Topological Prior-Embedded Spatial-Temporal Tensor Approximation Model for Infrared Small-Target Detection

Mingjing Zhao, Qi Ming, Yingjie Ma, Lu Li,
Wei Li, *Senior Member, IEEE*, and Ran Tao, *Senior Member, IEEE*

***Abstract*— Infrared small-target detection is crucial in both military and civilian applications, yet it remains challenging to distinguish true targets from complex backgrounds and target-like flickering noise. To address this issue, a novel topological prior-embedded spatial-temporal tensor approximation (TP-STTA) model is proposed. The framework first design multi-topological feature extraction (MTFE) model, in which area, diagonal, and volume attributes are explored, which effectively characterize the spatial properties of small targets and suppress background clutter and flickering noise. These features not only provide an auxiliary detection response but, more importantly, serve as target-aware priors to guide the subsequent tensor-based optimization. Building upon this, a multi-prior tensor approximation (MPTA) model jointly exploits spatial-temporal correlations by integrating the multi-attribute topological priors from MTFE with background-oriented smoothness and low-rank constraints, leading to stable and robust target enhancement. Finally, the responses from MTFE and MPTA branches are fused, and temporal trajectory consistency is exploited to further suppress false alarms. Experimental results on six real infrared sequences demonstrate that the proposed TP-STTA achieves accurate and reliable small-target detection with significant background suppression.**

***Index Terms*— Infrared Image, Small-Target Detection, Multi-Attribute Topological Features Extraction, Multi-prior Tensor Approximation**

This work was supported by the National Natural Science Foundation of China (62501025, 62471049), the Fundamental Research Funds for the Central Universities (3282025011), and the National Natural Science Foundation of China (62476013) (*Corresponding author: Lu Li)*.

M. Zhao and Y. Ma are with the Beijing Electronic Science and Technology Institute, Beijing 100070 China (e-mail: zmj@besti.edu.cn; dmzm12@163.com).

Q. Ming is with the School of College of Computer Science, Beijing University of Technology, Beijing 100124 China (e-mail: chaser.ming@gmail.com).

L. Li is with the School of Automation, Beijing Information science and Technology University, Beijing 100101 China (e-mail: 20192380@bistu.edu.cn).

W. Li and R. Tao are with the School of Information and Electronics, Beijing Institute Technology, Beijing 100081 China (e-mail: liwei089@ieee.org; rantao@bit.edu.cn).

I. Introduction

Infrared detection systems provide reliable imaging under both daytime and nighttime conditions and remain effective in the presence of cloud, smoke, or camouflage, making them a strong complement or even an alternative to traditional visible-light and radar systems [1, 2]. Owing to these advantages, infrared sensors have been widely adopted in various applications [3–5]. In many practical scenarios, especially in long-range observation or wide-area surveillance, the targets of interest occupy only a few pixels. Such small targets also lack distinctive structural or textural information. These characteristics make infrared small-target detection particularly challenging [6].

A wide range of methods have been proposed for infrared small-target detection, and these methods can generally be divided into two categories: single-frame infrared small target detection methods and multi-frame infrared small target detection methods [7, 8]. Single-frame methods rely solely on the spatial information within each individual image, whereas multi-frame methods exploit both spatial cues and temporal correlations across consecutive frames, thereby enhancing the reliability of target detection [9].

Single-frame infrared small-target detection methods mainly include filter-based methods, local information-based methods, data structure-based methods, and deep learning-based methods. Traditional filter-based methods separate targets by exploiting the observation that discrete objects disturb the spatial smoothness and structural continuity of infrared backgrounds, such as Top-hat [10], Max-mean/Max-median [11]. While these methods offer high computational efficiency, their applicability is generally restricted to scenarios characterized by relatively homogeneous background distributions.

Local information-based methods utilize the local differences between targets and their neighborhood. Motivated by the properties of the human visual system, Chen et al. [12] proposed an local contrast method (LCM), which assumed that the target was brighter than

its neighbors. Subsequently, several refined variants of LCM were proposed, including improved LCM (ILCM) [13], novel LCM (NLCM) [14], relative LCM (RLCM) [15], multiscale LCM (MLCM) [16], global contrast measure (GCM) [17], and so on. Although these methods can effectively enhance the target, strong edges and corners in complex scenes are also enhanced, resulting in residual clutter in the detection results.

Data structure-based methods separate targets from background by exploiting their distinct structural characteristics. Specifically, these methods utilize the low-rank of the background and the sparsity of the target. Gao et al. [18] innovatively extended the traditional infrared image model to an infrared image patch (IPI) model, which characterized the background and target through low-rank and sparse subspaces, respectively. Later, several enhanced variants were introduced, such as weighted IPI (WIPI) [19] and reweighted IPI (ReWIPI) [20]. In addition, tensor-based methods [21–23] were also proposed, which exploited low-rank tensor decomposition to model the intrinsic structural correlations within spatial patches while separating the sparse target component. Huang et al. [24] proposed a joint partial tensor correlated total variation and k-means clustering (PTCTV-KMC) method, which integrated local and global features. However, the performance deteriorates when encountering noise with the same sparse characteristics as the target.

Deep learning-based methods have attracted increasing attention. Some methods based on the convolutional neural networks (CNN) have been proposed [25, 26], and further improvements have been achieved by integrating CNNs with transformers to exploit both local and global features [27, 28]. Diffusion models [29–32] also show strong potential in low-level vision tasks, particularly for image enhancement. However, existing deep learning-based methods are still constrained by the scarcity of large-scale, high-quality annotated datasets. The limited diversity and low target density in current datasets hinder model generalization and lead to unstable performance in complex real-world scenarios.

Single-frame infrared small target detection methods rely on extracting spatial features from a single image, without depending on the temporal continuity between frames. This makes this type of methods struggle to ensure robustness in challenging environments, which highlights the necessity of multi-frame infrared small target detection methods.

Multi-frame infrared small-target detection methods primarily encompass filter-based methods, difference-based methods, spatial-temporal local information-based methods, spatial-temporal tensor-based methods, and deep learning-based methods. Conventional filter-based methods exploit their motion consistency across consecutive frames while suppressing background clutter through temporal or spatial-temporal filter, such as particle filter-based methods [33] and 3-Dimensional (3D) matched filter [34, 35]. These methods often rely on explicit motion models and predefined target templates, making them sensitive to model mismatch and ineffective when the target exhibits irregular motion or strong background fluctuations.

Difference-based methods identify targets by comparing pixel intensities across adjacent frames and detecting significant temporal changes in gray values. Kim et al. [36] eliminated the background by performing a difference operation on adjacent image frames. Gao et al. [37] combined the spatial domain contrast information in a single-frame image with the difference concept of adjacent multi-frame images. Although the principle of performing differential operations between consecutive frames is simple and easy to implement, in practical scenarios, relative motion between the infrared sensor and the background can lead to background components being falsely detected as targets.

Spatial-temporal local information-based methods, inspired by single-frame local information-based methods, detect targets by combining local spatial contrast with temporal changes across consecutive frames. Deng et al. [38] proposed a spatial-temporal local contrast filter (STLCF), which defined new spatial local contrast (SLC) and temporal local contrast (TLC). However, the local contrast calculation is easily affected by noise. Li et al. [39] proposed a spatial-temporal saliency model that generated spatial saliency maps, temporal saliency maps, and trajectory prediction maps. However, due to the presence of noise, targets may still be lost. Du and Hamdulla [40] designed a spatial-temporal local difference measure (STLDM) method based on the variation characteristics. Li et al. [41] introduced a coupled spatialCtemporal tensor model, combining global low-rank properties with local smoothness and the local minimum absolute gradient (GLLS-LMAG) regularization term. These methods still struggle with noise sensitivity and unstable target continuity, often leading to missed detections or target loss across frames.

Spatial-temporal tensor-based methods model infrared sequences as multi-dimensional tensors and separate sparse targets from low-rank backgrounds by exploiting spatial and temporal correlations. Zhang et al. [42] designed an edge and corner awareness-based spatial-temporal tensor (ECA-STT) model, which distinguished targets from background residuals by adaptively measuring the importance of edges and corners. Liu et al. [43]

employed asymmetric spatial-temporal total variation (ASTTV) to improve background estimation accuracy. Luo et al. [44] proposed STRL-LBCM, a low-rank and sparse spatial-temporal tensor model using local binary contrast, reformulating target detection as the decomposition of background, target, and noise. Wu et al. [45] constructed a 4-Dimensional tensor and decomposed it into lower-dimensional tensors using tensor train, referred to as 4DTT. Zhou et al. [46] proposed a sparse differential directionality (SDD) prior, using targets' directional features and sparse constraints on Tucker-derived directional and temporal components. Then, Liu et al. [47] combined fast graph Laplacian regularization with the minimax concave penalty function (FGLR-MCP) for target detection. These methods still face limitations in handling complex, nonstationary backgrounds and extremely weak targets, making them prone to confusion with structured noise and causing reduced detection robustness.

Deep learning-based methods learn hierarchical spatial-temporal features from sequences to capture both local and global context for target detection. Tong et al. [48] introduced a spatial-temporal Transformer architecture that models long-range dependencies across frames to enhance the detection of dim and small infrared targets. Yan et al. [49] enhanced motion cues and suppressed background clutter by combining multi-scale feature extraction with differential attention across frames. Cui et al. [50] introduced a multi-scale temporal-context learning framework, which jointly models spatial details and temporal relations across frames. However, these methods still have weak generalization, causing performance to drop sharply when encountering unseen scenes or target patterns.

Compared with single-frame methods, multi-frame methods generally achieve better performance by leveraging spatial-temporal information across frames. However, existing multi-frame methods still face two key challenges. First, these methods often fail to effectively capture spatial-temporal features, as well as adequately exploit prior knowledge of both target and background, thereby limiting their detection accuracy in complex background scenarios. Second, these existing methods are sensitive to noise, frequently misidentifying target-like flickering noise or other background structures as targets, resulting in high false-alarm rates.

To address these issues, a topological prior-embedded spatial-temporal tensor approximation (TP-STTA) model for infrared small-target detection is proposed. The overall flowchart of TP-STTA is illustrated in Fig. 1. To more comprehensively extract features, two complementary strategies are applied to the input consecutive frames. On the one hand, each frame of the input infrared sequence is processed to extract multi-topological features, including area, diagonal, and volume attributes. The area attribute captures the spatial size of the target. The diagonal attribute encodes combined information of size and shape, while the volume attribute represents a joint measure of size and intensity difference. By jointly exploiting multiple topological attributes, the true target is effectively enhanced, while target-like flickering noise and background interference are simultaneously suppressed. On the other hand, the consecutive infrared images are stacked in chronological order to form a spatial-temporal tensor. Similar to low-rank and sparse matrix decomposition, but extended from matrices to tensors, the spatial-temporal tensor is decomposed into a background tensor and a target tensor under multiple physically meaningful priors. For the background tensor, a low-rank prior along the temporal dimension and a piecewise-smooth prior along the spatial dimensions are jointly imposed. In addition, the extracted multi-topological features are incorporated as target-aware priors for the target tensor. All priors are integrated into a unified convex framework, and alternating direction method of multipliers (ADMM) is employed to iteratively optimize the two tensors until convergence. Finally, the results of two complementary strategies results are fused, and trajectory information is further leveraged to finalize target detection.

The contributions can be summarized as follows. (1) A novel TP-STTA model is designed, in which two complementary strategies are employed, including multi-topological feature extraction (MTFE) for capturing spatial-topological features, and multi-prior tensor approximation (MPTA) for exploiting spatial-temporal features. (2) MTFE extracts area, diagonal, and volume attributes to comprehensively characterize spatial size, shape, and intensity differences, effectively separating small targets from complex backgrounds and flickering noise. The MTFE output serves a dual role: it provides an independent detection result and generates target-aware priors to guide the MPTA model. (3) MPTA jointly exploits spatial and temporal dimensions, integrating multiple physically meaningful and complementary priors, including MTFE-derived target-aware priors, together with background-oriented smoothness and low-rank priors, within a unified tensor optimization framework, thereby enhancing detection stability and accuracy. The outputs of MTFE and MPTA are further fused, and temporal trajectory information is exploited to suppress false alarms and enhance temporal consistency. Comprehensive experiments on six real infrared sequences demonstrate the effectiveness of the proposed TP-STTA.

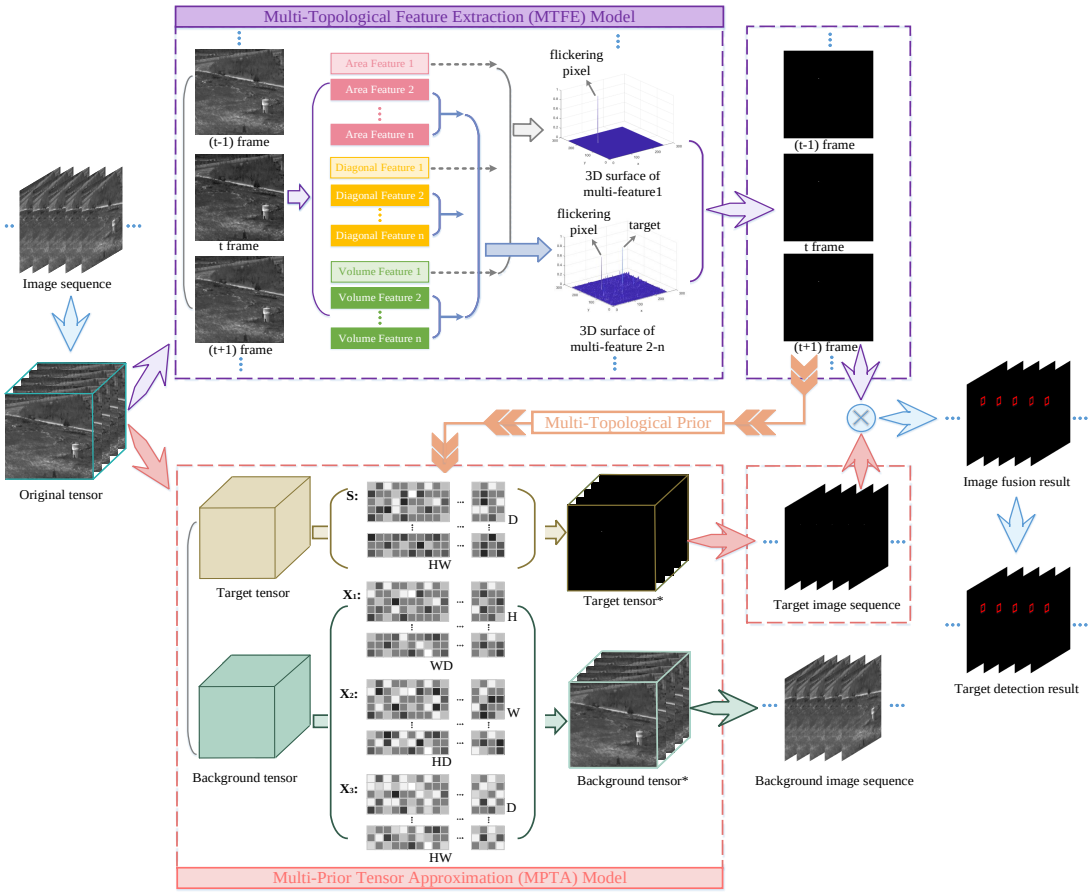


Fig. 1. Illustration of the topological prior-embedded spatial-temporal tensor approximation model for infrared small-target detection.

The remainder of this paper is organized as follows. Preliminary and notations are introduced in Section II. In Section III, a detailed description of the proposed TP-STTA is given. Section IV provides experimental results. Conclusions are summarized in Section V.

II. PRELIMINARY AND NOTATIONS

Uppercase bold letters denote matrices (e.g., $\mathbf{X}$); lowercase bold letters denote vectors (e.g., $\mathbf{x}$); lowercase letters denote scalars (e.g., x); and calligraphic symbols are used for tensors (e.g., $\mathcal{X}$). The consecutive infrared images are stacked in chronological order to form a spatial-temporal tensor with H rows, W columns, and temporal frames T, denoted by $\mathcal{X} \in \mathbb{R}^{H\times W\times T}$, in which T means consecutive T frames in the temporal order. $\mathbf{x}_{ij}$ represents the (i, j)th element of the matrix $\mathbf{X}$, and $\mathbf{x}_{ijk}$ represents the (i, j, k)th element of the tensor $\mathcal{X}$.

For a stacked spatial-temporal tensor, unfolding along any mode converts the tensor into a matrix. For instance, the stacked spatial-temporal tensor $\mathcal{X}$ can be unfolded along the first mode as $\text{unfold}_1(\mathcal{X}) = X_1 \in \mathbb{R}^{H\times(WT)}$. Likewise, unfolding along the second and third modes yields $\text{unfold}_2(\mathcal{X}) = X_2 \in \mathbb{R}^{W\times(HT)}$ and $\text{unfold}_3(\mathcal{X}) = X_3 \in \mathbb{R}^{T\times(HW)}$. The corresponding inverse operators are written as $\text{fold}_k(\mathbf{X}_k) = \mathcal{X}, \{k = 1, 2, 3\}$. The Frobenius norm of $\mathcal{X}$ is defined as $\|\mathcal{X}\|_F = (\sum_{i,j,k} |\mathbf{x}_{i,j,k}|^2)^{1/2}$.

III. PROPOSED TARGET DETECTION FRAMEWORK

Flowchart of the proposed TP-STTA is shown in Fig. 1, which mainly includes two complementary strategies to comprehensively extract features. First, each frame of the input infrared sequence is processed to extract multi-topological features by MTFE model. At the same time, a spatial-temporal tensor is constructed from the consecutive frames and analyzed by the MPTA model, where multiple physically meaningful priors are jointly incorporated, including target-oriented multi-topological priors as well as background-oriented smoothness and low-rank priors. It should be noted that the multi-topological priors are derived from MTFE in the first stage. Subsequently, the outputs of these two complementary strategies are fused, and temporal trajectory information is exploited to determine the final detection results. The details are introduced as follows.

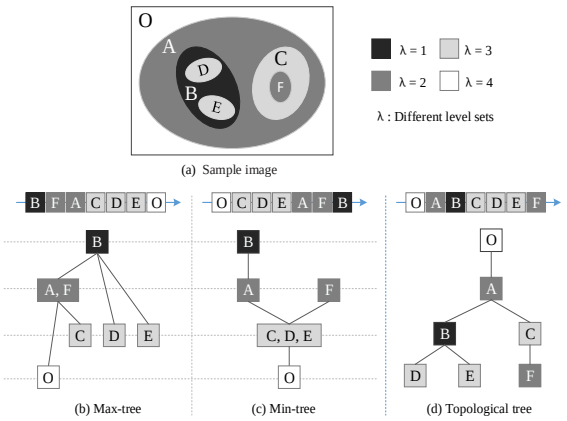


Fig. 2. Compare the topological-tree to the max-tree/min-tree.

A. Multi-topological Feature Extraction (MTFE)

Natural images can be represented using various tree-based structures, including component tree, max/min-tree, and topological-tree, which capture essential morphological information [51]. Due to their effectiveness, these representations have received considerable attention [52]. The transformation process generally involves three stages: topological-tree construction, filtering, and image reconstruction [53, 54].

(a) Topological-tree construction

Different from our previous work [54], which employed max-tree and min-tree for feature extraction, this work utilizes topological-tree instead. In [54], the max-tree and min-tree processing were applied to detect bright targets with dim surroundings and the opposite case. Topological trees, however, possess a dual characteristic, making no assumptions about the contrast of targets. Consequently, constructing a topological-tree enables the simultaneous detection of both bright and dim targets. In short, it can be regarded as the result of merging the dual component trees, min-tree and max-tree, into a single unified tree [55].

Fig. 2(a) shows a sample image, where different colors represent intensity variations, and λ denotes the corresponding level sets. From Fig. 2(a), it is clear that a max-tree or min-tree is organized by sorting gray levels in ascending or descending order, as indicated by the blue arrows. In contrast, a topological-tree does not follow this intensity-based ordering. Instead, it is constructed based on the inclusion relationships among connected components belonging to the same level set of the image. Therefore, topological-tree is also known as tree of shapes or inclusion tree [56]. Each node in the topological-tree corresponds to a connected component, defined as a region composed of 4-connected or 8-connected pixels, and each connected component is represented by only one node in the tree. The topological-tree is constructed from the connected components of the level set and captures the inclusion relationships among these components. Please refer to [55, 57] for more details.

(b) Filtering Process

Once the tree is constructed, each node is characterized by specific attributes. As illustrated in Fig. 3, the original image is first transformed into its corresponding topological-tree, where the target is marked with a red box and the target-like flickering noise is marked with a green box. It is evident that the flickering noise not only resembles the target in size but also exhibits similar intensity characteristics, making it difficult to distinguish between them using a single attribute. This motivates the use of multi-attributes together with appropriate pruning values (ϑ) to effectively separate true targets from flickering noise.

In addition to the area attribute ($\mathfrak{A}_A$) used in our work [54], which is mainly employed to capture size information. Two other increasing attributes, the volume attribute ($\mathfrak{A}_V$) and the diagonal attribute ($\mathfrak{A}_D$), are also incorporated. $\mathfrak{A}_V$ extracts intensity difference information between target and its neighbors,

$$\mathfrak{A}_V(\zeta) = \mathfrak{A}_A(\zeta) \times \left(g(p) - \sum_{\max_{p\in\zeta}} g(p) \right) \right) \tag{1}$$

where ζ is the node of the constructed topological-tree, p denote the pixels in ζ, and $g\,(p)$ denotes the gray value of p.

$\mathfrak{A}_D$ denotes the bounding box that encloses the nodes connected component, effectively integrating both size and shape information,

$$\mathfrak{A}_D\,(\zeta) = \sqrt{(x_{i,max} - x_{i,min})^2 + (y_{i,max} - y_{i,min})^2} \tag{2}$$

where $x_{i,max}$, $x_{i,min}$, $y_{i,max}$, and $y_{i,min}$ represent the coordinates of the bounding box.

For each individual attribute, appropriate ϑ are then selected according to whether the flickering noise or the target is suppressed, thereby determining whether the corresponding node is removed or preserved. In this way, pruning strategies provide an effective means of evaluating the persistence of regional extrema.

Taking $\mathfrak{A}_A$ as an example, the influence of different values of ϑ is illustrated in Fig. 3. Two representative settings are considered, corresponding to pruning value 1 ($\vartheta 1$) and pruning value 2 (ϑ_2). Their corresponding pruning results are shown in Fig. 3(c) and (f), respectively. When an appropriate ϑ_1 is selected, only the node containing flickering noise is removed, while

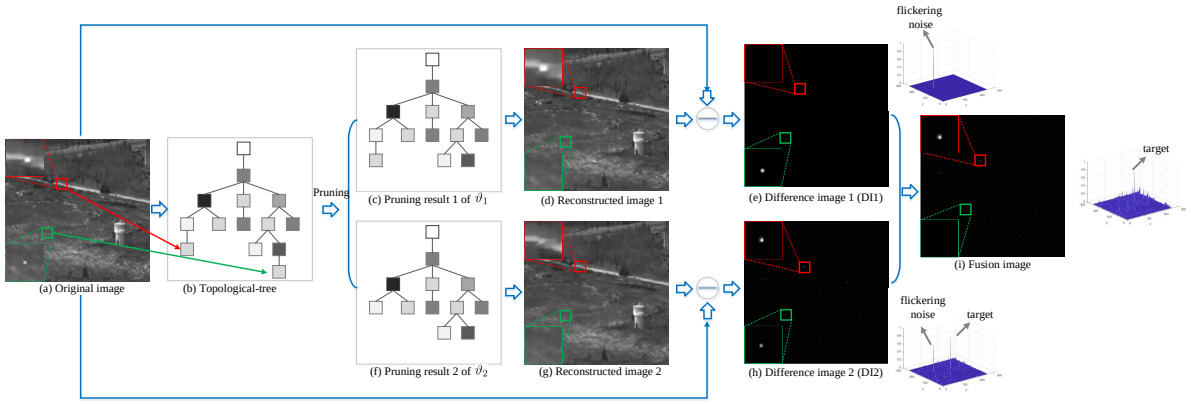


Fig. 3. Illustration of the influence of different pruning values ϑ.

the target node is preserved. Consequently, the target remains in the reconstructed image and the flickering noise is suppressed. Taking the difference between the original image and reconstructed image 1, difference image 1 (DI1) is obtained, in which the flickering noise is effectively highlighted, as shown in Fig. 3(e).

In contrast, when ϑ_2 is chosen, both the nodes corresponding to flickering noise and those associated with the target are removed, as illustrated in Fig. 3(f). As a result, neither the target nor the flickering noise appears in the reconstructed image, as shown in Fig. 3(g). Under this pruning setting, differencing the original image and the reconstructed image 2 yields difference image 2 (DI2), in which both the flickering noise and the target are simultaneously enhanced, as shown in Fig. 3(h).

At last, DI1 and DI2 are fused to obtain the fusion image, ensuring that flickering noise appearing simultaneously across frames is removed, while targets that do not appear simultaneously are preserved, as shown in Fig. 3(f).

(c) Multi-attribute Multi-pruning Fusion

Although the constructed topological-tree is different from the max-tree in [54], the conclusion is the same: a attribute and a pruning value cannot achieve satisfactory detection results. Therefore, to effectively distinguish the target not only from flickering noise but also from complex background clutter, a multi-attribute multi-pruning fusion strategy is adopted.

Since extensive analysis across multiple datasets indicates that flickering noise exhibits relatively consistent characteristics across different images, a single ϑ is sufficient for feature extraction with respect to the $\mathfrak{A}_A$, $\mathfrak{A}_V$, and $\mathfrak{A}_D$. The corresponding results are denoted as R_A^0, R_V^0, and R_D^0, respectively. The fusion image $R_1(t)$ of tth frame is as follows,

$$R_1(t) = \prod_{k\in\{\mathfrak{A}_A,\mathfrak{A}_V,\mathfrak{A}_D\}} R_k^0 \tag{3}$$

in which only the flickering noise preserved.

In contrast, targets vary significantly in size, type, and intensity across different scenarios; therefore, multiple ϑ are employed to different attributes in a more robust manner. The corresponding results are denoted as $\left\{R^1_{\mathfrak{A}_A}, R^2_{\mathfrak{A}_A}, ...R^n_{\mathfrak{A}_A}\right\}$, $\left\{R^1_{\mathfrak{A}_V}, R^2_{\mathfrak{A}_V}, ..., R^n_{\mathfrak{A}_V}\right\}$, and $\left\{R^1_{\mathfrak{A}_D}, R^2_{\mathfrak{A}_D}, ..., R^n_{\mathfrak{A}_D}\right\}$. Then, they are fused as below,

$$R_2(t) = \prod_{k\in\{\mathfrak{A}_A,\mathfrak{A}_V,\mathfrak{A}_D\}} \frac{1}{N}\sum_{n=1}^{N} R_k^{(1,2,...,n)} \tag{4}$$

where N denotes the number of the ϑ. Note that in R_2, both the flickering noise and the target have been enhanced.

Then, $R_1(t)$ and $R_2(t)$ are fused to suppress the flickering noise appearing in both results, while preserving the targets that appear only in R_2. The mask of flickering noise is defined

$$M(t) = \begin{cases} 1, & x \in \arg\max R_1(t), \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The final output of MTFE is as follows,

$$R_o(t) = R_2(t) \times (1 - M(t)). \tag{6}$$

By incorporating the area, volume, and diagonal attributes, multiple characteristics, including size, shape, and intensity, can be jointly extracted. Moreover, by selecting different ϑ for these attributes, target can be effectively distinguished not only from flickering noise but also from complex background clutter.

On the one hand, the feature extraction results produced by MTFE can be incorporated into the MPTA model as prior information to guide its optimization. On the other hand, these MTFE-derived features can

also serve as an independent single-frame representation, which is fused with the output of the MPTA model to generate the final result.

B. Multi-Prior Tensor Approximation (MPTA) Model

The consecutive input infrared frames are stacked into a tensor $\mathcal{Y}$ with dimension $H \times W \times T$, which can be decomposed into a background tensor $\mathcal{X}$ and a target tensor $\mathcal{S}$, such that $\mathcal{Y} = \mathcal{X} + \mathcal{S}$. Various priors, including background and target priors, can be embedded in this tensor model. First, the background tensor $\mathcal{X}$ unfolds the matrix $\mathbf{X}_3$ along the temporal dimension, using a truncated nuclear norm regularization (TNNR [58]) to encourage the $\mathbf{X}_3$ to be low-rank. Then, $\mathcal{X}$ is unfolded along the spatial dimensions to obtain $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. A linear TV-norm regularization is applied to both $\mathbf{X}_1$ and $\mathbf{X}_2$ to further enforce the piecewise smoothness of the background. For the target tensor, it is unfolded along its temporal dimension as $\mathbf{S}$. Therefore, the proposed MPTA can be formulated

$$\begin{aligned} \min_{\mathcal{X},\mathcal{S}} \quad & \frac{1}{2}\left(\|\nabla_H \mathbf{X}_1\|_F^2 + \|\nabla_W \mathbf{X}_2\|_F^2\right) + \alpha\|\mathbf{X}_3\|_r + \beta\|\mathcal{S}\|_{1,w}, \\ \text{s.t.} \quad & \begin{cases} \mathcal{Y} = \mathcal{X} + \mathcal{S}, \\ \mathbf{X}_1 = \text{unfold}_1(\mathcal{X}), \\ \mathbf{X}_2 = \text{unfold}_2(\mathcal{X}), \\ \mathbf{X}_3 = \text{unfold}_3(\mathcal{X}), \\ \mathbf{S}_3 = \text{unfold}_3(\mathcal{S}). \end{cases} \end{aligned} \tag{7}$$

where $\nabla_H \in \mathbb{R}^{(H-1)\times H}$ and $\nabla_W \in \mathbb{R}^{(W-1)\times W}$ are defined

$$\nabla_H = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \tag{8}$$

and

$$\nabla_W = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \tag{9}$$

respectively.

The first two terms in the objective Eq. (7) correspond to the linear total variation (LTV)-norm regularization (see Section III-B(a)), while the third and fourth terms represent the TNNR (see Section III-B(b)) and the multi-topological features guidance (see Section III-B(c)), respectively. Since the row-wise and column-wise effects in spatial domain are considered equally important, the trade-off parameter of the smoothness prior is set to 1.

(a) Background-aware Piecewise Smoothness Prior

In infrared image sequences, the background component usually exhibits strong spatial continuity and local smoothness along both horizontal and vertical directions. To characterize this property, a LTV norm is adopted to impose a piecewise smooth prior on the background tensor.

$$\text{LTV}(\mathbf{X}) = \sum_{i=1}^{H}\sum_{j=1}^{W} \mathbf{X}_v(i,j)^2 + \mathbf{X}_h(i,j)^2, \tag{10}$$

where

$$\mathbf{X}_v(i,j) = \begin{cases} \mathbf{X}(i+1,j) - \mathbf{X}(i,j), & \text{if } i < H \\ 0, & \text{if } i = H \end{cases} \tag{11}$$

and

$$\mathbf{X}_h(i,j) = \begin{cases} \mathbf{X}(i,j+1) - \mathbf{X}(i,j), & \text{if } j < H \\ 0, & \text{if } j = W. \end{cases} \tag{12}$$

This regularization encourages spatial smoothness of the background while preserving local structural transitions, which is particularly suitable for infrared scenes with slowly varying backgrounds.

(b) Background-aware Low-Rank Prior

A low-rank prior on $\mathbf{X} \in \mathbb{R}^{H\times W}$ is commonly characterized by the rank function $\text{rank}(\mathbf{X})$ as a regularization term. However, directly minimizing $\text{rank}(\mathbf{X}_3)$ leads to an NP-hard optimization problem. To circumvent this difficulty, the nuclear norm $\|\mathbf{X}\|_*$ is employed as a convex surrogate. The nuclear norm is defined as the sum of the singular values of $\mathbf{X}$. Accordingly, the proximal operator of the nuclear norm, denoted by $\text{prox}\|\mathbf{X}\|_{*,\alpha}(\mathbf{X})$, admits a closed-form solution via the singular value shrinkage operator,

$$\text{prox}_{\|\bullet\|_*,\alpha}(\mathbf{Y}) = \text{D}_\alpha(\mathbf{Y}) = \arg\min_{\mathbf{X}} \left\{ \alpha\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2 \right\} \tag{13}$$

where $\text{D}_\alpha(\mathbf{Y}) = \mathbf{U}\mathbf{S}_\alpha(\mathbf{\Sigma})\mathbf{V}'$. Here, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$ are obtained from the singular value decomposition of $\mathbf{Y}$, i.e., $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$. The operator $\mathbf{S}_\alpha(\mathbf{\Sigma}) = \text{diag}(\max(\sigma_i(\mathbf{Y}) - \alpha, 0))$ is a soft-thresholding operator on the diagonal of the matrix $\mathbf{\Sigma}$.

However, the nuclear norm is only an imperfect surrogate for the rank function. In $\text{rank}(\mathbf{X})$, all nonzero singular values contribute equally, whereas their contributions to the nuclear norm $\|\mathbf{X}\|_*$ in infrared images vary significantly, often spanning several orders of magnitude. Motivated by this observation, TNNR is adopted, where $\|\mathbf{X}\|_r$ defined as the sum of $\min(H, W)$-$r$minimum singular values

$$\|\mathbf{X}\|_r = \sum_{i=r+1}^{\min(H,W)} \sigma_i(\mathbf{X}). \tag{14}$$

Eq. (14) is equal to Eq. (15) as follows:

$$\|\mathbf{X}\|_r = \|\mathbf{X}\|_* - \max_{\mathbf{A}\mathbf{A}^{\mathrm{T}}=\mathbf{B}\mathbf{B}^{\mathrm{T}}=\mathbf{I}_{\mathrm{r}}} \mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}^T). \tag{15}$$

Thus, the proximal operator of TNNR, denoted as $\mathrm{prox}_{\|\bullet\|_*,\alpha}(\mathbf{Y})$, can be efficiently computed via a two-stage strategy. In the first stage, the leading r singular components are extracted through singular value decomposition to estimate the factor matrices $\mathbf{A}$ and $\mathbf{B}$. In the second stage, with $\mathbf{A}$ and $\mathbf{B}$ fixed, the remaining subproblem reduces to a nuclear-norm regularized optimization, which admits a closed-form solution via singular value thresholding. These two stages are alternately executed until convergence in the proposed algorithm.

(c) Target-aware Sparse Prior

Unlike the background component, small infrared targets are spatially sparse and manifest as localized anomalies across frames. To further enhance target awareness during the sparse decomposition process, the results of MTFE are incorporated. Specifically, a multi-topological prior tensor $\mathcal{P}$ is constructed from consecutive responses R_o via preliminary target enhancement. Each element of $\mathcal{P}$ represents the likelihood of target presence, with higher values indicating a greater probability of belonging to a true target region. This prior tensor is then embedded to guide the extraction of the sparse component. Based on $\mathcal{P}$, a spatial-temporal weight map is constructed as

$$w = 1 + \gamma\left(1 - \mathcal{P}\right), \tag{16}$$

where $\gamma > 0$ controls the strength of the multi-topological guidance. Pixels with strong responses are assigned smaller weights, reducing their penalization, while background regions are suppressed by larger weights.

The sparse prior is then formulated as

$$\beta\|\mathcal{S}\|_{1,w}, \tag{17}$$

and the weighted l_1-norm is defined as

$$\|\mathcal{S}\|_{1,w} = \sum_{i,j,t} w_{i,j,t}|\mathcal{S}_{i,j,t}|. \tag{18}$$

Incorporating multi-attributes topological priors as weighting factors in the sparse tensor update guides the algorithm to focus on potential target regions, enhancing the accuracy and robustness of infrared small-target detection while effectively suppressing background interference.

(d) ADMM Optimization for Proposed MPTA

To efficiently solve Eq. (7), the ADMM framework is adopted. By introducing auxiliary variables $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$, and $\mathbf{V}_4$, the augmented Lagrangian function is written as

$$\begin{aligned}
&\mathrm{L}(\mathcal{X},\mathcal{S},\mathbf{V}_1,\mathbf{V}_2,\mathbf{V}_3,\mathbf{V}_4,\mathbf{D}_1,\mathbf{D}_2,\mathbf{D}_3,\mathbf{D}_4,\mathbf{A},\mathbf{B})\\
&= \frac{1}{2}\left(\|\mathbf{D}_H\mathbf{V}_1\|_F^2 + \|\mathbf{D}_W\mathbf{V}_2\|_F^2\right)\\
&+ \alpha\left(\|\mathbf{V}_3\|_* - \max_{\mathbf{A}\mathbf{A}^{\mathbf{T}}=\mathbf{B}\mathbf{B}^{\mathbf{T}}=\mathbf{I}_{\mathbf{r}}} \mathrm{Tr}(\mathbf{A}\mathbf{V}_4\mathbf{B}^T)\right) + \beta\|\mathcal{S}\|_{1,w}\\
&+ \frac{\mu}{2}\sum_{i=1}^{3}\|\mathcal{X}_i - \mathbf{V}_i + \mathbf{D}_i\|_F^2 + \frac{\mu}{2}\|\mathbf{V}_4 - \mathbf{V}_3 + \mathbf{D}_4\|_F^2\\
&+ \frac{\delta}{2}\left\|\mathcal{S} - \left(\mathcal{Y} - \mathcal{X}^k + \mathbf{D}_1^k\right)\right\|_F^2,
\end{aligned} \tag{19}$$

where $\{\mathbf{D}_i\}_{i=1}^4$ are the scaled Lagrange multipliers and $\mu > 0$ and $\delta > 0$ are the penalty parameters. Eq.(19) is solved by alternatively updating each variable while keeping the others fixed. Specifically, at the k-th iteration, the optimization procedure can be carried out through the following steps.

1) Update of $\mathbf{V}_1$ and $\mathbf{V}_2$: The update of $\mathbf{V}_1$ is obtained by solving

$$\mathbf{V}_1^{k+1} = \arg\min_{\mathbf{V}_1} \frac{1}{2}\left\|\nabla_H\mathbf{V}_1^k\right\|_F^2 + \frac{\mu}{2}\left\|\mathcal{X}_1^k - \mathbf{V}_1^k + \mathbf{D}_1^k\right\|_F^2, \tag{20}$$

We can obtain a closed-form solution by updating $\mathbf{M}_1^{k+1}$

$$\mathbf{V}_1^{k+1} = \mu\left(\nabla_H^T\nabla_H + \mu\mathbf{I}_H\right)^{-1}\mathcal{X}_1^k + \mathbf{D}_1^k, \tag{21}$$

where $\mathbf{I}_H$ is an identity matrix with $H \times H$.

Similarly, the closed-form solution is obtained by updating $\mathbf{V}_2^{k+1}$ is updated by

$$\mathbf{V}_2^{k+1} = \mu\left(\nabla_W^T\nabla_W + \mu\mathbf{I}_W\right)^{-1}\mathcal{X}_2^k + \mathbf{D}_2^k, \tag{22}$$

where $\mathbf{I}_W$ is an identity matrix with $W \times W$.

2) Update of $\mathbf{V}_3$: The objective function can be written as

$$\begin{aligned}
\mathbf{V}_3^{k+1} &= \arg\min_{\mathbf{V}_3} \alpha\left\|\mathbf{V}_3^k\right\|_* + \frac{\mu}{2}\left(\left\|\mathcal{X}_3^k - \mathbf{V}_3^k + \mathbf{D}_3\right\|_F^2\right.\\
&\quad \left.+ \left\|\mathbf{V}_4^k - \mathbf{V}_3^k + \mathbf{D}_4\right\|_F^2\right)\\
&= \arg\min_{\mathbf{V}_3} \alpha\left\|\mathbf{V}_3^k\right\|_*\\
&\quad + \mu\left\|\mathbf{V}_3^k - \frac{\mathcal{X}_3^k + \mathbf{D}_3^k + \mathbf{V}_4^k + \mathbf{D}_4^k}{2}\right\|_F^2
\end{aligned} \tag{23}$$

3) Update of $\mathbf{V}_4$: The objective function can be written as

$$\begin{aligned}
\mathbf{V}_4^{k+1} &= \operatorname*{argmin}_{\mathbf{V}_4} -\mathrm{Tr}(\mathbf{A}^k\mathbf{M}_4^k\mathbf{B}^{\mathrm{k}\mathrm{T}})\\
&\quad + \frac{\mu}{2}\left\|\mathbf{V}_4^k - \mathbf{V}_3^{k+1} + \mathbf{D}_4^k\right\|_F^2,
\end{aligned} \tag{24}$$

and the closed-form solution can be derived as

$$\mathbf{V}_4^{k+1} = \mathbf{V}_3^{k+1} - \mathbf{D}_4^k + \frac{\mathbf{A}^k\mathbf{B}^{\mathrm{k}\mathrm{T}}}{\mu}. \tag{25}$$

4) Update of $\mathbf{A}$ *and* $\mathbf{B}$*:* $\mathbf{A}$ and $\mathbf{B}$ can be derived as below

$$\left[\mathbf{A}^{k+1}, \Sigma^{k+1}, \mathbf{B}^{k+1}\right] = \text{TSVDr}\left(\mathbf{V}_4^{k+1}\right). \tag{26}$$

TSVDr denotes the rank-r truncated singular value decomposition, which computes the leading r singular values and corresponding singular vectors.

5) Update of $\mathcal{X}$*:* Fixing other variables, the subproblem with respect to $\mathcal{X}$ is quadratic and admits the following closed-form solution:

$$\mathcal{X}^{k+1} = \frac{1}{4}\left(\mathcal{Y} - \mathcal{S}^k + \mathbf{D}_1^k + \sum_{i=1}^{3}(\mathcal{V}_i^k + \mathbf{D}_{i+1}^k)\right). \tag{27}$$

6) Update of $\mathcal{X}_1$*,* $\mathcal{X}_2$*, and* $\mathcal{X}_3$*:*

$$\begin{aligned} \mathcal{X}_1^{k+1} &= \text{unfold}_1\left(\mathcal{X}^{k+1}\right), \\ \mathcal{X}_2^{k+1} &= \text{unfold}_2\left(\mathcal{X}^{k+1}\right), \\ \mathcal{X}_3^{k+1} &= \text{unfold}_3\left(\mathcal{X}^{k+1}\right). \end{aligned} \tag{28}$$

7) Update of $\mathcal{S}$ *with multi-topological features guidance:* The sparse component $\mathcal{S}$ is updated by solving

$$\mathcal{S}^{k+1} = \arg\min_{\mathcal{S}} \beta\|\mathcal{S}\|_{1,w} + \frac{\mu}{2}\left\|\mathcal{S} - \left(\mathcal{Y} - \mathcal{X}^k + \mathbf{D}_1^k\right)\right\|_F^2. \tag{29}$$

8) Multiplier Update and Convergence: The Lagrange multipliers are updated according to

$$\begin{aligned} \mathbf{D}_1^{k+1} &= \mathbf{D}_1^k + (\mathcal{Y} - \mathcal{X}^{k+1} - \mathcal{S}^{k+1}), \\ \mathbf{D}_2^{k+1} &= \mathbf{D}_2^k + (\mathbf{V}_1^{k+1} - \mathcal{X}^{k+1}), \\ \mathbf{D}_3^{k+1} &= \mathbf{D}_3^k + (\mathbf{V}_2^{k+1} - \mathcal{X}^{k+1}), \\ \mathbf{D}_4^{k+1} &= \mathbf{D}_4^k + (\mathbf{V}_4^{k+1} - \mathbf{V}_3^{k+1}). \end{aligned} \tag{30}$$

Finally, the optimization of the low-rank background tensor $\mathcal{X}^*$ and sparse target tensor $\mathcal{S}^*$ can be derived when convergence has been achieved. The whole procedure is summarized in Algorithm 1.

Algorithm 1 MPTA Algorithm

Input: Infrared image sequence tensor $\mathcal{Y}$; morphological prior tensor $\mathcal{P}$; truncation rank r; hyperparameters setting of spatial-temporal tensor α, β, δ and μ; maximum iteration $K_{\max}$.
Output: Low-rank background tensor $\mathcal{X}^*$; sparse target tensor $\mathcal{S}^*$.
Initialization: $\mathcal{X} := \mathbf{0}, \mathcal{S} := \mathbf{0}$, $\mathbf{V}_i, \mathbf{D}_i := \mathbf{0}$, $\mathcal{X}_i = unfold_i(\mathcal{X})$, $\varepsilon = 10^{-6}$, $k = 1$.
while not converged and $k < K_{\max}$ **do**
  Update $\mathbf{V}_1$ by Eq.(21).
  Update $\mathbf{V}_2$ by Eq.(22).
  Update $\mathbf{V}_3$ by Eq.(23).
  Update $\mathbf{V}_4$ by Eq.(25).
  Update $\mathbf{A}$ and $\mathbf{B}$ according to Eq.(26).
  Update $\mathcal{X}$ by solving Eq.(27).
  Update $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{X}_3$ by solving Eq.(28).
  Construct w Eq.(16).
  Update $\mathcal{S}$ by solving Eq.(29).
  Update $\mathcal{D}_i$ by solving Eq.(30).
  if $\|\mathcal{Y} - \mathcal{X}^{k+1} - \mathcal{S}^{k+1}\|_F \le \varepsilon$ or $k > K_{\max}$ **then**
    break
  end if
  $k \leftarrow k + 1$
end while
$\mathcal{X}^* = \mathcal{X}^k$, $\mathcal{S}^* = \mathcal{S}^k$.

C. Prior-Guided Spatial-Temporal Feature Fusion

The single-frame spatial-topological results of the MTFE method excel at capturing the fine-grained, high-contrast structural characteristics of potential targets in individual frames, thus enabling precise spatial localization of targets. In contrast, the spatial-temporal results derived from the MPTA model capitalize on inter-frame spatial-temporal correlations and the structural features of targets across consecutive frames, which effectively enhances the consistency of target representation while suppressing background clutter. Subsequently, we intend to fuse these two types of results to obtain more satisfactory performance.

Specifically, for each frame image, the fusion response $\widetilde{\mathcal{R}}_{(:,:,t)}$ is computed as follows

$$\widetilde{\mathcal{R}}_{(:,:,t)} = \mathcal{S}^*_{(:,:,t)} \odot R_{o(:,:,t)}, \tag{31}$$

where $\mathcal{S}^*_{(:,:,t)}$ denotes the t-th slice of the recovered sparse tensor from MPTA, $R_{o(:,:,t)}$ represents the corresponding t-th frame result of MTFE, and $\odot$ denotes the Hadamard product.

Continuous $\widetilde{\mathcal{R}}_{(:,:,t)}$ are then aggregated across the temporal dimension to obtain a final saliency map. Let $\left\{\widetilde{\mathcal{R}}_{i,j,t}\right\}_{t=1}^{T}$ represent the enhanced target tensor accumulated from the first frame up to the t-th frame. The aggregated target responses map is constructed as

$$\mathcal{O}(i,j) = \mathcal{F}\left(\left\{\widetilde{\mathcal{R}}_{i,j,t}\right\}_{t=1}^{T}\right), \tag{32}$$

where $\mathcal{F}(\cdot)$ denotes a temporal fusion operator designed to integrate aggregated maps.

In practice, the fusion is realized through sliding-window temporal integration, enabling the enhancement of temporally stable target responses while mitigating intermittent noise. As a result, persistent small targets are emphasized, whereas background clutter and isolated false detections are effectively removed.

IV. Experimental Results and Discussion

In this section, we compare the proposed method with six real infrared image sequences from 10 other

methods to demonstrate its effectiveness and robustness. All experiments were conducted on a computer with 16 GB of RAM and an Intel(R) Core(TM) i7-1260P processor featuring Intel(R) Iris(R) Xe integrated graphics. All the comparison methods were executed in MATLAB R2021b. The proposed method was implemented in Visual Studio 2017 and MATLAB R2021b, in which MTFE was conducted in Visual Studio 2017 using C++ language, and MPTA and small target detection was performed in MATLAB R2021b.

A. Experimental Settings

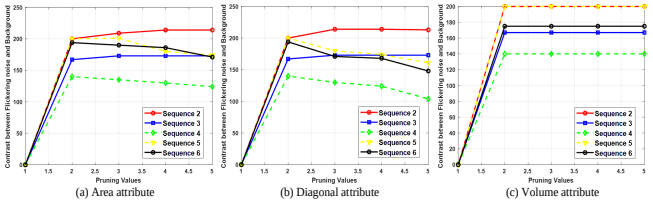
1) Measurement Metrics: To enable a more comprehensive evaluation of the proposed TP-STTA, a 3D receiver operating characteristic (3D ROC) curve [59, 60] is employed, which extends the conventional ROC analysis by explicitly incorporating the cut-off threshold dimension. Based on the 3D ROC representation, three 2-Dimensional (2D) ROC curves, namely $(\mathrm{P_D}, \mathrm{P_F})$, $(\mathrm{P_D}, \tau)$, and $(\mathrm{P_F}, \tau)$, can be derived. For the $(\mathrm{P_D}, \mathrm{P_F})$ and $(\mathrm{P_D}, \tau)$ curves, better detection performance is indicated by curves closer to the upper-right corner. Conversely, for the $(\mathrm{P_F}, \tau)$ curves, curves approaching the lower-left corner reflect stronger background suppression capability. Furthermore, eight area-under-the-curve (AUC) metrics are obtained, including

$$\begin{cases} \mathrm{AUC}_{(\mathrm{D,F})} \in [0,1], \\ \mathrm{AUC}_{(\mathrm{D},\tau)} \in [0,1], \\ \mathrm{AUC}_{(\mathrm{F},\tau)} \in [0,1], \\ \mathrm{AUC_{TD}} = \mathrm{AUC}_{(\mathrm{D,F})} + \mathrm{AUC}_{(\mathrm{D},\tau)} \in [0,2], \\ \mathrm{AUC_{BS}} = \mathrm{AUC}_{(\mathrm{D,F})} - \mathrm{AUC}_{(\mathrm{F},\tau)} \in [-1,1], \\ \mathrm{AUC_{SNPR}} = \left(\mathrm{AUC}_{(\mathrm{D},\tau)} / \mathrm{AUC}_{(\mathrm{F},\tau)}\right) \in [0,+\infty], \\ \mathrm{AUC_{TD\text{-}BS}} = \mathrm{AUC}_{(\mathrm{D},\tau)} - \mathrm{AUC}_{(\mathrm{F},\tau)} \in [-1,1], \\ \mathrm{AUC_{ODP}} = \mathrm{AUC}_{(\mathrm{D,F})} + \mathrm{AUC}_{(\mathrm{D},\tau)} - \mathrm{AUC}_{(\mathrm{F},\tau)} \in [-1,2]. \end{cases} \tag{33}$$

Benefiting from the 3D ROC, the performance of different methods can be evaluated from multiple perspectives, including target detection (TD), background suppression (BS), and threshold sensitivity. For a detailed introduction of the 3D ROC, readers are referred to [59, 60]. Except for the metrics related to $\mathrm{AUC}_{(\mathrm{P_F},\tau)}$, where smaller values indicate superior performance, all other AUC-based metrics are positively correlated with detection performance and favor larger values.

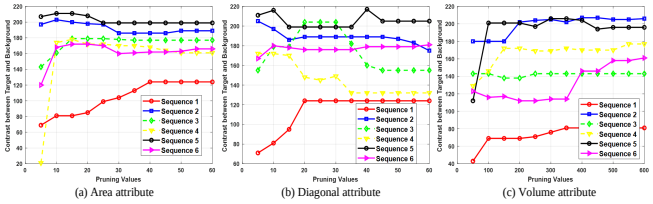
2) Baseline Comparison Methods: The proposed TP-STTA is compared with ten different infrared small-target detection methods, including three single-frame methods (GCM [17], NTFRA [23], and PTCTV-KMC [24]) and seven multi-frame methods (GLLS-LMAG [41], ECA-STT [42], ASTTV [43], STRL-LBCM [44], 4DTT [45], SDD [46], and FGLR-MCP [47]). The detailed parameter settings of all the methods involved in the experiments are summarized in Table I.

3) Experimental Datasets: To thoroughly validate the effectiveness of the proposed method, six real infrared sequences, referred to as Sequence 1 through Sequence 6, are employed in the experiments. All sequences have a spatial resolution of 256×256, and detailed descriptions are provided in Table II. The sequences cover diverse scene types. Owing to the limited texture and spatial structural information present in real infrared imagery, these sequences pose significant challenges for infrared small-target detection. Notably, all datasets are obtained from the publicly available repository of the National University of Defense Technology [1].



(a) Area attribute (b) Diagonal attribute (c) Volume attribute

Fig. 4. Contrast between flickering noise and background using different ϑ_s.



(a) Area attribute (b) Diagonal attribute (c) Volume attribute

Fig. 5. Contrast between target and background using different ϑ_s.

B. Parameter Setting

For MTFE, appropriate ϑ plays a critical role. In this work, two sets of parameters, denoted as ϑ_1 and ϑ_2, need to be determined. Specifically, the optimal ϑ_1 is employed to suppress flickering noise, whereas the optimal ϑ_2 is used for target feature extraction. The experimental results obtained with different values of ϑ_1 and ϑ_2 are presented in Fig. 4 and Fig. 5. In both figures, the horizontal axis corresponds to different pruning values, while the vertical axis denotes the contrast metric defined in [54]. In particular, Fig. 4 depicts the contrast between the target and the background, whereas Fig. 5 illustrates

[1] http://www.sciencedb.cn/dataSet/handle/902

TABLE I

PARAMETER SETTINGS FOR ELEVEN METHODS IN THE EXPERIMENTS.

No.	Methods	Year	Parameter settings
1	GCM [17]	2023	$R_o = 3, R_i = 2.$
2	NTFRA [23]	2022	Patch size: 40×40, step :40, $\lambda = 1/\sqrt{\min(n_1, n_2) * n_3}$, $\beta = 0.01$, $\mu = 200$.
3	PTCTV-KMC [24]	2025	Patch size: 50×50, $N = 1$, $\lambda = \alpha\frac{L}{\sqrt{\frac{n_1*n_2*n_3}{\max(n_1,n_2)}}}$, $L = 0.85$, $\mu = 0.15$.
4	GLLS-LMAG [41]	2025	$n_3 = 20, \lambda_1 = 0.02, \lambda_2 = 1/\sqrt{\prod(n_1, n_2, n_3)/min(n_1, n_2)}, \varepsilon = 10^{-7}, \mu = 10^{-3}, \beta = 10^{-5}, \rho = 1.1.$
5	ECA-STT [42]	2021	$\beta = 0.1, t = 3, \lambda_1 = 0.009, \lambda_2 = 5.0/\sqrt{\min(m, n) \times t}, \lambda_3 = 100, \varepsilon = 10^{-7}.$
6	ASTTV [43]	2021	$L = 3, H = 6, \lambda_{tv} = 0.005, \lambda_s = \frac{H}{\sqrt{\max(M,N)*L}}, \lambda_3 = 100.$
7	STRL-LBCM [44]	2023	frame $= 16, \lambda = 1, \lambda_1 = 10^3\lambda_L/\sqrt{\min(I_1, I_2, I_3)}, \lambda_2 = 50\lambda_1, \lambda_3 = 0.1\lambda_1, \lambda_L = 10,$ $\mu = 0.05, \alpha = 0.5, \beta = 1, \eta = 1.5, r_\Lambda = 0.1, h = 0.011, \varepsilon = 10^{-4}.$
8	4DTT [45]	2023	Patch size: 10×10, temporal size: 15, $\lambda_1 = \sum_{i=1}^{m-1} 2/\sqrt{\max\left(\prod_{i=1}^{k} N_i, \prod_{i=k+1}^{m-1} N_i\right)}$, $\tau = 1 \times 10^{-7}$.
9	SDD [46]	2024	Frame no.: 30, R=10, $\lambda \in [0.5, 5]$, $\gamma \in [0.1, 1]$, i_{outer}: 50, i_{inner}: 10,$\beta = 15000$, $\mu = 0.05$, $\varepsilon = 0.01$.
10	FGLR-MCP [47]	2025	$L = 8, H = 8, r_1 = 4.5, r_2 = 1, \rho = 1.5, \lambda = \frac{H}{\sqrt{\max(M,N)*L}}, \text{tol} = 1 \times 10^{-6}, \mu = 0.0027.$
11	The Proposed		$\mathfrak{A}_A$, $\mathfrak{A}_D$, $\mathfrak{A}_V$: $\vartheta_1 = \{1, 2, 3, 4, 5\}$, $\mathfrak{A}_A$, $\mathfrak{A}_D$: $\vartheta_2 = \{10, 15, 20\}$, $\mathfrak{A}_V$ = $\vartheta_2 = \{300, 350, 400\}$, r = 1, l = 6, γ = 200, β = 0.1, δ = 0.1, α = 1, μ = 0.001

TABLE II

DETAILS OF SIX INFRARED SMALL TARGET DETECTION SEQUENCES USING IN THE EXPERIMENTS.

	Frame	Image Resolution	Target Description	Background Description
Sequence 1	60	256×256	• A long imaging distance. • An airplane with changing size.	• Heterogeneous background containing forest regions and highly reflective structures.
Sequence 2	60	256×256	• A long imaging distance. • Slowly moving target.	• Significant road and high contrast buildings. • Target-like flickering noise.
Sequence 3	120	256×256	• A long imaging distance. • An airplane with small size.	• Changing background. • Target-like flickering noise.
Sequence 4	120	256×256	• A long imaging distance. • An airplane with quite small size.	• Complex forest background. • Target-like flickering noise.
Sequence 5	180	256×256	• The target brightness is not obvious. • The target occupies a small number of pixels.	• Complex sky-ground background. • Target-like flickering noise.
Sequence 6	180	256×256	• An airplane with small size. • The target with changing size and distance.	• Target-like flickering noise. • Sky-ground background changes quickly.

the contrast between flickering noise and the background. Higher contrast values indicate superior performance.

1) Analysis of pruning value ϑ_1: Different values of ϑ_1 are evaluated on Seqs. 2-6, since Seq. 1 does not contain flickering noise. The corresponding results are illustrated in Fig. 4, where Figs. 4(a), (b), and (c) present the outcomes obtained using $\mathfrak{A}_A$, $\mathfrak{A}_D$, and $\mathfrak{A}_V$, respectively. The parameter ϑ_1 is varied over the set $\{1, 2, 3, 4, 5\}$. It can be clearly observed that $\vartheta_1 = 2$ consistently achieves the best performance for all three attributes. Since the characteristics of flickering noise exhibit little variation across different images, a unified parameter setting is sufficient to effectively suppress such noise. Therefore, ϑ_1 is fixed to 2 for all three attributes in the subsequent experiments.

2) Analysis of pruning value ϑ_2: Similarly, different values of ϑ_2 are tested on all six sequences. As shown in Fig. 5, the candidate set $\{5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60\}$ is examined for $\mathfrak{A}_A$ and $\mathfrak{A}_D$, while a broader range $\{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600\}$ is explored for $\mathfrak{A}_V$. For different attribute, if ϑ is set to a single large value, both target regions and target-like background clutter may be removed during the pruning process, whereas a small value may fail to eliminate target nodes effectively. As a result, the optimal ϑ_2 is highly dependent on scene characteristics. To enhance robustness across diverse scenarios, multiple nonconsecutive pruning values are jointly adopted. Specifically, $\vartheta_2 = \{10, 15, 20\}$ is selected for for $\mathfrak{A}_A$ and $\mathfrak{A}_D$, while $\vartheta_2 = \{300, 350, 400\}$ is used for $\mathfrak{A}_V$.

There are several parameters of MPTA that need to be adjusted, including truncated low-rank r in TNNR, adjacent frame number l, the strength of morphological guidance γ, and hyperparameters setting of spatial-temporal tensor, i.e., α, β, μ, and δ. To explore how these parameters significantly influence the proposed method, we perform comprehensive parameter analysis on Seqs. 1-6. Limited by space, we present the evaluation metrics derived from the 3D ROC curves on Seq. 3 in Table III.

1) Analysis of truncated low-rank r: r controls the rank of the background subspace in the proposed model. To evaluate its influence, we conduct experiments with $r \in \{1, 2, 3, 4, 5\}$. The results are reported in Table III. Experimental results indicate that the best performance is achieved when $r = 1$, suggesting that the background in infrared small-target scenes can be effectively represented by a very low-rank structure. Therefore, r is fixed

TABLE III

3D ROC RESULTS OF THE PROPOSED METHOD ON SEQUENCE 3 WITH VARYING r, l, α, β, γ, μ, AND δ.

α	β	γ	μ	δ	r	l	$\mathrm{AUC}_{(\mathrm{D,F})}$	$\mathrm{AUC}_{(\mathrm{D},\tau)}$	$\mathrm{AUC}_{(\mathrm{F},\tau)}$	$\mathrm{AUC}_{\mathrm{TD}}$	$\mathrm{AUC}_{\mathrm{BS}}$	$\mathrm{AUC}_{\mathrm{TDBS}}$	$\mathrm{AUC}_{\mathrm{ODP}}$	$\mathrm{AUC}_{\mathrm{SNRP}}$
1	0.1	200	0.001	0.1	1	6	**0.9925**	**0.9923**	**0**	**1.9848**	**0.9924**	**0.9922**	**1.9847**	**20941.3747**
1	0.1	200	0.001	0.1	2	6	**0.9925**	0.9905	0.0001	1.9829	0.9924	0.9904	1.9828	9450.3568
1	0.1	200	0.001	0.1	3	6	**0.9925**	0.9841	0.0001	1.9765	0.9924	0.984	1.9764	9799.3028
1	0.1	200	0.001	0.1	4	6	**0.9925**	0.984	0.0001	1.9765	0.9924	0.9839	1.9763	8634.3094
1	0.1	200	0.001	0.1	5	6	**0.9925**	0.9844	0.0001	1.9769	0.9924	0.9843	1.9768	9109.8875
1	0.1	200	0.001	0.1	1	3	0.9924	0.9344	**0**	1.9268	**0.9924**	0.9344	1.9268	**43143.0791**
1	0.1	200	0.001	0.1	1	4	**0.9925**	0.9608	0.0001	1.9533	**0.9924**	0.9607	1.9532	18259.4624
1	0.1	200	0.001	0.1	1	5	**0.9925**	0.9707	**0**	1.9632	**0.9924**	0.9707	1.9631	20545.6554
1	0.1	200	0.001	0.1	1	6	**0.9925**	**0.9923**	**0**	**1.9848**	**0.9924**	**0.9922**	**1.9847**	20941.3747
1	0.1	200	0.001	0.1	1	7	**0.9925**	0.974	**0**	1.9665	**0.9924**	0.9739	1.9664	22504.5513
1	0.1	200	0.001	0.1	1	8	**0.9925**	0.9754	0.0001	1.9679	**0.9924**	0.9754	1.9678	19264.8479
0.001	0.1	200	0.001	0.1	1	6	**0.9925**	0.9872	**0**	1.9797	**0.9924**	0.9872	1.9797	**22047.4334**
0.01	0.1	200	0.001	0.1	1	6	**0.9925**	0.9873	**0**	1.9798	**0.9924**	0.9872	1.9797	22201.0818
0.1	0.1	200	0.001	0.1	1	6	**0.9925**	0.9847	**0**	1.9772	**0.9924**	0.9847	1.9771	24135.3996
1	0.1	200	0.001	0.1	1	6	**0.9925**	**0.9923**	**0**	**1.9848**	**0.9924**	**0.9922**	**1.9847**	20941.3747
10	0.1	200	0.001	0.1	1	6	**0.9925**	0.9664	0.0001	1.9589	**0.9924**	0.9663	1.9588	11312.8748
100	0.1	200	0.001	0.1	1	6	0.9031	0.8136	0.0002	1.7167	0.9029	0.8134	1.7165	3319.3751
1	0.001	200	0.001	0.1	1	6	**0.9925**	0.9829	0.0001	1.9754	**0.9924**	0.9828	1.9753	19392.5382
1	0.01	200	0.001	0.1	1	6	**0.9925**	0.9587	0.0001	1.9512	**0.9924**	0.9587	1.9511	19040.7507
1	0.1	200	0.001	0.1	1	6	**0.9925**	**0.9923**	**0**	**1.9848**	**0.9924**	**0.9922**	**1.9847**	**20941.3747**
1	1	200	0.001	0.1	1	6	0.6036	0.1588	0.0023	0.7624	0.6013	0.1565	0.7601	69.3513
1	10	200	0.001	0.1	1	6	0.9761	0.9235	0.0001	1.8996	0.976	0.9235	1.8996	14231.7358
1	100	200	0.001	0.1	1	6	0.6375	0.2387	0.0018	0.8762	0.6357	0.2369	0.8744	132.4164
1	0.1	1	0.001	0.1	1	6	0.9925	0.9791	**0**	1.9715	0.9924	0.979	1.9715	21725.9585
1	0.1	50	0.001	0.1	1	6	0.9789	0.9631	0.0001	1.942	0.9788	0.963	1.9419	12964.5600
1	0.1	100	0.001	0.1	1	6	0.9731	0.9649	0.0001	1.938	0.973	0.9648	1.9379	15225.3657
1	0.1	150	0.001	0.1	1	6	**0.9925**	0.9897	0.0001	1.9821	0.9924	0.9896	1.9821	18251.4683
1	0.1	200	0.001	0.1	1	6	**0.9925**	**0.9923**	**0**	**1.9848**	**0.9924**	**0.9922**	**1.9847**	20941.3747
1	0.1	250	0.001	0.1	1	6	**0.9925**	0.9915	**0**	1.9839	0.9924	0.9914	1.9839	**24159.8708**
1	0.1	200	0.001	0.001	1	6	0.8273	0.6248	0.0002	1.4521	0.8272	0.6246	1.4519	3400.5363
1	0.1	200	0.001	0.01	1	6	0.8711	0.6438	0.0005	1.5149	0.8706	0.6433	1.5144	1336.0173
1	0.1	200	0.001	0.1	1	6	**0.9925**	**0.9923**	**0**	**1.9848**	**0.9924**	**0.9922**	**1.9847**	**20941.3747**
1	0.1	200	0.001	1	1	6	0.9924	0.9537	0.0001	1.9462	**0.9924**	0.9537	1.9461	13576.7763
1	0.1	200	0.001	10	1	6	**0.9925**	0.9903	0.0001	1.9828	**0.9924**	0.9903	1.9828	17776.5979
1	0.1	200	0.001	100	1	6	**0.9925**	0.9841	0.0001	1.9766	**0.9924**	0.984	1.9765	11057.9367
1	0.1	200	0.0001	0.1	1	6	**0.9925**	0.9843	0.0001	1.9768	**0.9924**	0.9842	1.9767	9278.2228
1	0.1	200	0.001	0.1	1	6	**0.9925**	**0.9923**	**0**	**1.9848**	**0.9924**	**0.9922**	**1.9847**	**20941.3747**
1	0.1	200	0.01	0.1	1	6	**0.9925**	0.983	0.0001	1.9755	**0.9924**	0.9829	1.9754	10596.4776
1	0.1	200	0.1	0.1	1	6	**0.9925**	0.983	0.0001	1.9755	**0.9924**	0.9829	1.9754	10596.4776
1	0.1	200	1	0.1	1	6	**0.9905**	0.9382	**0**	1.9287	0.9905	0.9381	1.9286	20452.9696

* Bold texts indicate the best results.

to 1 in all subsequent experiments.

2) Analysis of adjacent frame number l: In Section III-B, l adjacent frames are stacked to form a 3D spatial-temporal tensor. The parameter l is critical, as it determines the amount of temporal information incorporated into the tensor and directly affects both detection performance and computational efficiency. To investigate the impact of l, we evaluate the proposed method with $l \in \{3, 4, 5, 6, 7, 8\}$. The corresponding results are reported in Table III. Based on the overall performance, l is set to 6 in all subsequent experiments.

3) Analysis of the strength of multi-topological guidance γ: γ determines the influence of the multi-topological priors. To investigate its effect, we test the proposed method with $\gamma \in \{1, 50, 100, 150, 200, 250\}$. The results are summarized in Table III. Considering both detection accuracy and background suppression, γ is set to 200 for all subsequent experiments.

4) Analysis of hyperparameters setting of spatial-temporal tensor α, β, μ, and δ: The parameters α, β and δ are empirically selected from the range $\{0.001, 0.01, 0.1, 1, 10, 100\}$. μ is chosen from $\{0.0001, 0.001, 0.01, 0.1, 1\}$. According to the results reported in Table III, α, β, δ are set in the range $\{0.001, 0.01, 0.1, 1, 10, 100\}$, and μ is set in the range $\{0.0001, 0.001, 0.01, 0.1, 1\}$. From Table III, β and δ are set equal to 0.1, α and μ are set to 1 and 0.001, respectively. These settings provide a good trade-off between background suppression, target preservation, and convergence stability.

C. Experiments Comparison

To further demonstrate the capability of the proposed TP-STTA in simultaneously enhancing targets and suppressing noise, a randomly selected frame from each sequence is presented in Fig. 6, together with the corresponding filtering results and 3D surface representations. For improved visualization, the highlighted regions are magnified and shown as close-up views in the left-bottom or right-bottom part.

For Seq. 1, NTFRA and SDD suffer from severe target blurring, resulting in low target-background contrast and only faintly discernible target contours. Although GCM, PTCTV-KMC, GLLS-LMAG, ECA-STT, ASTTV, and 4DTT achieve a certain degree of target enhancement,

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2026.3684452

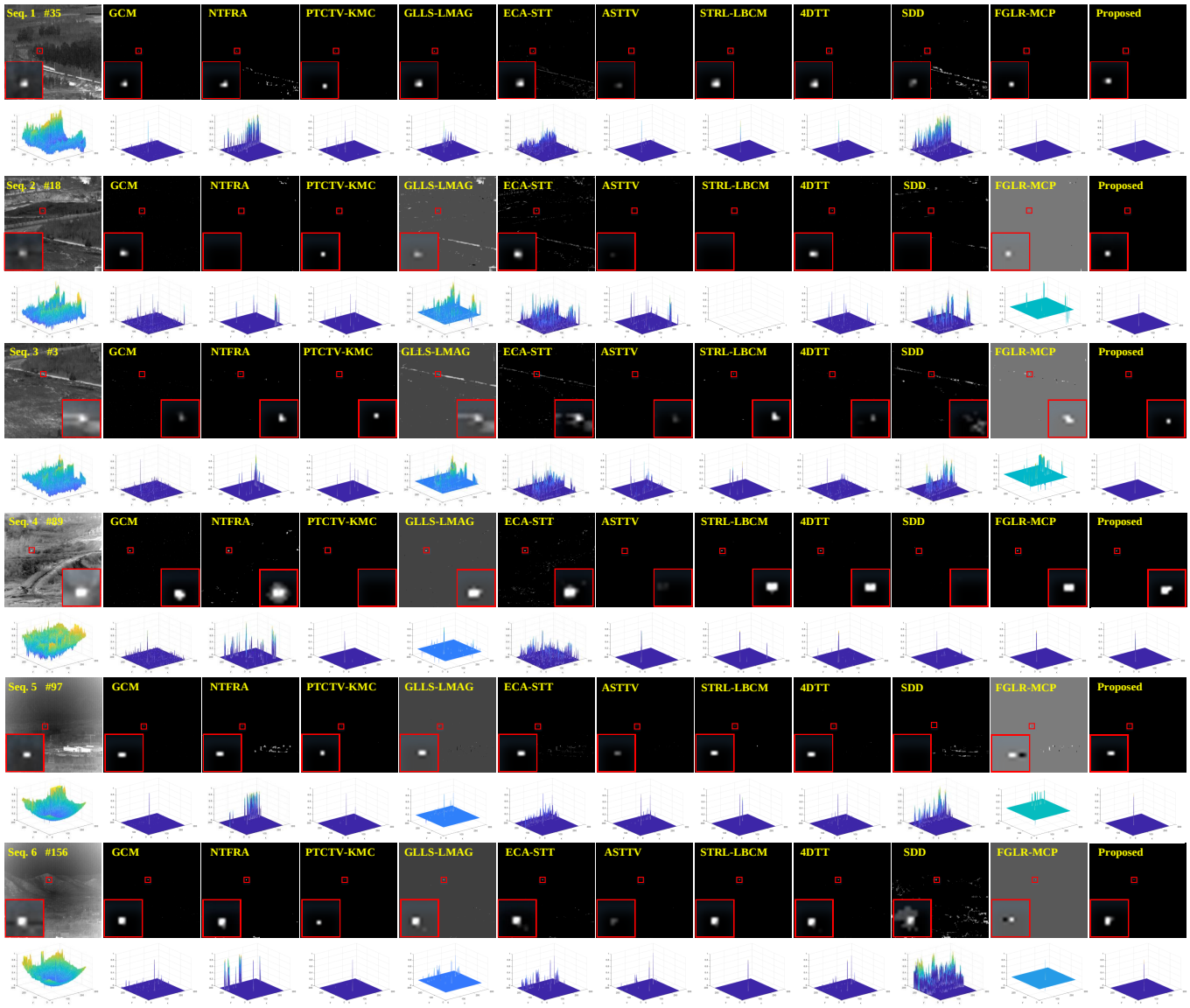


Fig. 6. Target enhancement results obtained by different methods using Seqs. 1-6. For better visualization, the demarcated area is enlarged, a close-up version is shown in the left/right-bottom part.

residual background interference still persists. It is worth noting that STRL-LBCM, FGLR-MCP, and the proposed method produce sharp and uniformly enhanced targets without background redundancy, enabling clear target-background separation. For Seq. 2, several methods, including NTFRA, ASTTV, STRL-LBCM, and SDD, fail to detect the target. Moreover, methods such as GCM, NTFRA, GLLS-LMAG, ECA-STT, 4DTT, SDD, and FGLR-MCP excessively enhance background responses, causing the target to be overwhelmed. Although PTCTV-KMC enhances the target to some extent, residual noise remains in the background. In contrast, the proposed method achieves satisfactory detection results. For Seq. 3, several methods mistakenly enhance flickering noise as targets, resulting in noise responses that are stronger than the true target, such as GCM, ECA-STT, ASTTV, and 4DTT. In addition, some methods exhibit limited background suppression capability, including NTFRA, GLLS-LMAG, SDD, and FGLR-MCP. Noticeable clutter also remains in the results of PTCTV-KMC and STRL-LBCM. Remarkably, the proposed method effectively enhances the target while leaving no residual clutter.

For Seq. 4, PTCTV-KMC and SDD mistakenly enhance flickering noise as targets, causing the true target to be overwhelmed. A considerable amount of residual noise remains in the results of GCM, NTFRA, GLLS-LMAG, and ECA-STT, while sporadic noise artifacts persist in STRL-LBCM and 4DTT. In contrast, both

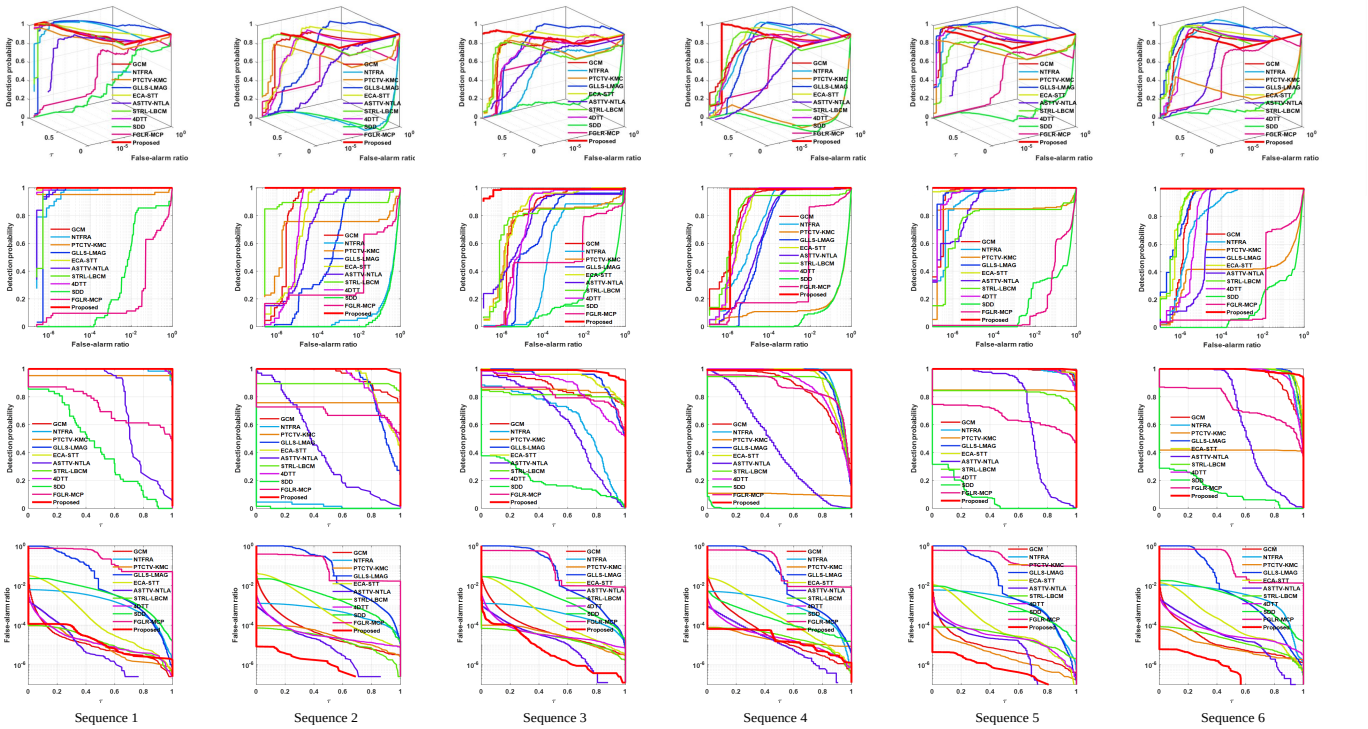


Fig. 7. ROC curves achieved by different methods on the six sequences. For better visualization, close-up versions are shown in blue rectangles in the bottom part.

FGLR-MCP and the proposed method achieve satisfactory detection performance. For Seq. 5, flickering noise continues to hinder accurate target detection for several methods. Specifically, GLLS-LMAG, ECA-STT, ASTTV, STRL-LBCM, and 4DTT suffer from noticeable noise interference, which degrades their detection performance. Meanwhile, residual noise is still present in the outputs of NTFRA, SDD, FGLR-MCP, and PTCTV-KMC. In comparison,, both GCM and the proposed method achieve noise-free target enhancement within the region of interest. Similarly, in the subsequent sequence, GLLS-LMAG, ECA-STT, ASTTV, and 4DTT exhibit stronger enhancement of flickering noise than of the true target. NTFRA and SDD retain a substantial amount of background noise, while FGLR-MCP fails to provide sufficient target-background contrast for clear target highlighting. GCM and PTCTV-KMC still suffer from minor residual noise. Among all the evaluated methods, only STRL-LBCM and the proposed method achieve effective target enhancement without any residual noise.

In addition, the 3D ROC metric is adopted to quantitatively assess the performance of the eleven methods from multiple perspectives, including detection probability, background suppression capability, and false alarm rate. Fig. 7 illustrates the 3D ROC curves together with their corresponding 2D ROC projections for Seqs. 1-6. Tables IV and V report the AUC values for Seqs.1-3 and Seqs.4-6 obtained from the 3D ROC analysis, together with the average runtime of each method. Bold, italic, and underlined entries indicate the best, second-best, and third-best results, respectively.

For Seq. 1, most comparison methods achieve high detection-related metrics under relatively simple background conditions, but several of them exhibit degraded robustness, as reflected by larger $\mathrm{AUC}_{(\mathrm{F},\tau)}$ values or lower $\mathrm{AUC}_{\mathrm{SNRP}}$. In contrast, the proposed TP-STTA consistently attains optimal or near-optimal performance across all metrics, including $\mathrm{AUC}_{\mathrm{TD}}$, $\mathrm{AUC}_{\mathrm{BS}}$, and $\mathrm{AUC}_{\mathrm{TD\text{-}BS}}$, while maintaining a minimal $\mathrm{AUC}_{(\mathrm{F},\tau)}$. The significantly higher $\mathrm{AUC}_{\mathrm{SNRP}}$ further demonstrates its superior target enhancement and background suppression capability. In the more challenging Seq. 2, several methods suffer from severe performance degradation, particularly in $\mathrm{AUC}_{(\mathrm{D},\tau)}$ and $\mathrm{AUC}_{\mathrm{TD}}$, indicating limited robustness to complex background interference. By comparison, the proposed TP-STTA maintains stable and superior performance across all evaluation metrics, achieving optimal detection accuracy, strong background suppression, and zero $\mathrm{AUC}_{(\mathrm{F},\tau)}$. Its leading $\mathrm{AUC}_{\mathrm{SNRP}}$ highlights its robustness under low signal-to-noise conditions. For Seq. 3, some competing methods show partial improvements in false alarm control but sacrifice background suppression

TABLE IV

QUANTITATIVE COMPARISON OF THE TESTED METHODS FOR THE SEQUENCES 1–3.

	Metrics	$AUC_{(D,F)}$	$AUC_{(D,\tau)}$	$AUC_{(F,\tau)}$	AUC_{TD}	AUC_{BS}	AUC_{TD-BS}	AUC_{ODP}	AUC_{SNRP}	Time(s)
	GCM	**1.0000**	0.9998	0.0003	1.9998	0.9997	*0.9996*	*1.9996*	3897.5955	0.0045
	NTFRA	**1.0000**	0.9959	0.0028	1.9959	0.9972	0.9931	1.9931	360.9197	3.4231
	PTCTV-KMC	*0.9750*	0.9516	**0.0000**	1.9266	0.9749	0.9516	1.9266	**34816.6590**	2.6565
	GLLS-LMAG	**1.0000**	0.9987	0.2400	1.9987	0.7600	0.7587	1.7587	4.1609	6.3752
	ECA-STT	**1.0000**	*0.9999*	0.0045	*1.9999*	0.9955	*0.9954*	*1.9954*	221.6535	6.5477
Seq.1	ASTTV-NTLA	**1.0000**	0.7511	*0.0001*	1.7511	*0.9999*	0.7509	1.7509	5507.5723	3.8891
	STRL-LBCM	**1.0000**	0.9987	**0.0000**	1.9987	**1.0000**	0.9987	1.9987	*30286.7591*	1.2492
	4DTT	**1.0000**	0.9987	*0.0001*	1.9987	*0.9999*	0.9986	1.9986	10200.4139	1.0678
	SDD	0.8792	0.4335	0.0083	1.3127	0.8709	0.4252	1.3044	51.9551	5.6897
	FGLR-MCP	0.7586	0.7299	0.3902	1.4885	0.3684	0.3397	1.0983	1.8706	0.0687
	Proposed	**1.0000**	**1.0000**	**0.0000**	**2.0000**	**1.0000**	**1.0000**	**2.0000**	23350.0218	4.2275
	GCM	**1.0000**	*0.9501*	0.0007	*1.9501*	0.9993	*0.9494*	*1.9493*	1344.4566	0.0046
	NTFRA	0.5165	0.0198	0.0006	0.5363	0.5159	0.0191	0.5356	30.9157	2.9905
	PTCTV-KMC	0.8676	0.7576	**0.0000**	1.6251	0.8675	0.7575	1.6251	16636.3857	3.2023
	GLLS-LMAG	*0.9992*	0.8962	0.3725	1.8954	0.6267	0.5238	1.5229	2.4062	3.5220
	ECA-STT	**1.0000**	0.9211	0.0068	1.9211	0.9932	0.9143	1.9143	135.8926	3.8992
Seq.2	ASTTV-NTLA	0.9930	0.4231	*0.0001*	1.4162	0.9929	0.4231	1.4161	5430.5709	2.5864
	STRL-LBCM	0.9644	0.8917	**0.0000**	1.8561	0.9644	0.8917	1.8561	*29368.8258*	0.7105
	4DTT	**1.0000**	0.9131	*0.0001*	1.9131	*0.9999*	0.9129	1.9129	7182.0053	0.8698
	SDD	0.4937	0.0015	0.0065	0.4952	0.4872	-0.0050	0.4887	0.2287	4.3238
	FGLR-MCP	0.7890	0.6874	0.1769	1.4765	0.6122	0.5106	1.2996	3.8870	0.0887
	Proposed	**1.0000**	**0.9983**	**0.0000**	**1.9983**	**1.0000**	**0.9983**	**1.9983**	**455407.6392**	4.5055
	GCM	*0.9940*	0.8753	0.0007	1.8693	*0.9933*	0.8746	1.8686	1258.8375	0.8277
	NTFRA	0.9044	0.6488	0.0006	1.5533	0.9038	0.6482	1.5527	1113.8316	4.8823
	PTCTV-KMC	0.9388	0.8424	**0.0000**	1.7812	0.9387	0.8424	1.7812	18107.3134	2.7311
	GLLS-LMAG	0.9709	0.9341	0.3829	1.9050	0.5879	0.5512	1.5221	2.4395	3.3220
	ECA-STT	**0.9996**	*0.9392*	0.0042	*1.9388*	**0.9954**	*0.9351*	*1.9347*	226.1394	6.2852
Seq.3	ASTTV-NTLA	0.9637	0.6199	*0.0001*	1.5836	0.9636	0.6198	1.5835	6531.8247	6.7929
	STRL-LBCM	0.8917	0.8138	**0.0000**	1.7055	0.8916	0.8138	1.7055	**25750.0709**	1.8350
	4DTT	0.9884	0.8627	*0.0001*	1.8511	0.9883	0.8626	1.8510	6491.3199	1.8693
	SDD	0.6813	0.2019	0.0080	0.8831	0.6733	0.1939	0.8751	25.3225	7.8319
	FGLR-MCP	0.8660	0.8135	0.2763	1.6795	0.5897	0.5373	1.4033	2.9448	0.0668
	Proposed	0.9925	**0.9923**	**0.0000**	**1.9848**	0.9924	**0.9922**	**1.9847**	*20941.3747*	4.3192

* Bold, italics, and underlined texts indicate the best results, the second-best results, and the third-best results, respectively.

TABLE V

QUANTITATIVE COMPARISON OF THE TESTED METHODS FOR THE SEQUENCES 4–6.

	Metrics	$AUC_{(D,F)}$	$AUC_{(D,\tau)}$	$AUC_{(F,\tau)}$	AUC_{TD}	AUC_{BS}	AUC_{TD-BS}	AUC_{ODP}	AUC_{SNRP}	Time(s)
	GCM	0.9954	0.8568	0.0009	1.8522	0.9945	0.8559	1.8513	977.8679	0.0062
	NTFRA	**1.0000**	*0.9448*	0.0020	*1.9447*	0.9979	*0.9427*	*1.9427*	463.0336	3.0377
	PTCTV-KMC	0.5206	0.1025	**0.0000**	0.6231	0.5206	0.1024	0.6230	3720.0668	0.6184
	GLLS-LMAG	*0.9999*	0.9330	0.3452	1.9329	0.6547	0.5878	1.5877	2.7029	4.0490
	ECA-STT	**1.0000**	0.9189	0.0040	1.9189	0.9960	0.9150	1.9149	230.7072	3.5630
Seq.4	ASTTV-NTLA	0.9972	0.4456	*0.0001*	1.4428	0.9971	0.4455	1.4427	4670.9118	4.6960
	STRL-LBCM	0.9684	0.8967	**0.0000**	1.8650	0.9683	0.8966	1.8650	*24644.8312*	1.0835
	4DTT	**1.0000**	0.8830	*0.0001*	1.8830	**0.9999**	0.8829	1.8829	8818.4968	0.9746
	SDD	0.5232	0.0025	0.0006	0.5257	0.5226	0.0020	0.5251	4.4868	3.8354
	FGLR-MCP	0.9299	0.8631	0.2880	1.7931	0.6419	0.5751	1.5050	2.9965	0.0458
	Proposed	0.9987	**0.9917**	**0.0000**	**1.9904**	*0.9987*	**0.9916**	**1.9903**	**30639.8435**	4.5845
	GCM	**1.0000**	0.9937	*0.0001*	1.9937	*0.9999*	0.9936	*1.9936*	8663.1091	0.0085
	NTFRA	**1.0000**	0.9898	0.0028	1.9897	0.9972	0.9870	1.9869	352.2636	3.0789
	PTCTV-KMC	*0.9071*	0.8480	**0.0000**	1.7551	0.9071	0.8480	1.7551	*98825.8745*	0.6658
	GLLS-LMAG	**1.0000**	0.9957	0.2842	1.9957	0.7158	0.7115	1.7115	3.5039	4.8846
	ECA-STT	**1.0000**	*0.9958*	0.0014	*1.9958*	0.9986	*0.9944*	*1.9944*	725.5002	3.5437
Seq.5	ASTTV-NTLA	**1.0000**	0.7035	**0.0000**	1.7035	*0.9999*	0.7035	1.7035	14995.0483	2.6248
	STRL-LBCM	0.9016	0.8307	**0.0000**	1.7323	0.9016	0.8307	1.7323	21299.2260	0.7130
	4DTT	**1.0000**	0.9912	*0.0001*	1.9912	*0.9999*	0.9911	1.9911	9101.0196	1.8404
	SDD	0.6261	0.0786	0.0026	0.7047	0.6235	0.0761	0.7022	30.8101	4.1850
	FGLR-MCP	0.7104	0.6610	0.3322	1.3714	0.3782	0.3287	1.0391	1.9895	0.0941
	Proposed	**1.0000**	**0.9996**	**0.0000**	**1.9996**	**1.0000**	**0.9996**	**1.9996**	**765102.2775**	4.7511
	GCM	**1.0000**	0.9519	0.0002	1.9519	*0.9998*	0.9517	1.9517	4444.1895	0.0030
	NTFRA	**1.0000**	0.9715	0.0052	1.9714	0.9948	0.9663	1.9662	186.5649	1.1473
	PTCTV-KMC	0.7084	0.4173	**0.0000**	1.1257	0.7084	0.4173	1.1257	29222.5508	0.6288
	GLLS-LMAG	**1.0000**	0.9825	0.2754	1.9825	0.7246	0.7070	1.7070	3.5670	4.8220
	ECA-STT	**1.0000**	*0.9853*	0.0019	*1.9853*	0.9981	*0.9834*	*1.9834*	514.7555	3.7090
Seq.6	ASTTV-NTLA	**1.0000**	0.6377	0.0002	1.6377	*0.9998*	0.6375	1.6375	3235.0951	5.0281
	STRL-LBCM	**1.0000**	0.9793	**0.0000**	1.9793	**1.0000**	0.9792	1.9792	*33799.7119*	0.8296
	4DTT	**1.0000**	0.9649	*0.0001*	1.9649	*0.9998*	0.9647	1.9647	6558.6621	1.2662
	SDD	0.6115	0.1256	0.0058	0.7372	0.6058	0.1198	0.7314	21.7484	4.2712
	FGLR-MCP	*0.8185*	0.7451	0.3316	1.5636	0.4869	0.4135	1.2320	2.2471	0.0472
	Proposed	**1.0000**	**0.9899**	**0.0000**	**1.9899**	**1.0000**	**0.9899**	**1.9899**	**562936.6702**	4.9178

* Bold, italics, and underlined texts indicate the best results, the second-best results, and the third-best results, respectively.

or detection stability. In contrast, the proposed TP-STTA achieves a well-balanced performance, delivering optimal or near-optimal results in $AUC_{(D,F)}$, $AUC_{(D,\tau)}$, AUC_{TD}, and $AUC_{TD\text{-}BS}$, while maintaining the minimum $AUC_{(F,\tau)}$. These results indicate strong robustness against threshold variations and clutter interference.

In Seq. 4, although several comparison methods perform reasonably well on individual metrics, their overall robustness is limited, especially under strong noise interference. The proposed TP-STTA consistently outperforms all competitors across nearly all metrics, achieving superior detection accuracy, background suppression, and the highest AUC_{SNRP}, which confirms its effectiveness in challenging scenarios. For Seq. 5, many existing methods exhibit reduced background suppression capability, leading to compromised overall performance despite acceptable detection accuracy. In contrast, the proposed TP-STTA achieves the best or joint-best results across almost all metrics, including AUC_{TD}, AUC_{BS}, and $AUC_{TD\text{-}BS}$, while maintaining the minimum $AUC_{(F,\tau)}$. Its leading AUC_{SNRP} demonstrates strong robustness in low-contrast environments. As for Seq. 6, several methods show competitive performance on isolated metrics but lack overall stability. The proposed TP-STTA consistently achieves the best comprehensive performance, with optimal detection accuracy, strong background suppression, minimal $AUC_{(F,\tau)}$, and the highest AUC_{SNRP}. These results further verify the robustness and generalization ability of the proposed TP-STTA across diverse scenarios.

At last, computational efficiency is compared. The results are listed in Tables IV and V, based on one randomly chosen image in Seqs. 1-6. From the runtime perspective, the methods exhibit noticeable differences in computational efficiency. Several methods such as GCM, SDD, and FGLR-MCP consistently achieve very low runtimes across most sequences, indicating high computational efficiency. However, as shown by the other evaluation metrics, these methods generally suffer from inferior detection performance and limited robustness, particularly under complex background conditions. Methods including PTCTV-KMC, STRL-LBCM, and 4DTT present moderate runtimes, but their computational costs vary significantly across different sequences, suggesting a strong sensitivity to data characteristics. Although they may achieve competitive performance on certain metrics or specific sequences, their overall stability and consistency remain limited. The proposed TP-STTA exhibits a moderate runtime level: it is not the fastest among all compared methods, yet it avoids excessive computational overhead and demonstrates stable execution times across all six sequences. More importantly, the proposed TP-STTA consistently achieves the best or near-best performance across almost all evaluation metrics.

Overall, although the proposed TP-STTA does not prioritize minimal runtime, it achieves a favorable trade-off between computational efficiency and detection performance, making it the most satisfactory method when all evaluation metrics are jointly considered.

V. Conclusions

In this paper, a novel TP-STTA model is proposed for infrared small-target detection. The TP-STTA framework consists of two complementary parts: MTFE and MPTA. To the best of our knowledge, this is the first attempt to introduce multi-topological features into infrared small-target detection, including area, diagonal, and volume attributes, which allows for the extraction of comprehensive features. On the one hand, MTFE effectively distinguishes small targets from complex backgrounds and flickering noise, and its output can be directly used as an independent detection result. On the other hand, the extracted multi-topological features serve as informative priors for the MPTA model, guiding it to further localize and confirm true targets. As for MPTA, it jointly exploits the spatial and temporal dimensions of the tensor by incorporating multiple physically meaningful priors, including target-aware priors from MTFE and background-oriented smoothness and low-rank priors, within an integrated optimization framework to ensure stable detection performance. Finally, the outputs of MTFE and MPTA are further fused, and temporal trajectory information is exploited to suppress false alarms and enhance temporal consistency, thereby ensuring reliable detection of infrared small targets. Experimental results on six real infrared sequences demonstrate that TP-STTA not only successfully detects small targets but also significantly enhances target-background contrast while effectively suppressing flickering noise and background clutter.

The proposed TP-STTA also has some limitations. TP-STTA incurs moderate computational overhead due to its multi-term optimization and iterative scheme. Its performance is also influenced by the selection of hyperparameters. These aspects will be further optimized in future work.

References

[1] M. Zhao, W. Li, L. Li, J. Hu, P. Ma, and R. Tao, "Single-frame infrared small-target detection: A survey," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 2, pp. 87–119, 2022.

[2] J. Han, S. Moradi, B. Zhou, W. Wang, Q. Zhao, and Z. Luo, "A true global contrast method for IR small target detection under complex background," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, no. 5004424, 2025.

[3] Y. Shi, Y. Lin, P. Wei, X. Xian, T. Chen, and L. Lin, "Diff-Mosaic: Augmenting realistic representations in infrared small target detection via diffusion prior," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, no. 5004311, 2024.

[4] Y. Xiao, Q. Yuan, K. Jiang, J. He, X. Jin, and L. Zhang, "DiffS-R: An efficient diffusion probabilistic model for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, no. 5601514, 2024.

[5] Z. Liu, Y. Shang, T. Li, G. Chen, Y. Wang, Q. Hu, and P. Zhu, "Robust multi-drone multi-target tracking to resolve target occlusion: A benchmark," *IEEE Transactions on Multimedia*, vol. 25, pp. 1462–1476, 2023.

[6] R. Kou, C. Wang, Z. Peng, Z. Zhao, Y. Chen, J. Han, F. Huang, Y. Yu, and Q. Fu, "Infrared small target segmentation networks: A survey," *Pattern Recognition*, vol. 143, no. 109788, 2023.

[7] J. Yin, H. Li, Y. Zheng, G. Gao, Y. Hu, and R. Tao, "S-patialctemporal weighted and regularized tensor model for infrared dim and small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, no. 5005317, 2024.

[8] H. Sun, J. Bai, F. Yang, and X. Bai, "Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, no. 5000513, 2023.

[9] Y. Luo, X. Li, Y. Yan, and C. Xia, "Spatial-temporal tensor representation learning with priors for infrared small target detection," *EEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 6, pp. 9598–9620, 2023.

[10] Y. Zhang, C. Yang, and X. Xiang, "IR small target detection based on morphological top-hat filter," *Computer Measurement and Control*, vol. 19, no. 6, pp. 1269–1272, 2011.

[11] R. Venkateswarlu, "Max-mean and max-median filters for detection of small targets," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 3809, pp. 74–83, 1999.

[12] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Tang, "A local contrast method for small infrared target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 574–581, 2013.

[13] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 12, pp. 2168–2172, 2014.

[14] Q. Yao and B. Li, "Effective infrared small target detection utilizing a novel local contrast method," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1890–1894, 2016.

[15] J. Han, K. Liang, B. Zhou, X. Zhu, J. Zhao, and L. Zhao, "Infrared small target detection utilizing the multiscale relative local contrast measure," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 4, pp. 612–616, 2018.

[16] Y. Shi, Y. Wei, H. Yao, D. Pan, and G. Xiao, "High-boost-based multiscale local contrast measure for infrared small target detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, p. 33C37, 2018.

[17] K. X. Y. Tang and C. Wang, "Fast infrared small target detection based on global contrast measure using dilate operation," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, no. 8000105, 2023.

[18] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4996–5009, 2013.

[19] Y. Dai, Y. Wu, and Y. Song, "Infrared small target and background separation via column-wise weighted robust principal component analysis," *Infrared Physics and Technology*, vol. 77, p. 421C430, 2016.

[20] J. Guo, Y. Wu, and Y. Dai, "Small target detection based on reweighted infrared patch-image model," *IET Image Processing*, vol. 12, no. 1, p. 70C79, 2017.

[21] Y. Dai and Y. Wu, "Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 8, p. 3752C3767, 2017.

[22] L. Zhang and Z. Peng, "Infrared small target detection based on partial sum of the tensor nuclear norm," *Remote Sensing*, vol. 11, no. 4, p. 382, 2019.

[23] S. C. C. L. X. Kong, C. Yang and Z. Peng, "Infrared small target detection via nonconvex tensor fibered rank approximation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 60, no. 5000321.

[24] Z. Huang, E. Zhao, W. Zheng, Y. Wen, X. Peng, W. Niu, and Z. Yang, "PTCTV-KMC: Infrared small target detection using joint partial tensor correlated total variation and k-means clustering," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 1252–1271, 2025.

[25] M. Zhang, R. Zhang, Y. Yang, H. Bai, J. Zhang, and J. Guo, "ISNet: Shape matters for infrared small target detection," in *Proceeding IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, p. 877C886.

[26] 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 Transactions on Image Processing*, vol. 32, pp. 1745–1758, 2023.

[27] M. Qi, L. Liu, S. Zhuang, Y. Liu, K. Li, Y. Yang, and X. Li, "FTC-Net: Fusion of transformer and CNN features for infrared small target detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 8613–8623, 2022.

[28] J. Lin, S. Li, L. Zhang, X. Yang, B. Yan, and Z. Meng, "IR-TransDet: Infrared dim and small target detection with IR-transformer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, no. 5004813, 2023.

[29] T. Hu, L. Wu, W. Dong, P. Wu, J. Sun, X. Xu, Q. Yan, and Y. Zhang, "Boosting HDR image reconstruction via semantic knowledge transfer," *IEEE Transactions on Image Processing*, vol. 35, pp. 1910 – 1922, 2026.

[30] Q. Yan, T. Hu, P. Wu, D. Dai, S. Gu, W. Dong, and Y. Zhang, "Efficient image enhancement with a diffusion-based frequency prior," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 9, pp. 8452 – 8465, 2025.

[31] X. Ma, K. Shi, D. Chen, Y. Cao, Q. Yan, and Y. Zhang, "A unified framework for deblurring and HDR imaging in dynamic scenes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 19, no. 8, pp. 1579–1588, 2025.

[32] Q. Yan, L. Zhang, Y. Liu, Y. Zhu, J. Sun, Q. Shi, and Y. Zhang, "Deep HDR imaging via a non-local network," *IEEE Transactions on Image Processing*, vol. 29, pp. 4308–4322, 2020.

[33] Z. Wang, C. Mo, and H. Dai, "Target detection method before tracking based on particle filter," *Procedia Computer Science*, vol. 107, 2017.

[34] B. Porat and B. Friedlander, "A frequency domain algorithm for multiframe detection and estimation of dim targets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 4, pp. 398 – 401, 1990.

[35] M. Li, T. Zhang, W. Yang, and X. Sun, "Moving weak point target detection and estimation with threedimensional double directional filter in IR cluttered background," *Optical Engineering*, vol. 44, no. 10, p. 107007, 2005.

[36] S. Kim, S. Sun, and K. Kim, "Highly efficient supersonic small infrared target detection using temporal contrast filter," *Electronics Letters*, vol. 50, no. 2, pp. 81–83, 2014.

[37] J. Gao, Z. Lin, and W. An, "Infrared small target detection using a temporal variance and spatial patch contrast filter," *IEEE Access*, vol. 7, pp. 32 217–32 226, 2019.

[38] L. Deng, H. Zhu, C. Tao, and Y. Wei, "Infrared moving point target detection based on spatial-temporal local contrast filter," *Infrared Physics and Technology*, vol. 76, pp. 168–173, 2016.

[39] Y. Li, Y. Zhang, J. Yu, Y. Tan, J. Tian, and J. Ma, "A novel spatio-temporal saliency approach for robust dim moving target detection from airborne infrared image sequences," *Information Sciences*, vol. 369, pp. 548–563, 2016.

[40] P. Du and A. Hamdulla, "Infrared moving small-target detection using spatial-temporal local difference measure," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 10, pp. 1817–1821, 2019.

[41] J. Li, X. Hou, and Y. Zeng, "Infrared small target detection using the global low-rank and local smoothness coupled representation with local structure," *Neurocomputing*, p. 130546, 2025.

[42] 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 Transactions on Geoscience and Remote Sensing*, vol. 59, no. 12, pp. 10 708–10 724, 2021.

[43] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2021.

[44] Y. Luo, X. Li, Y. Yan, and C. Xia, "Spatial-temporal tensor representation learning with priors for infrared small target detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 6, pp. 9598–9620, 2023.

[45] F. Wu, H. Yu, A. Liu, J. Luo, and Z. Peng, "Infrared small target detection using spatiotemporal 4-d tensor train and ring unfolding," *IEEE transactions on geoscience and remote sensing*, vol. 61, pp. 1–22, 2023.

[46] F. Zhou, M. Fu, Y. Qian, J. Yang, and Y. Dai, "Sparse prior is not all you need: when differential directionality meets saliency coherence for infrared small target detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–18, 2024.

[47] T. Liu, Y. Liu, J. Yang, B. Li, Y. Wang, and W. An, "Graph Laplacian regularization for fast infrared small target detection," *Pattern Recognition*, vol. 158, p. 111077, 2025.

[48] X. Tong, Z. Zuo, S. Su, J. Wei, X. Sun, P. Wu, and Z. Zhao, "ST-trans: Spatial-temporal transformer for infrared small target detection in sequential images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, no. 5001819, 2024.

[49] P. Yan, R. Hou, X. Duan, C. Yue, X. Wang, and X. Cao, "STDMANet: Spatio-temporal differential multiscale attention network for small moving infrared target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, no. 5602516, 2023.

[50] R. Cui, N. Li, J. Liu, and H. Zhao, "MSTCNet: Toward generalization improving for multiframe infrared small target detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 8416–8437, 2025.

[51] P. Monasse and G. Frederic, "Fast computation of a contrast-invariant image representation," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 860–872, 2002.

[52] P. Ghamisi, R. Souza, J. A. Benediktsson, X. Xiao, and R. A. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 1–15, 2016.

[53] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3747–3762, 2010.

[54] M. Zhao, L. Li, W. Li, R. Tao, L. Li, and W. Zhang, "Infrared small-target detection based on multiple morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 6077–6091, 2021.

[55] T. Graud, E. Carlinet, S. Crozet, and L. Najman, "A quasi-linear algorithm to compute the tree of shapes of nD images," pp. 98–110, 2013.

[56] P. Monasse and F. Guichard, "Fast computation of a contrast-invariant image representation," *IEEE transactions on image processing*, vol. 9, no. 5, pp. 860–872, 2000.

[57] W. Li, Z. Wang, L. Li, and Q. Du, "Feature extraction for hyperspectral images using local contain profile," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 5035–5046, 2019.

[58] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, p. 2117C2130, 2013.

[59] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 5131–5153, 2021.

[60] C. Xia, S. Chen, R. Huang, J. Hu, and Z. Chen, "Separable spatial-temporal patch-tensor pair completion for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–20, 2024.