Sparse Prior Is Not All You Need: When Differential Directionality Meets Saliency Coherence for Infrared Small Target Detection

Fei Zhou, Maixia Fu, Yulei Qian, Jian Yang, Yimian Dai

Abstract—Infrared small target detection is crucial for the efficacy of infrared search and tracking systems. Current tensor decomposition methods emphasize representing small targets with sparsity but struggle to separate targets from complex backgrounds due to insufficient use of intrinsic directional information and reduced target visibility during decomposition. To address these challenges, this study introduces a *Sparse Differential Directionality prior (SDD)* framework. SDD leverages the distinct directional characteristics of targets to differentiate them from the background, applying mixed sparse constraints on the differential directional images and continuity difference matrix of the temporal component, both derived from Tucker decomposition. We further enhance target detectability with a saliency coherence strategy that intensifies target contrast against the background during hierarchical decomposition. A Proximal Alternating Minimization-based (PAM) algorithm efficiently solves our proposed model. Experimental results on several real-world datasets validate our method's effectiveness, outperforming ten state-of-the-art methods in target detection and clutter suppression. Our code is available at https://github.com/GrokCV/SDD.

Index Terms—Target detection, tensor decomposition, spatial-temporal regularization, saliency map, proximal optimization

I. INTRODUCTION

INFRARED small target detection is crucial for the early identification and localization of objects of interest [1], enhancing the effectiveness of subsequent search and tracking operations. This technique has a wide range of practical applications, including early warning systems [2], anti-missile

Fei Zhou and Maixia Fu are with Key Laboratory of Grain Information Processing and Control (Henan University of Technology), Ministry of Education; Henan Key Laboratory of Grain Photoelectric Detection and Control, Henan University of Technology, Zhengzhou, China. (e-mail: hellozf1990@163.com,fumaixia@126.com).

Yulei Qian is with Nanjing Marine Radar Institute, Nanjing, China (e-mail: yuleifly@126.com).

Jian Yang and Yimian Dai are with PCA Lab, Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, and Jiangsu Key Lab of Image and Video Understanding for Social Security, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. (e-mail: csjyang@mail.njust.edu.cn; yimian.dai@gmail.com).

defense mechanisms [3], marine intelligent defense [4], and automated defect detection [5], among others.

A. Prior Work on Infrared Small Target Detection

Despite the remarkable advancements in infrared small target detection, two significant challenges persist:

1) **Limited pixel representation**: Objects of interest often occupy only a few pixels in the scene, a consequence of the requirement for long-distance perception. The limited number of pixels fails to convey the intrinsic shape and texture characteristics needed for target identification. This paucity of information poses a significant challenge in accurately detecting and recognizing small targets.

2) **Salient background interferences**: Infrared scenes frequently contain salient background elements, such as thick clouds or sea clutters, which can obscure small targets. These interferences weaken the correlation between backgrounds, making it difficult to distinguish targets from their surroundings. Effectively suppressing background interferences while preserving target information remains a critical challenge in infrared small target detection.

These challenges continue to be the key factors impeding progress in downstream tasks and require solutions to enhance the performance of infrared small target detection.

In recent years, deep learning methods have made significant strides in small target detection [6], with notable works such as YOLO series [7, 8]. The advent of these methodologies have eclipsed the detection performance of traditional knowledge-driven approaches [6, 9–11]. However, these data-driven approaches are constrained by limitations such as dataset scarcity, difficulty in representing intrinsic patterns, and poor interpretability, leading to a bottleneck in performance enhancement. To address these issues, researchers have explored deep unfolding methods that incorporate domain knowledge into data-driven approaches, leveraging the advantages of feature representation inherent in both domain knowledge-driven and data-driven methods [12–15], for example RPCANet [13], ALCNet [12]. From this, we observe that precise delineation of domain-specific knowledge can significantly contribute to the augmentation of data-driven models. Therefore, this paper endeavors to further explore the development of models underpinned by such specialized domain knowledge.

Over the past decades, numerous domain knowledge-driven detection methods have been proposed. These techniques focus on extracting small targets from infrared backgrounds by
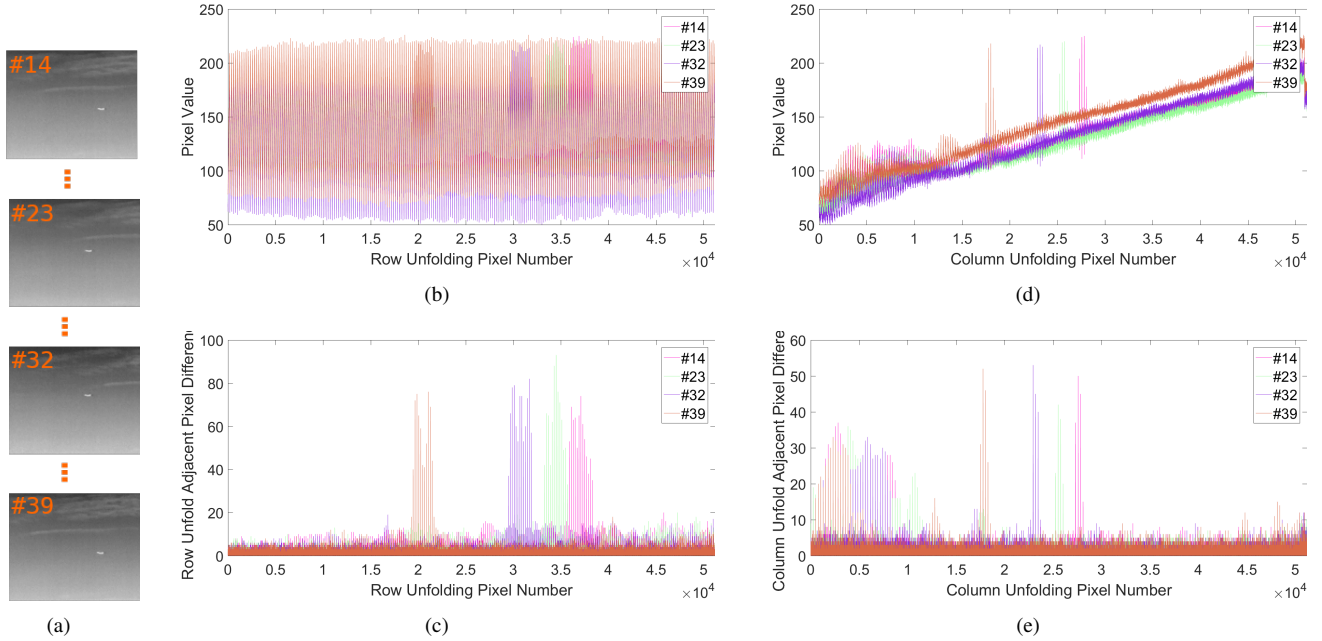
Fig. 1. Illustration of directional information of background and target in infrared sequences. (a) Infrared Sequence. (b) Unfolding the four scenes in the sequence horizontally. (c) Difference between adjacent pixels during horizontal unfolding. (d) Unfolding the four scenes in the sequence vertically. (e) Difference between adjacent pixels during vertical unfolding.

modeling target priors, such as sparsity [16], saliency [17, 18], or constructing background knowledge [19–24]. Despite their effectiveness, early low-rank methods struggled to distinguish targets from background interference [16, 25, 26] due to the presence of sparse distractors in the background. Consequently, improvement schemes have been proposed, with sequential detection offering an attractive alternative due to its ability to integrate inter-frame priors.

In sequence detection, methods enhance the low-rank nature of the background using different feature structures, such as spatio-temporal patch images [23, 27], tensor cube models [28–31], and connected multi-frame patch groups [32]. These methods apply tensor constraints [33, 34] to decompose the low-rank and sparse components representing small targets. Additionally, researchers have incorporated domain expert priors, such as tophat regularization [28], total variation [35, 36], and its variants [37, 38]. However, these methods are complex, computationally expensive, and lack stability in dynamic scenes due to the decrease in inter-frame correlation and non-local auto-correlation of the background over time.

From the above analysis, infrared small target detection methods are broadly classified into two groups: background characterization and target characterization methods. Background characterization methods, such as low-rank sparse modeling [39, 40], filtering [20], and transform-domain approaches [41], leverage the intrinsic geometric structures and mathematical statistical properties of infrared images for target-background separation. While effective in suppressing globally repetitive sparse components, these methods struggle with rare sparse components in dynamic scenes due to the similar sparsity between disturbances and targets, and the weakened interframe correlation from scene changes.

Target characterization methods, for instance, local contrast [17, 42], gradient vector fields [43, 44], and random walker [45], exploit the difference information of local pixels to highlight target saliency. However, their inability to fully utilize the background's structural information limits their effectiveness in extracting dim targets and suppressing repetitive sparse regions, especially when high-brightness background areas exhibit greater prominence than targets.

In summary, existing methods primarily consider either global structural priors or local directional differential priors, leading to performance limitations. A more efficient approach should merge the complementary strengths of both methods to better manage complex infrared scenes with both rare and repetitive sparse components.

B. Motivation

In this study, we have identified a latent prior that augments the effectiveness of low-rank plus sparse decomposition methods: sparse targets exhibit non-directionality, while background interference possesses inherent directionality. Background directionality pertains to the phenomenon wherein components of clutter and interference in the background exhibit varying sparse representations across different directions. In contrast, the non-directionality of the target implies that the target components maintain consistent sparse representations regardless of direction. Our approach utilizes the contrast between the non-directionality of targets and the directionality of backgrounds. By applying differential methods across various axes in infrared imagery, we accurately delineate the inherent directionality of the background compared to the non-directionality of targets, enabling effective target-background segregation. Different from previous works, our study is dedicated to incorporate the concept of directionality as a prior to distinguish targets

TABLE I
DIRECTIONAL VERIFICATION ON THE SIRST DATASET

Indicators	Homogeneous Region	Small Target Area	Clutter Edge
Variance range	$0 \sim 10$	$5 \sim 20$	≥ 20
Variance percentage	97.2%	96.3%	95.6%

TABLE II
SUMMARY OF MATHEMATICAL ANNOTATION

Notation	Interpretation
$x, \mathbf{x}, X, \mathcal{X}$	scalar, vector, matrix, N-dimensional tensor
$x_{i_1}, x_{i_2}, ..., x_{i_n}$	the $(x_{i_1}, x_{i_2}, ..., x_{i_n})$ element of $\mathcal{X}$
$x_{i:k}, x_{:jk}, x_{ij:}$	row, column and tube fibers of a 3-D tensor $\mathcal{X}$
$X_{:j:}, X_{i::}, X_{::k}$	lateral, horizontal, and frontal slides of a 3-D tensor $\mathcal{X}$
$X_{(n)}$	mode-n matricization of tensor $\mathcal{X}$, obtained by arrangingthe mode-n fibers as the columns of the resulting matrix of size $\mathbb{R}^{I_n \times \prod_{k \neq n} I_k}$
$\|\mathcal{X}\|_F$	Frobenius norm of tensor $\mathcal{X}$, defined as $\|\mathcal{X}\|_F = \sqrt{\sum_{i_1, i_2, ..., i_n} \|x_{i_1, i_2, ..., i_N}\|^2}$
$\|\mathcal{X}\|_1$	l_1 norm of tensor $\mathcal{X}$, defined as $\|\mathcal{X}\|_1 = \sum_{i_1, i_2, ..., i_n} \|x_{i_1, i_2, ..., i_N}\|$
$\langle \mathcal{X}_1, \mathcal{X}_2 \rangle$	the inner product for two tensors, defined as $\langle \mathcal{X}_1, \mathcal{X}_2 \rangle = \sum_{i_1, i_2, ..., i_n} x_{1 i_1, i_2, ..., i_N} x_{2 i_1, i_2, ..., i_N}$
$\otimes$	Kronecker product
$\odot$	component-wise multiplication
$\times_n$	mode-n tensor-matrix product

from the background in the low-rank sparse decomposition model.

As depicted in Fig. 1, small targets maintain a significant sparse distribution, regardless of the unfolding direction. Conversely, the background displays notable differences across different directions. The differential information between adjacent pixels in both unfolding directions further accentuates the anisotropy of the background and the isotropy of the targets, as shown in Fig. 1 (c) and (e).

To validate the non-directionality of the target and the directionality of background clutter interference, we applied a multi-scale local contrast strategy [18] on the public dataset SIRST [12]. A statistical analysis of the differential variance information in the horizontal, vertical, and two diagonal directions for the selected homogeneous background region, clutter edge region, and target region was conducted. Table I presents the differential variances of different regions in various directions, with smaller variance values indicating weaker directionality. The statistical results show that 95.6% of the dataset exhibits differential variances greater than 20 in the background clutter edge, 96.3% of the dataset shows target differential variances between 5 and 20, and 97.2% of the dataset has differential variances between 0 and 10 in homogeneous background regions.

In summary, the main contributions can be summarized as follows:

1) We introduce the SDD prior, which distinguishes sparse small targets from sparse background interference based on their differential directionality.
2) We propose an infrared small target detection method that incorporates the SDD prior into sparse and low-rank decomposition with mixed sparsity regularization, effectively exploiting the differential directionality information.
3) We design a saliency coherent map to mitigate the issues of brightness attenuation in small targets and weak saliency coherence among background clutters.
4) We develop an efficient PAM algorithm to solve the proposed model, which has improved detection efficiency by at least 50% compared to similar algorithms [28, 37, 38, 46].

II. RELATED WORK

A. Notations and Preliminaries

In this section, we introduce several notations related to the proposed method, as listed in Table II. Additionally, we provide definitions for the mode-n tensor-matrix product, the Tucker decomposition, and the Tucker rank. For a more comprehensive understanding of tensors, we recommend consulting [47].

Definition 1 (Mode-n tensor-matrix product): the n-mode (matrix) product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times \cdots \times I_N}$ with a

matrix $U \in \mathbb{R}^{J \times I_N}$ is defined as $\mathcal{X} \times_n U$ with size $I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$, which is written as:

$$(\mathcal{X} \times_n U)_{i_1...i_{n-1}ji_{n+1}...i_N} = \sum_{i_n} x_{i_1 i_2 ... i_N} u_{j i_n}$$

According to the mode-n fiber multiplication, it can also be rewritten as:

$$\mathcal{Y} = \mathcal{X} \times_n U \Leftrightarrow Y_{(n)} = U X_{(n)}$$

Definition 2 (Tucker decomposition and Tucker rank): Tucker decomposition is formulated as a high order principal component analysis. It decomposes a tensor into a core tensor multiplied by a matrix along each mode, for example in a 3-D tensor, as shown in Fig. 2. For a N-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times \cdots \times I_N}$, let $U^{(n)} \in \mathbb{R}^{J_n \times I_n}$ in any $n \in \{1, ..., N\}$, we obtain:

$$\mathcal{Y} = \mathcal{X} \times_1 U^{(1)} \times U^{(2)} \cdots \times_N U^{(N)}$$
$$\Leftrightarrow Y_{(n)} = U^{(n)} X_{(n)} \left(U^{(N)} \otimes \cdots \otimes U^{(1)} \right)^{\mathrm{T}}$$

And the tucker rank is defined as:

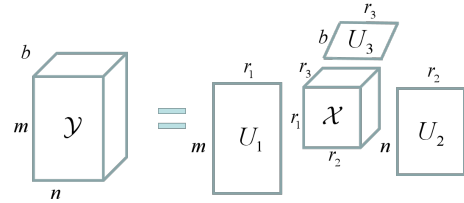$$Rank_N(\mathcal{X}) = (Rank(U_{(1)}), Rank(U_{(2)}), ..., Rank(U_{(N)}))$$

Fig. 2. Illustration of 3-D tensor Tucker decomposition.

B. Tensor Decomposition for Infrared Small Target Detection

For an infrared small target image, it can be mathematically expressed as a linear superposition model by target, background, and noise components:

$$f_Y = f_F + f_T + f_N \tag{1}$$

Numerous methods have been developed to distinguish targets from background elements, extensively reviewed in [48]. Our focus is on tensor decomposition-based methods, particularly highlighted by the Infrared Patch Tensor (IPT) model introduced in [21]. This model arranges overlapping image patches into a third-order tensor and achieves target-background separation by applying optimally regularized constraints where $J_1(\mathcal{F})$ and $J_2(\mathcal{T})$ are constraint functions representing background and target priors, respectively, balanced by λ and influenced by Gaussian noise density variance η. This model is formulated as:

$$\mathcal{Y} = \mathcal{F} + \mathcal{T} + \mathcal{N} \tag{2}$$

where $\mathcal{Y}, \mathcal{F}, \mathcal{T}, \mathcal{N}$ denotes the tensor model of infrared image, background component, target component and noise component, respectively. The purpose of target-background separation is achieved by introducing optimal regularized constraints based on the characteristics of different components. Then, the generalized form can be written as:

$$\min_{\mathcal{F}, \mathcal{T}} J_1(\mathcal{F}) + \lambda J_2(\mathcal{T}) \, s.t. \, \|\mathcal{Y} - \mathcal{F} - \mathcal{T}\|_F^2 \leq \eta \tag{3}$$

where $J_1(\mathcal{F})$ and $J_2(\mathcal{T})$ represent constraint functions used to encode the background prior and target prior, respectively. λ is a positive tradeoff between the two components, while η denotes the Gaussian noise density variance. Consequently, numerous studies have explored methods to precisely extract small targets by investigating regularization techniques. Among these, the sum of the nuclear norm and reweighted l_1 norm are employed to constrain the background and small target components in the IPT model [21]. However, this approach is susceptible to sparse interference, similar to small targets, due to the approximation error of the nuclear norm constraint.

To address this issue, improvements have been developed by introducing non-convex surrogates [49, 50] or incorporating well-designed local spatial priors [51, 52]. Spatial-temporal tensor models have also been constructed to exploit interframe information [29, 30], employing various regularizations to describe underlying sequential priors in the spatial [30, 31] and temporal domains [29, 46]. TV regularization and saliency regularization are designed to exploit spatial-temporal saliency information [46, 50, 53].

Despite advancements, these methods still face challenges in fully utilizing background information and effectively suppressing interference. Our proposed model introduces differential directionality and mixed sparse constraints for a more intrinsic understanding, with saliency coherence factors to better differentiate between background interference and small targets.

III. DETECTION MODEL

A. Differential Directionality Mixed Sparse Regularization

In dynamic scenes, background variations reduce correlation across frames, challenging models reliant on low-rank sparse priors for effective target separation [54, 55]. Analysis reveals a latent directionality in background components: homogeneous regions and targets show no directionality, while clutter exhibits

directionality. This directionality is effectively captured by first-order differential operators in horizontal and vertical directions [56–58]. Despite gradual background changes, strong continuity exists between frames. Thus, we propose a spatio-temporal differential prior with hybrid sparse constraints to address both directionality and continuity.

The low-rank component $\mathcal{F}$ is further decomposed into a mode-3 tensor-matrix product:

$$\mathcal{F} = \mathcal{B} \times_3 A \tag{4}$$

where $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times r}$ represents the spatial factor, and $A \in \mathbb{R}^{n_3 \times r}$ is the temporal factor with $A^{\mathrm{T}} A = I$. We construct spatial-temporal differential images to explore background priors, using first-order difference operators as shown in Fig. 3, which displays heatmaps of differential images for spatial factors. The residual distribution shows notable disparity in horizontal and vertical images, with unfolding matrices presented for clarity. Distinct sparse patterns, namely unshared and shared, are illustrated, indicating significant non-zero element distribution differences.

To address this issue, a mixed sparse constraint approach is employed, with the l_1 norm constraint applied to the spatial horizontal mode and the $l_{2,1}$ norm constraint utilized for the spatial vertical mode. This enables the full exploitation of the sparse prior of different patterns, enhancing the algorithm's robustness across varied dynamic conditions. Additionally, differential information is constructed for the temporal factor and constrained by the l_2 norm to capture local temporal coherence in dynamic scenes.

The regularization on the low-rank background component can be modeled as follows:

$$J_1(\mathcal{F}) = \|\mathcal{B} \times_1 D_1\|_{2,1} + \|\mathcal{B} \times_2 D_2\|_1 + \lambda \|D_3 A\|_2^2 \tag{5}$$

where λ is the tradeoff between different items; $D_k(k = 1, 2, 3)$ are first-order difference matrices in three dimensions of infrared sequence. And the differential operation of each term can be obtained:

$$\begin{aligned} \mathcal{B} \times_1 D_1 &= \mathcal{B}(i+1, j, t) - \mathcal{B}(i, j, t) \\ \mathcal{B} \times_2 D_2 &= \mathcal{B}(i, j+1, t) - \mathcal{B}(i, j, t) \\ D_3 A &= A(:, t+1) - A(:, t) \end{aligned} \tag{6}$$

To enhance the sparsity of the difference images, a reweighted strategy is introduced in the mixed sparse regularization. The regularization on the low-rank component is rewritten as follows:

$$\begin{aligned} J_1(\mathcal{F}) = \|\mathcal{B} \times_1 D_1\|_{2,1,\mathcal{W}_1} + \|\mathcal{B} \times_2 D_2\|_{1,\mathcal{W}_2} \\ + \lambda \|D_3 A\|_2^2 \end{aligned} \tag{7}$$

where $\mathcal{W}_1$ and $\mathcal{W}_2$ are weight tensors.

B. Adaptive Saliency Coherence Map

Infrared sensing often experiences interference from elements like cloud clutter, sea waves, and ground structures. These structures exhibit significant inter-frame changes and similar sparsity to the target, causing misclassification during separation. To mitigate these interferences, previous studies have designed clutter suppression factors [21, 29, 37, 46], but these methods can still miss targets or leave residual salient
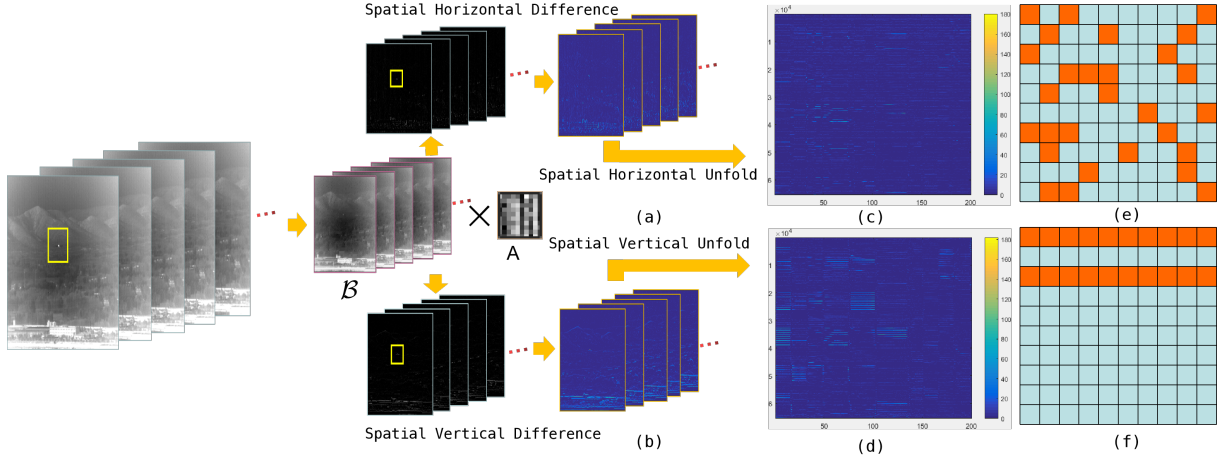
Fig. 3. The difference images along the temporal mode for infrared sequence data in both spatial modes. (a) The heatmaps of the spatial horizontal difference, (b) The heatmaps of the spatial vertical difference, (c) and (d) The matrices of corresponding difference images along the temporal mode, (e) and (f) Display of the shared and unshared sparsity.

points. This is primarily due to the limited capabilities of single-scale structure tensors in characterizing coherence between background components, and the overlooked decay of target energy during decomposition. To tackle this, we propose an adaptive saliency coherence exponent (ASCE) map, based on the pointwise structure tensor computed from the image.

Given an infrared image $Y \in \mathbb{R}^2$, the smoothed structure tensor (SST) [59] can be computed at every pixel:

$$K_\sigma\left(I\left(x,y\right)\right) = G_\sigma \star \left(\nabla I \nabla I^{\mathrm{T}}\right) = \begin{bmatrix} G_\sigma \star I_x^2 & G_\sigma \star I_x I_y \\ G_\sigma \star I_y I_x & G_\sigma \star I_y^2 \end{bmatrix} \quad (8)$$

where $G_\sigma = \left(\sigma\sqrt{2\pi}\right)^{-1} \exp\left(-|\mathrm{x}|^2/2\sigma^2\right)$ denotes the Gaussian kernel with the scale parameter $\sigma > 0$, and $\star$ is the 2D convolution operation. In general, K_σ contains a maximum-minimum eigenvalue pair $(e_+(x,y,\sigma), e_-(x,y,\sigma))$ and $e_+ \geq e_-$. The eigenvalues characterize the local background regions within a size σ neighborhood with the range of their values, specifically, indicating $e_+ \geq e_- \gg 0$ for corner region, $e_+ \gg e_- \approx 0$ for edge region, $e_+ \approx e_- \approx 0$ for the flat region. Based on this, the adaptive saliency coherence exponent function can be formulated as follows:

$$ASCE(x,y,\sigma) = 1 - \exp\left(-\sum_{i=1}^{3} \alpha_i C_i(e_+,e_-,\sigma)\right) \quad (9)$$

where $\alpha_i \geq 0 (i = 1,2,3)$ are the balance parameters. $\delta = 0.001$ is a smoothed parameter avoiding division by zero. The $ASCE(x,y,\sigma)$ represents the saliency coherence factor for portraying various background components at the pixel (x,y).

The first item is represented as:

$$C_1(e_+,e_-,\sigma) = \sqrt{|e_+e_-|} \quad (10)$$

The design is a corner measure [60, 61], used to capture the structure of non-target points with high spatial frequencies along two oriented directions, as shown in Fig. 4(b).

The second item is defined as:

$$C_2(e_+,e_-,\sigma) = \left|\frac{e_+ + e_-}{(e_+ - e_- + \delta)^{1/2}}\right| \quad (11)$$

It is deemed as coherence measure, where its value is close to one in edge pixels and close to zero in homogeneous or noisy regions, as presented in Fig. 4(c).

The third item is written as:

$$C_3(e_+,e_-,\sigma) = \frac{e_+e_-}{e_+ + e_- + \delta} \quad (12)$$

The item can be seen as edge indicator since its values encode the background components with strong gradient, such as the edge region tending to zero, the corner component tending to one, as presented in Fig. 4(d).

Furthermore, it can be observed that the $ASCE(x,y,\sigma)$ indicator ranges between 0 and 1. The greater the contrast of a background component, the closer its corresponding value approaches 1, while the values of other components gravitate toward 0. Generally, small targets exhibit higher local contrast areas, causing their corresponding $ASCE(x,y,\sigma)$ values to be nearer to 1 than those of other components, as illustrated in Fig. 4(e). During tensor decomposition, components with heightened contrast typically possess increased sparsity and can be more easily separated. However, if $ASCE(x,y,\sigma)$ is directly employed as the penalty factor, target saliency may diminish gradually throughout the decomposition iteration. Consequently, we propose a variant of $ASCE(x,y,\sigma)$:

$$W_{ASCE} = \begin{cases} 1 + ASCE(x,y,\sigma), & ASCE(x,y,\sigma) \geq T_S \\ ASCE(x,y,\sigma), & otherwise \end{cases} \quad (13)$$

where, T_s is a threshold defined as $mean(ASCE(x,y,\sigma)) + 5var(ASCE(x,y,\sigma))$.

The target enhancement factor of an infrared sequence cube can be obtained by stacking the improved $ASCE(x,y,\sigma)$ maps of each frame. We incorporate this factor and a reweighted strategy into the constraint of small target components:

$$\|\mathcal{W}_{ASCE} \odot \mathcal{T}\|_{1,\mathcal{W}_S} \quad (14)$$

where $\mathcal{W}_S$ denotes the weight of each element in the sparse component.

Finally, the proposed model is constructed by integrating the spatial-temporal difference prior with the saliency coherence
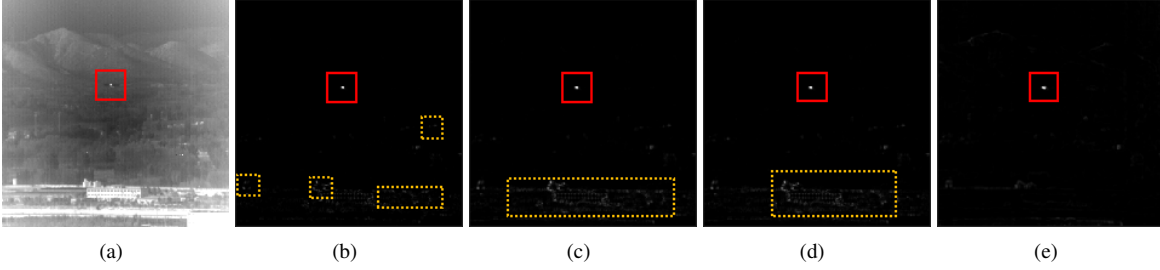
Fig. 4. Illustration of single and multiscale saliency coherence maps. (a) Original image (b)-(d) The saliency map obtained by the first to three saliency coherence items. (e) The saliency map obtained by the transformed enhancement factor.

exponent map, as follows:

$$\min_{A,\mathcal{B},\mathcal{T}} \|\mathcal{B}\times_1 D_1\|_{2,1,\mathcal{W}_1} + \|\mathcal{B}\times_2 D_2\|_{1,\mathcal{W}_2}$$
$$+ \lambda \|D_3 A\|_F^2 + \gamma \|\mathcal{W}_{ASCE} \odot \mathcal{T}\|_{1,\mathcal{W}_S} \quad (15)$$
$$s.t. \|\mathcal{Y} - \mathcal{B}\times_3 A - \mathcal{T}\|_F^2 \le \varepsilon$$

IV. MODEL SOLUTION BASED ON PAM

We solve the proposed model based on the PAM-based iterative algorithm by alternately updating each variable with the others fixed. First, the objective function is transformed into an equivalent form:

$$\min_{A,\mathcal{B},\mathcal{T}} \frac{1}{2}\|\mathcal{Y} - \mathcal{B}\times_3 A - \mathcal{T}\|_F^2 + \gamma\|\mathcal{W}_{ASCE} \odot T\|_{1,\mathcal{W}_S}$$
$$+ \lambda \|D_3 A\|_F^2 + (\|\mathcal{B}\times_1 D_1\|_{2,1,\mathcal{W}_1} + \|\mathcal{B}\times_2 D_2\|_{1,\mathcal{W}_2}) \quad (16)$$

By separating each variable, we can update them by minimizing the following sub-problems:

$$A^{k+1} = \arg\min_A f\left(A, \mathcal{B}^k, \mathcal{T}^k\right) + \frac{\rho}{2}\|A - A^k\|_F^2$$
$$\mathcal{B}^{k+1} = \arg\min_{\mathcal{B}} f\left(A^{k+1}, \mathcal{B}, \mathcal{T}^k\right) + \frac{\rho}{2}\|\mathcal{B} - \mathcal{B}^k\|_F^2$$
$$\mathcal{T}^{k+1} = \arg\min_{\mathcal{T}} f\left(A^{k+1}, \mathcal{B}^{k+1}, \mathcal{T}\right) + \frac{\rho}{2}\|\mathcal{T} - \mathcal{T}^k\|_F^2 \quad (17)$$

where $f(A, \mathcal{B}, \mathcal{T})$ is the equivalent objective function and $\rho > 0$ is a proximal parameter.

1) **Updating** A: by eliminating other items unrelated to A, the subproblem is defined as:

$$\arg\min_A \frac{1}{2}\|\mathcal{Y} - \mathcal{B}\times_3 A - \mathcal{T}\|_F^2$$
$$+ \lambda\|D_3 A\|_F^2 + \frac{\rho}{2}\|A - A^k\|_F^2 \quad (18)$$

whose solver can be directly obtained by solving the following Sylvester matrix equation:

$$AB_{(3)}^k(B_{(3)}^k)^{\mathrm{T}} + 2\lambda D_3^{\mathrm{T}} D_3 A + \rho A$$
$$= (Y_{(3)} - T_{(3)}^k)(B_{(3)}^k)^{\mathrm{T}} + \rho A^k \quad (19)$$

A fast solution of the above equation can be obtained by diagonalizing the circulant matrix $D_3^T D_3$ and a symmetric matrix $B_{(3)}^k(B_{(3)}^k)^{\mathrm{T}}$ through the 1-D fast Fourier transformation (FFT) and SVD, and its specific process can be referred to in Ref. [62].

2) **Updating** $\mathcal{B}$: the subproblem is obtained by picking out the items related to $\mathcal{B}$:

$$\min_{\mathcal{B}} \frac{1}{2}\|\mathcal{Y} - \mathcal{B}\times_3 A - \mathcal{T}\|_F^2 + \frac{\rho}{2}\|\mathcal{B} - \mathcal{B}^k\|_F^2$$
$$+ (\|\mathcal{B}\times_1 D_1\|_{2,1,\mathcal{W}_1} + \|\mathcal{B}\times_2 D_2\|_{1,\mathcal{W}_2}) \quad (20)$$

It can be solved by alternating direction method of multipliers(ADMM) [63]. By introducing two auxiliary valuables, the above equation is rewritten as:

$$\min_{\mathcal{B},\mathcal{Z}_1,\mathcal{Z}_2} \frac{1}{2}\|\mathcal{Y} - \mathcal{B}\times_3 A - \mathcal{T}\|_F^2 + \frac{\rho}{2}\|\mathcal{B} - \mathcal{B}^k\|_F^2$$
$$+ (\|\mathcal{Z}_1\|_{2,1,\mathcal{W}_1} + \|\mathcal{Z}_2\|_{1,\mathcal{W}_2}) \quad (21)$$
$$s.t. \mathcal{B}\times_1 D_1 = \mathcal{Z}_1, \mathcal{B}\times_2 D_2 = \mathcal{Z}_2$$

Its augmented Lagrangian function is:

$$\mathcal{L}_\beta\left(\mathcal{B}, \mathcal{Z}_1, \mathcal{Z}_2, \mathcal{P}_1, \mathcal{P}_2\right) = \frac{1}{2}\|\mathcal{Y} - \mathcal{B}\times_3 A - \mathcal{T}\|_F^2$$
$$+ \|\mathcal{Z}_1\|_{2,1,\mathcal{W}_1} + \|\mathcal{Z}_2\|_{1,\mathcal{W}_2} + \frac{\rho}{2}\|\mathcal{B} - \mathcal{B}^k\|_F^2$$
$$+ \langle \mathcal{B}\times_1 D_1 - \mathcal{Z}_1, \mathcal{P}_1 \rangle + \frac{\beta}{2}\|\mathcal{B}\times_1 D_1 - \mathcal{Z}_1\|_F^2 \quad (22)$$
$$+ \langle \mathcal{B}\times_2 D_2 - \mathcal{Z}_2, \mathcal{P}_2 \rangle + \frac{\beta}{2}\|\mathcal{B}\times_2 D_2 - \mathcal{Z}_2\|_F^2$$

where $\beta > 0$ is penalty factor. The solution of the function can be obtained by alternately updating each variable:

$$\mathcal{B}^{k+1,l+1} = \arg\min_{\mathcal{B}} \mathcal{L}_\beta(\mathcal{B}, \mathcal{Z}_1^l, \mathcal{Z}_2^l, \mathcal{P}_1^l, \mathcal{P}_2^l)$$
$$\mathcal{Z}_1^{l+1} = \arg\min_{\mathcal{Z}_1} \mathcal{L}_\beta(\mathcal{B}^{k+1,l+1}, \mathcal{Z}_1, \mathcal{Z}_2^l, \mathcal{P}_1^l, \mathcal{P}_2^l)$$
$$\mathcal{Z}_2^{l+1} = \arg\min_{\mathcal{Z}_2} \mathcal{L}_\beta(\mathcal{B}^{k+1,l+1}, \mathcal{Z}_1^{l+1}, \mathcal{Z}_2, \mathcal{P}_1^l, \mathcal{P}_2^l)$$
$$\mathcal{P}_1^{l+1} = \arg\min_{\mathcal{P}_1} \mathcal{L}_\beta(\mathcal{B}^{k+1,l+1}, \mathcal{Z}_1^{l+1}, \mathcal{Z}_2^{l+1}, \mathcal{P}_1, \mathcal{P}_2^l)$$
$$\mathcal{P}_2^{l+1} = \arg\min_{\mathcal{P}_2} \mathcal{L}_\beta(\mathcal{B}^{k+1,l+1}, \mathcal{Z}_1^{l+1}, \mathcal{Z}_2^{l+1}, \mathcal{P}_1^{l+1}, \mathcal{P}_2) \quad (23)$$

For $\mathcal{B}^{k+1,l+1}$, its updating is achieved by the following function:

$$\arg\min_{\mathcal{B}} \frac{1}{2}\|\mathcal{Y} - \mathcal{B}\times_3 A^{k+1} - \mathcal{T}^k\|_F^2 + \frac{\rho}{2}\|\mathcal{B} - \mathcal{B}^k\|_F^2$$
$$+ \frac{\beta}{2}\left\|\mathcal{B}\times_1 D_1 - \mathcal{Z}_1 + \frac{\mathcal{P}_1^l}{\beta}\right\|_F^2 + \frac{\beta}{2}\left\|\mathcal{B}\times_2 D_2 - \mathcal{Z}_2 + \frac{\mathcal{P}_2^l}{\beta}\right\|_F^2 \quad (24)$$

By taking the derivative of variable $\mathcal{B}$ and making it equal to zero, its solution can be computed by the following equation:

$$\mathcal{B}\times_3((A^{k+1})^{\mathrm{T}} A^{k+1}) + \beta(\mathcal{B}\times_1(D_1^{\mathrm{T}} D_1) + \mathcal{B}\times_2(D_2^{\mathrm{T}} D_2) = \mathcal{K} \quad (25)$$

where $\mathcal{K} = (\mathcal{Y} - \mathcal{T}^k)\times_3(A^{k+1})^{\mathrm{T}} + \beta((\mathcal{Z}_1^l - \frac{\mathcal{P}_1^l}{\beta}) + (\mathcal{Z}_2^l - \frac{\mathcal{P}_2^l}{\beta}) + \rho\mathcal{B}^k$.

The equation is further rewritten as:

$$B_{(3)}^{\mathrm{T}}((A^{k+1})^{\mathrm{T}} A^{k+1}) + \rho B_{(3)}^{\mathrm{T}} + C B_{(3)}^{\mathrm{T}} = K_{(3)}^{\mathrm{T}} \quad (26)$$

where $C = \beta[(I_{n_2} \otimes D_1^{\mathrm{T}} D_1) + (D_2^{\mathrm{T}} D_2 \otimes I_{n_1})]$ has a structure with circulant blocks. And the 2-D FFT and SVD are utilized to diagonalize the symmetric matrix $(A^{k+1})^{\mathrm{T}} A^{k+1}$ and C, respectively, and its specific process can be referred to in Ref. [62].

For $\mathcal{Z}_1^{l+1}$, $\mathcal{Z}_2^{l+1}$, we solve the following problem:

$$\arg \min_{\mathcal{Z}_1,\mathcal{Z}_2} \|\mathcal{Z}_1\|_{2,1,\mathcal{W}_1} + \frac{\beta}{2}\left\|\mathcal{B}\times_1 D_1 - \mathcal{Z}_1 + \frac{\mathcal{P}_1^l}{\beta}\right\|_F^2$$
$$+\|\mathcal{Z}_2\|_{1,\mathcal{W}_2} + \frac{\beta}{2}\left\|\mathcal{B}\times_2 D_2 - \mathcal{Z}_2 + \frac{\mathcal{P}_2^l}{\beta}\right\|_F^2 \quad (27)$$

which can be directly solved by the following closed solutions:

$$\mathcal{Z}_1^{l+1}(i,j,:) = shrink_{2,1}\left(\hat{\mathcal{Z}}_1(i,j,:), |W_1(i,j)|\cdot\frac{1}{\beta}\right)$$
$$\mathcal{Z}_2^{l+1}(i,j,v) = shrink_1\left(\hat{\mathcal{Z}}_2(i,j,v), |\mathcal{W}_2(i,j,v)|\cdot\frac{1}{\beta}\right) \quad (28)$$

And these threshold operators is defined as:

$$shrink_{2,1}(x,\xi) = \begin{cases} \frac{\|x\|_2 - \xi}{\|x\|_2}, if \ \ \xi<\|x\|_2 \\ 0, otherwise \end{cases},$$
$$[shrink_1(\mathcal{X},\xi)]_{i,j,v} = sign(x_{i,j,v})\max(|x_{i,j,v}| - \xi, 0) \quad (29)$$

where $\hat{\mathcal{Z}}_1 = \mathcal{B}^{k+1,l+1}\times_1 D_1 + \frac{\mathcal{P}_1^l}{\beta}$, $\hat{\mathcal{Z}}_2 = \mathcal{B}^{k+1,l+1}\times_2 D_2 + \frac{\mathcal{P}_2^l}{\beta}$, $W_1(i,j) = \frac{1}{\left\|\hat{\mathcal{Z}}_1(i,j,:)\right\|_2 + \varepsilon}$, $\mathcal{W}_2(i,j,v) = \frac{1}{|\hat{\mathcal{Z}}_2(i,j,v)| + \varepsilon}$. ε is a small constant for avoiding the appearance of singularities.

For $\mathcal{P}_1^{l+1}$, $\mathcal{P}_2^{l+1}$, they can be solved by:

$$\mathcal{P}_1^{l+1} = \mathcal{P}_1^l + \beta\left(\mathcal{B}^{k+1,l+1}\times_1 D_1 - \mathcal{Z}_1^{l+1}\right) \quad (30)$$

$$\mathcal{P}_2^{l+1} = \mathcal{P}_2^l + \beta\left(\mathcal{B}^{k+1,l+1}\times_2 D_2 - \mathcal{Z}_2^{l+1}\right) \quad (31)$$

3) **Updating** $\mathcal{T}$:the $\mathcal{T}$-subproblem is as follows:

$$\min_{\mathcal{T}} \frac{1}{2}\|\mathcal{Y} - \mathcal{B}\times_3 A - \mathcal{T}\|_F^2 + \gamma\|\mathcal{W}_{ASCE}\odot T\|_{1,\mathcal{W}_S}$$
$$+ \frac{\rho}{2}\left\|\mathcal{T} - \mathcal{T}^k\right\|_F^2 \quad (32)$$

which has the following solution:

$$\mathcal{T}^{k+1} = shrink_1(\mathcal{W}_{ASCE}\odot\hat{\mathcal{T}}, \mathcal{W}_S\odot\frac{\gamma}{1+\varepsilon}) \quad (33)$$

where $\hat{\mathcal{T}} = \frac{(\mathcal{Y} - \mathcal{B}^{k+1}\times_3 A^{l+1} + \varepsilon\mathcal{T}^k)}{1+\varepsilon}$, $\mathcal{W}_S = \frac{1}{|\hat{\mathcal{T}}(i,j,v)| + \varepsilon}$.

The pseudocode of the proposed PAM-based algorithm is summarized in Algorithm 1 to optimize the proposed model for detecting infrared small target.

A. Detection Procedure

Each step of the proposed model for detecting small targets is introduced in the overall schematic, as displayed in Fig. 5.

1) Intercepting the infrared sequence with a fixed number of frames and taking it as an input tensor cube $\mathcal{Y}$.

2) Calculating the $ASCE$ map of each frame of the clipping sequence cube and constructing the saliency enhancement factor $\mathcal{W}_{ASCE}$.

3) Separating the small target $\mathcal{T}$ and the background components $\mathcal{F}$ via the Algorithm 1.

4) Employing a simple threshold operator to extracting small target from each slice of the small target component, as following:

$$T_S = \max\left(c_{\min}, \mu + d\sigma\right) \quad (34)$$

where $c_{\min}$ and d denote the given constants empirically. μ and σ are the mean value and standard variance of each slice.

Algorithm 1 Solution of the SDD Model Based on PAM.

Input: The acquired infrared sequence $\mathcal{Y}\in\mathbb{R}^{n_1\times n_2\times n_3}$, rank r, parameters $\lambda, \beta, \gamma, \varepsilon = 0.01$;
1: **Initialize**: $k = 0$, $k_{max} = 30$, $l_{\max} = 10$, $A^0 = rand(n_3, r)$, $\mathcal{B}^0 = rand(n_1, n_2, r)$, $\mathcal{T}^0 = \mathcal{P}_1^0 = \mathcal{P}_2^0 = 0$.
2: **while** not converged and $k < k_{\max}$ **do**
3:     Solving A^{k+1} by (18)
4:     **Initialize**: $l = 0$, $\mathcal{P}_1^0 = \mathcal{P}_2^0 = \mathcal{Z}_1^0 = \mathcal{Z}_2^0 = 0$.
5:     **while** $l < l_{\max}$ **do**
6:         Solving $\mathcal{B}^{k+1,l+1}$ through (24)
7:         Solving $\mathcal{Z}_1^{l+1}$, $\mathcal{Z}_2^{l+1}$ through (28)
8:         Solving $\mathcal{P}_1^{l+1}$, $\mathcal{P}_2^{l+1}$ through (30,31)
9:         Updating $l: l = l + 1$
10:       Check the convergence condition:
        $\left\|\mathcal{B}^{k+1,l} - \mathcal{B}^{k+1,l-1}\right\|_F / \left\|\mathcal{B}^{k+1,l-1}\right\|_F < 10^{-5}$
11:     **end while**
12:     Under $\mathcal{B}^{k+1} = \mathcal{B}^{k+1,l}$
13:     Solving $\mathcal{T}^{k+1}$ through (33)
14:     Updating $k: k = k + 1$
15:     Check the convergence condition:
    $\left\|\mathcal{B}^k\times_3 A^k - \mathcal{B}^{k-1}\times_3 A^{k-1}\right\|_F / \left\|\mathcal{B}^{k-1}\times_3 A^{k-1}\right\|_F < 10^{-5}$
16: **end while**

Output: Infrared background $\mathcal{F} = \mathcal{B}\times_3 A$ and small target component $\mathcal{T}$

V. EXPERIMENTAL ANALYSIS

A. Experimental Settings

Datasets: The proposed model undergoes evaluation on an extensive array of real infrared sequences, encompassing both public and private datasets that span diverse scenes. Due to space constraints, we showcase only 18 representative sequences in Fig. 6. These sequences cover various scenarios, with backgrounds ranging from flat to complex and targets from prominent to dim. Considering that the greatest challenge for current detection algorithms is detecting dim targets from strong clutter, detection performance in extremely complex scenarios will be more convincing than in simple and uniform scenarios. Sequences a-f, furnished by Hui et al. [64], feature air-ground scenes. Sequences g-j, m-r, collected by our research group, display sky-cloud, deep-space and sea-air scenes. Sequence k, provided by Wang et al. [22], presents a ground-air scene, while sequence l, contributed by our collaborator, exhibits a scene with sea clutter and fish scale light. A summary of the detailed information pertaining to the showcased scenes can be found in Table III.

Baselines and Parameter Settings: In the experiments, we juxtapose the proposed method with 14 state-of-the-art solutions, assessing both objective indicators and visual effects. These approaches encompass 6 single-frame detection methods: Weighted Local Difference Measure (WLDM) [3], Multiscale Patch-based Contrast Measure (MPCM) [18], Feature Kernel-based Random Walk (FKRW) [45], Infrared Patch-Image model (IPI) [16], Reweighted Infrared Patch-Tensor model (RIPT) [21], and Dynamic Weight-guided Smooth-Sparse Decomposition (DWSSD) [36], and 8 multi-frame detection methods:
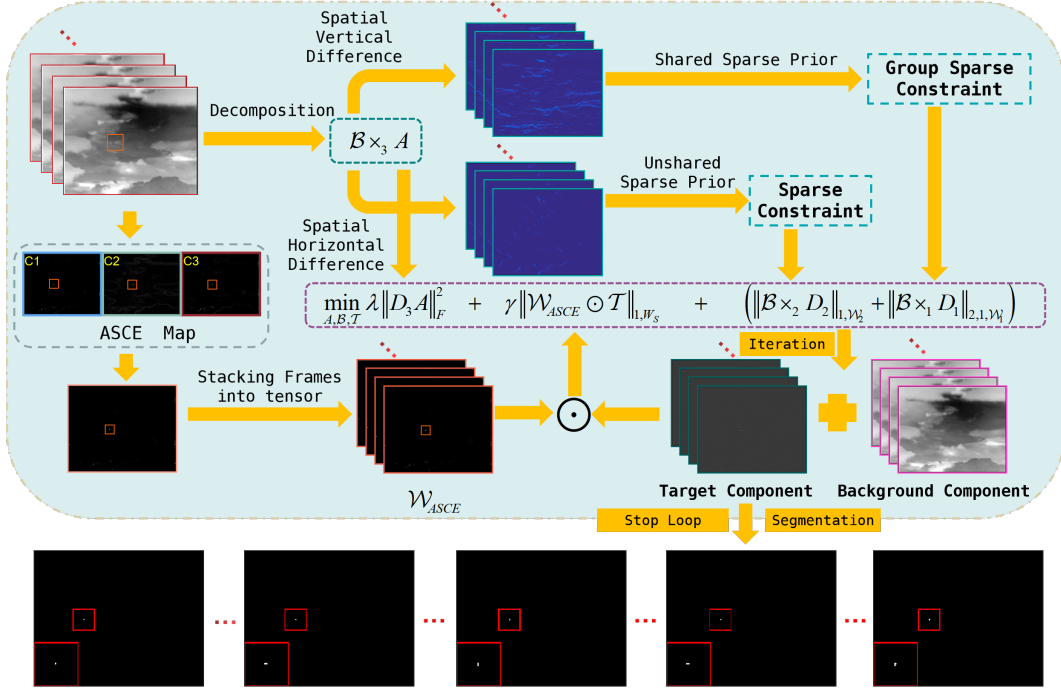
Fig. 5. The overall flow-process diagram of the proposed infrared small target detection model. During detection, the target saliency enhancement factor $\mathcal{W}_{ASCE}$, obtained by stacking single frame ASCE mapping, is first constructed and then integrated into the SDD model based on PAM optimization. During decomposition, the target components from each layer are multiplied by the enhancement factor to highlight targets and suppress background clutter.
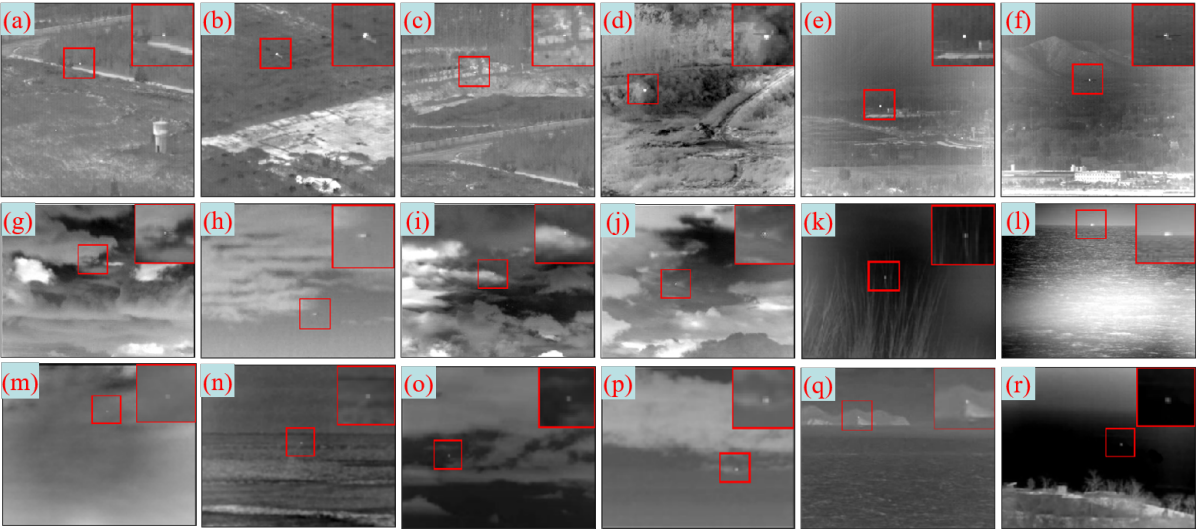

Fig. 6. Exhibition of the 18 representative scenes.

Spatial-Temporal Saliency Map (STSM) [24], spatial-temporal tensor modeling with Saliency Filter Regularization (SFR) [46], Tensor Completion with Top-Hat model (TCTH) [28], Edge and Corner Awareness-based spatial-temporal tensor model (ECA) [37], Non-convex Tensor Low-rank Approximation (NTLA) [38], Tensor Spectral k-support model (TSPK) [32], Sparse Regularization-based Spatial–Temporal Twist tensor (SRSTT) [34] and 4-D Tensor Ring (4DTR) [39]. To ensure equitable comparisons, we employ the original codes provided by the authors for FKRW, IPI, RIPT, ECA, NTLA, TSPK, 4DTR, SRSTT and DWSSD, adhering to the recommended parameter

settings delineated in their respective papers. For the remaining methods, we re-implement them based on the corresponding references. A summary of all methods and their parameter settings can be found in Table IV.

Evaluation Indicators: To provide an objective comparison of the performance among the tested methods, we utilize four evaluation indices. These indices include background suppression factor (BSF), signal to clutter ratio gain (G_{SCR}), contrast gain (CG), and receiver operating characteristic curve (ROC). The BSF quantifies the suppression effect in the vicinity of the small target's neighboring area and is defined

TABLE III
DETAILED FEATURES OF THESE PRESENTED INFRARED SEQUENCES

NO.	Frames	Resolution	Target Features	Background Features
Sequence a-f	400 per sequence	256×256	Large changes in brightness and scale, relatively high contrast	With numerous surface disturbances and nature noise, quite uneven
Sequence g-k	50 per sequence	320×256, 200×256,	Punctate structure with tiny size, extremely dim with low contrast	With numerous cloud clutters, brightness cloud edges, atmospheric interference
Sequence l	200	256×320	Saliency and size varying greatly, target edge blur	With brighter sea surface reflected light and fish scale light, strong waves
Sequence m-r	100 per sequence	256×200, 300×256, 520×430	Weak saliency and irregular size, low signal to clutter ratio	Significant background changes uneven ground surface, bright glitters.

TABLE IV
PARAMETER SETTING SUMMARY BY FINELY ADJUSTING THE TESTED METHODS

No.	Methods	Parameter Setting
1	WLDM [3]	$L = 4, m = 2, n = 2$.
2	MPCM [18]	Mean filter size: 3×3, local window size:$N = 3, 5, 7, 9$.
3	FKRW [45]	$K = 4, p = 6, \beta = 200$, window size: 11×11.
4	IPI [16]	Patch size: 50×50,sliding size: 10,$\lambda = L/\min(m, n)^{1/2}$, $L \in [1, 5]$, $\varepsilon = 10^{-7}$.
5	RIPT [21]	Patch size: 50×50, sliding size: 10, $\lambda = L/\min(I,J,P)^{1/2}$, $L \in [0.1, 2]$, $h = 10$, $\epsilon = 0.01$, $\varepsilon = 10^{-7}$.
6	STSM [24]	Patch size:4, overlap:2, local neighboring samples:8, $\lambda = 10^7$, frame number: 5
7	SFR [46]	Filter size:5×5, $\sigma_1 = 0.8$, $\sigma_2 = 2.0$, $t = 5$, $p = 0.9$, $\epsilon = 10^{-8}$, $\lambda = L/\left(\min(m, n)^{1/2} \times t\right)$, $L \in [2, 5]$.
8	TCTH [28]	$p = 0.1$, $\varepsilon = 10^{-7}$, $s2B = sB_O = 5 \times 5$, $s1B = 3 \times 3$, frame number: 10, $\beta = 1/\sqrt{\min(I, J, P)}$, $L \in [2, 10]$, $\alpha = 0.001$.
9	ECA [37]	$\beta = 0.5$, $t = 3$, $\lambda_1 = 0.005$, $\lambda_2 = L/\left(\min(m, n)^{1/2} \times t\right)$, $L \in [0.1, 2]$.
10	NLTA [38]	$L = 3$, $H \in [5, 10]$, $\lambda_{TV} = 0.005$, $\lambda_S = H/\left(\max(m, n)^{1/2} \times L\right)$, $\lambda_3 = 100$.
11	TSPK [32]	frame number: 10, patch size: 10×10, spectral factor: 5, $\lambda \in [1.5, 2]$, $\beta \in [0.5, 5]$, $\varepsilon = 5 \times 10^{-7}$.
12	4DTR [39]	Patch size:$N_1 \times N_2 : 70 \times 70$, temporal size:$N_3 = 15$,$\lambda_1 = \sum_{i=1}^{l} 2 \Big/ \sqrt{\max\left(\prod_{i=n}^{n+l-1} N_i, \prod_{i=n+l}^{n-1} N_i\right)}$, $\tau = 1 \times 10^{-7}$.
13	SRSTT [34]	$L = 30$, $\lambda_1 = 0.05$, $\lambda_2 = 0.1$, $\lambda_3 = 100$,$\varepsilon = 10^{-7}$,$\mu = 0.01$.
14	DWSSD [36]	$L = 0.015$, $\beta = 100\lambda$, $\lambda = L/\left(\max(m, n)^{1/2}\right)$, $k = 5$.
15	SDD	Frame number: 30, $R = 10$, $\lambda \in [0.5, 5]$, $\gamma \in [0.1, 1]$, outer iteration:50, inner iteration:10, $\beta = 15000$, $\mu = 0.05$, $\varepsilon = 0.01$.

as follows:

$$BSF = \frac{\sigma_{in}}{(\sigma_{out} + \omega)} \qquad (35)$$

where σ_{in} and σ_{out} are the standard deviation of target neighboring region of original and suppressed images, separately; $\omega = 0.01$ is a smoothing factor to avoid division zero. G_{SCR} measures the improvement in signal-to-clutter ratio before and after target separation and can be defined as:

$$G_{SCR} = \frac{out_{SCR}}{in_{SCR}} \qquad (36)$$

where in_{SCR} and out_{SCR} are the SCR values before and after target separation, separately, and $SCR = |M_t - \mu_b| / (\sigma_b + \omega)$. M_t denotes the maximum intensity of the target area. μ_b and σ_b are the average grayscale and standard deviation of the neighboring area around small targets. CG measures the contrast between the separated target and its local neighborhood and can be defined as:

$$CG = \frac{CON_{out}}{CON_{in}} \qquad (37)$$

where CON_{in} and CON_{out} represent the target local contrast before and after small target separation, respectively, and $CON = |M_t - \mu_b|$.

Herein, we assume that the target size is $a \times b$, and the neighboring area is defined as $(a + 2d) \times (a + 2d)$, with d set to 30 as the neighborhood width. The larger values

of the evaluation indicators corresponds to superior method performance.

The ROC curve depicts the dynamic range between the probability of detection (P_b) and the false alarm rate (F_a). P_b and F_a can be defined as follows:

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

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

In ROC curves, we regard small targets as detected if pixels exist within a 5×5 window surrounding the ground truth. Moreover, we compute the area under the curve (AUC) of the ROC curve to further appraise the detection performance. A larger AUC signifies superior detection performance.

B. Model Discussion

1) Inner Iteration Analysis: The number of inner iterations, denoted as l, influences not only the detection performance but also the computational complexity of the proposed model. To evaluate its impact, we vary l from 2 to 20 under scenes a and g, presenting the AUC and average computational time at a specific l in Table V. It is observed that the proposed model attains stable and superior performance when $l \geq 12$. However, as the number of iterations escalates, the computational time increases significantly, while the performance remains unaltered.

Consequently, we set the number of inner iterations to 12 in subsequent experiments to strike a balance between detection performance and computational complexity.

2) Model Parameter Analysis: The objective function of the proposed model involves four key parameters: the rank parameter R, continuity constraint factor λ, and sparse penalty μ. The selection of these parameters significantly impacts the performance of the proposed model. We subsequently investigate the influence of these parameters on typically complex scenes a and g. Fig. 7 illustrates the changes in the ROC curves of the two scenes when varying the parameter values.

The rank factor R primarily serves to measure the inter-frame correlation of infrared sequence cubes. The sensitivity of the proposed model's performance to changes in R is displayed in the first column of Fig. 7. It is evident that when R is set to 10 or 15, the proposed model achieves stable and superior performance. However, the model's performance diminishes when R assumes other values. Furthermore, a larger value of R engenders higher computational complexity. Considering these factors, we set R to 10 in our experiments.

The parameter λ governs the temporal continuity constraint. The second column of Fig. 7 demonstrates the changes in the ROC curves induced by varying the factor. The proposed model's behavior exhibits robustness to the parameters, with the ROC curves displaying a similar changing tendency under different scenes and only minor performance differences observed. Consequently, we set λ to 1.

The sparse penalty γ plays a pivotal role in controlling target sparsity due to its threshold attribute. A smaller value of γ results in the retention of some non-target components, while a larger value may lead to the loss of small targets. Therefore, striking an appropriate balance between detection probability and false alarm necessitates the careful selection of γ. Fig. 7's third column presents the ROC curves when varying γ from 0.005 to 5. It becomes apparent that the detection performance of the model deteriorates when the value of γ is excessively large or small. This outcome arises from the prevalence of high false alarms or low detection rates in these cases. When the value range is set between 0.1 and 1, the proposed model achieves stable and superior detection performance.

3) Ablation Experiment: In order to analyze the effectiveness of different spatial constraint configurations, we conduct a set of ablation experiments, as shown in Fig. 8. From the subfigures, it can be seen that using appropriate constraints for differential priors in different spatial directions can significantly improve the detection performance of the model. Especially, ignoring differential prior sparse patterns in different spatial directions and using the same constraints will affect the detection performance of the model. For example, in scene b and f, the P_d of SDD+HL_1+VL_1+W_S and SDD+HL_{21}+VL_{21}+W_S) significantly decrease when the false alarms are the same. Furthermore, it can be observed that the model's performance is better when sparse constraints are employed concurrently, compared to the simultaneous application of group sparse constraints. This is primarily attributed to the inability of group sparse constraints to effectively suppress interference that resembles the target point. The stability of the SDD+HL_{21}+VL_1+W_S

model is inadequate, as evidenced by its strong performance in scenario b and poor performance in the other two scenarios. Owing to the absence of saliency enhancement factors, the target components of the SDD+HL_1+VL_{21} model are prone to being lost during the decomposition process, thereby impeding further improvements in the target detection rate.

The experiment demonstrates that the proposed model can effectively suppress background clutter, highlight targets, and enhance robustness by appropriately allocating sparse and group sparse constraints, while integrating elaborately designed saliency enhancement factors.

4) Convergence Analysis: Owing to the non-convexity of the proposed model, existing convex solvers are rendered inapplicable. Consequently, the ADMM and reweighted strategy are incorporated into the PAM-based optimization framework. To empirically verify the convergence of the proposed algorithm, we employ the relative error change as a criterion. Fig. 9 presents the relative error change curves of the proposed algorithm concerning the iteration number across twelve distinct scenes. Upon examining the curves, it becomes apparent that the decrement values of the relative error gradually approach the convergence criterion after a sufficient number of iterations for all tested scenes. This observation numerically demonstrates the effective convergence of the solution.

C. Qualitative Evaluation of The Proposed Method

1) Evaluation on Different Scenes: To assess the robustness of the proposed model, we examine its performance on different scenes depicted in Fig. 6. In Fig. 10, we showcase the detection results of the proposed model on these scenes. It is evident that small targets are entirely separated without any residual background clutter. These findings demonstrate that the proposed model can consistently and proficiently tackle various scenarios.

To evaluate the noise-robust performance of the proposed method, we conduct experiments on scenes affected by different noise levels, as illustrated in Fig. 11. The first and third rows of the figure display six scenes with Gaussian noise, featuring a zero-mean and standard deviations of 5 and 10, respectively. The second and fourth rows present the detection results obtained using the proposed method.

A comparison of the detection results in the second and fourth rows reveals a substantial improvement in the former. For example, the small target in the results of noise scene l is nearly indiscernible, and a false alarm point emerges in noise scene h. These observations suggest that detection becomes increasingly challenging as the noise level escalates. Nevertheless, the proposed method can accurately locate the target, provided it is not entirely submerged in noise.

2) Visual Comparison with Other Competitors: We utilize six representative scenes a, b, j, m, l, r to compare the visual impact of the proposed model with 14 state-of-the-art competitors. These sequences not only include sea-air, land-sea, and deep-sea scenarios but also exhibit significant variations in target characteristics and background complexity. Fig. 12-17 display the detection results for all tested methods across the six scenes. Among the competitors, WLDM, MPCM, FKRW and

TABLE V
DETAILED INFORMATION OF THE CANDIDATE TARGET SETS AND THE ORIGINAL REAL BACKGROUNDS

Datasets	l	2	4	6	8	10	12	14	16	18	20
scene a	AUC	4.68	4.74	4.78	4.80	4.82	4.83	4.83	4.83	4.83	4.83
	Times	0.33	0.60	0.94	1.42	1.82	2.74	3.18	3.57	4.00	4.46
scene b	AUC	1.24	1.24	1.26	1.28	1.30	1.31	1.30	1.31	1.31	1.31
	Times	0.29	0.61	1.12	1.95	2.40	2.71	3.29	3.54	4.03	4.39
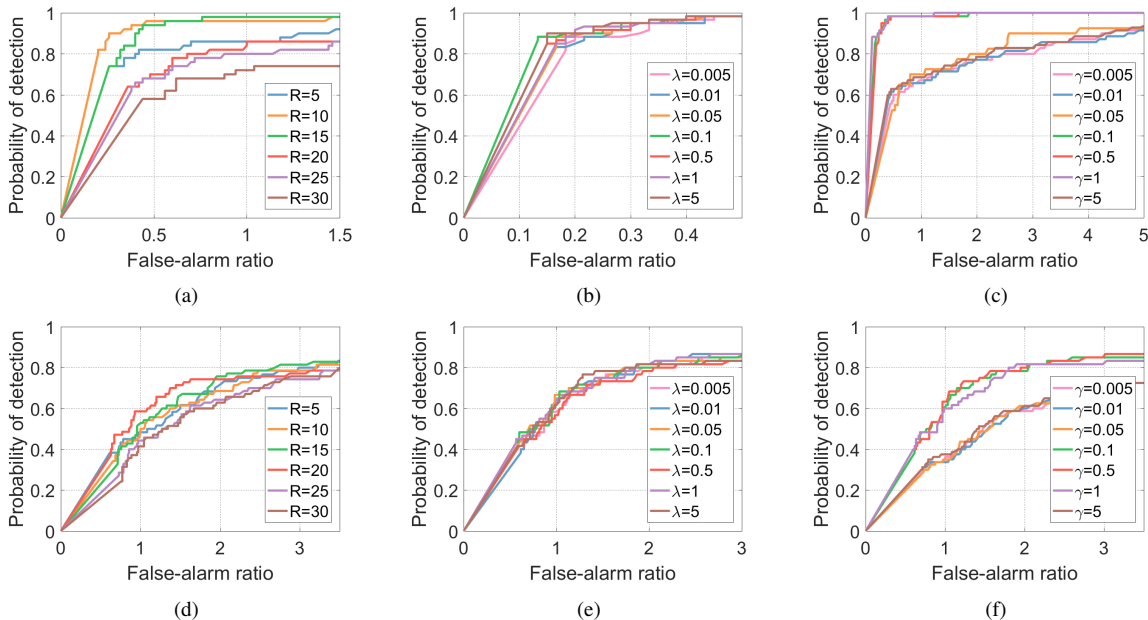
Fig. 7. Sensitivity analysis of the involved parameters by using the changing ROC curves under scenes (a) and (g).
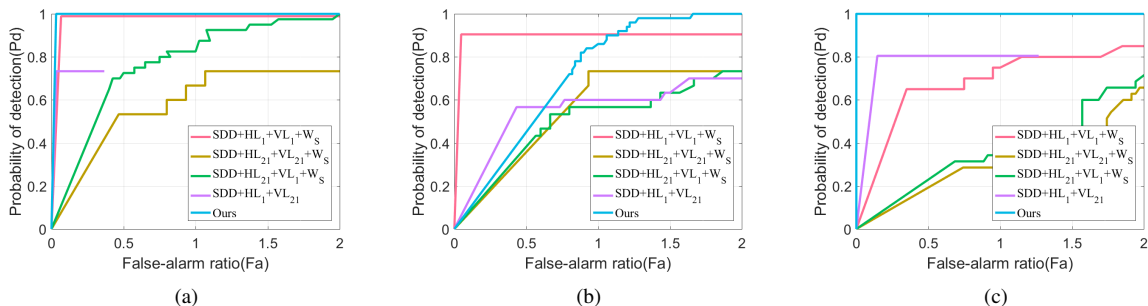

Fig. 8. Ablation studies with different assembles settings under Scenes b, c and f by the changing ROC curves. SDD+HL_1+VL_1+W_S: Nonshared l_1 constraints are used for both spatial horizontal and vertical modes. SDD+HL_{21}+VL_{21}+W_S: Shared l_{21} constraints are used for both spatial horizontal and vertical modes. SDD+HL_{21}+VL_1+W_S: Shared l_{21} constraint is used for the spatial horizontal mode while nonshared l_1 constraint for the vertical mode. SDD+HL_1+VL_{21}: Using the proposed constraint strategy without significant enhancement factors.

STSM belong to the saliency-based methods, which emphasize extracting small targets by eliminating background clutter. As evident from the results obtained by WLDM and MPCM in Fig. 12 - 15, 17, small targets are accentuated, but structures resembling target saliency are also preserved, leading to high false alarms, and in Fig. 16, low-contrast targets are not only unenhanced but also lost, leading to false detections. FKRW and STSM exhibit superior clutter suppression compared to WLDM and MPCM, with the exception of the detection results featuring either missing targets in Fig. 12, 14 and 16 or residual backgrounds in Fig. 13 and 17. The remaining methods are

founded on the low-rank and sparse decomposition framework. From their results, it is apparent that although fewer background residues are present in the target images compared to saliency-based methods, performance disparities still persist.

As illustrated in Fig. 12 and 13, IPI, RIPT, 4DTR, TSPK, NLTA and SRSTT exhibit no background residues, while SFR, TCTH, ECA and DWSSD contain some residues. The detection results in Fig. 14 reveal that NLTA and 4DTR lead to the omission of real targets and confusion with false ones, but the rest of these methods display relatively stable performance. Examining Fig. 15 and 17, it is clear that the performance

TABLE VI
THE QUANTITATIVE INDICATORS OF ALL COMPARED METHODS WITH RESPECT TO AVERAGE BSF, G_{SCR} AND CG PER FRAME UNDER SIX TESTED SCENES.

| Methods	Scene a			Scene b			Scene c			Scene d			Scene e			Scene f		
	BSF	G_{SCR}	CG	BSF	G_{SCR}	CG	BSF	G_{SCR}	CG	BSF	G_{SCR}	CG	BSF	G_{SCR}	CG	BSF	G_{SCR}	CG
WLDM[3]	4.66	2.96	1.42	6.16	8.55	1.31	6.14	2.29	1.25	6.71	3.82	1.93	5.03	3.22	1.15	7.97	13.38	1.34
MPCM[18]	9.00	7.55	2.81	15.62	11.94	5.64	6.29	5.91	2.85	2.49	1.82	2.24	14.58	6.53	1.13	16.18	8.05	2.41
FKRW[45]	5.55	4.79	0.96	6.89	1.77	0.69	7.12	3.46	0.88	9.29	2.37	1.66	3.49	1.71	0.98	7.86	3.44	0.74
IPI[16]	111.92	39.38	11.29	72.32	24.01	12.07	95.89	89.31	16.93	121.56	55.89	18.19	49.79	16.74	12.18	73.65	**76.54**	12.9
RIPT[21]	121.92	<u>76.48</u>	<u>29.16</u>	82.39	28.69	22.07	<u>102.88</u>	<u>91.37</u>	<u>26.49</u>	102.43	<u>65.47</u>	23.46	58.96	26.62	16.28	79.61	<u>73.66</u>	22.16
STSM[24]	5.667	11.30	7.668	6.965	17.13	1.334	8.640	7.384	3.360	13.22	13.95	1.944	5.117	9.508	1.239	7.985	20.71	1.34
SFR[46]	39.222	32.185	19.357	41.289	28.649	10.48	45.127	40.624	10.189	47.596	39.146	16.22	26.174	18.213	**18.956**	42.897	34.212	22.020
TCTH[28]	71.568	31.013	15.143	72.329	36.103	12.078	96.040	74.027	16.820	82.149	62.300	18.195	49.79	19.405	12.176	73.656	70.063	12.96
ECA[37]	37.937	15.884	16.286	71.817	43.005	12.213	23.408	48.658	14.584	**145.96**	58.474	<u>37.955</u>	37.517	17.665	11.342	69.21	32.74	22.56
NLTA[31]	60.712	32.575	14.967	56.513	27.291	<u>24.960</u>	41.311	29.003	22.110	59.540	27.58	24.19	27.419	9.953	11.565	29.447	31.945	<u>24.226</u>
TSPK[32]	**183.441**	76.191	12.867	**106.866**	<u>46.151</u>	14.282	87.332	55.796	12.011	132.720	36.601	14.182	**126.091**	<u>49.429</u>	12.211	<u>80.068</u>	41.458	14.126
4DTR[39]	3.909	1.278	4.468	13.372	14.453	5.820	6.611	1.617	4.225	6.292	1.311	6.912	10.765	1.263	3.626	14.446	3.232	4.116
SRSTT[34]	80.912	54.081	9.823	66.243	24.355	10.260	62.214	26.994	12.102	68.958	28.979	11.221	64.352	27.541	10.612	68.312	36.975	14.121
DWSSD[36]	9.347	1.244	3.016	8.999	1.989	2.936	6.3725	4.110	2.949	10.753	1.799	2.473	4.576	2.283	2.251	7.122	3.460	1.121
SDD	<u>123.733</u>	**84.66**	**35.092**	<u>90.222</u>	**49.428**	**26.917**	**106.679**	**97.322**	**34.781**	<u>136.174</u>	**70.430**	**39.026**	<u>79.391</u>	**49.946**	<u>14.827</u>	**88.234**	68.673	**42.436**

(a)   (b)
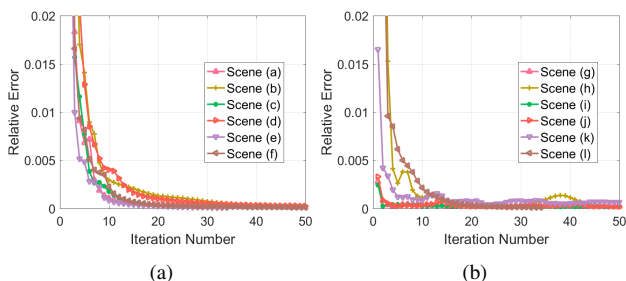
Fig. 9. Relative error change curves of the proposed model along with the number of iteration under different scenes.

of these methods in scene r is better than in scene l, mainly due to the interference of bright sea surface reflections in scene l. In Fig.16, dim targets are easily lost in RIPT, ECA, NLTA and DWSSD. In contrast, our proposed model achieves a significant performance improvement in eradicating background disturbances while accurately extracting small targets across different scenes, which is attributed to the integration of the spatial-temporal difference directionality prior and the saliency enhancement factor.

3) Quantitative Comparison with Other Competitors: In addition to visual illustrations, we offer objective evaluation indicators to thoroughly validate the proposed method. Three evaluation metrics are employed, namely the averages of BSF, G_{SCR}, and CG for each frame. The average results for all tested methods across six different scenes, with respect to the three indices, are presented in Table VI. The highest values are marked in bold, and the second highest values are underlined. The results clearly demonstrate that the proposed method surpasses its counterparts. In scene a, b, d, e, the BSF of our proposed method is slightly lower than the highest value, ranking second. Furthermore, the proposed method's G_{SCR} value is best in scenes a-e, but lower than RIPT and TCTH in scene f. Specifically, our approach attains the highest values in scenes a-d, f for CG, but slightly lower than SFR in scene e. This result indicates the model's efficacy in preserving the brightness of small targets during extraction.

The three aforementioned indicators primarily assess the local background suppression performance of the evaluated methods. In this section, we present the global ROC curves illustrating the dynamic changes in detection probability with respect to the false alarm rate for all tested methods, as depicted in Fig. 18. A steeper curve ascent indicates superior performance, signifying that higher detection performance can be attained with a reduced number of false alarms. Among all methods, our proposed approach achieves the highest P_d at lower F_a values. It is important to note that other methods demonstrate inconsistent detection performance across different scenes. For example, in scenes c and d, RIPT's P_d is 0.82 at an F_a of 2.5, which is considerably lower than its performance in other scenes. While SFR achieves a P_d of 1 in scenes a, d, and e, it does not maintain this level in the remaining tested scenes. SRSTT and 4DTR show promising results in scenes d-f, but its performance declines in scenes a-c. Based on this analysis, we can conclude that our proposed method effectively balances P_d and F_a across different scenes, thereby confirming its stability.

4) Computational Complexity and Time Consumption: In this subsection, we examine the computational complexity of the PAM-based solution applied to an infrared sequence cube $\mathcal{Y}^{n_1 \times n_2 \times n_3}$, as outlined in Algorithm 1. The primary computational expense at each loop iteration arises from updating A, $\mathcal{B}$, $\mathcal{Z}_1$, $\mathcal{Z}_2$, $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{T}$, and the number of inner iterations l. The cost of updating A is $\mathcal{O}\left(rn_1 n_2 n_3 + r^2 n_3 + rn_3 \log(n_3)\right)$, attributed to the involvement of SVD, 1-D FFT, and multiple matrix multiplications. Updating $\mathcal{B}$ incurs a computational expense of $\mathcal{O}\left((r + n_1 + n_2) n_1 n_2 n_3 + r^2 n_3 + rn_1 n_2 \log(n_1 n_2)\right)$, resulting from SVD, 2-D FFT, and several tensor-matrix product operations. For updates to $\mathcal{Z}_1$, $\mathcal{Z}_2$, and $\mathcal{T}$, simple threshold operators are employed, yielding a cost of $\mathcal{O}(n_1 n_2 n_3)$. The computational complexity of updating $\mathcal{P}_1$ and $\mathcal{P}_2$ is $\mathcal{O}\left(n_1^2 n_2 r\right)$. Ultimately, the computational complexity of the proposed optimization is $\mathcal{O}\left(rn_3 \log(n_3) + l\left((r + n_1 + n_2) n_1 n_2 n_3 + rn_1 n_2 \log(n_1 n_2)\right)\right)$.

To assess the detection efficiency of our proposed algorithm, we compare the average time consumption per frame for all tested scenes with 14 competing methods, as detailed in Table
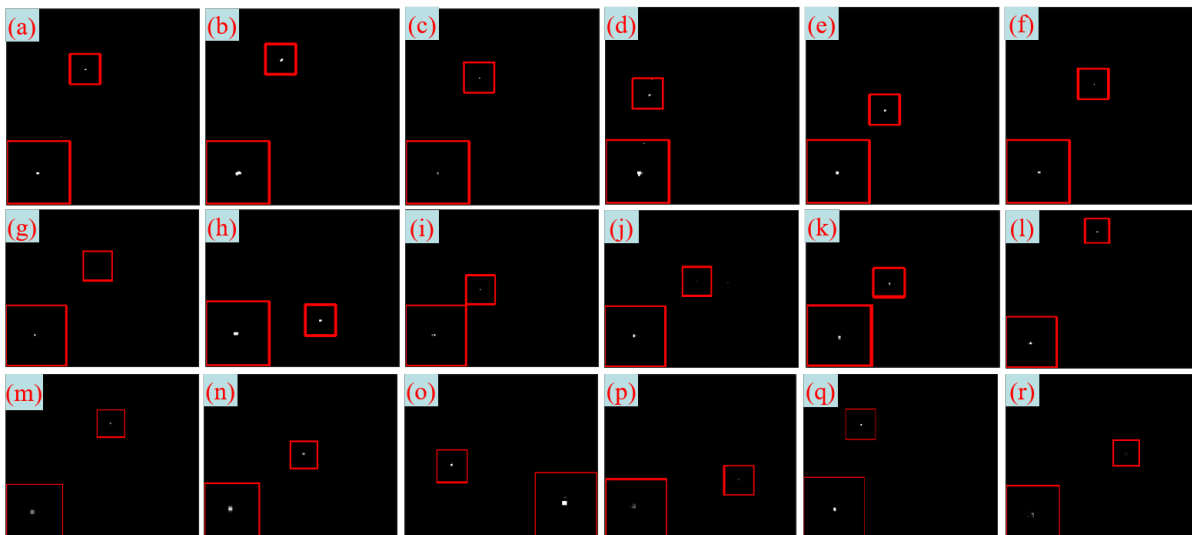
Fig. 10. Detection results obtained by the proposed model under twelve displayed scenes.
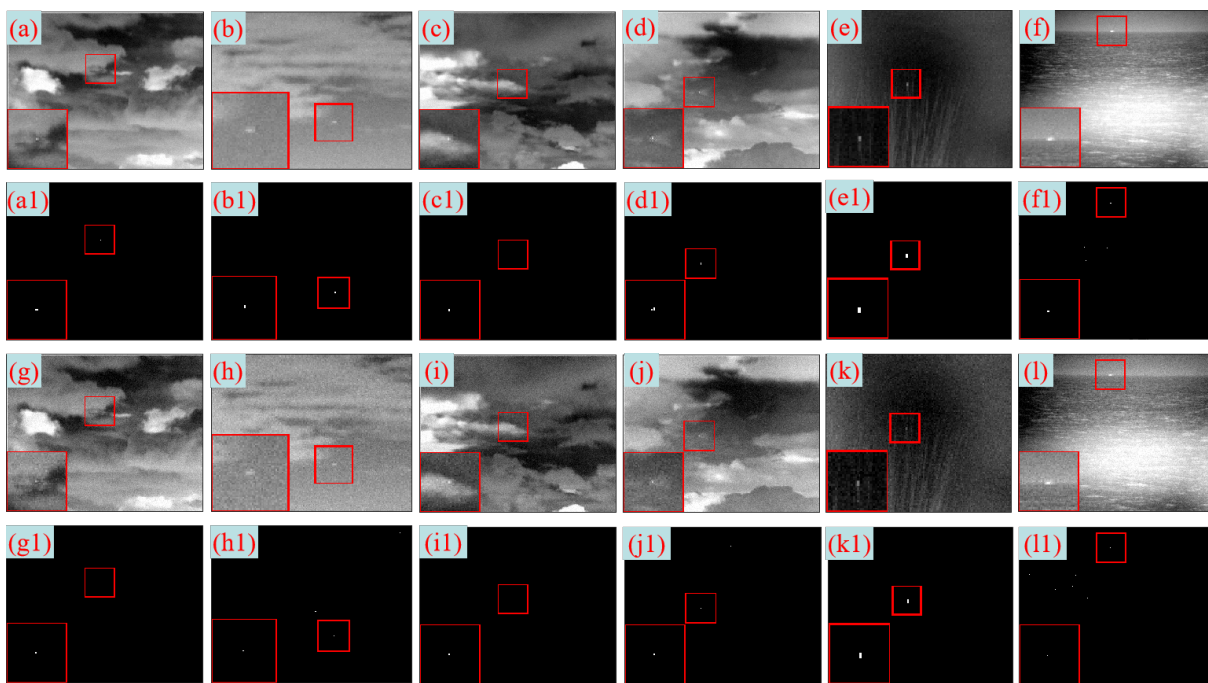

Fig. 11. Evaluation of the proposed method on polluted scenes with additive white Gaussian noise. The first and third rows show the noise images with standard deviation of 5 and 10. The second and fourth rows show the detecting results by the proposed method.
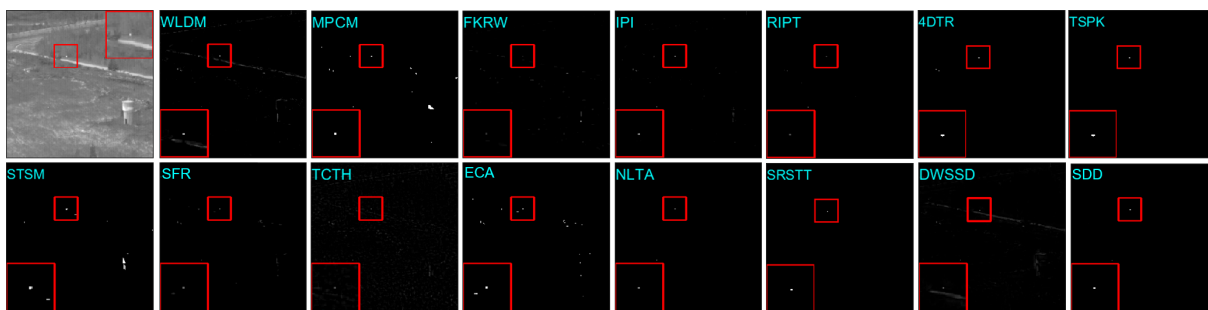

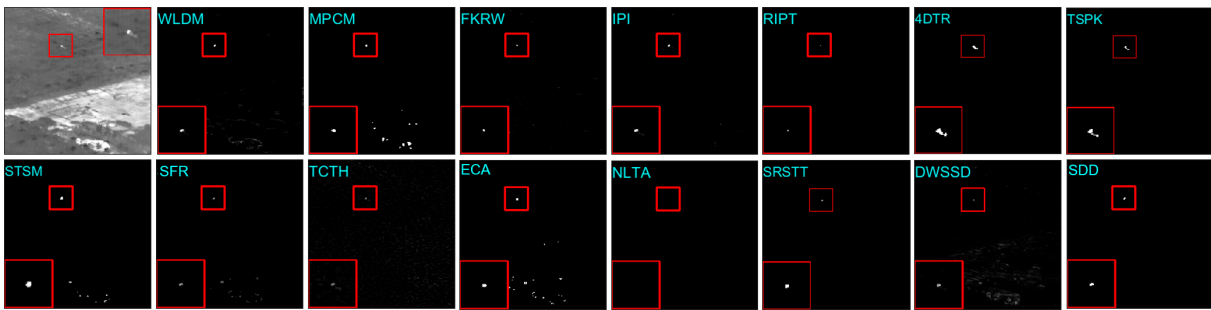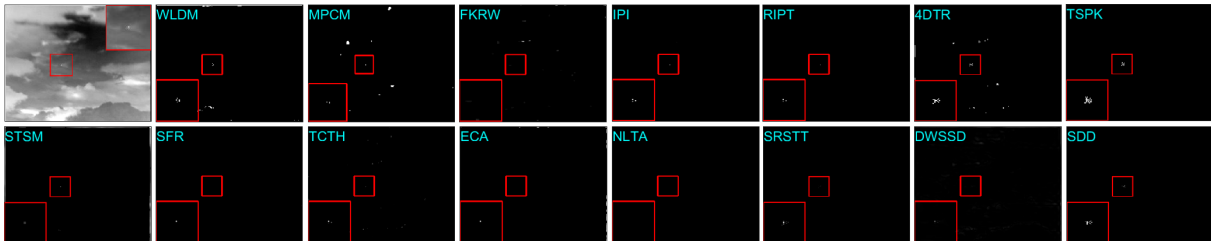Fig. 12. Detection results of ten comparison methods under scene a. The target area is zoomed in the lower left corner for better observation.

Fig. 13. Detection results of ten comparison methods under scene b. The target area is zoomed in the lower left corner for better observation.

Fig. 14. Detection results of ten comparison methods under scene j. The target area is zoomed in the lower left corner for better observation.
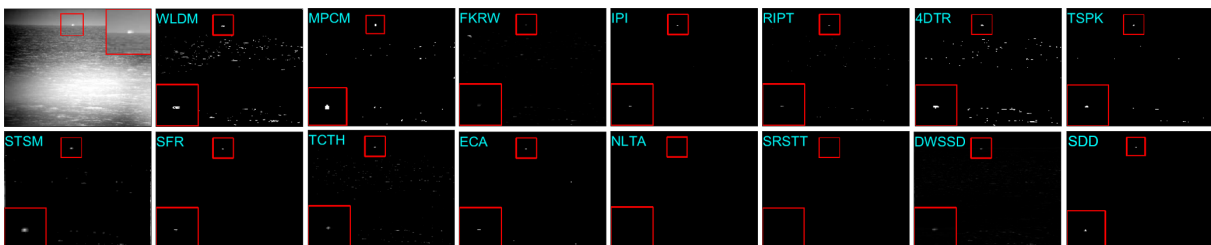

Fig. 15. Detection results of ten comparison methods under scene l. The target area is zoomed in the lower left corner for better observation.
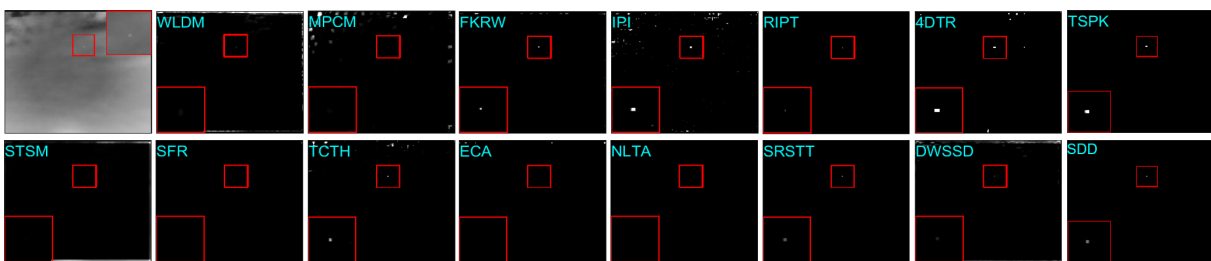

Fig. 16. Detection results of ten comparison methods under scene m. The target area is zoomed in the lower left corner for better observation.
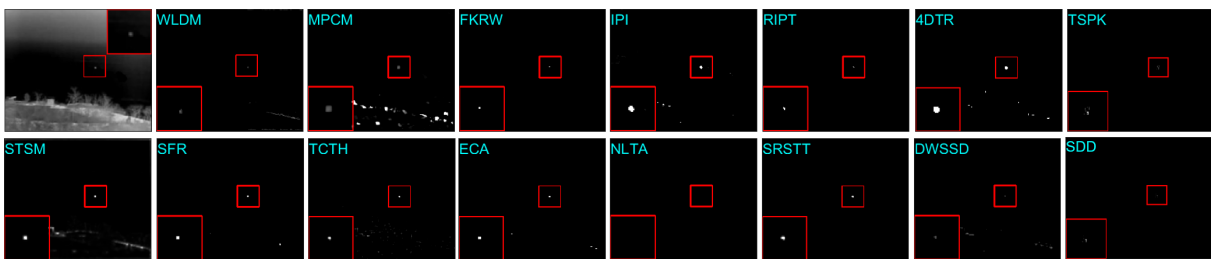

Fig. 17. Detection results of ten comparison methods under scene r. The target area is zoomed in the lower left corner for better observation.

VII. It can be observed that saliency-based approaches, such as WLDM, MPCM, FKRW, and STSM, exhibit faster performance than the rest of the low-rank decomposition-based methods. However, these methods demonstrate less stability across different scenes compared to their low-rank decomposition-based counterparts. Among the low-rank decomposition-based
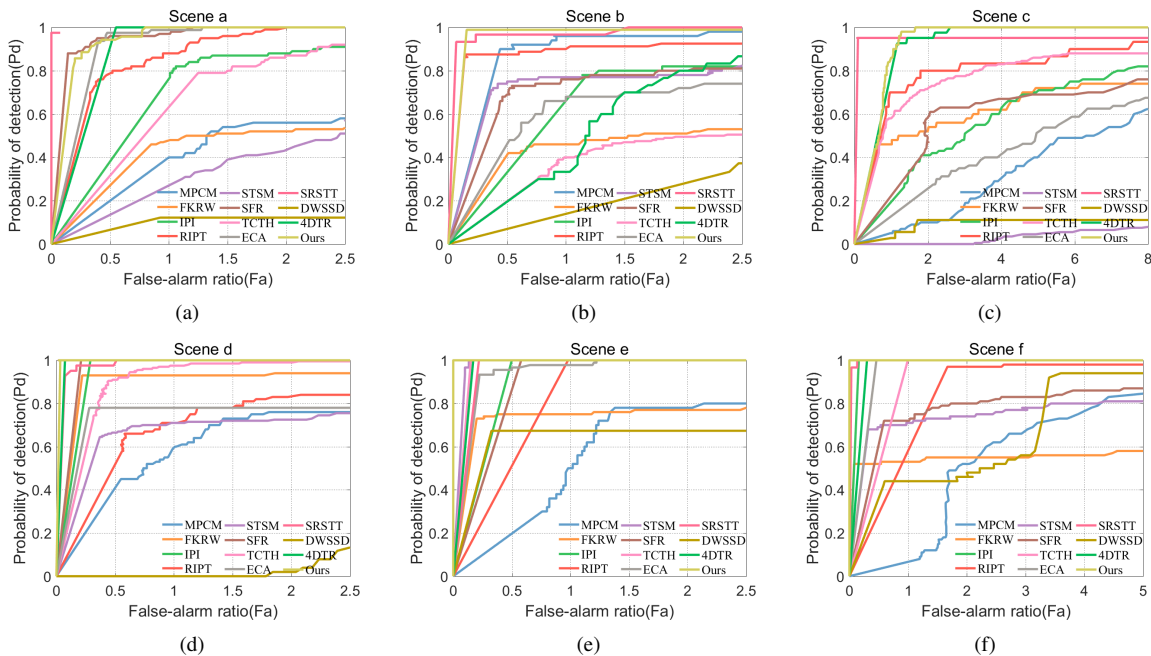
Fig. 18. ROC curves of dynamic change of P_d with F_a compared with eight competitors on scene a-f.

methods, our algorithm exhibits competitive detection efficiency. Although the proposed method is marginally slower than RIPT, NLTA, TSPK, 4DTR and DWSSD. The proposed method improves detection efficiency by at least 50% compared to similar algorithms SFR, TCTH, ECA, and SRSTT. In summary, although our method does not achieve the fastest detection efficiency, it demonstrates superior target detection and background suppression capabilities.

TABLE VII
AVERAGE RUNNING TIME (/S) ON PER FRAME OF ALL COMPARED METHODS
UNDER DIFFERENT TESTED SCENES.

Methods	Scene a	Scene b	Scene c	Scene d	Scene e	Scene f
Saliency-based Method						
WLDM[3]	1.5740	1.5322	1.6121	1.5985	1.5735	1.5624
MPCM[18]	0.1533	0.1581	0.1566	0.1676	0.1471	0.1265
FKRW[45]	0.9349	0.9794	1.056	0.9446	0.8350	0.8754
STSM[24]	0.3409	0.3527	0.3141	0.3050	0.3053	0.304
Low-rank Sparse Decomposition Method						
IPI[16]	8.634	8.646	9.449	8.893	8.101	8.423
RIPT[21]	2.511	2.117	2.485	1.815	1.943	2.155
SFR[46]	11.001	12.432	12.013	10.561	12.141	11.205
TCTH[28]	9.751	9.864	9.7519	9.573	9.704	9.67
ECA[37]	9.7297	9.7422	9.7673	9.753	9.6521	10.636
NLTA[38]	3.0719	3.3283	3.3271	3.3699	3.0874	3.0773
TSPK[32]	1.8916	1.9233	1.7931	1.6670	1.5580	1.7500
4DTR[39]	2.4384	2.4796	2.5130	2.5820	2.4374	2.4893
SRSTT[34]	14.3105	13.3884	14.7158	13.3774	13.1602	12.8718
DWSSD[36]	0.3616	0.3373	0.3756	0.3204	0.3008	0.3018
SDD	**4.0272**	**4.1055**	**4.0528**	**4.1315**	**4.0279**	**4.1326**

VI. CONCLUSION

In this study, we introduce a new potential prior, SDD, to distinguish targets from background interference. We integrate it into tensor decomposition using a mixed sparse constraint on the spatial factors' directional difference images. The temporal difference factor is regularized by a continuity constraint to leverage inter-frame correlations. Additionally, we devise a saliency enhancement coherence map, which we integrate using a reweighting strategy to enhance target saliency and reduce background interference. The proposed model is solved using a PAM-based algorithm for guaranteed numerical convergence. Extensive experiments compare our method with ten state-of-the-art approaches, demonstrating its superior performance in small target detection and background interference suppression. Remarkably, the proposed model consistently detects all targets in testing scenarios and surpasses other algorithms when allowing a false alarm rate of 1, outperforming the second-best algorithm by 10.6% in G_{SCR} and 75.1% in CG.
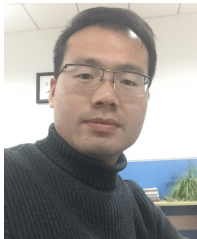
REFERENCES

[1] H. Deng, X. Sun, and X. Zhou, "A multiscale fuzzy metric for detecting small infrared targets against chaotic cloudy/sea-sky backgrounds," *IEEE Transactions on Cybernetics*, vol. 49, no. 5, pp. 1694–1707, 2019. 1

[2] X. Bai and Y. Bi, "Derivative entropy-based contrast measure for infrared small-target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2452–2466, 2018. 1

[3] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 4204–4214, 2016. 1, 7, 9, 12, 15

[4] H. Deng and Y. Zhang, "FMR-YOLO: Infrared ship rotating target detection based on synthetic fog and

multiscale weighted feature fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–17, 2024. 1

[5] R. Usamentiaga, Y. Mokhtari, C. Ibarra-Castanedo, M. Klein, M. Genest, and X. Maldague, "Automated dynamic inspection using active infrared thermography," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5648–5657, 2018. 1

[6] M. Zhang, R. Zhang, Y. Yang, H. Bai, J. Zhang, and J. Guo, "ISNet: Shape matters for infrared small target detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 867–876. 1

[7] M. Yang and X. Fan, "YOLOv8-Lit: A lightweight object detection model for real-time autonomous driving systems," *IECE Transactions on Emerging Topics in Artificial Intelligence*, vol. 1, no. 1, pp. 1–16, 2024. 1

[8] Z. Huang, P. Zhang, R. Liu, and D. Li, "An improved yolov3-based method for immature apple detection," *IECE Transactions on Internet of Things*, vol. 1, no. 1, pp. 9–14, 2023. 1

[9] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Asymmetric contextual modulation for infrared small target detection," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 949–958. 1

[10] H. Wang, L. Zhou, and L. Wang, "Miss Detection vs. False Alarm: Adversarial learning for small object segmentation in infrared images," in *IEEE International Conference on Computer Vision (ICCV)*, 2019, pp. 8508–8517.

[11] H. Fang, L. Ding, L. Wang, Y. Chang, L. Yan, and J. Han, "Infrared small uav target detection based on depthwise separable residual dense network and multiscale feature fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–20, 2022. 1

[12] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Attentional local contrast networks for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 11, pp. 9813–9824, 2021. 1, 3

[13] F. Wu, T. Zhang, L. Li, Y. Huang, and Z. Peng, "RP-CANet: Deep unfolding RPCA based infrared small target detection," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2024, pp. 4809–4818. 1

[14] Y. Dai, X. Li, F. Zhou, Y. Qian, Y. Chen, and J. Yang, "One-stage cascade refinement networks for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023.

[15] Z. Yang, T. Ma, Y. Ku, Q. Ma, and J. Fu, "DFFIR-net: Infrared dim small object detection network constrained by gray-level distribution model," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–15, 2022. 1

[16] 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 Transactions on Image Processing*, vol. 22, no. 12, pp. 4996–5009, 2013. 2, 7, 9, 12, 15

[17] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection,"

[18] Y. Wei, X. You, and H. Li, "Multiscale patch-based contrast measure for small infrared target detection," *Pattern Recognition*, vol. 58, pp. 216–226, 2016. 2, 3, 7, 9, 12, 15

[19] X. Zhang, A. Wang, Z. Yan, S. Mazhar, and Y. Chang, "A detection method with antiinterference for infrared maritime small target," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 3999–4014, 2024. 2

[20] J. Han, Y. Ma, J. Huang, X. Mei, and J. Ma, "An infrared small target detecting algorithm based on human visual system," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 452–456, 2016. 2

[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, pp. 3752–3767, 2017. 4, 7, 9, 12, 15

[22] X. Wang, Z. Peng, D. Kong, and Y. He, "Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scene," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5481–5493, 2017. 7

[23] D. Pang, T. Shan, W. Li, P. Ma, S. Liu, and R. Tao, "Infrared dim and small target detection based on greedy bilateral factorization in image sequences," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3394–3408, 2020. 2

[24] Y. Li, Y. Zhang, J. G. 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*, p. S0020025516305230, 2016. 2, 8, 9, 12, 15

[25] J. Guo, Y. Wu, and Y. Dai, "Small target detection based on reweighted infrared patch-image model," *IET Image Processing*, vol. 12, no. 1, pp. 70–79, 2017. 2

[26] Y. Dai, Y. Wu, Y. Song, and J. Guo, "Non-negative infrared patch-image model: Robust target-background separation via partial sum minimization of singular values," *Infrared Physics and Technology*, vol. 81, pp. 182–194, 2017. 2

[27] C. Gao, L. Wang, Y. Xiao, Q. Zhao, and D. Meng, "Infrared small-dim target detection based on markov random field guided noise modeling," *Pattern Recognition*, vol. 76, pp. 463–475, 2017. 2

[28] H. Zhu, S. Liu, L. Deng, Y. Li, and F. Xiao, "Infrared small target detection via low-rank tensor completion with top-hat regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1004–1016, 2020. 2, 3, 8, 9, 12, 15

[29] F. Zhou, Y. Wu, and Y. Dai, "Infrared small target detection via incorporating spatial structural prior into intrinsic tensor sparsity regularization," *Digital Signal Processing*, vol. 111, no. 5, p. 102966, 2021. 4

[30] Y. Sun, J. Yang, and W. An, "Infrared dim and small target detection via multiple subspace learning and spatial-

temporal patch-tensor model," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–16, 2020. 4

[31] H. Liu, L. Zhang, and H. Huang, "Small target detection in infrared videos based on spatio-temporal tensor model," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–12, 2020. 2, 4, 12

[32] F. Zhou, M. Fu, Y. Duan, Y. Dai, and Y. Wu, "Infrared small target detection via l_0 sparse gradient regularized tensor spectral support low-rank decomposition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 3, pp. 2105–2122, 2023. 2, 8, 9, 12, 15

[33] L. Deng, D. Xu, G. Xu, and H. Zhu, "A generalized low-rank double-tensor nuclear norm completion framework for infrared small target detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 4, pp. 3297–3312, 2022. 2

[34] J. Li, P. Zhang, L. Zhang, and Z. Zhang, "Sparse regularization-based spatial-temporal twist tensor model for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023. 2, 8, 9, 12, 15

[35] X. Wang, Z. Peng, D. Kong, and Y. He, "Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scene," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5481–5493, 2017. 2

[36] B. Dan, Z. Zhu, X. Qi, J. Zhang, Y. Ouyang, M. Li, and T. Tang, "Dynamic weight-guided smooth-sparse decomposition for small target detection against strong vignetting background," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–15, 2024. 2, 7, 9, 12, 15

[37] 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*, pp. 1–17, 2020. 2, 3, 4, 8, 9, 12, 15

[38] 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, 2022. 2, 3, 8, 9, 15

[39] 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. 2, 8, 9, 12, 15

[40] Y. Liu, X. Liu, X. Hao, W. Tang, S. Zhang, and T. Lei, "Single-frame infrared small target detection by high local variance, low-rank and sparse decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023. 2

[41] Q. Lu, Q. Li, L. Hu, and L. Huang, "An effective low-contrast sf_6 gas leakage detection method for infrared imaging," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021. 2

[42] Y. Xu, M. Wan, X. Zhang, J. Wu, Y. Chen, Q. Chen, and G. Gu, "Infrared small target detection based on local contrast-weighted multidirectional derivative," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023. 2

[43] D. Liu, L. Cao, Z. Li, T. Liu, and P. Che, "Infrared small target detection based on flux density and direction diversity in gradient vector field," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 7, pp. 2528–2554, 2018. 2

[44] J. Liu, J. Zhang, Y. Wei, and L. Zhang, "Infrared small target detection based on multidirectional gradient," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023. 2

[45] Y. Qin, L. Bruzzone, C. Gao, and B. Li, "Infrared small target detection based on facet kernel and random walker," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7104–7118, 2019. 2, 7, 9, 12, 15

[46] D. Pang, P. Ma, T. Shan, W. Li, R. Tao, Y. Ma, and T. Wang, "STTM-SFR: Spatial-temporal tensor modeling with saliency filter regularization for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022. 3, 4, 8, 9, 12, 15

[47] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009. 3

[48] 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. 4

[49] 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. 4

[50] X. Kong, C. Yang, S. Cao, C. Li, and Z. Peng, "Infrared small target detection via nonconvex tensor fibered rank approximation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. pp, no. 99, pp. 1–21, 2021. 4

[51] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022. 4

[52] L. Yang, P. Yan, M. Li, J. Zhang, and Z. Xu, "Infrared small target detection based on a group image-patch tensor model," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022. 4

[53] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022. 4

[54] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022. 4

[55] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Transactions on Cybernetics*, vol. 50, no. 8, pp. 3556–3570, 2020. 4

[56] S. Moradi, P. Moallem, and M. F. Sabahi, "Fast and robust small infrared target detection using absolute directional mean difference algorithm," *Signal Processing*, vol. 177, p. 107727, 2020. 4

[57] Y. Li, Z. Li, Y. Shen, and J. Li, "Infrared small target detection based on 1-D difference of guided filtering," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[58] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, Y. Chen, and W. He, "Double-factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8450–8464, 2020. 4

[59] V. B. S. Prasath, D. Vorotnikov, R. Pelapur, S. Jose, G. Seetharaman, and K. Palaniappan, "Multiscale Tikhonov-total variation image restoration using spatially varying edge coherence exponent," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5220–5235, 2015. 5

[60] C. Vicas and S. Nedevschi, "Detecting curvilinear features using structure tensors," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3874–3887, 2015. 5

[61] A. Akl, C. Yaacoub, M. Donias, J.-P. Da Costa, and C. Germain, "Texture synthesis using the structure tensor," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4082–4095, 2015. 5

[62] P. Kirrinnis, "Fast algorithms for the sylvester equation ax- xbt= c," *Theoretical Computer Science*, vol. 259, no. 1-2, pp. 623–638, 2001. 6

[63] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011. 6

[64] B. Hui, Z. Song, H. Fan, P. Zhong, W. Hu, X. Zhang, J. Ling, H. Su, W. Jin, Y. Zhang, and Y. Bai, "A dataset for infrared detection and tracking of dim-small aircraft targets under ground / air background," *Science Data Bank*, vol. 5, no. 3, 2020. 7

Maixia Fu received the B.S. degree in electronic science and technology from Information Engineering University, Zhengzhou, China, in 2003, the M.S. degree in information and communication engineering from Zhengzhou University, Zhengzhou, China, in 2009, and the Ph.D. degree in optical engineering from Capital Normal University, Beijing, China, in 2017. She is currently an Associate Professor and master's Supervisor with the College of Information Science and Engineering, Henan University of Technology, Zhengzhou, China. Her research interests include signal processing, spectrum analysis, optoelectronic devices designing, and remote sensing

Yulei Qian received his PhD degree from College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics in 2020. He is now an engineer in Nanjing Marine Radar Institute. His main research interests include target detection, radar imaging, satellite synthetic aperture radar, sparse recovery, deconvolution and radar waveform design. He has published 10+ paper in remote sensing journals and conferences.

Jian Yang received the PhD degree from Nanjing University of Science and Technology (NJUST) in 2002, majoring in pattern recognition and intelligence systems. From 2003 to 2007, he was a Postdoctoral Fellow at the University of Zaragoza, Hong Kong Polytechnic University and New Jersey Institute of Technology, respectively. From 2007 to present, he is a professor in the School of Computer Science and Technology of NJUST. Currently, he is also a visiting distinguished professor in the College of Computer Science of Nankai University. He is the author of more than 300 scientific papers in pattern recognition and computer vision. His papers have been cited over 40000 times in the Scholar Google. His research interests include pattern recognition and computer vision. Currently, he is/was an associate editor of Pattern Recognition, Pattern Recognition Letters, IEEE Trans. Neural Networks and Learning Systems, and Neurocomputing. He is a Fellow of IAPR.

Fei Zhou received his M.S. degree in electronic engineering from Xinjiang University in 2017, and received Ph.D. degree from College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics. He is currently a lecturer with College of Information Science and Engineering, Henan University of Technology, Zhengzhou, China. His main research interests are signal processing, target detection and image processing. He has published 10+ papers in journals and conferences, such as TGRS, TAES, WACV, etc.

Yimian Dai received the Ph.D. degree in signal and information processing from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2021, during which he honed his research skills as a visiting Ph.D. student at the University of Copenhagen and the University of Arizona between March 2018 and October 2020. Since 2021, he has been with the School of Computer Science and Engineering, Nanjing University of Science and Technology (NJUST), Nanjing, where he is currently an Assistant Researcher. His research interests include remote sensing, computer vision, and deep learning, with a focus on developing algorithms for object detection, data assimilation, and computational imaging to tackle real-world challenges. He has authored more than 20 peer-reviewed journal and conference papers such as IJCV, TGRS, TAES, etc.