated by incorporating these sparse regularizations and solved by LADMAP effectively. Finally, a simple thresholding operator is used to extract the real target from the target image reconstructed by the uniform average of estimator re-projection (UAE).
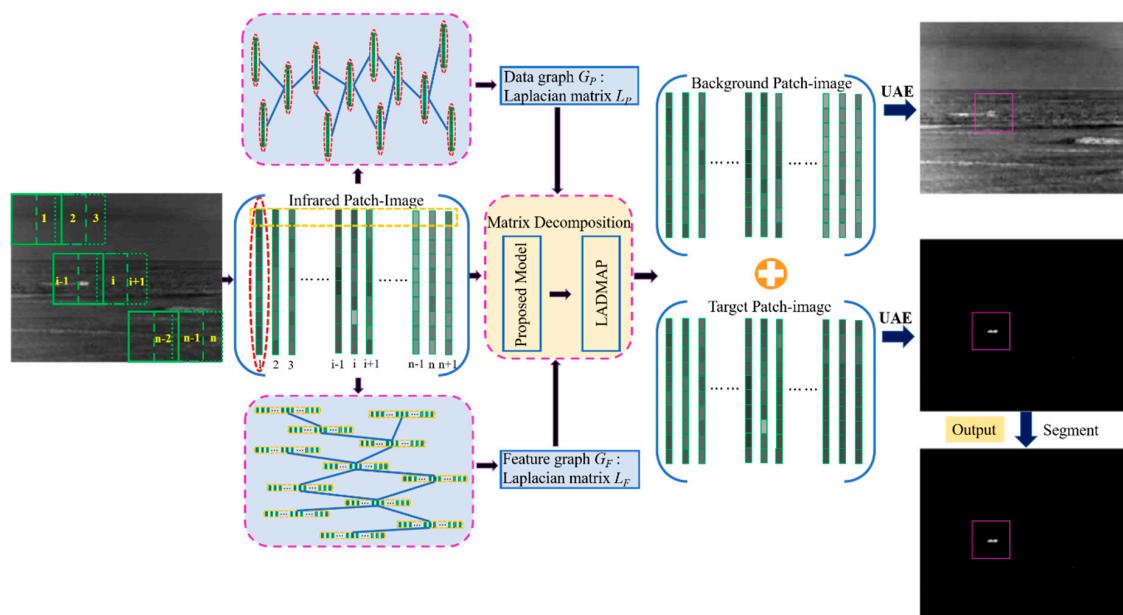
Figure 3. Illustration of the proposed model for infrared small target extraction.

3.1. Patch and Feature Graph Regularizations

Herein, we present in detail that patch and feature structural regularizations on sparse component are jointly introduced into an objective function. Constructing an infrared patch-image $D \in \mathbb{R}^{p \times n}$, let $G_P = (D, E_P, W_P)$ whose vertices $\{d_{.1}, d_{.2}, \cdots, d_{.n}\}$ are the column vectors of matrix D. W_P is the adjacency matrix, which encodes the edge weight and connectivity of the graph. The graph is constructed based on k-nearest neighboring strategy by using heat kernel, which involves searching for the closest neighbors of all the columns based on the Euclidean distance. In G_P, W_P contains the edge weight of each node connecting to its k-nearest neighbors, which can be defined as

$$W_P^{ij}(d_{.i}, d_{.j}) = \begin{cases} \exp\left(-\frac{\|d_{.i} - d_{.j}\|_2^2}{2\sigma^2}\right), & if\ d_{.i} \in N_k(d_{.j})\ or\ d_{.j} \in N_k(d_{.i}) \\ 0, & otherwise \end{cases} \tag{2}$$

where $i, j = 1, \ldots, n$. $N_k(d_{.i})$ represents the set of k-nearest neighbors of $d_{.i}$. The patch graph constraint on sparse component is designed as

$$\begin{aligned} \frac{1}{2} \sum_{i,j=1}^{k} \|t_{.i} - t_{.j}\|_2^2 w_P^{ij} &= \sum_{i=1}^{k} t_{.i}^{\top} t_{.i} h_P(i, i) - \sum_{i,j=1}^{k} t_{.i}^{\top} t_{.j} w_P^{ij} \\ &= Tr(TH_P T^{\top}) - Tr(TW_P T^{\top}) \\ &= Tr(TL_P T^{\top}) \end{aligned} \tag{3}$$

where $Tr(\bullet)$ is the trace function of matrix and $h_P(i, i)$ is a diagonal element of the degree matrix H_P. $t_{.i}$ and $t_{.j}$ are the column vectors of target patch-image T. The patch graph Laplacian matrix $L_P \in \mathbb{R}^{n \times n}$ is calculated by Equation (1).

Similarly, the feature graph $G_F = (D, E_F, W_F)$ is built by using the row vectors $d_{i.}$ and $d_{j.}$ of matrix D. W_F is formulated as

$$W_F^{ij}(d_{i.}, d_{j.}) = \begin{cases} \exp\left(-\frac{\|d_{i.} - d_{j.}\|_2^2}{2\sigma^2}\right), & if\ d_{i.} \in N_k(d_{j.})\ or\ d_{j.} \in N_k(d_{i.}) \\ 0, & otherwise \end{cases} \tag{4}$$

where $i, j = 1, \ldots, p$. Then, the feature graph constraint is denoted as

$$\begin{aligned} \frac{1}{2} \sum_{i,j=1}^{k} \|t_{i.} - t_{j.}\|_2^2 w_F^{ij} &= \sum_{i=1}^{k} t_{i.}^{\top} t_{i.} h_F(i, i) - \sum_{i,j=1}^{k} t_{i.}^{\top} t_{j.} w_F^{ij} \\ &= Tr(T^{\top} H_F T) - Tr(T^{\top} W_F T) \\ &= Tr(T^{\top} L_F T) \end{aligned} \tag{5}$$

where $t_{i.}$ and $t_{j.}$ are the row vectors of T. The information explored by the feature graph can refine the rare structure in the sparse component. The feature graph Laplacian matrix $L_F \in \mathbb{R}^{p \times p}$ is computed by W_F similar to L_P, as given by Equation (1).

3.2. $l_{1/2}$-Norm Regularization with Non-Negative Constraint

In heterogeneous scenes, corrupted components not only contain a small target but also an irregular flash point, which show similar sparsity to small targets under insufficient observation data. These components cannot be constrained equally as l_0-norm under l_1-norm measurement, causing biased suppression of sparse components. Then, some rare components will remain in the detection result and increase false alarms, directly undermining the robustness for the practical applications. Albeit employing l_q-norm ($0 < q < 1$) to constrain the sparse component can achieve a detection result with fewer false alarms, manual selection of q is an unwieldy process and reduces the adaptability of the method [48]. Fortunately, Xu et al. [25] validated the representativeness of $l_{1/2}$ regularization among

all l_q regularization $(0 < q < 1)$ by a phase diagram study. They creatively pointed that whenever q takes in [1/2, 1), the smaller the q, the sparser the solutions can be generated by l_q regularization and for q in (0,1/2), the performance of l_q regularization has no significant difference. Herein, we introduce $l_{1/2}$ regularization as a surrogate of l_1-norm to constrain the sparse part. Nevertheless, $l_{1/2}$ regularization only enhances the sparsity of target but neglects the basic physical fact that pixel values of target must be non-negative. Therefore, it is more reasonable to add a non-negative constraint on the sparse component. The non-negative sparse constraint can be defined as

$$\|T\|_{1/2,\geq 0}^{1/2} = \left(\sum_{i,j} \left(\max\left(T_{ij}, 0 \right) \right)^{1/2} \right)^2 \tag{6}$$

Finally, integrating the geometric manifold in patch and feature spaces in the form of a graph and $l_{1/2}$ regularization with a non-negative constraint into an overall framework, we propose a novel model for small target detection, which is formulated as

$$\min_{B,T} \|B\|_* + \lambda \|T\|_{1/2,\geq 0}^{1/2} + \gamma_1 Tr(TL_P T^\top) + \gamma_2 Tr(T^\top L_F T), \tag{7}$$
$$s.t \ D = B + T, T \geq 0$$

where λ, γ_1, and γ_2 are the tradeoffs to control the corresponding weight to each of the terms while optimizing the objective function.

4. LADMAP for Solving the Proposed Model

In recent years, many categories have been developed for solving low-rank optimization problems [49–51]. Especially, ADMM is frequently employed to handle the target-background separation under RPCA framework in infrared small target detection community. It can update the separable variables in convex programming by alternately minimizing, so that the optimization problem can be simplified in this way. However, observing the model in Equation (7), one can find that multiple auxiliary variables should be introduced to realize the separability of the augmented Lagrangian function. The computational complexity of the algorithm will then be correspondingly increased, because a certain amount of complexity is required to minimize each variable, and the number of iterations, as a key factor of influencing the computational efficiency, may be increased. It seriously erodes the algorithm in real time. To tackle this issue, we adopted a well-designed variant of ADMM, called the linearized ADMM with adaptive penalty (LADMAP) [31], to solve the proposed model effectively. For this purpose, the linear equality constraint in Equation (7) is removed by using the following augmented Lagragian function:

$$
\begin{aligned}
L(B, T, Y, \mu) \ &= \|B\|_* + \lambda \|T\|_{1/2,\geq 0}^{1/2} + \gamma_1 Tr(TL_P T^\top) + \gamma_2 Tr(T^\top L_F T) \\
&+ \langle Y, D - B - T \rangle + \tfrac{\mu}{2} \|D - B - T\|_F^2
\end{aligned}
\tag{8}
$$

where $Y \in \mathbb{R}^{p \times n}$ is a Lagrangian multiplier, and $\mu > 0$ assigns the penalty to the violation of the linear constraints. LADMAP is used to directly optimize the primary variables B, T, and Y by solving each variable alternatively while fixing other ones. It involves fewer auxiliary variables and converges faster than the initial ADMM [52]. The detail procedure of solving Equation (8) by LADMAP is provided in the following.

4.1. Solution of the Proposed Method

Solving B: Assuming that T and Y are fixed, the solution of B_{k+1} is provided by minimizing the following objective function:

$$
\begin{aligned}
L(B, T_k, Y_k, \mu_k) &= \|B\|_* + \langle Y_k, D - B - T_k \rangle + \frac{\mu_k}{2}\|D - B - T_k\|_F^2 \\
&= \frac{1}{\mu_k}\|B\|_* + \frac{1}{2}\|B - (D - T_k + Y_k/\mu_k)\|_F^2
\end{aligned}
\tag{9}
$$

The closed-form solution in Equation (9) can be obtained by singular value thresholding (SVT) operator, which is defined as

$$
B_{k+1} = U_k S_{1/\mu_k}(\Sigma)V_k^\top
\tag{10}
$$

where U, Σ, V are obtained using singular value decomposition (SVD) of $D - T_k + Y_k/\mu_k$. $S_\tau(x)$ is the thresholding algorithm written as $S_\tau(x) = \max(|x| - \tau)sign(x)$.

Solving T: Assuming that B and Y are fixed, the solution of T_{k+1} is provided by minimizing the following objective function:

$$
\begin{aligned}
L(B_{k+1}, T, Y_k, \mu_k) &= \lambda\|T\|_{1/2,\geq0}^{1/2} + \gamma_1 Tr(TL_P T^\top) + \gamma_2 Tr(T^\top L_F T) \\
&\quad + \langle Y_k, D - B_{k+1} - T \rangle + \frac{\mu_k}{2}\|D - B_{k+1} - T\|_F^2
\end{aligned}
\tag{11}
$$

which does not have a closed-form solution. In order to use the closed-form solution to the proximity operator of $l_{1/2}$-norm achieved by the *half*-thresholding operator [25], we take an ingenious strategy by further linearizing the smooth component of Equation (11) to simplify the subproblem. The smooth component of Equation (11) can be written as

$$
\begin{aligned}
s(B_{k+1}, T, Y_k, \mu_k) &= \gamma_1 Tr(TL_P T^\top) + \gamma_2 Tr(T^\top L_F T) \\
&\quad + \langle Y_k, D - B_{k+1} - T \rangle + \frac{\mu_k}{2}\|D - B_{k+1} - T\|_F^2
\end{aligned}
\tag{12}
$$

Then, motivated by the spirit of LADMAP, solving Equation (11) can be replaced by minimizing the following problem:

$$
\min_T \lambda\|T\|_{1/2,\geq0}^{1/2} + \langle \nabla_T s(T_k), T - T_k \rangle + \frac{\eta_1}{2}\|T - T_k\|_F^2
\tag{13}
$$

where $s(B_{k+1}, T, Y_k, \mu_k)$ can be approximated by the second-order Taylor expansion of the smooth components at T_k. $\nabla_T s(T_k)$ is the gradient of $s(B_{k+1}, T, Y_k, \mu_k)$ with respect to T. As long as $\eta_1 > 2\lambda\|T\|_2 + \mu(1 + \|Y\|_2^2)$, in which $\|T\|_2$ denotes the spectral norm of a matrix taking the largest singular value, the above replacement is valid. The subproblem (Equation (13)) is transformed into the minimization of $l_{1/2}$- regularization, expressed as

$$
\min_T \alpha\|T\|_{1/2,\geq0}^{1/2} + \|T - T_k + \nabla_T s(T_k)/\eta_1\|_F^2
\tag{14}
$$

where $\alpha = 2\lambda/\eta_1$.

According to [25], the solution of Equation (14) with a non-negative constraint can be computed with the help of *half*-thresholding operator, detailed as the following:

$$
T_{k+1} = \max\left(\mathcal{H}_{\alpha,\frac{1}{2}}(T_k - \nabla_T s(T_k)/\eta_1), 0\right)
\tag{15}
$$

where $\mathcal{H}_{\alpha,\frac{1}{2}}$ denotes the *half*-thresholding operator defined by Equations (16)–(19):

$$
\mathcal{H}_{\alpha,\frac{1}{2}}(X) = \left[h_{\alpha,\frac{1}{2}}(X_{ij})\right], X \in \mathbb{R}^{p\times n}
\tag{16}
$$

where

$$h_{\alpha,\frac{1}{2}}\left(X_{ij}\right) = \begin{cases} f_{\alpha,\frac{1}{2}}\left(X_{ij}\right), & \left|X_{ij}\right| > \frac{\sqrt[3]{54}}{4}(\alpha)^{2/3} \\ 0, & otherwise \end{cases} \tag{17}$$

with

$$f_{\alpha,\frac{1}{2}}\left(X_{ij}\right) = \frac{2}{3}X_{ij}\left(1 + \cos\left(\frac{2\pi}{3} - \frac{2}{3}\phi_\alpha\left(X_{ij}\right)\right)\right) \tag{18}$$

and

$$\phi_\alpha\left(X_{ij}\right) = \arccos\left(\frac{\alpha}{8}\left(\frac{\left|X_{ij}\right|}{3}\right)^{-3/2}\right) \tag{19}$$

Updating Y and μ:

$$Y_{k+1} = Y_k + \mu_k(D - B_{k+1} - T_{k+1}) \tag{20}$$

$$\mu_{k+1} = \min(\mu_{\max}, \rho_k\mu_k) \tag{21}$$

where $\mu_{\max}$ is a given positive constant, and

$$\rho_k = \begin{cases} \rho_0, & if \max\{\eta_1\|T_{k+1} - T_k\|, \mu_k\|B_{k+1} - B_k\|\} \le \varepsilon_1 \\ 1, & otherwise \end{cases} \tag{22}$$

Convergence Criteria:
According to KKT condition, the stopping criteria is designed as

$$\frac{\|D - B_{k+1} - T_{k+1}\|_F}{\|D\|_F} < \varepsilon_2 \ \ or \ \ \max\{\eta_1\|T_{k+1} - T_k\|, \mu_k\|B_{k+1} - B_k\|\} \le \varepsilon_1 \tag{23}$$

where ε_1 and ε_2 are the tolerance factors. Relying on the stopping criteria defined in Equation (23), the sequences (B, T, Y) yielded by the revised LADMAP converges to an optimal solution of the problem of Equation (7). The key steps of the proposed algorithm are summarized in Algorithm 1.

Algorithm 1: The revised LADMAP for Solving the Proposed Model

---
Input: Infrared small target image I, λ, γ_1, γ_2 and the number of nearest neighbors
Output: (B_k, T_k)
Initialize: Construct infrared patch-image $D \in \mathbb{R}^{p \times n}$; $B^0 = T^0 = 0$; $Y^0 = \frac{D}{\max(\|D\|_2, M\|\text{vec}(D)\|_{\inf})}$; $\mu_0 = \frac{1.25}{\|D\|_2}$; $\mu_{\max} = 10^7$;
$\rho = 1.1$; $\varepsilon_1 = 10^{-6}$; $\varepsilon_2 = 10^{-14}$; $k = 0$; Compute $L_F \in \mathbb{R}^{n \times n}$ and $L_P \in \mathbb{R}^{p \times p}$ from graph G_P and G_F.
While *not converged* **do**
　　　1: Compute B_{k+1} by Equation (10);
　　　2: Compute T_{k+1} by Equation (15);
　　　where $\nabla_{TS}(T_k) = 2(\gamma_1 T_k L_P + \gamma_2 L_F T_k) + \mu_k(B_{k+1} + T - D - Y_k/\mu_k)$;
　　　　　$\eta_1 = 2(\lambda_1\|L_P\|_2 + \lambda_2\|L_F\|_2) + \mu_k(1 + \|Y_k\|_2^2)$.
　　　3: Compute Y_{k+1} by Equation (20) and μ_{k+1} by Equation (21);
　　　4: Check convergence condition according to Equation (23);
　　　5: Update k: $k = k+1$
end

4.2. Complexity Analysis

The computational complexity of the proposed model is majorly dominated by the optimization of the revised LADMAP and the construction of the patch and feature graph. The construction of dual-graph based on nearest neighboring manner needs $O(p^2n + pn^2)$. Let k be the total number of iterations and r be the lowest rank of B. In each iteration, SVT is employed to compute the low-rank matrices in which its total complexity is $O(rn^2)$ under the usage of partial SVD. For *half*-thresholding operator, the complexity is $O(pn)$ since some productions between matrix and vector are only required. The overall computational complexity in all iterations is $O(pn^2 + p^2n + k(pn + rn^2))$. Hence, the primary computational complexity is $O(p^2n)$ in the case of $p > n$ or $O(pn^2)$ in the case of $n > p$.

5. Experimental Evaluation and Analysis

In this section, extensive experiments are conducted to test the effectiveness and robustness of the proposed model in terms of clutter suppression and sparse point elimination. Specially, we illustrate the validity of the proposed patch and feature sparse regularizations and provide sensitivity analysis of the parameters in the proposed model. After that, we also give the experimental comparisons with 11 competitive works on subjectively visual effect and objective indicators.

5.1. Datasets and Baselines

The experiments were carried out on numerous infrared images. Herein, 10 infrared small target sequences are displayed, which cover four typical scenes: deep space, sky-cloudy, sea-sky, terrain. To observe these sequences intuitively, Figure 4 exhibits representative frames randomly picked from each sequence, where the designated target areas are zoomed in for better observation for a dim small target. From the figures, it is easily found that these experimental scenes contain different interference components, which make them more complex than clearly simple background. The detailed features and scene classification of these sequences are presented in Table 2.
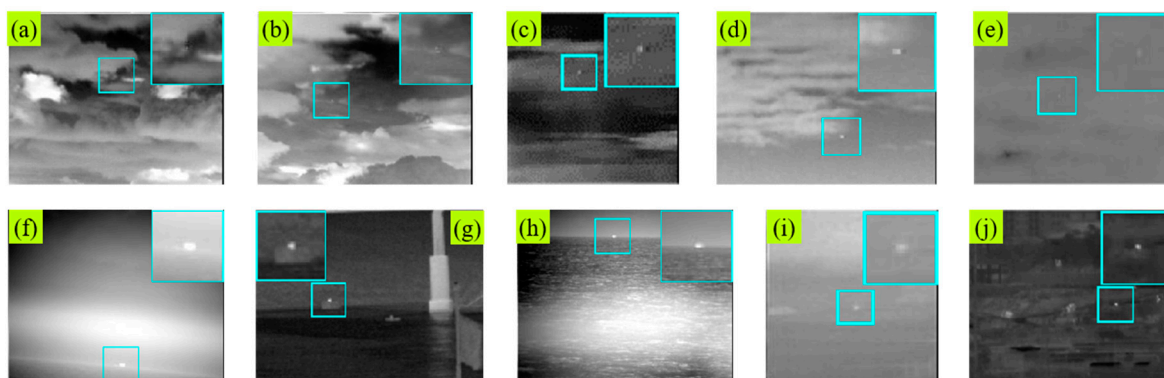

Figure 4. Exhibition of representative single image from each experimental sequence. The local area of dim target is enlarged for better observation. (**a**–**j**) represent Sequences 1 to 10.

Table 2. Detailed information of these real sequences.

Scenes	Sequences	Frames/Resolution	Target Features	Background Features
Deep-space	1, 2	$100,100/320 \times 256$, 320×256	Very small and weak with low contrast, moving along the cloud edge or buried in cloud.	Containing numerous irregular strong cloud clutter, and brightness changes greatly.
Sky-cloudy	3, 4, 5	$50,30,100/$ 128×128, 256×200, 256×200	Small with irregular shape, brightness varies greatly.	Containing substantial banded and floccus cloud and background noise. Low resolution.
Sea-sky	6, 7, 8	$100,100,200/$ 320×256, 320×256, 320×256	Target size changes greatly. Relatively high contrast. Emerging on sea-sky line.	Background with strong sea waves, bright glitters, and artificial buildings. Low signal-to-clutter.
Terrain	9, 10	$100,100/128 \times 128$, 256×220	Small square target with fuzzy contour, moving fast. Contrast changes obviously.	Background with heavy noise, plants, mountains, and manmade buildings. Low contrast.

The performance of the proposed model is compared with 11 state-of-the-art methods, comprising TDLMS [32], TopHat [9], MOG [7], MPCM [12], WLDM [13], FKRW [15], IPI [18], TVPCP [45], SMSL [20], GRLA [46], RIPT [21]. Among these methods, TDLMS, TopHat, and MOG belong to the type of

background prediction. MPCM and WLDM are the enhanced versions of LCM and achieve leading performance. FKRW is viewed as a model of multiple features integration. TVPCP, SMSL, GRLA, and RIPT are recently proposed methods based on subspace learning. For comparison, we used the original codes of MOG, IPI, TVPCP, SMSL, GRLA and RIPT provided by the authors, while the remaining methods were reimplemented according to their corresponding literature. Moreover, the parameters in these tested methods were adjusted to obtain better performance, as summarized in Table 3.

Table 3. All tested methods and their parameter settings.

No.	Methods	Parameter Settings
1	TDLMS	Support size: 5×5, step size: $\mu = 5 \times 10^{-8}$
2	TopHat	Structure shape: square, structure size: 3×3
3	MOG	Patch size: 50×50, step size: 10, noise component: 3, frames: 3, $k = 0.05$, $v_{\min} = 0.05$
4	WLDM	$L = 4$, $m = 2$, $n = 2$
5	MPCM	$N = 1, 3, \ldots, 9$
6	FKRW	$K = 4$, $p = 6$, $\beta = 200$, window size: 11×11
7	IPI	Patch size: 50×50, step size: 10, $\lambda = L/min(m,n)^{1/2}$, $L \in [2,5]$, $\varepsilon = 10^{-7}$
8	TVPCP	Patch size: 50×50, step size: 12, $\lambda_2 = L/min(m,n)^{1/2}$, $L \in [1,5]$, $\lambda = 0.005$, $\beta = 0.025$, $\gamma = 1.5$, $\varepsilon = 10^{-7}$
9	SMSL	Patch size: 50×50, step size: 50, $\lambda = L/min(m,n)^{1/2}$, $L \in [2,7]$, $\varepsilon = 10^{-7}$
10	GRLA	Patch size: 30×30, step size: 12, $\lambda_1 = L/max(m,n)^{1/2}$, $L \in [2,6]$, $\lambda_2 = G/min(m,n)^{1/2}$, $G \in [3,5]$, $\gamma = 0.01$, $\epsilon = 0.01$, $\varepsilon = 10^{-7}$
11	RIPT	Patch size: 50×50, step size: 10, $\lambda = L/min(I,J,P)^{1/2}$, $L \in [0.5,5]$, $h = 10$, $\epsilon = 0.01$, $\varepsilon = 10^{-7}$
12	Ours	Patch size: 20×20, step size: 10, $\lambda = L/min(m,n)^{1/2}$, $L \in (0,2]$, $\gamma_1 = \gamma_2 = 3$, $\varepsilon_1 = 10^{-6}$, $\varepsilon_2 = 10^{-14}$

5.2. Evaluation Indicators

The signal-to-clutter ratio gain (SCR_G) gives a measure of how much the complexity of the target area varies, defined as follows:

$$SCR_G = \frac{SCR_{out}}{SCR_{in}} \tag{24}$$

where SCR_{in} and SCR_{out} denote the SCR before and after background suppression separately, and it is defined as $SCR = |M_t - \mu_b|/(\sigma_b + \omega)$. M_t is the maximum intensity of target area. μ_b and σ_b are the average grayscale and standard deviation of the neighboring region around the small target. $\omega = 0.01$ denotes as a smoothing factor to avoid division zero. BSF gives some sense of how much the background suppression effect is presented, which is formulated as

$$BSF = \frac{\sigma_{in}}{(\sigma_{out} + \omega)} \tag{25}$$

where σ_{in} and σ_{out} denote the standard deviation of target neighboring region in original and suppressed images, respectively. Additionally, we introduce another metric to quantify the target contrast enhancement, namely contrast gain (CG) defined as

$$CG = \frac{CON_{out}}{CON_{in}} \tag{26}$$

where CON_{in} and CON_{out} denote the target local contrast before and after processing, respectively. The CON is defined as

$$CON = |M_t - \mu_b| \tag{27}$$

where M_t and μ_b are same as those in Equation (24). The higher the above three indicators, the better the background suppression performance of the detection method is. Both the original image and

the detection results are normalized to $[0, 1]$ when calculating these indicators. The three metrics are calculated in target local region, supposing that the target size is $a \times b$ and d is set to 20 as the neighborhood width. Furthermore, the probability of detection (P_d) and false alarm rate (F_a) are also very important indicators for wholly evaluating the detection performance, which are defined as

$$P_d = \frac{\text{number of true detections}}{\text{number of actual targets}} \tag{28}$$

$$F_a = \frac{\text{number of false detections}}{\text{number of images}} \tag{29}$$

In experiments, we deem that the detection of small target is correct under this case where there are pixels within a 5×5 window centering on the ground truth. A good detector owns high P_d under low F_a. The receiver operating characteristic (ROC) displays the dynamic relationship between P_d and F_a.

5.3. Validity of the Proposed Patch and Feature Sparse Regularizations

The importance of the patch and feature regularizations are validated with a series of experiments in four different scenarios. For Equation (7), the proposed model degenerates into a simple model with $l_{1/2}$-norm non-negative constraint by setting $\gamma_1 = \gamma_2 = 0$, whereas $\{\gamma_1 > 0, \gamma_2 = 0\}$ generates the patch regularized model (PRM), and $\{\gamma_1 = 0, \gamma_2 > 0\}$ produces the feature regularized model (FRM). Figure 5 shows the ROC results of four derived models for the tested scenes. From the figures, one can easily discover that the full model achieves the highest performance among other three variants of the proposed model, and the simple model without any regularizations obtains the lowest. For the deep-space scene, PRM performs better than FRM, as shown in Figure 5a. It is because the patch graph regularization can preserve clutter edges and recover them well, but the ability of feature graph is limited resulting in the performance degradation. With a sky-cloudy background (Figure 5b), FRM slightly outperforms PRM since the dim and weak target can be captured effectively by FRM but be ignored by PRM. With a sky-sea background (Figure 5c), PRM achieves higher P_d under the same F_a compared with FRM. The reason lies in that sea glitters that present similar sparsity to the small target may disrupt the dimension of feature space. FRM cannot effectively restrain them, and they then may be remained in the target images to raise the false alarm rate. For the terrain scenes, there is no significant difference between the performance of PRM and FRM. However, their probability of detection is prominently lower than that of the full model. By the above observations, it readily finds that the integration of patch and feature sparse regularizations contributes to an improvement of the detection performance in the proposed model.
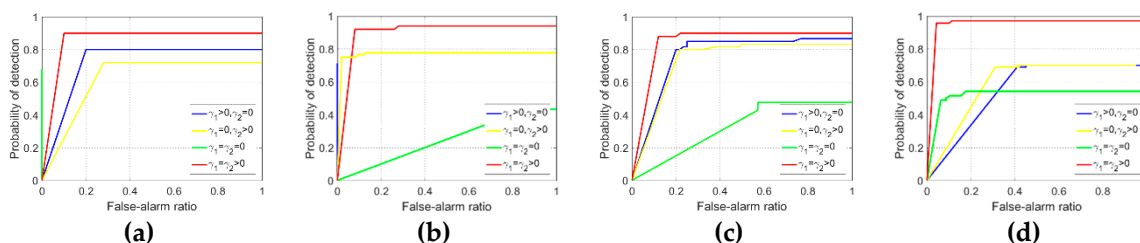

Figure 5. Performance of the proposed patch and feature sparse graph regularizations on different scenes, such as (**a**) deep-space, (**b**) sky-cloudy, (**c**) sea-sky, (**d**) terrain.

5.4. Sensitivity to Parameters

For the proposed model, there are several key parameters that affect its stability. Herein, we discuss the influence of these parameters on the model performance by investigating the variation of the detection probability with varying these parameters. The parameters include patch size, sliding step,

sparse penalty λ, patch graph regularization weight γ_1 and feature graph regularization weight γ_2. In order to better reflect the suitability of model parameters, we construct a tested dataset covering diverse scenes by randomly selecting 10 frames from each exhibited sequence. The variations in detection performance measured by ROC are visualized in Figure 6, where we vary one parameter while fixing others. Among them, patch size and sliding step are closely related to not only detection performance but also computational complexity, as shown in Figure 6a,b and Table 4. From the two subfigures, it is observed that larger patch and step exhibit the degradation of performance. It has largely been caused by a larger patch size and sliding step that undermines the relationship between the nonlocal patches. Then, the patch and feature graph may be inefficient in the preservation of the intrinsic structure of an image. On the other hand, although smaller patch and step may weaken the sparsity of a small target, the incorporation of dual-graph regularization is more conducive to maintaining the manifold structure of rare interferences. Then, the singularity of a small target can be highlighted well. In addition, we give the average execution time of the proposed method with different patch size and sliding step in Table 4. One can see that with a fixed-size patch, the larger the sliding step is, the shorter time the proposed model costs. Observing the two subfigures, the proposed model achieves the best in the tested dataset when the patch size is 20 and sliding step is set as 10.

Table 4. The average execution time (/s) of the proposed method with different patch sizes (*P*) and sliding steps (*S*).

S \ *P*	20	30	40	50	60	70
8	0.68	1.84	3.58	5.71	9.53	13.75
10	0.38	0.95	1.74	3.23	5.35	7.69
12	0.27	0.61	1.17	2.15	3.64	4.76
14	0.21	0.43	0.83	1.52	2.17	3.46
16	0.16	0.35	0.63	1.04	1.60	2.45
18	0.14	0.26	0.46	0.81	1.19	1.73
20	0.11	0.22	0.38	0.68	0.88	1.44

The sparse penalty λ plays a great significant role in controlling the balance between detection probability and false alarm ratio. To adjust this parameter more finely, we use $\lambda = L/min(m,n)^{1/2}$ varying L instead of directly varying λ. Figure 6c shows the impact of the parameter on the performance of the proposed method with changing L from 0.5 to 5. In the illustration, it is clearly observed that a larger penalty can effectively eliminate false alarms, but the detection probability is also decreased dramatically. For example, when L belongs to $(0, 2]$, the proposed method has a higher detection rate compared with the values outside the given interval. Although the ROC results of the proposed method become a straight line under larger penalty such as L in $[2.5, 5]$, meaning that it achieves zero false alarms, the detection rate is reduced seriously. In the experiments, when L takes $(0, 2]$, the proposed method is superior in robustness and effectiveness.

In spite of analyzing the importance of the proposed regularizations, we also performed experiments to evaluate the effect of the regularization parameters on the performance. Figure 6d–f shows the ROC results of the proposed method with regularization parameter variations for the constructed dataset. Among them, Figure 6d shows the result of varying γ_1 while fixing γ_2, and Figure 6e is obtained by varying γ_2 while fixing γ_1. For Figure 6f, it is obtained by varying both γ_1 and γ_2 while keeping $\gamma_1 = \gamma_2$. From these figures, we can see that as the regularized parameters increase, the P_d of the proposed method increases first and then decreases. It indicates that the penalty degree of different regularization has a certain impact on the performance of the proposed algorithm. In our test, $\gamma_1 = \gamma_2 = 3$ seems to be a good choice because it is the best in ROC.
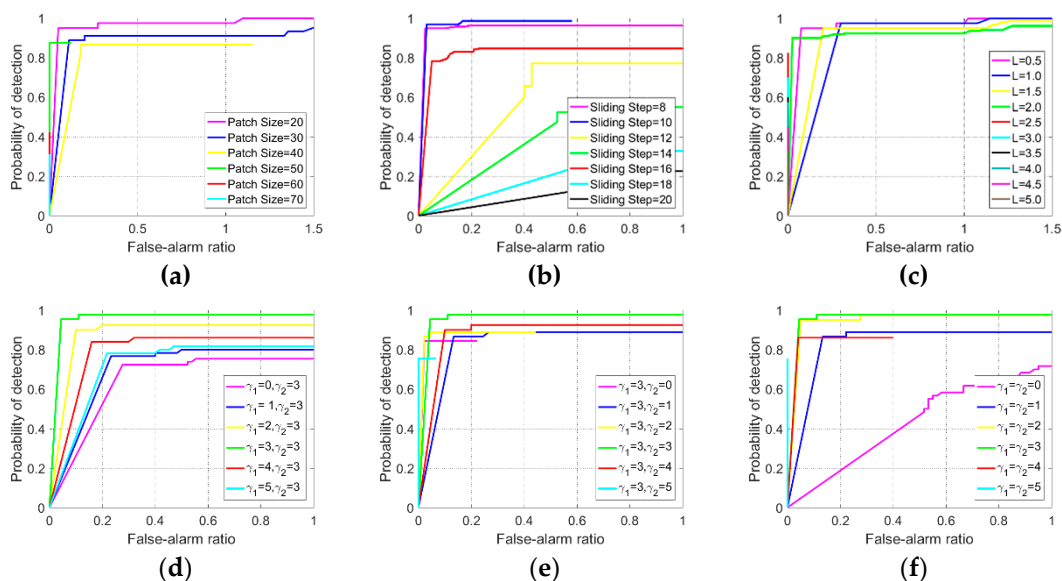
Figure 6. Receiver operating characteristic (ROC) results obtained by changing the parameters in the proposed model for the constructed test dataset. (**a**) Varying patch size. (**b**) Varying sliding step. (**c**) Varying sparse penalty. (**d**) Varying γ_1 while fixing γ_2. (**e**) Varying γ_2 while fixing γ_1. (**f**) Varying γ_1 and γ_2 while keeping $\gamma_1 = \gamma_2$.

5.5. Qualitative Evaluations

To further evaluate the performance of the proposed model, the visual comparisons with 11 state-of-the-art methods for different scenes are shown in Figures 7–10. In these results, only one representative image is selected from each sequence. For better observation, we enlarged the demarcated target area and creatively integrated target images with their 3D stereogram. From the figures, one can find that the typical filtering methods, such as TDLMS, TopHat, can highlight targets to a certain extent while retaining many background clutter residues in the detected results. MOG also belongs to the category of background prediction but different from the filtering methods. It can achieve relatively good performance on dim small target extraction in uniform scenes, especially for the sky-cloud background in Figure 8a–c. Compared with the results obtained by background prediction manners, WLDM and MPCM present marked improvement in enhancing the target, as illustrated in Figures 7–10. However, for the sky-cloud background with sparse noise, such as in Figure 8a, these two methods significantly highlight the small target while enhancing the background noise. Moreover, WLDM does not have enough power to deal with the border of image, as shown in Figure 9a,b and Figure 10a,b. FKRW first uses the filtering manner to eliminate high bright noise, and then employs the local saliency to suppress background clutter and enhance the small target. It is impressive at suppression of clutter and sparse noise in comparison with WLDM and MPCM. However, FKRW might suffer from incorrect detection, such as in Figure 7b. In addition, it is more susceptible to spot-like sea clutters in sea-sky scenes because the sea glitter has a similar appearance to a small target, as presented in Figure 9c. It is easily found that the methods derived on subspace learning perform better than other comparisons in background clutter removal. Meanwhile, we can clearly discover that the proposed method accurately extracts small targets from different complex scenes without any background residuals, which is attributed to the usage of the proposed patch-feature regularization. For the initial IPI model, it is less robust to complex scenes due to its above deficiencies. For example, in Figure 7a,b and Figure 9c, it incorrectly includes background clutter and sparse glitter into the target image, resulting in substantial false alarms. As extended versions of IPI, SMSL and RIPT improve the performance in eliminating false alarms. However, SMSL uses the manner of non-overlapping patch to construct the infrared patch-image. It will reduce the coherence of the sparse structure in patches. Some like-target points are left in the target image when using multi-subspace

learning strategy, as shown in Figure 7a,b, Figures 8a, 9c and 10b. Besides, RIPT shows sensitivity to the scenes with sparse points, such as in Figures 8a, 9c and 10b. TVPCP and GRLA can be regarded as enhancements of the IPI model. They integrate different regularization, such as total variation and graph regularization into the IPI model. TVPCP applies the total variation as the constraint of the background smooth component. From the results, one can obviously see that it successfully extracts the small target but some background interferences still remain, as shown in Figure 7a,b, Figures 8a, 9c and 10b. Among the two enhancements, GRLA can suppress background clutter better than TVPCP. The reason for this is that the graph regularization is more efficient in preserving the intrinsic structure within the background to address non-smooth component in complex scenes. Nevertheless, GRLA only employs one-side graph regularization, say in patch space. It achieves a comparable performance in static scenes, but it will lead to missing detection in dynamic scenes.
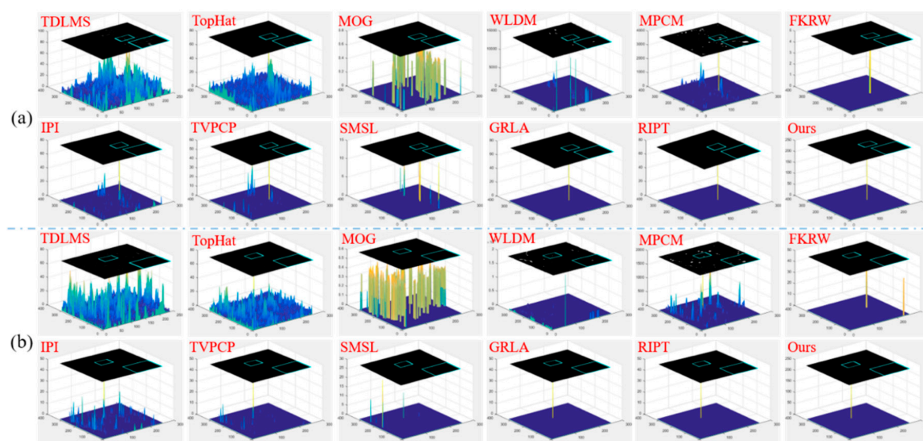

Figure 7. Detection results obtained by the proposed method over deep-space scenes compared with 11 comparative approaches. For better observation, the target images are integrated with their 3D stereogram. (**a**) Representative frame in Sequence 1. (**b**) Representative frame in Sequence 2.
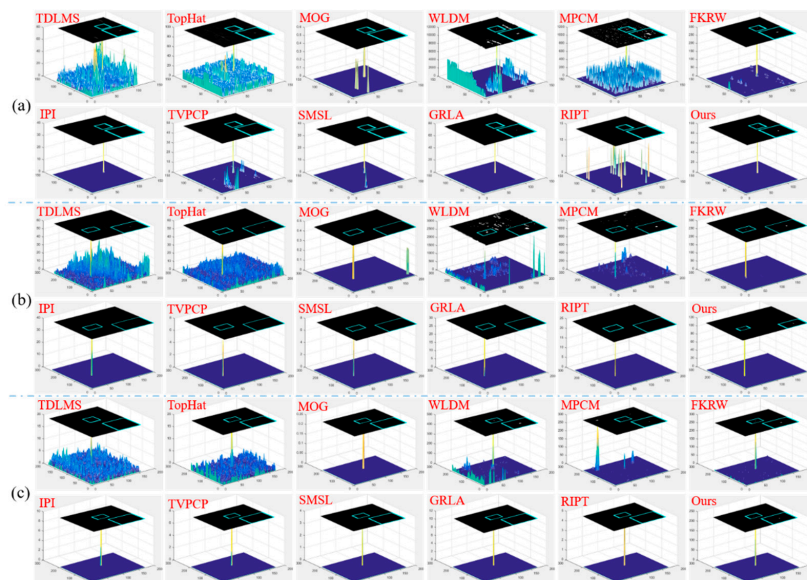

Figure 8. Detection results obtained by the proposed method over deep-space scenes compared with 11 comparative approaches. For better observation, the target images are integrated with their 3D stereogram. (**a**) Representative frame in Sequence 3. (**b**) Representative frame in Sequence 4. (**c**) Representative frame in Sequence 5.
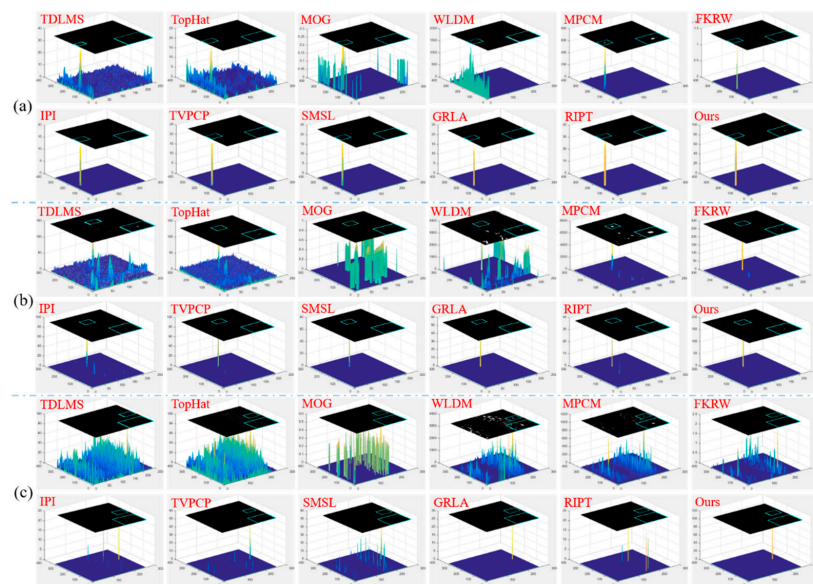
Figure 9. Detection results obtained by the proposed method over deep-space scenes compared with 11 comparative approaches. For better observation, the target images are integrated with their 3D stereogram. (**a**) Representative frame in Sequence 6. (**b**) Representative frame in Sequence 7. (**c**) Representative frame in Sequence 8.
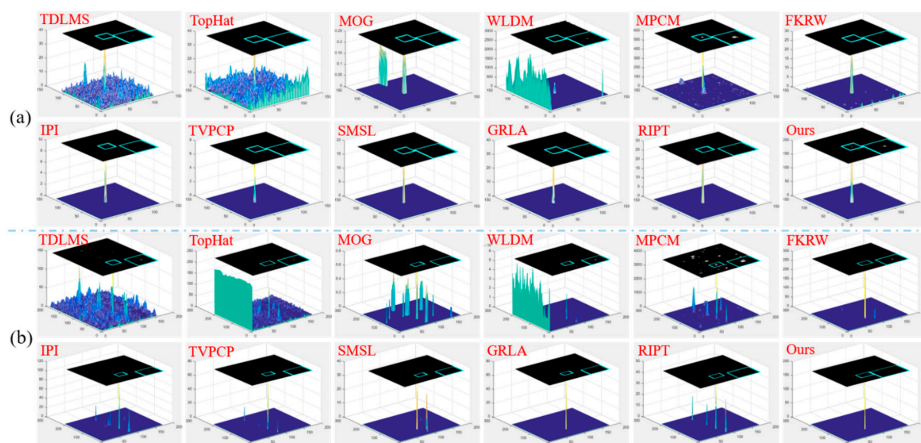

Figure 10. Detection results obtained by the proposed method over deep-space scenes compared with 11 comparative approaches. For better observation, the target images are integrated with their 3D stereogram. (**a**) Representative frame in Sequence 9. (**b**) Representative frame in Sequence 10.

5.6. Quantitative Evaluations

With the exception of the subjective visual evaluations, objective indicators are also important evaluation criteria. In this subsection, we provide the comparisons of the average SCR_G, BSF, and CG, which are obtained by all involved methods in the different scenes, as shown in Table 5. In experiments, the 10 infrared sequences are divided into four categories according to the type of scenes. In the table, we mark the best three results by red, blue, and green, respectively. The results present that, for most cases, the proposed method earns the first or second ranking place on these sequences across different indicators. Regarding the average SCR_G and BSF, the proposed method gets the highest scores in all tested methods for different scenarios, which means that our method has the best performance in clutter removal and background suppression. Additionally, GRLA is lower than RIPT in the deep-space scenes, but higher for the other three scenes. It is clear that methods based on subspace learning are generally better than other types of methods. This is because the IPI model takes

advantage of redundant information of the image patch to improve the robustness for diverse scenes. It is noteworthy that, for *CG*, the proposed method ranks the second in deep-space and sky-cloud scenes and the third in sea-sky and terrain-sky scenes. The reason lies in that the proposed method can separate small targets cleanly but does not enhance small targets. In contrast, although WLDM and MPCM do not completely suppress the background, they distinctly strengthen the target intensity. WLDM and MPCM obtain relative superior performance in *CG*, but their detection results contain numerous false alarms. In addition, the three indicators merely reflect the good performance in a local region, and do not necessarily represent the overall performance of the method.

Table 5. Comparison of the baselines with our method for average SCR_G, BSF, and CG for different scenarios. The red means the best, blue means the second and green means the third.

	Metrics	TopHat	MOG	WLDM	MPCM	FKRW	IPI	TVPCP	SMSL	GRLA	RIPT	Ours
Deep-space (Sequences 1 and 2)	SCR_G	1.75	4.26	18.42	23.75	102.19	82.14	112.03	196.64	296.46	348.13	512.06
	BSF	3.23	3.12	16.49	13.75	142.19	68.68	83.02	168.43	182.69	212.46	364.04
	CG	1.78	1.06	126.49	248.37	26.22	108.27	113.67	22.35	80.26	82.19	186.44
	Metrics	TopHat	MOG	WLDM	MPCM	FKRW	IPI	TVPCP	SMSL	GRLA	RIPT	Ours
Sky-cloud (Sequences 3–5)	SCR_G	4.54	3.15	10.81	32.28	55.62	126.81	124.26	212.06	318.29	316.42	586.24
	BSF	1.04	2.08	6.81	21.24	48.63	122.19	110.13	182.56	286.92	228.89	426.25
	CG	3.93	0.44	86.81	224.88	136.86	62.27	48.53	29.25	102.71	57.98	146.42
	Metrics	TopHat	MOG	WLDM	MPCM	FKRW	IPI	TVPCP	SMSL	GRLA	RIPT	Ours
Sea-sky (Sequences 6–8)	SCR_G	1.21	3.48	8.16	42.23	45.18	104.15	144.58	162.10	206.38	172.61	332.62
	BSF	2.65	3.68	4.25	22.23	25.81	128.76	156.36	109.24	166.73	118.09	306.16
	CG	2.76	2.18	116.23	186.43	32.52	65.87	75.35	36.62	48.67	69.08	108.91
	Metrics	TopHat	MOG	WLDM	MPCM	FKRW	IPI	TVPCP	SMSL	GRLA	RIPT	Ours
Terrain-sky (Sequences 9 and 10)	SCR_G	SCRG	1.42	2.69	15.35	58.39	70.92	96.00	108.66	134.69	266.73	232.56
	BSF	BSF	1.55	1.22	8.16	48.39	82.93	104.20	88.66	146.28	216.22	197.86
	CG	CG	2.82	0.67	286.18	257.39	69.28	81.40	52.11	30.28	66.62	86.43

To further assess the superiority of the algorithm globally, Figure 11 presents the ROC results obtained by different methods on the 10 displayed sequences. From the figure, we clearly identify that the proposed method achieves the best behaviors with respect to both detection accuracy and stability on different scenes compared with those competitive methods. Especially for the Sequences 2–7 (Figure 11b–g), our method attains the highest P_d at the lowest F_a. For Sequences 1 and 10 (Figure 11a,j), RIPT owns higher P_d when the F_a is less than 0.4 and 0.45, respectively. Nevertheless, with the increase of F_a, the proposed method reaches the highest P_d scores faster than RIPT. In addition, the significant changes in the performance of the compared algorithms can be easily discovered. For example, GRLA outperforms the others in Sequences 1, 6, 7, and 9 (Figure 11a,f,g,i), but is slightly inferior to some methods for other sequences. The P_d of RIPT in Sequences 1, 5, 7, 9, 10 (Figure 11a,e,g,i,j) arrives 1 within the range of F_a less than 1. However, it obtains lower P_d in remaining sequences, especially in Sequence 3. Figure 11e,g–i displays that TVPCP shows an impressive performance in these sequences while its performance declines seriously for other remaining sequences. Moreover, due to the limited capability to suppress background clutter, approaches based on background prediction and target saliency, such as TDLMS, TopHat, MOG, WLDM, and MPCM, have low values of P_d in the range of low F_a. Additionally, FKRW is less robust in Sequences 1, 2, 8, 10, obtaining low P_d. The conclusions drawn from objective evaluations demonstrate that our method is superior and more robust in detection performance against different scenes as compared with the baselines.
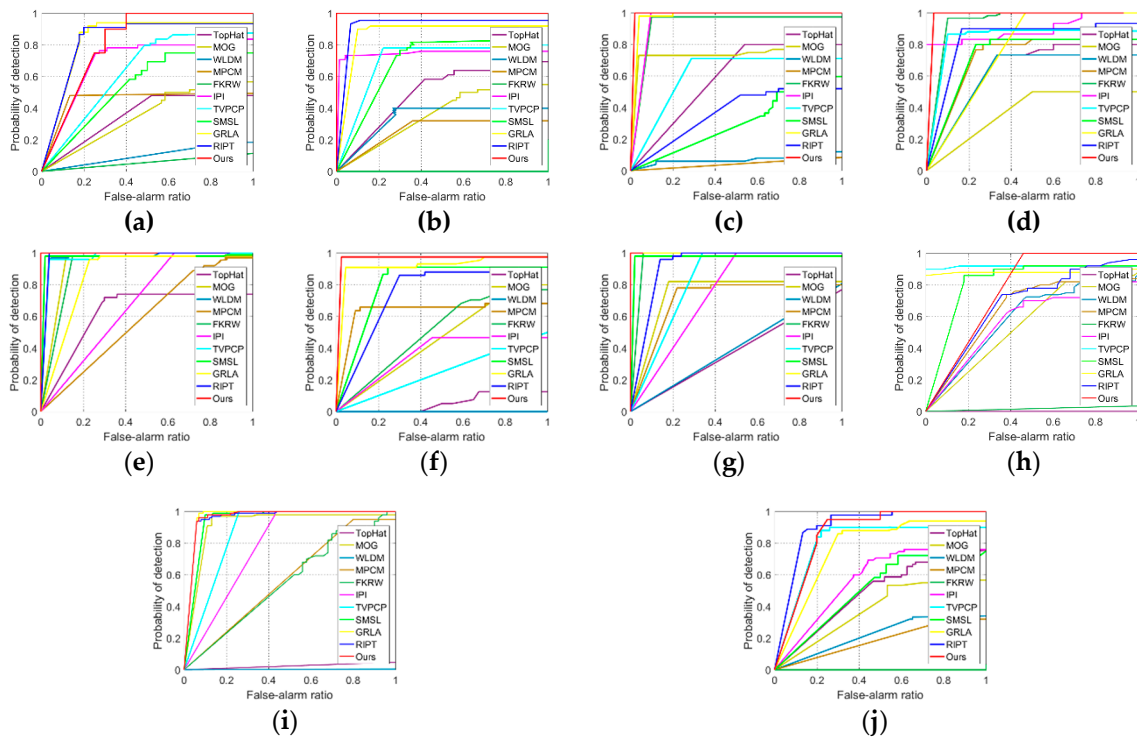
Figure 11. Comparison of the ROC results of different tested methods conducted on 10 sequences. (**a**–**j**): Sequences 1–10.

5.7. Convergence Analysis

The convergence of the proposed method solved by LADMAP was empirically studied on the real sequences, as illustrated in Figure 12. The relative error (r_e) in objective function served as the iteration stop criterion, which was computed in each iteration by $\|D - B_{k+1} - T_{k+1}\|_F / \|D\|_F < \varepsilon_2$. Figure 12a displays the changes of the relative error r_e in successive iterations of the objective function. Observing the figure, we can find that the declining rate of relative error is of great difference among Sequences 3, 5, 7, 9, and 10, and is almost uniform for the remaining sequences. Furthermore, the relative error drops sharply between iterations 1 and 2 that indicates a very fast convergence of the proposed method. Besides, the average number of iterations of the proposed method in all sequences is provided in Figure 12b, from which we can see that the convergence is obtained in less than 12 iterations.
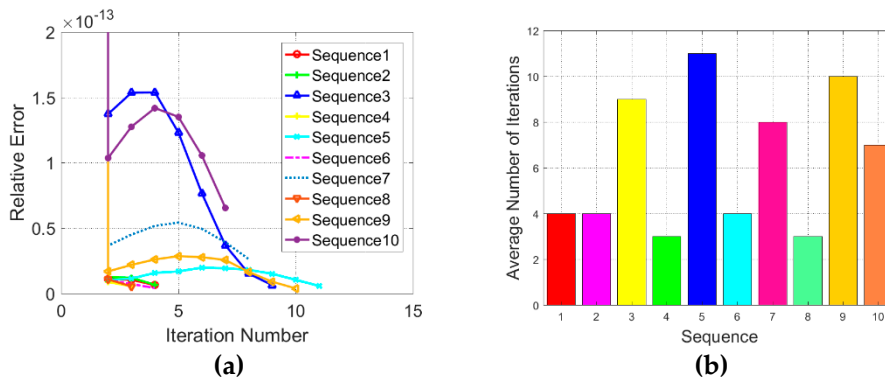

Figure 12. The empirical convergence analysis of the proposed method based on average number of iterations over 10 sequences. (**a**) Convergence curves. (**b**) Average number of iterations.

5.8. Execution Time Comparison

To evaluate the computational efficiency of the proposed method more intuitively, we provide a comparison of the average time executed by the proposed method and comparative methods on 10 sequences. All of the tested algorithms were implemented on a PC with Intel Core i5 CPU 3.4 GHz and 8GB RAM by MATLAB R2016b. Considering that the running time may be affected by the parameter settings of the tested methods, we used the installing parameters obtained under optimal performance, as given in Table 2. Table 6 provides the average running time of each tested method. From the table, one can find that the computational cost of MOG and TVPCP is much higher than other algorithms. TopHat takes the shortest time in all test methods, but its stability is the worst. Among subspace learning models, SMSL is the fastest since it employs the block coordinate descent method to avoid SVD in every iteration. As can be seen, although the proposed method is not the fastest, its execution time is at the same level as RIPT and GRLA. Furthermore, the execution time of the proposed method can be accelerated by parallelly implementing the construction of graph.

Table 6. The average execution time (frame/s) of different algorithms on 10 real sequences.

Methods	Seq 1	Seq 2	Seq 3	Seq 4	Seq 5	Seq 6	Seq 7	Seq 8	Seq 9	Seq 10
TopHat	0.069	0.063	0.034	0.059	0.052	0.085	0.065	0.071	0.042	0.056
MOG	269.9	274.5	7.73	121.7	77.5	156.2	235.7	172.4	1.64	126.32
WLDM	1.21	1.43	0.36	0.87	0.66	1.33	0.91	1.42	0.37	0.65
MPCM	0.58	0.56	0.42	0.36	0.58	0.69	0.62	1.07	0.47	0.15
FKRW	0.63	0.61	0.35	0.69	0.42	0.61	0.63	1.37	0.39	0.26
IPI	14.72	14.36	0.25	2.83	1.58	16.23	6.25	18.6	0.29	1.62
TVPCP	265.5	248.8	10.42	81.44	38.85	183.4	98.81	187.2	10.9	42.9
SMSL	0.39	0.35	0.14	0.21	0.19	0.36	0.59	0.38	0.10	0.49
GRLA	3.86	3.69	0.55	1.95	1.94	12.87	5.55	3.42	0.58	2.23
RIPT	1.81	1.91	0.18	0.85	0.37	1.54	1.22	2.09	0.20	0.60
Ours	2.34	2.09	0.22	1.41	0.73	1.99	1.39	2.31	0.37	0.86

6. Discussion

From the above analysis, we can see that the demand of real time detection can be satisfied by the methods based on local prior, e.g., local spatial consistency, local saliency. Additionally, these methods do really highlight a small target in the scenes with high signal-to-clutter ratio. However, they have high false alarm rates for the complex scenes due to the incapability to remove strong edges and bright spots. Methods based on subspace learning, by comparison, show the superiority in background suppression and sparse point elimination. It is mainly attributed to the usage of global prior including the nonlocal self-correlation of background and the sparsity of the small target. Nevertheless, the original IPI model still suffers from the performance degradation when facing extremely complicated scenes due to the aforementioned limitations. To overcome its flaws, some improvements have been developed subsequently along three lines: (1) selecting tighter rank approximation function to reduce the bias of background estimation, recovering background better [23,24]; (2) applying a sparse enhancement strategy such as reweighting to suppress non-target outliers in the background [21,22]; (3) imposing structural regularization to strengthen the correlation of background spatial structure, promoting the recovery of background [45,46]. These methods evidently reduce the false alarm rate and increase the detection probability for some complex scenes. However, these methods only employ the prior in patch space ignoring the one in feature space. Additionally, there is no closed form solution to the subproblems of some rank function surrogates. Hence, some of them may not only perform less robust in complex scenes with a dim small target, but have increased computational cost.

Different from the above methods, our proposed method integrates the prior in both data space and feature space in the form of dual graph regularization. This manner effectively constrains the background structures of deviating from low-rank subspace, boosting the suppression of clutters and edges. In addition, we employed $l_{1/2}$-norm instead of l_1-norm and its reweighting versions to constrain the target component, whose minimization realizes a sparser solution. It leads to a better removal of sparse outliers while preserving the small target. Finally, the solution of the objective function can be obtained by a well-designed optimization framework derived from LADMAP. By reducing the introduction of alternating multipliers, this framework not only reduces the number of iterations but also makes the computation friendly. The above experimental analysis has proved that the proposed method outperforms other tested methods in edge clutter suppression and sparse spot elimination. However, some issues still need to be considered. For example, the sparse constraint potentially assumes that the target elements are mutually independent, neglecting the spatial and pattern relation of some small targets. It will lead to the incompleteness of detection target. Furthermore, how to enhance the energy of small targets in the process of background separation is also a problem worthy of consideration. It will further increase the detection rate of dim and small targets. Following this, the future work can be concentrated on three lines: (1) designing a model to explicitly encode the spatial relation and feature similarity of small targets; (2) formulating a multi-frame detection model to fully explore the temporal cues in infrared sequences; (3) introducing appropriate deep learning framework to extract the target adaptively.

7. Conclusions

In this study, we gave an improvement on infrared small target detection when encountering complex scenes with cloudy clutter, sea glitter, and man-made interference. This is realized by incorporating the patch-feature sparse regularization into a RPCA framework with $l_{1/2}$-norm constraint. The $l_{1/2}$-norm can approximate a sparser solution than the traditional l_1-norm, and achieve effective removal of pixel-sized interferences. On the other hand, the knowledge embodied in the patch and feature space is imposed on the sparse component through the patch and feature graph Laplacian, where each graph preserves the corresponding manifold structure. The proposed model is efficiently solved by LADMAP. We demonstrate that the proposed model obtains excellent performance over different real infrared sequences compared with 11 state-of-the-art methods in terms of both objective evaluation and visual effect.

Author Contributions: F.Z. conceived the original idea, conducted the experiments, and wrote the manuscript. Y.D. and K.N. helped with data collection and revised the manuscript. Y.W. contributed to the writing, content, and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

References

1. Zhou, A.R.; Xie, W.X.; Pei, J.H. Background modeling in the fourier domain for maritime infrared target detection. *IEEE Trans. Circ. Syst. Vid.* **2019**, *99*, 1–16. [CrossRef]
2. Rozantsev, A.; Lepetit, V.; Fua, P. Detecting flying objects using a single moving camera. *IEEE Trans. Pattern Anal.* **2017**, *39*, 879–892. [CrossRef]
3. Li, Y.; Zhang, Y. Robust infrared small target detection using local steering kernel reconstruction. *Pattern Recogn.* **2017**, *77*, 113–125. [CrossRef]
4. Liu, D.L.; Li, Z.H.; Wang, X.R.; Zhang, J.Q. Moving target detection by nonlinear adaptive filtering on temporal profiles in infrared image sequences. *Infrared Phys. Technol.* **2015**, *73*, 41–48. [CrossRef]

5.　Rodriguez-Blanco, M.; Golikov, V. Multiframe GLRT-based adaptive detection of multipixel targets on a sea surface. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1–7. [CrossRef]

6.　Zhao, F.; Wang, T.T.; Shao, S.D.; Zhang, E.H.; Lin, G.F. Infrared moving small-target detection via spatiotemporal consistency of trajectory points. *IEEE Geosci. Remote Sens.* **2020**, *17*, 122–126. [CrossRef]

7.　Gao, C.Q.; Wang, L.; Xiao, Y.X.; Zhao, Q.; Meng, D.Y. Infrared small-dim target detection based on Markov random field guided noise modeling. *Pattern Recogn.* **2018**, *76*, 463–475. [CrossRef]

8.　Deshpande, S.D.; Meng, H.E.; Ronda, V.; Chan, P. Max-mean and Max-median filters for detection of small-targets. *Proc. Spie Int. Soc. Opt. Eng.* **1999**, *3809*, 74–83.

9.　Bai, X.Z.; Zhou, F.G. Analysis of new top-hat transformation and the application for infrared dim small target detection. *Pattern Recogn.* **2010**, *43*, 2145–2156. [CrossRef]

10.　Gu, Y.F.; Wang, C.; Liu, B.X.; Zhang, Y. A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications. *IEEE Geosci. Remote Sens.* **2010**, *7*, 469–473. [CrossRef]

11.　Chen, C.L.; Li, H.; Wei, Y.T.; Xia, T.; Tang, Y.Y. A local contrast method for small infrared target detection. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 574–581. [CrossRef]

12.　Wei, Y.T.; You, X.G.; Li, H. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recogn.* **2016**, *58*, 216–226. [CrossRef]

13.　Deng, H.; Sun, X.P.; Liu, M.L.; Ye, C.H.; Zhou, X. Small infrared target detection based on weighted local difference measure. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4204–4214. [CrossRef]

14.　Li, W.; Zhao, M.J.; Deng, X.Y.; Li, L.; Li, L.W.; Zhang, W.J. Infrared small target detection using local and nonlocal spatial information. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3677–3689. [CrossRef]

15.　Qin, Y.; Bruzzone, L.; Gao, C.Q.; Li, B. Infrared small target detection based on facet kernel and random walker. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7104–7118. [CrossRef]

16.　Yao, S.K.; Chang, Y.; Qin, X.J. A coarse-to-fine method for infrared small target detection. *IEEE Geosci. Remote Sens.* **2019**, *16*, 256–260. [CrossRef]

17.　Han, J.H.; Liu, S.B.; Qin, G.; Zhao, Q.; Zhang, H.H.; Li, N.N. A local contrast method combined with adaptive background estimation for infrared small target detection. *IEEE Geosci. Remote Sens.* **2019**, *16*, 1442–1446. [CrossRef]

18.　Gao, C.Q.; Meng, D.Y.; Yang, Y.; Wang, Y.T.; Zhou, X.F.; Hauptmann, A.G. Infrared patch-image model for small target detection in a single image. *IEEE Trans. Image Process.* **2013**, *22*, 4996–5009. [CrossRef]

19.　Wright, J.; Ganesh, A.; Rao, S.; Ma, Y. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. *Adv. Neural Inf. Process. Syst.* **2009**, *58*, 289–298.

20.　Wang, X.Y.; Peng, Z.M.; Kong, D.H.; He, Y.M. Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scenes. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5481–5493. [CrossRef]

21.　Dai, Y.M.; Wu, Y.Q. Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3752–3767. [CrossRef]

22.　Dai, Y.M.; Wu, Y.Q.; Song, Y. Infrared small target and background separation via column-wise weighted robust principal component analysis. *Infrared Phys. Technol.* **2016**, *77*, 421–430. [CrossRef]

23.　Zhang, L.D.; Peng, Z.M. Infrared small target detection based on partial sum of the tensor nuclear norm. *Remote Sens.* **2019**, *11*, 382. [CrossRef]

24.　Zhou, F.; Wu, Y.Q.; Dai, Y.M.; Wang, P. Detection of small target using schatten 1/2 quasi-norm regularization with reweighted sparse enhancement in complex infrared scenes. *Remote Sens.* **2019**, *11*, 2058. [CrossRef]

25.　Xu, Z.B.; Chang, X.; Xu, F.; Zhang, H. $L_{1/2}$ regularization: A thresholding representation theory and a fast solver. *IEEE Trans. Neur. Net. Learn.* **2012**, *23*, 1013–1027.

26.　Yin, M.; Gao, J.B.; Lin, Z.C. Laplacian regularized low-rank representation and its applications. *IEEE Trans. Pattern Anal.* **2016**, *38*, 504–517. [CrossRef]

27.　Shang, F.H.; Jiao, L.C.; Wang, F. Graph dual regularization non-negative matrix factorization for co-clustering. *Pattern Recogn.* **2012**, *45*, 2237–2250. [CrossRef]

28.　Javed, S.; Mahmood, A.; Al-Maadeed, S.; Bouwmans, T.; Jung, S.K. Moving object detection in complex scene using spatiotemporal structured-sparse RPCA. *IEEE Trans. Image Process.* **2019**, *28*, 1007–1022. [CrossRef]

29.　Tang, C.; Liu, X.W.; Zhu, X.Z.; Xiong, J.; Li, M.M.; Xia, J.Y.; Wang, X.K.; Wang, L.Z. Feature selective projection with low-rank embedding and dual laplacian regularization. *IEEE Trans. Knowl. Data. En.* **2019**, 1–14. [CrossRef]

30. Fan, B.J.; Cong, Y.; Tang, Y.D. Dual-graph regularized discriminative multitask tracker. *IEEE Trans. Multimed.* **2018**, *20*, 2303–2315. [CrossRef]

31. Lin, Z.C.; Liu, R.S.; Su, Z.X. Linearized alternating direction method with adaptive penalty for low-rank representation. *Adv. Neural Inf. Process. Syst.* **2011**, 612–620.

32. Zhao, Y.; Pan, H.B.; Du, C.P.; Peng, Y.R.; Zheng, Y. Bilateral two-dimensional least mean square filter for infrared small target detection. *Infrared Phys. Technol.* **2014**, *65*, 17–23. [CrossRef]

33. Sun, Y.; Yang, J.G.; Li, M.; An, W. Infrared small-faint target detection using non-i.i.d. mixture of gaussians and flux density. *Remote Sens.* **2019**, *11*, 2831. [CrossRef]

34. Qin, Y.; Li, B. Effective infrared small target detection utilizing a novel local contrast method. *IEEE Geosci. Remote Sens.* **2016**, *13*, 1890–1894. [CrossRef]

35. Bai, X.Z.; Bi, Y.G. Derivative entropy-based contrast measure for infrared small-target detection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2452–2466. [CrossRef]

36. Deng, H.; Sun, X.P.; Zhou, X. A multiscale fuzzy metric for detecting small infrared targets against chaotic cloudy/sea-sky backgrounds. *IEEE Trans. Cybern.* **2018**, *49*, 1694–1707. [CrossRef]

37. Liu, D.P.; Cao, L.; Li, Z.Z.; Liu, T.M.; Che, P. Infrared small target detection based on flux density and direction diversity in gradient vector field. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2528–2554. [CrossRef]

38. Lu, Y.; Huang, S.C.; Zhao, W. Sparse representation based infrared small target detection via an online-learned double sparse background dictionary. *Infrared Phys. Technol.* **2019**, *99*, 14–27. [CrossRef]

39. Wang, X.; Shen, S.Q.; Ning, C.; Xu, M.X.; Yan, X.J. A sparse representation-based method for infrared dim target detection under sea–sky background. *Infrared Phys. Technol.* **2015**, *71*, 347–355. [CrossRef]

40. Gao, J.Y.; Guo, Y.L.; Lin, Z.P.; An, W.; Li, J. Robust infrared small target detection using multiscale gray and variance difference measures. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 5039–5052. [CrossRef]

41. Wu, S.C.; Zuo, Z.R. Small target detection in infrared images using deep convolutional neural networks. *Int. J. Wavelets. Multi.* **2019**, *38*, 03371.

42. Lin, L.K.; Wang, S.Y.; Tang, Z.X. Using deep learning to detect small targets in infrared oversampling images. *J. Syst. Eng. Electron.* **2018**, *29*, 947–952.

43. Zhao, D.; Zhou, H.X.; Rang, S.H.; Jia, X.P. An adaptation of cnn for small target detection in the infrared. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 669–672.

44. Dey, M.; Rana, S.P.; Siarry, P. A robust flir target detection employing an auto-convergent pulse coupled neural network. *Remote Sens. Lett.* **2019**, *10*, 639–648. [CrossRef]

45. Wang, X.Y.; Peng, Z.M.; Kong, D.H.; Zhang, P.; He, Y.M. Infrared dim target detection based on total variation regularization and principal component pursuit. *Image Vis. Comput.* **2017**, *63*, 1–9. [CrossRef]

46. Zhou, F.; Wu, Y.Q.; Dai, Y.M.; Wang, P.; Ni, K. Graph-regularized laplace approximation for detecting small infrared target against complex backgrounds. *IEEE Access* **2019**, *7*, 85354–85371. [CrossRef]

47. Zhu, H.; Liu, S.M.; Deng, L.Z.; Li, Y.S.; Xiao, F. Infrared small target detection via low-rank tensor completion with top-hat regularization. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1004–1016. [CrossRef]

48. Deng, Y.; Dai, Q.H.; Liu, R.S.; Zhang, Z.K.; Hu, S.Q. Low-rank structure learning via nonconvex heuristic recovery. *IEEE Trans. Neur. Net. Learn.* **2013**, *24*, 383–396. [CrossRef]

49. Lin, Z.; Chen, M.; Ma, Y. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv* **2010**, arXiv:1009.5055.

50. Zhou, X.W.; Yang, C.; Zhao, H.Y.; Yu, W.C. Low-rank modeling and its applications in image analysis. *ACM Comput. Surv. (CSUR)* **2014**, *47*, 1–33. [CrossRef]

51. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Rin Mach. Learn.* **2011**, *3*, 1–122.

52. Wen, Z.W.; Goldfarb, D.; Yin, W.T. Alternating direction augmented lagrangian methods for semidefinite programming. *Math. Program. Comput.* **2010**, *2*, 203–230. [CrossRef]