

# IMNN-LWEC: A Novel Infrared Small Target Detection based on Spatial-Temporal Tensor Model

Yuan Luo, Xiaorun Li, Shuhan Chen, *Member, IEEE*, Chaoqun Xia, Liaoying Zhao

**Abstract**—Despite many state-of-the-art methods have been proposed for infrared (IR) small target detection, target detectability (TD) and background suppressibility (BS) cannot be significantly improved simultaneously, especially in complex situations. This paper proposes a novel IR small target detection method named improved multi-mode nuclear norm joint local weighted entropy contrast (IMNN-LWEC), which represents the IR target detection task as an optimization problem for tensor decomposition of three components in spatial-temporal domain, including background tensor, target tensor and sparse structure tensor. First, to utilize the spatial and temporal information in an IR sequence effectively, we transform the original IR sequence into a nonoverlapping spatial-temporal patch-tensor. Second, a nonconvex approximation of tensor rank called improved multi-mode weighted tensor nuclear norm (IMWTNN) is proposed to estimate background tensor rank, which is of benefit to separate the background component more completely from the original image. Third, based on structure tensor theory, we introduce a new sparse prior map called local weighted entropy contrast (LWEC) via a designed image entropy operator and a new prior information filter, which can further preserve the target and suppress the background simultaneously. Besides, a novel tube-wise sparse regularization term is designed to identify linear sparse structures. And Frobenius norm is used to characterize noise. Finally, to solve the proposed model, an efficient optimization scheme utilizing the alternating direction multiplier method (ADMM) is designed to retrieve the small targets. Comprehensive experiments on five datasets witness the superior TD and BS performance of the proposed method compared with nine state-of-the-art detection methods.

**Index Terms**—Infrared (IR) small target detection, spatial-temporal patch-tensor, improved multi-mode weighted tensor nuclear norm (IMWTNN), local weighted entropy contrast (LWEC), alternating direction multiplier method (ADMM), target detectability (TD), background suppressibility (BS).

## I. INTRODUCTION

INFRARED search and tracking (IRST) system has widespread applications in many aspects, such as buried landmine detection, low-altitude security, missile tracking and other military and civilian fields [1]–[5]. The robust and

Manuscript received September 18, 2022; revised November 23, 2022; accepted December 14, 2022. This work was supported in part by the National Natural Science Foundation of China (Grant No.62171404), and in part by the Joint Fund Project of Chinese Ministry of Education (Grant No.8091B022120). (Corresponding author: Xiaorun Li.)

Yuan Luo, Xiaorun Li and Shuhan Chen are with the College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: lyhzxs@zju.edu.cn; lxr@zju.edu.cn; 11410057@zju.edu.cn).

Chaoqun Xia is with the College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325035, China (e-mail: auto\_xia@foxmail.com).

Liaoying Zhao is with the Institute of Computer Application Technology, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: zhaoly@hdu.edu.cn).

efficient detection of infrared (IR) dim and small targets plays a significant role in IRST systems. However, the size of small targets in IR images generally ranges from  $2 \times 2$  to  $9 \times 9$  pixels [6], lacking concrete shapes, texture details and other structure information because of standoff imaging. Meanwhile, due to the complex imaging environment, the intricate backgrounds of IR images are usually mixed with kinds of interference, such as dense clouds, ocean clutters and so on. And the small targets with a low signal-to-clutter ratio (SCR) are usually submerged in the background. Therefore, it is urgent to use the characteristics of the target as well as the relationship between the background and the small target to detect the IR dim and small targets.

Currently, a variety of approaches have been proposed to detect IR dim and small targets in complex background, including traditional detection methods and deep learning-based detection methods. The former can be divided into single-frame detection methods and sequential detection methods. The temporal information of IR images in spatial-temporal domain is generally ignored in single-frame detection methods, while the current multiframe methods cannot achieve excellent performance in the complicated background with stubborn noise and clutters. Therefore, this paper aims to design an efficient algorithm, which can enhance targets and suppress background simultaneously, and achieve satisfactory detection performance.

## A. Related Works

We have mentioned that traditional detection methods and deep learning-based methods are used for infrared small target detection. According to the number of the input IR image frames, IR dim and small target detection methods are divided into two categories: single-frame-based detection methods relying on only the spatial information and multiframe-based sequential detection methods with both intraframe spatial information and interframe temporal information.

1) *Single-frame-Based Detection Methods*: Traditional single-frame-based detection methods assume that the target distorts local correlation of the background, and filters are designed to distinct backgrounds and targets. [7]–[11] are typical detection methods. In fact, these methods enhance not only IR small targets but also the edges of some structures (e.g., architecture edges), raising false alarms.

Considering that the IR small target is more visually salient than its surrounding background, Kim *et al.* pioneerily proposed Laplacian of Gaussian (LoG) filter-based method [12], which motivated a broad range of work on the Human Visual System (HVS)-based methods. Chen *et al.* [13]

developed a descriptor called local contrast measure (LCM). Afterwards, a series of improved LCM methods are proposed, such as tri-layer window LCM (TLLCM) [14], weighted three-layer window LCM (WTLLCM) [15], strengthened robust LCM (SRLCM) [16], halo structure prior (HSP)-based LCM (HSPLCM) [17]. Unfortunately, these descriptors will fall into a dilemma in complicated ground or high-brightness background because clutters are similar to the target, causing false detection.

Recently, low-rank and sparse decomposition-based model (LRSD), which assumes that the IR image is composed of the low-rank background, the sparse target image and the noise image, has injected new vitality into IR small target detection. Gao *et al.* [18] proposed IR patch image (IPI) model. Li *et al.* [19] established the multiscale-window IPI model. Zhang *et al.* [20] introduced a method based on nonconvex rank approximation minimization (NRAM). Regrettably, the process of reconstructing an image matrix will destroy the spatial structure information of the original IR image, which greatly limits the performance of target detection.

Meanwhile, numerous IR small target detection methods based on low-rank and sparse tensor recovery have been proposed. Dai *et al.* [21] extended IPI model and established reweighted infrared patch tensor (RIPT) model. Zhang *et al.* [22] proposed a novel model based on inter-patch correlation enhancement (IPCE). Yang *et al.* [23] presented a novel group image-patch tensor model (GIPT). To approximate tensor rank more realistically, Kong *et al.* [24] adopted the tensor fiber kernel norm based on the Log operator (LogTFNN), and Cao *et al.* [25] defined multi-mode extension tensor tubal rank (METTR). In IRST system, spatial and temporal information in sequence images is essential for detection performance. Whereas, more false alarms or target missing may occur in many single-frame detection methods because only spatial information is used and temporal information is ignored.

2) *Multiframe-Based Sequential Detection Methods*: For the reason that single-frame-based detection methods ignore the temporal information in IR sequence, some LRSD-based methods in matrix domain seriously destroy the structural information of IR images. In contrast with the single-frame-based detection methods, the multiframe-based sequential detection methods can excavate more spatial-temporal domain information to identify small targets. Many researchers have witnessed that the sequential methods show a great deal of vitality, especially in complex backgrounds with low SCR.

Typical sequential detection methods include 3-D directional matched filter [26], multistage hypothesis testing [27], maximum-likelihood estimation [28]. Recently, a multitude of sequential detection methods in tensor domain have been proposed. To fully utilize both spatial and temporal information, Sun *et al.* [29] stacked the input image sequence to construct a holistic spatial-temporal tensor model (STTM). However, it needs more constraints in the procedure of optimization. From the perspective of edge-corner combinations, Zhang *et al.* [30] proposed an edge and corner awareness-based spatial-temporal tensor (ECA-STT) model. Owing to the influence of strong edges and other nontarget objects, Liu *et al.* [31] constructed a STTM based on the sparse prior and

the local correlation of the background, and obtained the target image via tensor decomposition. Nevertheless, there exists redundant information in STTM because of overlapping patches. Thereby, taking advantage of local and global correlation of images, Wang *et al.* [32] presented a nonoverlapping patch spatial-temporal tensor (NPSTT) model based on nonconvex optimization with the tensor capped nuclear norm (TCNN-NPSTT), which suppresses background and achieves better performance. Liu *et al.* [33] proposed a nonconvex tensor low-rank approximation (NTLA) method with asymmetric spatial-temporal total variation (ASTTV) regularization to estimate background part more accurately in complex scenarios. In [34], a new model combining facet derivative-based multidirectional edge awareness with spatial-temporal tensor (FDMDEA-STT) is presented for IR small target detection. Sun *et al.* [35] presented a novel optimization algorithm via multisubspace learning strategy and spatial-temporal tensor data structure (MSLSTIPT) to improve its robustness to heterogeneous and noisy scenes. And Pang *et al.* [36] established a spatial-temporal tensor model with saliency filter regularization (STTM-SFR) but did not achieve satisfactory real-time performance. Liu *et al.* [37] believed that the temporal profiles of pixels had local similarity and proposed a similarity model for pixel temporal profile (TP), and eliminated false alarms via a maximum suppression strategy. However, the existing multiframe-based detection methods are somewhat time-consuming and cannot sufficiently mine spatial-temporal features in sequence images, lacking information utilization.

3) *Deep learning-Based Detection Methods*: Recently, deep learning (DL) methods have developed rapidly and present great effectiveness in the IR target detection field. Different from traditional methods, convolutional neural networks (CNN) can learn IR small target representations in a data-driven manner. DL-based methods can be divided into detection-based methods and segmentation-based methods. As for detection-based methods, Du *et al.* [38] proposed a spatial-temporal feature-based detection framework with the small IOU strategy and IFEA mechanism. Bai *et al.* [39] proposed a cross-connected bidirectional pyramid network for IR small target detection. Segmentation-based methods can generate pixel-level classification and localization to achieve small IR target detection. Wang *et al.* [40] proposed MDvsFA-cGAN, which used a generative adversarial network (GAN) with two generators to optimize the two opposite sub-tasks of miss detection (MD) and false alarm (FA). Dai *et al.* [41] proposed Attentional Local Contrast Network (ALCNet) with a dilated local contrast measure. Li *et al.* [42] improved target detectability by designing a dense nested attention network (DNA-Net) based on UNet++. However, as a data-driven approach, large amounts of training data is critical for feature extraction and learning. Unfortunately, it is hard to obtain a great quantity of marked IR data because of heavy labor costs. Although the performance has been improving, the phenomenon of feature loss in deep layers still exists, leading to poor robustness to changing scenarios (e.g., targets with different SCR, size, shape and noisy background). And the cross-layer fusion may be ineffective or even reduce the detection performance. Hence, some traditional methods still

maintain a certain advantage in spite of traditional methods are usually sensitive to hyperparameters such as the kernel size, etc.

### B. Motivation

Although many of the existing methods have achieved relatively satisfactory performance in IRST systems, there are still several problems to be solved.

The most critical issue is that both the existing single-frame-based and multiframe-based methods cannot promote target detectability (TD) and background suppressibility (BS) simultaneously due to the limitations of algorithms, resulting in high false alarms, shrinkage of target even missed detection in the target image. In essence, improving TD and strengthening BS are mutually conflicting tasks. To balance and improve TD and BS, it urgently requires us to accurately depict the physical characteristics of the target, background and noise in IR sequences from global and local perspectives. Noting that detection performance is extremely affected by background and noise, so we deeply excavate their intrinsic characteristics, and explore influence mechanism towards targets of each component in IR images to improve the TD and BS.

The other issue lies in that the existing methods cannot accurately estimate the rank of background and inaccurate rank surrogate is difficult for completely separating background from the original image, leading to target over-shrinking or background residuals in the target image. Unfortunately, many rank approximation methods cannot take full advantage of the valuable information from different dimensions of tensors because they rely on single-mode extension. Some inappropriate rank surrogate will even destroy the tensor's internal structure. Meanwhile, singular values of each frontal slice have various physical meanings [43]. If treated equally, it can also cause the background rank approximation to be more deviated from the actual situation, leading to dilemmas such as target over-shrinkage problem. Therefrom, in view of the correlation in the inter-patch and intra-patch dimensions as well as the actual physical meanings of IR image contents, we are motivated to propose an improved nonconvex low-rank approximation for comprehensively extracting and combining different structural information among multiple modes of the constructed tensor. A satisfactory nonconvex rank surrogate can further separate the low-rank background from the original image.

Besides, there exist many line-features structures in complex scenarios, making it tough to separate them from the target only by global sparse features, leading to false alarms. It is necessary to introduce local feature, i.e., prior information. But existing local saliency priors only based on original images and structure tensor theory contain not only target priors but also invalid priors, which are unsatisfactory to guide the detection method to further suppress background under complex situations. Therefore, we are motivated to extract saliency prior information with more valid target priors and less invalid priors more finely. As a sparsity constraint, the prior “penalizes” the optimization algorithm to promote target saliency learning and restore a clean target image more thoroughly. Essentially, image entropy is a measurement of the degree of randomness

in an image. With an image entropy operator, the salient areas (e.g., target) can be highlighted and the flat areas can be suppressed, which is beneficial to extract the target priors via structure tensor theory. However, some invalid priors produced by significant structures may still appear. The phenomenon requires us to take measures to further suppress the invalid priors (e.g., line-features) while maintaining or even further enhancing the original valid priors. Noting that some point-featured and line-featured structures in the original IR image will be more prominent by contrast measuring. And theoretically, the structure tensors  $J$  of different objects (point, line and flat) in an image present specific determinants  $\text{Det}(J)$  and traces  $\text{Tr}(J)$ . More specifically, both  $\text{Det}(J)$  and  $\text{Tr}(J)$  of structure tensors of point/flat areas are large/small, and  $\text{Det}(J)$  of line areas are relatively small but its  $\text{Tr}(J)$  are large, which have been verified by experiments. The region recognition mapping derived from contrast measurement with structure tensor theory is regarded as a weight to reweight the acquired prior information. And the weight works as a prior information filter, which is referred to as a criterion to further enhance the valid priors and suppress the invalid priors.

In addition, noise will disturb target detection, making the background and target seriously affected by noise. Simultaneously, many scenarios include real objects with the streamlined appearances, such as buildings with perpendicular edges. These linear structures, which are of linear sparsity with respect to the whole image, may be separated into a part of target component and influence the overall detection performance.  $\ell_{2,1}$  norm of matrix has been used to deal with the rare structure effect [44]. As for the low-rank and sparse decomposition problem in tensor domain, inspired by the linear sparsity of these linear sparse structures and the statistical distribution characteristics of Gaussian noise, we are motivated to accurately characterize these linear sparse structures and noise respectively to reduce the possibility of these two components being retained in the target component.

In view of these abovementioned issues, we propose a novel IR small target detection method to promote small target detection performance.

### C. Contributions

Several contributions and innovations of this paper originate from addressing the mentioned issues in Section I-B, which can be summarized as follows:

- 1) The inherent characteristics of each component in IR images are characterized mathematically from the global and local perspectives of spatial-temporal. And an IR small target detection model is established, which comprehensively considers the influences of background, linear sparse structures and noise on small targets, and alleviates the contradiction between TD and BS.

- 2) Organically combining multi-mode tensor extension theory and the sum of the Laplace function of all singular values, a novel nonconvex approximation of tensor rank named improved multi-mode weighted tensor nuclear norm (IMWTNN) is proposed, which takes the advantages of the complete internal structural information of spatial-temporal tensor in multi-views.

3) Based on image information entropy theory and multiscale local contrast measurement based on structure tensor theory, an innovative local saliency prior map named local weighted entropy contrast (LWEC) is designed. And we theoretically and experimentally prove that the sparse constraint (LWEC) benefits to extract priors more finely.

4) We propose a novel IR target detection method called improved multi-mode nuclear norm joint local weighted entropy contrast (IMNN-LWEC) with an efficient optimization algorithm, which integrates the novel nonconvex approximation tensor rank, the original sparse prior weight, the novel tube-wise sparse regularization and Frobenius norm into the spatial-temporal tensor decomposition model. Comprehensive experiments demonstrate the superiority of IMNN-LWEC.

The remainder of this paper is organized as follows. Section II presents related mathematical symbols and preliminaries. In Section III, we propose a nonoverlapping spatial-temporal patch-tensor model named IMNN-LWEC and our optimization algorithm for IR small target detection by the spatial-temporal tensor decomposition. Section IV gives the comprehensive experiments and detailed analysis to verify the effectiveness of our method. Section V briefly summarizes this study.

## II. SYMBOLS AND PRELIMINARIES

In this section, we briefly introduce some basic concepts related to the tensor and the mathematical theories. We denote lowercase letter  $x$ , capital letter  $X$ , and Euler script letter  $\mathcal{X}$  to represent scalar, matrix and tensor, respectively. For a 3-order tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ ,  $n_1$ -th horizontal,  $n_2$ -th lateral and  $n_3$ -th frontal slices are denoted as  $\mathcal{X}(n_1, :, :)$ ,  $\mathcal{X}(:, n_2, :)$  and  $\mathcal{X}(:, :, n_3)$ , respectively. Especially, we denote  $\mathcal{X}_{(i)}$  as the  $i$ -th frontal slice of  $\mathcal{X}$ . And tensor  $\bar{\mathcal{X}}$  is generated by performing fast Fourier transformation (FFT) on  $\mathcal{X}$ , i.e.,  $\bar{\mathcal{X}} = fft(\mathcal{X}, [], 3)$ .

**Definition 1** (tensor nuclear norm, TNN). [45] The tensor nuclear norm of  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is denoted as  $\|\mathcal{X}\|_*$ , which is the average of the nuclear norm of all the frontal slices of  $\bar{\mathcal{X}}$ , namely,  $\|\mathcal{X}\|_* = \frac{1}{n_3} \sum_{i=1}^{n_3} \|\bar{\mathcal{X}}_{(i)}\|_*$ . And  $\bar{\mathcal{X}}_{(i)}$  is the  $i$ -th frontal slice of  $\bar{\mathcal{X}}$ .

**Definition 2** (mode- $k_1 k_2$  tensor unfolding). [46] Given an  $N$ -order tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_N}$ , its mode- $k_1 k_2$  extension is a 3-order tensor, which can be expressed by t-unfold  $(\mathcal{X}, k_1, k_2) = \mathcal{X}_{k_1 k_2} \in \mathbb{R}^{n_{k_1} \times n_{k_2} \times \prod_{s \neq k_1, k_2} n_s}$ . And the  $(i_1, i_2, \dots, i_N)$ -th element of  $\mathcal{X}$  corresponds to the  $(i_{k_1}, i_{k_2}, j)$ -th element of  $\mathcal{X}_{k_1 k_2}$ , where

$$j = 1 + \sum_{s=1, s \neq k_1, k_2}^N (i_s - 1) \times J_s, J_s = \prod_{m=1, m \neq k_1, k_2}^{s-1} n_m. \quad (1)$$

After mode- $k_1, k_2$  tensor unfolding, we can get  $N(N-1)/2$  three-order tensors ( $\mathcal{X}_{12}, \mathcal{X}_{13}, \dots, \mathcal{X}_{1N}, \mathcal{X}_{23}, \dots, \mathcal{X}_{2N}, \dots, \mathcal{X}_{(N-1)N}$ ). And the inverse operation of mode- $k_1, k_2$  extension is denoted as  $\mathcal{X} = t\text{-fold}(\mathcal{X}_{k_1 k_2}, k_1, k_2)$ . For a 3-order tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , although mode- $k_1, k_2$  extension does not reduce the dimension of  $\mathcal{X}$ ,  $\mathcal{X}$  can be extended into 3 tensors with different structures, thus, the structural information of  $\mathcal{X}$  can be dug out.

**Theorem 1** (t-SVD). [47] Based on the singular value decomposition (SVD) of matrix, the tensor-SVD (t-SVD) of  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  can be described as (2).

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \quad (2)$$

where, both  $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$  and  $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$  are orthogonal tensors, and  $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  is a f-diagonal tensor, “\*” is t-product. Its computation process is shown in Algorithm 1.

---

### Algorithm 1 t-SVD for tensor $\mathcal{X}$

---

**Input:**  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ .

1: Compute  $\bar{\mathcal{X}} = fft(\mathcal{X}, [], 3)$ .

2: Compute each frontal slice of  $\bar{\mathcal{U}}, \bar{\mathcal{S}}$  and  $\bar{\mathcal{V}}$  from  $\bar{\mathcal{X}}$  by

**For**  $i = 1, 2, \dots, n_3$  **do**

$[\bar{\mathcal{U}}_{(i)}, \bar{\mathcal{S}}_{(i)}, \bar{\mathcal{V}}_{(i)}] = svd(\bar{\mathcal{X}}_{(i)})$

**End for**

3.  $\mathcal{U} = ifft(\bar{\mathcal{U}}, [], 3)$ .

4.  $\mathcal{S} = ifft(\bar{\mathcal{S}}, [], 3)$ .

5.  $\mathcal{V} = ifft(\bar{\mathcal{V}}, [], 3)$ .

**Output:** t-SVD components  $\mathcal{U}, \mathcal{S}$  and  $\mathcal{V}$  of  $\mathcal{X}$ .

---

## III. METHODOLOGY

In this section, we present the IMNN-LWEC model for IR dim and small target detection in complex scenarios.

### A. Construction of the Spatial-Temporal Tensor Model

Given an IR image sequence, each IR image in sequence can be modeled linearly as [48]

$$\mathbf{D} = \mathbf{B} + \mathbf{T} + \mathbf{N}, \quad (3)$$

where,  $\mathbf{D}, \mathbf{B}, \mathbf{T}$  and  $\mathbf{N} \in \mathbb{R}^{m_1 \times m_2}$  represent input image, background component, target component and stochastic noise component in matrix domain, respectively.

In the IPT model [21], a sliding window is used to traverse the image and each patch is used as a frontal slice to construct a tensor. To fully excavate more spatial and structural information, we construct a novel nonoverlapping spatial-temporal patch-tensor of the IR image sequence.

Supposed that the IR sequence contains  $K$  frames of images. As shown in Fig. 1, the input  $t$ -th frame image  $D^t (t = 1, 2, \dots, K)$  is divided into a series of patches with a window sliding from left to right and top to bottom in “S-shaped” path. In this way, we can guarantee the local spatial correlation between adjacent patches. The size of each patch is  $ps \times ps$ , and the sliding step ( $ss$ ) is equal to patchsize, i.e.,  $ss = ps$ . Thereupon, there are no overlapping parts between adjacent patches, which reduces redundant information and improves computation efficiency to a certain extent. Finally, we get  $n (= n_r \times n_c)$  patches from the  $t$ -th frame and stack them together to obtain a 3-order tensor  $\mathcal{D}^t (t = 1, 2, \dots, K)$ . And each frontal slice in  $\mathcal{D}^t$  denotes as  $\mathcal{D}_{(i)}^t$ , corresponding to the  $i$ -th patch in the  $t$ -th frame image  $D^t$ .

To excavate more structural spatial and temporal information, for consecutive  $k (1 < k \leq K)$  frames of IR images,

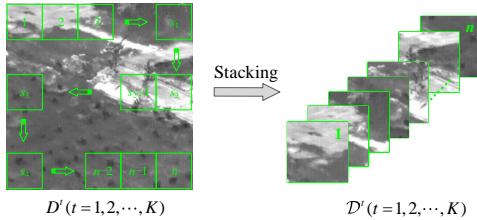


Fig. 1. Tensor construction of the  $t$ -th frame IR image  $D^t(t = 1, 2, \dots, K)$ .

we can obtain  $k$  cubes ( $\mathcal{D}^1, \mathcal{D}^2, \dots, \mathcal{D}^k$ ) as the stacking way in Fig. 1. Then the  $k$  cubes are stacked in order and a nonoverlapping spatial-temporal patch tensor  $\mathcal{D}$  is obtained, as shown in Fig. 2. In addition to being rich in spatial information, the constructed nonoverlapping patch tensor also fully incorporates temporal characteristics of sequential images. Thus, model (3) can be re-modeled as a holistic spatial-temporal tensor model with spatial structure preserved as follows [21], [24], [32]–[34]:

$$\mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{N}, \quad (4)$$

where,  $\mathcal{D}$ ,  $\mathcal{B}$ ,  $\mathcal{T}$ ,  $\mathcal{N} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$  represent the spatial-temporal patch-tensor of raw IR image, background tensor, target tensor and noise tensor, respectively.  $n_1 (= ps)$  and  $n_2 (= ps)$  are the height and the width of each local patch and  $n_3 (= n \times k)$  represents the number of patches.

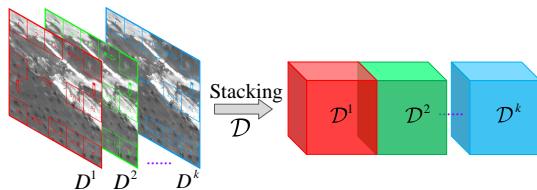


Fig. 2. Spatial-temporal tensor construction of multi-frame images.

According to [31], the background patch tensor  $\mathcal{B}$  is of low-rank property, and the target patch-tensor  $\mathcal{T}$  can be considered as a sparse tensor because the small targets only occupy a bit pixels in the whole image. In image processing, it usually assumes that noise component obeys some statistical distributions [49]. For example, Gaussian noise can be characterized by Frobenius norm. Taking Gaussian noise into consideration, the tensor decomposition based small target detection problem can be converted into an optimization problem, defined by (5).

$$\begin{aligned} \min_{\mathcal{T}, \mathcal{B}, \mathcal{N}} & \text{rank}(\mathcal{B}) + \lambda_1 \|\mathcal{T}\|_0 + \lambda_2 \|\mathcal{N}\|_F^2 \\ \text{s.t.} & \mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{N}, \end{aligned} \quad (5)$$

where,  $\lambda_1$  and  $\lambda_2$  are positive trade-off parameters,  $\|\cdot\|_0$  is  $\ell_0$  norm,  $\|\cdot\|_F$  is Frobenius norm. However, solving problem (5) is an NP-hard problem [50], because it is almost unachievable to solve the rank and  $\ell_0$  norm of tensor directly. In [51],  $\|\mathcal{T}\|_0$  is replaced by  $\|\mathcal{T}\|_1$  to make the above problem tractable, but the precise approximation of tensor rank remains to be solved.

### B. Nonconvex Approximation of Tensor Rank

We have mentioned that the background component in IR image is low-rank and tensor rank approximation is NP-hard.

Thus, it is necessary to propose a suitable tensor rank surrogate method to alleviate this NP problem and measure the non-local correlation property of the background in IR image. And the more significant the non-local correlation of the tensor, the smaller the tensor rank.

For the constructed 3-order spatial-temporal patch-tensor  $\mathcal{D}$  in section III-A, whose first two dimensions are rich in spatial information and jointly describe the local correlation of each patch. The third dimension of  $\mathcal{D}$  consists of multi-temporal and multi-position patches, which contains sufficient temporal and spatial characteristics. Also, the non-local feature of the background can be mined from the third dimension of  $\mathcal{D}$ .

However, current TNN-based rank surrogate methods have limitations. According to Theorem 1, t-SVD is a generalization of singular value decomposition in tensor domain. In fact, t-SVD considers a 3-order tensor  $\mathcal{X}$  as a special matrix, whose elements are tubes, i.e., mode-3 fibers  $\mathcal{X}(i, j, :)$  in  $\mathcal{X}$ . And TNN calculated via t-SVD is often based on a single mode. Especially for the 3-order spatial-temporal patch-tensor  $\mathcal{D}$ , tensor rank surrogate based on a single mode lacks the flexibility to handle different correlations along different modes. Besides, it is hard to retain the structural information among multiple modes and improve the computational efficiency simultaneously.

Therefore, we introduce a novel tensor unfolding operation rather than single mode extension to transform background tensor  $\mathcal{B}$  by reordering its slices along any two modes, i.e., mode- $k_1 k_2$  extension in definition 2. Because for a 3-order background tensor, we can get three multi-views of itself after mode- $k_1 k_2$  extension,  $\mathcal{B}_{12} \in \mathbb{R}^{n_1 \times n_2}$ ,  $\mathcal{B}_{13} \in \mathbb{R}^{n_1 \times n_3}$  and  $\mathcal{B}_{23} \in \mathbb{R}^{n_2 \times n_3}$ . Not only the spatial-temporal structural information is not destroyed, but also the relationship between mode- $k_1$  and mode- $k_2$  is more accurately characterized. Particularly, all the frontal slices in  $\mathcal{B}_{13}$  and  $\mathcal{B}_{23}$  directly present both spatial and temporal information.

As shown in Fig. 3, three tensors from multiple views can be obtained via mode- $k_1k_2$  extension of background tensor  $\mathcal{B}$ , illustrating that mode- $k_1k_2$  extension can directly depict the correlation between  $k_1$ -mode and  $k_2$ -mode. And these tensors obtained from mode- $k_1k_2$  extension satisfy the low-rank property because the singular values of each frontal slice decrease to zero rapidly and it can be mathematically characterized by (6). The phenomenon indicates that the three tensors can be used to approximate tensor rank more accurately.

$$\text{rank}(\mathcal{B}_{12}) < r_1, \text{rank}(\mathcal{B}_{13}) < r_2, \text{rank}(\mathcal{B}_{23}) < r_3. \quad (6)$$

Another shortcoming of TNN-based rank approximation is that all singular values are treated equally, which is not a better surrogate to tensor rank. Since the singular values in an IR image have clear physical meanings with different importance [43], once the singular values are assigned with same weights, image information loss will occur. Thanks to the Laplace function, a better approximation to the  $\ell_0$  norm [52], which is defined as  $g(x) = 1 - e^{-\frac{x}{\epsilon}}$ , and  $\epsilon$  is a relative small positive constant. Noting that  $\lim_{\epsilon \rightarrow 0} g(x) = 1$  when  $x > 0$ , and  $\lim_{\epsilon \rightarrow 0} g(x) = 0$  iff  $x = 0$ . Given that  $\bar{\mathcal{X}}_{(i)}$  is the  $i$ -th frontal slice of the tensor  $\bar{\mathcal{X}} (= fft(\mathcal{X}, [], 3))$ , whose  $j$ -th singular value denotes as  $\sigma_j(\bar{\mathcal{X}}_{(i)})$ . And the rank of the matrix  $\bar{\mathcal{X}}_{(i)}$

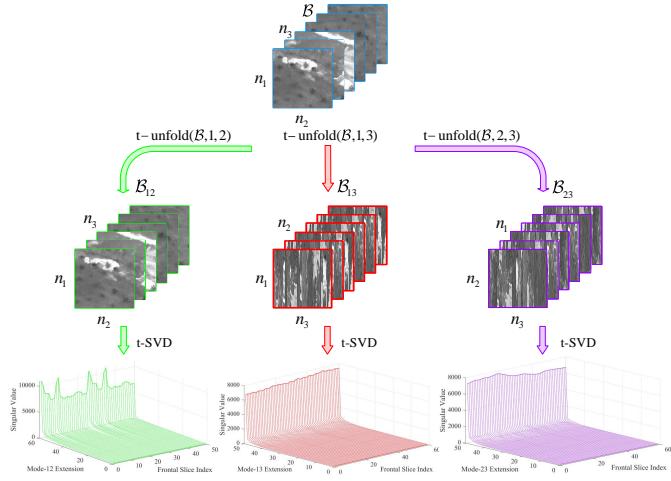


Fig. 3. Mode- $k_1 k_2$  extension and singular values of the extended tensors.

is equal to the number of its non-zero singular values, thus,  $\lim_{\epsilon \rightarrow 0} \sum_{j=1}^{\min(n_1, n_2)} g(\sigma_j(\bar{\mathcal{X}}_{(i)})) = \text{rank}(\bar{\mathcal{X}}_{(i)})$ .

Compared with the  $\ell_1$  norm, which only adds up all the singular values, a better alternative nonconvex surrogate for tensor rank can be obtained via the sum of the Laplace function of singular values. We denote the improved tensor rank as  $\|\cdot\|_{\epsilon,*}$ , defined as

$$\begin{aligned} \|\mathcal{B}\|_{\epsilon,*} &= \sum_{i=1}^{n_3} \sum_{j=1}^{\min(n_1, n_2)} g(\sigma_j(\bar{\mathcal{B}}_{(i)})) \\ &= \sum_{i=1}^{n_3} \sum_{j=1}^{\min(n_1, n_2)} 1 - e^{-\frac{\sigma_j(\bar{\mathcal{B}}_{(i)})}{\epsilon}}, \end{aligned} \quad (7)$$

where,  $\sigma_j(\bar{\mathcal{B}}_{(i)})$  is the  $j$ -th singular value of the  $i$ -th frontal slice of  $\bar{\mathcal{B}}$ , and  $\epsilon$  is a positive constant.

Since tensor  $\mathcal{B}_{12}$ ,  $\mathcal{B}_{13}$  and  $\mathcal{B}_{23}$  have been proved to satisfy the low-rank property. And the singular values of their frontal slices are of different levels of importance. Combining the mode- $k_1 k_2$  tensor extension and the sum of the Laplace function of all singular values in (7), we propose a novel nonconvex approximation norm of tensor rank, called improved multi-mode weighted tensor nuclear norm (IMWTNN), which is actually the linear weighted sum of the improved tensor rank of several tensors  $\mathcal{B}_{k_1 k_2}$  obtained by mode- $k_1 k_2$  extension of  $\mathcal{B}$ , defined as

$$\begin{aligned} \|\mathcal{B}\|_{IMWTNN} &= \sum_{1 \leq k_1 < k_2 \leq N} \alpha_{k_1 k_2} \|\mathcal{B}_{k_1 k_2}\|_{\epsilon,*} \\ &= \sum_{1 \leq k_1 < k_2 \leq N} \alpha_{k_1 k_2} \sum_{i=1}^{n_3} \|\mathcal{B}_{k_1 k_2(i)}\|_{\epsilon,*}, \end{aligned} \quad (8)$$

where,  $\sum_{1 \leq k_1 < k_2 \leq N} \alpha_{k_1 k_2} = 1$ ,  $N$  is 3 because of the 3-order tensor,  $\mathcal{B}_{k_1 k_2(i)}$  is the  $i$ -th frontal slice of  $\mathcal{B}_{k_1 k_2}$ . By setting different values of coefficients  $\alpha_{k_1 k_2}$ , the temporal information and spatial information can be used to varying degrees. From the above mathematical analysis, IMWTNN can more realistically characterize the low-rank characteristics of background.

### C. Local Saliency Prior Map

Image entropy can reflect the degree of difference between pixel values in IR images. The stronger the degree of grayscale mutation in the local area, the higher the local image entropy. IR targets are usually small with the size of only one or a few pixels, and presents structural discontinuities and brightness differences. Image neighborhoods, where small targets appear, usually have larger image entropy when the contrast between target and background is significant. In order to improve the visual saliency of IR small targets, we design an image entropy operator. [53] has proposed local weighted information entropy. Assuming that the center of the local entropy template is located at  $(x, y)$ ,  $M$  is an  $N \times N$  neighborhood and contains  $m$  kinds of grayscale values  $f_i (i = 1, 2, \dots, m)$ , whose corresponding probability distributions are  $P_i (i = 1, 2, \dots, m)$ . The local weighted information entropy is as follows

$$e(x, y) = - \sum_{i=1}^m \left[ (f_i - f(x, y))^2 P_i \log P_i \right]. \quad (9)$$

If the difference between pixel point  $(x, y)$  and its surrounding neighborhood pixels is significant, pixel value  $f(x, y)$  will be enhanced by (9). Otherwise, the pixel point will be suppressed. Therefore, (9) can effectively enhance IR small targets. Meanwhile, high-brightness pixels belonging to edges in complex background are enhanced as well, which interferes detection and causes false alarms. By introducing the local standard deviation  $\sigma_{\text{std}}$  to measure the uniformity and consistency of the pixels in  $M$ , targets are enhanced and edges are suppressed. The designed image entropy operator is as follows:

$$\begin{aligned} E(x, y) &= e(x, y) / (\sigma_{\text{std}}(x, y) + \epsilon_1) \\ &= e(x, y) \left[ \frac{1}{N^2} \sum_{(i,j) \in M} [f(i, j) - M(x, y)]^2 + \epsilon_1 \right]^{-\frac{1}{2}}, \end{aligned} \quad (10)$$

where,  $\epsilon_1$  is a small positive constant to avoid the denominator being zero, and we set  $M$  to be a  $7 \times 7$  neighborhood.

In this way, the point-like areas are visually significantly enhanced, and some flat areas are well suppressed. Since structure tensor theory is a useful tool to describe the spatial structure information of an image. Hence, computing the classical structure tensor  $J_\ell$  via (11) at each pixel of  $E$  in (10) rather than the original image  $D$  can better extract corners and strong edges:

$$\begin{aligned} J_\ell &= G_\ell * (\nabla E_\sigma \otimes \nabla E_\sigma) \\ &= \begin{pmatrix} G_\ell * I_x^2 & G_\ell * I_x I_y \\ G_\ell * I_y I_x & G_\ell * I_y^2 \end{pmatrix} = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}, \end{aligned} \quad (11)$$

where,  $G_\ell$  is a Gaussian kernel function with variance  $\ell$ ,  $E_\sigma$  is a Gaussian smoothing filter with variance  $\sigma$  on  $E$ , “\*” is a 2D convolution operator,  $\otimes$  is Kronecker product,  $\nabla$  is gradient operator, and  $I_x = \frac{\partial E_\sigma}{\partial x}$ ,  $I_y = \frac{\partial E_\sigma}{\partial y}$ .

The structure tensor in (11) at each position  $(x, y)$  of  $E$  has two eigenvalues ( $\lambda_1$  and  $\lambda_2$ ), which reflect the features of  $E$  in different situations: i) edge,  $\lambda_1 \gg \lambda_2 \approx 0$ ; ii) corner,  $\lambda_1 \geq \lambda_2 \gg 0$ ; iii) flat,  $\lambda_1 \approx \lambda_2 \approx 0$ . Zhang [54] proposed

a local prior map combined with background prior  $w_{\text{bp}}$  and target prior  $w_{\text{tp}}$  as follows:

$$W_l = w_{\text{bp}} \cdot w_{\text{tp}} = \max(\lambda_1, \lambda_2) \cdot \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}. \quad (12)$$

In this paper, we take the harmonic mean of the background prior  $w_{\text{bp}}$  and the improved target prior  $w_{\text{tp}}$  as the improved local prior map:

$$W_l = \frac{w_{\text{bp}} \cdot w_{\text{tp}}}{w_{\text{bp}} + w_{\text{tp}} + \epsilon_3}, \quad (13)$$

where,  $w_{\text{tp}} = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2 + \epsilon_2}$ ,  $\epsilon_2$  and  $\epsilon_3$  are to prevent the denominator to be zero. Then,  $W_l$  is normalized via (14).

$$W'_l = \frac{W_l - w_{\min}}{w_{\max} - w_{\min}}, \quad (14)$$

where,  $w_{\min}$  and  $w_{\max}$  denote the minimum and maximum of  $W_l$ , respectively.

However, some target-like areas and stubborn edges are also magnified in complex environments after the above procedure, the prior map obtained from (13) will contain edge component. Therefore, a new prior information filter is required. To this end, we propose a prior information filter as a new criteria, which is based on multiscale local contrast measure and structure tensor theory. By assigning different weights to invalid areas of line-features, invalid prior information in the aforementioned prior map  $W'_l$  can be filtered.

Multiscale local contrast measure can highlight salient targets and edge areas, so a template shown in Fig. 4 is designed to measure the grayscale difference between targets and background to get a multiscale local contrast map. The template is composed of several square boxes centered at  $(x, y)$ , and its  $g$ -level neighborhood is defined as

$$\Omega_g = \{(u, v) | \max(|u - x|, |v - y|) \leq g\}. \quad (15)$$

Then, we can get the grayscale mean of the inner box  $c_0$  and  $n$  adjacent square rings  $c_i (i = 1, 2, \dots, n)$ .

$$c_0 = \frac{1}{N_g} \sum_{(u, v) \in \Omega_g} f(u, v), \quad (16)$$

$$c_i = \frac{1}{N_{g+i} - N_{g+i-1}} \sum_{(u, v) \in R_i} f(u, v), \quad (17)$$

$$i = 1, 2, \dots, n.$$

where,  $N_g$  is the number of pixels contained in neighborhood  $\Omega_g$ ,  $(u, v) \in R_i$  represents pixel  $(u, v)$  belongs to  $i$ -th square ring  $R_i$ , i.e.,  $(u, v) \in \Omega_{g+i}$  and  $(u, v) \notin \Omega_{g+i-1}$ .

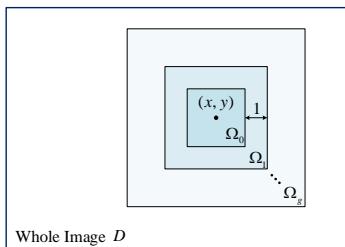


Fig. 4. Local Contrast Coefficient Calculation Template.

With the grayscale means of the inner box and the pixel mean of the square ring areas, the local contrast map  $F$  can be obtained as follows:

$$F(x, y) = f(x, y)^{c(x, y)}, \quad (18)$$

and

$$c(x, y) = \max \left\{ \frac{c_0}{c_1}, \frac{c_0}{c_2}, \dots, \frac{c_0}{c_n} \right\}, \quad (19)$$

where,  $c(x, y)$  in (19) is the local contrast coefficient centered at  $(x, y)$ ,  $g$  is set to 1, and  $n$  is set to 3 in this paper. Although there exists works to detect dim targets with brighter surroundings [55], we are aimed at detecting bright targets with dim surroundings. And (18) is based on the premise that the grayscale of the target is greater than or even equal to the grayscale of its surroundings.

Similarly, we calculate the structure tensor of the contrast image via (11), and the corresponding eigenvalues are denoted by  $\lambda'_1$  and  $\lambda'_2$ . The trace of the structure tensor at each pixel is  $\Phi_t = \lambda'_1 + \lambda'_2$ , and the determinant is  $\Phi_d = \lambda'_1 \lambda'_2$ . Given an image, the mathematical relationship between the determinant value and trace value of the structure tensor of each pixel can reflect point-features, line-features and flat-features. And the image visualization is shown in Fig. 5.

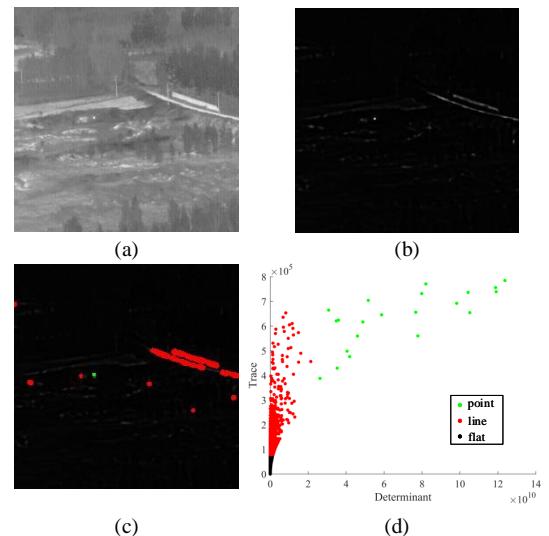


Fig. 5. (a) Raw infrared image; (b) multiscale contrast image; (c) contrast image with point-features and line-features; and (d) image visualization of the relationship between determinant and trace of structure tensor of each pixel. The red dots, the green dots and the black dots in (d) correspond to the line-featured areas, point-featured and flat-featured areas in (c), respectively.

Based on the abovementioned analysis, to further distinguish the three situations above accurately by contrast images, we introduce a new criterion.

$$\begin{cases} \Phi_t(x, y) \geq t_1, \Phi_d(x, y) \leq d_1, & (x, y) \in \text{edge}, \\ \Phi_t(x, y) \geq t_2, \Phi_d(x, y) \geq d_2, & (x, y) \in \text{point}, \\ \text{otherwise}, & (x, y) \in \text{flat}, \end{cases} \quad (20)$$

where,  $d_1 \leq d_2$ . And  $t_1, t_2, d_1, d_2$  are appropriately thresholds as the judgment as follows:

$$\begin{aligned} t_1 &= \min(\tilde{\Phi}_t) + \psi_1(\max(\tilde{\Phi}_t) - \min(\tilde{\Phi}_t)), \\ t_2 &= \min(\tilde{\Phi}_t) + \psi_2(\max(\tilde{\Phi}_t) - \min(\tilde{\Phi}_t)), \\ d_1 &= \min(\tilde{\Phi}_d) + \psi_3(\max(\tilde{\Phi}_d) - \min(\tilde{\Phi}_d)), \\ d_2 &= \min(\tilde{\Phi}_d) + \psi_4(\max(\tilde{\Phi}_d) - \min(\tilde{\Phi}_d)), \end{aligned} \quad (21)$$

where,  $\psi_i(i = 1, 2, 3, 4)$  are empirical values,  $\tilde{\Phi}_t$  and  $\tilde{\Phi}_d$  are remaining values after removing infinite numbers.

In a complex environment, strong edges are magnified. To further enhance the small targets of point-features and suppress the strong edges of line-features, we propose two weight matrices  $W_{\text{edge}}$  in (22) and  $W_{\text{point}}$  in (23) and joint assign them to the obtained prior map  $W_l'$ . The final local saliency prior map  $W_p$  named local weighted entropy contrast (LWEC) is calculated via (24).

$$W_{\text{edge}}(x, y) = \begin{cases} 0, & \text{if } (x, y) \in \text{edge}, \\ 1, & \text{else.} \end{cases} \quad (22)$$

$$W_{\text{point}}(x, y) = \begin{cases} 1, & \text{if } (x, y) \in \text{point}, \\ 0, & \text{else.} \end{cases} \quad (23)$$

$$W_p = W_l' \cdot (W_{\text{edge}} + \gamma W_{\text{point}}), \quad (24)$$

where,  $\gamma$  is a positive constant and is set to 3.

The different local saliency prior maps are shown in Fig. 6. Invalid interference information can be well suppressed in this paper. So far, the local saliency prior map  $W_p$  can be constructed to a new spatial-temporal prior tensor  $\mathcal{W}_p$ . And the construction process is the same as the procedure of constructing tensor  $\mathcal{D}$  in Section III-A. As a sparse prior,  $\mathcal{W}_p$  is integrated into the spatial-temporal tensor model.

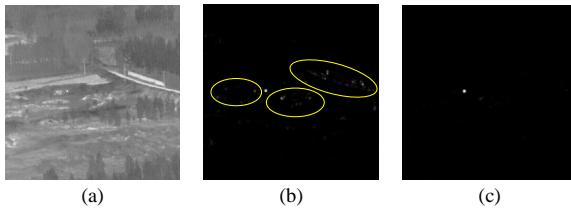


Fig. 6. (a) Raw infrared image; (b) local prior weight map of LogTFNN, where there are some strong edge residuals left in the yellow circle; and (c) the proposed local saliency prior map.

#### D. Tube-wise Sparse Regularization Term

Many existing methods suffer from rare structure effect [44]. Objects with streamlined appearances in complex situations usually present the property of linear sparsity, and outliers produced by these structures (e.g., stubborn building edges) may be separated as a part of target component. In matrix domain,  $\ell_{2,1}$  norm can identify the sample outliers on each row, most of which are related to linear sparse structures [20]. In fact,  $\ell_{2,1}$  norm can characterize row sparsity of an image matrix. And  $\ell_{1,1,2}$  norm of a third-order tensor [45], [56] is similar to  $\ell_{2,1}$  norm of matrix, which is utilized as a tube-wise sparse regularization to separate outliers in linear sparse

structures on all tubes of the third dimension. In this paper, after transforming the original IR sequence into a tensor, linear sparse structures will still exist in some patches. Because of the linear sparsity of most interference sources, we introduce a novel tube-wise sparse regularization for a 3-order tensor as the sum of  $\ell_2$  norm of all tubes  $\mathcal{E}(i, :, l)(1 \leq i \leq n_1, 1 \leq l \leq n_3)$  and  $\mathcal{E}(:, j, l)(1 \leq j \leq n_2, 1 \leq l \leq n_3)$  of the linear sparse structure tensor  $\mathcal{E}$ , which is effective to identify the sample outliers on linear sparse structures on each frontal slice of the tensor  $\mathcal{D}$  and better suppress strong sparse edges in different rows and columns. Based on row sparsity and column sparsity of these linear sparse structures, the definition of the designed norm utilized in the novel tube-wise sparse regularization is as follows:

$$\begin{aligned} \|\mathcal{E}\|_{tw} &= \sum_{i,l} \|\mathcal{E}(i, :, l)\|_2 + \sum_{j,l} \|\mathcal{E}(:, j, l)\|_2 \\ &= \sum_{i,l} \left( \sum_j |\mathcal{E}(i, j, l)|^2 \right)^{\frac{1}{2}} + \sum_{j,l} \left( \sum_i |\mathcal{E}(i, j, l)|^2 \right)^{\frac{1}{2}}, \end{aligned} \quad (25)$$

where,  $\|\cdot\|_2$  is  $\ell_2$  norm,  $\|\cdot\|_{tw}$  is the norm utilized in the novel tube-wise sparse regularization.

Consequently, the spatial-temporal patch-tensor  $\mathcal{D}$  can be re-decomposed into background tensor  $\mathcal{B}$ , target tensor  $\mathcal{T}$ , linear sparse structure tensor  $\mathcal{E}$  and noise tensor  $\mathcal{N}$ , and (4) can be equivalently transformed into the following (26).

$$\mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{E} + \mathcal{N}. \quad (26)$$

#### E. The Proposed IMNN-LWEC Model

Generally, the retained target requires small regularization parameters, edges and corners need to be suppressed, we obtain  $\mathcal{W}_{pr}$  by taking the reciprocal of the corresponding elements in tensor  $\mathcal{W}_p$ . By integrating nonconvex approximation of tensor rank, local sparse prior tensor and tube-wise sparse regularization term into the model (5), and combined with (26), we propose a novel optimization model as follows:

$$\begin{aligned} \min \|\mathcal{B}\|_{IMWTNN} + \lambda_1 \|\mathcal{W}_{pr} \odot \mathcal{T}\|_1 + \lambda_2 \|\mathcal{E}\|_{tw} + \lambda_3 \|\mathcal{N}\|_F^2 \\ \text{s.t. } \mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{E} + \mathcal{N}, \end{aligned} \quad (27)$$

where,  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  are positive trade-off parameters.

In [57], Candès proposed a reweighted  $\ell_1$  minimization to enhance sparsity, whose weight is inversely proportional to the real signal value. To speed up the convergence, we adopt the sparse reweighted scheme.

$$\mathcal{W}_{sw}^{p+1} = \frac{a}{|\mathcal{T}^p| + b}, \quad (28)$$

where,  $a$  is a positive constant and is set to 2 in this paper,  $b$  is a small positive constant to avoid division by zero, and  $p$  is the number of iterations. Thus, the final sparse prior weight tensor is

$$\mathcal{W} = \mathcal{W}_{pr} \odot \mathcal{W}_{sw}, \quad (29)$$

where,  $\odot$  denotes Hadamard product. The small weights preserve nonzero elements, and the large weights discourage nonzero elements.

Finally, the model (27) is rewritten as follows:

$$\begin{aligned} \min & \| \mathcal{B} \|_{IMWTNN} + \lambda_1 \| \mathcal{W} \odot \mathcal{T} \|_1 + \lambda_2 \| \mathcal{E} \|_{tw} + \lambda_3 \| \mathcal{N} \|_F^2 \\ \text{s.t. } & \mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{E} + \mathcal{N}, \end{aligned} \quad (30)$$

where,  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  are positive trade-off parameters.

#### F. ADMM Solver for Proposed Model

The optimization problem (30) can be solved effectively via Alternating Direction Method of Multipliers (ADMM) [58] approach. By introducing three auxiliary variables  $\mathcal{Z}_{k_1 k_2}$  ( $1 \leq k_1 < k_2 \leq N$ ,  $N = 3$ ), the equivalent expression of (30) is given as follows:

$$\begin{aligned} \min & \| \mathcal{Z} \|_{IMWTNN} + \lambda_1 \| \mathcal{W} \odot \mathcal{T} \|_1 + \lambda_2 \| \mathcal{E} \|_{tw} + \lambda_3 \| \mathcal{N} \|_F^2 \\ \text{s.t. } & \mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{E} + \mathcal{N}, \mathcal{Z}_{k_1 k_2} = \mathcal{B}_{k_1 k_2}. \end{aligned} \quad (31)$$

Then, the augmented Lagrangian function of (31) is given as follows:

$$\begin{aligned} & \mathcal{L}_{\beta_{k_1 k_2}, \rho} (\mathcal{Z}_{k_1 k_2}, \mathcal{P}_{k_1 k_2}, \mathcal{M}, \mathcal{B}, \mathcal{T}, \mathcal{E}, \mathcal{N}) \\ &= \sum_{1 \leq k_1 < k_2 \leq N} \left\{ \alpha_{k_1 k_2} \sum_{i=1}^{n_3} \| \mathcal{Z}_{k_1 k_2(i)} \|_{\epsilon, *} + \frac{\beta_{k_1 k_2}}{2} \| \mathcal{B}_{k_1 k_2} - \mathcal{Z}_{k_1 k_2} + \frac{\mathcal{P}_{k_1 k_2}}{\beta_{k_1 k_2}} \|_F^2 \right\} + \lambda_1 \| \mathcal{W} \odot \mathcal{T} \|_1 + \lambda_2 \| \mathcal{E} \|_{tw} \\ &+ \lambda_3 \| \mathcal{N} \|_F^2 + \frac{\rho}{2} \left\| \mathcal{D} - \mathcal{B} - \mathcal{T} - \mathcal{E} - \mathcal{N} + \frac{\mathcal{M}}{\rho} \right\|_F^2 + \mathcal{C}, \end{aligned} \quad (32)$$

where,  $\rho$  and  $\beta_{k_1 k_2}$  are penalty parameters,  $\mathcal{M}$  and  $\mathcal{P}_{k_1 k_2}$  are Lagrange multipliers, and  $\mathcal{C}$  is an independent variable. In this paper, we concentrate more on temporal information utilization and set  $\alpha_{12} = w/(2+w)$ ,  $\alpha_{23} = \alpha_{13} = 1/(2+w)$ , and  $w$  is set to 0.001.

Since it is difficult to solve all variables in (32) simultaneously, we decompose (32) into seven subproblems, and then alternately update variables. The specific process is as follows:

##### 1) Subproblem of $\mathcal{Z}_{k_1 k_2}$ :

With other parameters fixed,  $\mathcal{Z}_{k_1 k_2}$  can be updated by solving the following subproblem:

$$\begin{aligned} & \mathcal{Z}_{k_1 k_2}^{(j+1)} \\ &= \arg \min_{\mathcal{Z}_{k_1 k_2}} \mathcal{L} (\mathcal{Z}_{k_1 k_2}, \mathcal{P}_{k_1 k_2}^{(j)}, \mathcal{M}^{(j)}, \mathcal{B}^{(j)}, \mathcal{T}^{(j)}, \mathcal{E}^{(j)}, \mathcal{N}^{(j)}) \\ &= \arg \min_{\mathcal{Z}_{k_1 k_2}} \sum_{1 \leq k_1 < k_2 \leq N} \left\{ \alpha_{k_1 k_2} \sum_{i=1}^{n_3} \| \mathcal{Z}_{k_1 k_2(i)} \|_{\epsilon, *} + \frac{\beta_{k_1 k_2}}{2} \| \mathcal{B}_{k_1 k_2}^{(j)} + \frac{\mathcal{P}_{k_1 k_2}^{(j)}}{\beta_{k_1 k_2}} - \mathcal{Z}_{k_1 k_2} \|_F^2 \right\}. \end{aligned} \quad (33)$$

The Laplace function  $g(\cdot)$  is concave, continuous and differentiable w.r.t singular value  $\sigma$ . By linearizing the Laplace

function, we get the following relaxation problem of (33).

$$\begin{aligned} & \mathcal{Z}_{k_1 k_2}^{(j+1)} \\ &= \arg \min_{\mathcal{Z}_{k_1 k_2}} \sum_{1 \leq k_1 < k_2 \leq N} \left\{ \alpha_{k_1 k_2} \sum_{s=1}^{\min(n_1, n_2)} \sigma_{(i)s} \cdot \nabla g (\sigma_{(i)s}) \right. \\ &\quad \left. + \frac{\beta_{k_1 k_2}}{2} \left\| \mathcal{B}_{k_1 k_2}^{(j)} + \frac{\mathcal{P}_{k_1 k_2}^{(j)}}{\beta_{k_1 k_2}} - \mathcal{Z}_{k_1 k_2} \right\|_F^2 \right\}, \end{aligned} \quad (34)$$

where,  $\nabla g(\sigma) = \frac{1}{\epsilon} e^{-\frac{1}{\epsilon} \sigma}$ ,  $\sigma_{(i)1} \geq \sigma_{(i)2} \geq \dots \geq \sigma_{(i)s}$  are singular values of the  $i$ -th frontal slice  $\mathcal{Z}_{k_1 k_2(i)}$ .

Inspired by Theorem 2 in [59],  $\mathcal{Z}_{k_1 k_2}^*$  can be solved by

$$\mathcal{Z}_{k_1 k_2}^{(j+1)} = \text{t-fold} \left[ \mathcal{U} * \mathcal{D}_{\frac{\alpha_{k_1 k_2}}{\beta_{k_1 k_2}}} \left( \mathcal{B}_{k_1 k_2}^{(j)} + \frac{\mathcal{P}_{k_1 k_2}^{(j)}}{\beta_{k_1 k_2}} \right) * \mathcal{V}^T, k_1, k_2 \right], \quad (35)$$

where,  $\mathcal{D}_\tau$  is the singular value shrinkage operator,  $\mathcal{D}_\tau(\sigma) = \text{diag}[(\sigma - \tau)_+]$ , thus,  $\mathcal{D}_{\frac{\alpha_{k_1 k_2}}{\beta_{k_1 k_2}}} \left( \mathcal{B}_{k_1 k_2}^{(j)} + \frac{\mathcal{P}_{k_1 k_2}^{(j)}}{\beta_{k_1 k_2}} \right) = \text{diag} \left[ \left( \mathcal{Q}_{(i)s} - \frac{\alpha_{k_1 k_2}}{\beta_{k_1 k_2}} \nabla g(\sigma_{(i)s}) \right)_+ \right]$ ,  $\mathcal{B}_{k_1 k_2}^{(j)} + \frac{\mathcal{P}_{k_1 k_2}^{(j)}}{\beta_{k_1 k_2}} = \mathcal{U} * \mathcal{Q} * \mathcal{V}^T$ ,  $\zeta_+ = \max(\zeta, 0)$ , and t-fold( $\cdot$ ) is inverse operation of t-unfold( $\cdot$ ) in Definition 2.

##### 2) Subproblem of $\mathcal{B}$ :

After getting  $\mathcal{Z}_{k_1 k_2}$ , with the other parameters fixed, the solution of  $\mathcal{B}$  is as follows:

$$\begin{aligned} & \mathcal{B}^{(j+1)} \\ &= \arg \min_{\mathcal{B}} \mathcal{L} (\mathcal{Z}_{k_1 k_2}^{(j+1)}, \mathcal{P}_{k_1 k_2}^{(j)}, \mathcal{M}^{(j)}, \mathcal{B}, \mathcal{T}^{(j)}, \mathcal{E}^{(j)}, \mathcal{N}^{(j)}) \\ &= \frac{\sum_{1 \leq k_1 < k_2 \leq N} \beta_{k_1 k_2} \left( \mathcal{Z}_{k_1 k_2}^{(j+1)} - \frac{\mathcal{P}_{k_1 k_2}^{(j)}}{\beta_{k_1 k_2}} \right)}{\rho + \sum_{1 \leq k_1 < k_2 \leq N} \beta_{k_1 k_2}} \\ &+ \frac{\rho \left( \mathcal{D} - \mathcal{T}^{(j)} - \mathcal{E}^{(j)} - \mathcal{N}^{(j)} + \frac{\mathcal{M}^{(j)}}{\rho} \right)}{\rho + \sum_{1 \leq k_1 < k_2 \leq N} \beta_{k_1 k_2}}. \end{aligned} \quad (36)$$

##### 3) Subproblem of $\mathcal{T}$ :

Updating  $\mathcal{T}$  with the other parameters fixed:

$$\mathcal{T}^{(j+1)}$$

$$\begin{aligned} &= \arg \min_{\mathcal{T}} \mathcal{L} (\mathcal{Z}_{k_1 k_2}^{(j+1)}, \mathcal{P}_{k_1 k_2}^{(j)}, \mathcal{M}^{(j)}, \mathcal{B}^{(j+1)}, \mathcal{T}, \mathcal{E}^{(j)}, \mathcal{N}^{(j)}) \\ &= \arg \min_{\mathcal{T}} \lambda_1 \| \mathcal{W}^{(j)} \odot \mathcal{T} \|_1 + \frac{\rho}{2} \left\| \mathcal{D} - \mathcal{B}^{(j+1)} - \mathcal{T} - \mathcal{E}^{(j)} - \mathcal{N}^{(j)} + \frac{\mathcal{M}^{(j)}}{\rho} \right\|_F^2. \end{aligned} \quad (37)$$

And (37) can be solved efficiently by soft threshold shrinkage operator:

$$\mathcal{T}^{(j+1)} = \mathcal{S}_{\frac{\lambda_1 \mathcal{W}^{(j)}}{\rho}} \left( \mathcal{D} - \mathcal{B}^{(j+1)} - \mathcal{E}^{(j)} - \mathcal{N}^{(j)} + \frac{\mathcal{M}^{(j)}}{\rho} \right), \quad (38)$$

where,  $\mathcal{S}(\cdot)$  denotes the soft threshold shrinkage operator:

$$\mathcal{S}_\xi(x) = \text{sign}(x) \cdot \max(|x| - \xi, 0). \quad (39)$$

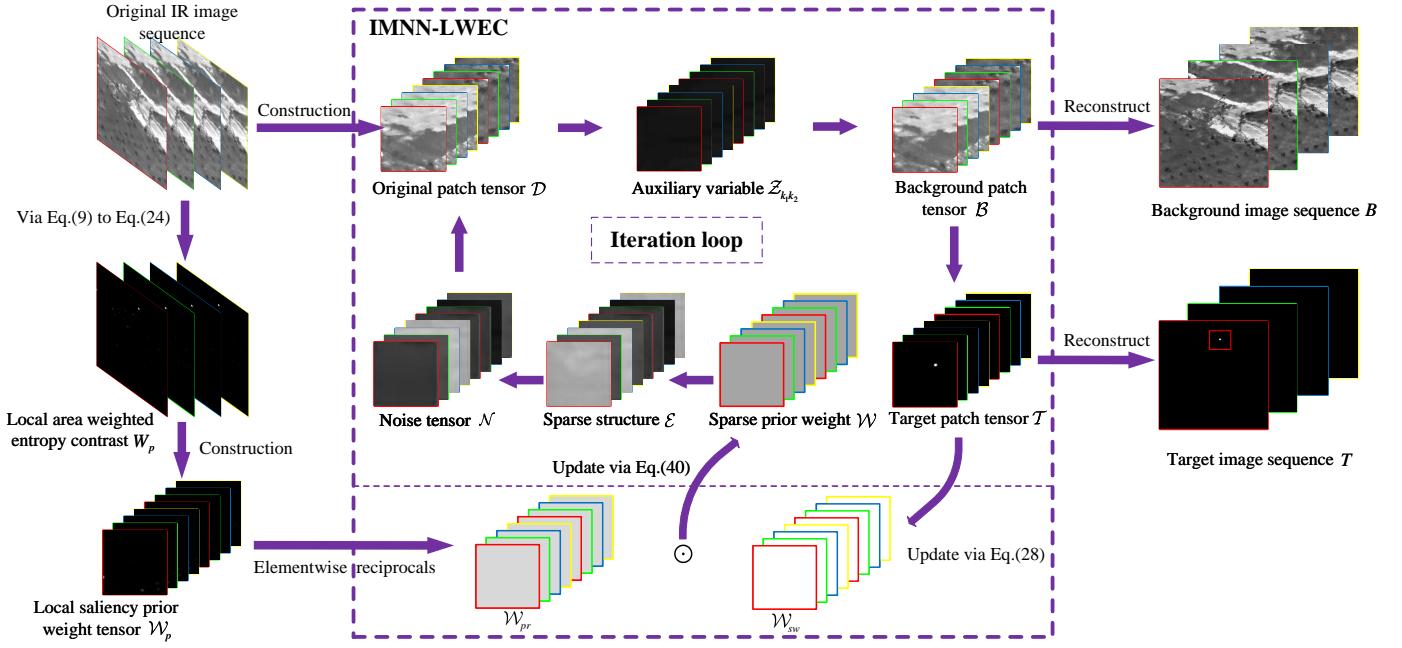


Fig. 7. Overview of the proposed IMNN-LWEC model in this paper.

4) Subproblem of  $\mathcal{W}$  :

The weight tensor  $\mathcal{W}$  can be updated by

$$\mathcal{W}^{(j+1)} = \mathcal{W}_{pr} \odot \mathcal{W}_{sw}^{(j)}. \quad (40)$$

5) Subproblem of  $\mathcal{E}$  :

Updating  $\mathcal{E}$  with the other parameters fixed, and the  $\mathcal{E}$ -subproblem is as follows:

$$\begin{aligned} \mathcal{E}^{(j+1)} &= \arg \min_{\mathcal{E}} \mathcal{L} \left( \mathcal{Z}_{k_1 k_2}^{(j+1)}, \mathcal{P}_{k_1 k_2}^{(j)}, \mathcal{M}^{(j)}, \mathcal{B}^{(j+1)}, \mathcal{T}^{(j+1)}, \mathcal{E}, \mathcal{N}^{(j)} \right) \\ &= \arg \min_{\mathcal{E}} \lambda_2 \|\mathcal{E}\|_{tw} \\ &\quad + \frac{\rho}{2} \left\| \mathcal{D} - \mathcal{B}^{(j+1)} - \mathcal{T}^{(j+1)} - \mathcal{E} - \mathcal{N}^{(j)} + \frac{\mathcal{M}^{(j)}}{\rho} \right\|_F^2. \end{aligned} \quad (41)$$

Let  $\mathcal{I} = \mathcal{D} - \mathcal{B}^{(j+1)} - \mathcal{T}^{(j+1)} - \mathcal{N}^{(j)} + \frac{\mathcal{M}^{(j)}}{\rho}$ , the solution of updating  $\mathcal{E}$  in (41) is given by

$$\mathcal{E}^{(j+1)}(:, n, :) = \max \left( 1 - \frac{\lambda_2}{\rho \|\mathcal{I}(:, n, :)\|_2}, 0 \right) * \mathcal{I}(:, n, :), \quad (42)$$

and

$$\mathcal{E}^{(j+1)}(m, :, :) = \max \left( 1 - \frac{\lambda_2}{\rho \|\mathcal{I}(m, :, :)\|_2}, 0 \right) * \mathcal{I}(m, :, :), \quad (43)$$

where,  $m = 1, \dots, n_1$ ,  $n = 1, \dots, n_2$ .

6) Subproblem of  $\mathcal{N}$ :

Updating  $\mathcal{N}$  with the other parameters fixed:

$$\begin{aligned} \mathcal{N}^{(j+1)} &= \arg \min_{\mathcal{N}} \mathcal{L} \left( \mathcal{Z}_{k_1 k_2}^{(j+1)}, \mathcal{P}_{k_1 k_2}^{(j)}, \mathcal{M}^{(j)}, \mathcal{B}^{(j+1)}, \mathcal{T}^{(j+1)}, \mathcal{E}^{(j+1)}, \mathcal{N} \right) \\ &= \arg \min_{\mathcal{N}} \lambda_3 \|\mathcal{N}\|_F^2 + \frac{\rho}{2} \left\| \mathcal{D} - \mathcal{B}^{(j+1)} - \mathcal{T}^{(j+1)} - \mathcal{E}^{(j+1)} - \mathcal{N} + \frac{\mathcal{M}^{(j)}}{\rho} \right\|_F^2 \\ &= \frac{\rho (\mathcal{D} - \mathcal{B}^{(j+1)} - \mathcal{T}^{(j+1)} - \mathcal{E}^{(j+1)}) + \mathcal{M}^{(j)}}{2\lambda_3 + \rho}. \end{aligned} \quad (44)$$

7) Subproblem of  $\mathcal{P}_{k_1 k_2}$ ,  $\mathcal{M}$ ,  $\rho$  and  $\beta$ :

Lagrange multipliers  $\mathcal{P}_{k_1 k_2}$ ,  $\mathcal{M}$  and penalty parameters  $\rho$  and  $\beta$  can be updated by

$$\mathcal{P}_{k_1 k_2}^{(j+1)} = \mathcal{P}_{k_1 k_2}^{(j)} + \beta_{k_1 k_2} \left( \mathcal{B}^{(j+1)} - \mathcal{Z}_{k_1 k_2}^{(j+1)} \right), \quad (45)$$

$$\mathcal{M}^{(j+1)} = \mathcal{M}^{(j)} + \rho \left( \mathcal{D} - \mathcal{B}^{(j+1)} - \mathcal{T}^{(j+1)} - \mathcal{E}^{(j+1)} - \mathcal{N}^{(j+1)} \right), \quad (46)$$

$$\rho = \min(\kappa\rho, \rho_{\max}), \quad (47)$$

$$\beta = \min(\kappa\beta, \beta_{\max}). \quad (48)$$

## 8) Iteration Criterion:

When the relative error  $\|\mathcal{D} - \mathcal{B} - \mathcal{T} - \mathcal{E} - \mathcal{N}\|_F / \|\mathcal{D}\|_F$  is less than  $\delta$ , iteration terminates. Besides, to promote computational efficiency, once the number of nonzero elements in target tensor  $\mathcal{T}$  no longer changes (i.e.,  $\|\mathcal{T}\|_0$ ), iteration ends.

---

**Algorithm 2** Proposed IMNN-LWEC Model Solved by ADMM
 

---

**Input:** Original IR image sequence  $D_1, D_2, \dots, D_s \in \mathbb{R}^{m_1 \times m_2}$ .

**Output:**  $\mathcal{B}, \mathcal{T}, \mathcal{E}, \mathcal{N}$ .

- 1: Transform the IR image sequence into a tensor  $\mathcal{D}$ .
- 2: Calculate local saliency prior maps via **III-C**.
- 3: Construct  $t$  frame prior maps to obtain tensor  $\mathcal{W}_p$ .
- 4: Calculate the reciprocal of the corresponding elements in  $\mathcal{W}_p$  to obtain  $\mathcal{W}_{pr}$ .
- 5: Obtain the final optimization model (30) by regularizing the model with  $\mathcal{W}$ .
- 6: **Initialization:**  $\mathcal{B}^{(0)} = \mathcal{T}^{(0)} = \mathcal{E}^{(0)} = \mathcal{N}^{(0)} = \mathcal{Z}_{k_1 k_2}^{(0)} = \mathcal{P}_{k_1 k_2}^{(0)} = \mathcal{M}^{(0)} = j = 0, \kappa = 1.5, \delta = 1e - 4, \beta_{max} = \rho_{max} = 1e5, J = 400$ .
- 7: **While** not converged **do**.
- 8: Fix the others and update  $\mathcal{Z}_{k_1 k_2}^{(j+1)}$  by (35).
- 9: Fix the others and update  $\mathcal{B}^{(j+1)}$  by (36).
- 10: Fix the others and update  $\mathcal{T}^{(j+1)}$  by (38).
- 11: Fix the others and update  $\mathcal{W}^{(j+1)}$  by (40).
- 12: Fix the others and update  $\mathcal{E}^{(j+1)}$  by (42) and (43).
- 13: Fix the others and update  $\mathcal{N}^{(j+1)}$  by (44).
- 14: Fix the others and update  $\mathcal{P}_{k_1 k_2}^{(j+1)}$  by (45).
- 15: Fix the others and update  $\mathcal{M}^{(j+1)}$  by (46).
- 16: Update  $\rho$  by (47).
- 17: Update  $\beta$  by (48).
- 18: Check the convergence conditions.  

$$\frac{\|\mathcal{D} - \mathcal{B}^{(j+1)} - \mathcal{T}^{(j+1)} - \mathcal{E}^{(j+1)} - \mathcal{N}^{(j+1)}\|_F}{\|\mathcal{D}\|_F} < \delta$$
or  $\|\mathcal{T}^{(j+1)}\|_0 = \|\mathcal{T}^{(j)}\|_0$ .
- 19: Let  $j = j + 1$ .
- 20: **End While**
- 21:  $\mathcal{B}^{(j+1)}, \mathcal{T}^{(j+1)}, \mathcal{E}^{(j+1)}, \mathcal{N}^{(j+1)}$ .

---

### G. Detection Procedure

The detailed process of the ADMM solving the optimization model is summarized in Algorithm 2. And in Fig. 7, we present the whole solving procedure of the proposed IMNN-LWEC model in this paper, which can be described as follows:

(1) Given an IR sequence  $f_D$ , sliding a window of size  $ps \times ps$  to traverse each frame and obtaining  $n$  nonoverlapping patches in each frame, followed by constructing a spatial-temporal patch-tensor  $\mathcal{D}$  in a certain way mentioned in **III-A**.

(2) Calculating the local structure feature maps  $\mathcal{W}_p$  via (9)-(24) and transform them into a spatial-temporal tensor  $\mathcal{W}_p$ .

(3) Establish the model (30). And according to Algorithm 2, the spatial-temporal patch-tensor  $\mathcal{D}$  is decomposed into target patch tensor  $\mathcal{T}$  and background patch tensor  $\mathcal{B}$  through iteration.

(4) The final target image sequence  $T$  and the background image sequence  $B$  are reconstructed from sparse tensor  $\mathcal{T}$  and low-rank tensor  $\mathcal{B}$ .

## IV. EXPERIMENTS AND ANALYSIS

In this section, we first introduce the experimental settings and evaluation indicators derived from 3-D ROC curve. Then,

we conduct parameter analysis and discuss the influences of the different parameters on the proposed method. To verify the effectiveness of IMNN-LWEC, it is compared with nine state-of-the-art approaches on comprehensive experiments. Besides, we test the detection performance of our method in noisy scenes. To testify the function of each designed module, we also conduct sufficient ablation study. Finally, we briefly analyze the convergence of our algorithm.

### A. Experimental Setting

1) **Data Sets:** To evaluate the effectiveness and feasibility of the proposed method more comprehensively, we test the proposed method on five extensive real IR image sequences to cover different complex scenarios. For the simplicity, the representative images of five sequences are shown in Fig. 8. For better visualization, the three-dimensional surface diagram of each representative frame is exhibited, and the target area marked in red box is enlarged, which can be better seen in the right-bottom part by zooming on a computer screen. It is clearly that the contrast between target and background is not large and target is even submerged in the background. The detailed description of the test data sets is given in Table I. And these datasets with complex backgrounds are original from National University of Defense Technology (NUDT) online, whose URL is <http://www.sciedb.cn/dataSet/handle/902>.

2) **Competitive Methods:** To reflect the advantages of the proposed method, we make a comparison with nine state-of-the-art single-frame and multi-frame detection methods for IR small target detection in the following experiments. Single-frame detection methods include TLLCM [14], IPI [18], NRAM [20], NIRPS [60], RIPT [21] and LogTFNN [24]. Other three sequential detection methods are WSNM-STIPT [61], TCNN-NPSTT [32] and ASTTV-NTLA [33].

The codes of all competitive methods are provided by authors and parameter settings of these methods are given in Table II. All experiments are implemented with MATLAB R2021b software in Windows 10 with CPU Intel Core i7-11700K @3.60GHz, 16-GB memory.

### B. 3-D ROC Curve-Derived Evaluation Metrics

IR dim and small target detection is essentially an anomaly detection task. Background suppression capability and target detection capability are significant criteria for evaluating the performance of target detection methods. Nowadays, the background suppression factor (BSF) [62], the signal-to-clutter ratio gain (SCRG) [63], the receiver operating characteristic (ROC) curve of detection probability ( $P_D$ ) versus false alarm rate ( $P_F$ ) [64], and the area under an ROC curve (AUC) are common evaluation metrics in the field of IR small target detection. BSF is the ratio of the standard deviation of clutters in input image and output image. When both of targets and clutters are strongly suppressed, BSF is satisfactory but recall is disappointed. Hence, BSF is a misleading statistic for background suppression. SCRG is defined by the SCR of the detection result image  $SCR_{detection}$  and the original input image  $SCR_{origin}$ . The process of calculating the SCR utilizes a fixed target region and a local background neighborhood, which is

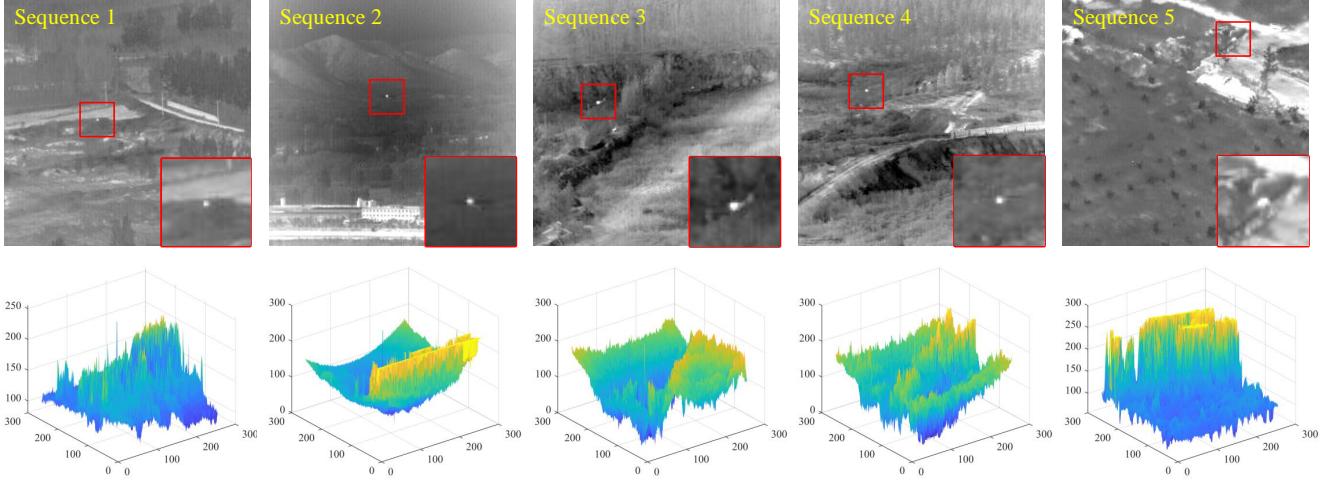


Fig. 8. Representative images of five sequences with corresponding three-dimensional surface diagrams.

TABLE I  
CHARACTERISTICS OF THE DATASETS

	Frames	Image Size	Target and Background Description
Sequence 1	100	$256 \times 256$	<ul style="list-style-type: none"> <li>• Many linear structures and target-like clutters</li> <li>• Low local contrast and heavy noise</li> <li>• Tiny aircraft with background moving slowly</li> </ul>
Sequence 2	100	$256 \times 256$	<ul style="list-style-type: none"> <li>• Low-altitude background</li> <li>• Background moving slowly</li> <li>• Strong radiation building clutters and noise</li> </ul>
Sequence 3	100	$256 \times 256$	<ul style="list-style-type: none"> <li>• Air-ground background</li> <li>• Tiny helicopter with low local contrast</li> <li>• Salient suspected targets</li> </ul>
Sequence 4	400	$256 \times 256$	<ul style="list-style-type: none"> <li>• Sky-ground background</li> <li>• Background moving slowly</li> <li>• Strong radiation interference</li> </ul>
Sequence 5	56	$256 \times 256$	<ul style="list-style-type: none"> <li>• Ground-sky background with suspicious targets</li> <li>• Bright and strong radiation clutter</li> <li>• High brightness building edges</li> </ul>

TABLE II  
DETAILED PARAMETER SETTINGS FOR COMPETITIVE METHODS

Methods	Parameter Settings
TLLCM [14]	$K = 9$ , Gaussian filter size: $3 \times 3$ .
IPI [18]	Patchsize: $45 \times 45$ , Sliding Step: 10, $\lambda = \frac{1}{\sqrt{\min(m,n)}}$ , $\varepsilon = 10^{-4}$ .
NRAM [20]	Patchsize: $30 \times 30$ , Sliding Step: 25, $\lambda = \frac{1}{\min(m,n)}$ , $\varepsilon = 10^{-7}$ , $\mu^0 = 3\sqrt{\min(m,n)}$ , $\gamma = 0.003$ , $C = \sqrt{\min(m,n)}/2.5$ .
NIRPS [60]	$s = 2$ , $\lambda = \frac{2}{\sqrt{\min(m,n)}}$ , $\mu = 1.25/\ D'\ _2$ , $\rho = 1.5$ , $\varepsilon = 10^{-7}$ .
RIPT [21]	Patchsize: $30 \times 30$ , Sliding Step: 10, $\lambda = \frac{L}{\sqrt{\min(I,J,P)}}$ , $L = 0.7$ , $\varepsilon = 10^{-8}$ , $h = 1$ .
LogTFNN [24]	Patchsize: $40 \times 40$ , Sliding Step: 40, $\lambda = \frac{L}{\sqrt{\max(I,J)*P}}$ , $L = 0.15$ , $\beta = 10^{-4}$ , $\varepsilon = 10^{-6}$ .
WSNM-STIPT [61]	$\mu = 0.02$ , $p = 0.8$ , $C = 5$ , $\lambda = \frac{5}{\sqrt{\max(m,n)*L}}$ , $L = 4$ , $\varepsilon = 10^{-7}$ .
TCNN-NPSTT [32]	Patchsize: $45 \times 45$ , Sliding Step : 45, frame : 3, $\lambda = \frac{L}{\sqrt{\min(n_1,n_2)*n_3}}$ , $\alpha = 10^{-3}$ , $\varepsilon = 0.01$ .
ASTTV-NTLA [33]	$L = 3$ , $H = 8$ , $\lambda_{tv} = 0.005$ , $\lambda_s = \frac{H}{\sqrt{\max(M,N)*L}}$ , $\lambda_3 = 100$ .
Proposed	frame: 3, Patchsize: $15 \times 15$ , Sliding Step: 15, $\omega: 5$ , $\lambda_1 = \frac{\lambda_L}{\sqrt{\max(n_1,n_2)*n_3}}$ , $\lambda_2 = \lambda_3 = 50\lambda_1$ , $\lambda_L = 1.4$ , $\beta_{k_1 k_2} = \rho = \frac{1}{\omega}$ , $\kappa = 1.5$ , $\varepsilon = 10^{-4}$ .

tough to adapt to varying size of moving targets. Besides, Prof. Chang and Chen [65]–[67] have proved that only relying on true positive rate (TPR) and false positive rate (FPR) is not convincing but rather misleading, because TPR and FPR are determined by the same cut-off threshold  $\tau$ , making  $P_D$  and  $P_F$  tied together simultaneously and cannot work independently. Therefore, the ROC curve and AUC are not creditable enough to measure the effectiveness of an anomaly detector, i.e., TPR and FPR cannot be alone to determine the target detectability (TD) and background suppressibility (BS). To tackle these limitations above, an effective 3-D ROC analysis-based evaluation tool [65], which contains an additional changing parameter  $\tau$  and a series of related detection evaluation metrics, is more effective to evaluate the performance of TD and BS of an anomaly detector. In summary, the metrics derived from AUC are summarized as follows:

(1)  $AUC_{(D,F)}$ : AUC of a  $(P_D, P_F)$  curve, an effectiveness evaluation of a detector, defined by

$$AUC_{(D,F)} \in [0, 1]. \quad (49)$$

(2)  $AUC_{(D,\tau)}$ : AUC of a  $(P_D, \tau)$  curve, target detectability of a detector, defined by

$$AUC_{(D,\tau)} \in [0, 1]. \quad (50)$$

(3)  $AUC_{(F,\tau)}$ : AUC of a  $(P_F, \tau)$  curve, background suppression of a detector, defined by

$$AUC_{(F,\tau)} \in [0, 1]. \quad (51)$$

(4)  $AUC_{TD}$ : Joint target detectability (TD) of a detector, defined by

$$AUC_{TD} = AUC_{(D,F)} + AUC_{(D,\tau)} \in [0, 2]. \quad (52)$$

(5)  $AUC_{BS}$ : Joint background suppression (BS) of a detector, defined by

$$AUC_{BS} = AUC_{(D,F)} + AUC_{(F,\tau)} \in [-1, 1]. \quad (53)$$

(6)  $AUC_{SNPR}$ : Signal-to-noise probability ratio (SNPR) of a detector, defined by

$$AUC_{SNPR} = \frac{AUC_{(D,\tau)}}{AUC_{(F,\tau)}} \in [0, +\infty). \quad (54)$$

(7)  $AUC_{TDBS}$ : Joint target detectability and background suppression (TDBS) of a detector, defined by

$$AUC_{TDBS} = AUC_{(D,\tau)} - AUC_{(F,\tau)} \in [-1, 1]. \quad (55)$$

(8)  $AUC_{ODP}$ : Overall detection probability (ODP) of a detector, defined by

$$AUC_{ODP} = AUC_{(D,F)} + AUC_{(D,\tau)} - AUC_{(F,\tau)} \in [-1, 2]. \quad (56)$$

### C. Parameters Analysis

The key parameters of the proposed method include the number of frames  $k$ , the patch size  $ps$ , the trade-off factors  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ , as well as the penalty factors  $\beta_{k_1 k_2}$  and  $\rho$ . The above parameters will greatly affect the effectiveness and the robustness of the algorithm. Then, we conduct 3-D ROC analysis of varying parameters on Sequence 1. Since it is hard to obtain the optimal values of all parameters simultaneously, we adjust one parameter while fixing other parameters to obtain the local optimal or suboptimal value of the model under each parameter.

1) *Analysis of frame number  $k$* :  $k$  frame images are used to construct nonoverlapping spatial-temporal patch-tensor. Noting that  $k$  is a key parameter, which determines the structure of the spatial-temporal patch-tensor and algorithm complexity. In the experiments, we test the algorithm effectiveness with the value of  $k$  set of 2, 3, 4, 5, 6 and 7. Parameter  $k$  should make the algorithm have better target detectability and background suppressibility. Finally, we choose  $k = 3$  according to the corresponding 3-D ROC analysis results in Table III.

2) *Analysis of patchsize  $ps$* : The size of each image patch is another critical parameter, which affects the time complexity of the algorithm directly. Large patchsize can ensure the sparsity of targets, but too large patchsize will mistake the sparse structures for targets and cause false alarms. Conversely, small patchsize will weaken the sparseness of small targets. Obviously, it is essential to guarantee that IR target is sufficiently sparse. In the experiments, we traverse  $ps$  from 10 to 40 with an interval of 5, and tabulate the AUC results in Table III. Small patchsize can strengthen TD of IMNN-LWEC, but its BS is weakened. And a relatively large patchsize produces balanced performance of background suppression and target detectability. Based on the experimental results, we determine the patchsize  $ps$  to be 15.

3) *Analysis of trade-off factors  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$* : The trade-off factors  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  are important parameters for IMNN-LWEC as well. A larger  $\lambda_1$  will make more details remain in the background, but part of targets may also be preserved in the background, i.e., missing detection occurs. On the contrary, the smaller  $\lambda_1$ , the stronger target detectability, while false alarm is likely to occur. Accordingly, it's necessary to choose appropriate trade-off factors to balance TD and BS. We set  $\lambda_1$  as  $10^5 \lambda_L [\max(n_1, n_2) \times n_3]^{-1/2}$  and  $\lambda_2 = \lambda_3 = 50\lambda_1$ . Thus, we only need to search the optimal or suboptimal trade-off coefficient  $\lambda_L$  to obtain trade-off factors  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ . In the experiments, we select  $\lambda_L$  from 0.6 to 2.2 with an interval of 0.4. Table III shows that the most suitable  $\lambda_L$  is 1.4.

4) *Analysis of penalty factors  $\beta_{k_1 k_2}$  and  $\rho$* : There are four penalty parameters  $\rho$ ,  $\beta_{k_1 k_2}$  ( $1 \leq k_1 < k_2 \leq 3$ ) in Algorithm 2 and they greatly influence the detection performance. To facilitate the execution of the algorithm, assuming that these penalty parameters ( $\beta_{k_1 k_2}$  and  $\rho$ ) are inversely proportional to the penalty coefficient  $\omega$ :

$$\beta_{k_1 k_2} = \frac{\beta_L}{\omega}, \quad (57)$$

$$\rho = \frac{1}{\omega}. \quad (58)$$

TABLE III  
3-D ROC-DERIVED AUC RESULTS OF PARAMETER ANALYSIS FOR SEQUENCE 1.

$k$	$ps$	$\lambda_L$	$\omega$	$AUC_{(D,F)}$	$AUC_{(D,\tau)}$	$AUC_{(F,\tau)}$	$AUC_{TD}$	$AUC_{BS}$	$AUC_{SNPR}$	$AUC_{TDBS}$	$AUC_{ODP}$
2	15	1.4	5	<b>1.0000</b>	0.9901	<b>0.0050</b>	1.9900	<b>0.9950</b>	1.9793e2	0.9850	1.9850
3	15	1.4	5	<b>1.0000</b>	<b>1.0000</b>	<b>0.0050</b>	<b>2.0000</b>	<b>0.9950</b>	<b>1.9913e2</b>	<b>0.9950</b>	<b>1.9950</b>
4	15	1.4	5	<b>1.0000</b>	0.9999	0.0051	1.9999	0.9949	1.9681e2	0.9948	1.9948
5	15	1.4	5	<b>1.0000</b>	<b>1.0000</b>	0.0052	<b>2.0000</b>	0.9948	1.9417e2	0.9948	1.9948
6	15	1.4	5	<b>1.0000</b>	<b>1.0000</b>	0.0052	<b>2.0000</b>	0.9948	1.9197e2	0.9948	1.9948
7	15	1.4	5	<b>1.0000</b>	<b>1.0000</b>	0.0053	<b>2.0000</b>	0.9947	1.8990e2	0.9947	1.9947
3	10	1.4	5	<b>1.0000</b>	<b>1.0000</b>	0.0052	<b>2.0000</b>	0.9948	1.9368e2	0.9948	1.9948
3	15	1.4	5	<b>1.0000</b>	<b>1.0000</b>	<b>0.0050</b>	<b>2.0000</b>	<b>0.9950</b>	<b>1.9913e2</b>	<b>0.9950</b>	<b>1.9950</b>
3	20	1.4	5	<b>1.0000</b>	0.9901	<b>0.0050</b>	1.9900	<b>0.9950</b>	1.9796e2	0.9850	1.9850
3	25	1.4	5	<b>1.0000</b>	0.9901	<b>0.0050</b>	1.9900	<b>0.9950</b>	1.9800e2	0.9850	1.9850
3	30	1.4	5	<b>1.0000</b>	0.9901	<b>0.0050</b>	1.9901	<b>0.9950</b>	1.9801e2	0.9851	1.9851
3	35	1.4	5	<b>1.0000</b>	0.9901	<b>0.0050</b>	1.9901	<b>0.9950</b>	1.9801e2	0.9851	1.9851
3	40	1.4	5	<b>1.0000</b>	0.9005	<b>0.0050</b>	1.9005	<b>0.9950</b>	1.8010e2	0.8955	1.8955
3	15	0.6	5	<b>1.0000</b>	<b>1.0000</b>	0.0053	<b>2.0000</b>	0.9947	1.8858e2	0.9947	1.9947
3	15	1.0	5	<b>1.0000</b>	<b>1.0000</b>	0.0051	<b>2.0000</b>	0.9949	1.9593e2	0.9949	1.9949
3	15	1.4	5	<b>1.0000</b>	<b>1.0000</b>	<b>0.0050</b>	<b>2.0000</b>	<b>0.9950</b>	<b>1.9913e2</b>	<b>0.9950</b>	<b>1.9950</b>
3	15	1.8	5	<b>1.0000</b>	0.9901	<b>0.0050</b>	1.9900	<b>0.9950</b>	1.9782e2	0.9850	1.9850
3	15	2.2	5	<b>1.0000</b>	0.9901	<b>0.0050</b>	1.9900	<b>0.9950</b>	1.9795e2	0.9850	1.9850
3	15	1.4	1	<b>1.0000</b>	<b>1.0000</b>	0.0057	<b>2.0000</b>	0.9943	1.7669e2	0.9943	1.9943
3	15	1.4	5	<b>1.0000</b>	<b>1.0000</b>	<b>0.0050</b>	<b>2.0000</b>	<b>0.9950</b>	<b>1.9913e2</b>	<b>0.9950</b>	<b>1.9950</b>
3	15	1.4	9	<b>1.0000</b>	0.9901	<b>0.0050</b>	1.9900	<b>0.9950</b>	1.9798e2	0.9850	1.9850
3	15	1.4	13	<b>1.0000</b>	0.9901	<b>0.0050</b>	1.9901	<b>0.9950</b>	1.9801e2	0.9851	1.9851
3	15	1.4	17	<b>1.0000</b>	0.9901	<b>0.0050</b>	1.9901	<b>0.9950</b>	1.9801e2	0.9851	1.9851
3	15	1.4	21	<b>1.0000</b>	0.0050	<b>0.0050</b>	1.0050	<b>0.9950</b>	1.0000e0	0.0000	1.0000

Note: the bold data represent the best performance for the corresponding AUC results.

In this paper, we set  $\beta_L = 1$  and fine-tune the coefficient  $\omega$ . The larger  $\omega$ , the smaller penalty parameters, and the stronger penalty for sparse component. Therefrom, the target detection algorithm has the effect of denoising and reducing false alarms, i.e., the target-like pixels will be suppressed, but may cause the over-shrinking of the targets. Conversely, the smaller the penalty coefficient  $\omega$ , the larger the penalty parameters, resulting in increasing false alarms. We vary  $\omega$  from set  $\{1, 5, 9, 13, 17, 21\}$ , and list the 3-D ROC curve-derived metrics in Table III. It claims that the compromising  $\omega$  to get effective detection performance is 5 for practical scenarios.

#### D. Analysis and Comparison of Target Detection Performance

1) *Qualitative Comparison:* Fig. 9 presents the visual results of the five datasets via nine competitive methods and ours. Real targets are marked with red boxes in the grayscale images, whose corners are enlarged view of the red boxes. For a more intuitive comparison, the 3-D gray distribution of each detection result is also presented. For the simplicity, we only show the detection result of a representative frame of each dataset.

TLLCM is based on human visual saliency (HVS). It mistakes many bright point-like clutters for targets and there still exist weak clutters in the 3-D gray distribution map. Besides, the target in Sequence 1 is significantly suppressed by TLLCM. And IPI, NRAM, NIRPS, RIPT and LogTFNN are single-frame methods based on optimization. Detection results on five sequences jointly demonstrate that IPI is extremely weak in background suppression because the background radiation values are much higher than 0. Detection results of LogTFNN have heavy residual background of line-featured

edge and architecture areas. NRAM and NIRPS have a certain effect in IR small target detection, but highlights and edges in images will cause false alarms, and target distortion phenomenon appears in NRAM. As for RIPT, combining the detection maps and the 3-D gray distribution maps, it is weak in background suppression and leads to target over-shrinking dilemma. WSNM-STIPT, TCNN-NPSTT, ASTTV-NTLA and IMNN-LWEC are sequential detection methods based on optimization. These methods utilize spatial-temporal information to detect IR small targets, but background clutters are remained in the target image and the targets are severely distorted by WSNM-STIPT and TCNN-NPST in Sequence 4 to 5. Particularly, the corresponding 3-D gray distribution maps of ASTTV-NTLA's detection results indicate that this approach basically cannot completely suppress the background, especially bright building edges and line-featured areas.

In contrast, our proposed method is able to retain relatively complete shaped targets and has the best visual detection effect on five image sequences. The target level is significant and the background level is almost 0 in 3-D gray distribution maps, illustrating that IMNN-LWEC can not only accurately extract the targets but also suppress almost all substantial and bright edge interference and clutters. Compared with other nine competitive methods, the proposed IMNN-LWEC in this paper achieves the best visual detection effect.

2) *Quantitative Comparison:* The proposed method produces visually superior results in IV-D1. Furthermore, we perform a detailed quantitative analysis of the competitive methods and IMNN-LWEC via the 3-D ROC analysis. Table IV-VIII tabulate the AUC results derived from 3-D ROC analysis-based evaluation tool, i.e.,  $AUC_{(D,F)}$ ,  $AUC_{(D,\tau)}$ ,  $AUC_{(F,\tau)}$ ,  $AUC_{TD}$ ,  $AUC_{BS}$ ,  $AUC_{SNPR}$ ,  $AUC_{TDBS}$  and  $AUC_{ODP}$ . And the

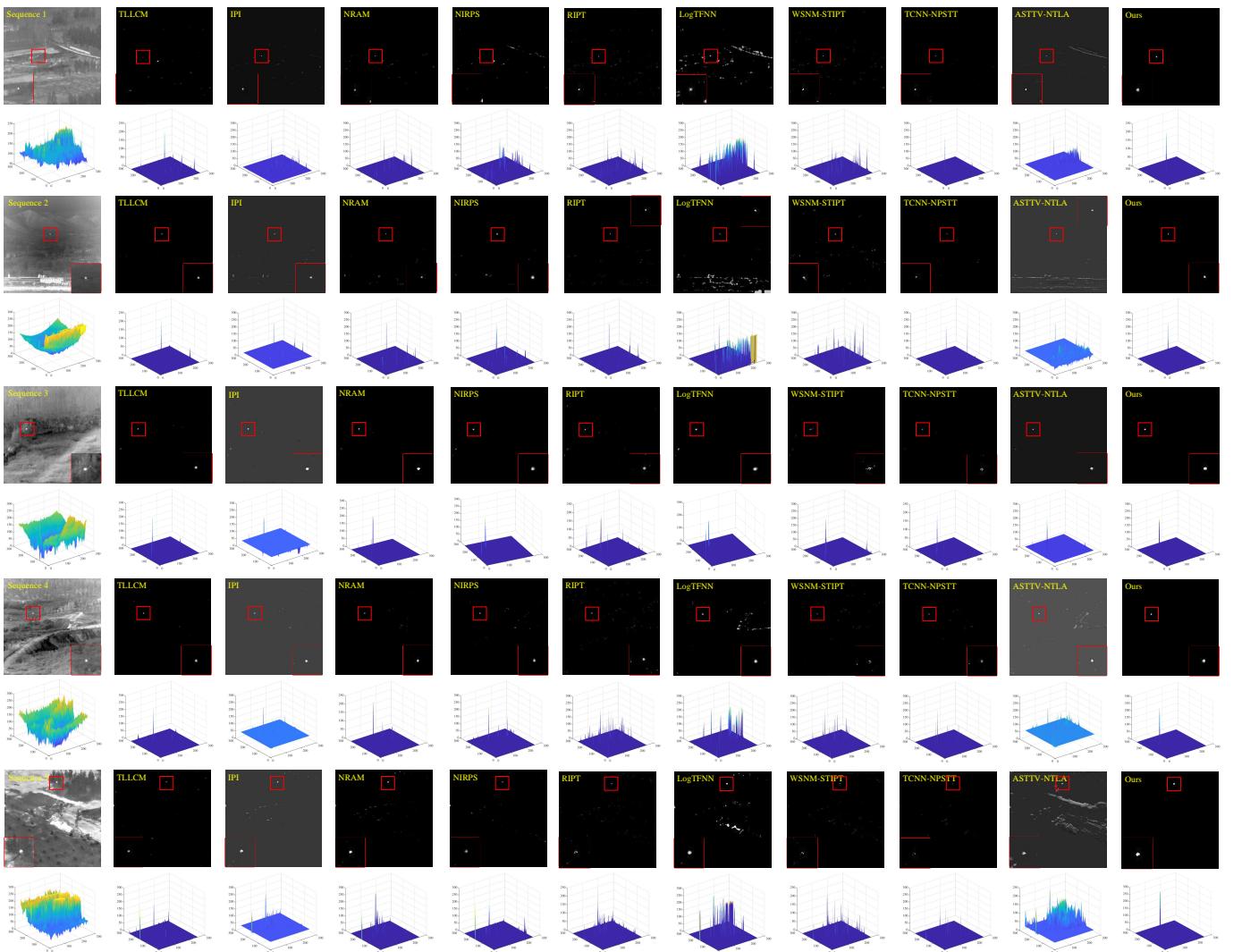


Fig. 9. Comparison of qualitative results and their corresponding intuitive 3-D displays for Sequence 1 to Sequence 5. Targets are marked with red boxes and enlarged in the corners.

bold and underlined data represent the optimal and suboptimal performance for the corresponding AUC results, respectively. Besides, Fig. 10-14 present three 2-D ROC curves of  $(P_D, P_F)$ ,  $(P_D, \tau)$ ,  $(P_F, \tau)$  and 3-D ROC curves  $(P_D, P_F, \tau)$ . To make the comparison more obvious, the log scale is used on the  $x$ -axis of 2-D ROC curves, and a partial enlarged view of  $(P_F, \tau)$  curve is placed in the left-bottom corner.

For Sequence 1, whose images contain lots of bright edge structures, target-like points and heavy noise. The 3-D ROC-derived curves and the corresponding AUC metrics for Sequence 1 are given in Fig. 10 and Table IV. Statistically speaking, the  $AUC_{(D,F)}$  of eight methods such as IMNN-LWEC and IPI achieve 1.0000. And we hardly determine which anomaly detector is the best, directly verifying that the traditional ROC curve has limitations to measure detection methods. Fortunately, by introducing a threshold  $\tau$ , the 3-D ROC analysis tool is suitable to evaluate the effectiveness of an anomaly detector. For any target detection methods, both satisfactory TD and BS are required. Noting that BS and TD are contradictory indicators, in other word, increasing BS will

sacrifice TD, and increasing TD will lead to lower BS instead. Although the target detection ability of IPI, RIPT and TCNN-NPSTT are as pretty as ours, their background suppression is insufficient. It shows that IMNN-LWEC has the best results of 3-D ROC curves along with the corresponding 2-D ROC curves and eight 3-D ROC-derived AUC results. Since there are many high-brightness edge structures in Sequence 1, the satisfactory results illustrate that IMNN-LWEC can alleviate the contradiction between BS and TD to a certain extent, and can improve them at the same time.

There are strong bright buildings and noise in Sequence 2. Fig. 11 and Table V are the 3-D ROC-derived curves and AUC results, respectively. In Fig. 11, IMNN-LWEC outperforms other competitive methods on the  $(P_F, \tau)$  curve, corresponding to its  $AUC_{(F,\tau)}$  in Table V achieves the optimal value. In Table V, TLLCM, IPI, NRAM and LogTFNN achieve outstanding target detectability at the expense of background suppression, resulting in serious false alarms. Other sequential methods like WSNM-STIPT and TCNN-NPSTT are weak in TD. And ASTTV-NTLA is quite weak in terms of BS. As for IMNN-

TABLE IV  
3-D ROC-DERIVED AUC RESULTS OF COMPETITIVE METHODS FOR SEQUENCE 1.

Methods	AUC <sub>(D,F)</sub>	AUC <sub>(D,<math>\tau</math>)</sub>	AUC <sub>(F,<math>\tau</math>)</sub>	AUC <sub>TD</sub>	AUC <sub>BS</sub>	AUC <sub>SNPR</sub>	AUC <sub>TDBS</sub>	AUC <sub>ODP</sub>
TLLCM	0.9981	0.3513	0.0055	1.3494	0.9926	6.4188e1	0.3458	1.3439
IPI	<b>1.0000</b>	<b>1.0000</b>	0.0762	<b>2.0000</b>	0.9238	1.3118e1	0.9238	1.9238
NRAM	<b>1.0000</b>	0.9988	0.0055	1.9988	0.9945	1.8171e2	0.9933	1.9933
NIRPS	0.9984	0.5097	0.0071	1.5081	0.9913	7.2256e1	0.5026	1.5010
RIPT	<b>1.0000</b>	<b>1.0000</b>	0.0057	<b>2.0000</b>	0.9943	1.7439e2	0.9943	1.9943
LogTFNN	<b>1.0000</b>	0.9993	0.0137	<u>1.9993</u>	0.9863	7.2988e1	0.9856	1.9856
WSNM-STIPT	<b>1.0000</b>	0.9610	0.0053	1.9610	0.9947	1.8295e2	0.9557	1.9557
TCNN-NPSTT	<b>1.0000</b>	<b>1.0000</b>	<u>0.0051</u>	<b>2.0000</b>	<u>0.9949</u>	<u>1.9671e2</u>	<u>0.9949</u>	<u>1.9949</u>
ASTTV-NTLA	<b>1.0000</b>	0.9970	0.0683	1.9970	0.9317	1.4602e1	0.9287	1.9287
Proposed	<b>1.0000</b>	<b>1.0000</b>	<b>0.0050</b>	<b>2.0000</b>	<b>0.9950</b>	<b>1.9913e2</b>	<b>0.9950</b>	<b>1.9950</b>

Note: the bold and underlined data represent the optimal and suboptimal performance for the corresponding AUC results, respectively.

TABLE V  
3-D ROC-DERIVED AUC RESULTS OF COMPETITIVE METHODS FOR SEQUENCE 2.

Methods	AUC <sub>(D,F)</sub>	AUC <sub>(D,<math>\tau</math>)</sub>	AUC <sub>(F,<math>\tau</math>)</sub>	AUC <sub>TD</sub>	AUC <sub>BS</sub>	AUC <sub>SNPR</sub>	AUC <sub>TDBS</sub>	AUC <sub>ODP</sub>
TLLCM	<b>1.0000</b>	<b>1.0000</b>	<u>0.0051</u>	<b>2.0000</b>	<b>0.9949</b>	<u>1.9516e2</u>	<b>0.9949</b>	<u>1.9949</u>
IPI	<b>1.0000</b>	<b>1.0000</b>	0.1959	<b>2.0000</b>	0.8041	5.1056	0.8041	1.8041
NRAM	<b>1.0000</b>	<b>1.0000</b>	0.0053	<b>2.0000</b>	0.9947	1.8888e2	0.9947	1.9947
NIRPS	<b>1.0000</b>	0.9747	0.0055	1.9746	0.9945	1.7777e2	0.9692	1.9692
RIPT	<b>1.0000</b>	<u>0.9895</u>	0.0055	<u>1.9894</u>	0.9945	1.7918e2	0.9839	1.9839
LogTFNN	<b>1.0000</b>	<b>1.0000</b>	0.0116	<b>2.0000</b>	0.9884	8.5991e1	0.9884	1.9884
WSNM-STIPT	0.9999	0.8617	0.0052	1.8616	0.9947	1.6518e2	0.8564	1.8564
TCNN-NPSTT	<b>1.0000</b>	0.9573	<u>0.0051</u>	1.9573	<u>0.9949</u>	1.8882e2	0.9522	1.9522
ASTTV-NTLA	<b>1.0000</b>	0.9496	0.1413	1.9495	0.8587	6.7192	0.8082	1.8082
Proposed	<b>1.0000</b>	<b>1.0000</b>	<b>0.0050</b>	<b>2.0000</b>	<b>0.9950</b>	<b>2.0000e2</b>	<b>0.9950</b>	<b>1.9950</b>

TABLE VI  
3-D ROC-DERIVED AUC RESULTS OF COMPETITIVE METHODS FOR SEQUENCE 3.

Methods	AUC <sub>(D,F)</sub>	AUC <sub>(D,<math>\tau</math>)</sub>	AUC <sub>(F,<math>\tau</math>)</sub>	AUC <sub>TD</sub>	AUC <sub>BS</sub>	AUC <sub>SNPR</sub>	AUC <sub>TDBS</sub>	AUC <sub>ODP</sub>
TLLCM	<b>1.0000</b>	<b>1.0000</b>	0.0052	<b>2.0000</b>	0.9948	1.9404e2	0.9948	1.9948
IPI	<b>1.0000</b>	<b>1.0000</b>	0.1903	<b>2.0000</b>	0.8097	5.2538	0.8097	1.8097
NRAM	<b>1.0000</b>	0.9903	<u>0.0051</u>	1.9902	<u>0.9949</u>	1.9436e2	0.9852	1.9852
NIRPS	<b>1.0000</b>	<b>1.0000</b>	<u>0.0051</u>	<b>2.0000</b>	<b>0.9949</b>	1.9487e2	<b>0.9949</b>	<b>1.9949</b>
RIPT	<b>1.0000</b>	0.9745	0.0053	1.9745	0.9947	1.8376e2	0.9692	1.9692
LogTFNN	<b>1.0000</b>	<b>1.0000</b>	<u>0.0051</u>	<b>2.0000</b>	<u>0.9949</u>	<u>1.9572e2</u>	<b>0.9949</b>	<u>1.9949</u>
WSNM-STIPT	<b>1.0000</b>	0.9634	<u>0.0051</u>	1.9633	<u>0.9949</u>	1.8980e2	0.9583	1.9583
TCNN-NPSTT	<b>1.0000</b>	0.9649	<u>0.0051</u>	1.9648	<u>0.9949</u>	1.9067e2	0.9598	1.9598
ASTTV-NTLA	<b>1.0000</b>	<u>0.9942</u>	0.0589	<u>1.9941</u>	0.9411	1.6891e1	0.9353	1.9353
Proposed	<b>1.0000</b>	<b>1.0000</b>	<b>0.0050</b>	<b>2.0000</b>	<b>0.9950</b>	<b>2.0000e2</b>	<b>0.9950</b>	<b>1.9950</b>

TABLE VII  
3-D ROC-DERIVED AUC RESULTS OF COMPETITIVE METHODS FOR SEQUENCE 4.

Methods	AUC <sub>(D,F)</sub>	AUC <sub>(D,<math>\tau</math>)</sub>	AUC <sub>(F,<math>\tau</math>)</sub>	AUC <sub>TD</sub>	AUC <sub>BS</sub>	AUC <sub>SNPR</sub>	AUC <sub>TDBS</sub>	AUC <sub>ODP</sub>
TLLCM	<b>1.0000</b>	<b>1.0000</b>	0.0052	<b>2.0000</b>	0.9948	<u>1.9118e2</u>	<b>0.9948</b>	<b>1.9948</b>
IPI	<b>1.0000</b>	0.9990	0.2586	<u>1.9990</u>	0.7414	3.8632	0.7404	1.7404
NRAM	<b>1.0000</b>	0.9820	0.0053	1.9820	0.9947	1.8646e2	0.9767	1.9767
NIRPS	<b>1.0000</b>	0.9974	0.0052	1.9974	0.9948	<u>1.9118e2</u>	0.9922	1.9922
RIPT	<b>1.0000</b>	0.9702	0.0055	1.9702	0.9945	1.7649e2	0.9647	1.9647
LogTFNN	<b>1.0000</b>	<b>1.0000</b>	0.0073	<b>2.0000</b>	0.9927	1.3696e2	<b>0.9927</b>	<b>1.9927</b>
WSNM-STIPT	<b>1.0000</b>	0.9247	0.0052	1.9247	0.9948	1.7941e2	0.9196	1.9196
TCNN-NPSTT	<b>1.0000</b>	0.8585	<u>0.0051</u>	1.8585	<u>0.9949</u>	1.6774e2	0.8534	1.8534
ASTTV-NTLA	<b>1.0000</b>	0.9872	0.1185	1.9872	0.8815	8.3321	0.8687	1.8687
Proposed	<b>1.0000</b>	0.9973	<b>0.0050</b>	1.9973	<b>0.9950</b>	<b>1.9899e2</b>	0.9923	1.9923

TABLE VIII  
3-D ROC-DERIVED AUC RESULTS OF COMPETITIVE METHODS FOR SEQUENCE 5.

Methods	AUC <sub>(D,F)</sub>	AUC <sub>(D,τ)</sub>	AUC <sub>(F,τ)</sub>	AUC <sub>TD</sub>	AUC <sub>BS</sub>	AUC <sub>SNPR</sub>	AUC <sub>TDBS</sub>	AUC <sub>ODP</sub>
TLLCM	0.9977	0.6245	0.0056	1.6221	0.9921	1.1182e2	0.6189	1.6166
IPI	<b>1.0000</b>	0.9336	0.2682	1.9335	0.7317	3.4806	0.6653	1.6653
NRAM	0.9998	0.8608	0.0057	1.8606	0.9942	1.5159e2	0.8551	1.8550
NIRPS	0.9997	0.7567	0.0059	1.7563	0.9938	1.2888e2	0.7508	1.7505
RIPt	<u>0.9999</u>	0.6513	0.0056	1.6512	0.9943	1.1675e2	0.6457	1.6456
LogTFNN	<b>1.0000</b>	<b>0.9968</b>	0.0101	<b>1.9967</b>	0.9898	9.8358e1	<b>0.9867</b>	<b>1.9866</b>
WSNM-STPT	<b>1.0000</b>	0.8994	<u>0.0053</u>	1.8994	0.9947	<b>1.7023e2</b>	0.8941	1.8941
TCNN-NPSTT	<u>0.9999</u>	0.3751	<b>0.0051</b>	1.3750	<u>0.9948</u>	7.3312e1	0.3700	1.3698
ASTTV-NTLA	<b>1.0000</b>	0.9839	0.1983	1.9839	0.8017	4.9613	0.7856	1.7856
Proposed	<b>1.0000</b>	<b>0.9982</b>	<b>0.0051</b>	<b>1.9982</b>	<b>0.9949</b>	<b>1.9751e2</b>	<b>0.9932</b>	<b>1.9932</b>

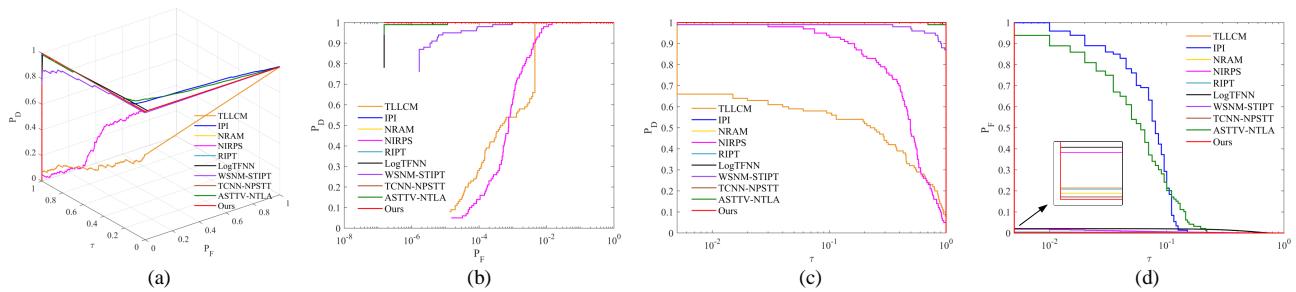


Fig. 10. 3-D ROC curves along with the corresponding 2-D ROC curves of the competitive methods for Sequence 1. And the log scale is used on the x-axis of 2-D ROC curves for the convenience of visual comparison.

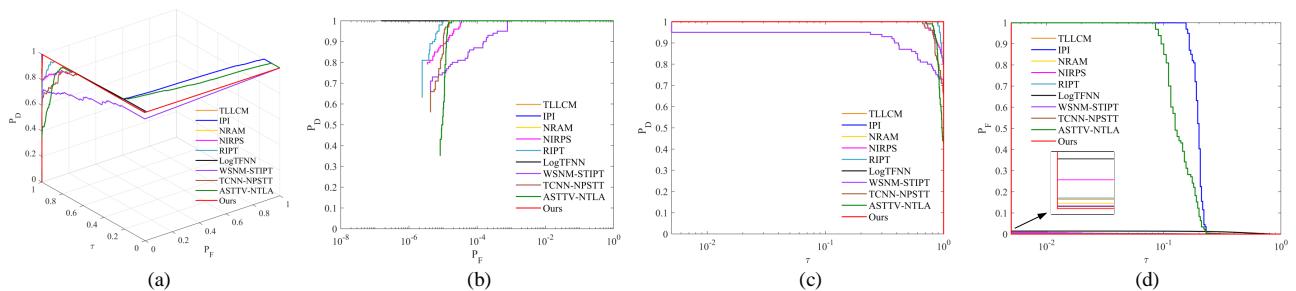


Fig. 11. 3-D ROC curves along with the corresponding 2-D ROC curves of the competitive methods for Sequence 2. And the log scale is used on the x-axis of 2-D ROC curves for the convenience of visual comparison.

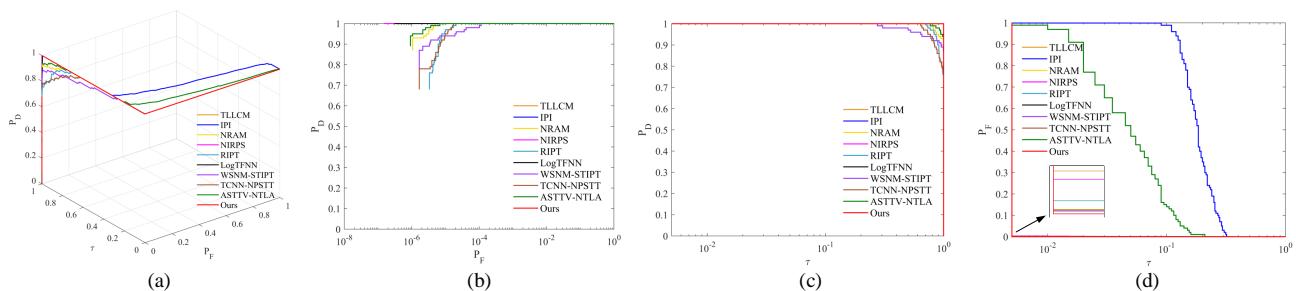


Fig. 12. 3-D ROC curves along with the corresponding 2-D ROC curves of the competitive methods for Sequence 3. And the log scale is used on the x-axis of 2-D ROC curves for the convenience of visual comparison.

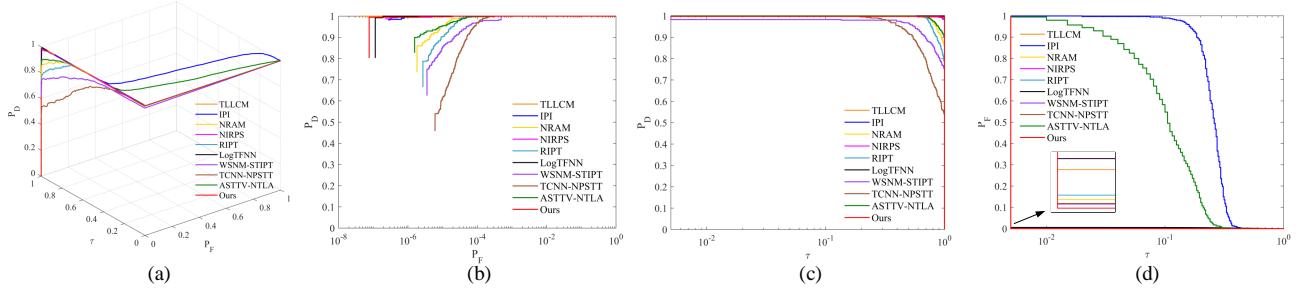


Fig. 13. 3-D ROC curves along with the corresponding 2-D ROC curves of the competitive methods for Sequence 4. And the log scale is used on the x-axis of 2-D ROC curves for the convenience of visual comparison.

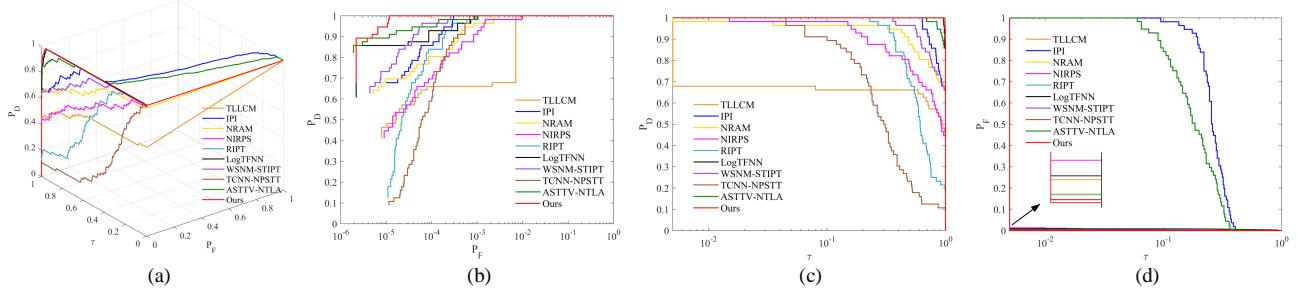


Fig. 14. 3-D ROC curves along with the corresponding 2-D ROC curves of the competitive methods for Sequence 5. And the log scale is used on the x-axis of 2-D ROC curves for the convenience of visual comparison.

LWEC, whose 3-D ROC curve-derived evaluation metrics obtain the best results, shows its excellent target detection performance in scenarios with stubborn building clutters and noise.

The images in Sequence 3 contain thick clouds and heavy noise. In Fig. 12, IMNN-LWEC outperforms all competitive methods in the 2-D ROC curves of  $(P_F, \tau)$ ,  $(P_D, \tau)$ ,  $(P_D, P_F)$  as well as 3-D ROC curve  $(P_D, P_F, \tau)$ . Table VI draws the similar conclusion from statistics. Similar to our method, TLLCM, IPI, NIRPS and LogTFNN are robust enough to detect small targets, but they will fall into a dilemma when faced with bright suspected targets. Besides, sequential detection methods are inferior to our method in both TD and BS. It can be concluded that our method is more suitable for small target detection in ground-sky scenarios with clouds and noise.

Sequence 4 belongs to low altitude motion background with significant roads and running trains, making target detection more challenging. In Fig. 13, IMNN-LWEC presents the best results in 2-D ROC curve of  $(P_F, \tau)$ , while it shows a little unsatisfactory TD performance in 2-D ROC curve of  $(P_D, \tau)$ . The above phenomenon can be further explained in Table VII. Except that  $AUC_{(D,F)}$  of each method is 1.0000, IMNN-LWEC does best on  $AUC_{(F,\tau)}$ ,  $AUC_{BS}$  and  $AUC_{SNPR}$ . In order to greatly suppress the strong radiation noise and target-suspected points carried by Sequence 4 as much as possible, IMNN-LWEC reduces its TD capability to some extent. Although the  $AUC_{(D,\tau)}$  of four methods (e.g., TLLCM and IPI) slightly exceed that of IMNN-LWEC, their BS is really weak according to  $AUC_{(F,\tau)}$ ,  $AUC_{BS}$  and  $AUC_{SNPR}$ . Moreover, IMNN-LWEC exceeds the chosen sequential methods on all 3-D ROC-derived AUC results. Overall, IMNN-LWEC favorably outperforms competitive methods on Sequence 4.

The flying aircraft in Sequence 5 is quite similar to the highlighted noisy background. Combining Fig. 13 and Table VIII, most of the competitive methods, especially IPI and ASTTV-NTLA, are weak in background expression. By comparison, TCNN-NPSTT does not work well in target detection because of its  $AUC_{(D,\tau)}$  (0.3751). All the results of AUC driven from 3-D ROC analysis tool rank first. And these 3-D ROC curves along with the corresponding 2-D ROC curves prove the superiority of IMNN-LWEC. Therefore, our algorithm has certain advantages in complex ground-sky backgrounds.

In Section IV-B, we have analyzed the shortcomings of only using traditional 2-D ROC to evaluate TD and BS of small target detection methods. Through sufficient experiments on five complex IR sequences and comprehensive evaluation using 3-D ROC analysis tool, it is convincing that IMNN-LWEC is robust in complex scenes and outperforms the most existing state-of-the-art methods in the aspect of target detectability and background suppression.

#### E. Robustness to Noisy Scene

Affected by the sensor itself and the surrounding environment, the IR images are likely to be noisy. Hence, we further verify the robustness and target detection performance of the proposed method in scenarios containing different levels of noise. As shown in Fig. 15, the standard deviations of Gaussian white noise in the first and third rows are 0.001 and 0.005, respectively. To test the robustness of IMNN-LWEC in more complex noisy scenarios, we also randomly add strip noise and Gaussian white noise of standard deviations 0.005 to IR images. The corresponding results are shown in the sixth row of Fig. 15. The detection results show that the proposed model

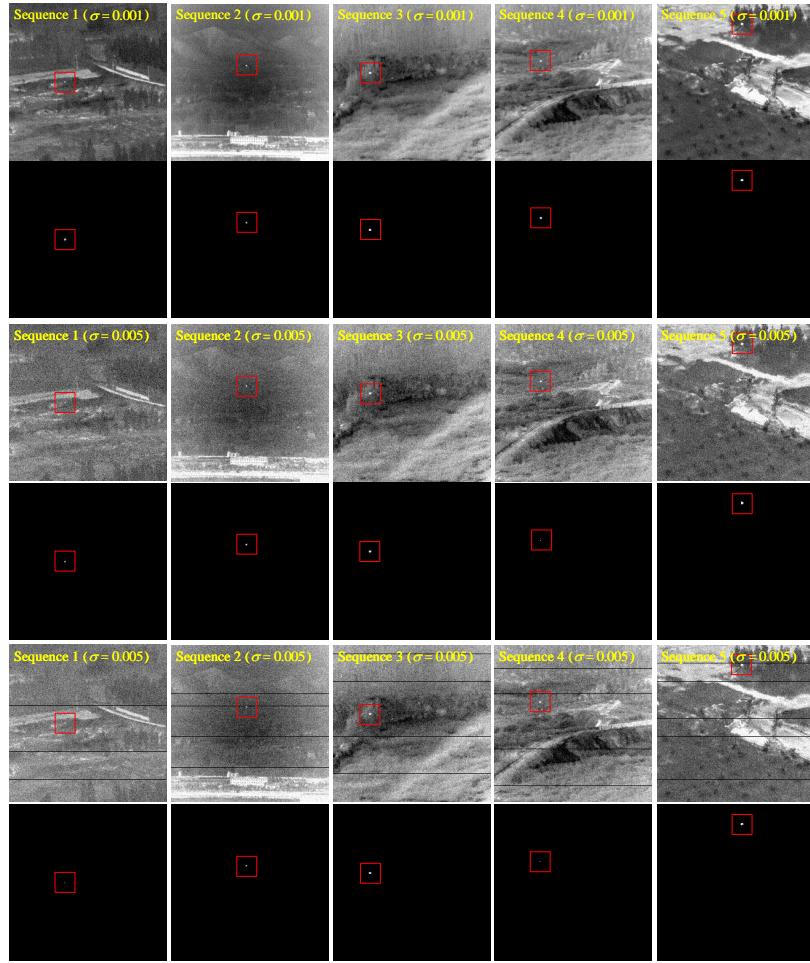


Fig. 15. Detection results in noisy scenarios obtained by IMNN-LWEC. The variance of Gaussian white noise in the first and third rows are 0.001 and 0.005, respectively. The images in fifth row contain random strip noise and Gaussian white noise of standard deviations 0.005. And images in the second, fourth and sixth rows are the corresponding detection results.

can accurately extract the target component without retaining the background residuals, which indicates the IMNN-LWEC has strong robustness in noisy backgrounds.

#### F. Ablation Study

As shown in Table IX, the ablation study is conducted on Sequence 1 to test the effect of each module in IMNN-LWEC model. It should be noted that the bold and underlined data represent the optimal and suboptimal performance for the corresponding AUC results, respectively. Baseline denotes only using one frame to construct tensor as shown in Fig. 1 without saliency sparse prior weight tensor  $\mathcal{W}$ , temporal information and the proposed novel tube-wise regularization. Besides, we replace IMWTNN with TNN in the baseline. Symbol “✓” means adding the specific module to the baseline. For the 2<sup>nd</sup> to 7<sup>th</sup> rows, IMWTNN is used to approximate the background rank. The 3<sup>rd</sup> to 7<sup>th</sup> rows in Table IX represent that three consecutive frames are used to construct spatial-temporal tensor for small target detection. The first three rows do not consider prior information  $\mathcal{W}$ . And the first six rows do not consider tube-wise regularization.

Comparing the results in the first row of Table IX with Table IV, it can be clearly seen that TD and BS of the baseline still exceed some competitive methods, such as TLLCM, which illustrates that the proposed solution framework is effective. The results in the first two rows indicate that the designed IWMTNN can better estimate the rank of the background tensor and keep the background component as much as possible to reduce false alarms in the final target images. The results in the 2<sup>nd</sup> and 3<sup>rd</sup> rows demonstrate that the temporal information can improve BS to some extent. All of the eight 3-D ROC curve-derived evaluation metrics in the 6<sup>th</sup> row significantly exceed that of in the 3<sup>rd</sup> row, reflecting that the prior  $\mathcal{W}$  as a sparse weight tensor can strengthen TD and BS. Specifically, the image entropy operator and the new criterion (i.e., prior information filter) are integrated into the prior information. To testify the role of the image entropy operator, we use the original images to calculate the improved local prior maps  $W_l$  via (11)-(14), ignoring (10). The results in 4<sup>th</sup> and 6<sup>th</sup> rows of Table IX confirm that the image entropy operator can improve TD and BS, because it enhances the targets with high information entropy and suppresses the flat area with low information entropy. As for the results in the 3<sup>rd</sup> and 5<sup>th</sup>

TABLE IX  
ABLATION STUDY OF DIFFERENT MODULES

Baseline	IMWTNN	Temporal information	$\mathcal{W}$	$\ \cdot\ _{tw}$	AUC <sub>(D,F)</sub>	AUC <sub>(D,<math>\tau</math>)</sub>	AUC <sub>(F,<math>\tau</math>)</sub>	AUC <sub>TD</sub>	AUC <sub>BS</sub>	AUC <sub>SNPR</sub>	AUC <sub>TDBS</sub>	AUC <sub>COPD</sub>
✓					0.9998	0.9046	0.0062	1.9043	0.9936	1.4640e2	0.8984	1.8981
✓	✓				<u>0.9999</u>	0.9519	0.0063	1.9518	0.9936	1.5206e2	0.9456	1.9455
✓	✓	✓			<b>1.0000</b>	0.9403	<u>0.0055</u>	1.9402	<u>0.9945</u>	1.7122e2	0.9348	1.9347
✓	✓	✓	✓	✓	<b>1.0000</b>	0.9889	0.0060	1.9889	0.9940	1.6399e2	0.9829	1.9829
✓	✓	✓	✓	✓	<b>1.0000</b>	0.9204	<b>0.0050</b>	1.9204	<b>0.9950</b>	1.8408e2	0.9154	1.9154
✓	✓	✓	✓	✓	<b>1.0000</b>	<u>0.9992</u>	<b>0.0050</b>	<u>1.9991</u>	<b>0.9950</b>	<u>1.9885e2</u>	<u>0.9941</u>	<u>1.9941</u>
✓	✓	✓	✓	✓	<b>1.0000</b>	<b>0.0050</b>	<b>2.0000</b>	<b>0.9950</b>	<b>1.9913e2</b>	<b>0.9950</b>	<b>1.9950</b>	

TABLE X  
RUNNING TIME OF COMPARED METHODS IN FIVE SCENES

Methods	Sequence 1	Sequence 2	Sequence 3	Sequence 4	Sequence 5
TLLCM	0.2436	0.3464	0.3564	0.3503	0.3445
IPI	2.8824	3.4652	3.1747	2.9014	3.2207
NRAM	0.0829	0.0737	0.0625	0.0719	0.0795
NIRPS	0.2560	0.2512	0.2538	0.2548	0.2666
RIPT	0.6309	0.4891	0.4281	0.4837	0.5318
LogTFNN	0.9053	0.8421	0.8558	0.8942	0.8741
WSNM-STIPT	0.0335	0.0348	0.0773	0.0435	0.0436
TCNN-NPSTT	6.7173	7.4548	6.5167	7.1519	7.0838
ASTTV-NTLA	1.3921	1.3213	1.3319	1.3381	1.4003
Proposed	1.8168	1.8896	1.9639	1.8585	1.9595

rows, only using local information entropy image to construct  $\mathcal{W}$  is slightly beneficial for BS. Comparing the 5<sup>th</sup> row with the 6<sup>th</sup> row, adding the new criteria based on multiscale local contrast measure and structure tensor theory to  $W_l'$  in (14) can further improve the target recall. The 4<sup>th</sup> to the 6<sup>th</sup> rows jointly illustrate that combining the image information entropy operator with the new criteria can greatly improve the performance of IMNN-LWEC. What's more, the results of the 6<sup>th</sup> and 7<sup>th</sup> rows illustrate that the proposed tube-wise regularization can further improve the detection performance.

In summary, the ablation experiments in Table IX prove that each module takes effect in multiple aspects and improves the comprehensive performance of IMNN-LWEC.

#### G. Running Time

In addition to the above 3-D ROC curve-derived evaluation metrics, running time is also a crucial factor. In Table X, the average running time per frame of the ten detection algorithms on the five sequences is summarized. Among the multi-frame detection algorithms, TCNN-NPSTT takes the longest time. Generally, the single-frame detection methods are faster than the multi-frame detection methods, but inter-frame temporal information is ignored. Among the multi-frame detection methods, TCNN-NPSTT costs the most time. Combined with the 3D-ROC analysis results in Table IV-VIII, our method works better than the competitive methods. Given that our algorithm requires to traverse each pixel to obtain saliency prior maps and approximate the rank of multi-mode tensors, the extra runtime of our method is acceptable to get excellent performance. In the future, we consider optimizing algorithm or using GPU to improve the real-time performance of our method.

#### H. Convergence Analysis

The convergence of the optimization problem (30) has been proved in [68]. Taking Sequence 1 as an example, it costs approximately 1.8168s per frame to realize IR small target detection. Furthermore, the convergence curve is shown in Fig. 16. Under certain constrains, the value of the objective function  $\|\mathcal{D} - \mathcal{B} - \mathcal{T} - \mathcal{E} - \mathcal{N}\|_F / \|\mathcal{D}\|_F$  (i.e., relative error), tends to be stable after several iterations. Consequently, IMNN-LWEC can easily achieve the convergence and successfully obtain an optimal or suboptimal solution.

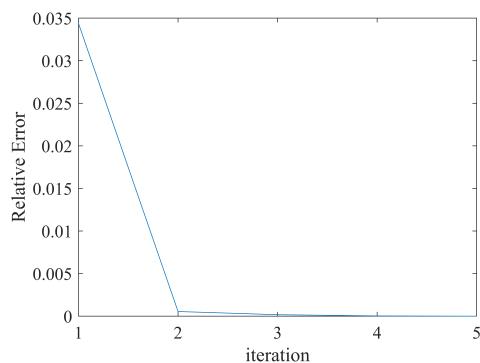


Fig. 16. The convergence curve of the proposed algorithm in Sequence 1.

#### V. CONCLUSION

In this paper, we proposed a novel IR small target detection method named IMNN-LWEC. Considering the spatial-temporal correlation information, we construct a new nonoverlapping spatial-temporal tensor model and introduce an improved multi-mode weighted tensor nuclear norm (IMWTNN)

to describe the low-rank property of the background. Besides, a local sparse saliency prior weight named local weighted entropy contrast (LWEC), which is based on a local weighted image entropy operator and a new prior information filter simultaneously, is proposed to further suppress the impact of complicated background. And a novel tube-wise sparse regularization is designed to reduce rare structure effect. Thus, the IR small target detection task is transformed into a tensor-based low-rank and sparse decomposition problem. An efficient optimization algorithm based on ADMM is promoted to solve this problem. Extensive experiments witness the superior ability of IMNN-LWEC in aspects of target detectability and background suppressibility compared with other nine state-of-the-art competitive methods. And IMNN-LWEC is robust to noisy scenarios. Besides, the specific role of each designed module is demonstrated by ablation experiments. However, our method needs to calculate multiscale local contrast maps and weighted entropy maps, which are time-consuming processes.

#### ACKNOWLEDGMENTS

The authors would like to thank the editor and the anonymous reviewers for their constructive comments to improve the quality of this paper.

#### REFERENCES

- [1] S. Bourennane, C. Fossati, and A. Cailly, "Improvement of target-detection algorithms based on adaptive three-dimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1383–1395, 2010.
- [2] L. Peng, T. Zhang, S. Huang, T. Pu, Y. Liu, Y. Lv, Y. Zheng, and Z. Peng, "Infrared small-target detection based on multi-directional multi-scale high-boost response," *Opt. Rev.*, vol. 26, no. 6, pp. 568–582, 2019.
- [3] X. Bai, Z. Chen, Y. Zhang, Z. Liu, and Y. Lu, "Infrared ship target segmentation based on spatial information improved fcm," *IEEE T. Cybern.*, vol. 46, no. 12, pp. 3259–3271, 2015.
- [4] C. Xia, S. Chen, and Y. Luo, "Omnidirectional mirror gradient dissimilarity for infrared small target detection," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.* IEEE, 2022, pp. 3255–3258.
- [5] C. Xia, S. Chen, X. Zhang, Z. Chen, and Z. Pan, "Infrared small target detection via dynamic image structure evolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2022.
- [6] X. Wang, Z. Peng, D. Kong, and Y. He, "Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scene," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5481–5493, 2017.
- [7] V. T. Tom, T. Peli, M. Leung, and J. E. Bondaryk, "Morphology-based algorithm for point target detection in infrared backgrounds," in *Proc. SPIE*, vol. 1954. SPIE, 1993, pp. 2–11.
- [8] L. Deng, G. Xu, J. Zhang, and H. Zhu, "Entropy-driven morphological top-hat transformation for infrared small target detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 2, pp. 962–975, 2021.
- [9] S. D. Deshpande, M. H. Er, R. Venkateswarlu, and P. Chan, "Max-mean and max-median filters for detection of small targets," in *Proc. SPIE*, vol. 3809. SPIE, 1999, pp. 74–83.
- [10] M. M. Hadhoud and D. W. Thomas, "The two-dimensional adaptive lms (tdlms) algorithm," *IEEE Trans. Circuits Syst.*, vol. 35, no. 5, pp. 485–494, 1988.
- [11] G. Davidson and H. Griffiths, "Wavelet detection scheme for small targets in sea clutter," *Electron. Lett.*, vol. 38, no. 19, p. 1, 2002.
- [12] S. Kim, Y. Yang, J. Lee, and Y. Park, "Small target detection utilizing robust methods of the human visual system for irst," *J. Infrared Millim. Terahertz Waves*, vol. 30, no. 9, pp. 994–1011, 2009.
- [13] C. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 574–581, 2013.
- [14] J. Han, S. Moradi, I. Faramarzi, C. Liu, H. Zhang, and Q. Zhao, "A local contrast method for infrared small-target detection utilizing a tri-layer window," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1822–1826, 2019.
- [15] H. Cui, L. Li, X. Liu, X. Su, and F. Chen, "Infrared small target detection based on weighted three-layer window local contrast," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.
- [16] Z. Li, S. Liao, and T. Zhao, "Infrared dim and small target detection based on strengthened robust local contrast measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.
- [17] J. Liu, H. Wang, L. Lei, and J. He, "Infrared small target detection utilizing halo structure prior-based local contrast measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [18] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4996–5009, 2013.
- [19] W. Li, M. Zhao, X. Deng, L. Li, L. Li, and W. Zhang, "Infrared small target detection using local and nonlocal spatial information," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3677–3689, 2019.
- [20] L. Zhang, L. Peng, T. Zhang, S. Cao, and Z. Peng, "Infrared small target detection via non-convex rank approximation minimization joint 1 2, 1 norm," *Remote Sens.*, vol. 10, no. 11, p. 1821, 2018.
- [21] Y. Dai and Y. Wu, "Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3752–3767, 2017.
- [22] C. Zhang, Y. He, Q. Tang, Z. Chen, and T. Mu, "Infrared small target detection via interpatch correlation enhancement and joint local visual saliency prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2021.
- [23] L. Yang, P. Yan, M. Li, J. Zhang, and Z. Xu, "Infrared small target detection based on a group image-patch tensor model," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [24] X. Kong, C. Yang, S. Cao, C. Li, and Z. Peng, "Infrared small target detection via nonconvex tensor fibered rank approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–21, 2021.
- [25] Z. Cao, X. Kong, Q. Zhu, S. Cao, and Z. Peng, "Infrared dim target detection via mode-k1k2 extension tensor tubal rank under complex ocean environment," *ISPRS-J. Photogramm. Remote Sens.*, vol. 181, pp. 167–190, 2021.
- [26] I. S. Reed, R. M. Gagliardi, and L. B. Stotts, "Optical moving target detection with 3-d matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 327–336, 1988.
- [27] A. P. Tzannes and D. H. Brooks, "Detecting small moving objects using temporal hypothesis testing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 2, pp. 570–586, 2002.
- [28] B. Balasingam, Y. Bar-Shalom, P. Willett, and K. Pattipati, "Maximum likelihood detection on images," in *Proc. Inf. Fusion.* IEEE, 2017, pp. 1–8.
- [29] Y. Sun, J. Yang, Y. Long, and W. An, "Infrared small target detection via spatial-temporal total variation regularization and weighted tensor nuclear norm," *IEEE Access*, vol. 7, pp. 56667–56682, 2019.
- [30] P. Zhang, L. Zhang, X. Wang, F. Shen, T. Pu, and C. Fei, "Edge and corner awareness-based spatial-temporal tensor model for infrared small-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10708–10724, 2020.
- [31] H.-K. Liu, L. Zhang, and H. Huang, "Small target detection in infrared videos based on spatio-temporal tensor model," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8689–8700, 2020.
- [32] G. Wang, B. Tao, X. Kong, and Z. Peng, "Infrared small target detection using nonoverlapping patch spatial-temporal tensor factorization with capped nuclear norm regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2021.
- [33] T. Liu, J. Yang, B. Li, C. Xiao, Y. Sun, Y. Wang, and W. An, "Nonconvex tensor low-rank approximation for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2021.
- [34] D. Pang, T. Shan, W. Li, P. Ma, R. Tao, and Y. Ma, "Facet derivative-based multidirectional edge awareness and spatial-temporal tensor model for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.
- [35] Y. Sun, J. Yang, and W. An, "Infrared dim and small target detection via multiple subspace learning and spatial-temporal patch-tensor model," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 3737–3752, 2020.
- [36] D. Pang, P. Ma, T. Shan, W. Li, R. Tao, Y. Ma, and T. Wang, "Sstm-sfr: Spatial-temporal tensor modeling with saliency filter regularization for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2022.
- [37] X. Liu, L. Li, L. Liu, X. Su, and F. Chen, "Moving dim and small target detection in multiframe infrared sequence with low scr based on

- temporal profile similarity," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [38] J. Du, H. Lu, L. Zhang, M. Hu, S. Chen, Y. Deng, X. Shen, and Y. Zhang, "A spatial-temporal feature-based detection framework for infrared dim small target," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2021.
- [39] Y. Bai, R. Li, S. Gou, C. Zhang, Y. Chen, and Z. Zheng, "Cross-connected bidirectional pyramid network for infrared small-dim target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [40] H. Wang, L. Zhou, and L. Wang, "Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 8509–8518.
- [41] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Attentional local contrast networks for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9813–9824, 2021.
- [42] B. Li, C. Xiao, L. Wang, Y. Wang, Z. Lin, M. Li, W. An, and Y. Guo, "Dense nested attention network for infrared small target detection," *IEEE Trans. Image Process.*, 2022.
- [43] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, 2017.
- [44] H. Wang, F. Yang, C. Zhang, and M. Ren, "Infrared small target detection based on patch image model with local and global analysis," *Int. J. Image Graph.*, vol. 18, no. 01, p. 1850002, 2018.
- [45] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-svd," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3842–3849.
- [46] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-Y. Ji, and T.-H. Ma, "Tensor n-tubal rank and its convex relaxation for low-rank tensor recovery," *Inf. Sci.*, vol. 532, pp. 170–189, 2020.
- [47] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Alg. Appl.*, vol. 435, no. 3, pp. 641–658, 2011.
- [48] Y. Gu, C. Wang, B. Liu, and Y. Zhang, "A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 469–473, 2010.
- [49] X. Peng, C. Lu, Z. Yi, and H. Tang, "Connections between nuclear-norm and frobenius-norm-based representations," *IEEE Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 218–224, 2016.
- [50] H. Johan, "Tensor rank is np-complete," *J. Algorithms*, vol. 11, no. 4, pp. 644–654, 1990.
- [51] D. Goldfarb and Z. Qin, "Robust low-rank tensor recovery: Models and algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 1, pp. 225–253, 2014.
- [52] W.-H. Xu, X.-L. Zhao, T.-Y. Ji, J.-Q. Miao, T.-H. Ma, S. Wang, and T.-Z. Huang, "Laplace function based nonconvex surrogate for low-rank tensor completion," *Signal Process.-Image Commun.*, vol. 73, pp. 62–69, 2019.
- [53] X. Qu, H. Chen, and G. Peng, "Novel detection method for infrared small targets using weighted information entropy," *J. Syst. Eng. Electron.*, vol. 23, no. 6, pp. 838–842, 2012.
- [54] L. Zhang and Z. Peng, "Infrared small target detection based on partial sum of the tensor nuclear norm," *Remote Sens.*, vol. 11, no. 4, p. 382, 2019.
- [55] M. Zhao, W. Li, L. Li, P. Ma, Z. Cai, and R. Tao, "Three-order tensor creation and tucker decomposition for infrared small-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2021.
- [56] X. Guan, L. Zhang, S. Huang, and Z. Peng, "Infrared small target detection via non-convex tensor rank surrogate joint local contrast energy," *Remote Sens.*, vol. 12, no. 9, p. 1520, 2020.
- [57] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l-1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.
- [58] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
- [59] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optimiz.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [60] Y. Liu and Z. Peng, "Infrared small target detection based on resampling-guided image model," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.
- [61] Y. Sun, J. Yang, M. Li, and W. An, "Infrared small target detection via spatial-temporal infrared patch-tensor model and weighted schatten p-norm minimization," *Infrared Phys. Technol.*, vol. 102, p. 103050, 2019.
- [62] C. Xia, X. Li, Y. Yin, and S. Chen, "Multiple infrared small targets detection based on hierarchical maximal entropy random walk," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.
- [63] C. Xia, X. Li, L. Zhao, and R. Shu, "Infrared small target detection based on multiscale local contrast measure using local energy factor," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 157–161, 2019.
- [64] C. Xia, X. Li, and L. Zhao, "Infrared small target detection via modified random walks," *Remote Sens.*, vol. 10, no. 12, p. 2004, 2018.
- [65] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3d receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5131–5153, 2020.
- [66] S. Chen, C.-I. Chang, and X. Li, "Component decomposition analysis for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–22, 2021.
- [67] C.-I. Chang, "Hyperspectral anomaly detection: A dual theory of hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–20, 2021.
- [68] Q. Gao, P. Zhang, W. Xia, D. Xie, X. Gao, and D. Tao, "Enhanced tensor rpca and its application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 2133–2140, 2020.

**Yuan Luo** received the B.S. degree from Jianxing Honors College, Zhejiang University of Technology, Hangzhou, China, in 2021. He is currently pursuing the Ph.D. degree in control theory and control engineering from Zhejiang University, Hangzhou, China. His research interests include image processing and pattern recognition.



**Xiaorun Li** received the B.S. degree from the National University of Defense Technology, Changsha, China, in 1992, and the M.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 1995 and 2008, respectively. Since 1995, he has been with Zhejiang University, where he is currently a Professor with the College of Electrical Engineering. His research interests include hyperspectral image processing, signal and image processing, and pattern recognition.



**Shuhan Chen** received the B.S. degree from the Ludong University, Yantai, China, in 2011, and the M.S. degree from Liaoning Technical University, Huludao, China, in 2014 and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2020. She was also doing research as a visiting faculty research assistant in the Remote Sensing Signal and Image Processing laboratory (RSSIPL), Department of Computer Science and Electrical Engineering at University Maryland, Baltimore County (UMBC), Baltimore, USA, in 2018–2020. She is currently a Post-doc in the department of Control Science and Engineering from Zhejiang University, Hangzhou, China, and her research interests included image registration, hyperspectral image processing and pattern recognition.

Post-doc in the department of Control Science and Engineering from Zhejiang University, Hangzhou, China, and her research interests included image registration, hyperspectral image processing and pattern recognition.





**Chaoqun Xia** received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2016, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2021. He is currently a lecturer with the College of Computer Science and Artificial Intelligence, Wenzhou University. His research interests include image processing, machine learning, and pattern recognition.



**Liaoying Zhao** received the B.S. and M.S. degrees from Hangzhou Dianzi University, Hangzhou, China, in 1992 and 1995, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, in 2004. Since 1995, she has been with Hangzhou Dianzi University, where she is currently a Professor with the College of Computer Science. Her research interests include hyperspectral image processing, signal and image processing, pattern recognition, and machine learning.