*remote sensing*

MDPI

*Article*

# Infrared Dim and Small Target Detection Based on Superpixel Segmentation and Spatiotemporal Cluster 4D Fully-Connected Tensor Network Decomposition

**Wenyan Wei [1,2,3,4], Tao Ma [1,2,3,4], Meihui Li [1,2,3] and Haorui Zuo [1,2,3,*]**

[1] National Key Laboratory of Optical Field Manipulation Science and Technology, Chinese Academy of Sciences, Chengdu 610209, China; limeihui@ioe.ac.cn (M.L.)
[2] Key Laboratory of Optical Engineering, Chinese Academy of Sciences, Chengdu 610209, China
[3] Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China
[4] School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100149, China
* Correspondence: zuohaorui@ioe.ac.cn

**Abstract:** The detection of infrared dim and small targets in complex backgrounds is very challenging because of the low signal-to-noise ratio of targets and the drastic change in background. Low-rank sparse decomposition based on the structural characteristics of infrared images has attracted the attention of many scholars because of its good interpretability. In order to improve the sensitivity of sliding window size, insufficient utilization of time series information, and inaccurate tensor rank in existing methods, a four-dimensional tensor model based on superpixel segmentation and statistical clustering is proposed for infrared dim and small target detection (ISTD). First, the idea of superpixel segmentation is proposed to eliminate the dependence of the algorithm on selection of the sliding window size. Second, based on the improved structure tensor theory, the image pixels are statistically clustered into three types: corner region, flat region, and edge region, and are assigned different weights to reduce the influence of background edges. Next, in order to better use spatiotemporal correlation, a Four−Dimensional Fully-Connected Tensor Network(4D−FCTN) model is proposed in which 3D patches with the same feature types are rearranged into the same group to form a four-dimensional tensor. Finally, the FCTN decomposition method is used to decompose the clustered tensor into low-dimensional tensors, with the alternating direction multiplier method (ADMM) used to decompose the low-rank background part and sparse target part. We validated our model across various datasets, employing cutting-edge methodologies to assess its effectiveness in terms of detection precision and reduction of background interference. A comparative analysis corroborated the superiority of our proposed approach over prevailing techniques.

**Keywords:** infrared small target detection (ISTD); spatio-temporal tensor; sparse and low-rank decomposition; superpixel segmentation; fully-connected tensor network decomposition (FCTN); cluster 4D−FCTN method

check for updates

**Citation:** Wei, W.; Ma, T.; Li, M.; Zuo, H. Infrared Dim and Small Target Detection Based on Superpixel Segmentation and Spatiotemporal Cluster 4D Fully-Connected Tensor Network Decomposition. *Remote Sens.* **2024**, *16*, 34. https://doi.org/10.3390/rs16010034

Academic Editor: Lionel Bombrun

Received: 1 November 2023
Revised: 15 December 2023
Accepted: 16 December 2023
Published: 21 December 2023

**Copyright:** © 2023 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/).

## 1. Introduction

The detection of small infrared targets under complex backgrounds plays an essential part in Infrared Search and Track (IRST) systems [1]. However, infrared dim and small targets have weak brightness, small size (according to research convention, generally no more than 80 pixels), and lack obvious shape, texture, and colour information, making them difficult to detect directly [2]. In addition, false alarms are a difficult problem to solve in real scenarios [3,4]. Therefore, an infrared small target detection algorithm has essential research significance and value. This section briefly discusses the existing infrared dim and small target detection methods [5], where the method based on low-rank sparse decomposition is introduced in detail [6].

In recent years, many approaches for Infrared Small Target Detection (ISTD) have emerged, with a specific emphasis on single frame-based and sequence-based techniques. Section 1.1 provides a concise overview of the methods based on individual frames, followed by an exploration of sequence-based methods in Section 1.2. Finally, Section 1.3 delves into the approaches centered around low-rank sparse decomposition.

### *1.1. Single-Frame Detection Methods*

The existing single-frame algorithms can be divided into two categories: local information and nonlocal information. According to the idea of local information single-frame algorithms, the gray level of the background pixel is usually close to that of its local neighborhood pixels, while the gray level of the target pixel is different from that of its local neighborhood pixels. By extracting the difference information between each pixel in the image and its neighborhood reference pixels, the target can be successfully selected. Such algorithms mainly include background estimation methods [7,8], morphological methods [9,10], directional derivative/gradient methods [11,12], and local contrast methods [13]. However, using local information to detect the target must satisfy a basic premise, that is, the target must be the most remarkable in the local area, which may not hold in the complex real background. A very possible situation is that if the background is very complex, the target may be close to some extremely bright background and easily submerged by it. In this case, it is very difficult to use local information to detect the target [14]. This leads to the problem that the detection performance of local information algorithms decreases in complex real backgrounds.

Algorithms based on non-local single-frame information use the differences between the target image and the background/noise image in terms of the frequency band, data space, and other aspects. Such algorithms use the frequency domain method [15,16], classifier method [17,18], over-complete sparse representation method [19,20], and sparse low-rank decomposition method to separate the target image from the whole original image. Among these, the standard classifier methods include clustering, support vector machines, neural networks, etc. Zhou Weina [21] selected radial basis neural networks and nonlinear autoregressive networks to realize infrared target detection of single-frame images. In [22], the authors used the YOLO (You Only Look Once) depth neural network architecture to process infrared and visible dual-channel images to detect fast low-altitude UAV targets. Based on deep learning theory, Hou Qingyu et al. [23] proposed a convolution neural network called RISTDnet. In [24], Dai Yimian et al. made the LCM algorithm into a module and then embedded it into the whole convolution divine sutra network. Because these algorithms directly calculate the whole image and do not care about the saliency of the target in the local region, they can solve the problem of the object being submerged by the highlighted background in the neighborhood to a certain extent. A nonlocal information single-frame algorithm makes full use of all the information in the frame to detect a dim and small target, and does not care about the saliency of the target in the local area. Even if the target is not prominent because it is close to a highlighted background, it can still be successfully separated, which is a significant advantage compared with local information algorithms. However, there are shortcomings in the current nonlocal information single-frame algorithms which need to be further studied and improved. For example, the frequency domain method assumes that the target and the background occupy different frequency bands. However, when the background is complex, the frequency bands of the target and the background coincide to an extent, making it difficult to distinguish them accurately in the frequency domain. Most classifier methods need a large number of training samples; however, in the field of infrared dim and small target detection it is difficult to obtain samples, especially in the face of non-cooperative targets. The performance of the over-complete sparse representation depends on the accuracy of the over-complete dictionary. However, in practice, it is impossible to construct an over-complete dictionary that covers all situations.

### *1.2. Sequential Detection Methods*

Unlike single-frame algorithms, multi-frame algorithms use both spatial and temporal information to detect a dim and small target and predict the trajectory of the target in the sequence image. When there are multiple sources of interference within the field of view that are very similar to real objects (such as fragmenting clouds, etc.), this kind of method is usually more effective than a pure single-frame algorithm. The current multi-frame algorithms can be roughly divided into two categories. One is the association verification class [25,26], in which a single-frame algorithm is used to find suspicious targets in each frame, then the continuity of the target motion trajectory is used to check between multiple frames to further exclude false targets and extract the motion trajectory of the true target; this approach is sometimes called "Detect Before Track" (DBT). The other is to directly obtain the class [27,28] by using the gray fluctuation information of the moving target in the time domain to directly extract the target in multiple frames. This method is sometimes called "Track Before Detect" (TBD).

The performance of correlation check algorithms depends to a large extent on the detection performance of the preceding single frame algorithm itself. If the performance of the single frame algorithm is very poor (such as a large number of missed detection, etc.), it is difficult to effectively improve the final detection performance even if a subsequent time domain correlation check is added. It is even possible for the post-validation performance to be worse than the pre-validation performance. Theoretically, the advantage of direct acquisition algorithms is that they use the information of multiple frames to directly obtain the target in the time domain, meaning that their detection performance is generally better. Moreover, in certain cases the target position and the target trajectory can be determined simultaneously.

However, existing methods fail to effectively exploit spatial and temporal information simultaneously, particularly against complex backgrounds. Therefore, the objective of this research is to develop an algorithm that can efficiently leverage the spatiotemporal structure within infrared sequences while achieving exceptional performance on issues related to Infrared Small Target Detection (ISTD).

### *1.3. Related Work*

According to the sparse low-rank decomposition method, the background image and the target image come from different linear subspaces, and their characteristics are very different. The background image has certain nonlocal self-similarity and high internal data correlation, meaning that the background image has a low rank ratio and is the main component of an image. The size of the small target is not large, and the proportion of the whole image is extremely small; thus, the target image should be very sparse, which is the abnormal information in an image. Therefore, if the original image matrix can be divided into a matrix with the lowest rank and a matrix with the highest sparsity under certain optimization criteria, the background part can be used to achieve separation of the target parts.

After the image block matrix method (IPI) was first proposed by Gao et al. [29], this idea was quickly recognized and inherited by many researchers, and many further improvements were made to solve the problem of incorrect decomposition of the background edges. However the practice of "pulling vectors" for each image block in the image block matrix method destroys this spatial correlation, resulting in degradation of detection performance. Therefore, this kind of tensor method directly stacks each image block in order to form a new three-dimensional data cube, and this three-dimensional data cube is decomposed in the tensor space. Because the data cube can be expanded in the x, y, and z directions, it can describe more spatial correlation information between adjacent pixels. Dai and Wu [30] first introduced the tensor concept into the field of infrared dim and small target detection and proposed the reweighted IR patch tensor (RIPT), in which the local complexity is used as weighted feedback to the sparse foreground part during decomposition to ensure that the complex background edges and corners are decomposed into the low-rank background

part as much as possible. The motivation behind this approach stems from the Tensor Robust Principal Component Analysis (TRPCA) model.

Wang et al. [31] employed total variation regularization to effectively capture both internal smoothing and edge details within infrared (IR) backgrounds. This approach yields promising detection outcomes even when confronted with non-uniform or unsmooth IR scenes. To enhance detection speed, Zhang and Peng [32] utilized the partial sum of TNN (PSTNN) as a constraint for low-rank background tensors. Additionally, Cao et al. [33] introduced a novel approximation method called mode-k1k2 extension tensor tubal rank (METTR) specifically designed for estimating tensor ranks within IPT models. Kong et al. [34] proposed an enhanced version of the IPT model by incorporating LogTFNN non-convex tensor fibered rank approximation.

In order to exploit the spatiotemporal correlation of infrared image sequences, Hu et al. proposed a novel multi-frame spatial-temporal patch-tensor (MFSTPT) model for ISTD from complex scenes. In this approach, the Laplace function is utilized based on non-convex approximation to measure the low-rank properties of the background, and novel weighted prior information is introduced to suppress interference from the edges and highlighted areas. Moreover, a new tensor model is constructed using a simultaneous sliding window in space and time to satisfy the assumption of low-rank background. The ECA-STT model proposed by Zhang et al. [35] presents a novel approach for STT that incorporates edge and corner awareness. The ECA indicator is designed based on the ST to effectively differentiate between the target and sparse background residuals by adjusting its importance factor. Hu et al. [36] proposed a novel multi-frame spatial–temporal patch-tensor (MFSTPT) model that utilizes the Laplace method to approximate the rank of the tensor and established a prior weighted saliency map to suppress strong interference and sparse noise. Jie Li et al. [37] proposed a novel twist tensor model called sparse regularization-based SRSTT. This model effectively enhances the distinction between the target and complex background by introducing a distortion tensor model based on perspective transformation. However, the computational time of this method is prolonged due to the presence of abundant clutter in complex backgrounds. Wu et al. [38] proposed a four-dimensional infrared image patch tensor structure and decomposed into lower-dimensional tensors using the tensor train (4D−TT) and its extension-tensor ring (4D−TR) techniques. The proposed structure takes into account both local and global features in the spatial and temporal domains while avoiding discontinuities between patch sequences. However, this algorithm is time-consuming when detecting target-like items in fast-moving complex environments.

### *1.4. Motivation*

It is evident that the primary focus of single-frame tensor-based and sequence tensor-based infrared small target detection methods lies in exploring the 3D tensor structure. The single-frame tensor approach preserves local characteristics by arranging local patches within a 3D space, while incorporating a global prior to ensure object sparsity and low-ranking background. In terms of multi-frame tensor methods, certain approaches enhance the temporal aspect by overlaying the original image to aid in ISTD operations [35,39]. In [33,40], the authors partially adopted the concept from single frame tensors, where local blocks within a single frame are superimposed and combined across time to maintain localized information in both spatial and temporal domains.

Directly processing the entire sequence would result in significant computational time and disregard the spatial correlation of patches, making it impossible to simultaneously achieve spatiotemporal correlations of both local and global information within a 3D space. Although Fengyi Wu proposed a four-dimensional infrared image patch tensor structure, the structure is based on the patch size as the first two orders, the time as the third order, and the patch index as the fourth order. This structure preserves the continuity of the IR tensor in time sequence and the continuity in space and avoids the discontinuities between

the patch sequences; however, it is time-consuming to detect target-like items and a large number of coefficient background edges in fast-moving complex environments.

The crux of enhancing the detection rate for infrared small targets lies in effectively extracting pertinent information, as infrared images exhibit limited characteristics of small targets. Although the existing infrared small target detection techniques based on spatio-temporal tensor (STT) have good performance, the majority of existing methods fail to effectively leverage temporal information across consecutive multiple frames, necessitating the development of a novel detection model dataset construction approach. Moreover, compared to moving targets, the background of infrared images exhibits minimal changes and strong correlation within a short time frame, suggesting that combining similar regions in an appropriate manner could potentially emphasize the low-rank features of the background and enhance target visibility. Based on the problems of the above spatiotemporal tensor models, a four-dimensional tensor model based on superpixel segmentation and statistical clustering is established. The following section outlines the principal contributions of this study:

1. First, superpixel segmentation is employed to remove the reliance of conventional algorithms on the dimensions of sliding windows. For the first time, the application of superpixel segmentation to infrared dim and small target detection utilizing spatiotemporal tensor models is presented in this study.
2. Second, in order to make better use of the spatiotemporal correlation, the Cluster 4D-FCTN model is proposed. Based on the improved structure tensor theory, the image pixels are statistically clustered into three types: corner area, flat area, and edge area. The 3D patches with the same feature type are rearranged into the same group to form a four-dimensional tensor, and different weights are assigned to the image pixels with line prior and point prior to reduce the influence of strong background edges.
3. A fully-connected tensor network (FCTN) is proposed to detect small infrared targets using spatial and temporal correlation, which can better approximate the tensor rank. The FCTN decomposition is able to fully characterize the correlation between any two modes of a tensor. Additionally, the alternating direction multiplier method (ADMM), a highly efficient approach, has been developed to precisely address the proposed optimization model.

## 2. Notations and Preliminaries

In this segment, we present a succinct overview of the fundamental symbols and diverse explanations utilized within this manuscript.

### *2.1. Notations*

We employ $x$, x, X, and $\mathcal{X}$ to denote scalars, vectors, matrices, and tensors, respectively. For instance, a tensor of order N = 4 with size $M_1 \times M_2 \times M_3 \times M_4$ is denoted by $\mathcal{X} \in \mathbb{R}^{m_1 \times m_2 \times m_3 \times m_4}$. The representation of an element in a four-dimensional tensor is denoted as $\mathcal{X}(m_1, m_2, m_3, m_4)$ or $x_{m_1 m_2 m_3 m_4}$, while the index is $(m_1, m_2, m_3, m_4)$. $\mathcal{X}(m_1 \dots m_{n-1}, :, m_{n+1} \dots m_4)$ represents the mode fiber of $x$, $\mathcal{X}(\dots, m_n, \dots)$ indicates the mode-n slice, and the mode-n unfolded matrices of tensor $\mathcal{X}$ is $\mathbf{X}_{(n)} \in \mathbb{R}^{m_n \times \Pi_{k \neq n} m_k}$. The inner product of two tensors $\mathcal{X}$ and $\mathcal{Y}$ of equal dimensions is computed by summing the element-wise products, $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{m_1, m_2, \cdots, m_N} \mathcal{X}(m_1, m_2, \cdots, m_N) \mathcal{Y}(m_1, m_2, \cdots, m_N)$. The definitions of the $l_1$-norm and Frobenius norm for $\mathcal{X}$ are given as $\|\mathcal{X}\|_1 = \sum_{m_1, m_2, \cdots, m_N} |\mathcal{X}(m_1, m_2, \cdots, m_N)|$ and $\|\mathcal{X}\|_F = \sqrt{\sum_{m_1, m_2, \cdots, m_N} |\mathcal{X}(m_1, m_2, \cdots, m_N)|^2}$, respectively.

### *2.2. FCTN Decomposition*

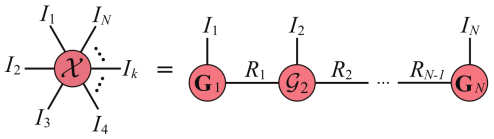
To effectively extract and utilize information from four-dimensional tensor data, a more suitable decomposition strategy than the traditional three-dimensional decomposition method is needed. This paper proposes a new tensor decomposition called Fully Connected Tensor Network (FCTN) decomposition [41], which facilitates the decomposition

of an nth-order tensor into a sequence of low-dimensional nth-order factors. Additionally, our method establishes an operation between any two factors. It breaks through the limitation that the exclusive role of TT and TR decomposition lies in the disclosure of the operation limited to two consecutive factors, rather than the connection between any two factors, resulting in their limited ability to measure the correlation.

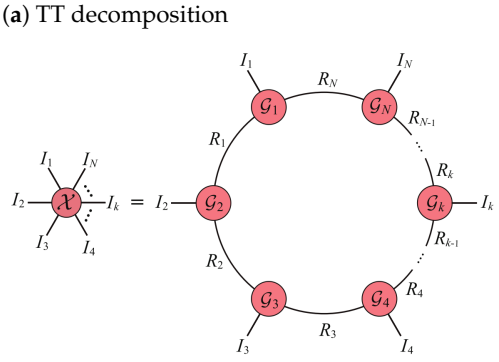
Instead of utilizing RPCA [42] based on matrices, TRPCA exhibits superior performance in handling high-dimensional data [43]. It breaks down a tensor of order m into a linear combination of a tensor with low rank and an element that is sparse. Unlike the matrix scenario, various types of tensor rank are available, including TTucker rank [29], multi-rank and tubal rank [31], TT rank [44], and TR rank [45] obtained from corresponding tensor decompositions. The Tucker decomposition method is employed to factorize an Nth-order tensor $\mathcal{X} \in \mathbb{R}^{M_1 \times M_2 \times \cdots \times M_N}$ into a compressed Nth-order core tensor, denoted as $\mathcal{G}$, which is multiplied by a matrix along each mode, resulting in $\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$. The Tucker rank is denoted by a vector, where the kth element indicates the ranking of the mode-k matricization of tensor $\mathcal{X}$. i.e., $\text{rank}_{\text{Tu}}(\mathcal{X}) := \left(\text{rank}\left(\mathbf{X}_{(1)}\right), \text{rank}\left(\mathbf{X}_{(2)}\right), \cdots, \text{rank}\left(\mathbf{X}_{(N)}\right)\right)$, where $\mathbf{X}_{(k)} \in \mathbb{R}^{M_k \times (M_1 \cdots M_{k-1} M_{k+1} \cdots M_N)}$ is the mode-k matricization of $\mathcal{X}$. Liu et al. introduced the concept of minimizing the Tucker rank by utilizing the sum of nuclear norms (SNN). However, it should be noted that SNN does not represent the most accurate convex envelope for calculating the sum of ranks from unfolding matrices of a tensor. The employment of an unbalanced matricization scheme to unfold a tensor into matrices along one mode exacerbates the complexity of this issue. Consequently, it can be argued that the Tucker rank may not effectively capture all global information pertaining to the tensor [46]. Introduced in the work of [47], Tensor Nuclear Norm (TNN) characterizes correlations along the first and second modes using tensor singular value decomposition (t-SVD). However, the main emphasis of t-SVD lies in third-order tensors, which might restrict its suitability for scenarios that involve high-dimensional data such as color videos and multi-temporal remote sensing images. In such cases, the incorporation of t-SVD and its related models based on TNN may not sufficiently capture the inherent low-dimensional structure present in the data. TT decomposition, as shown in Figure 1a [48], factorizes a tensor $\mathcal{X} \in \mathbb{R}^{M_1 \times M_2 \times \cdots \times M_N}$ of order N into two matrices and $\mathrm{N}-2$ tensors of the third order. TR decomposition, as depicted in Figure 1b, factorizes an Nth-order tensor $\mathcal{X} \in \mathbb{R}^{M_1 \times M_2 \times \cdots \times M_N}$ into a product formed by taking the circular multilinear combination of a set of third-order core tensors; its element-wise representation is denoted as $\mathcal{X}(m_1, m_2, \cdots, m_N) = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \cdots \sum_{r_N=1}^{R_N} \{\mathcal{G}_1(r_1, i_1, r_2)\mathcal{G}_2(r_2, i_2, r_3) \cdots \mathcal{G}_N(r_N, i_N, r_1)\}$. The vector $(R_1, R_2, \cdots, R_N)$ represents the TR rank. To address the low-rank tensor completion problem, TR nuclear norm minimization (TRNNM) [49] proposes utilizing the convex surrogate $\sum_{k=1}^{N} \beta_k \left\|\mathbf{X}_{<k,L>}\right\|_*$ as an alternative measure for the TR rank. Here, $\mathbf{X}_{<k,L>} \in \mathbb{R}^{\Pi_{i=k}^{k+L-1} I_i \times \Pi_{i=k+L}^{k-1} I_i}$ denotes the circular unfolding matrix of the tensor. The TRNNM-based RTC problem was investigated by Huang et al. [50]. To explore the interactions between two arbitrary factor tensors, Zheng et al. [51] proposed a decomposition technique known as Fully Connected Tensor Network (FCTN) decomposition, as illustrated in Figure 1c. This method breaks down an Nth-order tensor $\mathcal{X} \in \mathbb{R}^{M_1 \times M_2 \times \cdots \times M_N}$ into smaller Nth-order tensors, represented element-wise as

$$\begin{aligned} \mathcal{X}(m_1, m_2, \cdots, m_N) = & \sum_{r_{1,2}=1}^{R_{1,2}} \sum_{r_{1,3}=1}^{R_{1,3}} \cdots \sum_{r_{1,N}=1}^{R_{1,N}} \sum_{r_{2,3}=1}^{R_{2,3}} \cdots \sum_{r_{2,N}=1}^{R_{2,N}} \cdots \sum_{r_{N-1,N}=1}^{R_{N-1,N}} \\ & \{\mathcal{G}_1(m_1, r_{1,2}, r_{1,3}, \cdots, r_{1,N}) \cdots \mathcal{G}_N(r_{1,N}, r_{2,N}, \cdots, r_{N-1,N}, m_N)\}. \end{aligned} \tag{1}$$

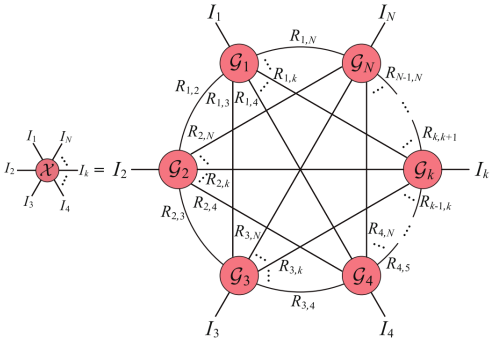
The rank of the FCTN is characterized as a vector $(R_{1,2}, \cdots, R_{1,N}, R_{2,3}, \cdots, R_{2,N}, \cdots, R_{N-1,N})$. The FCTN decomposition exhibits exceptional performance in addressing the issue of tensor completion, surpassing alternative tensor decompositions. This can be attributed to its effective capture of correlations across diverse modes.



(**a**) TT decomposition



(**b**) TR decomposition



(**c**) FCTN decomposition

**Figure 1.** Depiction of tensor network decomposition.

## 3. Methodology

In the current chapter, we initially explore the technique employed for constructing tensors, and employ the concept of superpixel segmentation to reduce the algorithm's reliance on sliding window size selection. Subsequently, we introduce background clustering based on the structure tensor. Furthermore, we delve into the Cluster 4D infrared tensor model and propose a 4D tensor-based 4D-FCTN method through FCTN decomposition and unfolding algorithms. Additionally, comprehensive explanations regarding the solution procedure derived from ADMM are readily accessible.

### *3.1. Superpixel Segmentation*

Because the method of using a fixed-size sliding window to intercept pixels does not consider the integrity of local features in the graph, it leads to a complete object being divided into several parts, which affects the subsequent decomposition. Therefore, this chapter first studies the means of tensor construction, and uses the idea of superpixel segmentation to reduce the dependence of the algorithm on selection the of sliding window size.

As shown in the first row of Figure 2, the size of the sliding window used in the first row is 40 × 40; the target image only contains the target, and the background residual of strong edges is completely removed. The size of the sliding window used in the second row is 20 × 20, and the background residual exists in the target image, which increases the likelihood of false alarms. In order to illustrate this problem, statistical analysis is carried out on two image sequences. The target size in the two image sequences is the same, meaning that the sliding window size to achieve optimal performance of the two sequences should be similar; however, the ROC curve in Figure 3 show that this is not the case. For the first sequence, the optimum is achieved when the sliding window is set to 50, while for the second sequence the effect is poor when the sliding window is set to 50; thus, it is obvious that the first sequence is more sensitive to the sliding window than the second one. The fundamental reason behind this result is that the grouping method is based on the probability of the block image belonging to a certain feature; thus, the size of the sliding window affects which group the block image is divided into. The method of using a fixed-size sliding window to intercept pixels does not consider the integrity of local features in the image, leading to a complete object being divided into several parts, which are divided into different groups. Therefore, a more reasonable segmentation method is needed to more accurately describe the characteristics of the block image in order to maintain the integrity of the local features of the image.

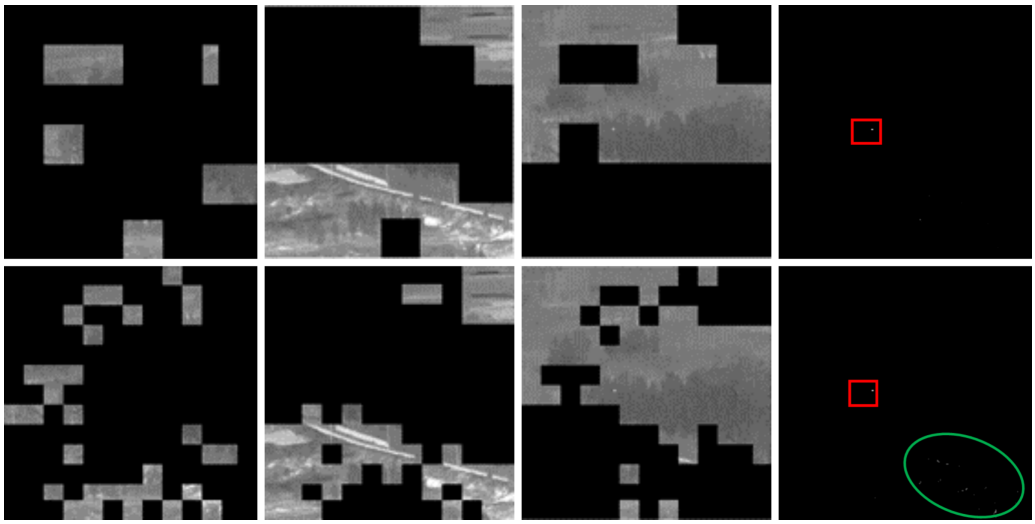


**Figure 2.** Clustering and separated target images obtained using different patch sizes.

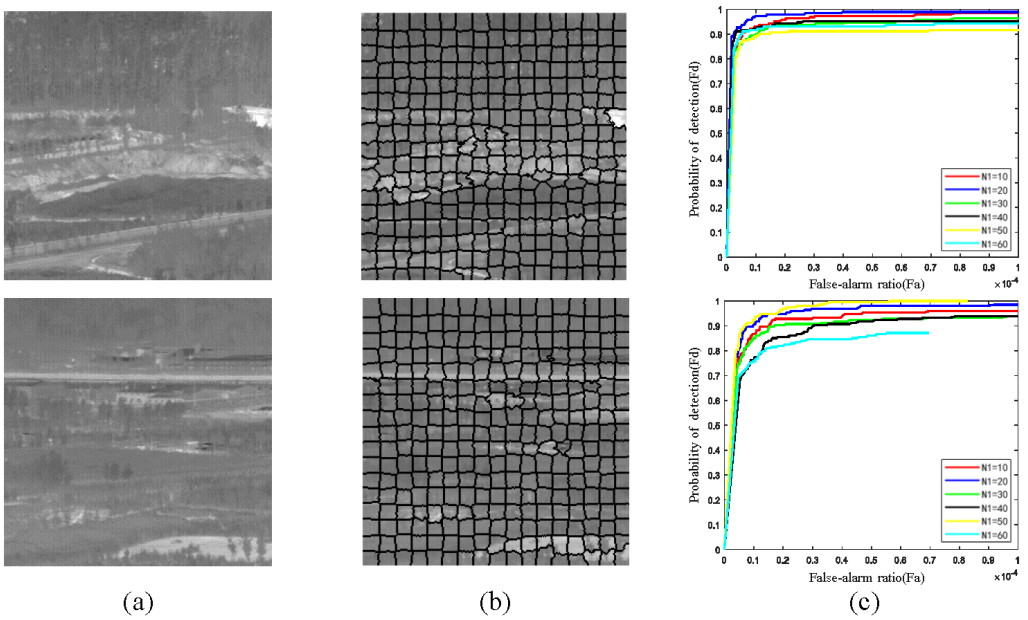



**Figure 3.** Illustration of the sensitivity of the detection results to the sliding window size: (**a**) two original images with similar complexity(**b**) results of superpixel segmentation and (**c**) ROC curves obtained with sliding windows of different sizes.

Superpixels are irregular regions composed of pixels with similar features. These small areas retain the relevant information of the local area; furthermore, because of the settings in the algorithm, the boundary information in the object is usually well preserved and is not destroyed. Therefore, superpixels are widely used in image preprocessing, object detection, image classification, and other fields. Among the available methods, Simple Linear Iterative Clustering (SLIC) [52] has the advantages of simplicity and efficiency, and has attracted the attention of many scholars. For an input image, the main idea of the SLIC algorithm is as follows: first, the image is divided into an $S \times S$ regular grid; in order to avoid the edge parts or noisy points in the image becoming the center of the current superpixel, the cluster center is taken as the minimum gradient position in its $3 \times 3$ neighborhood during each calculation, then different features are selected as the similarity measure. As shown in Figure 4, the similarity value is usually computed by searching in the neighborhood of $2S \times 2S$ from the center of superpixel, and it is assigned to the nearest cluster center. Finally, after iteration, when the residual of the cluster center in the last iteration is less

than the threshold, the iteration is stopped and the final superpixel segmentation result is obtained. Because infrared images do not have color information, the gray distance and spatial distance are selected as similarity measures in this paper. For a cluster center $C_i = \begin{bmatrix} l_i & x_i & y_i \end{bmatrix}$, the Euclidean distance D between it and each pixel in its neighborhood is calculated as follows:

$$d_c = \sqrt{(l_p - l_q)^2} \tag{2}$$

$$d_s = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2} \tag{3}$$

$$\mathrm{D} = \sqrt{d_c^2 + \alpha\left(\frac{d_s}{S}\right)^2} \tag{4}$$

where $d_c$ represents the grayscale distance, $d_s$ represents the spatial distance, and $\alpha$ is a constant.

As shown in Figure 4, irregular regions are obtained after superpixel segmentation, and the size of image blocks needs to be consistent during tensor calculation. Therefore, we take the maximum size of the region obtained by superpixel segmentation as the uniform size of the image block when constructing the tensor, and expand the region smaller than this size from the center to the periphery to fill with the original pixel. Thus, the construction of the tensor is completed.

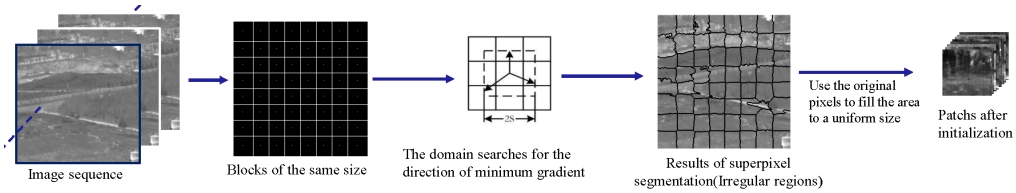



**Figure 4.** The whole process of sliding window selection for adaptive superpixel segmentation.

### *3.2. Description of Features Exploiting the Structure Tensor*

The structure tensor, in contrast to the gradient, enhances local structural characteristics by employing a smoothing procedure and assessing the coherence of the spatial orientations surrounding each pixel. Assuming that D represents the original image, we can calculate the conventional structure tensor with a linear configuration as described in Equation (5) [53,54]:

$$\begin{aligned} \mathbf{J} &= \mathbf{K}_\rho * (\nabla \mathbf{D}_\sigma \otimes \nabla \mathbf{D}_\sigma) \\ &= \begin{pmatrix} \mathbf{K}_\rho * \mathbf{I}_\mathbf{x}^2 & \mathbf{K}_\rho * \mathbf{I}_\mathbf{x}\mathbf{I}_\mathbf{y} \\ \mathbf{K}_\rho * \mathbf{I}_\mathbf{x}\mathbf{I}_\mathbf{y} & \mathbf{K}_\rho * \mathbf{I}_\mathbf{y}^2 \end{pmatrix} = \begin{pmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{pmatrix} \end{aligned} \tag{5}$$

where $\mathbf{K}_\rho$ represents a kernel function with a Gaussian distribution and variance $\rho$, while $\mathbf{D}_\sigma$ denotes a Gaussian smoothing filter with variance $\sigma$ applied to the original image. The gradients of $\mathbf{D}_\sigma$ with respect to $\mathbf{x}$ and $\mathbf{y}$, denoted as $\mathbf{I}_\mathbf{x} = (\partial \mathbf{D}_\sigma / \partial \mathbf{x})$ and $\mathbf{I}_\mathbf{y} = (\partial \mathbf{D}_\sigma / \partial \mathbf{y})$, respectively, are obtained through a convolution operation. The gradient operator is represented by $\nabla$, and the Kronecker Product is used. It should be noted that the structure tensor has dimensions of $2 \times 2$. The local image features can be enhanced by employing principal axis transformation in the structure tensor, thereby improving their discriminative power and facilitating more accurate analysis:

$$\mathbf{J} = \mathbf{S}^\mathrm{T} \mathbf{\Lambda} \mathbf{S} = \sum_\mathrm{m} \lambda_\mathrm{m} \mathbf{j}_\mathrm{m} \mathbf{j}_\mathrm{m}^\mathrm{T} \tag{6}$$

where $\mathbf{S} = [\mathbf{j}_1, \mathbf{j}_2]^{\mathrm{T}}$ and $\mathbf{j}_1 \perp \mathbf{j}_2$ is the structure tensor J characteristic vector, while $\mathbf{\Lambda} = \mathrm{diag}(\lambda_{\mathrm{m}})$ is a diagonal matrix of eigenvalues. The computation of the structure tensor's two eigenvalues is performed through

$$\lambda_1, \lambda_2 = \mathbf{J}_{11} + \mathbf{J}_{22} \pm \sqrt{(\mathbf{J}_{22} - \mathbf{J}_{11})^2 + 4\mathbf{J}_{12}^2}. \tag{7}$$

The characteristics of the image can be represented by the top two eigenvalues of the structure tensor under various conditions involving $\lambda_1$ and $\lambda_2$. First, $\lambda_1 \geq \lambda_2 \gg 0$, indicating a corner region. It can be observed that the partial derivative amplitude is similar in two vertical directions; however, there is a noticeable difference in their respective orientations. This discrepancy indicates a variation in the location of this performance, particularly towards the corner. Second, $\lambda_1 \gg \lambda_2 \approx 0$, indicating an edge region. This implies that the partial derivative at this particular position exhibits a minimal change in one direction, approaching zero, while displaying a substantial variation in the perpendicular direction. This pronounced dissimilarity signifies a distinct orientation, thereby indicating a high likelihood of the point being situated on an edge. Third, $\lambda_1 \approx \lambda_2 \approx 0$, indicating a flat region. This implies that the partial derivative at this particular position exhibits minimal variation in all directions, approaching zero. Consequently, the grayscale transition occurs gradually, indicating a high likelihood of the point being situated within a flat region.

Although both the RIPT model [30] and PSTNN model [32] utilize the structure tensor, they do not simultaneously incorporate all three types of feature points; instead, only one or two of them are utilized. In order to maximize the utilization of these three types of features, we have devised the following functions to depict the characteristics exhibited in infrared images:

$$\omega_a = \frac{1}{2} e^{\frac{\lambda_1}{\lambda_2 + 0.003} - 1} \left[ e^{\lambda_1(\lambda_2 - \lambda_1)} + e^{\lambda_2(\lambda_2 - \lambda_1)} \right], \omega_b = e^{\frac{\lambda_1}{\lambda_2 + 0.003} - 1} \tag{8}$$

where $\omega_b$ serves as a measure of the amplification factor indicating the difference between $\lambda_1$ and $\lambda_2$. When $\omega_a$ approaches infinity, this indicates that a pixel lies in an edge region and should be categorized as a line feature. Conversely, when $\omega_a$ approximates 1, it can be regarded as a flat feature while considering other parts of the initial area as point features. Nevertheless, careful selection of suitable threshold values becomes crucial due to noise interference and potential over-amplification caused by $\omega_b$. The decision-making process for thresholds can be summarized accordingly:

$$\mathrm{class}_p = a_1 \max(P), \quad \mathrm{class}_f = a_2 \min(P) \tag{9}$$

$$\mathrm{class}_{p1} = a_3 \min(Q), \quad \mathrm{class}_{f1} = a_4 \min(Q) \tag{10}$$

where $a_1$, $a_2$, $a_3$, and $a_4$ are fixed values, while P and Q denote the remaining elements obtained by removing infinity from sets $\omega_a$ and $\omega_b$, respectively. By conducting an exploration to satisfy the aforementioned criteria

$$\begin{cases} \omega_a(x,y) > \mathrm{Class}_p \text{ and } \omega_b(x,y) \leq \mathrm{Class}_{\mathrm{p1}}, & (x,y) \in \text{ point} \\ \omega_a(x,y) = \infty \text{ and } \omega_b(x,y) > \mathrm{Class}_{p1}, & (x,y) \in \text{ line} \\ \omega_a(x,y) \leq \mathrm{Class}_f \text{ and } \omega_b(x,y) \leq \mathrm{Class}_{f1}, & (x,y) \in \text{ flat} \end{cases} \tag{11}$$

the $(x, y)$ coordinates represent the pixel position. The corner region contains point features, the edge region contains line features, and the flat region contains flat features.

### *3.3. The Cluster 4D-FCTN Model*

#### 3.3.1. Four-Dimensional Infrared Image Tensor Model

IR small target detection models based on low-rank and sparse decomposition generally assume the presence of target, noise, and background in the original image. This assumption allows for an approximate modeling of the image, as follows:

$$f_D = f_B + f_T + f_N \tag{12}$$

where $f_D$, $f_B$ , $f_T$, and $f_N$ represent the initial image, backdrop image, desired image, the noise image, respectively. The IPI model represents an optimal approach for maximizing the utilization of this characteristic, as illustrated below:

$$D = B + T + N \tag{13}$$

where $T$, $B$, $N$, and $D$ represent the target, background, noise, and the original infrared image.

The presence of spatial correlation between adjacent pixel data in the image is disrupted by employing the "pulling vector" approach for each image block in the image block matrix method, leading to a decline in detection performance. Conversely, utilizing the "pulling vector" technique for each image patch within the image patch matrix method disrupts this spatial correlation and results in a decrease in detection performance. To enhance the model's performance, we introduce a novel approach called image patch tensor (IPT). This method shares similarities with the IPI model, and is represented in Equation (14):

$$\mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{N} \tag{14}$$

where $\mathcal{D}, \mathcal{B}, \mathcal{T}, \mathcal{N} \in \mathbb{R}^{m_1 * m_2 * m_3}$ stand for the original image tensor, the background tensor, the target tensor, and the noise tensor, respectively.

However, both methods exhibit limited effectiveness in complex infrared environments. Moreover, the computational complexity is further exacerbated by the need for restacking after tensor generation, hindering the efficiency of execution. In order to boost local correlation and retain temporal data, Wu [38] recently presented a novel 4D infrared tensor model that eliminates the need for restacking and enhances the tensor order, enabling effective extraction of spatial–temporal information through high-order decomposition-based optimization. Consequently, the proposed formulation of the 4D infrared tensor $\mathcal{D}^{4D}$ can be expressed as follows:

$$\mathcal{D}^{4D} = \mathcal{B}^{4D} + \mathcal{T}^{4D} + \mathcal{N}^{4D} \tag{15}$$

where $\mathcal{D}^{4D}$, $\mathcal{B}^{4D}$, $\mathcal{T}^{4D}$, and $\mathcal{N}^{4D} \in \mathbb{R}^{M_1 \times M_2 \times M_3 \times M_4}$. Here, $M_3$ represents the number of consecutive frames in an image sequence, while $M_1$ and $M_2$ are patch sizes ($M_1 = M_2$), Additionally, we define both the sliding step $M_s$ and window size $M_1$ to have equal values. Lastly, we use $M_4$ to indicate the total quantity of generated patches.

It is possible to place a constraint on $\mathcal{D}^{4D}$ by taking advantage of the low-rank feature of the 4D background tensor[38]:

$$\operatorname{rank}\left(\mathcal{B}^{4D}\right) \leq r \tag{16}$$

where the complexity of the background is denoted by a positive constant r.

Usually, spatially sparse distributions characterize tiny infrared targets which occupy a limited area. Despite environmental fluctuations occurring rapidly, these targets only cover a small fraction of their trajectory zone over time. As a result, they maintain their sparsity even within temporal patch volumes. This property guarantees that the 4D target tensor $\mathcal{T}^{4D}$ satisfies the following condition:

$$\left\|\mathcal{T}^{4D}\right\| \leq \alpha \tag{17}$$

where the value of $\alpha$ is determined by the characteristics and number of the targets. Subsequently, the low-ranking background and sparse target can be effectively separated through TRPCA modeling:

$$\begin{aligned} &\min_{\mathcal{B}^{4D},\mathcal{T}^{4D}} \operatorname{rank}\left(\mathcal{B}^{4D}\right) + \lambda_1 \left\|\mathcal{T}^{4D}\right\|_0 \\ &\text{s.t. } \mathcal{D}^{4D} = \mathcal{B}^{4D} + \mathcal{T}^{4D} + \mathcal{N}^{4D} \end{aligned} \tag{18}$$

where the weighting parameter $\lambda_1$ regulates the overall tradeoff between the target patch-tensor $\mathcal{T}^{4D}$ and the background patch-tensor $\mathcal{B}^{4D}$.

However, the fourth dimension of the 4D tensor constructed by the algorithm does not fully maintain the spatial continuity and correlation of patches, that is, all patches are arranged into a row to form the fourth dimension, which destroys the spatial structure of the patches.

#### 3.3.2. ISTD Based on C4D-FCTN

Based on the problems of the ordinary 4D tensor model, we propose a feasible classification and combination of 3D patch tensors to form several kinds of more relevant 4D tensors to solve such limitations.

In a complex background, strong edges sometimes only occupy a small portion of the image with distinct edge characteristics. However, more blocks may be required to adequately capture the low-rank property of the background in such cases. To address this issue, we propose modeling and grouping similar feature blocks that satisfy the low-rank property within their respective spaces. This approach aligns better with the low-ranking nature of the background and enables parallel computation through block grouping, thereby reducing computational time.

The infrared image features can be clustered into three groups based on the tensor feature description, which is the foundation for calculating the probability of belonging to each group within a block. The image patches exhibiting significant similarities are naturally grouped together. The probability of the i-th patch belonging to the three features can be computed as follows:

$$P_i(\text{point}, \text{line}, \text{flat}) = \frac{k_i(\text{piont}, \text{line}, \text{flat})}{\max\left(k_{\text{point}}, k_{\text{line}}, k_{\text{flat}}\right)} \tag{19}$$

where $k_i$ (point, line, flat) represents the number of pixels belonging to point features, line features, and flat features in the i-th image patch, while $k_{\text{point}}$, $k_{\text{line}}$, and $k_{\text{flat}}$ represent the number of pixels belonging to point features, line features, and flat features in all image patches.

After grouping the block images, we tested the low-rank property of tensors in each cluster and performed singular value decomposition on them. The results are presented in Figure 5, where (a) denotes the original image, (b) represents the singular value curve of the point feature cluster tensor, (c) shows the singular value curve of the line feature cluster tensor, and (d) displays the singular value curve of the flat feature cluster tensor. It is evident that different clusters possess their own intra-cluster low-rank properties; hence, the background can be modeled on this basis.

After the completion of the aforementioned clustering process, instead of stacking the spatial tensor patches in series, we rearrange the 3D patches with the same feature type into the same cluster to form the 4D tensor (that is, the four dimensions of the 4D tensor are the length of the sliding window, width of the sliding window, sequence index, and sliding window index within a cluster). Specifically, as described in Algorithm 1, the initialized patch is obtained by adaptively selecting the sliding window size through superpixel segmentation of the initial frame. Subsequently, the feature detection function in Formula (11) is used to extract three kinds of feature points on the initialized patch. The clustering criterion in Formula (19) is then applied to calculate the probability of each pixel in the sliding window belonging to these three types of features based on the extracted

feature points. This probability is further divided into a cluster of tensors with maximum probability, resulting in three 3D tensors: $\left(D^{3D}\right)_p$, $\left(D^{3D}\right)_l$, and $\left(D^{3D}\right)_f$. Subsequently, each patch in the $\mathbf{N}_3$ frames are classified as the corresponding patch at the same position as the initial frame, thereby stacking the image patches into three 4D tensors: $\left(D^{4D}\right)_p$, $\left(D^{4D}\right)_l$, and $\left(D^{4D}\right)_f$.

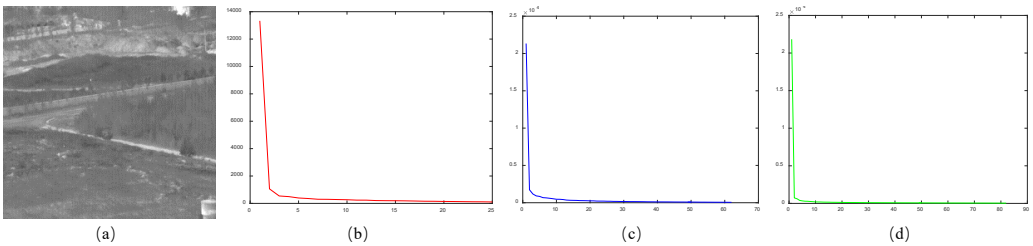



**Figure 5.** Intra-cluster low-rank properties of different clusters (**a**) represents the original image, (**b**) represents the singular value curve of the point feature group tensor, (**c**) represents the singular value curve of the line feature group tensor, and (**d**) represents the singular value curve of the flat feature group tensor.

The stacking is performed in the order of the third-order coordinates of the frame index $\mathbf{N}_3$, and the patch index $\mathbf{N}_4$ is used as the fourth-order axis. The patches with the same feature type in the third dimension patch of a 4D tensor are rearranged into the same cluster to form three 4D tensors. Using MATLAB, we start by utilizing sliding window techniques to produce $\mathbf{N}_1$ localized patches from a single image (the starting frame) for the purpose of constructing sub-three-way tensors. Subsequently, we utilize the operator for multidimensional arrays to accomplish the stacking process of the 4D tensor.

**Algorithm 1:** The process of building the 4D infrared image tensor for clustering

**Input:** Image sequence $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$, ...., $\mathbf{S}_N \in R^{m\times n}$
**Output:** Original 4D tensor $\mathcal{D}^{4D}$
**Initialization:** The initialized patch is obtained by superpixel segmentation of the initial frame. Initialize patch size $\mathbf{N}_1$, time size $\mathbf{N}_3$
**for** $i = 1 : \mathbf{N}_3$ **do**
1.Use formula (11) to extract three kinds of feature points on the initialized patches;
2.Use formula (19) to calculate the probability of each pixel in the sliding window belonging to these three types of features based on the extracted feature points. This probability is further divided into a cluster of tensors with maximum probability, resulting in three 3D tensors: $(\mathcal{D}_i^{3D})_p \in R^{M_1\times M_2\times M_4}$, $(\mathcal{D}_i^{3D})_l \in R^{M_1\times M_2\times M_4}$, $(\mathcal{D}_i^{3D})_f \in R^{M_1\times M_2\times M_4}$;
3.each patch in the $\mathbf{N}_3$ frames are clustered as the corresponding patch at the same position as the initial frame, thus along the direction of the third dimension $\mathbf{N}_3$, stacking the image patches into three 4D tensors: $(\mathcal{D}^{4D})_p(:,:,i,:) = \mathcal{D}_i^{3D}$, $(\mathcal{D}^{4D})_l(:,:,i,:) = \mathcal{D}_i^{3D}$, $(\mathcal{D}^{4D})_f(:,:,i,:) = \mathcal{D}_i^{3D}$.
**end**

A recent investigation [51] has unveiled the efficacy of the FCTN-based TRPCA technique in handling high-dimensional data, providing an accurate estimation of tensor rank. In comparison to alternative tensor decompositions, FCTN decomposition demonstrates superior performance when addressing challenges related to tensor completion. This advantage arises from its inherent ability to flexibly capture correlations among different modes. In this study, we propose a novel Cluster 4D-FCTN model that employs a new FCTN nuclear norm as a convex surrogate for quantifying mode correlations within our

4D model. For an Nth-order tensor $\mathcal{X} \in \mathbb{R}^{M_1 \times M_2 \times \cdots \times I_N}$, we formulate the FCTN nuclear norm as follows:

$$\left\|\mathcal{B}^{4D}\right\|_{4D-FCTN} = \sum_{k=1}^{L} \beta_k \left\|\mathbf{B}_{k\left[\mathbf{n}_1^k;\mathbf{n}_2^k\right]}\right\|_* \tag{20}$$

where the weighting factors $\beta_k$ satisfy the condition $\sum_{k=1}^{L} \beta_k = 1$, which are all positive, $\boldsymbol{n}^k$ is the k-th rearrangement of the vector $[1, 2, \cdots, N]$, $\boldsymbol{n}_1^k = \boldsymbol{n}_{1:\lfloor N/2 \rfloor}^k$, $\boldsymbol{n}_2^k = \boldsymbol{n}_{\lfloor N/2 \rfloor+1:N}^k$, and $L = C_N^{\lfloor N/2 \rfloor}$.

To ensure low rank within clusters while maintaining each cluster's independence from the others and global sparsity of targets, our proposed Cluster 4D Fully-Connected Tensor Network model (C4D-FCTN) is

$$\min_{\mathcal{B}^{4D},\mathcal{T}^{4D},\mathcal{N}^{4D}} \sum_{i=1}^{3} \left\|\left(\mathcal{B}^i\right)^{4D}\right\|_{4D-FCTN} + \lambda_1 \|\mathcal{T}^{4D}\|_1 + \lambda_2 \|\mathcal{N}^{4D}\|_F^2 \tag{21}$$

$$\text{s.t. } \mathcal{D}^{4D} = \mathcal{B}^{4D} + \mathcal{T}^{4D} + \mathcal{N}^{4D}.$$

Our proposed method effectively utilizes the low-rank characteristics of the background and the sparse properties of the target, with the positive penalty parameter $\lambda_2$ responsible for regulating noise and $\left(\mathcal{B}^i\right)^{4D}$ denoting the background tensor of cluster $i$.

$$\begin{aligned} \min_{\mathcal{B}^{4D},\mathcal{T}^{4D},\mathcal{N}^{4D}} & \sum_{k=1}^{L} \sum_{k=1}^{3} \beta_k \left\|\mathbf{B}^i_{k\left[\mathbf{n}_1^k:\mathbf{n}_2^k\right]}\right\|_* + \lambda_1 \left\|\mathcal{T}^{4D}\right\|_1 + \lambda_2 \left\|\mathcal{N}^{4D}\right\|_F^2 \\ & \text{s.t. } \mathcal{D}^{4D} = \mathcal{B}^{4D} + \mathcal{T}^{4D} + \mathcal{N}^{4D} \\ & \mathcal{E} = \mathcal{T}^{4D} \\ & \mathcal{B}_k = \mathcal{B}^{4D} \end{aligned} \tag{22}$$

### *3.4. Resolution of the Proposed Model*

We have developed the Cluster 4D-FCTN model for representing IR images. Subsequently, an Alternating Direction Method of Multipliers (ADMM) approach is employed to effectively address the provided optimization problem as a resilient tensor recovery task. By introducing auxiliary variables, Equation (22) is reformulated as follows:

$$\begin{aligned} & L\left(\mathcal{B}^{4D}, \mathcal{T}^{4D}, \mathcal{N}^{4D}, \mathcal{E}, \mathcal{B}_k, \mathcal{X}_k, \mathcal{U}, \mathcal{V}\right) \\ & = \sum_{k=1}^{N} \sum_{i=1}^{3} \sum_{k}^{3} \left\{ \beta_k \left\|\mathbf{B}^i_{k\left[\mathbf{n}_1^k:\mathbf{n}_2^k\right]}\right\|_* + < \mathcal{X}_k, \mathcal{B}_k^i - \left(\mathcal{B}^i\right)^{4D} > + \frac{\omega_k}{2} \left\|\mathcal{B}_k^i - \mathcal{B}^{i4D}\right\|_F^2 \right\} + \lambda_1 \left\|\mathcal{T}^{4D}\right\|_1 \\ & + \lambda_2 \left\|\mathcal{N}^{4D}\right\|_F + < \mathcal{U}, \mathcal{F} > + \frac{\sigma}{2} \|\mathcal{F}\|_F^2 + < \mathcal{V}, \mathcal{E} - \mathcal{T}^{4D} > + \frac{\gamma}{2} \left\|\mathcal{E} - \mathcal{T}^{4D}\right\|_F^2 \\ & \text{s.t. } \mathcal{F} = \mathcal{D}^{4D} - \mathcal{B}^{4D} - \mathcal{T}^{4D} - \mathcal{N}^{4D} \end{aligned} \tag{23}$$

where the variables $\omega_k$, $\sigma$, and $\gamma$ represent penalty parameters and where $\mathcal{X}_k$, $\mathcal{U}$, and $\mathcal{V}$ denote Lagrangian multipliers. According to the ADMM framework [55], $\mathcal{B}_k$, $\mathcal{E}$, $\mathcal{B}^{4D}$, and $\mathcal{T}^{4D}$ can be divided into two subsets, with the variables in each subset being updated alternately.

$$\left(\mathcal{B}_k^{t+1}, \mathcal{E}^{t+1}\right) = \arg\min_{\mathcal{B}_k,\mathcal{E}} L\left(\left[\sum_{i=1}^{3}\left(\mathcal{B}^i\right)^{4D}\right]^t, \mathcal{T}^{4Dt}, \mathcal{E}, \mathcal{B}_k, \mathcal{X}_k^t, \mathcal{U}^t, \mathcal{V}^t\right) \tag{24}$$

$$\left(\mathcal{B}^{4Dt+1}, \mathcal{T}^{4Dt+1}\right) = \arg\min_{\mathcal{B}^{4D},\mathcal{T}^{4D}} L\left(\sum_{i=1}^{3}\left(\mathcal{B}^i\right)^{4D}, \mathcal{T}^{4D}, \mathcal{S}^{t+1}, \mathcal{B}_k^{t+1}, \mathcal{X}_k^t, \mathcal{U}^t, \mathcal{V}\right)^t \tag{25}$$

Next, we provide additional information regarding each individual subproblem.

(1) Update $\mathcal{B}_k$. The formulation of subproblem $\mathcal{B}_k$ (where k = 1, 2, …, L) can be readily converted into an equivalent representation:

$$\begin{aligned}\mathbf{B}_k^{t+1} &= \arg\min_{\mathbf{B}_k} \alpha_k \sum_{i=1}^{3} \left\| \left(\mathbf{B}^i\right)_{k\left[\mathbf{n}_1^k;\mathbf{n}_2^k\right]} \right\|_* + < \mathcal{X}_k, \mathbf{B}_k - \mathcal{B}^{4D} > + \frac{\omega_k}{2} \left\| \mathcal{B}_k - \mathcal{B}^{4D} \right\|_F^2 \\ &= \arg\min_{\mathbf{B}_k} \alpha_k \sum_{i=1}^{3} \left\| \left(\mathbf{B}^i\right)_{k\left[\mathbf{n}_1^k;\mathbf{n}_2^k\right]} \right\|_k + \frac{\omega_k}{2} \left\| \mathbf{B}_{k\left[\mathbf{n}_1^k;\mathbf{n}_2^k\right]} - \left( \mathcal{B}^{4Dt} - \frac{\mathcal{X}_k^t}{\omega_k} \right)_{\left[\mathbf{n}_1^k;\mathbf{n}_2^k\right]} \right\|_F^2\end{aligned} \tag{26}$$

and the given equation possesses a closed-form solution

$$\mathbf{B}_{k\left[\mathbf{n}_1^k;\mathbf{n}_2^k\right]}^{t+1} = \mathbf{U}\Sigma_{\beta_k/\omega_k}\mathbf{V}^T \tag{27}$$

where $\left( \mathcal{B}^{4Dt} - \frac{\mathcal{X}_k^t}{\omega_k} \right)_{\left[\mathbf{n}_1^k;\mathbf{n}_2^k\right]} = \mathbf{U}\Sigma\mathbf{V}^T$, $\Sigma_{\beta_k/\omega_k} = \mathrm{diag}(\max(\Sigma_{r,r} - \beta_k/\omega_k, 0))$ and $\Sigma_{r,r}$ is the rth singular value of $\Sigma$.$\mathcal{B}_k^{t+1} = \Upsilon\left(\mathbf{B}_k^{t+1}\right)$.

(2) Update $\mathcal{E}$. The $\mathcal{E}$-subproblem is

$$\begin{aligned}\mathcal{E}^{t+1} &= \arg\min_{\mathcal{E}} \lambda\|\mathcal{E}\|_1 + \left\langle \mathcal{V}^t, \mathcal{E} - \mathcal{T}^{4Dt} \right\rangle + \frac{\gamma}{2} \left\| \mathcal{E} - \mathcal{T}^{4Dt} \right\|_F^2 \\ &= \arg\min_{\mathcal{E}} \lambda\|\mathcal{E}\|_1 + \frac{\sigma}{2} \left\| \mathcal{E} - \mathcal{T}^{4Dt} + \frac{\mathcal{V}^t}{\gamma} \right\|_F^2,\end{aligned} \tag{28}$$

which can be solved analytically using the following formula:

$$\mathcal{E}^{t+1} = \mathrm{soft}_{\lambda/\sigma}\left( \mathcal{T}^{4Dt} - \frac{\mathcal{V}^t}{\sigma} \right) \tag{29}$$

where the operator $\mathrm{soft}_{\lambda/\sigma}$ represents a soft shrinkage operation using a threshold value of $\lambda/\sigma$.

(3) Update $\left(\mathcal{B}^{4Dt}, \mathcal{T}^{4Dt}\right)$. This subproblem $(\mathcal{B}^{4Dt}, \mathcal{T}^{4Dt})$ can be formulated as a problem of minimizing the sum of squared residuals:

$$\begin{aligned}\left(\mathcal{B}^{4D(t+1)}, \mathcal{T}^{4D(t+1)}\right) &= \arg\min_{\mathcal{B}^{i4D}, \mathcal{T}^{4D}} \sum_{k=1}^{L} \left\{ \left\langle \mathcal{X}_k, \mathcal{B}_k - \mathcal{B}^{4D} \right\rangle + \frac{\omega_k}{2} \left\| \mathcal{B}_k - \mathcal{B}^{4D} \right\|_F^2 \right\} \\ &\quad + < \mathcal{U}, \mathcal{F} > + \frac{\gamma}{2}\|\mathcal{F}\|_F^2 + < \mathcal{V}, \mathcal{E} - \mathcal{T}^{4D} > + \frac{\sigma}{2} \left\| \mathcal{E} - \mathcal{T}^{4D} \right\|_F^2 \\ &= \arg\min_{\mathcal{B}^{4D}, \mathcal{T}^{4D}} \sum_{k=1}^{L} \left\{ \frac{\omega_k}{2} \left\| \mathcal{B}_k^{t+1} - \mathcal{B}^{4D} + \frac{\mathcal{X}_k^t}{\omega_k} \right\|_F^2 \right\} + \frac{\sigma}{2} \left\| \mathcal{D}^{4D(t+1)} - \mathcal{B}^{4D} - \mathcal{T}^{4D} + \frac{\mathcal{U}^t}{\sigma} \right\|_F^2 \\ &\quad + \frac{\gamma}{2} \left\| \mathcal{E}^{t+1} - \mathcal{T}^{4D} + \frac{\mathcal{V}^t}{\sigma} \right\|_F^2.\end{aligned} \tag{30}$$

and is acquired through

$$\mathcal{B}^{4D(t+1)} = \left(\sigma\mathcal{K}^t - (\sigma+\gamma)\mathcal{J}^t\right) / \left( \sigma^2 - \left( \sum_{k=1}^{L} \omega_k + \sigma \right)(\gamma+\sigma) \right) \tag{31}$$

and

$$\mathcal{T}^{4D(t+1)} = \left( \sigma\mathcal{J}^t - \left( \sum_{k=1}^{L} \omega_k + \sigma \right)\mathcal{K}^t \right) / \left( \sigma^2 - \left( \sum_{k=1}^{L} \omega_k + \sigma \right)(\gamma+\sigma) \right), \tag{32}$$

where $\mathcal{J}^t = \sum_{k=1}^{L} \omega_k \left( \mathcal{B}_k^{4D(t+1)} + \mathcal{X}_k^t / \omega_k \right) + \sigma \left( \mathcal{D}^{4D(t+1)} + \mathcal{U}^t / \sigma \right)$ and $\mathcal{K}^t = \sigma(\mathcal{D}^{4D(t+1)} + \mathcal{U}^t / \sigma) + \gamma(\mathcal{S}^{t+1} + \mathcal{V}^t / \gamma)$

(4) Update the multipliers. The Lagrangian multipliers are modified using the following procedure:

$$\begin{cases} \mathcal{X}_k^{t+1} = \mathcal{X}_k^t + \alpha \omega_k \left( \mathcal{B}^{4D(t+1)} - \mathcal{B}_k^{t+1} \right), k = 1, 2, \cdots, L \\ \mathcal{U}^{t+1} = \mathcal{U}^t + \alpha \gamma \left( \mathcal{D}^{4D} - \mathcal{B}^{4D} - \mathcal{T}^{4D} \right) \\ \mathcal{V}^{t+1} = \mathcal{V}^t + \alpha \sigma \left( \mathcal{T}^{4D} - \mathcal{E} \right) \end{cases} \tag{33}$$

where $\alpha$ is the step length.

The comprehensive procedural overview of the ADMM-based method for addressing the Cluster 4D-FCTN is presented in Algorithm 2.

**Algorithm 2:** Algorithm utilizing ADMM to solve the 4D-FCTN clustering problem.

**Input:** Original infrared tensor $\mathcal{D}^{4D}$. parameter $\lambda_1$, $\lambda_2$, $\tau$.
**Output:** Background tensor $\mathcal{B}^{4D}$, target tensor $\mathcal{T}^{4D}$, and noise tensor $\mathcal{N}^{4D}$.
**Initialization:** $\mathcal{B}^{4D} = \mathcal{D}^{4D}$, $\mathcal{T}^{4D}$, $\mathcal{N}^{4D}$, Multiplier of the Lagrangian function $\mathcal{X}_k^0 (k = 1, 2, \cdots, L)$, $\mathcal{U}$, $\mathcal{V}$, parameters $\omega_k (k = 1, 2, \cdots, L)$,$\gamma$.
Update $\mathbf{B}_k^{t+1}$ by Formula(27);
Update $\mathcal{E}^{t+1}$ by Formula (29);
Update $\mathcal{B}^{4Dt+1}$ and $\mathcal{T}^{4Dt+1}$ by Formula (31) and (32);
Update Lagrangian multiplies by Formula (33);
Verify the condition of convergence:
$\left\| \mathcal{D}^{4D} - \mathcal{B}^{4D^*} - \mathcal{T}^{4D^*} - \mathcal{N}^{4D^*} \right\|_F / \left\| \mathcal{D}^{4D} \right\|_F < \tau$;
Should the convergence criteria not be satisfied, increment t by 1 and proceed to Step 1.

In summary, Figure 6 depicts the comprehensive process of identifying infrared targets through the proposed model. The subsequent stages are expounded upon as follows:

(1) For a sequence of original images consisting of N frames, denoted as $\{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_N \in \mathbb{R}^{m \times n}\}$, we create three Cluster 4D infrared image tensors $\left(D^{4D}\right)_p$, $\left(D^{4D}\right)_l$, and $\left(D^{4D}\right)_f \in \mathbb{R}^{N_1 \times N_2 \times N_3 \times N_4}$ by organizing a sequence of $N_3$ frames in the correct temporal sequence using Algorithm 1, as described in Step 1 of Figure 6.

(2) For the three obtained 4D tensors, each is decomposed into target 4D tensors $\mathcal{T}^{4D}$, background 4D tensors $\mathcal{B}^{4D}$, and noise 4D tensors $\mathcal{N}^{4D}$ obtained through $\mathcal{D}^{4D}$ by Cluster 4D-FCTN decomposition using Algorithm 2, as described in Step 2 of Figure 6.

(3) We reconstruct the target images $\left\{ \mathbf{S}_1^T, \mathbf{S}_2^T, \mathbf{S}_3^T, \ldots, \mathbf{S}_{N_3}^T \in \mathbb{R}^{m \times n} \right\}$ from the 4D target tensor of the point feature $(\mathcal{T}^{4D})_p$, the 4D target tensor of the line feature $(\mathcal{T}^{4D})_l$, and the 4D target tensor of the flat feature $(\mathcal{T}^{4D})_f$ using the reverse process outlined in Algorithm 1, as described in Step 3 of Figure 6.

### *3.5. Complexity Analysis*

This section briefly touches upon the computational complexities associated with our proposed approach. To ensure uniformity, we denote the true dimensions of our 4D tensor as $M_1 \times M_2 \times M_3 \times M_4$. For Algorithm 2, we establish the constant $x_i$ as $x_i = \prod_{i=1}^{l} M_{\mathbf{n}_i}$, $y_i$ as $y_i = \prod_{i=l+1}^{N} M_{\mathbf{n}_i}$, $l = \lfloor N/2 \rfloor$, and $\mathbf{n}$ as the rearrangement of vector $(1, 2, \cdots, N)$. Thus, the update complexity of $\mathcal{B}_k$ is $\mathcal{O}\left( \sum_{i=1}^{L} x_i y_i \min(x_i, y_i) \right)^n$, and the updating of $\mathcal{B}^{4D}$, $\mathcal{T}^{4D}$, and $\mathcal{N}^{4D}$ are within the complexity of $\mathcal{O}\left( \prod_{i=1}^{N} M_i \right)$. In addition, the computing complexity of $\mathcal{X}_k$, $\mathcal{U}$, and $\mathcal{V}$ are $\mathcal{O}\left( \prod_{i=1}^{N} M_i \right)$.

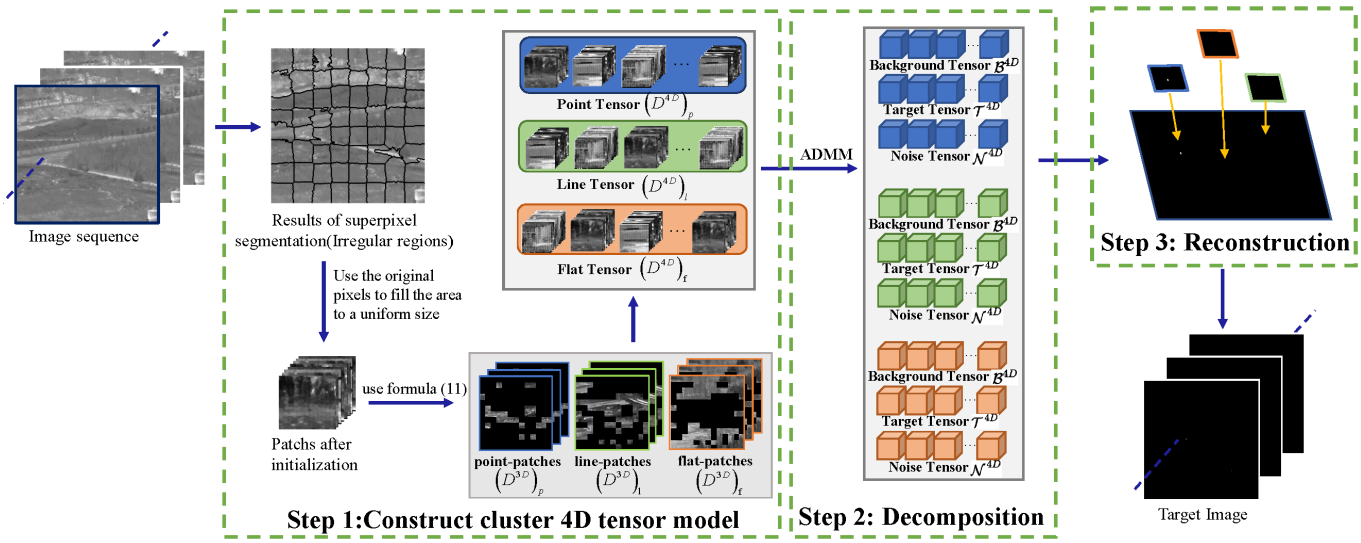



**Figure 6.** Overall procedure of the Cluster 4D−FCTN model.

## 4. Experiment and Results

The effectiveness of the proposed the Cluster 4D-FCTN method is assessed in this section through a series of experiments. Initially, we present the experimental setup, which includes evaluation criteria, datasets, and baseline techniques. Subsequently, various crucial parameters are examined and the resilience of our model is evaluated. Additionally, a comparative analysis is conducted between our method and other competitive approaches. Lastly, a series of ablation studies demonstrates that techniques elevating the dimensionality to 4D and conducting clustering computations yield more favorable outcomes compared to analogous methods based on 3D data and 4D approaches without clustering. This section additionally encompasses a discussion on limitations and future directions.

### *4.1. Data and Experiment Settings*

To provide a more objective assessment of the Cluster 4D-FCTN model's performance and facilitate comparisons with other models, we performed trials on five different sets of images captured in diverse environments, as listed in Table 1. The dataset utilized in our research comprised infrared images, as suggested by Hui et al. [56]. Figure 7 exhibits representative images captured from these sequences, featuring expanded target regions showcased in the upper right-hand corner of every picture. Sequences 1 and 2 exhibit significant local contrast variations while featuring background consisting of trees and man-made structures, posing a moderate detection challenge. Sequence 3 presents low contrast levels and noticeable background radiation, making it particularly challenging to detect targets. In Sequence 4, several radiation sources emitting high-intensity levels in the surrounding environment can significantly increase false alarm rates. On the other hand, Sequence 5 is relatively straightforward; however, it encompasses a large target area that necessitates algorithms capable of preserving target integrity.

**Table 1.** Data description (sizes in pixels).

| Seq. | Image Size | Frames | Average SCR | Target Description | Background Description |
|---|---|---|---|---|---|
| 1 | 256 × 256 | 300 | 0.83 | Single, tiny, low nonlocal contrast | Artifical structures, clutter |
| 2 | 256 × 256 | 200 | 6.08 | Small, slow-moving airplane | Village, reflective road, forest |
| 3 | 256 × 256 | 100 | 0.71 | Small, slow-moving airplane | village, reflective road, forest |
| 4 | 256 × 256 | 300 | 14.52 | Single, slow-moving airplane | Forest, bright rock, noise |
| 5 | 256 × 256 | 200 | 7.55 | Small, dim, regular shape | River, noise, heavy clutter |

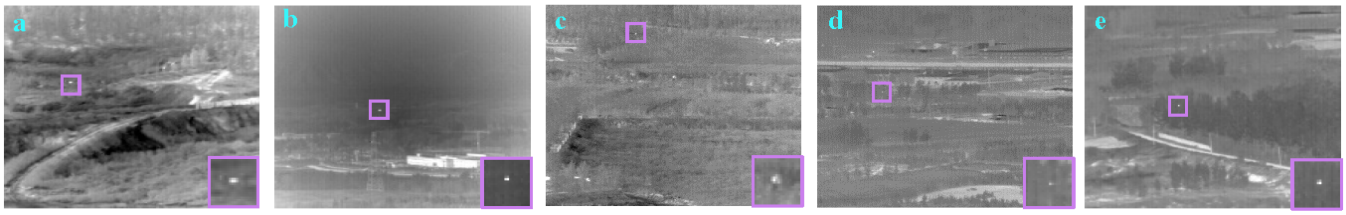



**Figure 7.** Images (**a**–**e**) represent five real sequences employed in the experiments.

### *4.2. Evaluation Metrics and Baselines*

In this study, a performance assessment of infrared target detection algorithms employing five prevalent metrics is presented. These measurements include the ROC curve for receiver operating characteristic, the SCRG, which represents the signal-to-clutter ratio gain, the LSNRG, which denotes the local signal-to-noise ratio gain, the BSF, indicating the background suppression factor, and CG, which represents the contrast gain:

$$\mathrm{SCRG} = \frac{\mathrm{SCR_{out}}}{\mathrm{SCR_{in}}}, \mathrm{BSF} = \frac{\sigma_{\mathrm{out}}}{\sigma_{\mathrm{in}}} \tag{34}$$

where $\mathrm{SCR_{in}}$ and $\mathrm{SCR_{out}}$ are the SCR values of the original image and the recovered target image. Similarly, the standard deviations between the original image and the reconstructed target image are represented by $\sigma_{\mathrm{out}}$ and $\sigma_{\mathrm{in}}$ . The ROC curves are depicted based on the false alarm ratio (Fa) and the probability of detection (Pd), as outlined below:

$$\mathrm{Fa} = \frac{\text{the quantity of identified false targets}}{\text{the total number of pixels throughout the entire image.}} \tag{35}$$

$$\mathrm{Pd} = \frac{\text{the quantity of detected true targets}}{\text{the aggregate count of authentic targets.}} \tag{36}$$

where Fa represents the abscissa and Pd denotes the ordinate.

In addition, we employ the LSNRG for evaluating the ability to enhance the target, quantified as follows:

$$\mathrm{LSNRG} = \frac{\mathrm{LSNR_{out}}}{\mathrm{LSNR_{in}}} \tag{37}$$

where $\mathrm{LSNR_{out}}$ and $\mathrm{LSNR_{in}}$ denote the respective LSNR values of the initial image and the modified image,

$$\mathrm{LSNR} = \frac{P_T}{P_B} \tag{38}$$

where $P_T$ and $P_B$ denote the highest pixel values of the surrounding area and the target, respectively. Lastly, we present the CG, which is expressed as follows:

$$\mathrm{CG} = \frac{\mathrm{CON_{out}}}{\mathrm{CON_{in}}} \tag{39}$$

where $\mathrm{CON_{out}}$ and $\mathrm{CON_{in}}$ represent the contrast of the initial and modified image, respectively, with CON defined as

$$\mathrm{CON} = |\sigma_{\mathrm{in}} - \sigma_{\mathrm{out}}|, \tag{40}$$

where $\sigma_{\mathrm{in}}$ and $\sigma_{\mathrm{out}}$ have identical values to those in Equation (34).

In our comparative analysis, we incorporated six state-of-the-art ISTD methods as benchmarks. Specifically, these methods included a matrix-based IPI approach [29], a tensor-singleframe-based RIPT technique [30], a PSTNN algorithm [32], a tensor sequence-based ECASTT method [35], an MFSTPT model [36], and SRSTT, a sparse regularization-based twist tensor model [37]. The detailed parameter configurations are presented in Table 2.

**Table 2.** Parameter settings of the seven comparison methods.

| Method | Parameters |
|---|---|
| IPI | Patch size: $m_1 \times n_1 : 50 \times 50$, sliding size: 10, $\lambda = 1/\sqrt{\min(m,n)}$, $\varepsilon = 1 \times 10^{-7}$ |
| RIPT | Patch size: $m \times n : 30 \times 30$, sliding size: 10, $\lambda = 1/\sqrt{\min(m_1,n_1)}$, $\varepsilon = 1 \times 10^{-7}$ |
| PSTNN | Patch size: $m_1 \times n_1 : 40 \times 40$, sliding size: 40, $\lambda = 1/\sqrt{\max(m_1,n_1) * n_3}$, $\varepsilon = 1 \times 10^{-7}$ |
| ECASTT | $\beta = 0.1$, t = 3, $\lambda_1 = 0.009$, $\lambda_2 = 5/\sqrt{\min(m,n) \times t}$, $\lambda_3 = 100$, $\varepsilon = 1 \times 10^{-7}$ |
| MFSTPT | Patch size: $m_1 \times n_1 : 60 \times 60$, sliding size: 60, $\lambda = 1/\sqrt{\max(n1,n2) * n3}$ |
| SRSTT | L = 30, $\lambda_1 = 0.05$, $\lambda_2 = 0.1$, $\lambda_3 = 100$, $\varepsilon = 1 \times 10^{-7}$, $\mu = 0.01$ |
| Ours | Patch size: $N_1 \times N_2$, temporal size: $N_3 = 20$, $\varepsilon = 1 \times 10^{-7}$ |

### *4.3. Parameter Settings*

The $\lambda_1$ parameter affects the resilience of different scenarios. A larger $\lambda_1$ is required to reduce false alarms in the target image, while a smaller $\lambda_1$ is needed to enhance the rate of identification for the target image. To evaluate the influence of $\lambda_1$ on the detection accuracy, we conducted tests and plotted ROC curves for different values of $\lambda_1$, which are presented in the second column of Figure 8. The results indicate that there is an initial improvement in detection ability as $\lambda_1$ increases, and that this trend reverses beyond a certain threshold. Thus, we imposed a constraint on the upper limit of $\lambda_1$ when it reached 2. We correlated it with the number of image patches to establish the following:

$$\lambda_1 = \min\left(\frac{L * \left\|N_p - N_{pl}\right\|}{\left\|N_p - N_{pp}\right\|}, 2\right) \tag{41}$$

where $N_p$ is the total number of patches, $N_{pl}$ is the number of patches of line type, and $N_{pp}$ is the number of patches of point type.

According to the findings presented in the third column of Figure 8, the ROC curves achieve their peak initial value when L is set at 0.6. Consequently, we opted for this value for L. The empirical setting for $\lambda_2$ was established as 100, following [35]. It has been documented that penalty factors escalate with the number of iterations. Consequently, we decide to set $\omega_k^*$ as $\sigma\omega_k$ and $\gamma^*$ as $\sigma\gamma$, where $\rho$ represents a numerical constant and its value corresponds to $\sigma = 1.2$. Furthermore, in our experiments both $\omega_k$ and $\gamma$ were assigned a value of $1 \times 10^{-4}$.

#### 4.3.1. Patch Size $N_1$

Traditional algorithms may need to test the size of the sliding window and the sliding step in order to choose an appropriate size. In order to enhance the robustness of the sliding window size and sliding step length, in this paper we propose an algorithm that draws inspiration from superpixel segmentation. This algorithm uses a specific sliding step size, and only requires initialization of the mesh size S. As depicted in the first column of Figure 8, images were individually tested with grid initialization sizes ranging from 5–50 using a uniform selection of ten parameters. It is evident from the ROC curve that this algorithm is not sensitive to variations in grid initialization size, even when scaling the horizontal coordinate down to the 105 differences observed among the ROC curves under different initialization sizes. This finding confirms that utilizing superpixel segmentation enhances the algorithm's resilience towards changes in window size.

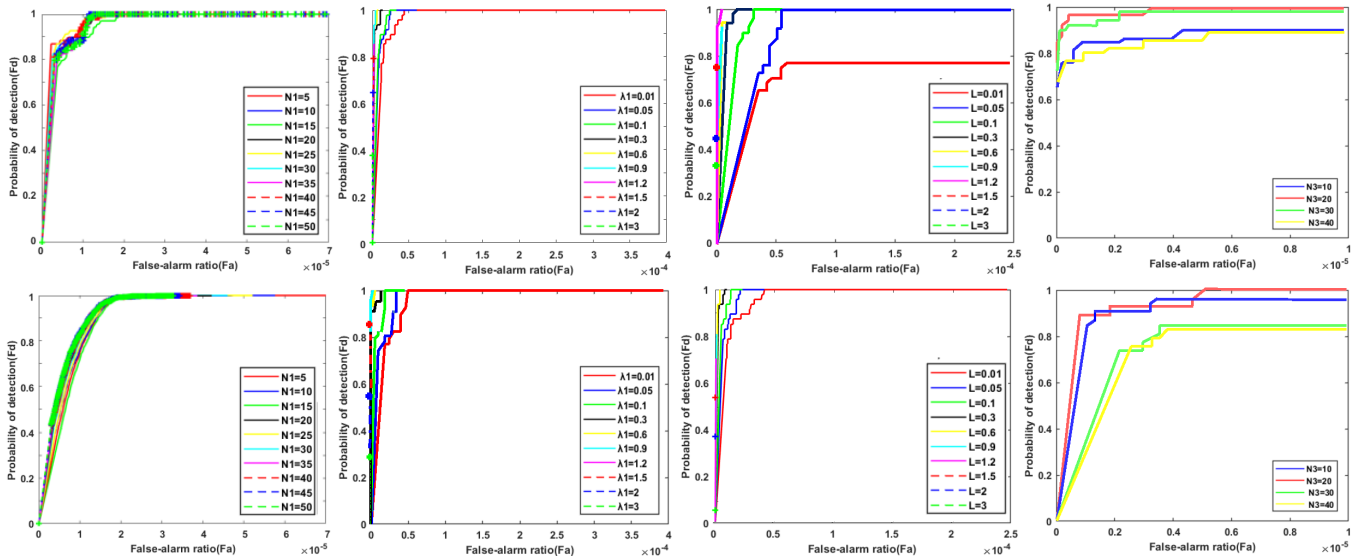



**Figure 8.** ROC curves for Sequences 1 and 4 with respect to varying parameters. The first row corresponds to Sequence 1, while the second row represents Sequence 4. Column 1 pertains to patch size $N_1$, Column 2 encompasses different $\lambda_1$ values, Column 3 shows various L values, and Column 4 concerns the temporal size $N_3$.

#### 4.3.2. Temporal Size $N_3$

The temporal dimension $N_3$ plays a crucial role in the construction and optimization of 4D tensors. Insufficient frames can lead to the loss of low-rank features and inefficient utilization of time domain data, while an excessive number of frames may result in increased time complexity, redundancy, repetition, and diminished time sparsity. To address this issue, we conducted experiments with varying values of L (10, 20, 30, and 40). The results are presented in the final column of Figure 8. Our findings indicate that G4D−FCTN performs optimally when $N_3$ is set to 20. Consequently, we chose $N_3$ as 20 for subsequent experiments.

### *4.4. Robustness of Scene Perception in Real-World and Synthetic Noisy Environments.*

In this section, we assess the robustness of the Cluster 4D-FCTN model across diverse environments and varying noise intensities.

#### 4.4.1. Robustness in Diverse Environmental Conditions

The first row (a)–(f) of Figure 9 displays four infrared (IR) images depicting different scenes and sizes: 320 × 240, 256 × 172, 252 × 213, 256 × 256, 320 × 240, and 320 × 240, respectively. Figure 9a,b represent sky scenes, Figure 9c,d depict land scenes, and Figure 9e,f show ocean scenes. The second row (g)–(l) corresponds to the test results for these images; red boxes have been used to highlight all objects. The clustered 4D−FCTN model accurately detects small targets across various scenes while effectively suppressing background interference. This observation underscores the versatility of this model in adapting to diverse scenarios and datasets of different sizes.

#### 4.4.2. Robustness against Noise

The application may result in noise presence within the acquired IR images, thereby necessitating the implementation of a test to evaluate the model's ability to cope with such noise. To address this, we extracted six images from Figure 9 and injected Gaussian white noise with standard deviations of 10 and 20. This led to the generation of the corresponding noise maps presented in the first and third rows of Figure 10 a–f. The second and fourth rows of Figure 10g–l display the recovered target images, which were obtained by applying the Cluster 4D-FCTN model to these noise maps. It is evident that

the introduction of noise significantly exacerbates image blurriness, making it increasingly challenging to discern target features. Nonetheless, in spite of these obstacles, our Cluster 4D-FCTN model demonstrates robust detection capabilities while maintaining overall stability throughout various robustness tests conducted on single-frame images alone. Moreover, the employment of image sequences yields even more accurate detection results, as demonstrated by these experiments. Consequently, it can be concluded that the Cluster 4D-FCTN algorithm is able to effectively process single-frame images.

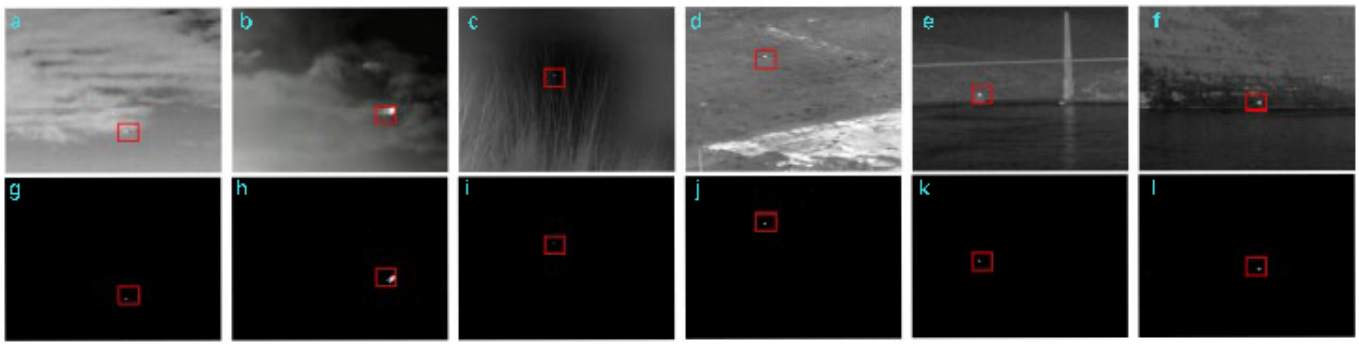



**Figure 9.** Different scenes (**a**–**f**) and results (**g**–**l**).

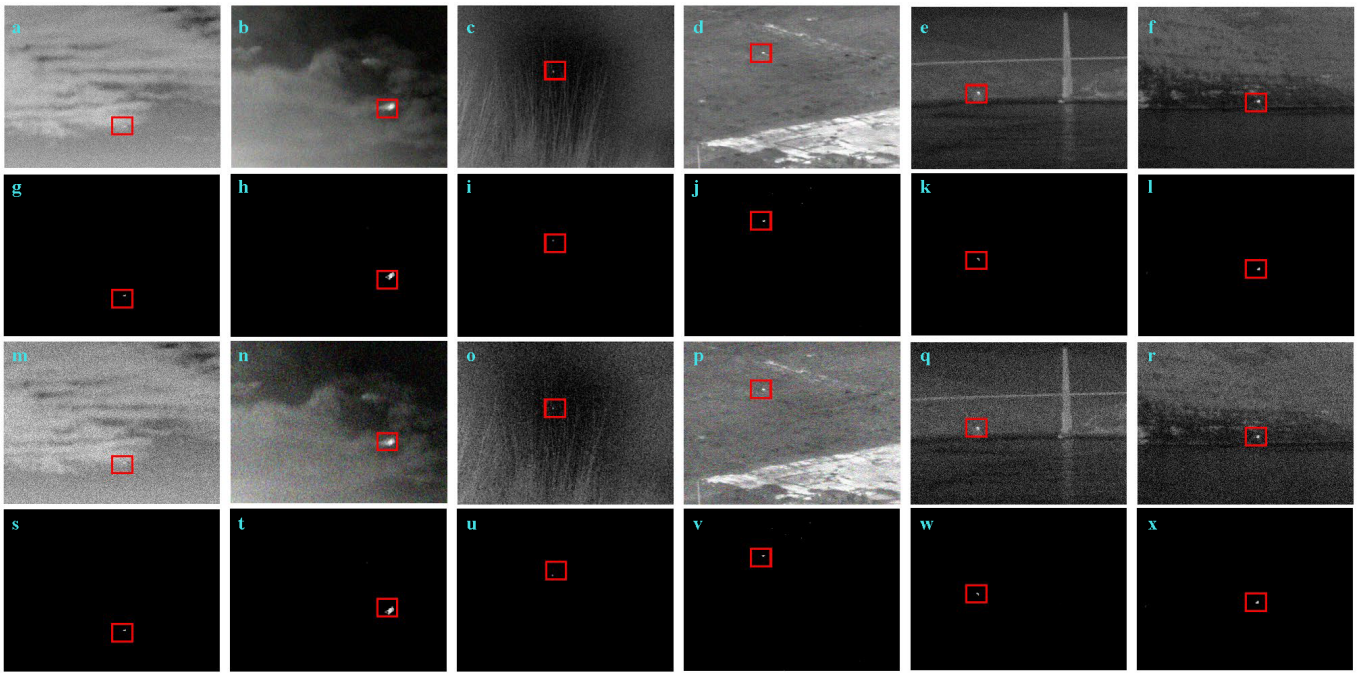



**Figure 10.** Scenarios and outcomes related to noise. (**a**–**f**) represents the white Gaussian noise image with standard deviation 10 added, and (**g**–**l**) represents its corresponding object detection results using the C4D–FCTN method.(**m**–**r**) represents the white Gaussian noise image with standard deviation 10 added, and (**s**–**x**) represents its corresponding object detection results using the C4D–FCTN method.

### *4.5. Comparison with Other Typical Methods*

#### 4.5.1. Visual Comparison

To provide a more objective performance evaluation of the Cluster 4D-FCTN model, we conducted tests employing various comparative algorithms on five distinct sequences of infrared images. Figure 11 displays the outcome images via three-dimensional surfaces.Additionally, the results for Sequences 1–5, which were not subjected to threshold segmentation, are presented in Figure 12. In the upper right corner, a magnified view (scaled by a factor of 2) accentuates the target zone, which is represented by a purple

rectangle. Moreover, instances of failed detection are indicated by a dotted purple box. The IPI model demonstrates effectiveness in detecting targets across various sequences while maintaining their shapes; however, it encounters challenges with non-smooth backgrounds, resulting in images with pronounced edge residuals. The reason of this problem is that the nuclear norm is not accurate enough to constrain the strong edge background; treating all singular values equally leads to uniform shrinking, thereby compromising crucial details of these backgrounds. Consequently, strong edges are incorporated into the detected target images. In comparison to matrix-based methods, the RIPT and PSTNN tensor-based models exhibit superior ability to suppress complex backgrounds; however, they struggle with accurately reconstructing target shapes, especially when faced with scenarios such as Sequences 1–5, which contain numerous sparse objects within the backgrounds. Notably, neither PSTNN nor RIPT could detect any targets in Sequence 4. Sequence-based methods exhibit superior ability to suppress the background in intricate environments compared to single-frame approaches. ECASTT, MFSTPT, and SRSTT exhibit better preservation of the target's edge shape. As observed in Sequences 1–3 and Sequence 4, the algorithms demonstrate accurate target detection. However, because of the relatively expansive size of the compact objectives in this series, the detected results partially suppress or magnify the targets. Moreover, MFSTPT and SRSTT fail to detect the target in Sequence 4. Our model and the IPI model exhibit the highest level of target shape retention integrity. Detecting artificial road structures in Sequences 1, 4, and 5 presents a challenge, in particular the presence of high-brightness background radiation in Sequence 2, which results in several methods retaining these backgrounds in their detection results. Among Sequences 1–5, only the Cluster 4D-FCTN model effectively suppresses clutter and noise across all frames. Sequence 4 poses the greatest difficulty, involving a small target with low contrast; the RIPT, PSTNN, and MFSTPT methods fail to detect the target, while the IPI and ECASTT methods retain significant noise areas. In contrast, our model effectively eliminates radiation sources with high-intensity background while striving to preserve the target's integrity. In general, our model exhibits exceptional performance in detecting targets and suppressing background in all five sequences.

#### 4.5.2. Qualitative Analysis

We conducted a quantitative assessment of the detection outcomes of the aforementioned methods across five sets of sequences utilizing the average values of SCRG, BSF, LSNRG, and CG calculated for each sequence. The resulting metric values are presented in Table 3, with the maximum value for each metric on each sequence highlighted in bold. It should be emphasized that if a single frame exhibits an Inf test value, this can result in an Inf average value for the entire sequence, encompassing the detection impact of other frames. Consequently, when calculating SCRG and BSF, we only consider them as Inf if the algorithm exhibits Inf values throughout the entire sequence; otherwise, we compute only the average of non-Inf values. Furthermore, Sequence 4 does not have targets detected by the PSTNN, RIPT, SRSTT, and MFSTPT methods; hence, they are excluded from the calculations. By definition, BSF evaluates the effectiveness of an algorithm in suppressing background across the entire image, while SCRG and CG assess its ability to enhance target presence and suppress background within local regions. Notably, the Cluster 4D-FCTN model demonstrates remarkably high BSF and SCRG values across all test sequences, indicating its exceptional capability to effectively remove background structures both locally and globally. Compared to its remarkable capacity for background suppression, our Cluster 4D-FCTN model demonstrates slightly inferior performance in target enhancement. The SRSTT model exhibits impressive outcomes in CG, particularly for Sequences 3–4. Nevertheless, when considering all algorithms across time, the Cluster 4D-FCTN model surpasses others in terms of overall evaluation based on the CG and LSNRG metrics. By integrating these three evaluation metrics, it becomes evident that the proposed Cluster 4D-FCTN model achieves superior outcomes by effectively suppressing background and enhancing targets, particularly in terms of background suppression.

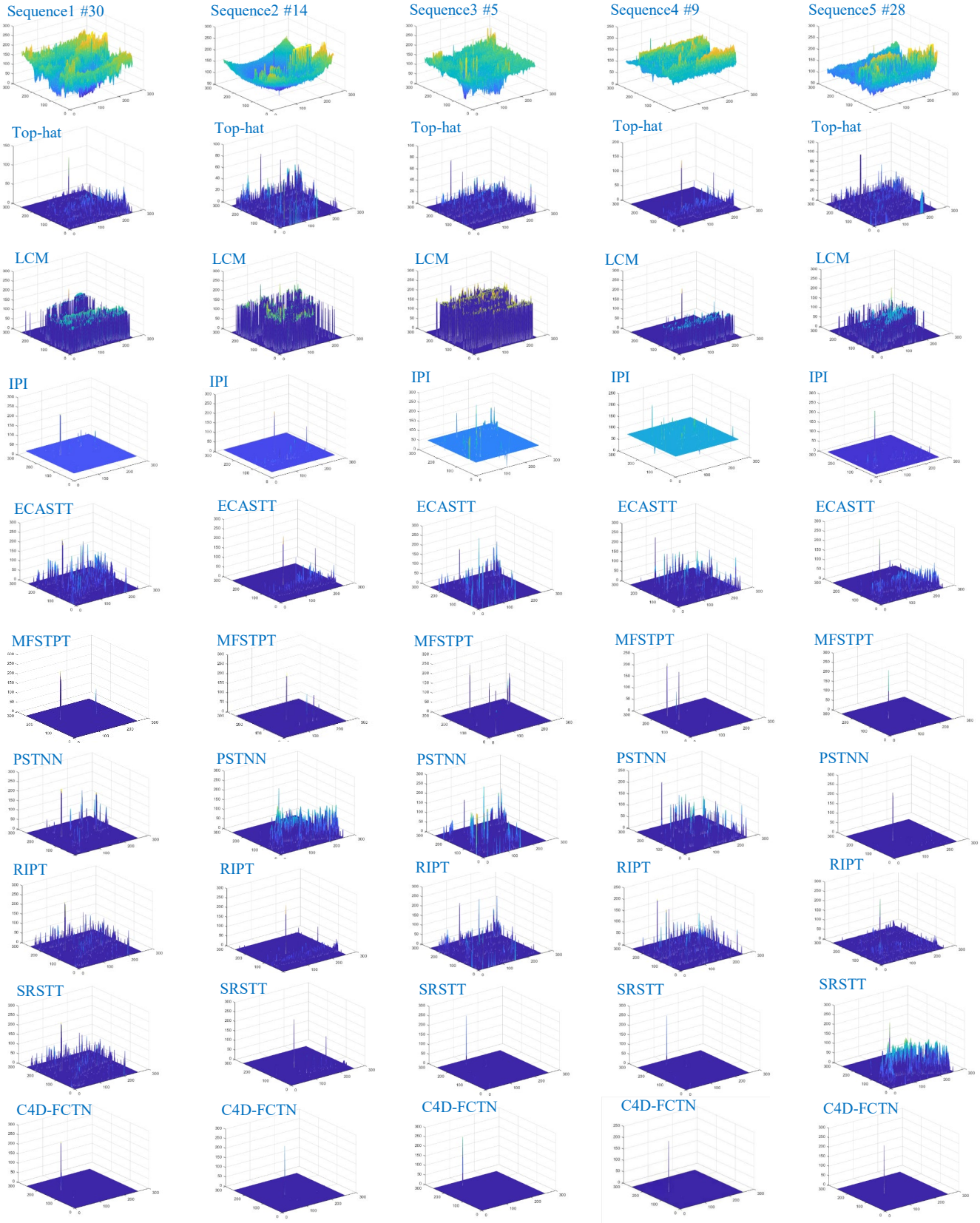



**Figure 11.** Different testing methods were employed to detect the results on Sequences 1 to 5 in the three-dimensional surface.

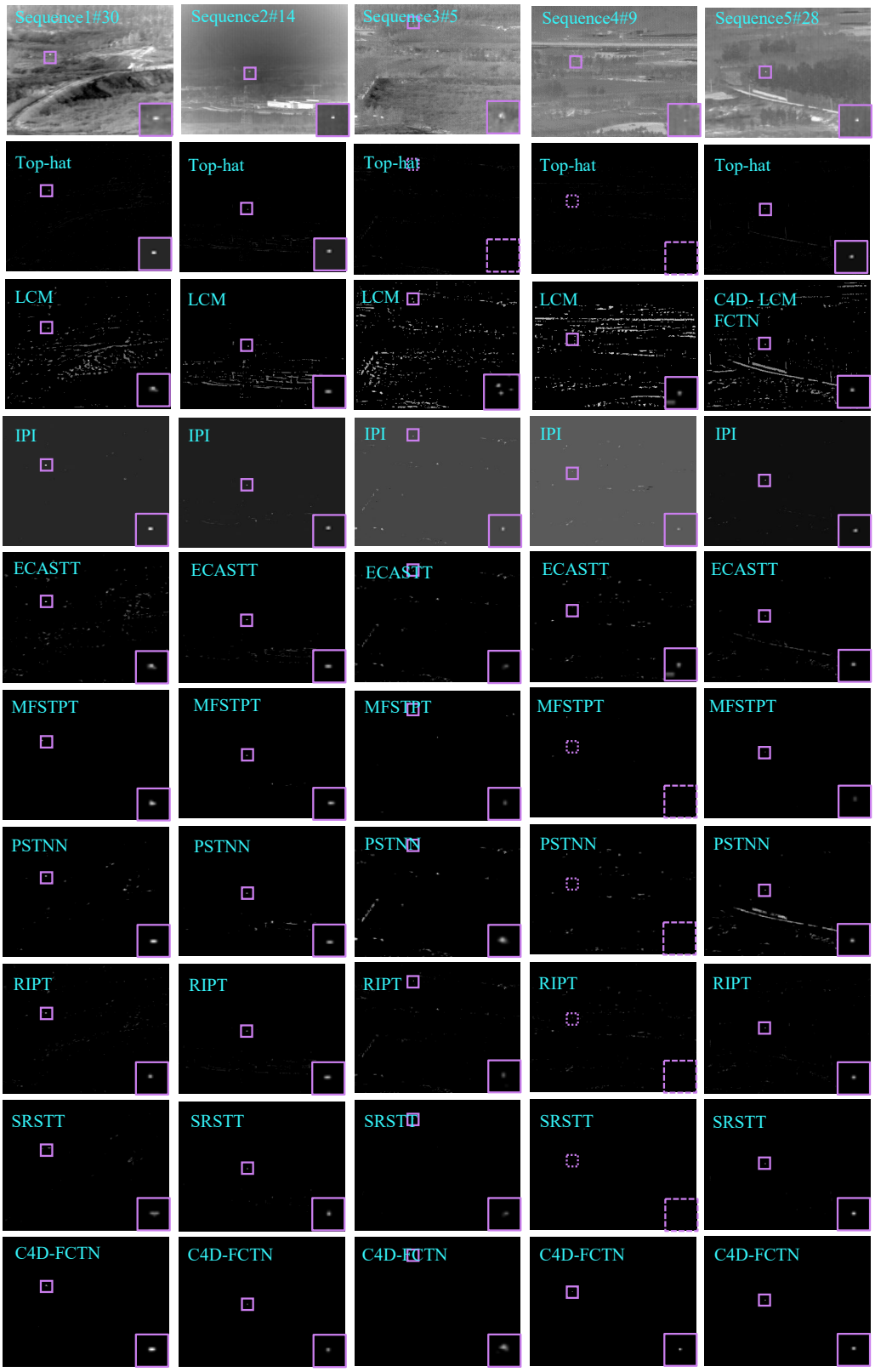



**Figure 12.** Detection outcomes for Sequences 1 to 5 employing seven different testing methodologies. The target is represented by a purple box, while a detection failure is indicated by a dotted box.

The performance evaluation of our algorithm in complex and noisy scenarios is illustrated by the five representative ROC curves shown in Figure 13, corresponding to Sequence 1 to 5. Our C4D-FCTN method consistently produces stable results even under challenging scenes, positioning itself relatively close to the top left corner, particularly evident in Sequence.4 and Sequence.5 where the presence of target-like objects may result in high false positive rates. Notably, single-frame methods such as RIPT and PSTNN fail when dealing with dark and small targets. In contrast, multi-frame methods outperform their single-frame counterparts across all sequences (Sequence.1–Sequence.5), highlighting the significance of temporal information for background noise suppression and object detection within sequence-based approaches. Nevertheless, some multi-frame methods like SRSTT and MFSTPT exhibit detection failures due to suboptimal utilization of time domain information, further emphasizing the effectiveness of our four-dimensional clustering tensor construction approach.

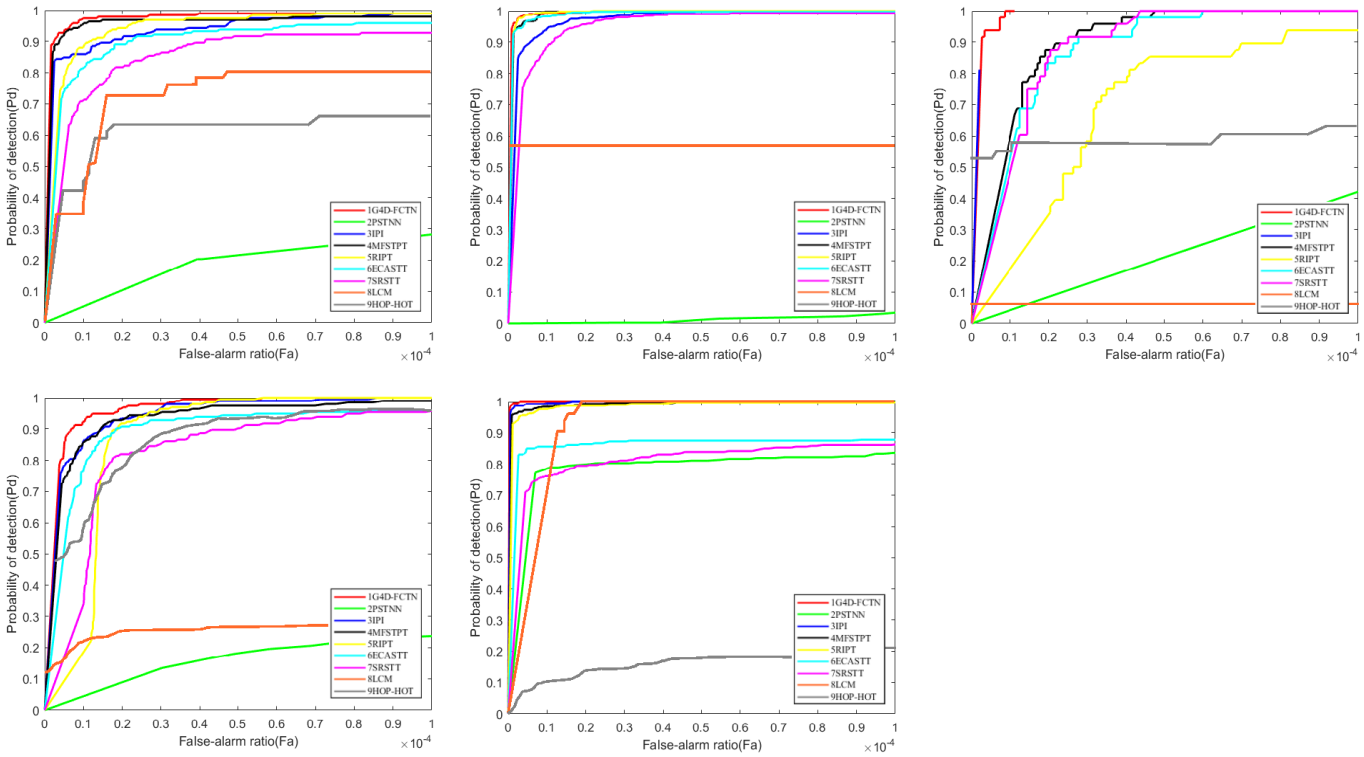



**Figure 13.** ROC curves of the compared and proposed methods on different sequences.

Figure 12 illustrates the assessment of various algorithms in intricate and turbulent surroundings (Sequences 1 through 5), with all of them exhibiting acceptable performance in simple scenarios. Our method consistently generates stable results in demanding situations, positioning itself proximate to the uppermost left corner, particularly in Sequences 4 and 5 in locations with objects resembling the target that might elevate false positive rates. Certain single-frame techniques, such as RIPT and PSTNN, manifest commendable detection outcomes on Sequences 1 and 2. Nonetheless, when confronting dim and small targets, strategies based on multiple frames surpass their single-frame counterparts on Sequence 3 to 5, precisely where our proposed technique outperforms them. This finding suggests that incorporating temporal information bolsters sequence-based methods in reducing background noise and augmenting target detection capabilities.

**Table 3.** Performance comparison of seven methods in terms of SCGR, BSF, CG, and LSNRG.

| Method | Seq1 SCRG | Seq1 BSF | Seq2 SCRG | Seq2 BSF | Seq3 SCRG | Seq3 BSF | Seq4 SCRG | Seq4 BSF | Seq5 SCRG | Seq5 BSF |
|---|---|---|---|---|---|---|---|---|---|---|
| IPI | 7.63 | 2.38 | 8.63 | 7.63 | 5.30 | 2.87 | 10.62 | 5.62 | 25.64 | 7.17 |
| ECASTT | 60.16 | 15.81 | 15.29 | 23.47 | 11.91 | 26.49 | 11.34 | 9.75 | **84.64** | 14.45 |
| MFSTPT | 24.60 | 19.96 | 102.79 | 295.14 | 13.51 | 207.24 | 1.02 | 0.69 | 34.68 | 12.45 |
| PSTNN | 26.17 | 23.30 | 0.95 | 59.55 | 82.49 | 10.75 | 274.37 | 11.26 | 6.45 | 25.45 |
| RIPT | 6.71 | 1.32 | 10.41 | 2.18 | 4.61 | 4.21 | 9.42 | 1.92 | 15.14 | 9.65 |
| SRSTT | 56.96 | **262.48** | 346.98 | 245.78 | 206.97 | 94.52 | 49.09 | 98.97 | 64.52 | 246.54 |
| Ours | **70.17** | 221.32 | **408.55** | **1330.8** | **860.65** | **301.44** | **767.19** | **374.39** | 56.96 | **285.62** |

| Method | Seq1 CG | Seq1 LSNRG | Seq2 CG | Seq2 LSNRG | Seq3 CG | Seq3 LSNRG | Seq4 CG | Seq4 LSNRG | Seq5 CG | Seq5 LSNRG |
|---|---|---|---|---|---|---|---|---|---|---|
| IPI | 1.27 | 2.75 | **19.54** | 1.43 | 1.10 | 1.22 | 1.90 | 2.73 | 1.23 | **17.08** |
| ECASTT | 0 | 32.67 | 1.35 | 1.30 | 1.21 | 1.12 | 0 | 2.87 | 1.62 | 4.44 |
| MFSTPT | 1.57 | 1.03 | 1.07 | 4.60 | 1.09 | 2.31 | 1.57 | 0 | 1.05 | 28.38 |
| PSTNN | 0.11 | 2.87 | 0.18 | 0.07 | 0.03 | 0.05 | 0.11 | 0 | 0.84 | 0.86 |
| RIPT | 2.31 | 2.87 | 1.40 | 2.04 | 1.25 | 1.43 | 2.30 | 0 | 1.31 | 4.27 |
| SRSTT | 2.83 | 47.08 | 1.28 | 20.01 | **2.81** | **2.89** | **2.87** | 0 | 1.56 | 8.36 |
| Ours | **3.75** | **68.45** | 2.25 | **35.64** | 1.14 | 0.98 | 2.32 | **47.08** | **1.60** | 4.62 |

#### 4.5.3. Comparison of Time Required for Computation

The runtime (per frame) of the seven algorithms across the five sequences is presented in Table 4. PSTNN emerges as the most proficient tensor-based method. Our approach exhibits a reduced computational time in the spatial–temporal tensor-domain.

**Table 4.** Comparison of average computing time (in seconds) of the seven methods.

| Method | Seq1 | Seq2 | Seq3 | Seq4 | Seq5 |
|---|---|---|---|---|---|
| IPI | 10.53 | 11.706 | 12.272 | 17.341 | 23.929 |
| ECASTT | 9.205 | 14.823 | 15.006 | 16.087 | 13.821 |
| MFSTPT | 0.956 | 1.103 | 1.035 | 1.324 | 0.998 |
| PSTNN | 0.113 | **0.125** | **0.213** | 0.341 | **0.245** |
| RIPT | 2.800 | 2.341 | 2.515 | 2.641 | 3.069 |
| SRSTT | 18.625 | 17.633 | 19.562 | 20.065 | 17.326 |
| Ours | **0.351** | 0.247 | 0.650 | **0.324** | 0.231 |

### *4.6. Ablation Study*

To demonstrate the efficacy and value of our dimensionality expansion and clustering strategies, we conducted an ablation study to evaluate the contributions of the Cluster 4D−FCTN model. We compared it to two other approaches: a spatial–temporal tensor construction-based 3D model (3DST−TCTN) akin to [57], and a model constructed by spatial–temporal tensor without clustering (4D−FCTN). The experiments were carried out on Sequences 1 to 5 with individually optimized parameter settings. As illustrated in Figure 14, the detection performance of Cluster 4D−FCTN surpasses both comparative methods in terms of ROC performance. Moreover, although 4D−FCTN achieves notable results across most datasets, the overall performance of Cluster 4D−FCTN outperforms both it and the 3D-based spatial–temporal tensor models. In conclusion, the integration of four-dimensional tensors provides enhanced detection capability in four-dimensional fields; however, our cluster four-dimensional tensor model demonstrates superior handling of ISTD tasks. Therefore, this experiment provides compelling evidence indicating that the employment of clustering calculations on four-dimensional data is both practical and highly effective.

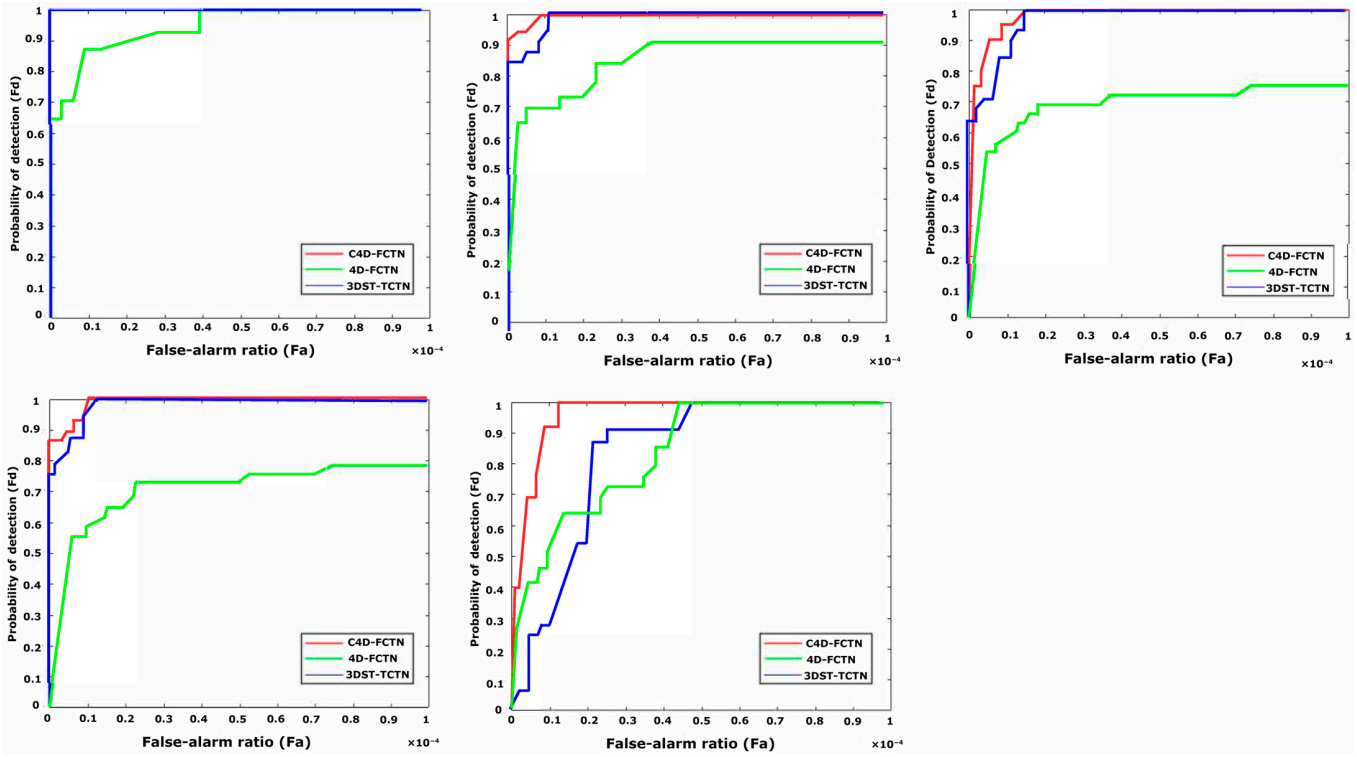



**Figure 14.** Detection results of the ablation study, showing ROC curves for 3DST−TCTN, 4D−FCTN, and C4D−FCTN on Sequences 1 to 5.

## 5. Discussion

Infrared detection systems exhibit superior stealth capabilities and robust anti-jamming resistance, making them extensively applicable in both the military and civilian sectors. Infrared dim and small target detection constitutes a crucial component of these detection systems, and has emerged as a focal point of current research.

Considering that the background of infrared images has nonlocal autocorrelation and that the target occupies only a small number of pixels, scholars have proposed the application of low-rank sparse decomposition to the field of infrared small target detection. In this approach, the low rank of the background and the sparsity of the target are used to transform the target detection problem into a mathematical optimization problem to restore low-rank components and sparse components. This can satisfy the global prior of a low-rank sum target that is sparse in both simple and complex backgrounds. In recent years, a method of fusing local prior and temporal information to construct a low-rank sparse decomposition model that can effectively suppress false alarms has been developed. This method based on low-rank sparse decomposition is more suitable for complex and changeable scenes, and has better detection performance than other methods.

Although infrared dim and small target detection algorithms based on low-rank sparse decomposition have achieved a series of promising research results, there are several challenges that demand further study, mainly reflected in the following four aspects:

1. The method based on low-rank sparse decomposition involves many parameters, such as the size of the sliding window, the moving step of the sliding window, the weight coefficient in the objective function, and more. These parameters have an important impact on the performance of the algorithm.
2. At present, the construction of the tensor is relatively simple, generally using the sliding window to traverse the original infrared image or multi-frame sequence images to construct the tensor directly, which involves a large amount of redundant information that has nothing to do with the target.
3. The background tensor rank approximation is not accurate.

4. When utilizing time-domain information, existing algorithms based on low-rank sparse decomposition mainly fuse the time domain information into the image matrix or tensor data, and use multi-frame images instead of the original single-frame images to construct new data. However, the sequence information between the frames is not fully mined.

## 6. Conclusions

To address the issues related to sensitivity of sliding window size, inadequate utilization of time series data, and imprecise approximation of tensor rank in current methods for detecting infrared dim and small targets, this study introduces a novel approach based on superpixel segmentation and statistical clustering to construct a four-dimensional tensor model. Our model diverges from existing approaches that rely solely on information within the spatial domain at either a local or global level. Instead, our method utilizes both temporal and spatial information, and uses their correlation to construct a new model. By organizing 3D patches with the same feature type into the same group, we effectively exploit temporal characteristics; moreover, we connect local patches in sequential order to further capitalize on these temporal properties.

In addition, we propose a new tensor decomposition, called the Fully Connected Tensor Network (FCTN) decomposition, which facilitates the decomposition of an nth-order tensor into a sequence of low-dimensional nth-order factors. It breaks through the limitation that the exclusive role of TT and TR decomposition lies in the disclosure of the operation limited to two consecutive factors rather than the connection between any two factors, resulting in limited ability to measure the correlation. These issues are effectively resolved by leveraging the ADMM framework. In this paper, we extensively evaluate our methods on several datasets containing infrared sequences, employing five cutting-edge algorithms. Our results showcase superior detection accuracy and robustness in diverse scenarios compared to existing approaches. Moreover, a series of ablation studies demonstrate that the proposed techniques elevating the dimensionality to 4D and conducting clustering computations yield more favorable outcomes compared to analogous methods based on 3D data and 4D approaches without clustering. Finally, the Section 5 encompasses a discussion on limitations and future directions. Notably, the algorithm occasionally detects target-like items in fast-moving complex environments due to the absence of constraints on prior information regarding the target. Therefore, our future work aims to incorporate the prior knowledge of objects contained in the spatiotemporal features in order to facilitate object detection and mitigate false positives.

**Author Contributions:** W.W. was responsible for the conception and design of the experiments. W.W. conducted the experiments and drafted the manuscript. T.M. critically reviewed the manuscript and made revisions. M.L. provided guidance and edited the manuscript. H.Z. contributed computational resources and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Bingwei Hui et al., "A dataset for infrared image dim-small aircraft target detection and tracking under ground/air background". Science Data Bank, 28 October 2019. Available Online: https://doi.org/10.11922/sciencedb.902 (accessed on 2 December 2022).

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

## References

1. 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]
2. Ren, X.; Wang, J.; Ma, T.; Zhu, X.; Bai, K.; Wang, J. Review on infrared dim and small target detection technology. *J. Zhengzhou Univ. (Nat. Sci. Ed.)* **2020**, *52*, 1–21.
3. Bai, X.; Bi, Y. Derivative entropy-based contrast measure for infrared small-target detection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2452–2466. [CrossRef]

4. Li, J.; Zhang, P.; Wang, X.; Huang, S. Infrared small-target detection algorithms: A survey. *J. Image Graph.* **2020**, *25*, 1739–1753.
5. Jinhui, H.; Yantao, W.; Zhenming, P.; Qian, Z.; Yaohong, C.; Yao, Q.; Nan, L. Infrared dim and small target detection: A review. *Infrared Laser Eng.* **2022**, *51*, 20210393-1.
6. Liu, C.; Wang, H. Research on infrared dim and small target detection algorithm based on low-rank tensor recovery. *J. Syst. Eng. Electron.* **2023**, *34*, 861–872.
7. Yang, W.P.; Shen, Z.K. Small target detection and preprocessing technology in infrared image sequences. *Infrared Laser Eng.* **1998**, *27*, 23–28.
8. Deshpande, S.D.; Er, M.H.; Venkateswarlu, R.; Chan, P. Max-mean and max-median filters for detection of small targets. In *Signal and Data Processing of Small Targets 1999 Denver, CO, United States*; SPIE: Bellingham, DC, USA, 1999, Volume 3809, pp. 74–83.
9. Yu, N.; Wu, C.Y.; Tang, X.Y.; Li, F.M. Adaptive background perception algorithm for infrared target detection. *Acta Electonica Sin.* **2005**, *33*, 200.
10. Liu, Y.; Tang, X.; Li, Z. A new top hat local contrast based algorithm for infrared small target detection. *Infrared Technol.* **2015**, *37*, 544–552.
11. Bi, Y.; Chen, J.; Sun, H.; Bai, X. Fast detection of distant, infrared targets in a single image using multiorder directional derivatives. *IEEE Trans. Aerosp. Electron. Syst.* **2019**, *56*, 2422–2436. [CrossRef]
12. 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]
13. Dong, L.; Wang, B.; Zhao, M.; Xu, W. Robust infrared maritime target detection based on visual attention and spatiotemporal filtering. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3037–3050. [CrossRef]
14. Han, J.; Liu, C.; Liu, Y.; Luo, Z.; Zhang, X.; Niu, Q. Infrared small target detection utilizing the enhanced closest-mean background estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 645–662. [CrossRef]
15. Yang, L.; Yang, J.; Yang, K. Adaptive detection for infrared small target under sea-sky complex background. *Electron. Lett.* **2004**, *40*, 1. [CrossRef]
16. Qi, S.; Ma, J.; Li, H.; Zhang, S.; Tian, J. Infrared small target enhancement via phase spectrum of quaternion Fourier transform. *Infrared Phys. Technol.* **2014**, *62*, 50–58. [CrossRef]
17. Sheng, Z.; Jian, L.; Jinwen, T. Research of SVM-based infrared small object segmentation and clustering method. *Signal Process.* **2005**, *21*, 515–519.
18. 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]
19. Zhao, J.; Tang, Z.; Yang, J.; Liu, E. Infrared small target detection using sparse representation. *J. Syst. Eng. Electron.* **2011**, *22*, 897–904. [CrossRef]
20. Liu, D.; Li, Z.; Liu, B.; Chen, W.; Liu, T.; Cao, L. Infrared small target detection in heavy sky scene clutter based on sparse representation. *Infrared Phys. Technol.* **2017**, *85*, 13–31. [CrossRef]
21. Xu, K.; Hu, W.; Zhou, W.; Zheng, H. Target detection based on the artificial neural network technology. In Proceedings of the 2006 9th International Conference on Control, Automation, Robotics and Vision, Orlando, FL, USA, 15–19 May 2006; pp. 1–5.
22. Ma, Q.; Zhu, B.; Cheng, Z.; Zhang, Y. Detection and recognition method of fast low-altitude unmanned aerial vehicle based on dual channel. *Acta Opt. Sin.* **2019**, *39*, 105–115.
23. Hou, Q.; Wang, Z.; Tan, F.; Zhao, Y.; Zheng, H.; Zhang, W. RISTDnet: Robust infrared small target detection network. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
24. 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]
25. Zhang, X.; Ren, K.; Wan, M.; Gu, G. Infrared small target tracking based on sample constrained particle filtering and sparse representation. *Infrared Phys. Technol.* **2017**, *87*, 72–82.
26. Fan, X.; Li, J.; Chen, H.; Min, L.; Li, F. Dim and small target detection based on improved hessian matrix and F-Norm collaborative filtering. *Remote Sens.* **2022**, *14*, 4490. [CrossRef]
27. Liu, D.; Li, Z.; Wang, X.; Zhang, J. Moving target detection by nonlinear adaptive filtering on temporal profiles in infrared image sequences. *Infrared Phys. Technol.* **2015**, *73*, 41–48. [CrossRef]
28. Fan, X.; Li, J.; Min, L.; Feng, L.; Yu, L.; Xu, Z. Dim and Small Target Detection Based on Energy Sensing of Local Multi-Directional Gradient Information. *Remote Sens.* **2023**, *15*, 3267. [CrossRef]
29. Gao, C.; Meng, D.; Yang, Y.; Wang, Y.; Zhou, X.; Hauptmann, A.G. Infrared patch-image model for small target detection in a single image. *IEEE Trans. Image Process.* **2013**, *22*, 4996–5009. [CrossRef] [PubMed]
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]
31. 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]
32. Zhang, L.; Peng, Z. Infrared small target detection based on partial sum of the tensor nuclear norm. *Remote Sens.* **2019**, *11*, 382. [CrossRef]
33. Cao, Z.; Kong, X.; Zhu, Q.; Cao, S.; Peng, Z. Infrared dim target detection via mode-k1k2 extension tensor tubal rank under complex ocean environment. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 167–190. [CrossRef]

34. 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]
35. 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]
36. 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. [CrossRef]
37. Li, J.; Zhang, P.; Zhang, L.; Zhang, Z. Sparse Regularization-Based Spatial–Temporal Twist Tensor Model for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–17. [CrossRef]
38. Wu, F.; Yu, H.; Liu, A.; Luo, J.; Peng, Z. Infrared Small Target Detection Using Spatio-Temporal 4D Tensor Train and Ring Unfolding. *IEEE Trans. Geosci. Remote Sens.* **2023**.
39. Sun, Y.; Yang, J.; Long, Y.; An, W. Infrared small target detection via spatial-temporal total variation regularization and weighted tensor nuclear norm. *IEEE Access* **2019**, *7*, 56667–56682. [CrossRef]
40. 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]
41. Liu, Y.Y.; Zhao, X.L.; Song, G.J.; Zheng, Y.B.; Huang, T.Z. Fully-connected tensor network decomposition for robust tensor completion problem. *arXiv* **2021**, arXiv:2110.08754.
42. Vaswani, N.; Bouwmans, T.; Javed, S.; Narayanamurthy, P. Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery. *IEEE Signal Process. Mag.* **2018**, *35*, 32–55. [CrossRef]
43. 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]
44. Grasedyck, L.; Hackbusch, W. An introduction to hierarchical (H-) rank and TT-rank of tensors with examples. *Comput. Methods Appl. Math.* **2011**, *11*, 291–304. [CrossRef]
45. Wang, W.; Aggarwal, V.; Aeron, S. Efficient low rank tensor ring completion. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5697–5705.
46. Liu, J.; Musialski, P.; Wonka, P.; Ye, J. Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 208–220. [CrossRef]
47. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 5249–5257.
48. Oseledets, I.V. Tensor-train decomposition. *SIAM J. Sci. Comput.* **2011**, *33*, 2295–2317. [CrossRef]
49. Yu, J.; Li, C.; Zhao, Q.; Zhao, G. Tensor-ring nuclear norm minimization and application for visual: Data completion. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3142–3146.
50. Huang, H.; Liu, Y.; Long, Z.; Zhu, C. Robust low-rank tensor ring completion. *IEEE Trans. Comput. Imaging* **2020**, *6*, 1117–1126. [CrossRef]
51. Zheng, Y.B.; Huang, T.Z.; Zhao, X.L.; Zhao, Q.; Jiang, T.X. Fully-connected tensor network decomposition and its application to higher-order tensor completion. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 11071–11078.
52. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef] [PubMed]
53. Di Zenzo, S. A note on the gradient of a multi-image. *Comput. Vision Graph. Image Process.* **1986**, *33*, 116–125. [CrossRef]
54. Förstner, W.; Gülch, E. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, 2–4 June 1987; Volume 6, pp. 281–305.
55. 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]
56. Hui, B.; Song, Z.; Fan, H.; Zhong, P.; Hu, W.; Zhang, X.; Lin, J.; Su, H.; Jin, W.; Zhang, Y.; et al. A dataset for infrared image dim-small aircraft target detection and tracking under ground/air background. *Sci. Data Bank* **2019**, *5*, 12.
57. Wang, G.; Tao, B.; Kong, X.; Peng, Z. Infrared small target detection using nonoverlapping patch spatial–temporal tensor factorization with capped nuclear norm regularization. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5001417. [CrossRef]


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