IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 March 2024, accepted 18 March 2024, date of publication 21 March 2024, date of current version 29 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3380164

RESEARCH ARTICLE

Infrared Object Detection Based on Improved Twist Tensor Model

GAOSHAN FENG[1], **WENLIN QIN**[2], **ENYONG XU**[1], **ZIJUN SUN**[3,4], **XIANGSUO FAN**[2,4], **AND HUAJIN CHEN**[1,3,4,5]

[1]Dongfeng Liuzhou Motor Company Ltd., Liuzhou 545005, China
[2]School of Automation, Guangxi University of Science and Technology, Liuzhou 545006, China
[3]School of Electronic Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China
[4]Guangxi Key Laboratory of Multidimensional Information Fusion for Intelligent Vehicles, Liuzhou 545006, China
[5]School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Huajin Chen (hjchen@gxust.edu.cn)

This work was supported by Guangxi Science and Technology Project under Grant 2023GXNSFFA026002, Grant 2021GXNSFDA196001, Grant 2022AC21108, Grant AD22080042, and Grant AB21220052.

ABSTRACT Infrared object detection holds significant importance in automatic target search and tracking system under complex background. The conventional structural tensor models have not harnessed the full potential of spatio-temporal domain information in sequence scenes, and the strong edge contours in the image often lead to false alarms. In order to tackle this problem, we propose an improved twist tensor model based on the optimization of background constraints. Firstly, we propose a diffusion function according to the gradient difference between the target and backgrounds, which preserves the target signal to a great extent. Secondly, the spatio-temporal information of the sequence images is used to construct the twist tensor objective constraint optimization function, and the improved twist tensor effectively distinguishes the sparse components of the target and the low-rank components of the background. Finally, the optimized model is solved using ADMM to obtain the final target signal. Eight sequence images and nine comparison methods are performed for experimental validation, after improvement, the mean SSIM value reaches 0.9921, the mean BSF value attains 126.1710, and the detection rate also surpasses 85%, experiment results demonstrate that the proposed algorithm can effectively suppress the complex background while retaining the target well.

INDEX TERMS Complex background, twist tensor, background constraint, object detection.

I. INTRODUCTION

The utilization of infrared object detection extends to various fields, including military, medical, and aerospace. Visible light-based detection tends to perform poorly in detecting surrounding obstacles during nighttime driving, making infrared imaging technology a preferred choice due to its all-weather capabilities and strong resistance to interference. Infrared object detection is extensively utilized in applications like forest fire monitoring, wildlife tracking, and unmanned vehicles for pedestrian and vehicle detection [1], [2]. Rapid advancements in imaging techniques and complex and dynamic background in sequence scenes pose increasing challenges in object detection and tracking. Over the past few years, numerous scholars have dedicated their efforts to infrared object detection and have achieved significant results. These detection algorithms can mainly be categorized into conventional methods and deep learning approaches [3].

Traditional object detection algorithms include frame difference methods [4], [5], background modeling methods [6], [7], and detection methods based on image features. Frame difference methods detect targets by subtracting two consecutive frames to highlight differences in target positions. For instance, Gao proposed a motion object detection algorithm that integrates the three-frame differencing method with an optimized hybrid Gaussian background model. This method effectively suppresses interference from sea ripples on object detection by detecting changing regions and using adaptive learning strategies. Experimental results demonstrate its strong performance [4]. Li et al. introduced an infrared

The associate editor coordinating the review of this manuscript and approving it for publication was Cesar Vargas-Rosales.

motion object detection algorithm that combines morphological top-hat transformation, frame difference methods, adaptive region growing, and grayscale-level adaptive threshold segmentation. This approach suppresses background interference and allows for the extraction and detection of single and multiple infrared moving targets [5].

Background modeling methods predict the grayscale value of a central pixel by using the grayscale distribution between target and its corresponding neighbor regions. The background prediction image is then subtracted from the original image, a residual image only containing the target and a few noise points is then generated. This process is usually followed by threshold segmentation to accomplish object detection. For instance, He and Huang proposed a moving object detection algorithm based on Top-Hat filtering results. They achieved high-performance and speed improvements for detecting moving small targets using grayscale morphology and background modeling. Compared to classical algorithms such as MoG2 and ViBe+, this method demonstrates significant improvements in both performance and speed [6]. To address issues related to noise, breakpoints, and internal holes in motion object detection using the three-frame difference method, Ding and Lu introduced the Vibe background modeling method that incorporates color and edge features. This approach combines the Vibe algorithm with an enhanced three-frame difference method for real-time motion object detection. Experiments indicate a significant improvement in the performance of this approach compared to the traditional three-frame difference algorithm [7].

In addition to grayscale and motion features, differentiation between targets and backgrounds can be achieved through gradient and texture features, as well as scale invariance. In regions where the gradient difference is substantial, typically associated with target areas, gradients can be used to extract target edges. Then target regions can be filled to enhance target extraction. For instance, Lu and Chen proposed a method for detecting weak small targets based on gradient feature extraction. This method effectively suppresses the background and enhances the targets, resulting in significantly improved detection rates and background suppression capabilities compared to other algorithms [8]. Zhang and Zhang introduced a prominent object detection algorithm based on texture and color features. This algorithm combines Gabor filters for texture feature extraction, color contrast, and Bayesian enhancement methods. Experimental results demonstrate superior comprehensive performance compared to other algorithms in various aspects [9]. Luo and Liu presented an infrared object detection algorithm based on fast spectral scale space, dynamic pipeline filtering, and visual saliency analysis. They used global and local saliency analysis and Gaussian difference theory to detect weak small moving targets efficiently. The algorithm exhibits excellent detection performance in complex background conditions [10].

With advancements in computer vision technology, object detection methods based on convolutional neural networks have gained significant attention from experts and scholars. Unlike traditional manual feature extraction, these methods leverage convolutional neural networks (CNNs) to automatically extract object features, and they train on samples to achieve automatic detection. For instance, Acremont proposed a compact fully convolutional neural network(cfCNN), with global average pooling (GAP). Through testing on various datasets, the results demonstrated that GAP contributes to enhanced robustness in handling interrupted inputs [11]. Xu et al. introduced a deep learning-based infrared object detection framework called TF-SSD, which uses transposed convolutions to enhance feature extraction and detection efficiency. They improved the network structure through a visual approach and implemented a multi-scale feature fusion model to establish a connection between high-level and low-level networks. Experiments showed that the TF-SSD effectively identifies infrared targets at various flight attitudes, achieving a high level of detection accuracy [12]. Ma et al. developed an infrared small object detection network called GLFM, featuring scale-adaptive feature extraction and a multi-layer joint up-sampled feature mapping network for sparse feature extraction and background suppression. They introduced a 2D Gaussian label generation strategy during model training to address sample imbalance issue. Experimental results demonstrated that the network effectively detects infrared objects in various complex backgrounds, regardless of size and low SNR, and exhibited better performance and robustness [13]. Dai et al. proposed a novel deep network for infrared small object detection, combining a discriminative network with traditional model-driven approaches. They emphasized and preserved small target features through feature map cyclic movement and bottom-up attention adjustment. Through ablation studies and performance comparisons, the effectiveness and efficiency of this network architecture were confirmed [14]. Yang et al. introduced two models, CMF Net and CMF-3DLSTM, to address the challenges of target multi-scale feature extraction and occlusion issues. CMF Net, based on the VGG16 network, utilized a multi-scale feature extraction mechanism and fused low-level visual features with high-level semantic features. On the other hand, CMF-3DLSTM improved the classification network by employing a 3D long short-term memory (LSTM) network to tackle target occlusion. By integrating multi-scale and contextual features through an attention mechanism, these models effectively leveraged spatiotemporal features [15].

However, deep learning-based methods for object detection rely heavily on extensive training data. Current datasets often lack the diversity to cover a wide range of scenes. When the scene changes, retraining the model with new samples becomes necessary, posing challenges to real-time applicability. Additionally, the lack of comprehensive datasets hinders the progress of deep learning. Based on a comprehensive analysis of prior research efforts, this paper presents a novel infrared object detection approach that combines spatiotemporal information from sequence images with the low-rank sparsity theory. The main contributions of this paper are as follows:

(1) The background-constrained optimization algorithm we presented in this paper aims to enhance the low-rank properties of the backgrounds. Leveraging the advantages of anisotropy for better background representation, a new diffusion equation is proposed, further improving the constraint capability on the background.

(2) Building upon the background-constrained optimization, this paper combines spatio-temporal information from sequence images to construct the twist tensor model. By changing the observation perspective, this model effectively suppresses strong edge contours in the images, enabling the detection and extraction of targets.

(3) Finally, we solve the improved twist tensor model by ADMM to extract the final target signal components and realize the object detection.

II. METHODS

A. TWIST TENSOR MODEL

In traditional infrared object detection methods, single-frame detection approaches rely solely on the spatial information of the target. These methods perform well in the background with relatively stable gray levels. However, in real-world scenes, image backgrounds are becoming increasingly complex, presenting challenges such as high-intensity backgrounds, random noise, and low-contrast targets. Conventional multi-frame methods combine the spatiotemporal information of the object, which can effectively improve the detection rate of the algorithm. However, compared to single-frame detection methods, they also require longer detection times and may lead to varying degrees of missed detections in some scenes. To address the issue of infrared object detection in the presence of strong clutter interference in the complex background and to incorporate both spatial and temporal information, a model based on spatiotemporal twist tensor was constructed, as described in reference [16]. This model constrains the low-rank background to achieve infrared object detection. The expression based on tensor is as follow:

$$D = B + T + N \tag{1}$$

In the equation above, $D, B, T, N \in R^{n_1 * n_2 * n_3}$, D denotes the original tensor, B is the background tensor, Trepresents the target tensor and N denotes the noise tensor. The twist tensor model transforms the original front view slices into side view slices, allowing for a change in the viewing perspective, which can reduce some unnecessary interference. The non-target sparse components in the front view slices of the sequential images, when transformed through a change in perspective, no longer exhibit sparse characteristics in non-local areas. This reduces the complexity of the image background, making it more conducive to object detection. Assuming that X is a structure tensor of size $n_1 * n_2 * n_3$, rotating the direction of X to obtain the twist tensor $Y = \overrightarrow{X}$, and the size of Y is $n_1 * n_3 * n_2$, i.e., $Y(:, k, :) = X(:, :, k)$. To this end, the detection model based on the twist tensor is obtained as follow:

$$\overrightarrow{D} = \overrightarrow{B} + \overrightarrow{T} + \overrightarrow{N} \tag{2}$$

$\overrightarrow{D}$ represents original twist tensor, $\overrightarrow{B}$ denotes background twist tensor, $\overrightarrow{T}$ denotes the target twist tensor and $\overrightarrow{N}$ represents the noise twist tensor. The corresponding object function is as follow:

$$\begin{cases} \min\limits_{\overrightarrow{B}, \overrightarrow{T}, \overrightarrow{N}} \; rank(\overrightarrow{B}) + \lambda_1 \left\| \overrightarrow{T} \right\|_1 + \lambda_2 \left\| \overrightarrow{N} \right\|_F^2 \\ \text{s.t. } \overrightarrow{D} = \overrightarrow{B} + \overrightarrow{T} + \overrightarrow{N} \end{cases} \tag{3}$$

λ_1 represents sparsity weights and λ_2 represents noise weights. To more effectively and comprehensively represent the target, a joint regularization strategy is employed, combining the l_1 norm with structured sparsity-inducing norms. Considering that utilizing the tensor nuclear norm (TNN) based on the discrete cosine transform as a linear transformation yields superior results compared to the tensor nuclear norm based on the discrete Fourier transform, a linear transformation constraint-induced tensor nuclear norm is applied to the infrared background. Therefore, the object function is updated as follow:

$$\begin{cases} \min\limits_{\overrightarrow{B}, \overrightarrow{T}, \overrightarrow{N}} \left\| \overrightarrow{B} \right\|_{\text{TNNL}} + \lambda_1 \left\| \overrightarrow{T} \right\|_{l_1/l_\infty} + \lambda_2 \left\| \overrightarrow{T} \right\|_1 + \lambda_3 \left\| \overrightarrow{N} \right\|_F^2 \\ \text{s.t. } \overrightarrow{D} = \overrightarrow{B} + \overrightarrow{T} + \overrightarrow{N} \end{cases} \tag{4}$$

B. TWIST TENSOR MODEL BASED ON BACKGROUND CONSTRAINT OPTIMIZATION

1) BACKGROUND CONSTRAINT OPTIMIZATION MODEL

When there is prominent high-intensity noise or sparse strong edge contours in the image, the low-rank characteristic of the background is disrupted, which often leads to an increased number of false alarms in object detection. Considering that anisotropy has better representation capability in terms of background features, this paper is based on anisotropy to construct a more similar background tensor model, thus enhancing the discrimination between the target and background for more effective object detection. Anisotropy was initially applied to background modeling of weak infrared small targets. Since weak infrared small targets occupy only a small number of pixels, based on the different grayscale distribution of the different regions in the images, the target could be separated from the smooth background. In contrast to weak small infrared targets, pedestrian target regions are relatively large. If the diffusion functions and step sizes employed in the large object detection as same as using in weak and small object detection, some of the target information will be lost. The key to anisotropy lies in the selection of the diffusion equation and step size, as the calculation of diffusion coefficients depends on the construction of the diffusion function. If the step size is too large, more background information will be retained, which can interfere with subsequent object detection. If the step size is too small, some target information will be smoothed out just like the smooth background. Zhang and Ling respectively improved the diffusion functions and proposed new diffusion functions for background modeling of weak small targets under a smooth background [17], [18]. These diffusion functions achieved good results in those scenes. However, when applying them to infrared pedestrian detection, the varying scale of the targets results in a decrease in background modeling effectiveness. Therefore, based on an analysis of previous work and taking advantage of the characteristics of anisotropy, this paper improves a background-constrained optimization model to further enhance its ability for background constraint. The specific approach is detailed in the following expressions.

The first step is to compute the gradient between pixels, and in this paper we chose 3 as a moving step size.

$$\begin{cases} \nabla_{up} = H(i,j) - H(i-l,j) \\ \nabla_{down} = H(i,j) - H(i+l,j) \\ \nabla_{left} = H(i,j) - H(i,j-l) \\ \nabla_{right} = H(i,j) - H(i,j+l) \end{cases} \tag{5}$$

$H(i,j)$ denotes the center pixel position, l denotes the moving step, ∇_{up}, ∇_{down}, ∇_{left}, ∇_{right} denotes the gradient in the direction of the four neighbors of the center pixel, respectively. The improved diffusion function in this paper is as follow:

$$C = \frac{e^{-(\nabla f/k)^2}}{e^{-(\nabla f/k)^2} + 1} \tag{6}$$

∇_f denotes the gradient, k denotes a constant, and in this paper $k = 120$ was chosen in the experiment. Substituting the gradient into the diffusion equation to calculate the diffusion coefficients:

$$\begin{cases} c_u = (e^{-(\nabla_{up}/k)^2})/(e^{-(\nabla_{up}/k)^2} + 1) \\ c_d = (e^{-(\nabla_{down}/k)^2})/(e^{-(\nabla_{down}/k)^2} + 1) \\ c_l = (e^{-(\nabla_{left}/k)^2})/(e^{-(\nabla_{left}/k)^2} + 1) \\ c_r = (e^{-(\nabla_{right}/k)^2})/(e^{-(\nabla_{right}/k)^2} + 1) \end{cases} \tag{7}$$

c_u, c_d, c_l, c_r denotes the diffusion coefficients corresponding to the gradients in the four directions. According to Equation (8), taking the reciprocal of the diffusion coefficient and multiplying it with the corresponding gradient, then averaging them, the background-constrained optimization function is obtained.

$$\|B\|_{\text{ANIS}} = \frac{(\frac{1}{c_u} * \nabla_{up} + \frac{1}{c_d} * \nabla_{down} + \frac{1}{c_l} * \nabla_{left} + \frac{1}{c_{right}} * \nabla_{right})}{4} \tag{8}$$

In the above equation, due to the large scale of the infrared target, small diffusion weight value may lead to the loss of target information. Therefore, in this article, the calculation of diffusion coefficient is taken as the inverse of the diffusion coefficient to obtain a larger weight to highlight the target signal. Here, $\|B\|_{\text{ANIS}}$ represents the background-constrained optimization function.

2) IMPROVED TWIST TENSOR MODEL(ITTM)

The improved background-constrained optimization model rewrites the objective function to better describe the degree of feature correlation among backgrounds and enhance the low-rank properties of backgrounds. The specific improved model is as follow:

$$\begin{cases} \min\limits_{\overrightarrow{B},\overrightarrow{T},\overrightarrow{N}} \left\|\overrightarrow{B}\right\|_{\text{ANIS}} + \lambda_1 \left\|\overrightarrow{T}\right\|_{l_1/l_\infty} + \lambda_2 \left\|\overrightarrow{T}\right\|_1 + \lambda_3 \left\|\overrightarrow{N}\right\|_F^2 \\ \text{s.t. } \overrightarrow{D} = \overrightarrow{B} + \overrightarrow{T} + \overrightarrow{N} \end{cases} \tag{9}$$

The above equation shows the improved twist tensor model in this paper, and auxiliary variables $\overrightarrow{Z}$ and $\overrightarrow{S}$ are introduced to facilitate the solution of $\overrightarrow{T}$. The target function is rewritten as the following expression (10), as shown at the bottom of the next page.

The corresponding augmented Lagrangian function of the above equation could be expressed as follow:

$$\begin{aligned} L(\overrightarrow{B}, \overrightarrow{T}, \overrightarrow{N}, \overrightarrow{Z}, \overrightarrow{S}) &= \left\|\overrightarrow{B}\right\|_{\text{ANIS}} + \lambda_1 \left\|\overrightarrow{Z}\right\|_1 + \lambda_2 \left\|\overrightarrow{S}\right\|_{l_1/\infty} \\ &\quad + \lambda_3 \left\|\overrightarrow{N}\right\|_F^2 + \left\langle y_1, \overrightarrow{Z} - \overrightarrow{T}\right\rangle + \left\langle y_2, \overrightarrow{S} - \overrightarrow{T}\right\rangle \\ &\quad + \left\langle y_3, \overrightarrow{D} - \overrightarrow{B} - \overrightarrow{T} - \overrightarrow{N}\right\rangle \\ &\quad + \frac{\mu}{2}(\left\|\overrightarrow{Z} - \overrightarrow{T}\right\|_F^2 + \left\|\overrightarrow{S} - \overrightarrow{T}\right\|_F^2 \\ &\quad + \left\|\overrightarrow{D} - \overrightarrow{B} - \overrightarrow{T} - \overrightarrow{N}\right\|_F^2) \end{aligned} \tag{11}$$

$\mu > 0$ is a penalty parameter. By solving the object function using the Alternating Direction Method of Multipliers (ADMM), we can reconstruct the target twist tensor $\overrightarrow{T}$. When compressing $\overrightarrow{T}$ back to the frontal view, the corresponding frontal slice represents the detected target image. Equation (11) is decomposed into a number of sub-problems, one sub-problem corresponds to one variable, all the variables are updated in each round of iteration, and the other variables should be kept unchanged when solving one of them, and the specific solution process is as follow:

(1) The solution of $\overrightarrow{B}^{k+1}$

$$\begin{aligned} \overrightarrow{B}^{k+1} &= \arg\min_{\overrightarrow{B}} \left\|\overrightarrow{B}\right\|_{\text{ANIS}} + \left\langle y_3^k, \overrightarrow{D} - \overrightarrow{B} - \overrightarrow{T}^k - \overrightarrow{N}^k\right\rangle \\ &\quad + \frac{\mu^k}{2}\left\|\overrightarrow{D} - \overrightarrow{B} - \overrightarrow{T}^k - \overrightarrow{N}^k\right\|_F^2 \end{aligned}$$

$$= \arg\min_{\overrightarrow{B}} \left\| \overrightarrow{B} \right\|_{\text{ANIS}} + \frac{\mu^k}{2} \left\| \overrightarrow{B} - (\overrightarrow{D} - \overrightarrow{T}^k - \overrightarrow{N}^k + \frac{y_3^k}{\mu^k}) \right\|_F^2 \tag{12}$$

$\overrightarrow{B}^{k+1}$ can be solved by the tensor singular value threshold operator $D_\tau(\cdot)$ [19], the resulting solution is:

$$\overrightarrow{B}^{k+1} = D_{\frac{1}{\mu^k}}(\overrightarrow{D} - \overrightarrow{T}^k - \overrightarrow{N}^k + \frac{y_3^k}{\mu^k}) \tag{13}$$

(2) The solution of $\overrightarrow{Z}^{k+1}$

$$\begin{aligned}\overrightarrow{Z}^{k+1} &= \arg\min_{\overrightarrow{Z}} \lambda_1 \left\| \overrightarrow{Z} \right\|_1 + \left\langle y_1^k, \overrightarrow{Z} - \overrightarrow{T} \right\rangle \\ &\quad + \frac{\mu^k}{2} \left\| \overrightarrow{Z} - \overrightarrow{T} \right\|_F^2 \\ &= \arg\min_{\overrightarrow{Z}} \lambda_1 \left\| \overrightarrow{Z} \right\|_1 + \frac{\mu^k}{2} \left\| \overrightarrow{Z} - \overrightarrow{T}^k + \frac{y_1^k}{\mu^k} \right\|_F^2 \end{aligned} \tag{14}$$

$\overrightarrow{Z}^{k+1}$ can be solved by the soft threshold operator $\text{Soft}_\tau(\cdot)$ [20], the resulting solution is:

$$\overrightarrow{Z}^{k+1} = \text{Soft}_{\frac{\lambda_1}{\mu^k}}(\overrightarrow{T}^k - \frac{y_1^k}{\mu^k}) \tag{15}$$

(3) The solution of $\overrightarrow{S}^{k+1}$

$$\begin{aligned}\overrightarrow{S}^{k+1} &= \arg\min_{\overrightarrow{S}} \lambda_2 \left\| \overrightarrow{S} \right\| l_1/l_\infty + \left\langle y_2^k, \overrightarrow{S} - \overrightarrow{T}^k \right\rangle \\ &\quad + \frac{\mu^k}{2} \left\| \overrightarrow{S} - \overrightarrow{T}^k \right\|_F^2 = \arg\min_{\overrightarrow{S}} \lambda_2 \left\| \overrightarrow{S} \right\| l_1/l_\infty \\ &\quad + \frac{\mu^k}{2} \left\| \overrightarrow{S} - \overrightarrow{T}^k + \frac{y_2^k}{\mu^k} \right\|_F^2 \end{aligned} \tag{16}$$

$\overrightarrow{S}^{k+1}$ can be solved by the proximal operator $\text{Prox}_g(\cdot)$ [21], the resulting solution is:

$$\overrightarrow{S}^{k+1} = \text{Prox}_g(\overrightarrow{T} - \frac{y_2^k}{\mu^k}) \tag{17}$$

(4) The solution of $\overrightarrow{T}^{k+1}$

$$\begin{aligned}\overrightarrow{T}^{k+1} &= \arg\min_{\overrightarrow{T}} \left\langle y_1^k, \overrightarrow{Z}^{k+1} - \overrightarrow{T} \right\rangle + \left\langle y_2^k, \overrightarrow{S}^{k+1} - \overrightarrow{T} \right\rangle \\ &\quad + \left\langle y_3^k, \overrightarrow{D} - \overrightarrow{B}^{k+1} - \overrightarrow{T} - \overrightarrow{N}^k \right\rangle \\ &\quad + \frac{\mu^k}{2} \left\| \overrightarrow{Z}^{k+1} - \overrightarrow{T} \right\|_F^2 \\ &\quad + \frac{\mu^k}{2} \left\| \overrightarrow{S}^{k+1} - \overrightarrow{T} \right\|_F^2 \\ &\quad + \frac{\mu^k}{2} \left\| \overrightarrow{D} - \overrightarrow{B}^{k+1} - \overrightarrow{T} - \overrightarrow{N}^k \right\|_F^2 \end{aligned} \tag{18}$$

Derivation of $\overrightarrow{T}$ makes the equation equal to 0, and the resulting solution is:

$$\begin{aligned}\overrightarrow{T}^{k+1} &= \frac{1}{3\mu^k}(y_1^k + y_2^k + y_3^k) \\ &\quad + \frac{1}{3}(\overrightarrow{Z}^{k+1} + \overrightarrow{S}^{k+1} + \overrightarrow{D} - \overrightarrow{B}^{k+1} - \overrightarrow{N}^k) \end{aligned} \tag{19}$$

(5) The solution of $\overrightarrow{N}^{k+1}$

$$\begin{aligned}\overrightarrow{N}^{k+1} &= \arg\min_{\overrightarrow{N}} \lambda_3 \left\| \overrightarrow{N} \right\|_F^2 \\ &\quad + \left\langle y_3^k, \overrightarrow{D} - \overrightarrow{B}^{k+1} - \overrightarrow{T}^{k+1} - \overrightarrow{N} \right\rangle \\ &\quad + \frac{\mu^k}{2} \left\| \overrightarrow{D} - \overrightarrow{B}^{k+1} - \overrightarrow{T}^{k+1} - \overrightarrow{N} \right\|_F^2 \end{aligned} \tag{20}$$

Derivation of $\overrightarrow{N}$ makes the equation equal to 0, and the resulting solution is:

$$\overrightarrow{N}^{k+1} = \frac{y_3^k + \mu^k(\overrightarrow{D} - \overrightarrow{B}^{k+1} - \overrightarrow{T}^{k+1})}{\mu^k + 2\lambda_3} \tag{21}$$

(6) The solution of $y_i^{k+1}(i = 1, 2, 3)$ and μ^{k+1}

$$\begin{cases} y_1^{k+1} = y_1^k + u^k(\overrightarrow{Z}^{k+1} - \overrightarrow{T}^{k+1}) \\ y_2^{k+1} = y_2^k + u^k(\overrightarrow{S}^{k+1} - \overrightarrow{T}^{k+1}) \\ y_3^{k+1} = y_3^k + u^k(\overrightarrow{D} - \overrightarrow{B}^{k+1} - \overrightarrow{T}^{k+1} - \overrightarrow{N}^{k+1}) \end{cases} \tag{22}$$

$$\mu^{k+1} = \mu^k * \rho \tag{23}$$

The above describes the solution process of the improved twist tensor model(ITTM) we proposed in the article. In Figure 1, (A1) and (A2) are detection results of the pre-improved twist tensor model and their corresponding 3D diagrams, respectively. (B1) and (B2) are object detection

$$\begin{cases} \min_{\overrightarrow{B}, \overrightarrow{T}, \overrightarrow{N}, \overrightarrow{Z}, \overrightarrow{S}} \left\| \overrightarrow{B} \right\|_{\text{ANIS}} + \lambda_1 \left\| \overrightarrow{Z} \right\|_1 + \lambda_2 \left\| \overrightarrow{S} \right\|_{l_1/l_\infty} + \lambda_3 \left\| \overrightarrow{N} \right\|_F^2 \\ \text{s.t. } \overrightarrow{Z} = \overrightarrow{T} \\ \overrightarrow{S} = \overrightarrow{T} \\ \overrightarrow{D} = \overrightarrow{B} + \overrightarrow{T} + \overrightarrow{N} \end{cases} \tag{10}$$

results of the ITTM based on the background constraint optimization model and their corresponding 3D diagrams in this paper. The pre-improved twist tensor model, despite removing most of the background clutters, enhances the signal intensity of the objects. However, for targets with little difference in gray level compared to the background, strong edge contours result in a large number of false alarms due to the disruption of the background low-rank property in the image structure. After applying the background constraint optimization algorithm we presented to further restrict the background and then solving with the twist tensor model, the ITTM has a stronger background suppression ability, eliminating background clutter interference and achieving infrared object detection with low false alarms. Table 1 shows the pseudo-code for the solving process of our algorithm.

TABLE 1. The model solving process.

```
Input: $\overrightarrow{D}$, $\lambda_1$, $\lambda_2$,$\lambda_3$; Ouput:$B$,$T$,$N$
1. Initialization: $\overrightarrow{B}^0 = \overrightarrow{D}$, $\overrightarrow{T}^0 = \overrightarrow{N}^0 = 0$, $y_1^0 = y_2^0 = y_3^0 = 0$, $\mu^0 = 1e^{-3}$, $\varepsilon = 1e^{-7}$, $\rho = 1.5$, $k = 0$
2. Update $\|B\|_{\mathrm{ANIS}}$ according to Formula (5)-(8).
3. While not converged do
4. Update $\overrightarrow{B}^{k+1}$ according to Formula (12).
5. Update $\overrightarrow{Z}^{k+1}$ according to Formula (14).
6. Update $\overrightarrow{S}^{k+1}$ according to Formula (16).
7. Update $\overrightarrow{T}^{k+1}$ according to Formula (18).
8. Update $\overrightarrow{N}^{k+1}$ according to Formula (20).
9. Update $y_i^{k+1}(i = 1,2,3)$ and $\mu^{k+1}$ according to Formula (22)-(23).
10. Determine if it has converged: $\left(\left\|\overrightarrow{D} - \overrightarrow{B}^{k+1} - \overrightarrow{T}^{k+1} - \overrightarrow{N}^{k+1}\right\|_F / \left\|\overrightarrow{D}\right\|_F\right) \leq \varepsilon$
11. $k = k + 1$
12. End while
13. $B = squeeze(\overrightarrow{B}^k)$, $T = squeeze(\overrightarrow{T}^k)$, $N = squeeze(\overrightarrow{N}^k)$
```

III. EXPERIMENT

A. EXPERIMENTAL SETTING

The previous section introduced specific methodology we proposed. This section focuses on introducing the datasets used in the experiments [22], more information shows in Table 2. Furthermore, to validate the effectiveness and feasibility of our algorithm, we compared it with nine commonly used infrared object detection algorithms, including bilateral Filtering [23], multiscale gray difference weighted entropy(MGDWE) [24], top-hat transform(Top-hat) [25], partial sum of the tensor nuclear norm(PSTNN) [26], absolute directional mean difference(ADMD) [27], nonconvex tensor fibered rank approximation(NTFRA) [28], and tri-layer template local difference measure(TLLDM) [29], YOLO v5 [30] and YOLO v7 [31], the parameter settings are shown in Table 3.Three commonly used evaluation metrics for object detection are selected for evaluation, which are background structural similarity (SSIM), background suppression factor (BSF), detection rate P_d, and false alarm rate P_F, the related formulas are as follows [22]:

$$SSIM = \frac{(2\mu_R\mu_F + \varepsilon_1)(2\sigma_{RF} + \varepsilon_2)}{(\mu_R^2 + \mu_F^2 + \varepsilon_1)(\sigma_R^2 + \sigma_F^2 + \varepsilon_2)} \tag{24}$$

$$BSF = \sigma_{in}/\sigma_{out} \tag{25}$$

$$\begin{cases} P_d = \dfrac{NTDT}{NT} \times 100\% \\ P_f = \dfrac{NFDT}{NP} \times 100\% \end{cases} \tag{26}$$

where μ_R denotes mean value of the original image, σ_R represents standard deviation of the original image. $\sigma_R F$ denotes the covariance between the original image and the background modeling image. ε_1 and ε_2 are small constants to ensure that the denominator is not 0. $\sigma_i n$ denotes standard deviation of the input image, $\sigma_o ut$ represents standard deviation of the difference map. BSF denotes background suppression factor. $NTDT$ represents the number of detected targets, and $NFDT$ denotes the number of false alarm, NT denotes the total number of real targets in the sequence scene, NP represents the sum of all pixels in the sequence scene.

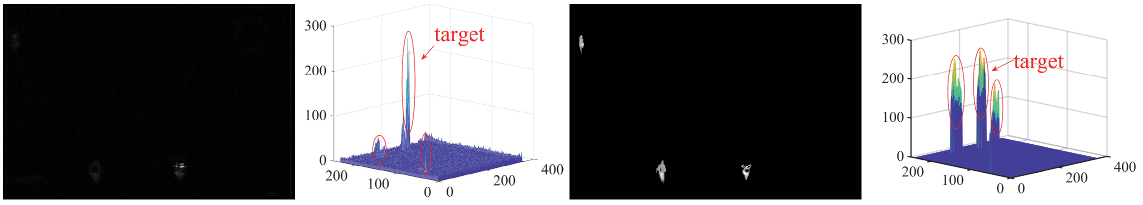
In Figure 2, scene 1 involves three pedestrians as targets, and the targets have significant differences from the background, making them easy to recognize. Scene 2 features six pedestrians as targets, similar to Scene 1, with significant differences between the targets and the background. However, some parts of the background are similar to the target features, making recognition more challenging. In Scene 3, there are two pedestrians as targets, but the background includes a car, making the target recognition process susceptible to interference. Scene 4 targets are three pedestrians, but one of them is similar to the background interference. Scene 5 targets are five easily recognizable pedestrians. Scene 6 targets are three people, one of which is almost submerged in the background. The targets in the Scene 7 are three airplanes, sometimes the targets are submerged by the clouds. In Scene 8, the target is an airplane, there is serious noises in the image.

TABLE 2. Detailed information of the sequences, includes the size of the target, the resolution and the number of images, and the types of the target.

	Target Size (pixel)	Resolution	Frame	Target description
Scene1	20*20	360*240	31	pedestrians
Scene2	20*20	360*240	18	pedestrians
Scene3	20*20	360*240	23	pedestrians
Scene4	20*20	360*240	28	pedestrians
Scene5	20*20	360*240	18	pedestrians
Scene6	20*20	360*240	24	pedestrians
Scene7	3*3	641*513	100	airplanes
Scene8	3*3	278*246	114	airplanes

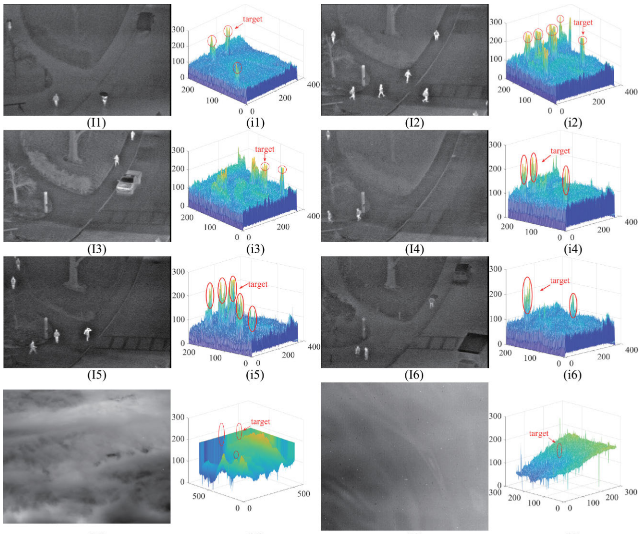
B. PARAMETER ANALYSIS

This section focuses on the analysis of the smoothing parameter K in the background constrained optimization



(A1) Pre-process (A2) (B1) Processed (B2)

FIGURE 1. Comparison of object detection effect before and after model improvement. The size of (A1) and (B1) is 360*240. (A1) result of the twist tensor model. (B1) result of the improved Twist Tensor Model(ITTM) proposed in this paper. (A2) corresponding 3D map of (A1). (B2) corresponding 3D map of (B1).



(I1) (i1) (I2) (i2)
(I3) (i3) (I4) (i4)
(I5) (i5) (I6) (i6)
(I7) (i7) (I8) (i8)

FIGURE 2. Input images and the corresponding 3D map. (I1)-(I8) is the original image, (i1)-(i8) is the corresponding 3D map of the original image, respectively. Where the image size of (I1)-(I6) is 360*240, the image size of (I7) is 641*513 and the image size of (I8) is 278*246.

TABLE 3. Parameter settings.

Algorithm	Parameter
Bilateral [23]	Filter size: 7×7, $\sigma_d = 2$, $\sigma_s = 0.1$
MGDWE [24]	Mean filter size: 7×7
Top-hat [25]	Structure size: 3×3 and 5×5
PSTNN [26]	Patch size: 40×40, sliding step: 40, $\lambda = 0.15$
ADMD [27]	$N = [3, 5, 7, 9]$
NTFRA [28]	Patch size: 40×40, sliding step: 40, $\beta = 0.01$, $\lambda = 0.1$
TLLDM [29]	Filter size: 15×15, $K = 3$
ITTM	$\lambda_1 = 0.04$, $\lambda_2 = 0.1$, $\lambda_3 = 100$, $\varepsilon = 10^{-7}$, $\mu = 0.01$

model and λ_1 in the ITTM. We introduce a new diffusion function in the background constrained optimization model, where the value of K plays a critical role. Generally, K is constrained within the ranges from 100 to 150. If K is too small, the diffusion function exhibits strong inhibition, resulting in excessive smoothing of the target. Conversely, a larger K weakens the inhibition, and the target information is retained while there are more false alarms, which interferes with the detection results. The relationship between the value of K and the detection rate P_d is depicted in Figure 3, and $K = 120$ is the optimal value. Similar to the role of K in the background optimization model, λ_1 is the regularization parameter. A larger λ_1 corresponds to a stronger smoothing ability of the background, which can easily result in the loss of target information. A smaller λ_1 is, the more false alarms will be retained, causing interference to the detection results. For this reason, different values of λ_1 are chosen to illustrate the relationship between λ_1 and detection rate P_d(as shown in Figure 4), and it is found that the detection rate P_d is the highest when $\lambda_1 = 0.04$. With the increase of λ_1, the rate P_d gradually decreases, therefore, $\lambda_1 = 0.04$ is chosen for the experiments in this paper. Furthermore, we plotted the relationship curve between λ_2 and BSF. In infrared object detection, a larger λ_2 can induce sparsity in the target region, making it easier for the model to distinguish it from the background, thus better suppressing the influence of non-target areas and highlighting the target region. Conversely, a smaller λ_2 may allow more non-target areas to be retained, including some areas that may be background or noise, potentially leading to these non-target areas being incorrectly identified as targets, thereby reducing the ability of background suppression and subsequently affecting the accuracy and reliability of object detection. From Figure 5, it can be observed that when $\lambda_2 = 0.16$, the background suppression effect reaches its optimum. At this point, both sparsity considerations and the accuracy and robustness of detection results are taken into account.

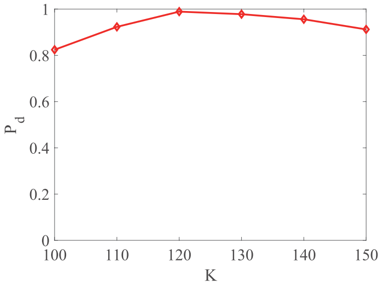


FIGURE 3. Relationship between *k* and detection rate.

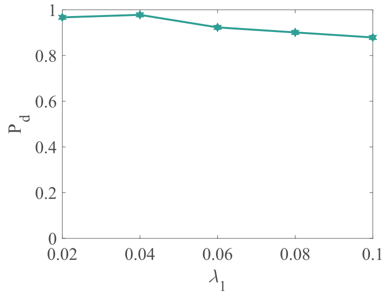


FIGURE 4. Relationship between λ_1 and detection rate.

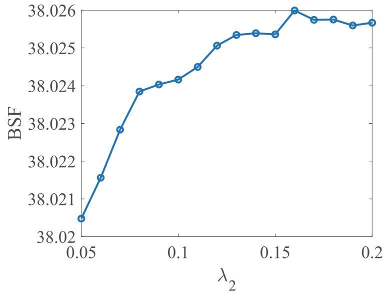


FIGURE 5. Relationship between λ_2 and *BSF*.

C. BACKGROUND MODELING EFFECT ANALYSIS

Figure 6 shows the comparison of background suppression effects for Scene 1. In this case, the details of the background prediction image obtained by bilateral filtering are blurred, and the target information retention ability in the difference map is limited. The extracted target shape is incomplete, and more clutter is retained in the 3D map. Similarly, prediction map of Top-hat filtering is poor. Although the target contour can be seen to be better retained in the difference map, there is more serious interference in the 3D map. Prediction map of MGDWE is better, with a significant increase in the target energy. TLLDM has some missing target information in the difference map, while the 3D map shows less background clutter. PSTNN, ADMD, and NTFRA have strong background clutter suppression ability, and there is almost no clutter in the 3D map. However, the strong suppression ability also weakens part of the target energy, which leads to less complete target information. The differential map obtained by the algorithm we proposed in the article has a complete shape of the target and less clutter, but the energy of the object is lower.

Figure 7 illustrates the comparison of background suppression effects for Scene 2. In this scene, the difference maps obtained by bilateral filtering and Top-hat filtering have incomplete target information, and strong edge contours are more pronounced. Additionally, the background prediction map of bilateral filtering retains more target information compared to the background prediction map of Top-hat filtering, indicating that Top-hat has better target retention ability.

The difference images of MGDWE, PSTNN, and NTFRA show complete target shapes, and their background prediction is more effective. However, the 3D map of MGDWE contain some background clutter. In the background prediction images of ADMD and TLLDM, the target information is corrupted. The difference maps have incomplete target shape contours and the background suppression ability is better without more clutter interference. The difference map obtained by the ITTM has complete target shapes, however, its suppression capability for strong edge contours is poor, and the target energy is weak.

Figure 8 compares the background suppression effects for Scene 3. Bilateral filtering, MGDWE, and Top-hat leave considerable strong edge contours in the difference maps. However, the 3D map of MGDWE contains fewer clutters. On the other hand, PSTNN, ADMD, NTFRA, and TLLDM effectively remove background clutter. Due to the similarity between the target and some background regions, the difference maps show that target information is missing, and the 3D maps retain varying degrees of clutter. The difference image of ITTM has a clearer extracted target contour. However, its ability to suppress the background with strong edge contours is slightly insufficient, and the target energy is lower.

Figure 9 shows the comparison of the background suppression effect for scene 4. Bilateral filtering target information retention ability has limitations, although the target shape is not complete enough, but it effectively removes the strong clutter interference.Top-hat, on the other hand, can better retain the target texture information, but the image also retains more background edges, which will affect the further detection work. While algorithms such as MGDWE, ADMD, PSTNN, etc., all obtain better background suppression, but these algorithms can also retain only part of the target shape, which is due to the presence of background interference in the scene that is very similar to the target, which leads to false alarms. Meanwhile, the background modeling effect of ITTM in this scene is poor, as can be clearly seen from the 3D image, only one target signal is stronger, while the other two targets are almost smoothed.

Figure 10 shows the comparison of the background suppression effect of scene 5. The contrast of the target is high in the scene, and all algorithms achieve better background modeling effect. However, due to the characteristics of the spatio-temporal filtering algorithms themselves, which result in the inability to completely remove the clutter from the image, there are still more clutter residues, such as bilateral filtering and Top-hat algorithms. In contrast, algorithms such as PSTNN, TLLDM, and NTFRA show strong background suppression ability, but it should be noted that the target texture shapes obtained by ADMD and TLLDM algorithms are incomplete. The ITTM also achieve better background results.

Figure 11 shows the comparison of the background suppression effect of scene 6. The target has low contrast and large gray span in the scene. All algorithms successfully localize the target even though the target texture shape is less complete.Top-hat and MGDWE algorithms have better background modeling effect in this scene. PSTNN, TLLDM,

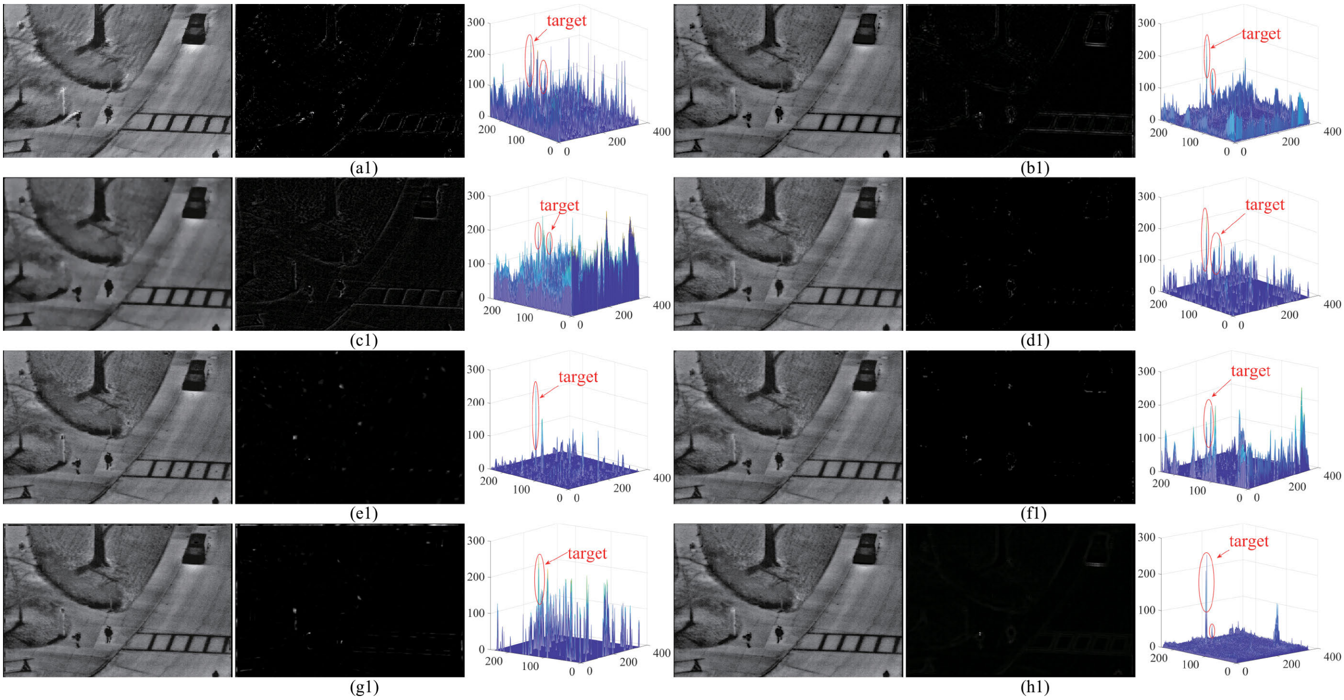


FIGURE 6. Background modeling effect of scene 1. a-h represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], and ITTM, respectively.

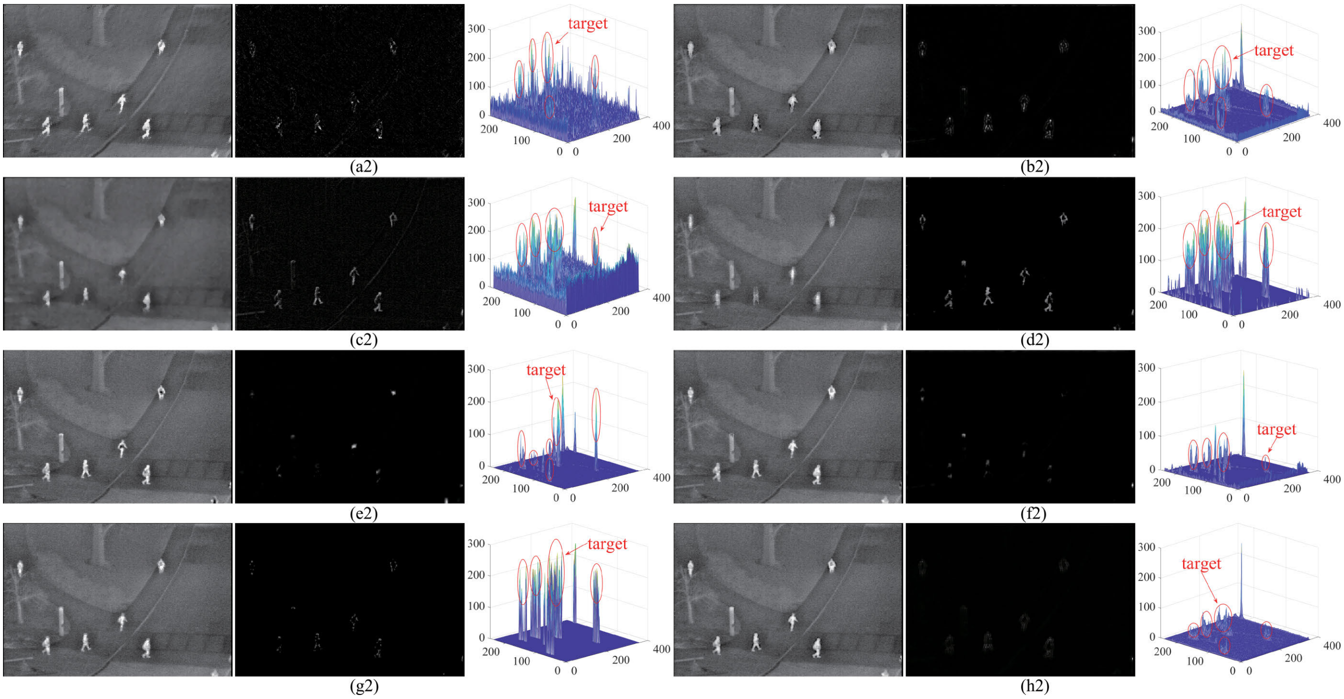


FIGURE 7. Background modeling effect of scene 2. a-h represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], and ITTM, respectively.

and NTFRA algorithms show strong background suppression ability, but there is still a small amount of noise interference. In contrast, the bilateral filtering effect is poor and the clutter interference is serious. And the algorithm of this paper has better clutter suppression effect, but some targets have lower energy.

Figure 12 presents the comparison of background suppression effects in scene 7. In this scene, the target scale is

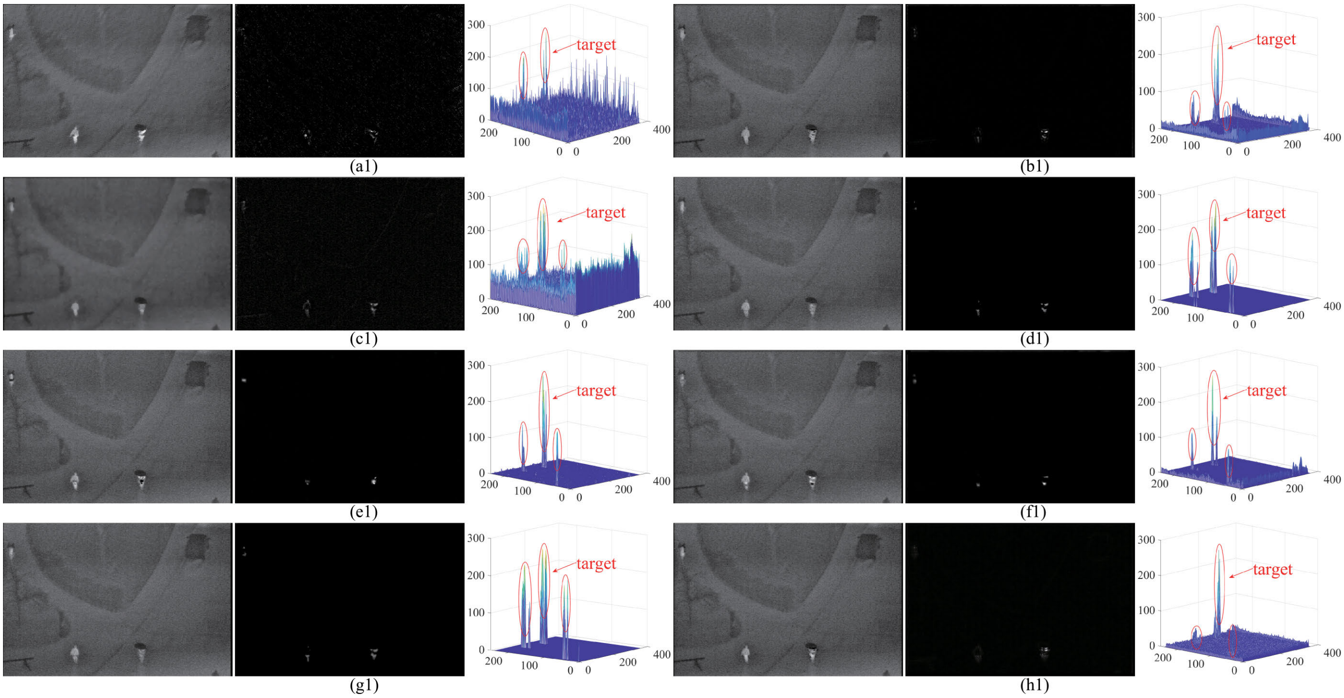


FIGURE 8. Background modeling effect of scene 3. a-h represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], and ITTM, respectively.

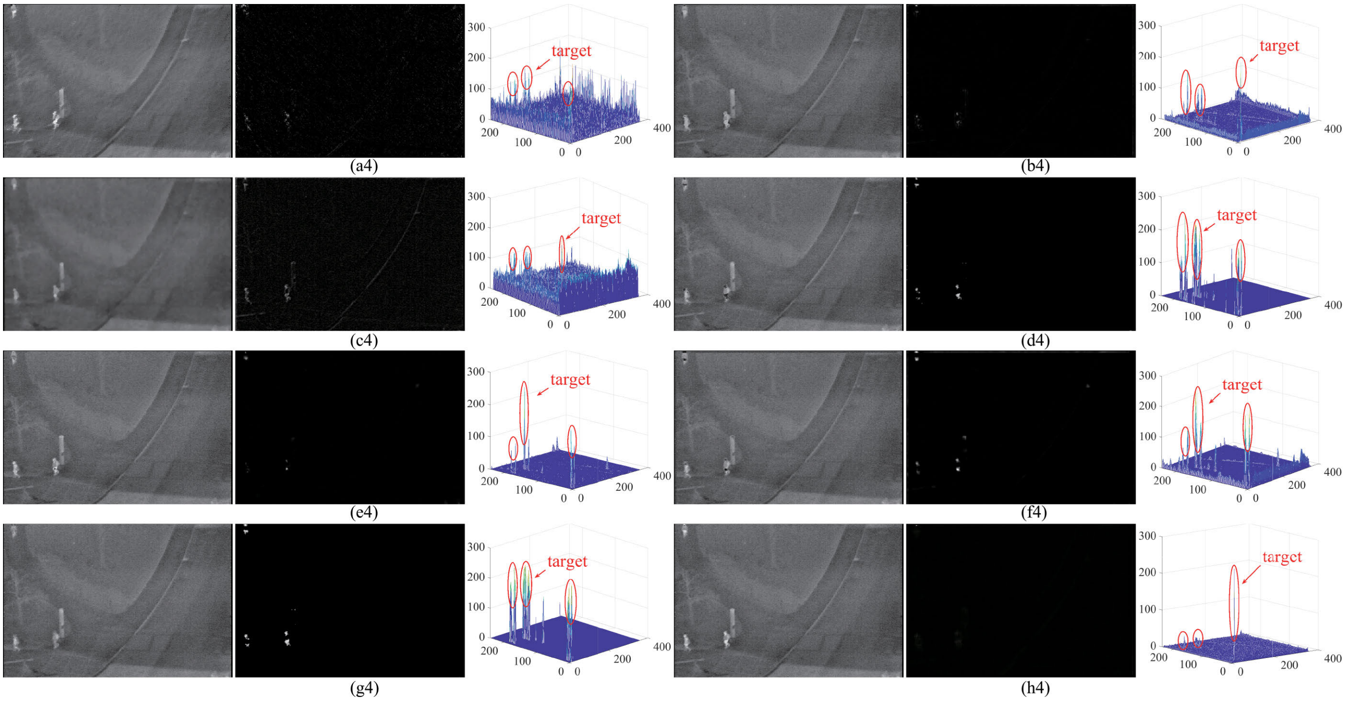


FIGURE 9. Background modeling effect of scene 4. a-h represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], and ITTM, respectively.

relatively small, and there is significant interference from cloud, which often obscures the target during its motion due to complex backgrounds. MGDWE and Top-hat algorithms exhibit poor clutter suppression performance, whereas bilateral filtering, PSTNN, ADMD, and TLLDM algorithms demonstrate better background suppression effect. However, due to the small target scale and low grayscale intensity, some algorithms tend to smooth out low-intensity targets, such as

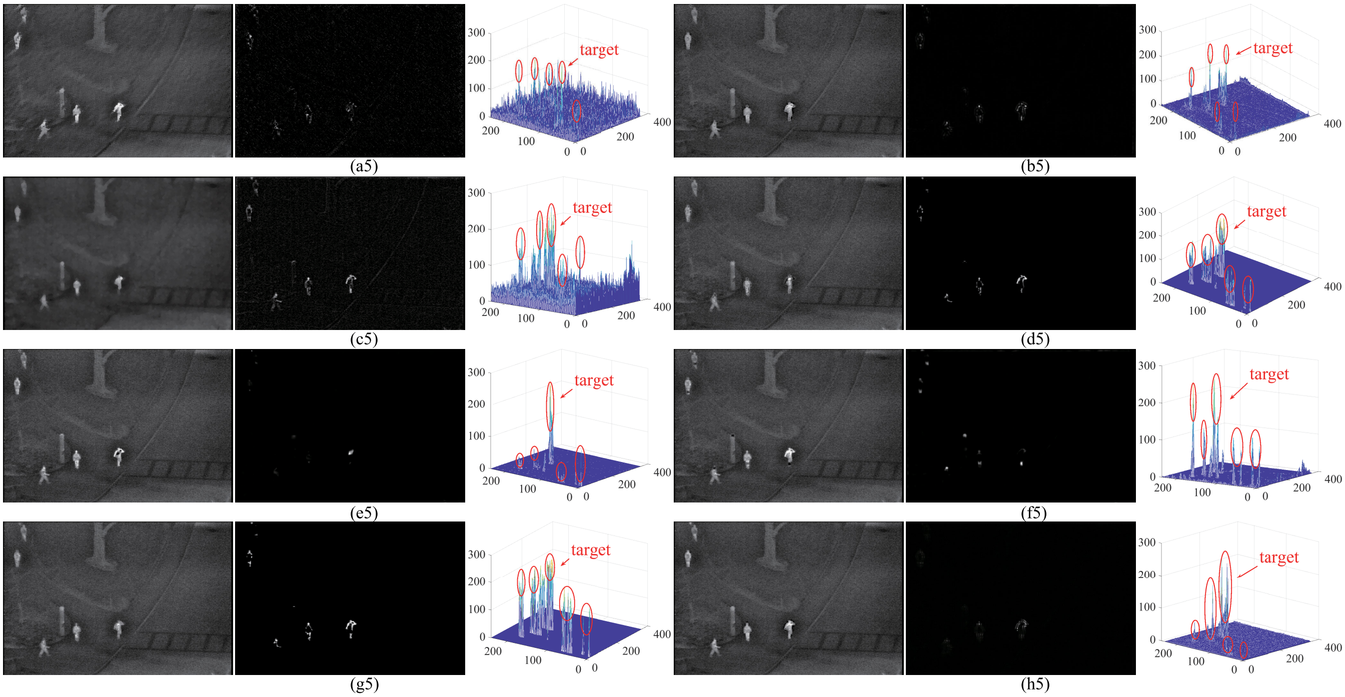


FIGURE 10. Background modeling effect of scene 5. a-h represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], and ITTM, respectively.

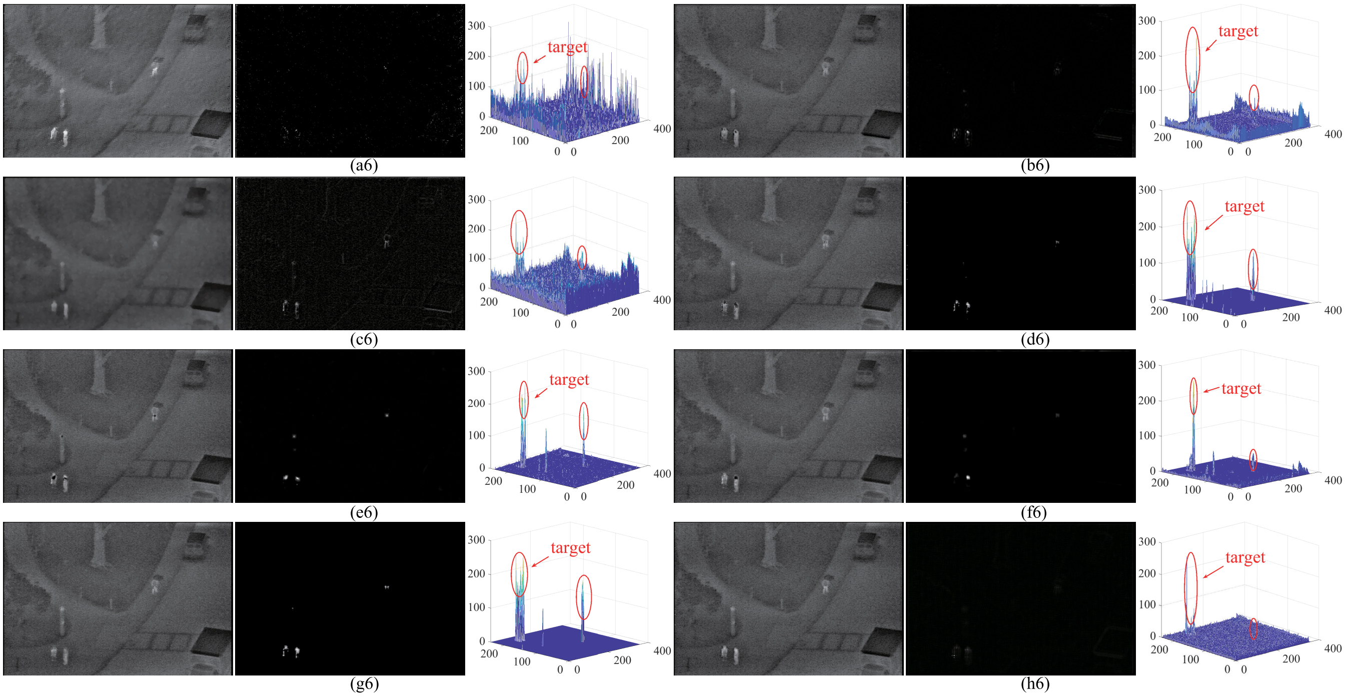


FIGURE 11. Background modeling effect of scene 6. a-h represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], and ITTM, respectively.

NTFRA failing to correctly identify targets, leaving behind substantial residual background clutter. The ITTM achieves effective background suppression in this scene, eliminating strong clutter interference while preserving the target.

Figure 13 presents the comparison of background suppression effectiveness in scene 8. In this scene, the target size is relatively small, and the image contains a significant amount of high-intensity noise, which can lead

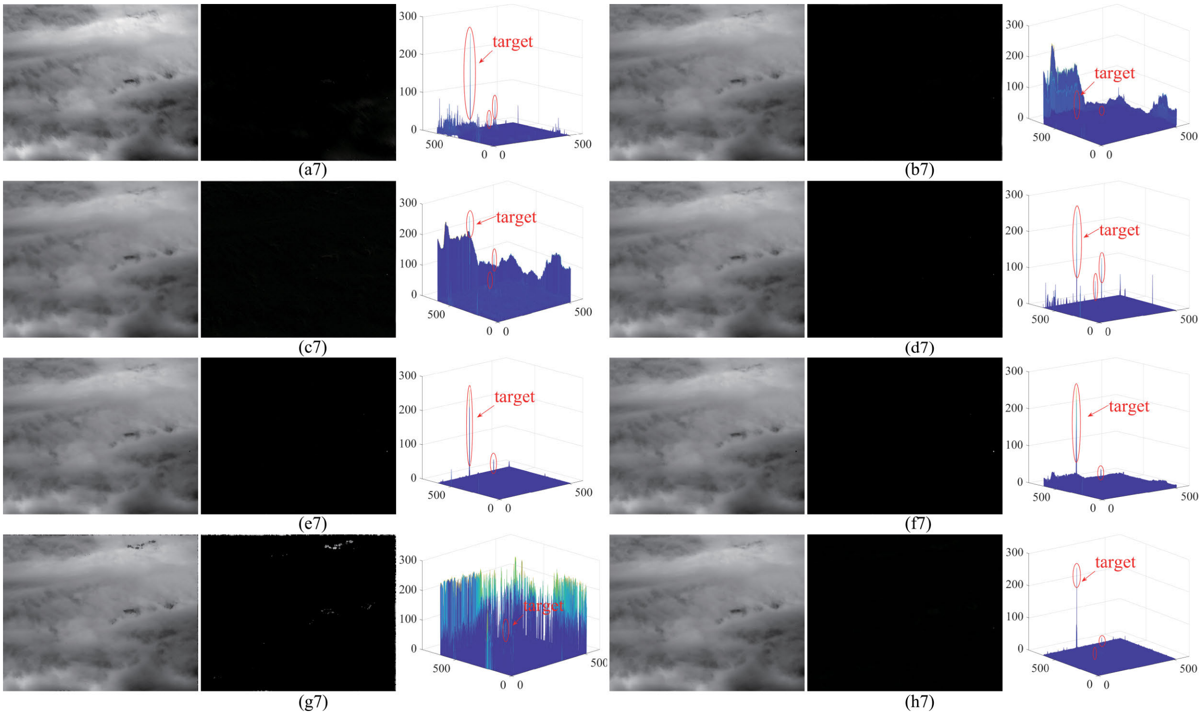


FIGURE 12. Background modeling effect of scene 7. a-h represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], and ITTM, respectively.

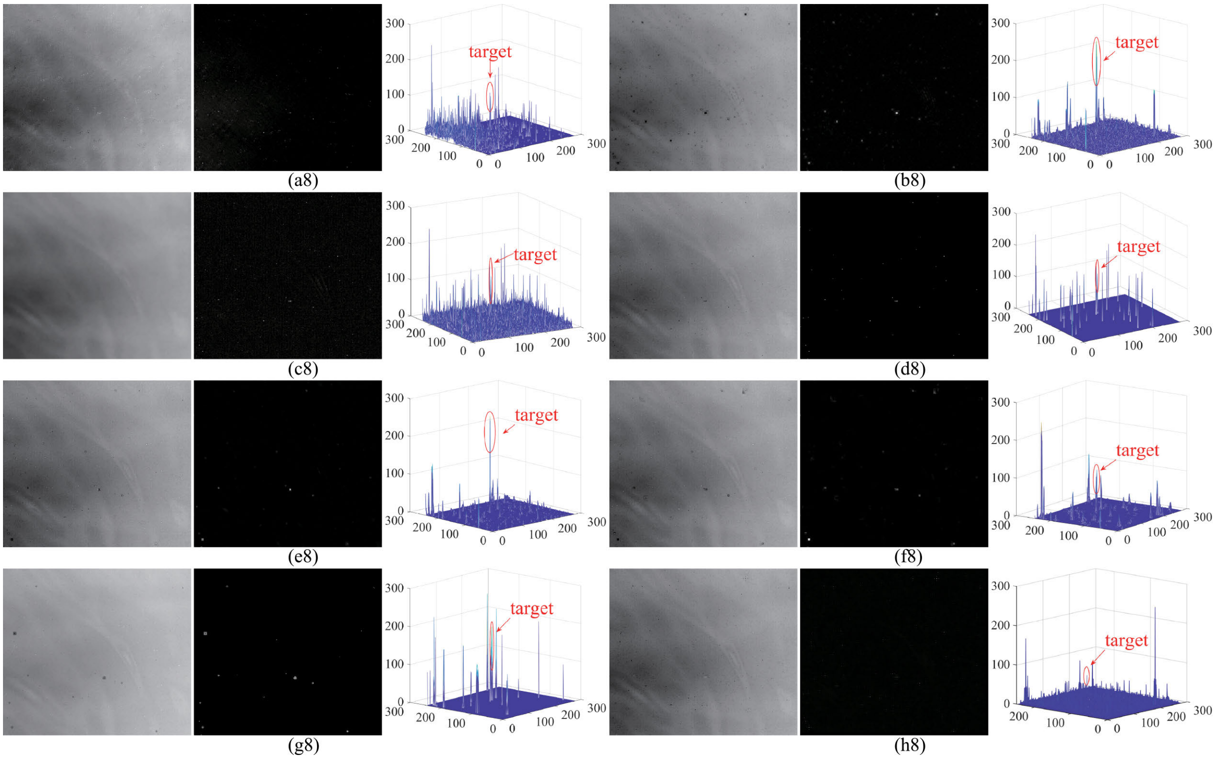


FIGURE 13. Background modeling effect of scene 8. a-h represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], and the proposed algorithm, respectively.

to issues such as misjudgment. Consequently, although all algorithms effectively preserve target information in this scene, there is also a noticeable amount of residual clutter, as observed in algorithms like bilateral filtering

TABLE 4. Evaluation indicators.

		Bilateral [23]	MGDWE [24]	Top-hat [25]	PSTNN [26]	ADMD [27]	TTLDM [29]	NTFRA [28]	ITTM
Seq.1	SSIM	0.9187	0.974	0.8102	0.98	0.9715	0.9724	0.978	**0.9819**
	BSF	17.1554	38.6854	18.9016	**50.8906**	40.9638	42.797	48.6384	50.0513
Seq.2	SSIM	0.9395	0.9611	0.8213	0.9702	0.9718	0.9841	0.9846	**0.9907**
	BSF	20.4801	34.8834	19.3015	42.1805	42.7683	57.3195	58.4531	**75.0252**
Seq.3	SSIM	0.9458	0.9622	0.8554	0.9922	0.9858	0.9587	0.9908	**0.9943**
	BSF	20.2821	31.1387	21.47	80.8062	59.7327	35.1846	74.6834	**96.4068**
Seq.4	SSIM	0.9307	0.9842	0.8303	0.9498	0.991	0.9497	0.9812	**0.9968**
	BSF	20.6147	47.6896	19.5438	32.0511	75.3148	31.8428	52.4492	**126.838**
Seq.5	SSIM	0.9129	0.9738	0.7865	0.921	0.9835	0.963	0.9669	**0.9869**
	BSF	19.2217	43.9026	18.8754	26.3824	52.3287	37.2552	40.5448	**57.2765**
Seq.6	SSIM	0.8934	0.9713	0.8109	0.9642	0.9722	0.9767	0.9869	**0.9892**
	BSF	21.5634	34.8003	19.4737	37.8306	39.9386	46.2967	62.5343	**70.1558**
Seq.7	SSIM	0.9541	0.9859	0.9853	**0.9997**	**0.9997**	0.9995	0.9838	**0.9997**
	BSF	104.3994	55.12	60.8561	396.2152	380.0704	259.7239	55.5128	**405.6008**
Seq.8	SSIM	0.9641	0.9927	0.9903	0.996	**0.9977**	0.9964	0.9951	0.9973
	BSF	45.0348	73.9085	88.2434	108.8469	120.2592	89.4316	100.5868	**128.0142**

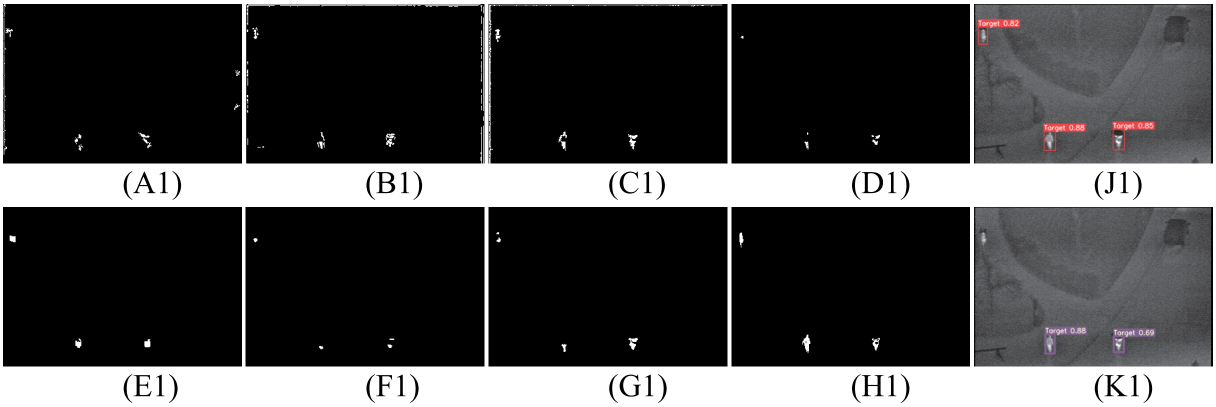


FIGURE 14. Detection result of scene 1. A-K represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], ITTM, YOLO v5 [30] and YOLO v7 [31], respectively.

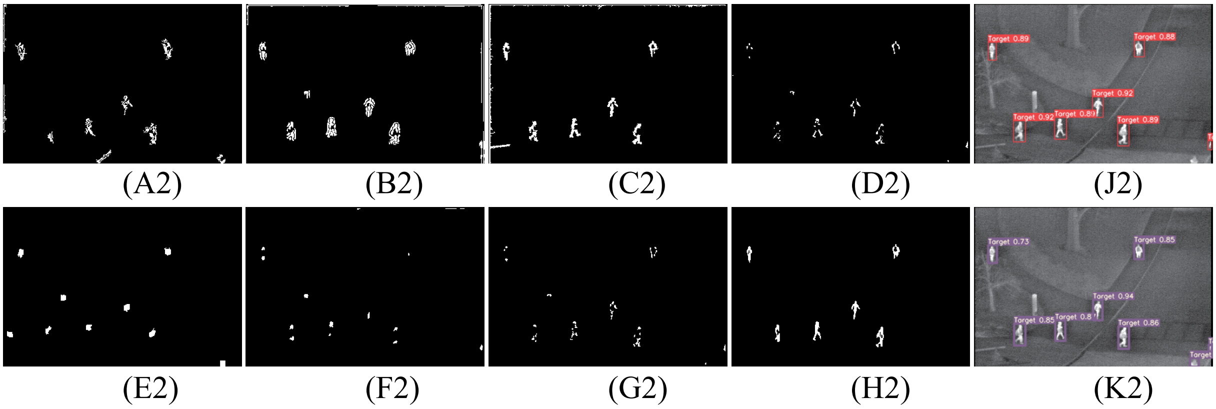


FIGURE 15. Detection result of scene 2. A-K represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], ITTM, YOLO v5 [30] and YOLO v7 [31], respectively.

and Top-hat. The object detection like YOLO v5 and YOLO v7 yield detection results directly and therefore does not involve background modeling. ITTM demonstrates strong clutter suppression capabilities, effectively reducing clutter while slightly attenuating target energy, resulting in fewer clutter artifacts. Additionally, comparative experiments on evaluation metrics have been conducted in the paper(Table 4). Except for the BSF being lower than

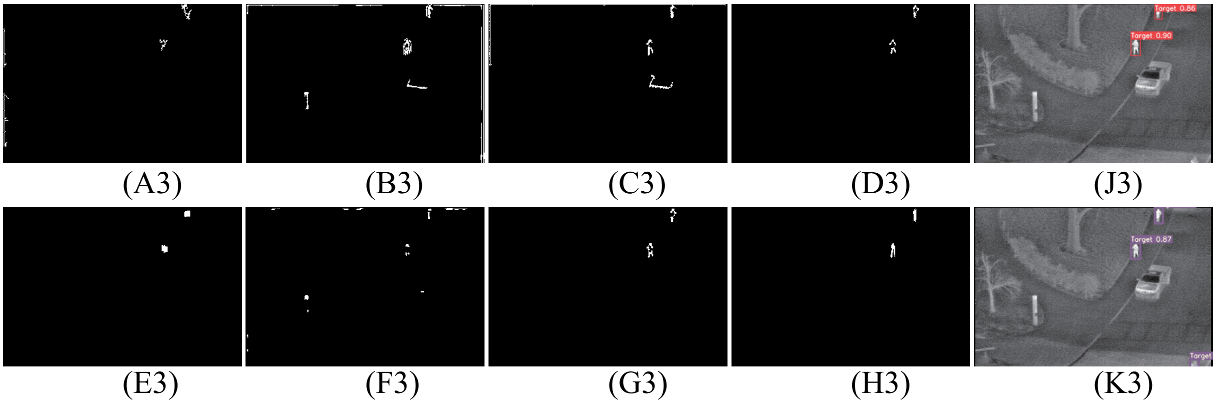


FIGURE 16. Detection result of scene 3. A-K represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], ITTM, YOLO v5 [30] and YOLO v7 [31], respectively.

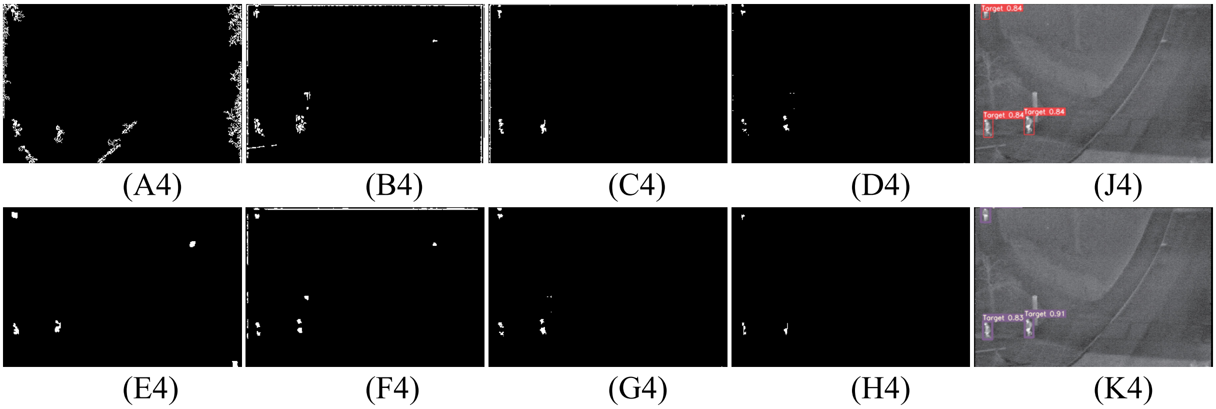


FIGURE 17. Detection result of scene 4. A-K represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], ITTM, YOLO v5 [30] and YOLO v7 [31], respectively.

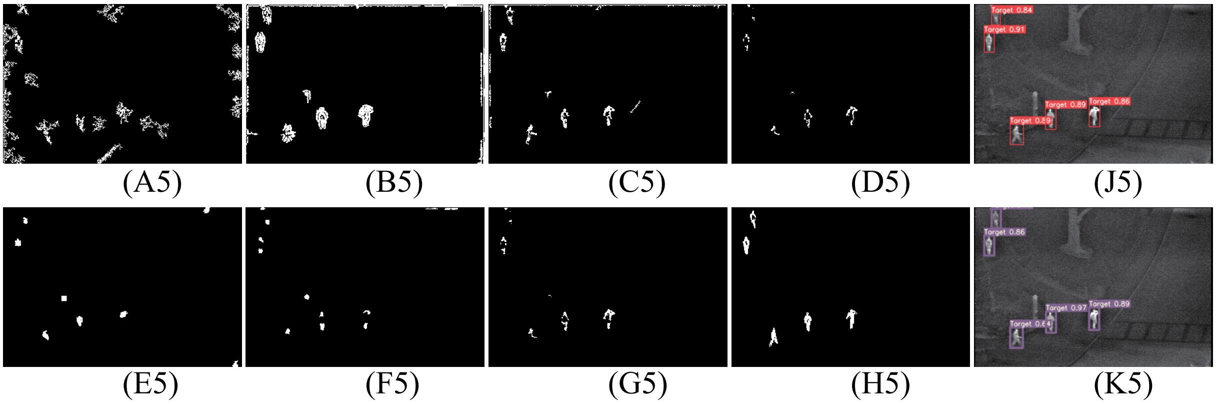


FIGURE 18. Detection result of scene 5. A-K represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], ITTM, YOLO v5 [30] and YOLO v7 [31], respectively.

PSTNN in scene 1 and SSIM being lower than the ADMD algorithm in scene 8, ITTM outperforms other comparative algorithms in terms of evaluation metrics across other scenes. Overall, the ITTM exhibits favorable background suppression effectiveness across the eight experimental scenes.

D. DETECTION RESULT

The previous section analyzed the background modeling effects, this section provides an analysis of the detection results for the ITTM and other comparison methods, the comparison methods align with Section III-B. Figure 14 depicts the detection results of Scene 1. Bilateral filtering,

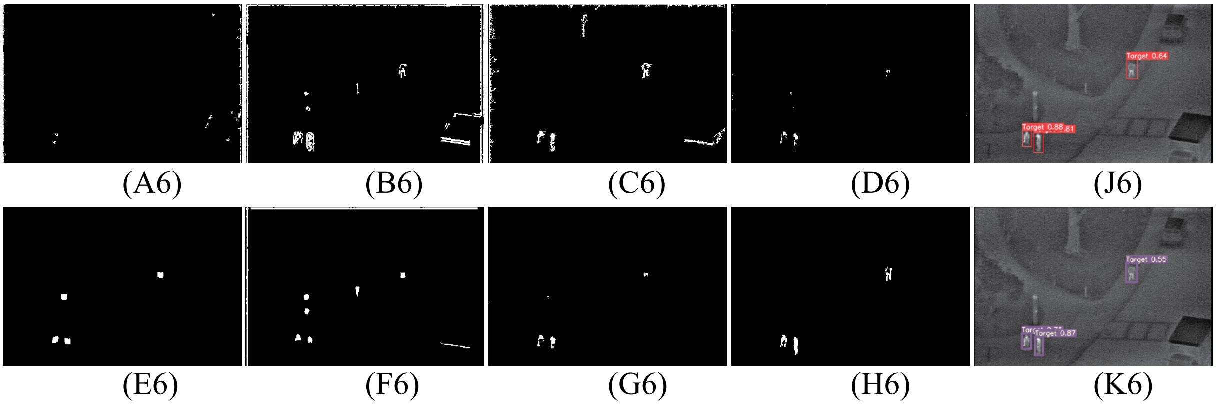


FIGURE 19. Detection result of scene 6. A-K represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], ITTM, YOLO v5 [30] and YOLO v7 [31], respectively.

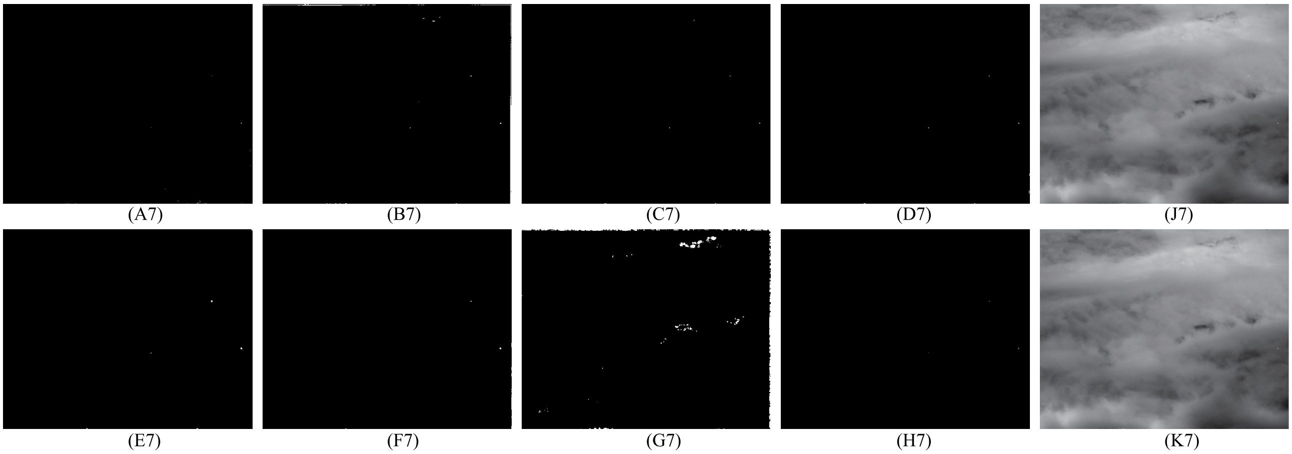


FIGURE 20. Detection result of scene 7. A-K represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], ITTM, YOLO v5 [30] and YOLO v7 [31], respectively.

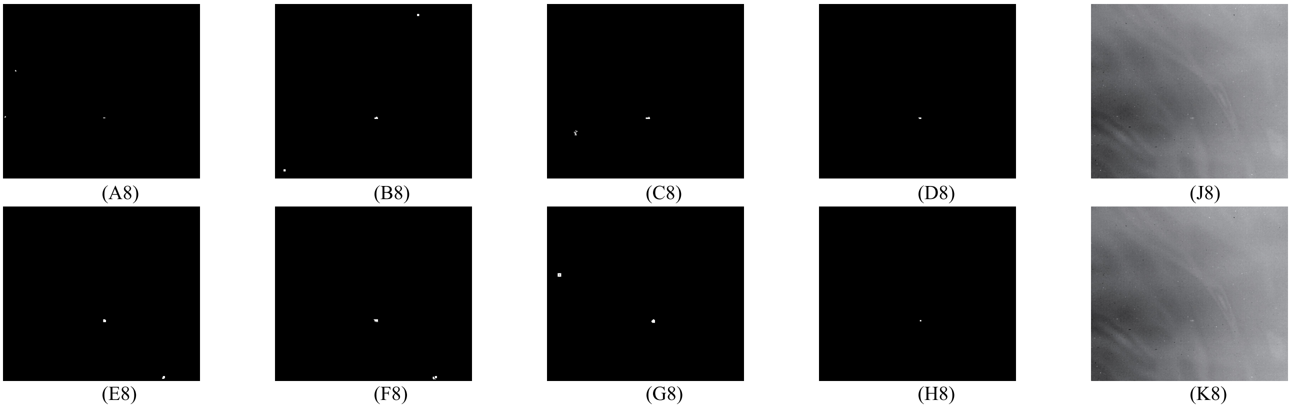
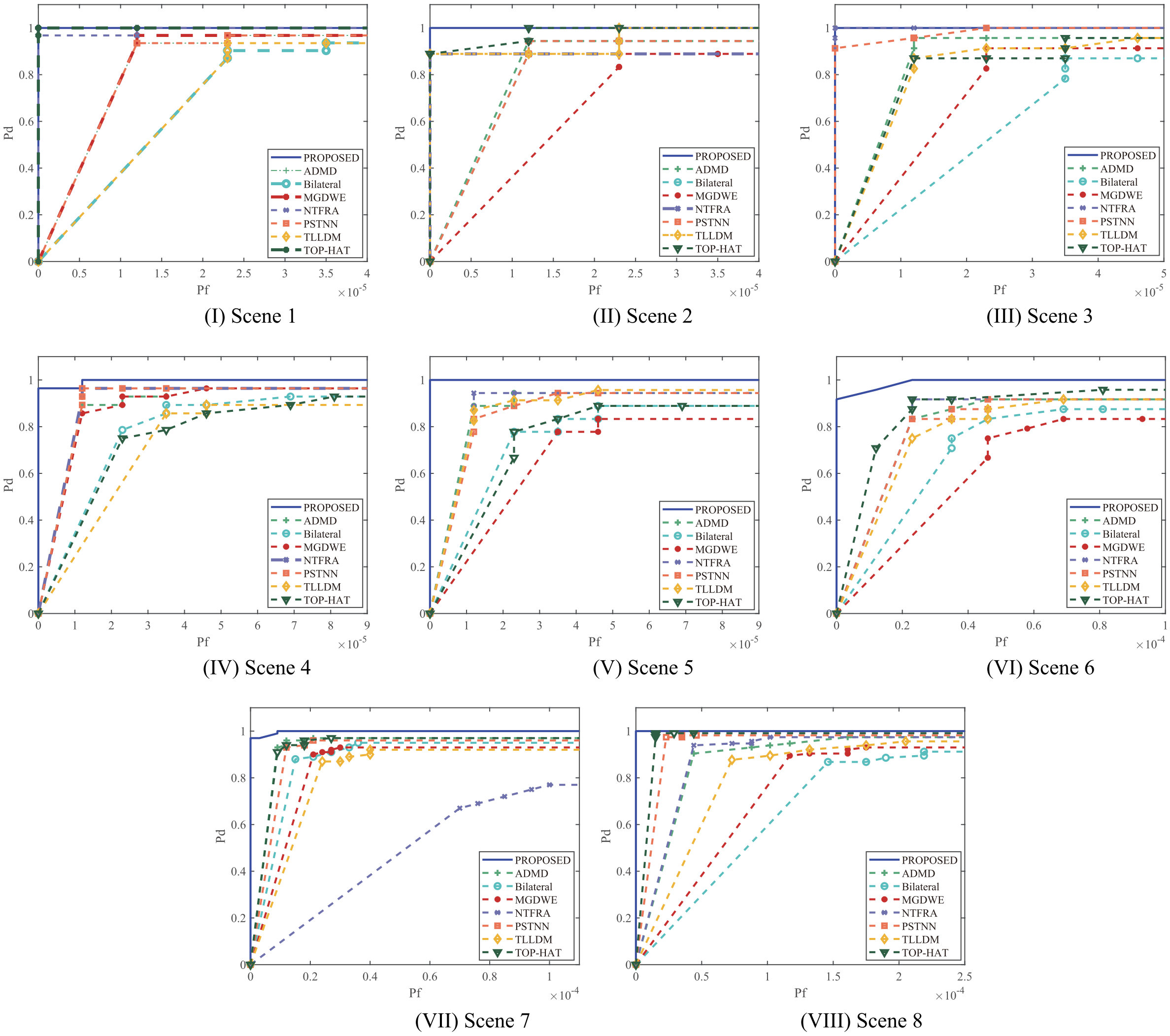


FIGURE 21. Detection result of scene 8. A-K represent bilateral filtering [23], MGDWE [24], Top-hat [25], PSTNN [26], ADMD [27], TLLDM [29], NTFRA [28], ITTM, YOLO v5 [30] and YOLO v7 [31], respectively.

MGDWE, and Top-hat filtering can detect the target well, but there is some loss of target texture information, and residual noise interference remains. On the other hand, PSTNN, ADMD, TLLDM, and NTFRA exhibit strong clutter suppression abilities, which enables them to remove more background clutter interference. However, this comes at the cost of retaining only a portion of the target information. Figure 15 displays the detection results of Scene 2. In this



(I) Scene 1

(II) Scene 2

(III) Scene 3

(IV) Scene 4

(V) Scene 5

(VI) Scene 6

(VII) Scene 7

(VIII) Scene 8

FIGURE 22. ROC curves. (I)ROC curve of scene 1, (II)ROC curve of scene 2, (III)ROC curve of scene 3, (IV)ROC curve of scene 4, (V)ROC curve of scene 5, (VI)ROC curve of scene 6, (VII)ROC curve of scene 7, (VIII)ROC curve of scene 8.

case, bilateral filtering, MGDWE, Top-hat filtering, PSTNN, and NTFRA can effectively detect the object while preserving the target's texture and shape. However, these methods still contain varying degrees of noise interference. On the contrary, ADMD and TLLDM can also detect the target but tend to have more noise interference, which can be mistaken as the real target. Figure 16 presents the detection results for Scene 3. Bilateral filtering, MGDWE, Top-hat filtering, and TLLDM can detect the target but contain minor noise interference. However, the target contour in bilateral filtering, Top-hat, and TLLDM is not as clear or complete. ADMD, PSTNN, and NTFRA can effectively eliminate noise and effectively detect the target. However, target texture shape obtained by ADMD is incomplete. Figure 17 shows the detection results of Scene 4. Bilateral filtering and MGDWE can detect the target better, but the false alarm rate is higher. Except for ADMD as well as TLLDM algorithms that leave a small amount of noise interference, the other algorithms detect the target more completely. Figure 18 shows the detection results of Scene 5, which has an obvious target, all algorithms can detect the target, but the bilateral filtering, ADMD and TLLDM algorithms have a small amount of false alarms. Figure 19 shows the detection results of Scene 6, bilateral filtering fails to detect all targets successfully, and the detection results of MGDWE and Top-hat filtering have serious clutter interference and high false alarm rate. The other comparison algorithms can detect the targets, but contain a small amount of noise in different degrees.

Figure 20 shows the detection results of Scene 7. In this scene, although there are some grayscale differences among the targets, most of the comparative algorithms can correctly identify them. Algorithms such as bilateral filtering, MGDWE, and Top-hat perform well in object detection, while PSTNN and ADMD algorithms exhibit more satisfactory detection results. In contrast, the TLLDM fails to completely detect the targets, and the NTFRA

TABLE 5. Execution efficiency of different algorithms in a frame(Seq.8).

	Bilateral [23]	MGDWE [24]	Top-hat [25]	PSTNN [26]	ADMD [27]	TTLDM [29]	NTFRA [28]	YOLO v5 [28]	YOLO v7 [28]	ITTM
Times(seconds)	0.8247	3.0514	**0.0567**	0.2462	0.1122	0.1263	1.4793	0.1070	0.3990	10.4862

fails to converge rapidly due to the similarity between the target grayscale and the cloud grayscale, resulting in detection failure. Figure 21 displays the detection results of Scene 8. Despite significant noise interference in this scene, all algorithms successfully detect target points. However, there are varying numbers of false targets present, which can affect target identification to some extent. From the above figures, it is evident that deep learning-based object detection algorithms like YOLO v5 and YOLO v7 achieve satisfactory detection results in the first six scenes, while failing to detect objects in the last two scenes. This is because the objects in the first six scenes are relatively large, allowing convolutional neural networks to achieve good results with limited training samples. However, in the last two scenes, the objects have fewer texture details and weaker signal intensity. Additionally, both datasets have fewer samples, which prevents convolutional neural networks from adequately learning the target features and thus leading to detection failure. Therefore, in infrared small object detection, traditional detection methods are still necessary to complete the detection task. As shown in Table 5, the YOLO object detection algorithm has a fast model inference speed. However, due to its direct output of detection results, it cannot generate its corresponding ROC curve in Figure 22.

Compared with the other algorithms, the ITTM is based on the optimization of the background constraints, it can better retain the target texture shape while removing the noise interference. In addition, this paper also depicts the ROC curve, in which the horizontal coordinate indicates the false alarm rate and the vertical coordinate indicates the detection rate, and the larger the area enclosed by the horizontal and vertical coordinates, the better the detection effect. In Figure 22, compared with other comparative algorithms, the ITTM achieves better results in both detection rate and false alarm rate in eight scenes. Finally, the abundance of experimental findings underscores the effectiveness of our proposed algorithm. However, it remains subject to certain limitations. Notably, our algorithm exhibits a longer processing time compared to all comparative algorithms(Table 5), primarily due to the construction and resolution of the twist tensor model. This observation serves as a guiding principle for future endeavors, emphasizing the necessity to prioritize the selection of optimal parameters to attain a more efficient iteration of our algorithm.

IV. CONCLUSION

To enhance the detection capability of a photoelectric detection system for infrared targets, this paper introduces a novel detection method that combines spatio-temporal information from the image with the low-rank sparse theory for infrared object detection. Initially, to improve the low-rank characteristics of the infrared image, a background constrained optimization model is constructed based on anisotropy. Then the background-constrained optimization model is combined with the low-rank sparse theory to construct an improved twist tensor model based on background-constrained optimization. This effectively suppresses strong edge clutter to achieve more accurate object detection. Finally, the model of this paper is solved to obtain the effective target signal components. Through experimental validation, the ITTM demonstrated SSIM of 0.9818, 0.9907, 0.9943, 0.9968, 0.9869, 0.9892, 0.9997 and 0.9973, respectively. The BSF of 50.0513, 75.0252, 96.4068, 126.8379, 57.2765, 70.1558, 405.6008 and 128.0142, respectively. And the detection rate exceeding 85% in the eight scenes. The experimental results confirm the feasibility of our approach. In future endeavors, there is a necessity for optimizing the solving methodology of the ITTM.

REFERENCES

[1] S. Zhao and C. Ning, "Review on small-scale pedestrian detection technology for complex pavement," *Comput. Syst. Appl.*, vol. 31, no. 7, pp. 1–11, 2022.

[2] M. Jia, G. Li, S. He, H. Li, Y. Kang, and C. Wang, "Review of two-band infrared target detection," *Flight Control Detection*, vol. 6, no. 2, pp. 60–69, 2023.

[3] D. Zhang, J. Wu, and P. Li, "A summary of moving target detection algorithm based on machine vision," *Intell. Comput. Appl.*, vol. 10, no. 3, pp. 192–195, 2020.

[4] H. Gao and D. Xianhua, "A method for detecting maritime moving targets based on three-frame difference method and improved hybrid Gaussian background model," *Comput. Digit. Eng.*, vol. 47, no. 5, pp. 1140–1144, 2019.

[5] B. Li, F. Li, J. Xie, and F. Huang, "Infrared moving target detection based on frame difference method and adaptive threshold region growing," *Semicond. Optoelectron.*, vol. 38, no. 1, pp. 156–160, 2017.

[6] Z. He and Y. Huang, "Analysis of moving small target detection algorithm using top-hat background modeling," *Electron. Technol.*, vol. 51, pp. 64–66, 2022.

[7] Z. Ding and W. Lu, "Moving target detection algorithm based on vibe background modeling," *Comput. Syst. Appl.*, vol. 28, no. 4, pp. 183–187, 2019.

[8] M. Lu and Z. Chen, "Dim target detection based on gradient feature," *Laser Infr.*, vol. 52, no. 1, pp. 129–135, 2022.

[9] Y. Zhang and F. Zhang, "Salient object detection based on texture and color features," *Comput. Digit. Eng.*, vol. 49, no. 9, pp. 1793–1798, 1877.

[10] Q. Luo and J. Liu, "Infrared target detection algorithm based on fast spectral scale space and dynamic pipeline filtering," *J. Terahertz Sci. Electron. Inf. Technol.*, vol. 20, no. 4, pp. 346–353, 2022.

[11] A. d'Acremont, R. Fablet, A. Baussard, and G. Quin, "CNN-based target recognition and identification for infrared imaging in defense systems," *Sensors*, vol. 19, no. 9, p. 2040, Apr. 2019.

[12] L. Xu, B. Cao, P. Xu, and F. Zhao, "Infrared target detection using deep learning algorithms," *Signal, Image Video Process.*, vol. 17, no. 8, pp. 3993–4000, Nov. 2023.

[13] T. Ma, Z. Yang, J. Wang, S. Sun, X. Ren, and U. Ahmad, "Infrared small target detection network with generate label and feature mapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[14] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Attentional local contrast networks for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9813–9824, Nov. 2021.

[15] L. Yang, S. Liu, and Y. Zhao, "Deep-learning based algorithm for detecting targets in infrared images," *Appl. Sci.*, vol. 12, no. 7, p. 3322, Mar. 2022.

[16] J. Li, P. Zhang, L. Zhang, and Z. Zhang, "Sparse regularization-based spatial–temporal twist tensor model for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 3234608.

[17] Q. Zhang, J. Cai, and Q. Zhang, "Anisotropic infrared background prediction method," *High Power Laser Part. Beams*, vol. 24, no. 2, pp. 301–306, 2012.

[18] Q. Ling, S. Huang, X. Wu, and Z. Yu, "Infrared small target detection based on kernel anisotropic diffusion," *High Power Laser Part. Beams*, vol. 27, no. 1, p. 11014, 2015.

[19] C. Lu, X. Peng, and Y. Wei, "Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5989–5997.

[20] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for ℓ_1-minimization: Methodology and convergence," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1107–1130, Jan. 2008.

[21] C. Gao, T. Zhang, and Q. Li, "Small infrared target detection using sparse ring representation," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 27, no. 3, pp. 21–30, Mar. 2012.

[22] E. Xu, A. Wu, J. Li, H. Chen, X. Fan, and Q. Huang, "Infrared target detection based on joint Spatio–Temporal filtering and L1 norm regularization," *Sensors*, vol. 22, no. 16, p. 6258, Aug. 2022.

[23] Y. Zeng and Q. Chen, "Dim and small target background suppression based on improved bilateral filtering for single infrared image," *Infr. Technol.*, vol. 33, no. 9, pp. 537–540, 2011.

[24] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Infrared small-target detection using multiscale gray difference weighted image entropy," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 1, pp. 60–72, Feb. 2016.

[25] F. Lu, Y. Li, X. Chen, G. Chen, and P. Rao, "Weak target detection for PM model based on top-hat transform," *Syst. Eng. Electron.*, vol. 40, no. 7, pp. 1417–1422, 2018.

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

[27] S. Moradi, P. Moallem, and M. F. Sabahi, "Fast and robust small infrared target detection using absolute directional mean difference algorithm," *Signal Process.*, vol. 177, Dec. 2020, Art. no. 107727.

[28] X. Kong, C. Yang, S. Cao, C. Li, and Z. Peng, "Infrared small target detection via nonconvex tensor fibered rank approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 3068465.

[29] J. Mu, W. Li, J. Rao, F. Li, and H. Wei, "Infrared small target detection using tri-layer template local difference measure," *Opt. Precis. Eng.*, vol. 30, no. 7, pp. 869–882, 2022.

[30] X. Fan, W. Ding, W. Qin, D. Xiao, L. Min, and H. Yuan, "Fusing self-attention and CoordConv to improve the YOLOV5s algorithm for infrared weak target detection," *Sensors*, vol. 23, no. 15, p. 6755, Jul. 2023.

[31] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOV7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

GAOSHAN FENG is currently a Professorate Senior Engineer with Dongfeng Liuzhou Motor Company Ltd. His research interests include unmanned driving and signal and information processing.

WENLIN QIN received the bachelor's degree from the School of Automation, Guangxi University of Science and Technology, Liuzhou, China, where he is currently pursuing the degree. His research interests include signal processing and image detection and tracking algorithms.

ENYONG XU is currently a Senior Engineer with Dongfeng Liuzhou Motor Company Ltd. His main research interests include unmanned driving and signal and information processing.

ZIJUN SUN received the Ph.D. degree from the University of Science and Technology of China, in 2018. He is currently an Associate Professor with the School of Electronic Engineering, Guangxi University of Science and Technology, Liuzhou, China. His research interest includes signal processing.

XIANGSUO FAN received the B.S. degree in automation from Hainan Normal University, Haikou, China, in 2012, and the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 2015 and 2019, respectively. Since 2019, he has been an Associate Professor with the School of Automation, Guangxi University of Science and Technology, Liuzhou, China. His research interests include signal processing and image detection and tracking algorithms.

HUAJIN CHEN received the Ph.D. degree from Fudan University, in 2016. Since 2022, he has been a Professor with the School of Electronic Engineering, Guangxi University of Science and Technology, Liuzhou, China. His research interests include optical manipulations and image detection and tracking algorithms.