



Yuxin Hu^{1,2,3}, Yapeng Ma^{1,2,3}, Zongxu Pan^{1,2,3} and Yuhan Liu^{1,2,*}

- Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China; yxhu@mail.ie.ac.cn (Y.H.); mayapeng20@mails.ucas.ac.cn (Y.M.); zxpan@mail.ie.ac.cn (Z.P.)
- ² Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
- ³ School of Electronic, Electrical, and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
- * Correspondence: liuyuhan@aircas.ac.cn; Tel.: +86-010-58887208

Abstract: Infrared imaging plays an important role in space-based early warning and anti-missile guidance due to its particular imaging mechanism. However, the signal-to-noise ratio of the infrared image is usually low and the target is moving, which makes most of the existing methods perform inferiorly, especially in very complex scenes. To solve these difficulties, this paper proposes a novel multi-frame spatial-temporal patch-tensor (MFSTPT) model for infrared dim and small target detection from complex scenes. First, the method of simultaneous sampling in spatial and temporal domains is adopted to make full use of the information between multi-frame images, establishing an image-patch tensor model that makes the complex background more in line with the low-rank assumption. Secondly, we propose utilizing the Laplace method to approximate the rank of the tensor, which is more accurate. Third, to suppress strong interference and sparse noise, a prior weighted saliency map is established through a weighted local structure tensor, and different weights are assigned to the target and background. Using an alternating direction method of multipliers (ADMM) to solve the model, we can accurately separate the background and target components and acquire the detection results. Through qualitative and quantitative analysis, experimental results of multiple real sequences verify the rationality and effectiveness of the proposed algorithm.

Keywords: infrared image sequences; dim and small target detection; complex background

1. Introduction

Infrared imaging is flexible, convenient and easy to conceal due to its unique imaging mechanism. These advantages make infrared imaging of great significance in military applications such as early warning, infrared precision strikes, and space-based debris detection [1]. Among them, infrared dim and small target detection is the key step. Therefore, modern military warfare puts forward higher requirements for infrared target detection. However, the infrared target usually appears dim and small because of the long imaging distance, lacking obvious features, especially in complex backgrounds [2]. Therefore, how to detect the target effectively and accurately has become a hot topic for scholars from all over the world. Although there is a sea of infrared target detection approaches, the existing infrared target detection methods still face huge challenges, especially in the detection of rapidly moving targets and complex backgrounds when the quality of the infrared image is low.

1.1. Related Works

Generally speaking, infrared dim and small target detection methods can be divided into two categories [3]: detect-before-track (DBT) and track-before-detect (TBD). As the name suggests, DBT focuses on the information in a single frame, and its idea is to detect the



Citation: Hu, Y.; Ma, Y.; Pan, Z.; Liu, Y. Infrared Dim and Small Target Detection from Complex Scenes via Multi-Frame Spatial–Temporal Patch-Tensor Model. *Remote Sens.* 2022, *14*, 2234. https://doi.org/ 10.3390/rs14092234

Academic Editor: Angel D. Sappa

Received: 17 March 2022 Accepted: 30 April 2022 Published: 6 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). target from each frame, thereby detecting the entire target. TBD focuses on the information utilization between multiple frames. However, the TBD methods require the accumulation of multiple frames, which have low timeliness, and require better hardware equipment, so the practicability is relatively poor. At present, the DBT methods have higher computational efficiency and stronger applicability, which is the focus of various countries.

Typical methods of TBD mainly include 3D matched filtering [4], that spatial-temporal saliency model adaptive matched filtering [5], the spatial-temporal saliency model [6,7], and so on. However, the background and target of the infrared image are not often static. The rapid changes in the background and target make the accuracy of the detection method based on the TBD idea relatively low. Moreover, the TBD method requires the accumulation of multiple frames with low timeliness and demands better configuration, which gives TBD insufficient practicability.

The idea of the popular DBT method is to detect the target from every frame and is mainly divided into three categories [8]:

- (1) The first kind of detection method is based on the assumption of background consistency. This idea assumes that the background is consistent and the targets are destructive pixels in the uniform background so that the target pixels can be extracted by filtering. Commonly used methods include Tophat filtering [9], maximum mean filtering and maximum median filtering, etc. [10]. The principle of this kind of method is relatively simple, but the robustness to noise is not strong, and the detection performance is relatively poor.
- (2) In order to improve the detection accuracy and robustness of traditional filtering methods, scholars have combined human visual system (HVS) [11] to infrared dim and small target detection. Chen and others [12] firstly utilized saliency feature extraction and proposed the local contrast measure (LCM), which is used to calculate the local contrast saliency map of each pixel and detect the small targets on the saliency map. On this basis, many methods have been subsequently developed. Han et al. [13] proposed an improved local contrast measure (ILCM) by changing the method of taking the slider in LCM, which improved the detection performance. Wei et al. [14] proposed a multiscale patch-based contrast measure (MPCM) under the assumption that the background is uniform and the target is bright. Bai et al. [15] proposed a detection method based on derivative entropy-based contrast measure (DECM). Shi et al. [16] proposed a high-boost-based multiscale local contrast measure (HB-MLCM) method based on high-boost-based contrast detection. Lu et al. [17] proposed a new small target detection method based on multidirectional derivativebased weighted contrast measures (MDWCM). Han et al. [18], who improved the filtering window and introduced a three-layer filtering window, proposed a new detection framework named multiscale tri-layer local contrast measure (TLLCM). Hao et al. [19] considered the brightness characteristics of the target and proposed a method based on multiple morphological profiles (MMP). Zhang et al. [20] detected infrared targets by improving pixel growth using two-dimensional density-distance space. HVS-based methods have been widely applied because they require less prior information and the time consumption is low. However, these approaches are as sensitive to noises as previous filtering methods and the detection results severely depend on the choice of parameters, whose performances are poor, especially in the face of complex backgrounds.
- (3) To solve the difficulties in extracting the target features and the adaptability to the scenes for previous algorithms, a method based on the low-rank and sparse decomposition (LRSD) framework is proposed. Based on the characteristics of infrared images, these methods avoid extracting the characteristics of the target itself. Instead, it makes low-rank and sparse assumptions for the background and the target as a whole, respectively, and models the image as consisting of noise, background, and target components. Through the establishment of the objective function and the optimization algorithm, the final detection result is obtained. Gao et al. [21] firstly proposed the

method of infrared patch image (IPI), which assumes that the background is lowrank, and the target is sparse, using nuclear norm minimization (NNM) to replace the rank function of the matrix. However, NNM has a problem of excessive target shrinkage. To solve this problem, Dai et al. [22] introduced the concept of re-weighting each patch and proposed a weighted infrared patch image (WIPI), using singular value partial sum minimization [23] to approximate the image rank. By introducing the γ norm to approximate the rank of the non-convex function and adding the $L_{2,1}$ norm to reduce the false alarm of strong edges, the non-convex rank approximation minimization (NRAM) model was proposed [24]. Zhang et al. [25] also used the L_P norm to approximate the rank and proposed NOLC model. Inspired by the total variation (TV) norm, Wang et al. [26] improved the robustness of the algorithm in non-uniform scenes by introducing the TV regularization term into the constructed model. Rawat et al. [27] replaced NNM with partial sum minimization (PSM) of singular values based on IPI and introduced the TV norm (TV-PSMSV) for infrared target detection. In addition, some subspace learning models have also been applied to the detection of small infrared targets, such as self-regularized weighted sparse (SRWS) [28], the stable multi-subspace learning (SMSL) [29] method and so on, which have also achieved good results.

To further improve the computational efficiency and the detection effect, Dai et al. [30] introduced the theory of tensor into the low-rank sparse model by transforming the matrix construction method [31,32], and proposed a reweighted infrared patch tensor (RIPT) model. Because tensors can make better use of the structural information between pixels, the research on tensors has since received extensive attention. Zhang et al. [33] proposed an image-block model using the information between pictures. Through adding the prior weight information of the corner and edges and using the partial sum of tensor singular values to approximate the rank of the non-convex function, partial sum of tensor nuclear norm (PSTNN) was proposed [34]. The PSTNN model approximates the rank by preserving the summation of some singular values, while the reservation is defined by a fixed energy ratio, which should be different for different scenes so that the estimation of various images can be improved. Zhang et al. [35]. proposed an edge and corner awareness-based spatial-temporal tensor model (ECA-STT) by introducing an edge–corner awareness indicator and adding a tensor-based non-convex tensor low-rank approximation (NTLA) regularization term to the model. Liu [36] preserved more information in the spatial-temporal domain by giving different weights to the spatial TV norm and the temporal TV norm, and thus proposed a new model, which shows a better performance in complex scenes. Kong [37] used a Log operator to replace the L_0 norm to approximate the rank of the background, also adding the spatial-temporal TV norm and thus proposed infrared small-target detection via non-convex tensor-fibered nuclear norm rank approximation (LogTFNN).

With the development of deep learning networks, some scholars have introduced deep learning methods into infrared target detection. Dai et al. [38] added local contrast to the network, embedding low-level information into high-level feature maps, and proposed attentional local contrast networks (ALCnet) for infrared small-target detection. Wang [39] built a model by generating an adversarial network for false alarms and missed detections as well. To alleviate the problem of insufficient spatial information of target objects, Qi et al. [40] proposed a single-stage small-object detection network (SODNet) to detect small objects after integrating professional feature extraction and information fusion technology. Due to the lack of infrared datasets [41] and the characteristics of images [36], the development of deep learning in infrared target detection is relatively difficult.

1.2. Motivation

Through the related work, we can see that traditional IPT models usually apply the L_1 norm to approximate the L_0 norm, considering every singular value equally. The result

obtained by this inaccurate approximation may be sub-optimal. Although some methods in the rank approximation have been recently proposed, such as the sum of the singular values in PSTNN or the approximation by Log operator in LogTFNN, the approximation of the rank still needs to be improved. In this article, we propose a rank approximation by Laplace operator. The proposed method has a more accurate approximation than other methods. In addition, it can automatically assign different weights to each singular value.

On the other hand, the existing methods have good performance in dealing with simple background situations, but cannot achieve good performance when the target movement changes rapidly and the background has significant noises and highlighted areas that will produce a lot of false alarms [42]. In these complex scenes, the traditional sliding window sampling method of a single frame finds it difficult to meet the hypothetical requirement of low rank for the background. Only considering the information of a single frame image for constructing the model, the algorithm cannot achieve good detection results. Therefore, the use of spatial information and inter-frame information is necessary.

Moreover, noises and clutter edges can cause false alarms in detection results. Prior weights used in RIPT ignore the features of the corners, which causes missing detections, while the fixed prior weights used in PSTNN and LogTFNN are only applicable to some scenes. In different images, the importance of the edge information and the corner point information should be different. In order to solve this problem, we propose a new prior weight calculation method.

Inspired by these perspectives, we proposed a multi-frame spatial-temporal patchtensor (MFSTPT) model for infrared small-target detection in complex scenes. The main contributions of this article are as follows.

- (1) Tensor construction exploits spatial and temporal information. We propose an approach to combine both spatial information and temporal information to construct the tensor model. The constructed model satisfies the low-rank assumptions much better and can also help to remove the false alarm clutters.
- (2) Approximation of rank by the Laplace operator. The Laplace operator is introduced to approximate the rank that has better performance than other methods in this paper. It assigns different weights to each singular value, which helps us to obtain an accurate background estimation.
- (3) Weighted prior weights. We propose a method for computing prior weights by weighting, which can give different weights to the corner information and edge information. By adjusting the importance of two structures, it is better at dealing with dim and small targets.
- (4) We apply the tensor construction, the rank approximation and weighted prior weights to the IPT model for infrared small-target detection, and the process of applying alternating direction method of multipliers (ADMM) to solve the model are introduced in detail. Experimental results verify the superior performance of our method.

The rest of the article is arranged as follows. The second section describes the mathematical symbols and formulas used in this paper, and the third section introduces the proposed model, including the construction of local prior weights, the proposed sampling method, and the introduction of Laplace approximation. The optimization process is explained in detail. The fourth part introduces the process of conducting experiments and makes qualitative and quantitative evaluations. Finally, the discussions and conclusions are illustrated in the fifth section and the last part, respectively.

2. Notations

For some symbols and theorems that will be used in this work, firstly, we will give some specific explanations. The symbols involved are explained in Table 1.

Notation	Instruction
X/X/x/x	Tensor/matrix/vector/scalar
$X_{n,m,k}$	its (n, m, k) th element
$X_{:::k}$ or $X^{(i)} / X_{::m:} / X_{n:::}$	<i>k</i> -th frontal/ <i>m</i> -th lateral/ <i>n</i> -th horizontal slice
X^i	the <i>i</i> -th iteration of X
$\mathcal{X}_{(i)}$	the mode- <i>i</i> unfolding matrix of X
	The zero norm of X is the number of
$\ \mathbf{A}\ _{0}$	non-zero elements
$\ X\ _{1}$	The sum of all non-zero elements in X
¥	The kernel norm of ${\mathcal X}$ is the sum of all singular
<i>∥∩∥</i> ∗	values in the matrix
	The Frobenius norm of X is the sum of the
$\ X\ _F$	squares of all values in the tensor and then the
	square root
$\overline{X} = fft(X, [], 3)$	Fourier transform of X
$X = ifft(\overline{X}, [], 3)$	Inverse fourier transform of \overline{X}

Tał	ole	1.	Mathematical	sym	bols
-----	-----	----	--------------	-----	------

Theorem 1. *Tensor singular value decomposition (t-SVD) algorithm.*

The t-SVD [43] is not carried out in the original domain, like SVD. The properties of the matrix are calculated in the Fourier transform domain [44]. Given a three-dimensional tensor $X \in \mathbb{R}^{n1*n2*n3}$, it can be broken down into:

$$X = U * S * V^T \tag{1}$$

where $U \in R^{n1*n1*n3}$, $V \in R^{n2*n2*n3}$ are orthogonal tensors, that is $U^T * U = U * U^T$, $V * V^T = V^T * V$. $S \in R^{n1*n2*n3}$ is a diagonal tensor. Each frontal slice of S is an orthogonal matrix as shown in Figure 1. The algorithm flow is in Algorithm 1.



Figure 1. Illustration of t-SVD.

Algorithm 1: Three-dimensional tensor decomposition by t-SVD. t-SVD of 3D tensors.

Input: $X \in \mathbb{R}^{n1*n2*n3}$ Output: U, S, V after tensor decomposition 1. $\overline{X} = fft(X, [], 3)$ 2. Count each front slice $\overline{U}, \overline{S}, \overline{V}$ through for $i = 1, \dots, \lceil (n_3 + 1)/2 \rceil$ do $\lceil \overline{U}^{(i)}, \overline{S}^{(i)}, \overline{V}^{(i)} \rceil = SVD(\overline{X}^{(i)})$ end for $i = \lceil (n_3 + 1)/2 \rceil + 1, \dots, n3$ do $\overline{U}^{(i)} = conj(\overline{U}^{(n3-i+2)});$ $\overline{S}^{(i)} = (\overline{S}^{(n3-i+2)});$ $\overline{V}^{(i)} = conj(\overline{V}^{(n3-i+2)});$ end 3. Count $U = ifft(\overline{U}, [], 3), S = ifft(\overline{S}, [], 3), V = ifft(\overline{V}, [], 3)$ **Definition 1.** As mentioned in PSTNN [25], the conjugate transpose of a tensor is defined as follows [45]:

$$(X^T)^{(1)} = (X^{(1)})^T$$
(2)

$$(X^T)^{(1)} = (X^{(n_3+2-l)})^T \quad l = 2, \cdots, n3$$
 (3)

3. Proposed Model

3.1. Image Patch Tensor (IPT) Model

Through the analysis of the characteristics, the infrared image can be modeled by three parts, which are low-rank background components, sparse target components and additive noise components [21,25]. The model of images can be expressed as:

$$f_D = f_T + f_B + f_N \tag{4}$$

Among them, f_T , f_B , f_N , f_D represent target, background, noise component and original infrared image, respectively. The IPI model makes full use of this feature, and its model is expressed as

$$\mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{N} \tag{5}$$

where $\mathcal{T}, \mathcal{B}, \mathcal{N}, \mathcal{D}$ represent the target, background, noise and the original infrared image. This does not directly detect the small targets in infrared images, but constructs the image patches by sliding a window of a certain size and divides the obtained patches into multiple columns of the matrix. Due to the imaging mechanism of the infrared images, the pixels in the background are very similar and thus the built matrix model is fully non-locally correlated. Thereby, the background matrix can be considered as a low-rank component, while brighter small targets occupy little pixels, which can be considered sparse. In this way, the target detection problem is transformed into a robust principal component analysis (RPCA) problem, which has been widely used since it was proposed [46,47].

One of the big shortcomings in the IPI model is that it destroys the local characteristics between pixels [48]. In order to optimize the model, the image patch tensor (IPT) is proposed. The assumption of this model is similar to the IPI model and is shown in Equation (6):

$$D = B + T + N \tag{6}$$

where $D, B, T, N \in \mathbb{R}^{m * n * k}$ describe the input image tensor, background tensor, target tensor and noise tensor, respectively. According to the hypothesis, the target tensor is a sparse tensor, which satisfies $||T||_0 < A$, where A is a constant determined by image complexity and represents the degree of sparsity. At the same time, it is generally assumed that the noise is additive Gaussian noise, which satisfies $||N||_F < \delta$, and if $\delta > 0$ we have $||D - B - T||_F < \delta$. We make assumptions about the background as:

$$\operatorname{rank}(\mathcal{B}_{(1)}) \le q1, \operatorname{rank}(\mathcal{B}_{(2)}) \le q2, \operatorname{rank}(\mathcal{B}_{(3)}) \le q3 \tag{7}$$

where q_{1} , q_{2} , q_{3} are positive numbers related to the background, and $rank(\mathcal{B}_{(1)})$ represents the number of non-zero singular values of matrix $\mathcal{B}_{(1)}$. Figure 2a is the original background image, and (b) is the three sets of singular values obtained by expanding the front slice, horizontal slice, and side slice. No matter in which mode, the singular value drops very fast, which verifies our hypothesis and we can draw the conclusion that the background image is low-ranking. Moreover, the sparse nature of the target is obvious. Therefore, the detection model can be described as:

$$\min_{\substack{B,T\\s.t.}} rank(B) + \lambda \|T\|_0$$

$$s.t. \quad D = B + T + N$$
(8)

where λ is a trade-off factor. Solving the L_0 norm is an NP-hard problem, and the L_1 norm is generally used instead [49].

S



Figure 2. Illustration of the low-rank property of the unfolding matrices. (**a**) is the original background image, and (**b**) is the three sets of singular values obtained by expanding the front slice, horizontal slice, and side slice.

3.2. Information of Local Structure Tensor

Non-local correlation is robust to the entire background, but strong edges or some bright corners will cause false alarms in the target image [37] because the sparsity of these background residuals can be confused with the target. The structure tensor of each pixel position in the image contains two eigenvalues, λ_1 and λ_2 . When the pixel is located at the corner area, $\lambda_1 \ge \lambda_2 \gg 0$; when it is located at the edge area, $\lambda_1 \gg \lambda_2 \approx 0$; when the pixel is at the flat edge area, $\lambda_1 \approx \lambda_2 \approx 0$. The eigenvalues of the structure tensor can be calculated by the following formulas [50]:

$$J_{\rho} = K_{\rho} * \left(\Delta D_{\rho} \otimes \Delta D_{\rho}\right) = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} = \begin{pmatrix} K_{\rho} * I_{x}^{2} & K_{\rho} * I_{x} I_{y} \\ K_{\rho} * I_{x} I_{y} & K_{\rho} * I_{y}^{2} \end{pmatrix}$$
(9)

$$\lambda 1 = \frac{1}{2} (J_{11} + J_{22} + \sqrt{(J_{22} - J_{11})^2 + 4J_{12}^2}) \lambda 2 = \frac{1}{2} (J_{11} + J_{22} - \sqrt{(J_{22} - J_{11})^2 + 4J_{12}^2})$$
(10)

where Δ represents the gradient, I_x and I_y represent directional derivative, K_ρ represents the Gaussian kernel function with variance ρ , and \otimes represents kronecker product. RIPT uses Formula (11) as the prior weight:

$$E(x,y) = \lambda 1 - \lambda 2 \tag{11}$$

However, the problem with the prior weights used in RIPT is the trade-off relationship between the size of the retained target and the false alarm, which is greatly affected by the weight parameter. In addition, too much attention is paid to the edge prior information of the background without considering the target, which results in the loss of the target. We display the results obtained with different prior weights in Figure 3. The target position is marked with a yellow box, and the background clutter is marked with a red box. The (1) row shows the classic infrared small targets in complex and simple scenes. The (2)-(6)rows are the saliency maps obtained by different calculation methods. The results obtained according to Formula (11) are shown in the (2) row. It can be seen that not only is the target position highlighted, but the clutter is also largely highlighted. In order to overcome this shortcoming, the corner point index [51] is introduced as Formula (12) and the corresponding results are shown in row (3). We can see from Figure 3 that the corner pixels are more prominent, but the spot-like information in the background is also highlighted. Furthermore, the most direct thought is to use the calculation method of prior weight information represented by the Formula (13), but this method can only have obvious prominent effects on the target in a scene with a simple background, such as (Figure 3(b4)), which means

the image is in column b and row (4) in Figure 3. Once the background is complicated, its effect will become worse, as shown in (Figure 3(c4)). To comprehensively utilize the information of corner points and edges and reduce the influence of edges, the method used in PSTNN is shown as Formula (14), which abandons the difference between two eigenvalues and directly utilizes the maximum value to represent the edge, as displayed in row (5) in Figure 3. To a certain extent, the prior information obtained by Formula (14) can indeed suppress the residual effect of the edge, while it also highlights the information that is not the target, so it still needs to be improved.

Based on the above considerations, we believe that the indicators for edges and corners correspond to different prior information in the image, so their importance should not be treated equally [52]. Therefore, a weighted method is proposed to weigh the ratio of the edges and corners, which are our feature indicators computed by Formula (15). If q = 1 is fixed, when p > 1, we believe that the corner information is more important. When p < 1, we believe that the edge information is more important. When p > 1, we believe that the edge information is more important. When p = 1, we believe that the shown in Figure 3 with n = p + q in our simulation. From the last row in Figure 3, it can be seen that in a complex or simple background, this method has two advantages over other methods: (1) the target information we need is highlighted and (2) the background residual is greatly suppressed.

$$C(x,y) = w_{cs}(x,y) = \frac{\det(ST(x,y))}{tr(ST(x,y))} = \frac{\lambda 1 * \lambda 2}{\lambda 1 + \lambda 2}$$
(12)

$$W = C \odot E \tag{13}$$

$$W(x,y) = \max(\lambda 1, \lambda 2) \odot \frac{\lambda 1 * \lambda 2}{\lambda 1 + \lambda 2}$$
(14)

$$W_c(x,y) = \sqrt[n]{C^p E^q} \tag{15}$$

Among them, (x, y) represents the position of the pixel in the image, and \odot represents Hadamard product. det refers to the determinant of the matrix. tr means the trace of the matrix, and ST is the structure tensor. We normalize Formula (15) to the form of (16) as:

$$W_c = \frac{W_c - w_{\min}}{w_{\max} - w_{\min}} \tag{16}$$

Here, w_{max} , w_{min} represent the maximum and minimum values of W_c , respectively.

Solving the traditional L_0 norm is an NP-hard problem, and using L_1 norm instead is a popular method because the L_1 norm is the best convex approximation to the L_0 norm, and it is easier to solve. Therefore, for the sparsity of the target, referring to the methods of other models [21,25,36,53], we use the L_1 norm to measure. Therefore, our model is updated to:

$$\min_{\substack{B,T\\ s.t.}} rank(B) + \lambda \|T\|_1$$

$$s.t. \quad D = B + T + N$$

$$(17)$$

The re-weighted L_1 minimization scheme is adopted to solve the penalty problem of different coefficients and shorten the convergence time, so that the sparsity enhancement weight has been widely used in recent years [47,54,55]. The weight on sparsity is defined as:

$$W_{cw}^{k+1} = \frac{\beta}{\left|T^{k}\right| + \gamma} \tag{18}$$

where β is a constant. γ is a small positive number to ensure normal operation [56]. *k* is the number of iterations. Combined with (18), the weight established by our algorithm is:

$$W_F = W_{cw} \odot W_{cr} \tag{19}$$

 W_{cr} is the reciprocal of W_c . Then, the model turns into:

$$\min_{B,T} \operatorname{rank}(B) + \lambda \| W_F \odot T \|_1$$

$$s.t. \quad D = B + T + N$$
(20)



Figure 3. Different prior maps. Original images corresponding to the (1) row. Prior weight saliency maps calculated by Formula (11) corresponding to the (2) row. Prior weight saliency maps calculated by Formula (12) corresponding to the (3) row. Prior weight saliency maps calculated by Formula (13) corresponding to the (4) row. Prior weight saliency maps calculated by Formula (14) corresponding to the (5) row, and prior weight saliency maps of the proposed algorithm calculated by Formula (15). Columns (**a**–**c**) are the prior weights obtained by different calculation methods for the same sequence

3.3. Spatial-Temporal Low-Rank Tensor Construction

The traditional IPT model firstly obtains the input tensor by using sliders to sample the image patch from each frame [30]. Then, the resulting multiple sliders are combined into a three-dimensional tensor, which is the input of the model, as shown in Figure 4. Finally, the model is optimized and solved by the corresponding algorithm. This sampling method is suitable for the case of simple background, but when the background is complex, it is no longer in line with the low-rank constraint well [57]. The disadvantage of this method is that there are many background residuals in the final target image. In the infrared images, dim and small targets move faster than the background, and the background between adjacent frames can basically be considered stationary. Inspired by the multiframe sampling strategy mostly utilized in the field of detection from videos [46], we can also make use of inter-frame information. By sampling adjacent frames, thereby, the tensor will satisfy the low-rank assumption well and both temporal information and spatial information are fully utilized.



Figure 4. Traditional sliding window sampling method.

Inspired by using different frames [58], we propose a novel tensor construction method, aiming at the detection of dim and small targets from complex infrared images. Firstly, for the current slider of the current frame, we use n frames before and after, which means to solve the current slider we utilize in total $(2 \times n + 1)$ frames. Then, for each frame, we treat the current patch as the center and acquire other neighborhood patches with distance d. Finally, all of the sliders are ordered from top to bottom, left to right and front to back. These sliders make up the tensor of the current slider for the current frame, as shown in Figure 5. In this work, *n* is set to 2 and *d* takes 1. The image corresponding to the vellow boxes in Figure 5 is the current frame and the yellow filled part is the current slider, and the images with orange boxes are the adjacent frames. In Figure 5, a total of 45 block images are obtained for the current center slider to form the tensor to be solved. Figure 6 demonstrates the classical complex infrared backgrounds and on the right side is the singular value of the corresponding images obtained by different sliding window methods. The blue line is the classic sampling method, and the red line is ours. It can be seen that in a complex background environment, the rank of the tensor obtained by making full use of the effective information of space and time declines faster than that obtained by only using the spatial non-local correlation. Therefore, the tensor model proposed in this work is better satisfied with the constraint of low-rank background.



Figure 5. The proposed tensor construction method by fully using spatial and temporal information.



Figure 6. Singular values of different methods for tensor construction. The (**a**) column shows classical complex backgrounds, and on the (**b**) column are the results in descending order of singular values by different methods. The blue line is the classical sliding window sampling method, and the red line is the one proposed in this paper.

3.4. Rank Approximation Based on Laplace Operator

Measuring the rank of the background tensor is a critical problem. Recently, a tensor rank approximation based on the Laplace operator has been applied to low-rank tensor completion [59]. Tensor nuclear norm (TNN) [60] and sum of nuclear norm (SNN) [31] assign the same weights to each singular value, leading to the problem that the targets are being severely shrunk. Large singular values correspond to more detailed information when representing the background in the image [61]. Although the SNN in PSTNN has

made an improvement over the traditional approximation, it assigns the same weights to the reserved singular values, and the preset parameter N is determined by a fixed energy ratio, which is not suitable for different images. The Laplace operator not only can automatically assign different weights to singular values [62], but also can achieve smaller deviation than the Log operator when the singular value is relatively small, leading to a more accurate approximation than the L_0 norm. As shown in Figure 7, the black, yellow, red and green curves refer to the L_0 norm, L_1 norm, Log operator and Laplace operator, respectively. The Laplace function is defined as follows [59] and the green line displays the approximation with $\varepsilon = 1$:

$$\|X\|_{\varepsilon} = \sum_{k=1}^{n3} \sum_{j=1}^{n} \phi\left(\sigma_{j}\left(\bar{\mathcal{X}}^{(k)}\right)\right)$$

= $\sum_{k=1}^{n3} \sum_{j=1}^{n} \left(1 - e^{-\sigma_{j}\left(\bar{\mathcal{X}}^{(k)}\right)/\varepsilon}\right)$ (21)

Here, $\phi(\mathcal{X}) = 1 - e^{-\sigma_j(\mathcal{X})/\varepsilon}$, $n = \min(n1, n2)$, and ε is a positive constant. $\sigma_j(\mathcal{X})$ is the singular value of the tensor slice.

It can be seen from Figure 7 that the Laplace patch tensor nuclear norm (LPTNN) based on the Laplace operator is a more accurate measurement than other norms to approximate the rank of the background tensor. Therefore, our model is updated to:



Figure 7. Comparison of contributions of singular values to rank approximation measured by different methods.

3.5. Model Optimization

Model to Be Solved

In this work, we apply ADMM to solve the established model [63] as well, thereby obtaining the background and target components of the image. First, we write the problem as an augmented Lagrangian function, that is:

$$L_{\mu}(B,T,W,Y) = \|B\|_{LPTNN} + \lambda \|W_F \odot T\|_1 + \langle Y, B + T - D \rangle + \frac{\mu}{2} \|B + T - D\|_F^2$$
(23)

where *Y* is the Lagrange multiplier. μ is a trade-off factor greater than 0, and <• > represents the inner product. The cross-multiplier method is to fix one of the variables and solve the other variables; that is, the problem can be decomposed into two sub-problems:

$$T^{k+1} = \arg\min_{T} \lambda \left\| W_F^k \odot T \right\|_1 + \frac{\mu^k}{2} \left\| B^k + T - D + \frac{Y^k}{\mu^k} \right\|_F^2$$
(24)

$$B^{k+1} = \arg\min_{B} \|B\|_{LPTNN} + \frac{\mu^{k}}{2} \left\|B + T^{k+1} - D + \frac{Y^{k}}{\mu^{k}}\right\|_{F}^{2}$$
(25)

(1). To solve the problem specified in Equation (24), the soft threshold algorithm, according to the reference [64], is utilized. When the problem is:

$$\arg\min_{X} \alpha \|X\|_{1} + \frac{1}{2} \|X - Z\|_{F}^{2}$$
(26)

The solution can be obtained by thresholding the elements as (27)

$$S_{\tau}(x) = sign(x) * \max(|x| - \tau, 0) \tag{27}$$

Therefore, the sub-problem of background separation can be solved by [59]:

$$T^{k+1} = S_{\frac{\lambda W_F^k}{\mu^k}} \left(D - B^k - \frac{Y^k}{\mu^k} \right)$$
(28)

(2). We then transform the problem of the background tensor into an optimization problem as shown below:

$$\underset{X}{\arg\min} \|X\|_{LPTNN} + \frac{\beta}{2} \|X - Z\|_{F}^{2}$$
(29)

It can be seen from Algorithm 1 that the substituted non-convex rank of *LPTNN* is the combination of all front slices along $X_{(:,:,1...n3)}$ in the Fourier transform domain, so the optimization problem of (29) is transformed into the optimization problem of the sum of *n*3 matrices [59].

$$\arg\min_{\overline{X}^{(g)}} \sum_{i=1}^{n^3} \phi(\sigma_i(\overline{X}^{(g)})) + \frac{\beta}{2} \left\| \overline{X}^{(g)} - \overline{Z}^{(g)} \right\|_F^2$$
(30)

Here, $\overline{X}^{(g)}$, $\overline{Z}^{(g)} \in \mathbb{R}^{n1*n2}$, $g = 1, 2, \dots, n3$. Equation (30) can be solved by the generalized weighted singular value threshold operator [65]:

$$\overline{X}^{(g)} = \overline{U}^{(g)} * \overline{D}^{(g)} * \overline{V}^{(g)H}$$
(31)

$$\overline{D}_{\frac{\Delta\phi}{\beta}}^{(g)} = \max((\overline{S}^{(g)}(i,i) - \frac{\Delta\phi(\sigma_i^{k,g})}{\beta}), 0)$$
(32)

Among them, $\overline{Z}^{(g)} = \overline{U}^{(g)} * \overline{S}^{(g)} * \overline{V}^{(g)H}$ and $\Delta \phi(\sigma_i^{k,g}) = \frac{1}{\varepsilon} \exp(\frac{-\sigma_i^{k,g}}{\varepsilon})$. X can be then obtained by *ifft*. So, the sub-problem (25) can be solved. The algorithm flow is shown in Algorithm 2. Y and μ are updated as:

$$Y^{k+1} = Y^k + \mu^k (D - B^{k+1} - T^{k+1})$$
(33)

$$\mu^{k+1} = \rho \mu^k \tag{34}$$

Here, ρ is a positive constant. The process of the ADMM optimization solution is shown in Algorithm 3.

Algorithm 2: Optimization of problem (25).

Input: $Z^{k} = D - T^{k+1} - \frac{Y^{k}}{\mu^{k}} \in R^{n1*n2*n3}, \lambda, \mu^{k}$ Output: B^{k+1} 1. Calculate $\overline{Z}^{k} = fft(Z^{k}, [], 3)$ 2. Compute each frontal slice \overline{B}^{k+1} by for $i = 1, \dots, \lceil (n_{3} + 1)/2 \rceil$ do (1). $[\overline{U}^{(g)}, \overline{S}^{(g)}, \overline{V}^{(g)}] = \text{SVD}(\overline{Z}^{(g)})$ (2). Calculated by (32) (3). $(\overline{B}^{k+1})^{(g)} = \overline{U}^{(g)} * \overline{D}_{\frac{\Delta\phi}{\overline{B}}}^{(g)} * \overline{V}^{(g)}$ end for $i = \lceil (n_{3} + 1)/2 \rceil + 1, \dots, n3$ do $\overline{B}^{k+1}(g) = \operatorname{conj}(\overline{B}^{k+1}(n3 - g + 2))$ end 3. Count $B^{k+1} = ifft(\overline{B}^{k+1}, [], 3)$

Algorithm 3: Model optimization by ADMM algorithm.

Input: $D, W_F, \lambda, \mu^0, \varepsilon$ Output: B^k, T^k Initialization: $B^0 = T^0 = Y^0 = 0, W_{cw} = 1, W_F^0 = W_{cw} \odot W_{cr}, \mu^0 = 1 \times 10^{-3},$ $\rho = 1.15, k = 0, tol = 10^{-6}$ While $\frac{\|B^{k+1}+T^{k+1}-D\|_F}{\|D\|_F} > tol$ and $\|T^{k+1}\|_0 \neq \|T^k\|_0$ update T^{k+1} by Formula (28); update B^{k+1} by Algorithm 2; update W_F by Formula (19); update W_F by Formula (33); update μ by Formula (34); update k: k = k + 1; end While

The post-processing performs threshold segmentation by setting the threshold as mean $+ v^*$ std, in which mean is the mean value of the whole image. std is the standard deviation, and v is a constant.

3.6. Infrared Dim and Small-Target Detection Algorithm Based on Multi-Frame Spatial–Temporal Patch-Tensor Decomposition

Figure 8 shows the whole process of the proposed method in this work. The overall algorithm flow can be summarized as follows:

- 1. Prior weighted saliency map extraction. The saliency map is obtained by calculating the prior weight between adjacent frames in the sequence by Formula (19);
- 2. Construct a tensor. Through the sliding window of i * i, as shown in Figure 5, the sliders are formed into a three-dimensional tensor $X \in R^{(i*i*z)}$ in order, and z is the number of sliders obtained. Similarly, the above operations are performed on the prior weighted saliency map, acquiring the prior weight $W_F \in R^{(i*i*z)}$;
- 3. The input tensor is decomposed into a low-rank background tensor B and a sparse target tensor T by the ADMM algorithm;
- 4. Tensor reconstruction. Contrary to the construction process, the obtained sparse target tensor and low-rank background tensor are restored and reconstructed, and the overlapping position is sized by a one-dimensional median filter;



5. Image post-processing. The recovered sparse target image is processed by adaptive thresholding to obtain the final target image.

Figure 8. Flowchart of the proposed method. a and b in the picture above are connected to a and b in the picture below.

4. Experiments and Results

In this section, experiments are carried out based on the theory mentioned previously, mainly including the description of data, the influence of different parameters, enhancement degree of the target, detection accuracy, background suppression, and robustness to noise. We give a detailed experimental process and compare the proposed algorithm with eleven state-of-the-art approaches.

4.1. Evaluation Metrics

In order to evaluate the performance of our method, several typical indicators in the field of infrared dim and small-target detection are employed, including:

 Background suppression factor (BSF). The background inhibitory factor is a measure of the prominence of the target and the inhibitory ability of the background. BSF is defined as [66]:

$$BSF = \frac{\delta_{\rm in}}{\delta_{\rm out}} \tag{35}$$

 δ_{in} and δ_{out} represent the standard deviation of the whole background area of the input image and the processed image, respectively.

• Signal-to-clutter ratio gain (SCRG) is a measure of the image before and after processing to suppress the noise and clutter. It is related to signal clutter ratio (SCR). SCR is defined as follows:

$$SCR = \frac{|\mu_t - \mu_b|}{\sigma_b}$$
(36)

 μ_t and μ_b represent the mean value of pixels in the target area and the surrounding background area as shown in Figure 9. σ_b represents the variance in the background neighborhood pixels around the target. SCRG is then defined as:

$$SCRG = \frac{SCR_{out}}{SCR_{in}}$$
(37)

Among them, SCR_{in} represents the SCR of the input image, and SCR_{out} represents the SCR of the processed image.

 The detection probability and false alarm rate are used to measure the performance of the algorithm [36]. Detection probability is defined as:

$$P_d = \frac{DT}{AT} \tag{38}$$

where *DT* represents the number of detected targets, and *AT* represents the number of targets that exist in the image sequence. Moreover, the false alarm rate is described as:

$$F_a = \frac{FP}{NP} \tag{39}$$

where *FP* represents the number of pixels in the false alarm area, and *NP* represents the total number of pixels in the image sequence. Taking the detection probability as the abscissa and false-alarm probability as the ordinate, we can draw the receiver operation characteristic (ROC) curve and calculate the area between the curve and the coordinate axis to obtain the value of area under curve (AUC).

Contrast gain (CG) [67] is used to evaluate the ability to enhance the grayscale contrast
of the target and background. CG is calculated by:

$$CG = \frac{CON_{out}}{CON_{in}}$$
(40)

where CON_{out} is the contrast (CON) of the processed images, and CON_{in} is the contrast of the original images. The definition of the CON is:

$$CON = |\mu_t - \mu_b| \tag{41}$$

where μ_t and μ_b are defined as above. Among these metrics, the larger the SCRG, BSF, CG, and AUC, the better performance of the method is, and it should be noted that SCRG and CG are calculated in the local area, while BSF is computed in the whole image. Because the neighborhoods of the target that we use to evaluate the performance are varied, it is reasonable to consider multiple indicators at the same time.



Figure 9. A partial schematic diagram of the target area and the background area around the target. Assuming that the target size is a * b, then the size of the background area is (a + 2d) * (b + 2d).

4.2. Dataset Description

The dataset is provided by the 25th Research Institute of the Second Research Institute of China Aerospace Science and ATR Key Laboratory of the School of Electronic Sciences, National Defense University of Science and Technology. This dataset consists of several sequences with one or more fixed-wing unmanned aerial vehicles (UAVS) as targets, imaged in various backgrounds such as the sky, ground, etc. We demonstrate some scenes in the dataset and the target positions are marked with red boxes, as shown in Figure 10. To show the target clearly, the target position is enlarged and displayed in the corner of the picture and marked with a red square. It is worth mentioning that the SCR variances in (b) and (i) are quite different. Moreover, we describe five typical scenes of the dataset in Table 2, which include different backgrounds, different targets and different movements.



Figure 10. The (a–o) corresponds to 15 real sequences used in the experiments.

DATA	Total Number of Frames	Average SCR	Data Introduction
DATA 1 (g)	399	6.01	From far to near, single target, ground background
DATA 2 (h)	399	6.29	From near to far, single target, ground background
DATA 3 (n)	399	2.98	Target by far and near, single target, extended target, target maneuver, ground background
DATA 4 (e)	399	1.09	Low signal-to-noise ratio, target from far to near, single target, ground background
DATA 5 (o)	400	3.01	Single target, target maneuver, open space background

Table 2. An introduction to the representative 5 sequences.

4.3. Parameters Analysis

In Figure 11, we present the ROC curves of the results obtained with different parameters in five sequences. Each row represents different parameters in the same sequence and each column represents the results of the same parameter in different sequences.

The settings of various parameters in the model have a great impact on the performances, including running time and environmental robustness. To ensure the accuracy and validity of the experiment, we use the method of controlling variables. By analyzing a parameter within a certain range, we control other variables to remain unchanged at the same time. With various parameters, we can acquire different ROC curves and thus evaluate the performances. Overall, it is worth noting that the results at this part may not be the best. The parameters mainly include the size of the sliding window, the step size of the sliding window, the compromise factor λ , and the penalty factor μ . In addition, the value of the prior weight is determined by a preset experiment, which sets p = 1.5, q = 1, and n = 2.5. This is also determined to select the adjacent two frames before and after and build the tensor model.

The patch size will greatly affect the algorithm. If it is set too large, although the sparsity of the target will be enhanced, some highly disturbing noises will be mistaken for the target, and the motion information between frames will increase, resulting in more non-target components being regarded as the target, which degrades the accuracy. If it is set too small, the sparsity of the target will be greatly reduced, which will lead to more sliding windows. In order to analyze the influence of this parameter, the patch size of the five sequences is set from 20 to 70 in a step of 10. From the ROC curves shown as the first column of Figure 11, it can be seen that too-large and too-small patch size will both lead to inferior results, and the optimal results are obtained when the size is set to 60×60 .

Another parameter is the sliding step. When the step is relatively small, the sparsity of the target will be reduced as well as the utilization of inter-frame information, and the generation of more slices will also increase the time of the algorithm. Conversely, if the step is large, fewer slices will be generated, reducing the complexity of time consumption, but also reducing the redundancy of the tensor. In order to analyze the effect, the sliding step is set from 20 to 70 in a step of 10. At the same time, it should be noted that in order to avoid missing the target, the sliding step cannot be larger than the patch size. It can be seen from the second column of Figure 11 that when the sliding step is set to 60, we can achieve superior performance.

The penalty factor μ also plays a decisive role in the performance of the algorithm. This parameter controls the trade-off between the low-rank component and the sparse component. The background components contained in the target image will be greatly reduced, but the target will be shrunk, and the contrast of the target will be affected by a relatively small μ . Although a large μ can make the target clearer and retain the target

more completely, it will preserve more background residuals in the target image, declining the detection ability. In order to find the best parameters, this paper uses $\mu = 5 \times 10^{-4}$, $\mu = 7 \times 10^{-4}$, $\mu = 9 \times 10^{-4}$, $\mu = 1 \times 10^{-3}$, $\mu = 2 \times 10^{-3}$, $\mu = 3 \times 10^{-3}$, $\mu = 4 \times 10^{-3}$, $\mu = 6 \times 10^{-3}$, $\mu = 1 \times 10^{-2}$ for experimental research. It can be seen from the third column of Figure 11 that when μ is too large or too small, the performance is relatively poor, and when $\mu = 1 \times 10^{-3}$, the low-rank and sparse components can achieve a better trade-off.



Figure 11. The ROC curves formed under different parameters. The first column is the influence of the patch sizes. The second column is the influence of different sliding steps. The third column is the influence of the penalty factor μ . The fourth column is the influence of the trade-off factor λ . Each row represents different parameters in the same sequence.

The compromise parameter λ has a great influence on the performance. When λ is relatively large, in order to maintain the minimization criterion, the background components remaining in the target are largely suppressed, but the contrast of the target is reduced and the target is shrunk. When λ is relatively small, the target contrast will be improved, but the background residual will also become larger. Therefore, it is crucial to choose a suitable λ . Inspired by the predecessors [34,54], we set $\lambda = L/\sqrt{\max(n1, n2) * n3}$, and change λ only by changing the *L* in this work. By setting L to 0.5, 0.7, 1, 1.5, 2, 2.5, and 3.5, respectively, the fourth column of Figure 11 displays different results. We can conclude that when *L* is too large or too small, the ideal experimental results cannot be acquired while L = 1 achieves better performance. Therefore, L = 1 is selected to carry out subsequent experiments.

4.4. Detection Capability for Different Scenarios

As shown in Figure 10, 15 infrared sequences cover different real environments such as sky, ground, etc., as well as the targets with various motions, including moving from near to far, from far to near and so on. Moreover, the distance to the target can be short-range or long-range and the interference includes weak areas and strong areas. In addition to the targets and backgrounds, images with different qualities are also covered via various SCRs. The performance of an algorithm depends on its accuracy and robustness. Therefore, the detection results of the proposed algorithm in Figure 10 are shown in Figure 12. Its 3D display is demonstrated in Figure 13. The position of the target is marked with a red box. For clear demonstration, the result is enlarged and placed in the corner of the image, and the background residual is marked with a green ellipse. Two facts can be seen from our detection results: (1) The target is detected accurately in the 15 sequences; (2) the background residual is almost completely suppressed, and even if it is present, it is very weak, such as in Figure 12l,m.



Figure 12. Targets detected in 15 different sequences by the proposed method. (**a**–**o**) represent the corresponding scenes in Figure 10.



Figure 13. Three-dimensional displays of the detection results in Figure 12.

To illustrate the superiority of the proposed algorithm, two methods of recent years are compared for 15 sequences. Figure 14 shows the detection results of PSTNN, and Figure 15 shows the corresponding 3d displays. Figure 16 shows the detection results of LogTFNN, and Figure 17 shows the corresponding 3D displays. It can be seen from the results that when facing dim and small targets in complex backgrounds such as (c), (d), (g), (j), (n) in Figure 10, the detection results of these two methods are not very good, according to (c), (d), (j), (g), (n) in Figure 14, and (c), (d), (j), (g), (n) in Figure 16. From the comparisons, we can thereby believe that inaccurate selection of the prior weight will highlight the target and interference at the same time, as shown in Figure 16h,o, resulting in the existence of background residuals. In addition, the useful information of the inter-frame is ignored so that PSTNN and LogTFNN can only adapt to the simple scenes and the detection accuracy is decreased when the background is complex. However, the proposed algorithm in this paper can be applied in various complex backgrounds and the performance is competitive.



Figure 14. Targets detected in 15 different sequences by using PSTNN. (**a**–**o**) represent the corresponding scenes in Figure 10.



Figure 15. Three-dimensional displays of the detection results in Figure 14.



Figure 16. Targets detected in 15 different sequences by using LogTFNN. (**a**–**o**) represent the corresponding scenes in Figure 10.



Figure 17. Three-dimensional displays of the detection results in Figure 16.

Robustness to Noise

Another indicator to evaluate the performance of the algorithm is the robustness to noise. The 15 sequences have added noise with a mean value of 0 and a variance of 0.005 to obtain the data, as shown in Figure 18. The corresponding detection results are shown in Figure 19. It can be clearly seen that the signal-to-noise ratio (SNR) of the image is lower after adding the noise, and the contrast of the target is also affected, as shown in Figure 18e, f,h. Some targets are even submerged in the background, such as Figure 18d. Nevertheless, the proposed algorithm still can extract the targets accurately, except for (d), although the target shape is affected. The data obtained by assigning Gaussian noise with a mean value of 0 and a variance of 0.01 are shown in Figure 20. It can be seen that the targets of all images have become ambiguous, and the contrast of the targets has been greatly declined. The corresponding results are depicted in Figure 21, and we can see that the targets in (d) and (j) are missing. The targets detected in (h) are not clear, while the remaining images are relatively accurately detected. Although the shape does not exactly match the real shape of the target, the results are acceptable for the problem of infrared dim and small target detection. These results also illustrate that although the noise is added, the targets still can be extracted accurately in complex scenes.

4.5. Comparison with Other Typical Methods

In order to clearly show the performance of the algorithm proposed in this paper, we compare our approach with other eleven typical methods, and the obtained results are shown in Figures 22–26. Eleven state-of-the-art methods are utilized for comparison, and the experimental parameter settings of the methods are shown in Table 3. The eleven methods include Top-Hat among the methods based on the background consistency assumption. Among the HVS-based methods, LCM, HB-MLCM with better performance and TLLCM in the past three years have been included. The most classic methods based on low-rank sparse decomposition are matrix-based IPI and tensor-based RIPT. We also select improved NRAM based on IPI. PSTNN, NOLC, ECA-STT, and LogTFNN improved based on RIPT in the past three years are utilized as well.



Figure 18. The original images added noise with mean 0 and variance 0.005. (**a–o**) represent the corresponding scenes in Figure 10.



Figure 19. The detection results of the original images' added noise with mean 0 and variance 0.005. **(a–o)** represent the corresponding scenes in Figure 18.



Figure 20. The original images' added noise with mean 0 and variance 0.01. (**a–o**) represent the corresponding scenes in Figure 10.



Figure 21. The detection results of the original images' added noise with mean 0 and variance 0.01. **(a–o)** represent the corresponding scenes in Figure 20.

Methods	Parameter Settings
Top-Hat [9]	Structure size: square, size: 3×3
LCM [12]	Window size: 3×3
HB-MLCM [16]	Window size: 3×3 , 5×5 , 7×7 , 9×9
TLLCM [18]	Window size: 3×3
IPI [21]	patch size: 50 × 50, sliding step: $10, \lambda = 1/\sqrt{\min(m, n)}, \varepsilon = 10^{-7}$
RIPT [30]	patch size: 30×30 , sliding step: $10, \lambda = 1/\sqrt{\min(m, n)}, \varepsilon = 10^{-7}, h = 1$
NRAM [24]	patch size: 40×40 , step : 10 , $\lambda = 1/\sqrt{\min(m, n)}$, $\mu_0 =$
	$3 imes \sqrt{\min(m,n)}, \gamma = 0.002, C = 3/\sqrt{\min(m,n)}, \varepsilon = 10^{-7}$
PSTNN [34]	patch size: 40 × 40, step: 40, $\lambda = 1/\sqrt{\min(m, n)}$, $\varepsilon = -7$
NOLC [25]	patch size: 30×30 , Slide step: $10, \lambda = 1/\sqrt{\max(size(img))}, p = 0.6$
ECA-STT [35]	$\beta = 0.1, t = 3, \lambda_1 = 0.009, \lambda_2 = 5.0 / \sqrt{\min(m, n) \times t}, \lambda_3 = 100, \varepsilon = 10^{-7}$
LOGTNN [37]	patch size: 40×40 , step: 40,
	$\lambda = 0.4 / \sqrt{\max(n1, n2) * n3}, \beta = 0.05, \mu = 200$
Proposed	patch size: 60 × 60, step: 60, $\lambda = 1/\sqrt{\max(n1, n2) * n3}$

Table 3. Parameters of eleven state-of-the-art methods.

It can be seen from the results that the traditional TOPHAT method can highlight the target to a certain extent. However, the strong interference part cannot be suppressed, resulting in a poor final result. Although HVS-based methods such as LCM and HB-MLCM do highlight the target, the background is also enhanced, and the shape of the target is completely lost. After adding three layers of filter windows and improving the local contrast, the results obtained by TLLCM have good results. Compared with the low-rank sparse representation method, the performances of HVS-based methods are relatively low, which is mainly due to the working principle that makes the methods sensitive to the background, and the filter unit of a specific structure has very poor robustness to clutter. For the earliest proposed IPI algorithm, except the first sequence without background residuals, the backgrounds in all other cases are not approximated. Compared with the IPI algorithm, NRAM has a more accurate approximation of the rank. It can separate the low-rank and sparse components more accurately, and many background clutters are removed with strong constraints of the edges. However, for the influence of other categories, the detection performance of this method still needs to be improved, as shown in Figure 26. Similarly, the L_p norm selected by NOLC to approximate the rank of the matrix, compared with the sum of the general matrix kernel norm, can more accurately obtain the sparse components and the low-rank components, but the performance is inferior in the face of complex backgrounds, as shown in Figures 23 and 25. This is due to the way in which the image is unfolded, which destroys the correlation of information in the image. Compared with other types of methods, tensor-based algorithms have shown great advantages, such as RIPT, PSTNN, ECA-STT and LogTFNN. These algorithms can effectively use the information in the images to separate the target components. However, the simple use of single-frame information and inaccurate calculation of rank result in low robustness to the complex and changeable backgrounds, which still can be improved, as shown in the parts marked by green ellipses in Figures 23, 24 and 26. Compared with all other methods, the proposed approach can accurately detect the targets and suppress the background residuals to the greatest extent, which means our method achieves superior performance both in detection accuracy and the robustness of complex scenes.



Figure 22. The detection results of data 1.



Figure 23. The detection results of data 2.



Figure 24. The detection results of data 3.



Figure 25. The detection results of data 4.



Figure 26. The detection results of data 5.

For the convenience of the presentation, this paper only releases the detection results of 12 detection methods in these 5 sequences. In general, the detection performance of the method based on the background consistency assumption is the worst. The method based on HVS is the second, and the method based on the low-rank sparse model is the best. Among the low-rank sparse models, the method of obtaining the matrix model is inferior to the tensor model. Furthermore, the prior information saliency map adopted in this research can highlight the target information more accurately and suppress the background better. At the same time, the rank approximation by Laplace operator is closer to the real rank, and can more accurately decompose the image. However, the superior performance of the proposed method sacrifices the consuming time.

4.6. Quantitative Comparison of Eleven Methods

In previous sections, the proposed method is qualitatively compared with other methods, and the superiority of this algorithm can be acquired. For the sake of scientific rationality, the comparison results are quantitatively analyzed in this section. Seq1 to seq5 are data1 to data5 listed in Table 2. The measurement results of different indicators in sequences 1 to 5 are listed in Tables 4 and 5. As mentioned before, the larger the evaluation indicators in the table, the better performance the algorithm achieves. The best results are marked in red for each metric, and the second best results are marked in green. The inf in BSF means infinite; that is, in the target image, except for the area to be detected, the other background areas are eliminated.

It can be clearly seen that the proposed method has achieved very significant advantages in terms of BSF, which proves that the proposed method has the strongest ability to suppress the background. In the first sequence, the PSTNN method achieves the second best performance, which illustrates the rationality of using the sum of tensor partial nuclear norm to approximate the rank of the tensor. Due to the utilization of information between multi-frame images, the brightness of the separated target will be affected because the targets move slowly and the patches have many overlapping pixels. Therefore, the SCRG in sequences 4 and 5 are not optimal, but it also achieves competitive results. The CG in sequence 1, 2, 4, and 5 all reached the maximum, except for sequence 4. It is also proved that the proposed method achieves superior performance.

Methods	BSF	Seq.1 SCRG	CG	BSF	Seq.2 SCRG	CG	BSF	Seq.3 SCRG	CG
TOPHAT	1.161	2.070	36.609	1.766	1.664	0.741	2.866	1.496	11.931
LCM	0.910	1.755	0.924	0.962	0.404	26.159	1.818	1.226	7.829
HB-MLCM	4.437	4.577	34.967	16.089	2.192	7.018	56.681	1.472	12.021
TLLCM	6.214	1.493	3.617	22.339	2.022	8.461	28.455	1.539	9.364
IPI	8.736	3.352	20.871	11.527	2.272	28.227	19.255	1.699	21.354
RIPT	7.452	4.714	89.268	11.124	2.537	56.258	12.982	2.343	18.282
NRAM	15.920	5.045	82.644	23.164	2.230	60.592	29.559	2.176	54.072
PSTNN	16.424	5.056	98.977	44.915	2.315	65.115	26.359	2.594	39.214
NOLC	11.942	4.970	81.366	16.596	2.457	57.563	18.563	2.116	27.737
ECA-STT	13.355	4.359	54.090	34.858	1.999	9.346	46.425	1.550	7.126
LogTFNN	8.961	4.395	66.518	16.596	2.117	52.880	32.031	2.160	69.964
Proposed	33.017	5.207	101.415	inf	3.156	72.488	134.267	2.262	64.312

Table 4. Measurements for twelve detection methods.

Table 5. Measurements for twelve detection methods.

Mathada		Seq.4			Seq.5	
Methods	BSF	SCRG	CG	BSF	SCRG	CG
TOPHAT	2.026	1.806	26.217	3.310	1.544	7.131
LCM	1.473	0.775	15.414	1.630	0.546	21.853
HB-MLCM	36.361	2.177	1.016	10.049	1.887	15.365
TLLCM	30.425	2.240	6.940	27.922	1.995	23.396
IPI	26.342	2.287	0.647	14.562	1.693	4.035
RIPT	17.980	2.712	21.021	15.258	1.594	31.900
NRAM	68.585	2.828	48.054	22.962	1.791	38.960
PSTNN	inf	3.304	78.458	24.206	1.956	45.227
NOLC	26.973	2.665	56.447	18.036	2.561	34.274
ECA-STT	136.212	2.382	19.495	30.791	1.968	34.779
LogTFNN	inf	2.482	72.915	15.693	1.659	54.820
Proposed	inf	2.981	78.793	134.572	1.994	55.687

In addition, the proposed method is compared with eleven other methods, and the ROC curves of twelve methods in five sequences are obtained, which are placed in Figure 27, and the values of the AUC are shown in Table 6.

It can be seen from the ROC curves and the corresponding AUC values that the proposed method achieves the best performance. The anti-jamming abilities of the LCM and HB-MLCM methods are very poor, and the results have large fluctuations. They use filters, which find it particularly difficult to deal with objects submerged in the background, such as sequence 1 and 5. When there is a very strong interference in the background, such as in sequence 3, the performances of the other methods are inferior, while our algorithm can accurately detect the target. It can be concluded from the data analysis that under the same false-alarm rate, the proposed method has the highest probability of detection, and under the same accuracy, the proposed method has the lowest false-alarm rate.



Figure 27. ROC curves of the compared and proposed methods in different sequences.

Methods	Seq.1	Seq.2	Seq.3	Seq.4	Seq.5
TOPHAT	0.0576	0.6162	0.1894	0.5597	0.6216
LCM	0.0179	0.0372	0.0162	0.0048	0.0280
HB-MLCM	0.0969	0.9012	0.2360	0.9709	0.3513
TLLCM	0.0502	0.7259	0.9241	0.9839	0.9006
IPI	0.9687	0.7817	0.4418	0.7609	0.8576
RIPT	0.9668	0.7518	0.2290	0.5082	0.6450
NRAM	0.9596	0.6091	0.2471	0.7530	0.8658
PSTNN	0.9818	0.7539	0.4552	0.7858	0.9021
NOLC	0.9678	0.7552	0.4214	0.7650	0.7487
ECA-STT	0.8631	0.6031	0.2883	0.8489	0.8545
LogTFNN	0.6373	0.6519	0.6214	0.7247	0.6299
Proposed	0.9828	0.9092	0.9605	0.9970	0.9999

Table 6. AUC of different methods.

4.7. Computation Time

Table 7 summarizes the time consumption of each method. Among the algorithms, the filtering method takes the least time. NOLC is the least time-consuming among the matrix-based methods, and PSTNN is the least time-consuming among the tensor-based methods. When the size of the image to be processed is fixed, no matter how the complexity of the background in the image changes, the time spent in the presented methods will not fluctuate too much. The time spent by the proposed method is higher than other methods, mainly caused by the tensor construction. The time cost can be reduced by multi-threading, which should be improved in the future.

4.8. Validation Analysis

To enhance the persuasiveness of the superiority of the proposed algorithm, two sets of comparative experiments are added. Since the parameters used in the experiments are obtained by simulation analysis in the five sequences mentioned above, to enhance the effectiveness of the algorithm and fully verify its robustness, the other two sequences are detected with the same parameters. The obtained ROC curves are shown in Figure 28, and the corresponding AUC value is shown in Table 8. Through the results, it can be con-

cluded that the same parameters perform very well in the verified sequences, and the AUC values measured by the filtering method are very small, which is also in line with the experimental expectations, while the volatility of LogTFNN is relatively large. The performance of other tensor-based methods is at a moderate level. For the relatively easy-to-detect sequence 7, most of the methods have relatively large AUC values. In this case, the proposed method can still achieve the maximum value, which proves the efficiency and robustness of the algorithm.

Methods	Seq.1	Seq.2	Seq.3	Seq.4	Seq.5
ТОРНАТ	0.0137	0.6162	0.1894	0.5597	0.6216
LCM	0.0636	0.0545	0.0463	0.0452	0.0491
HB-MLCM	0.0172	0.0174	0.0145	0.0157	0.0155
TLLCM	1.0336	1.0293	1.0761	1.0164	0.9962
IPI	5.5765	6.1107	6.3536	6.0694	6.1535
RIPT	1.6991	1.5782	1.5513	1.5445	1.5946
NRAM	8.2516	7.7308	7.4015	7.4239	8.4935
PSTNN	0.0937	0.1420	0.0943	0.0935	0.1005
NOLC	4.8856	5.225	2.8089	2.5513	6.3097
ECA-STT	3.0169	3.1745	2.9984	3.0237	3.0713
LogTFNN	4.5226	4.6316	4.5990	4.5665	4.4655
Proposed	17.7711	17.8345	16.0337	13.7015	14.6458

Table 7. Computation time of different methods.



Figure 28. ROC curves of detection results in validation data.

 Table 8. AUC of different methods.

Methods	Seq.6	Seq.7
ТОРНАТ	0.8402	0.6381
LCM	0	0.2680
HB-MLCM	0.0528	0.4581
TLLCM	0.3354	0.9694
IPI	0.9706	0.9971
RIPT	0.9636	0.9974
NRAM	0.9787	0.9974
PSTNN	0.9734	0.9967
NOLC	0.9694	0.9974
ECA-STT	0.9148	0.9685
LogTFNN	0.3695	0.9838
Proposed	0.9928	0.9998

5. Discussion

With the development of modern military technology, infrared dim and small-target detection technology has received more and more attention. The current detection methods

have been greatly developed, but there is still a lot of space for further development. The detection method based on the background consistency assumption is fast, while the detection accuracy and robustness are poor. The method based on HVS requires less prior information, and the accuracy is improved as well. However, when the background is complex and changeable with submerged targets, the obtained detection results are unsatisfactory. The LRSD-based approach improves the infrared dim and small-target detection significantly. However, the earliest IPI model ignores the structural information between pixels, thus leading to the enhancement by the IPT model, which introduces the concept of a tensor that can make full use of the information among the image patches.

In the IPT model, most of the detection methods are based on the single frame, ignoring the useful information between frames, especially when the target and background are moving. In addition, in the traditional IPT model, such as inaccurate characterization of the background rank, the problem of poor robustness still exists. To solve the existing problem, we propose MFSTPT. Firstly, by modifying the tensor construction, the model that conforms to the assumption of low-rank background is obtained. In addition, to reduce the interference of noise and sparse edges, the prior saliency map is obtained by a novel weighted method. Finally, the rank of the tensor is represented by the Laplace rank, which can more accurately approximate the background. The model is optimized and solved by ADMM. We use TOPHAT, LCM, HB-MLCM, TLLCM, IPI, RIPT, NRAM, PSTNN, NOLC, ECA-STT and LOGTFNN as the comparison algorithms. It can be seen from Figures 22–26 and Figure 12 that the proposed method can detect the target more accurately while suppressing the background to the maximum, qualitatively. Figures 19 and 21 conclude that the proposed method is robust to noise. In the quantitative analysis, the SCRG, BSF, and CG in Tables 4 and 5 conclude that the proposed method can well highlight the target and suppress the background clutter. In Figure 27 and Table 6, the ROC curve of the proposed method performs better, and the AUC is closer to 1. Figure 28 and Table 8 further verify the superiority of the proposed method. In conclusion, the proposed method can highlight and detect objects more accurately while suppressing the background better. However, the trade-off between performance and time consumption also needs to be improved with the development.

6. Conclusions

In order to solve the existing problems of infrared small-target detection, especially the detection of moving targets in complex backgrounds, an improved IPT model, MFSTPT, is proposed. For measuring low-rank properties of the background, the Laplace function based on non-convex approximation is utilized. To suppress the interference of the edges and highlighted areas, novel weighted prior information is introduced. Moreover, a new tensor model is constructed by the simultaneous sliding window in space and time to satisfy the assumption of low-rank background. The established model is then solved by the ADMM algorithm, and experiments are carried out in different sequences. According to the results, it can be concluded that, except for the expense of consuming time, the detection accuracy and background suppression have been greatly improved, as well as the robustness to the environment.

Author Contributions: Y.H. and Y.M. proposed the original idea. Y.M. performed the experiments and wrote the manuscript. Z.P. reviewed and edited the manuscript. Y.L. contributed to the direction, content, and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

- Luo, K. Space-Based Infrared Sensor Scheduling with High Uncertainty: Issues and Challenges. Syst. Eng. 2015, 18, 102–113. [CrossRef]
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- 3. Peng, Z.; Zhang, Q.; Guan, A. Extended target tracking using projection curves and matching pel count. *Opt. Eng.* 2007, 46, 066401.
- 4. Kennedy, H.L. Multidimensional digital filters for point-target detection in cluttered infrared scenes. *J. Electron. Imaging* **2014**, 23, 063019. [CrossRef]
- 5. Grossi, E.; Lops, M.; Venturino, L. A novel dynamic programming algorithm for track-before-detect in radar systems. *IEEE Trans. Signal Process.* **2013**, *61*, 2608–2619. [CrossRef]
- Li, Y.; Zhang, Y. Robust infrared small target detection using local steering kernel reconstruction. *Pattern Recognit.* 2018, 77, 113–125. [CrossRef]
- 7. Dong, X.; Huang, X.; Zheng, Y.; Bai, S.; Xu, W. A novel infrared small moving target detection method based on tracking interest points under complicated background. *Infrared Phys. Technol.* **2014**, *65*, 36–42. [CrossRef]
- 8. Rawat, S.S.; Verma, S.K.; Kumar, Y. Review on recent development in infrared small target detection algorithms. *Procedia Comput. Sci.* 2020, *167*, 2496–2505. [CrossRef]
- 9. Gu, Y.; Wang, C.; Liu, B.; Zhang, Y. A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 469–473. [CrossRef]
- 10. Reed, I.S.; Gagliardi, R.M.; Stotts, L.B. Optical moving target detection with 3-D matched filtering. *IEEE Trans. Aerosp. Electron. Syst.* **1988**, 24, 327–336. [CrossRef]
- 11. Dong, X.; Huang, X.; Zheng, Y.; Shen, L.; Bai, S. Infrared dim and small target detecting and tracking method inspired by human visual system. *Infrared Phys. Technol.* **2014**, *62*, 100–109. [CrossRef]
- 12. Chen, C.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]
- 13. 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.
- 14. Wei, Y.; You, X.; Li, H. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recognit.* **2016**, *58*, 216–226. [CrossRef]
- 15. Bai, X.; Bi, Y. Derivative entropy-based contrast measure for infrared small-target detection. *IEEE Trans. Geosci. Remote Sens.* **2018**, 56, 2452–2466. [CrossRef]
- 16. Shi, Y.; Wei, Y.; Yao, H.; Pan, D.; Xiao, G. High-boost-based multiscale local contrast measure for infrared small target detection. *IEEE Geosci. Remote Sens. Lett.* **2017**, *15*, 33–37. [CrossRef]
- 17. Lu, R.; Yang, X.; Li, W.; Fan, J.; Li, D.; Jing, X. Robust infrared small target detection via multidirectional derivative-based weighted contrast measure. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 7000105. [CrossRef]
- 18. Han, J.; Moradi, S.; Faramarzi, I.; Liu, C.; Zhang, H.; Zhao, Q. A local contrast method for infrared small-target detection utilizing a tri-layer window. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1822–1826. [CrossRef]
- 19. Zhao, M.; Li, L.; Li, W.; Tao, R.; Li, L.; Zhang, W. Infrared small-target detection based on multiple morphological profiles. *IEEE Trans. Geosci. Remote Sens.* 2020, 59, 6077–6091. [CrossRef]
- Zhang, C.; Li, D.; Qi, J.; Liu, J.; Wang, Y. Infrared Small Target Detection Method with Trajectory Correction Fuze Based on Infrared Image Sensor. Sensors 2021, 21, 4522. [CrossRef]
- 21. 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]
- Dai, Y.; Wu, Y.; 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. 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]
- 24. Zhang, L.; Peng, L.; Zhang, T.; Cao, S.; Peng, Z. Infrared small target detection via non-convex rank approximation minimization joint l2, 1 norm. *Remote Sens.* 2018, 10, 1821. [CrossRef]
- 25. Zhang, T.; Wu, H.; Liu, Y.; Peng, L.; Yang, C.; Peng, Z. Infrared small target detection based on non-convex optimization with Lp-norm constraint. *Remote Sens.* **2019**, *11*, 559. [CrossRef]
- Wang, X.; Peng, Z.; Kong, D.; Zhang, P.; He, Y. Infrared dim target detection based on total variation regularization and principal component pursuit. *Image Vis. Comput.* 2017, 63, 1–9. [CrossRef]
- 27. Rawat, S.S.; Alghamdi, S.; Kumar, G.; Alotaibi, Y.; Khalaf, O.I.; Verma, L.P. Infrared small target detection based on partial sum minimization and total variation. *Mathematics* **2022**, *10*, 671. [CrossRef]
- Zhang, T.; Peng, Z.; Wu, H.; He, Y.; Li, C.; Yang, C. Infrared small target detection via self-regularized weighted sparse model. *Neurocomputing* 2021, 420, 124–148. [CrossRef]
- 29. Wang, X.; Peng, Z.; Kong, D.; He, Y. Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scene. *IEEE Trans. Geosci. Remote Sens.* 2017, 55, 5481–5493. [CrossRef]

- 30. Dai, Y.; Wu, Y. 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]
- Goldfarb, D.; Qin, Z. Robust low-rank tensor recovery: Models and algorithms. SIAM J. Matrix Anal. Appl. 2014, 35, 225–253. [CrossRef]
- Wu, Z.; Wang, Q.; Jin, J.; Shen, Y. Structure tensor total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising. *Signal Process.* 2017, 131, 202–219. [CrossRef]
- Zhang, X.; Ding, Q.; Luo, H.; Hui, B.; Chang, Z.; Zhang, J. Infrared small target detection based on an image-patch tensor model. *Infrared Phys. Technol.* 2019, 99, 55–63. [CrossRef]
- Zhang, L.; Peng, Z. Infrared small target detection based on partial sum of the tensor nuclear norm. *Remote Sens.* 2019, 11, 382. [CrossRef]
- Zhang, P.; Zhang, L.; Wang, X.; Shen, F.; Pu, T.; Fei, C. Edge and Corner Awareness-Based Spatial–Temporal Tensor Model for Infrared Small-Target Detection. *IEEE Trans. Geosci. Remote Sens.* 2020, 59, 10708–10724. [CrossRef]
- Liu, T.; Yang, J.; Li, B.; Xiao, C.; Sun, Y.; Wang, Y.; An, W. Non-Convex Tensor Low-Rank Approximation for Infrared Small Target Detection. *arXiv* 2021, arXiv:2105.14974.
- Kong, X.; Yang, C.; Cao, S.; Li, C.; Peng, Z. Infrared Small Target Detection via Nonconvex Tensor Fibered Rank Approximation. *IEEE Trans. Geosci. Remote Sens.* 2021, 60, 1–21. [CrossRef]
- Dai, Y.; Wu, Y.; Zhou, F.; Barnard, K. Attentional local contrast networks for infrared small target detection. *IEEE Trans. Geosci. Remote Sens.* 2021, 59, 9813–9824. [CrossRef]
- Wang, H.; Zhou, L.; Wang, L. Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 8509–8518.
- 40. Qi, G.; Zhang, Y.; Wang, K.; Mazur, N.; Liu, Y.; Malaviya, D. Small Object Detection Method Based on Adaptive Spatial Parallel Convolution and Fast Multi-Scale Fusion. *Remote Sens.* **2022**, *14*, 420. [CrossRef]
- Dai, Y.; Wu, Y.; Zhou, F.; Barnard, K. Asymmetric contextual modulation for infrared small target detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 950–959.
- 42. Pang, D.; Shan, T.; Ma, P.; Li, W.; Liu, S.; Tao, R. A novel spatiotemporal saliency method for low-altitude slow small infrared target detection. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
- 43. Kilmer, M.E.; Martin, C.D. Factorization strategies for third-order tensors. Linear Algebra Its Appl. 2011, 435, 641–658. [CrossRef]
- 44. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 925–938. [CrossRef] [PubMed]
- 45. Martin, C.D.; Shafer, R.; LaRue, B. An order-p tensor factorization with applications in imaging. *SIAM J. Sci. Comput.* **2013**, 35, A474–A490. [CrossRef]
- Liu, H.K.; Zhang, L.; Huang, H. Small target detection in infrared videos based on spatio-temporal tensor model. *IEEE Trans. Geosci. Remote Sens.* 2020, 58, 8689–8700. [CrossRef]
- Fang, H.; Chen, M.; Liu, X.; Yao, S. Infrared Small Target Detection with Total Variation and Reweighted Regularization. *Math. Probl. Eng.* 2020, 2020, 1529704. [CrossRef]
- 48. Sun, Y.; Yang, J.; Long, Y.; Shang, Z.; An, W. Infrared patch-tensor model with weighted tensor nuclear norm for small target detection in a single frame. *IEEE Access* **2018**, *6*, 76140–76152. [CrossRef]
- 49. Zhou, F.; Wu, Y.; Dai, Y. Infrared small target detection via incorporating spatial structural prior into intrinsic tensor sparsity regularization. *Digit. Signal Process.* **2021**, 111, 102966. [CrossRef]
- 50. Gao, C.Q.; Tian, J.W.; Wang, P. Generalised-structure-tensor-based infrared small target detection. *Electron. Lett.* 2008, 44, 1349–1351. [CrossRef]
- Brown, M.; Szeliski, R.; Winder, S. Multi-image matching using multi-scale oriented patches. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 510–517.
- 52. Liu, J.; He, Z.; Chen, Z.; Shao, L. Tiny and dim infrared target detection based on weighted local contrast. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1780–1784. [CrossRef]
- Nie, C.; Wang, H.; Lu, A. Infrared Small Target Detection Based on Prior Constraint Network and Efficient Patch-Tensor Model. In *Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 504–517.
- 54. Zhou, F.; Wu, Y.; Dai, Y.; 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]
- Candes, E.J.; Wakin, M.B.; Boyd, S.P. Enhancing sparsity by reweighted L₁ minimization. J. Fourier Anal. Appl. 2008, 14, 877–905. [CrossRef]
- Wang, H.; Yang, F.; Zhang, C.; Ren, M. Infrared small target detection based on patch image model with local and global analysis. *Int. J. Image Graph.* 2018, 18, 1850002. [CrossRef]
- Zhu, H.; Ni, H.; Liu, S.; Xu, G.; Deng, L. Tnlrs: Target-aware non-local low-rank modeling with saliency filtering regularization for infrared small target detection. *IEEE Trans. Image Process.* 2020, *29*, 9546–9558. [CrossRef] [PubMed]

- Nie, Y.; Li, W.; Zhao, M.; Ran, Q.; Ma, P. Infrared small target detection in image sequences based on temporal low-rank and sparse decomposition. In *Proceedings of the Twelfth International Conference on Graphics and Image Processing (ICGIP 2020)*; International Society for Optics and Photonics: Bellingham, WA, USA, 2021; Volume 11720, p. 117200A.
- Xu, W.H.; Zhao, X.L.; Ji, T.Y.; Miao, J.Q.; Ma, T.H.; Wang, S.; Huang, T.Z. Laplace function based nonconvex surrogate for low-rank tensor completion. *Signal Process. Image Commun.* 2019, 73, 62–69. [CrossRef]
- Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; Kilmer, M. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3842–3849.
- Guan, X.; Zhang, L.; Huang, S.; Peng, Z. Infrared small target detection via non-convex tensor rank surrogate joint local contrast energy. *Remote Sens.* 2020, 12, 1520. [CrossRef]
- Zhou, F.; Wu, Y.; Dai, Y.; Wang, P.; Ni, K. Graph-regularized laplace approximation for detecting small infrared target against complex backgrounds. *IEEE Access* 2019, 7, 85354–85371. [CrossRef]
- 63. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends. Mach. Learn.* **2011**, *3*, 1–122. [CrossRef]
- Hale, E.T.; Yin, W.; Zhang, Y. Fixed-point continuation for ℓ₁-minimization: Methodology and convergence. *SIAM J. Optim.* 2008, 19, 1107–1130. [CrossRef]
- Lu, C.; Tang, J.; Yan, S.; Lin, Z. Generalized nonconvex nonsmooth low-rank minimization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 4130–4137.
- 66. Guan, X.; Peng, Z.; Huang, S.; Chen, Y. Gaussian scale-space enhanced local contrast measure for small infrared target detection. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 327–331. [CrossRef]
- Gao, C.; Wang, L.; Xiao, Y.; Zhao, Q.; Meng, D. Infrared small-dim target detection based on Markov random field guided noise modeling. *Pattern Recognit.* 2018, 76, 463–475. [CrossRef]