



## Article

# Dim and Small Target Detection Based on Energy Sensing of Local Multi-Directional Gradient Information

Xiangsuo Fan <sup>1,2</sup> , Juliu Li <sup>1,\*</sup>, Lei Min <sup>3</sup>, Linping Feng <sup>4</sup>, Ling Yu <sup>1</sup> and Zhiyong Xu <sup>3</sup>

<sup>1</sup> School of Automation, Guangxi University of Science and Technology, Liuzhou 545006, China; 10002085@gxust.edu.cn (X.F.)

<sup>2</sup> Guangxi Collaborative Innovation Centre for Earthmoving Machinery, Guangxi University of Science and Technology, Liuzhou 545006, China

<sup>3</sup> Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China

<sup>4</sup> School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

\* Correspondence: 221055176@stdmail.gxust.edu.cn

**Abstract:** It is difficult for traditional algorithms to remove cloud edge contours in multi-cloud scenarios. In order to improve the detection ability of dim and small targets in complex edge contour scenes, this paper proposes a new dim and small target detection algorithm based on local multi-directional gradient information energy perception. Herein, based on the information difference between the target area and the background area in the four direction neighborhood blocks, an energy enhancement model for multi-directional gray aggregation (EMDGA) is constructed to preliminarily enhance the target signal. Subsequently, a local multi-directional gradient reciprocal background suppression model (LMDGR) was constructed to model the background of the image. Furthermore, this paper proposes a multi-directional gradient scale segmentation model (MDGSS) to obtain candidate target points and then combines the proposed multi-frame energy-sensing (MFESD) detection algorithm to extract the true targets from sequence images. Finally, in order to better illustrate the effect of the algorithm proposed in this paper in detecting small targets in a cloudy background, four sequence images are selected for detection. The experimental results show that the proposed algorithm can effectively suppress the edge contour of complex clouds compared with the traditional algorithm. When the false alarm rate  $P_f$  is 0.005%, the detection rate  $P_d$  is greater than 95%.

**Keywords:** small target detection; reciprocal gradient; multi-directional gradient information; background suppression; energy perception



**Citation:** Fan, X.; Li, J.; Min, L.; Feng, L.; Yu, L.; Xu, Z. Dim and Small Target Detection Based on Energy Sensing of Local Multi-Directional Gradient Information. *Remote Sens.* **2023**, *15*, 3267. <https://doi.org/10.3390/rs15133267>

Academic Editors: Danfeng Hong, Liang-Jian Deng and Gemine Vivone

Received: 4 May 2023

Revised: 16 June 2023

Accepted: 20 June 2023

Published: 25 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, infrared small target detection in the spatiotemporal domain has attracted much attention, as it has important significance for achieving early warning in the spatial domain, spatial surveillance, and so on [1,2]. However, due to the fact that small and weak targets are photoelectric signals for long-range imaging often accompanied by atmospheric turbulence and variable clouds, the targets are often submerged by these interferences [3–6], resulting in the target signal being too weak to be detected by the detector and ultimately leading to detection failure. Therefore, improving the detection efficiency of algorithm models and ensuring detection accuracy has become a key and difficult point in algorithm design research. In infrared image processing, images are often divided into background regions, target regions, and noise regions [7]. How to suppress the background and noise of images and preserve the target signal has become the main research direction. In the past decade, the detection of infrared small targets has mainly been divided into two major directions, namely detection before tracking and tracking before detection [8,9]. Due to the fact that pre-detection requires prior knowledge of the target's motion parameters and mode, and the computational process requires a significant

amount of time, pre-detection tracking algorithms do not require prior knowledge of the target's motion parameters and are widely used due to their simple operation. At present, methods for detecting small and weak targets can be divided into traditional filter detection models, visual saliency-based detection models, low rank sparse recovery theory-based detection models, and deep learning models.

Several background modeling algorithms use traditional filtering, such as top hat filtering [1,10–13], anisotropic filtering [14–17], TDLMS filtering [18–21], bilateral filter [22–24], gradient reciprocal weighted filtering [25–27], etc., and significant background modeling effects have been achieved. These preprocessing algorithms mainly suppress the image background to obtain differential and background images, but they perform poorly in images with more fluctuating clouds and lower signal-to-noise ratio (SNR) in the background. For instance, the anisotropic filtering infrared moving point target detection algorithm based on spatiotemporal fourth-order diffusion proposed by Hu et al. processes images [17]. This algorithm obtains the corresponding directional gradient size by taking derivatives of the three-dimensional coordinates  $x$ ,  $y$ , and  $z$ , expanding the gradient difference between each element and the background. It is then combined with an adaptive kernel diffusion function to obtain the corresponding gradient, and the gradient difference is used for image background prediction suppression, effectively solving the problem of traditional anisotropy not being able to predict the background of images with severe background fluctuations. Deng et al. proposed a top hat infrared small target detection algorithm based on an adaptive M-estimation loop [1], which has a good suppression effect on the image background and uses a new local weighted entropy function to obtain local image features to enhance the target signal. Overcoming the limitations of top hat filtering in filtering structural elements, it has achieved good results in combining local entropy to enhance target signals. In addition, for complex backgrounds with large areas, the gradient reciprocal weighted filtering proposed by Li Zhengzhou et al. [26] achieved good results in background suppression by utilizing the gradient differences between pixels, which is suitable for large-scale background suppression. However, its use of fixed filtering coefficients results in the suppression of the target as the background when processing scenes with complex cloud backgrounds, resulting in detection loss. Zhang et al. proposed a multi-scale gradient correlation filtering (MGCF) detection method based on non-parametric regression [28]. This method calculates the grayscale gradient of a single pixel in the image and designs a multi-scale gradient correlation template to distinguish targets based on the uniqueness of gradient features of weak and small targets, completing the background modeling of the image. This effectively utilizes the local correlation characteristics in the imaging process of small and weak targets, improving the information utilization rate of the target and thus improving the detection rate of the model. However, due to the weak perception of edge contours in nonlinear regression, there are still obvious spatial contours retained in the differential image after background modeling, resulting in a high false alarm rate of the model and low target discrimination.

A detection model based on visual saliency has been proposed to improve the detection rate of the algorithm based on the singularity of the target grayscale in the image: for example, the LCM model [29], ILCM model [30], TTLCM model [31], RLCM model [32], MPCM model [33], WSLCM model [34], etc. This type of algorithm is highly sensitive to grayscale and can effectively suppress background clutter while enhancing the target signal, thereby improving the effectiveness of background modeling. However, there are still many significant false alarm targets in the difference plot, which causes difficulties in target extraction. Therefore, in order to reduce false alarms in detection and improve the depth of complex background suppression, the research on background modeling using low rank sparse recovery theory is currently the main research direction for weak and small target detection. Classical algorithms include the RPCA model [35,36], IPI model [37], TV-PCP model [38], Joint Spatio-Temporal Filtering and L1 Norm Regularization model [39], MF-STPT model [40]. They divide the original image into two parts: target and background. In addition, they transform object detection into optimizing the sparse pixel matrix and

low rank moment matrix in the original image, overcoming many detection limitations such as discrete image grayscale distribution and unstable detection performance of traditional algorithms. Gao et al. proposed an infrared small target detection model (IPI model) for a single frame image [37], which divides the original image into two parts: target and background. The algorithm transforms target detection into optimizing the sparse pixel matrix and low rank moment matrix in the original image, overcoming many detection limitations such as the discrete distribution of image grayscale values and unstable detection performance of traditional algorithms. Sun et al. proposed a low-rank sparse detection model with multiple subspaces and spatiotemporal block tensors [41]. This method constructs norm tensors of multiple subspaces, effectively synergizing the intrinsic connections between spatial and temporal information of pixels in the image for low-rank analysis. The background modeling effect is significant, and the target signal is obvious. Fan et al. [35] proposed principal component analysis to analyze the pixel components of images and obtain low-rank sparse components for background modeling, achieving good background suppression effects. Furthermore, Rawat et al. proposed a new non-convex weighted kernel norm to enhance the robustness of principal component analysis detection models [42]. This method utilizes an adaptive weight operation approach, based on the generation of different weight norms for each pixel, to carry out background modeling operations for images and effectively suppress sharp contours in complex scenes. However, due to its high computational complexity and long cycle, this method is mainly used for single-frame image detection and sequence image detection. It has a low efficiency in processing complex background scenes with clouds, and it also has algorithmic shortcomings for the real-time detection and tracking of small targets.

With the development of science and technology, weak and small object detection methods based on deep learning [43–47] have become popular in object detection, and such algorithms can effectively improve the accuracy of detection. However, due to its close connection with the training data, it is necessary to continuously update the database for data training to adapt the algorithm framework to new scenarios, resulting in a longer computational cycle and poor real-time performance. Combining SSD as proposed by Huang et al., the MobileNetV2 network and feature pyramid structure are used to suppress the background of images [43], which effectively preserves the feature information of weak targets while suppressing the background. However, due to the need for a large dataset support, the detection rate of the algorithm is low, which is not conducive to real-time detection and tracking. The combination of Faster R-CNN and GAN proposed by Bai et al. for dim and small target detection [44] also requires a large amount of data training and calculation to achieve the background modeling of images. In order to improve the timeliness of deep learning detection models, Liu et al. proposed a heterogeneous parallel network with similar object enhancement to complete the detection of weak and small targets [48]. This model first obtains target features by calculating the target local area and background local area on the constructed dataset, greatly reducing the model's computational time and improving detection efficiency. Similarly, Xu et al. proposed a network model for feature extraction with multi-scale and multi-level features [49], which effectively improves target saliency by fusing target features through resampling, and the real-time detection effect of the model is outstanding.

This article combines traditional filtering detection models with visual saliency detection models to construct a dim and small target detection based on the energy sensing of a local multi-directional gradient information model. In response to the problem of significant cloud contours and weak target signals in multi-cloud scenarios, which results in poor background modeling performance, this paper first proposes an energy enhancement model for multi-directional gray aggregation (EMDGA) to weaken sharp cloud contours and preliminarily enhance the saliency of the target. Subsequently, in the background modeling process, using gradient reciprocal modeling has the advantage of suppressing the background in a large area, while only using a single pixel for background analysis results in incomplete background suppression. This article improves the processing method of a

single pixel to a local neighborhood and proposes a region block gradient reciprocal filtering model that integrates the local multi-directional gradient reciprocal background suppression model (LMDGR). Furthermore, considering the presence of noise in the differential image after background modeling, it poses a difficult problem for target discrimination. This article mobilizes the scale information during the target imaging process and proposes a multi-directional gradient scale segmentation model (MDGSS) combined with a dual window segmentation model to highlight the target and enhance its saliency in the image. Finally, to adapt to practical engineering applications, this paper proposes a multi-frame energy-sensing detection model (MFESD) to achieve target tracking and detection by utilizing the prominent features of target grayscale. The main contributions can be summarized as follows:

- (1) Construct an energy enhancement model for multi-directional grayscale aggregation, perform secondary energy aggregation operations on the original image, and enable the grayscale fusion of edge contours and noise in a multi-cloud background. In addition, adjust the enhancement strategy of the grayscale aggregation enhancement model based on the first energy aggregation, and perform a second grayscale aggregation processing on the target neighborhood to highlight the target and smooth the background.
- (2) Taking advantage of the advantages of traditional gradient reciprocal filtering algorithms with good background suppression, a region block gradient reciprocal filtering model integrating multi-directional information is proposed to model the background of sequence images with multi-cloud fluctuation interference, obtain differential images containing target information, and improve the utilization of local pixel information.
- (3) On the basis of background modeling, combined with the uneven distribution of target energy, a multi-directional and multi-scale segmentation model is constructed to segment the differential image to remove some noise. We enhance the saliency of the target in the image again and improve the detection rate of the algorithm.
- (4) Construct a multi-frame energy-sensing detection model for sequence images and perform real target determination operations on candidate targets based on the singularity of the target grayscale to improve the detection accuracy of the model.

After relevant experiments, the proposed local multi-directional gradient information energy perception dim small target detection model in this article achieves a structural similarity index (SSIM) over 99% for background restoration after background modeling, an average background suppression factor (BSF) of 373.1591, and a signal gain (IC) of 37.3615 dB. The organizational structure of this article is arranged as follows: Section 2 introduces the mathematical principles and overall algorithm flow of the model constructed in this article; Section 3 conducts energy aggregation analysis on the selected sequence of infrared images; Section 4 conducts experimental comparative analysis on the aggregated infrared sequence images to demonstrate the effectiveness of the proposed algorithm; and Section 5 provides a summary.

## 2. Materials and Methods

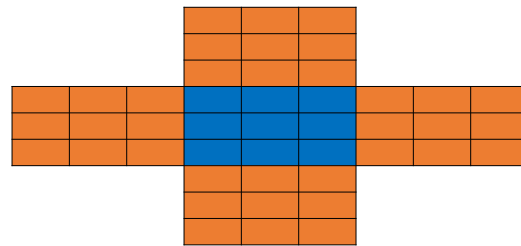
This section mainly provides a detailed introduction to the mathematical principles of the proposed energy enhancement model for multi-directional gray aggregation (EMDGA), local multi-directional gradient reciprocal background suppression model (LMDGR), multi-directional gradient scale segmentation model (MDGSS), and multi-frame energy-sensing detection model (MFESD). The relevant details are shown below.

### 2.1. Energy Enhancement Model for Multi-Directional Gray Aggregation (EMDGA)

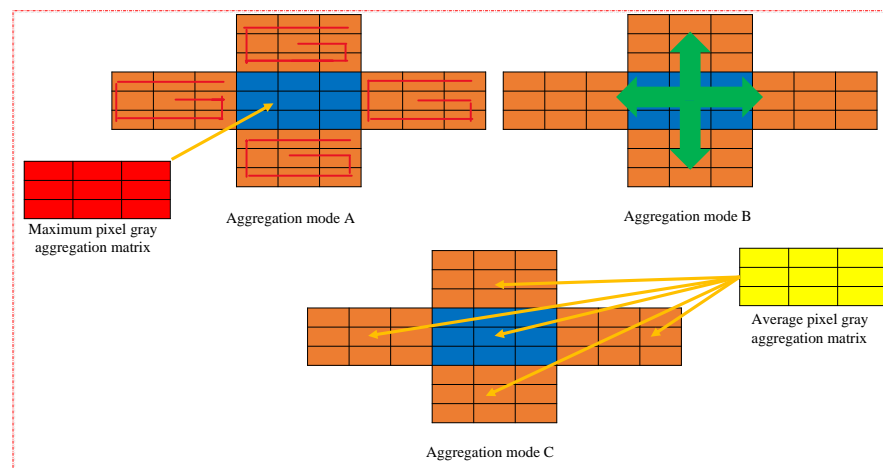
This section mainly provides a detailed introduction to the energy enhancement model for multi-directional gray aggregation (EMDGA) proposed in this article, aiming at the problem of small pixels occupied by weak targets in the imaging process and low target identification. This article uses the neighborhood of pixels to construct a multi-directional

grayscale energy aggregation model for region blocks to enhance the signal of the target and improve its discrimination. Firstly, the model for energy aggregation is defined as follows:

As shown in Figure 1, the central blue  $3 \times 3$  pixel block is used to build the  $3 \times 3$  area of the pixel neighborhood upwards, downwards, leftward and rightward to form the energy aggregation model. The model is used to preprocess the image with gray energy aggregation so as to improve the contrast of the target in the image and enhance the target signal. In order to make the target signal enhancement obvious, three multi-directional energy aggregation schemes are defined in combination with the above model to achieve signal enhancement. The specific aggregation form is shown in Figure 2.



**Figure 1.** Energy aggregation filling model.



**Figure 2.** Schematic diagram of gray energy aggregation mode.

As shown in Figure 2, three different energy aggregation modes are defined in this study. When the average pixel value of the blue area block in the middle of the aggregation model is at the maximum, it indicates that this area belongs to a candidate area including targets. Then, the maximum gray value in this area is extracted and formed into a  $3 \times 3$  maximum gray value matrix, and the maximum gray value matrix is replaced by the blue pixel area in the model. The four direction blocks in the model are filled with their pixel gray mean value to enhance the target signal, as shown in aggregation mode A in Figure 2. When the average pixel value of the blue area block in the middle of the aggregation model is at the minimum, it indicates that this area does not meet the characteristics of the isolated protrusions formed in the figure when small and weak targets are imaged, and this area belongs to the background area. Therefore, at this time, the average pixel value of the blue area block in the model is filled into the whole model to weaken the background effect, as shown in Figure 2. If the average pixel value of the middle blue area block in the aggregation model is not at the maximum or minimum, it indicates that the aggregation model may be at the edge contour of the background at this time, and the average pixel value of each area block in the aggregation model has little difference, which belongs to the background part. At this time, the two directions with the lowest mean value in the



four directions of up, down, left, and right are selected for pixel summation to obtain the mean value, and the results are filled into a new  $3 \times 3$  area block to form the average pixel gray-level aggregation matrix as shown in the above figure. Then, as shown in the above figure, aggregation mode C is used to replace the entire aggregation model with the pixel gray-level aggregation matrix to weaken the edge contour.

According to the energy aggregation model defined above, the detailed mathematical model is as follows:

$$\left\{ \begin{aligned} \Delta f_{U3 \times 3} &= \frac{1}{R \times R} \sum_{-\mu}^{\mu} \sum_{-\rho}^{\rho} f1(x + \mu, y - \rho - K) \\ \Delta f_{D3 \times 3} &= \frac{1}{R \times R} \sum_{-\mu}^{\mu} \sum_{-\rho}^{\rho} f1(x + \mu, y + K + \rho) \\ \Delta f_{T3 \times 3} &= \frac{1}{R \times R} \sum_{-\mu}^{\mu} \sum_{-\rho}^{\rho} f1(x + \mu, y + \rho) \\ \Delta f_{L3 \times 3} &= \frac{1}{R \times R} \sum_{-\mu}^{\mu} \sum_{-\rho}^{\rho} f1(x - K + \mu, y + \rho) \\ \Delta f_{R3 \times 3} &= \frac{1}{R \times R} \sum_{-\mu}^{\mu} \sum_{-\rho}^{\rho} f1(x + K + \mu, y + \rho) \end{aligned} \right. \quad (1)$$

where  $f1$  represents the whole energy aggregation model area selected from the input image  $f$ , and  $\Delta f_{U3 \times 3}$ ,  $\Delta f_{D3 \times 3}$ ,  $\Delta f_{T3 \times 3}$ ,  $\Delta f_{L3 \times 3}$ , and  $\Delta f_{R3 \times 3}$ , respectively, represent the regional mean values of the upper, lower, center, left and right in the aggregation model.  $R = 3$  represents the value of the regional size.  $(x, y)$  indicates the pixel position within the region,  $(\mu, \rho)$  represents the pixel position and  $\mu = \text{fix}(R/2), \rho = \text{fix}(R/2)$ , where  $\text{fix}$  is a function of Matlab R2012b.  $K = 3$  represents a parameter that controls the scale of pixel movement. For enhancing the grayscale signal of the local region of interest, this article compares the grayscale difference  $T$  between individual pixels and the mean difference  $T1$  between the two largest regions in Figure 2 with the set constant parameters  $S1$  and  $S2$  to determine the final filling strategy. The specific mathematical model is as follows:

$$\left\{ \begin{aligned} &\text{if } \max(\Delta f_{U3 \times 3}, \Delta f_{D3 \times 3}, \Delta f_{T3 \times 3}, \Delta f_{L3 \times 3}, \Delta f_{R3 \times 3}) = \Delta f_{T3 \times 3} \wedge T \geq S1 \wedge T1 \geq S2 \\ &up = \Delta f_{U3 \times 3} \\ &down = \Delta f_{D3 \times 3} \\ &left = \Delta f_{L3 \times 3} \\ &right = \Delta f_{R3 \times 3} \\ &Center = \max(f1(:)) \times x1 \\ &f_{Center} = Center \\ &F(\mu, \rho) = f_{Center} \\ &\text{if } \min(\Delta f_{U3 \times 3}, \Delta f_{D3 \times 3}, \Delta f_{T3 \times 3}, \Delta f_{L3 \times 3}, \Delta f_{R3 \times 3}) = \Delta f_{T3 \times 3} \\ &up = \Delta f_{T3 \times 3} \\ &down = \Delta f_{T3 \times 3} \\ &left = \Delta f_{T3 \times 3} \\ &right = \Delta f_{T3 \times 3} \\ &Center = \Delta f_{T3 \times 3} \\ &f_{Center} = Center \\ &F(\mu, \rho) = f_{Center} \\ &\text{others} \\ &up = down = left = right = Center = Data \\ &F(\mu, \rho) = f_{Center} \end{aligned} \right. \quad (2)$$

where  $f1$  represents the area occupied by the energy aggregation model, and  $\Delta f_{U3 \times 3}$ ,  $\Delta f_{D3 \times 3}$ ,  $\Delta f_{T3 \times 3}$ ,  $\Delta f_{L3 \times 3}$ , and  $\Delta f_{R3 \times 3}$  represent the regional mean values of the upper, lower, center, left, and right in the aggregation model, respectively.  $up$ ,  $down$ ,  $left$ ,  $right$ , and  $center$  represent the upper, lower, left, right, and center  $3 \times 3$  regions in the aggregation

model of Figure 2, respectively.  $f\_Center$  refers to the central area corresponding to the original image of the same size as the selected  $f1$  area.  $Data$  refers to the pixel gray value after the sum and average of the two least mean values in the four directions after sorting, and  $F(\mu, \rho)$  refers to the image that completes the energy aggregation processing.

$$\begin{cases} A = [\Delta f_{U_{3 \times 3}}, \Delta f_{D_{3 \times 3}}, \Delta f_{L_{3 \times 3}}, \Delta f_{R_{3 \times 3}}, \Delta f_{T_{3 \times 3}}] \\ data = sort(A, 'descend') \\ Data = [data(1) + data(2)]/2 \end{cases} \tag{3}$$

where  $A$  refers to the top, bottom, left and right three  $3 \times 3$ . The matrix is composed of regional mean values.  $data$  represents the sorted matrix.  $sort$  and  $descend$  are the sorting and descending functions of Matlab R2012b.  $data(1)$  and  $data(2)$  represent the values in the 2 directions with the highest grayscale mean of the pixel in the 4 directions.

The target signal enhancement can be completed by combining the above Equations (1)–(3). The model has preliminarily achieved the target signal enhancement. Aiming to further achieve the regional fusion of target energy to highlight the significance of the target, on the basis of the preliminary enhancement, the target signal is enhanced twice by combining Equation (1). The specific mathematical model is as follows:

$$\begin{cases} \Delta f_{U_{3 \times 3}} = \frac{1}{R \times R} \sum_{-\mu_1}^{\mu_1} \sum_{-\rho_1}^{\rho_1} ff1(x_1 + \mu_1, y_1 - \rho_1 - K) \\ \Delta f_{D_{3 \times 3}} = \frac{1}{R \times R} \sum_{-\mu_1}^{\mu_1} \sum_{-\rho_1}^{\rho_1} ff1(x_1 + \mu_1, y_1 + K + \rho_1) \\ \Delta f_{T_{3 \times 3}} = \frac{1}{R \times R} \sum_{-\mu_1}^{\mu_1} \sum_{-\rho_1}^{\rho_1} ff1(x_1 + \mu_1, y_1 + \rho_1) \\ \Delta f_{L_{3 \times 3}} = \frac{1}{R \times R} \sum_{-\mu_1}^{\mu_1} \sum_{-\rho_1}^{\rho_1} ff1(x_1 - K + \mu_1, y_1 + \rho_1) \\ \Delta f_{R_{3 \times 3}} = \frac{1}{R \times R} \sum_{-\mu_1}^{\mu_1} \sum_{-\rho_1}^{\rho_1} ff1(x_1 + K + \mu_1, y_1 + \rho_1) \end{cases} \tag{4}$$

In the formula,  $ff1$  represents the energy aggregation area selected in the preliminary enhancement image  $F$ .  $\Delta f_{U_{3 \times 3}}$ ,  $\Delta f_{D_{3 \times 3}}$ ,  $\Delta f_{L_{3 \times 3}}$ ,  $\Delta f_{R_{3 \times 3}}$ , and  $\Delta f_{T_{3 \times 3}}$  represent the mean values of the upper, lower, central, left, and right regions in the aggregation model.  $R$  represents the value of region size, with a value of 3 in the text,  $(x_1, y_1)$  represents the pixel position within the region,  $(\mu_1, \rho_1)$  represents the sequence number of pixels,  $K$  represents the parameter that controls the scale of pixel movement, and the value in the text is 3.

Due to the initial aggregation and enhancement of the target signal, some noisy signals were enhanced during the initial aggregation, resulting in detection confusion. This energy aggregation set a constant parameter  $S3$  to compare with the grayscale difference  $T2$  of a single pixel in the enhanced image  $F$  in Formula (1) in order to extract the enhanced region of interest. Correspondingly, in the selection of enhancement strategies, as preliminary enhancements have been made, the grayscale difference between the target and the background has been highlighted, and there is no need to fill the middle neighborhood to the surrounding neighborhood. Therefore, the secondary energy aggregation mode is only completed using mode A and mode C in Figure 2, and its corresponding mathematical expression is as follows:

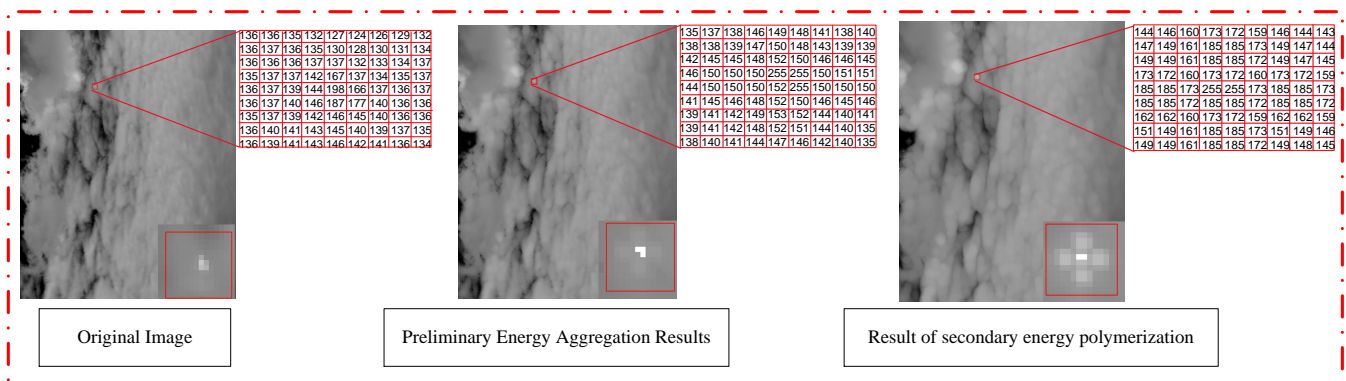
$$\left\{ \begin{array}{l} \text{if } \max(\Delta f_{U_{13 \times 3}}, \Delta f_{D_{13 \times 3}}, \Delta f_{T_{13 \times 3}}, \Delta f_{L_{13 \times 3}}, \Delta f_{R_{13 \times 3}}) = \Delta f_{T_{13 \times 3}} \wedge T2 \geq S3 \\ \quad \text{up} = \Delta f_{U_{13 \times 3}} \\ \quad \text{down} = \Delta f_{D_{13 \times 3}} \\ \quad \text{left} = \Delta f_{L_{13 \times 3}} \\ \quad \text{right} = \Delta f_{R_{13 \times 3}} \\ \quad \text{Center} = \max(ff1(:)) \times x_2 \\ \text{else} \\ \quad \text{up} = \text{down} = \text{left} = \text{right} = \text{Center} = \text{Data} \\ \quad FF(\mu, \rho) = \text{Center} \end{array} \right. \quad (5)$$

In the formula,  $ff1$  refers to the energy aggregation area selected in the preliminary aggregation image  $F$ ,  $FF$  refers to the image that completes the secondary energy aggregation, and  $x_2$  is the set constant parameter, with a value of 1.5 in the text.

$$\left\{ \begin{array}{l} A1 = [\Delta f_{U_{13 \times 3}}, \Delta f_{D_{13 \times 3}}, \Delta f_{L_{13 \times 3}}, \Delta f_{R_{13 \times 3}}, \Delta f_{T_{13 \times 3}}] \\ \text{data1} = \text{sort}(A1, 'descend') \\ \text{Data1} = \max(\text{data1}(:)); \end{array} \right. \quad (6)$$

$A1$  represents a set composed of the mean values of four  $3 \times 3$  regions by  $\Delta f_{U_{13 \times 3}}$ ,  $\Delta f_{D_{13 \times 3}}$ ,  $\Delta f_{L_{13 \times 3}}$ , and  $\Delta f_{R_{13 \times 3}}$ ,  $\text{data1}$  represents the sorted matrix, and  $\text{sort}$  and  $\text{descent}$  are functions of sorting and descending in Matlab, respectively.  $\text{Data1}$  represents the maximum value in the four directions.

In conclusion, after the preliminary verification as shown in Figure 3, the multi-directional gray level energy aggregation model proposed in this paper has been able to enhance the target signal and improve the target significance.



**Figure 3.** Effect of energy aggregation.

## 2.2. Improved Gradient Reciprocal Background Suppression Model Based on Region Information Fusion

### 2.2.1. Reciprocal Gradient Related Work

In the optical signal system of long-distance imaging, small and weak targets occupy few pixels and are often submerged by the undulating clouds, which makes the target often suppressed as the background, causing target detection failure. Therefore, the normalized gradient reciprocal filtering can suppress the edge contour of a complex background while retaining the target signal. This filter makes full use of the steep amplitude signal presented by the target in the image to calculate the gradient of the neighboring pixel and normalize it. Therefore, the gradient normalization at the target and strong noise points presents a two-stage phenomenon, which makes the weight coefficient obtained after normalization processing for the pixel with large difference (such as the target pixel) small, and the pixel with large difference can be well retained in the difference map. On the contrary, for pixels with little or no difference (such as background pixels), the weighting coefficient obtained after normalization processing is large, which can effectively weaken the pixels with little



difference in the difference map, so as to suppress the background and highlight the target. The traditional gradient reciprocal background modeling model is as follows [26]:

$$\begin{cases} \text{if } f(x + m, y + n, k) \neq f(x, y, k) \\ T_{x,y}(m, n, k) = 1/|f(x + m, y + n, k) - f(x, y, k)| \\ \text{if } f(x + m, y + n, k) = f(x, y, k) \\ T_{x,y}(m, n, k) = \theta \end{cases} \tag{7}$$

where  $(x, y)$  represents the center position of the selected neighborhood pixel,  $(m, n)$  represents the current neighborhood location, and  $T_{x,y}(m, n, k)$  represents the reciprocal gradient of the pixel at the  $(m, n)$  position in the neighborhood, generally  $r \times r$  select  $4 \times 4$ .  $\theta$  refers to the gradient reciprocal value given when the adjacent pixel is the same as the center pixel, which is generally greater than 1.  $k$  refers to the frame number of the currently processed image.

The processed image needs to be normalized to obtain the gradient weighting coefficient between pixels. The specific weighting model is as follows [26,27]:

$$\begin{cases} \text{if } f(m, n, k) = f(0, 0, k) \\ H_{x,y}(m, n, k) = 1/\theta \\ \text{else} \\ H_{x,y}(m, n, k) = \left(1 - \frac{1}{\theta}\right) \times \left[\frac{T_{x,y}(m, n, k)}{\sum_{f(m,n,k) \neq f(0,0,k)}^r T_{x,y}(m, n, k)}\right] \end{cases} \tag{8}$$

where  $f(m, n, k)$  is the position  $(m, n)$  of a point in the  $k$ -th frame image,  $H_{x,y}(m, n, k)$  represents the gradient weighting coefficient between pixels in the image,  $T_{x,y}(m, n, k)$  is the reciprocal gradient of the pixel, and  $r$  is the radius of the selected calculated neighborhood. Then, combining the gradient between pixels and the corresponding normalization coefficient, the inverse gradient value is output to obtain the predicted background image. The specific output formula is as follows [26]:

$$f_p(x, y, k) = \sum_{m=-r}^r \sum_{n=-r}^r H_{x,y}(m, n, k) \times f(x + m, y + n, k) \tag{9}$$

Finally, the difference image is obtained by subtracting the background image from the original image. The specific calculation formula is as follows [26]:

$$f_D(x, y, k) = f(x, y, k) - f_p(x, y, k) \tag{10}$$

where  $f_D(x, y, k)$  represents the difference image,  $f(x, y, k)$  represents the original image, and  $f_p(x, y, k)$  is the pixel prediction value.

Through relevant research, it is found that the gradient reciprocal filtering algorithm uses the singular property between the background and the target point to filter the selected area (generally  $r = 4$  and the matrix is  $9 \times 9$ ) to complete the background modeling. The algorithm achieves good results in sequential scenes with small background fluctuation. It shows that the gradient reciprocal model has certain advantages in suppressing the image background, but it is found through research that because of the filtering parameters of the gradient filtering model of a single pixel,  $\theta$  is a fixed parameter value, which results in a poor detection effect for multi-edge sequence images [50], and there are many false alarms in the differential image. Therefore, in document [50], References [26,27] use the correlation function  $R(x, y, k)$  to improve the filter parameters so that it can adaptively adjust the corresponding filter parameters to suppress the background, and the suppression effect is more obvious. The specific model is as follows [50]:

$$R(x, y, k) = \sum_{m=-r, m \neq 0}^r \sum_{n=-r, n \neq 0}^r \begin{cases} 1 & \text{if } |f(x + m, y + n, k) - f(x, y, k)| \geq T \quad T \geq 0 \\ 0 & \text{others} \end{cases} \tag{11}$$

In the formula, after comparing the difference between the current pixel  $f(x + m, y + n, k)$  and the central pixel  $f(x, y, k)$ , the image is binarized using the segmentation threshold  $T$  determined by CFAR criteria. If the pixel value of the current pixel is greater than the pixel value of the selected neighborhood center position, the area is filled with a number value of 1; otherwise,  $\theta$  [26] is filled, and finally, all the filled values are summed as the output of the correlation function  $R(x, y, k)$ . However, the value of correlation function  $R(x, y, k)$  in Reference [26] is determined by the segmentation threshold determined by CFAR, which cannot achieve adaptive effect in complex scenes. Finally, the improved correlation function and the given defined constant are multiplied to determine the filtering parameters  $\theta$  output in [50], as shown in the following formula:

$$\theta = E \times R(x, y) \quad (12)$$

In the formula,  $E$  represents the defining constant, which is a fixed value, and  $R(x, y)$  is the correlation function in the above formula.

To sum up, the improved gradient reciprocal filtering can be summarized as follows [50]:

$$\begin{cases} \text{if } |f(x + m, y + n, k) - f(x, y, k)| \leq c \\ T_{x,y}(m, n, k) = \theta \\ \text{other} \\ T_{x,y}(m, n, k) = 1/|f(x + m, y + n, k) - f(x, y, k)| \end{cases} \quad (13)$$

where  $T_{x,y}(m, n, k)$  represents the reciprocal of the pixel gradient,  $f(x + m, y + n, k)$  and  $f(x, y, k)$  represent the pixel values of the current pixel and the center pixel, respectively, and the filter coefficient  $\theta$  is determined by the difference between them. Based on this, we determine the normalized weighting coefficient of the pixel after and complete the background modeling of the image in combination with Equations (3) and (4) finally. The improved normalized model is as follows [50]:

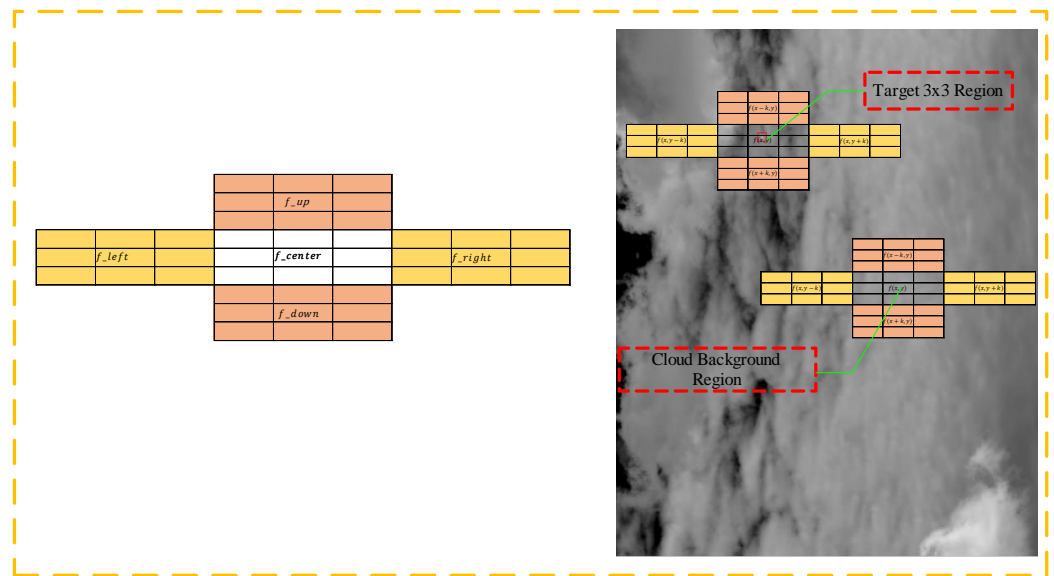
$$\begin{cases} \text{if } m = 0, n = 0 \\ H_{x,y}(m, n, k) = 1/\theta \\ \text{other} \\ H_{x,y}(m, n, k) = \left(1 - \frac{1}{\theta}\right) \times \left[\frac{T_{x,y}(m, n, k)}{\sum_{m=-r}^r \sum_{n=-r}^r T_{x,y}(m, n, k)}\right] \end{cases} \quad (14)$$

The relevant definitions in the formula are shown in Formula (8).

All things considered, traditional gradient reciprocal background modeling models mainly perform neighborhood calculations on a single pixel of the image to achieve background suppression, which cannot fully utilize the detailed information of the image. As a result, the model has poor background modeling performance and maintains a high false alarm rate when facing complex scenes, causing significant difficulties for target detection. Therefore, this article constructs a local multi-directional gradient reciprocal background suppression model (LMDGR) to achieve image background modeling, as detailed in the following section.

### 2.2.2. Local Multi-Directional Gradient Reciprocal Background Suppression Model (LMDGR)

The traditional gradient reciprocal only utilizes a single pixel's information to complete background modeling in the background modeling process, resulting in severe weakening of the target signal in the differential image. This article proposes a local multi-directional gradient reciprocal background suppression model (LMDGR) based on the energy aggregation model mentioned above to achieve image background modeling. This model improves the traditional model of gradient reciprocal calculation between individual pixels to a model that performs operations on local area pixels to achieve background suppression. We enable the retention of target information as regional blocks, enhance the saliency of the target, and lay the foundation for subsequent target detection. The specific model is shown in the following Figure 4.



**Figure 4.** Schematic diagram of local area block background modeling model.

As shown in the Figure 4, the model proposed in this article can fully utilize the information around the central pixel for gradient operation in background modeling, and it has good performance in suppressing large areas of background. On the basis of secondary energy aggregation, based on the strong correlation between the background and the small correlation between the target and the background, the region of interest is first extracted by comparing the grayscale difference  $L = |FF(i, j) - FF(i - 1, j)|$  between a single pixel with the set constant parameter  $Q$ , and the average value of the region is calculated. This lays the foundation for a background modeling model with a local multi-directional gradient reciprocal. The specific mathematical model definition is as follows:

$$\left\{ \begin{array}{l}
 \text{if } (L \geq Q) \\
 \Delta f_{U1} = \left[ \frac{1}{R \times R} \sum_{-T_1}^{T_1} \sum_{-T_2}^{T_2} FF(x + T_1, y - T_2 - K) \right] \\
 \Delta f_{D1} = \left[ \frac{1}{R \times R} \sum_{-T_1}^{T_1} \sum_{-T_2}^{T_2} FF(x + T_1, y + K + T_2) \right] \\
 \Delta f_{L1} = \left[ \frac{1}{R \times R} \sum_{-T_1}^{T_1} \sum_{-T_2}^{T_2} FF(x - K + T_1, y + T_2) \right] \\
 \Delta f_{R1} = \left[ \frac{1}{R \times R} \sum_{-T_1}^{T_1} \sum_{-T_2}^{T_2} FF(x + K + T_1, y + T_2) \right] \\
 \Delta f_{c1} = \left[ \frac{1}{R \times R} \sum_{-T_1}^{T_1} \sum_{-T_2}^{T_2} FF(x + T_1, y + T_2) \right] \\
 \text{else} \\
 \Delta f_{U1} = \Delta f_{D1} = \Delta f_{L1} = \Delta f_{R1} = \Delta f_{c1} = 0
 \end{array} \right. \quad (15)$$

where  $FF$  represents the image that completes the secondary energy aggregation in Formula (5), and  $\Delta f_{U1}$ ,  $\Delta f_{D1}$ ,  $\Delta f_{L1}$ , and  $\Delta f_{R1}$ ,  $\Delta f_{c1}$  represent the mean gray level of the local area, respectively. Corresponding to the aggregation model,  $R = 3$  represents the size of the neighborhood region.  $(T_1, T_2)$  indicates the sequence number of pixels and  $T_1 = fix(R/2)$ ,  $T_2 = fix(R/2)$ , where  $fix$  is a function of Matlab.  $(x, y)$  represents the pixel of interest whose gray-level gradient meets  $L \geq Q$ , and  $K$  represents the pixel moving step.  $FF(x + T_1, y - T_2 - K)$ ,  $FF(x + T_1, y + K + T_2)$ ,  $FF(x - K + T_1, y + T_2)$ ,  $FF(x + K + T_1, y + T_2)$ , and

$FF(x + T_1, y + T_2)$ , respectively, represent the upper, lower, left, right, and center region blocks in the model in Figure 4, and  $K$  is the movement step parameter of the pixel, which is set to 3.

On the basis of extracting regions of interest, fixed filtering coefficients are used for traditional gradient reciprocals  $\theta$  to determine the final background modeling weight of the gradient reciprocal model, resulting in poor modeling performance and an incomplete utilization of target signals. This article improves the filtering coefficients in Formula (12) and proposes gradient reciprocal filtering coefficients for region blocks  $\theta$ , enabling the model to adaptively update the corresponding filtering coefficients between each region block  $\theta$  to complete background modeling and increase the utilization of image detail information. This allows the region blocks containing target information to be given higher weights and retained during background modeling, achieving the goal of background modeling. The specific mathematical model is as follows:

$$\begin{cases} A2 = [\Delta f_{U1}, \Delta f_{D1}, \Delta f_{L1}, \Delta f_{R1}, \Delta f_{c1}] \\ \text{if } \max(A2(:)) == \Delta f_{c1} \\ \text{mark}_{up} = \text{mark}_{down} = \text{mark}_{left} = \text{mark}_{right} = 0 \\ \text{mark}_{Center} = 1 \\ \text{else} \\ \text{mark}_{up} = \text{mark}_{down} = \text{mark}_{left} = \text{mark}_{right} = \text{mark}_{Center} = 0 \end{cases} \quad (16)$$

where  $A2$  is the set of pixel mean values in the 5 regional blocks in Figure 4,  $\Delta f_{U1}$ ,  $\Delta f_{D1}$ ,  $\Delta f_{L1}$ ,  $\Delta f_{R1}$ , and  $\Delta f_{c1}$  represent the mean gray level of the local area,  $\text{mark}_{up}$ ,  $\text{mark}_{down}$ ,  $\text{mark}_{left}$ ,  $\text{mark}_{right}$ , and  $\text{mark}_{Center}$  represent the empty matrix corresponding to the five blocks, which is used to fill  $\theta$  and 1 to determine the filter coefficient  $\theta$ . Similar to the mathematical model determined in Reference [25], combined with constant parameter  $E$  (set to 10), the mathematical model determined by the filter coefficient in this paper is as follows:

$$\begin{cases} G = \text{sum}(\text{mark}(:)) \\ \theta = E \times G \end{cases} \quad (17)$$

According to traditional gradient reciprocal filtering, combined with filtering coefficient  $\theta$ , the background modeling of the image can be completed. However, the local region block gradient reciprocal filtering model proposed in this article processes images based on region block information, while the traditional gradient reciprocal reflects the information of a single pixel in the image, and the target information cannot be fully utilized. Therefore, this article redefines the gradient reciprocal background modeling model as follows:

$$\begin{cases} A2 = [\Delta f_{U1}, \Delta f_{D1}, \Delta f_{L1}, \Delta f_{R1}, \Delta f_{c1}] \\ \text{if } \max(A2(:)) == \Delta f_{c1} \\ FF(X + \omega_1, y - \omega_2 - K) = FF(X + \omega_1, Y + K + \omega_2) = \\ FF(X - K + \omega_1, Y + \omega_2) = FF(X + K + \omega_1, Y + \omega_2) = \\ FF(X + \omega_1, Y + \omega_2) = 1/\theta \\ \text{else} \\ FF(X + \omega_1, Y + \omega_2) = 1/L \end{cases} \quad (18)$$

where  $A1$  is defined as Formula (16), and  $FF(X + \omega_1, Y - \omega_2 - K)$ ,  $FF(X + \omega_1, Y + K + \omega_2)$ ,  $FF(X - K + \omega_1, Y + \omega_2)$ ,  $FF(X + K + \omega_1, Y + \omega_2)$ , and  $FF(X + \omega_1, Y + \omega_2)$  represent the upper, lower, left, right, and center area blocks in the model in Figure 4, respectively.  $FF$  represents the image after grayscale aggregation in Formula (5).  $(X, Y)$  represents the pixel coordinates in  $FF$ ,  $\theta$  represents the filtering coefficient determined in Formula (17), and  $K$  represents the step size of pixel movement, while  $(\omega_1, \omega_2)$  represents the sequence number of pixels in an  $FF$  image.  $L = |FF(i, j) - FF(i - 1, j)|$  is the gray difference between the above individual pixels, and  $(i, j)$  represents the pixel coordinates in image  $FF$ . According to the gradient reciprocal background modeling model, the normalization function of the

local multi-directional gradient reciprocal background modeling model proposed in this article is redefined as follows:

$$\begin{cases} \text{if } (\Delta f_{U1} \wedge \Delta f_{D1} \wedge \Delta f_{L1} \wedge \Delta f_{R1} \wedge \Delta f_{c1} \neq 0) \\ g = (1 - 1/\theta) / (\Delta f_{c1} / ((\Delta f_{U1} + \Delta f_{D1} + \Delta f_{L1} + \Delta f_{R1} + \Delta f_{c1}) / 5)) \\ \text{else} \\ g = 1 - 1/\theta \end{cases} \quad (19)$$

where  $\Delta f_{U1}$ ,  $\Delta f_{D1}$ ,  $\Delta f_{L1}$ ,  $\Delta f_{R1}$ , and  $\Delta f_{c1}$  represent the mean gray level of the local area,  $\theta$  is the filtering parameter and  $g$  is the normalized coefficient. Using Equations (18) and (19), the background modeling of the image can be completed to obtain the differential image  $D$ , and the formula is as follows:

$$D = (FF(X + \omega_1, Y - \omega_2 - K) + FF(X + \omega_1, Y + K + \omega_2) + FF(X - K + \omega_1, Y + \omega_2) + FF(X + K + \omega_1, Y + \omega_2) + FF(X + \omega_1, Y + \omega_2)) \times g \quad (20)$$

In the equation,  $D$  represents the differential image,  $FF$  represents the calculation result in Formula (18),  $g$  represents the normalization coefficient in Formula (19), and  $K$  is the movement step parameter of the pixel, which is set to 3.

### 2.3. Multi-Directional Gradient Scale Segmentation Model (MDGSS)

In the process of detecting small and weak targets, there are still some noise points and residual edge contours in the differential image after background modeling, which leads to the problem of target confusion. This article proposes a multi-directional gradient scale segmentation model (MDGSS) to eliminate the noise and edge contours of targets in differential images, thereby increasing the saliency of the targets and improving their discrimination. The dim and small multi-objective image segmentation method based on the circular window proposed by Jiang et al. [46] has been widely applied. This method first defines the size of the inner and outer windows of the circular window, then, it calculates the average gray value of the pixels in the corresponding window, and finally, it compares the difference between the average gray value of the two windows and the set threshold to complete image segmentation. The specific computational model is as follows [46]:

$$\begin{cases} P_I = \sum_{i=-\omega}^{\omega} \sum_{j=-\eta}^{\eta} f(\alpha + i, \beta + j) \\ P_O = \sum_{i=-(\omega+h)}^{\omega+h} \sum_{j=-(\eta+h)}^{\eta+h} f(\alpha + i, \beta + j) \\ P_{IAvg} = \frac{P_I}{\omega \times \eta} \\ P_{OAvg} = \frac{P_O}{\omega \times \eta} \end{cases} \quad (21)$$

where  $P_I$ ,  $P_O$  represents the mean value of pixels in the inner and outer windows of the looping window, respectively.  $f(\alpha, \beta)$  represents the grayscale value of pixels in the neighborhood,  $(i, j)$  denotes the row and column numbers of pixels in the neighborhood,  $\omega$ ,  $\eta$  represent the length and width of the return window,  $h$  denotes the neighborhood size of the window outside the circular window, and  $P_{IAvg}$  and  $P_{OAvg}$  represent the mean value of the pixels in the inner and outer windows of the circular window, respectively. The average value of the current outer window pixel is  $P_{OAvg}$  and that of the inner window pixel mean value is  $P_{IAvg}$ . When the difference value of  $P_{OAvg}$  and  $P_{IAvg}$  is greater than the set threshold  $Th$ , the current pixel  $f(\alpha, \beta)$  position is filled with 1 element, and the other pixel positions that do not meet the conditions are filled with 0 elements to complete image segmentation.

However, it is found in the research that the simple contour window segmentation algorithm is only applicable to the target segmentation of a single pixel, which is inconsistent with the feature that the target is distributed in blocks. After segmentation, the target usually occupies only 1–2 pixel positions, which will cause the loss of target detection when the segmentation threshold is properly raised. Therefore, combined with the detection idea of the area block research in this paper, the contour window is improved in

this paper. A segmentation model with multi-directional gradient scale is proposed to segment the target points. This model combines the uneven distribution of target signals in the imaging of small and weak targets, integrates the double window segmentation model with multi-scale, segments the target by adjusting different window sizes, retains the information of the target in multiple directions, and improves the target's attention. The defined computational model is as follows:

$$\left\{ \begin{array}{l} P_{upI} = \sum_{i=-\omega}^{\omega} \sum_{j=-\eta}^{\eta} D(\alpha + i - l, \beta + j) \\ P_{upO} = \sum_{i=-(\omega+h)}^{\omega+h} \sum_{j=-(\eta+h)}^{\eta+h} D(\alpha + i - l, \beta + j) \\ P_{downI} = \sum_{i=-\omega}^{\omega} \sum_{j=-\eta}^{\eta} D(\alpha + i + l, \beta + j) \\ P_{downO} = \sum_{i=-(\omega+h)}^{\omega+h} \sum_{j=-(\eta+h)}^{\eta+h} D(\alpha + i + l, \beta + j) \\ P_{leftI} = \sum_{i=-\omega}^{\omega} \sum_{j=-\eta}^{\eta} D(\alpha + i, \beta + j + l) \\ P_{leftO} = \sum_{i=-(\omega+h)}^{\omega+h} \sum_{j=-(\eta+h)}^{\eta+h} D(\alpha + i, \beta + j + l) \\ P_{rightI} = \sum_{i=-\omega}^{\omega} \sum_{j=-\eta}^{\eta} D(\alpha + i, \beta + j - l) \\ P_{rightO} = \sum_{i=-(\omega+h)}^{\omega+h} \sum_{j=-(\eta+h)}^{\eta+h} D(\alpha + i, \beta + j - l) \end{array} \right. \quad (22)$$

where  $P_{upI}$ ,  $P_{downI}$ ,  $P_{leftI}$ ,  $P_{rightI}$ ,  $P_{upO}$ ,  $P_{downO}$ ,  $P_{leftO}$ , and  $P_{rightO}$  represent the sum of the pixels in the upper, lower, left, and right inner and outer window areas of a single pixel, respectively, and  $l$  represents the scale range defined by the segmentation area.  $D$  represents the difference image modeled by the gradient reciprocal background in Formula (20).  $D(\alpha, \beta)$  represents the pixel grayscale, and  $(i, j)$  denotes the row and column numbers of pixels in the neighborhood. The size of the inner and outer windows is adjusted by setting the pixel scale to  $l$ , which conforms to the characteristics of uneven energy distribution when imaging weak targets. According to the traditional loopback window segmentation principle, this paper proposes segmentation models corresponding to each direction according to multi-scale characteristics, and the specific models are as follows:

$$\left\{ \begin{array}{l} P_{upIAvg} = \frac{P_{upI}}{\omega \times \eta}, P_{upOAvg} = \frac{P_{upO} - P_{upI}}{\omega \times \eta} \\ R = abs(P_{upIAvg} - P_{upOAvg}) \\ P_{downIAvg} = \frac{P_{downI}}{\omega \times \eta}, P_{downOAvg} = \frac{P_{downO} - P_{downI}}{\omega \times \eta} \\ R1 = abs(P_{downIAvg} - P_{downOAvg}) \\ P_{leftIAvg} = \frac{P_{leftI}}{\omega \times \eta}, P_{leftOAvg} = \frac{P_{leftO} - P_{leftI}}{\omega \times \eta} \\ R2 = abs(P_{leftIAvg} - P_{leftOAvg}) \\ P_{rightIAvg} = \frac{P_{rightI}}{\omega \times \eta}, P_{rightOAvg} = \frac{P_{rightO} - P_{rightI}}{\omega \times \eta} \\ R3 = abs(P_{rightIAvg} - P_{rightOAvg}) \end{array} \right. \quad (23)$$

where  $P_{upIAvg}$ ,  $P_{downIAvg}$ ,  $P_{leftIAvg}$ ,  $P_{rightIAvg}$ ,  $P_{upOAvg}$ ,  $P_{downOAvg}$ ,  $P_{leftOAvg}$ , and  $P_{rightOAvg}$  represent the pixel mean value of the inner and outer window areas of the upper, lower, left and right regions of the pixel, respectively, and  $R$ ,  $R1$ ,  $R2$ ,  $R3$  represent the pixel mean difference of the inner and outer window areas of the upper, lower, left and right regions, respectively. The image segmentation is completed by comparing the mean difference with the set threshold  $T$ . According to the principle of multi-scale gradient, the final segmentation result should be determined by summing up the four differences. However, considering that the target energy is distributed in regions, this paper determines the segmentation result by performing or operating on the above four results to preserve the target information and enhance the discrimination. The mathematical model is as follows:



$$\begin{cases} P = [R, R1, R2, R3] \\ \text{if } P(i_1, j_1) \geq T \\ \text{Result}(i_1, j_1) = 1 \\ \text{else} \\ \text{Result}(i_1, j_1) = 0 \\ F_g = \text{Result}(j_1) | \text{Result}(j_1 + 1) | \text{Result}(j_1 + 2) | \text{Result}(j_1 + 3) \end{cases} \quad (24)$$

where  $P$  represents the mean difference set of inner and outer window pixels in 4 directions,  $(i_1, j_1)$  represents the sequence number in the set  $P$ , and  $\text{Result}(i_1, j_1)$  represents the result of image filling 0 and 1 after segmentation in 4 directions, respectively.  $F_g$  represents the binary image after segmentation. The comparison diagram of the two segmentation algorithms is as follows.

As shown in the Figure 5, it can be observed that the traditional double window segmentation algorithm can only segment the central target when segmenting the target area, and the information after the target neighborhood segmentation is lost, which does not achieve the goal of maximum target signal retention. The segmentation model proposed in this article can segment targets based on different defined scales, with a complete preservation of target information and significant improvement in target discrimination, laying a good foundation for improving the model detection rate.

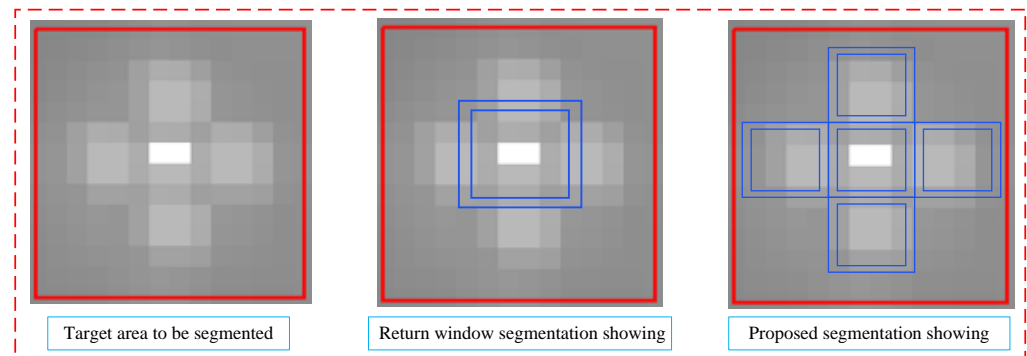


Figure 5. Image of loop window segmentation and multi-dimensional scale segmentation.

#### 2.4. Multi-Frame Energy-Sensing Detection Model (MFESD)

Intending to successfully extract the target and achieve multi-frame object detection in the sequence to output the target’s trajectory, this paper combines the idea of pipeline filtering to construct a multi-frame energy-sensing detection model (MFESD) to extract the target. First, we calculate the number of times that the target appears in the sequence of a binary image, and we calculate the average gray value of the target in each frame after segmentation according to the number of times. Next, we calculate the grayscale correlation value of the average grayscale, and finally, we average the grayscale based on the number of occurrences. The obtained value is the similar grayscale value of the target. The defined mathematical model is as follows:

$$\begin{cases} M_s = \sum_{t=1}^M \text{Mov}(F_g(x, y, t)) \\ L_b(x, y, t) = \sum_{-m_1}^{m_1} \sum_{-n_1}^{n_1} F_g(x + m_1, y + n_1, t) \times f(x + m_1, y + n_1, t) \\ \bar{L}_b(x, y, t) = \frac{1}{M} \times \sum_{t=1}^M L_b(x, y, t) \\ H_{GS} = \left\{ \sum_{-m_1}^{m_1} \sum_{-n_1}^{n_1} \left[ \prod_{i=1}^M \bar{L}_b(x + m_1, y + n_1, t) \right] \right\} / M_s \end{cases} \quad (25)$$

where  $M_s$  represents the total number of moves of the target in the figure,  $M$  represents the total number of frames, and  $F_g(x, y, n)$  represents the binary image that has been

segmented in Formula (24).  $(x, y)$  and  $t$  represent the coordinate position of the target in the image and the number of image frames, respectively.  $L_b(x, y, t)$  represents the gray value within the local range of the target, and  $(m_1, n_1)$  represents the coefficient of pixel position transformation,  $m_1 = n_1 = \varepsilon/2$ .  $\varepsilon$  represents the selection of neighborhood size for calculating pixel grayscale, with a value of 3 in the text.  $\overline{L}_b(x, y, t)$  represents the average gray value of the image, and  $H_{GS}$  represents the energy perception function, which is obtained by calculating the ratio of the correlation of multi-frame image energy to the number of target movements.

### 2.5. Overall Steps and Flow Diagram of Algorithm

To summarize, the pseudo-code of the overall steps can be summarized as follows in Algorithm 1:

---

#### Algorithm 1 Overall steps of the algorithm

---

Input original image  $f$ ;

Initialization parameters;

Step 1. Choose the calculated region  $f_1$  from original image  $f$ ;

Step 2. Use Formulas (1)–(6) to complete the initial aggregation and secondary aggregation of image energy, respectively, and output the preprocess image  $FF$ ;

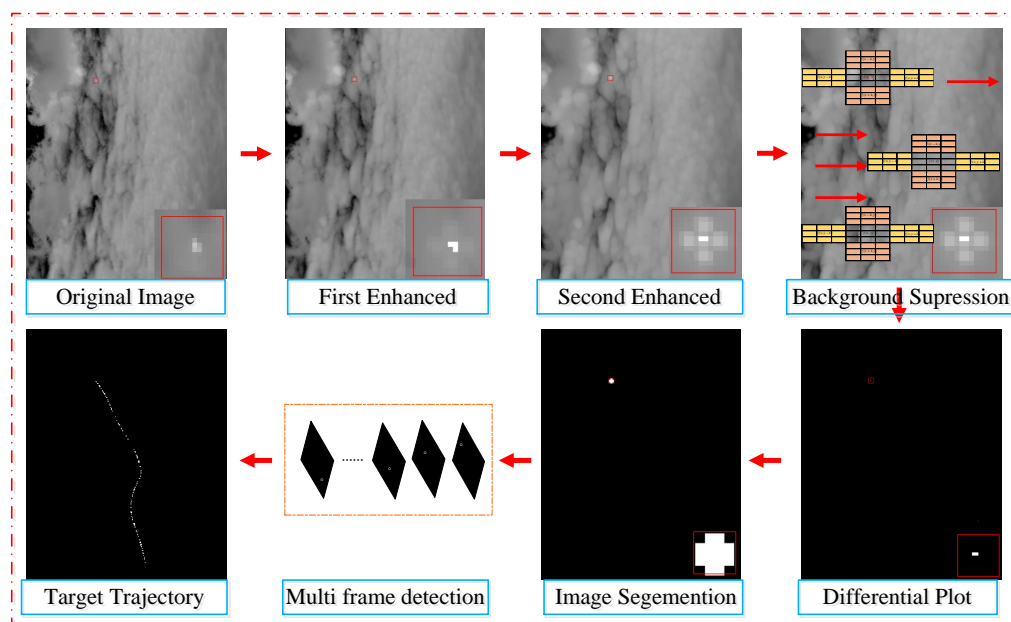
Step 3. Finish the background suppression by Formulas (14)–(19), and output the difference image  $D$ ;

Step 4. According to the background constrict, remove the noise to complete the segmentation by Formulas (21)–(23);

Step 5. Extract the target and output the trajectory result by Formula (24);

---

The algorithm flow chart of this paper is as follows Figure 6.



**Figure 6.** Overall flow chart of algorithm.

### 2.6. Evaluation Indicators

In terms of background processing, in order to better reflect the effect of the algorithm in this paper on the suppression of a multi-cloud background, the paper uses the signal gain (IC), background suppression factor (BSF), and background structure similarity (SSIM) of each algorithm to evaluate the prediction background of each algorithm. The specific definitions of the above three indicators are as follows [51–53]:

$$\left\{ \begin{array}{l}
 T_{in} = \frac{1}{l \times l} \sum_{x_g=-l}^{x_g=l} \sum_{y_g=-l}^{y_g=l} f_{in}(mt + x_g, nt + y_g) \\
 B_{in} = \frac{1}{l_1 \times l_1} \sum_{x_g=-l_1}^{x_g=l_1} \sum_{y_g=-l_1}^{y_g=l_1} f_{in}(mt + x_{g_1}, nt + y_{g_1}) \\
 T_{out} = \frac{1}{l \times l} \sum_{x_g=-l}^{x_g=l} \sum_{y_g=-l}^{y_g=l} f_{out}(mt + x_g, nt + y_g) \\
 B_{out} = \frac{1}{l_1 \times l_1} \sum_{x_g=-l_1}^{x_g=l_1} \sum_{y_g=-l_1}^{y_g=l_1} f_{out}(mt + x_{g_1}, nt + y_{g_1}) \\
 C_{in} = |T_{in} - B_{in}| / |T_{in} + B_{in}| \\
 C_{out} = |T_{out} - B_{out}| / |T_{out} + B_{out}| \\
 IC = C_{out} / C_{in}
 \end{array} \right. \tag{26}$$

In the equation,  $T_{in}$ ,  $B_{in}$  and  $T_{out}$ ,  $B_{out}$  represent the mean of neighboring pixels in the input image and output image, respectively.  $(mt, nt)$  represents the target coordinate,  $l = 1$ ,  $l_1 = 4$  represents different neighborhood radi, and  $C_{in}$  and  $C_{out}$  represent the signal gain of the target in the input and output images, while  $IC$  represents the signal gain of the target before and after image processing.

$$BSF = \sigma_{in} / \sigma_{out} \tag{27}$$

$$SSIM = \frac{(2\mu_R\mu_F + \varepsilon_1)(2\sigma_{RF} + \varepsilon_2)}{(\mu_R^2 + \mu_F^2 + \varepsilon_1)(\sigma_R^2 + \sigma_F^2 + \varepsilon_2)} \tag{28}$$

where  $\sigma_{in}$  and  $\sigma_{out}$  are the mean square deviations of the input image and difference image, respectively.  $\mu_R$  represents the average of the pixels in the entire image;  $\sigma_R$  represents the standard deviation of all pixels in the image;  $\sigma_{RF}$  represents the covariance between the original image and the background image after modeling; and  $\varepsilon_1$  and  $\varepsilon_2$  are a set of constant parameters.

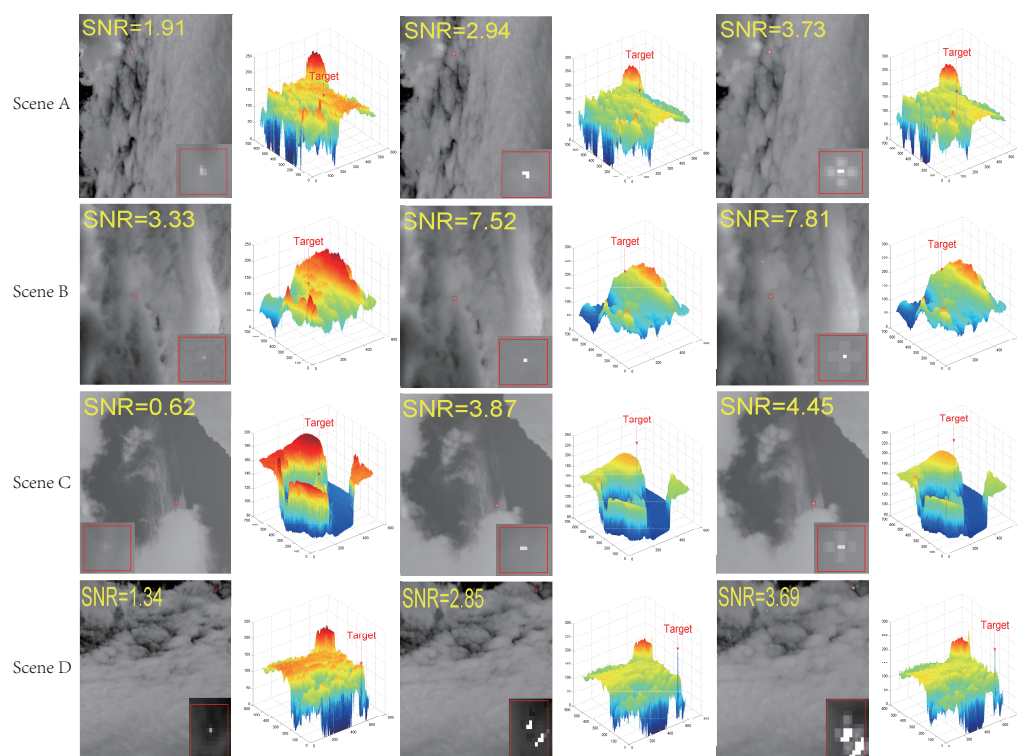
### 2.7. Scenario Selection and Preliminary Analysis

In order to reflect the effect of this algorithm on target detection, this paper selects 4 sequence images under complex clouds for preprocessing, background suppression, differential segmentation, and sequence target detection. The original image and the image processed by the energy aggregation model proposed in this paper are shown as follows in Figure 7.

As shown in the figure, in a multi-cloud scenario, the clouds fluctuate significantly and the target differentiation is not high. However, traditional spatiotemporal filtering and low-rank sparse recovery theories often suppress the target as a background in such scenarios, and after background modeling, the edge contours of multiple clouds are preserved, which will seriously affect the detection of the target. Therefore, this article proposes a local energy aggregation model, which aggregates the energy distribution of targets in different directions before conducting target detection. The detection rate of weak small targets in such scenes has been greatly improved, and the first frame of the four selected sequence scenes has the following effect after energy aggregation. The details of the four selected sequence images are shown in the following Table 1.

**Table 1.** Detail of sequence image.

Scene	Size	Target Size	Target Detail
Scene A	641 × 513	2 × 2	Moving birds in complex cloud backgrounds.
Scene B	641 × 513	3 × 3	Moving birds in complex cloud backgrounds.
Scene C	641 × 513	3 × 3	Moving birds in complex cloud backgrounds.
Scene D	180 × 180	3 × 3	UAV with complex cloud background motion.



**Figure 7.** Original sequence diagram and original three-dimensional diagram.

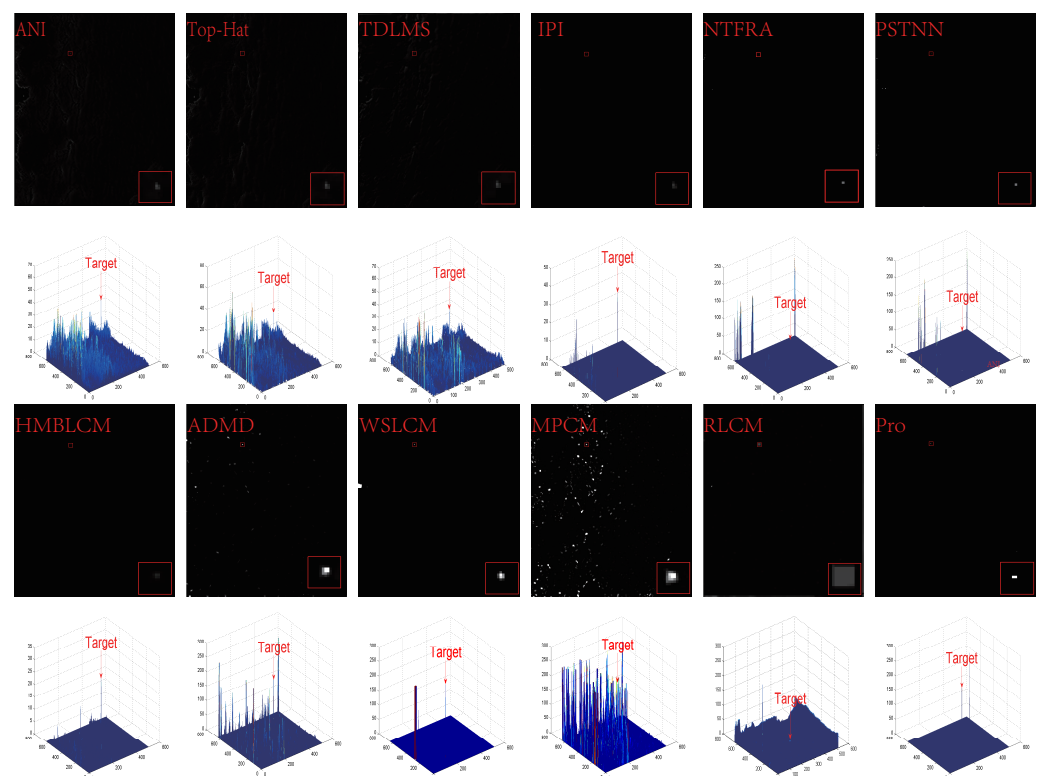
### 3. Results

#### 3.1. Analysis of Algorithm Background Suppression Results

Using the above four scenes, this paper applies the Anisotropic (ANI) model [54], improved top hat filter [11], TDLMS filter [22], IPI model [37], NTFRA model [55], PSTNN model [8], HMBLCM model [56], WSLCM model [34], MPCM model [33], RLCM model [32], ADMD model [57] and the algorithm proposed in this paper to suppress the background of the multi-cloud image, and then, it combines the multi-frame energy-sensing algorithm proposed in this paper to extract the target. In this section, we will give the background map, difference map and three-dimensional difference map of each algorithm on four sequences to evaluate and compare the algorithms; then, we analyze the relevant experimental data. The specific experimental results are as shown in Figures 8–11.

As shown in Figures 8–11, we can analyze the three-dimensional difference graph obtained from the four sequences under the processing of the seven algorithms in this paper. It can be seen from the figure that the algorithm proposed in this paper is superior to other algorithms in image background suppression and weak target signal retention. Among them are the background image, difference image, and three-dimensional image of the difference image obtained by anisotropic algorithm (ANI) processing. It can be seen from the figure that because the anisotropic algorithm uses the gradient information around a single pixel in combination with the diffusion function to suppress the background, it cannot effectively suppress the edge contour in the background, so the target points and more edge noise points in the difference map are retained. From the three-dimensional figure, it can be seen that the distribution of dendrites in the difference map is chaotic, and weak targets cannot be highlighted. The top hat filter preprocesses the image background to obtain the background image, the difference image and the three-dimensional structure of the difference image. As shown in the figure, it can be observed that the effects of top hat filtering in different sequences are obviously different, which indicates that the method of top hat filtering relies on the structural elements of the filter template to perform background processing and has a large defect in the target detection image with large cloud fluctuation and prominent edges. The method of using the relationship between adjacent pixels to distinguish between targets and backgrounds to complete background suppress-

sion is not applicable to images with strong edges. It will make the difference graph retain more edge information stronger than the target signal, making it difficult to distinguish between the target and the background. For images with large differences between image pixels and structural elements, the algorithm has a poor processing effect and large adaptability. Two-dimensional minimum mean square filtering (TDLMS) is used to process the background image, difference image and three-dimensional image of the difference image. This algorithm performs background processing on the information of adjacent pixels in the X and Y directions in the neighborhood pixel matrix. The processing effect is ideal in the image with a mild background, but the background processing is poor in the image with a large fluctuation in background transformation. Only calculating the information of adjacent pixels in two directions will cause confusion in distinguishing the target and background when predicting the position of the point transformation. As shown in the figure, this algorithm will sharpen the edge so that the energy of image edge information is far greater than the energy of the target location, which makes the detection target receive edge energy information, leading to target detection failure. The IPI model processes four sequences of background images, differential images and three-dimensional images of differential images. Similarly, the IPI model is also affected by the moving step size and filtering window, and it cannot perform benign iteration to suppress the background of the image, so that more false alarms remain in the image, and it is more sensitive to images with uneven light distribution, which makes the image form a dark and bright distinct fault, bringing difficulties to target detection. The NTFRA model combines the LogTFNN model, local tensor model, and HTV model to improve the IPT model. However, based on the scene difference 3D map, it was found that the model has a high false alarm rate and poor background modeling performance.



**Figure 8.** Background modeling results in scenario A.

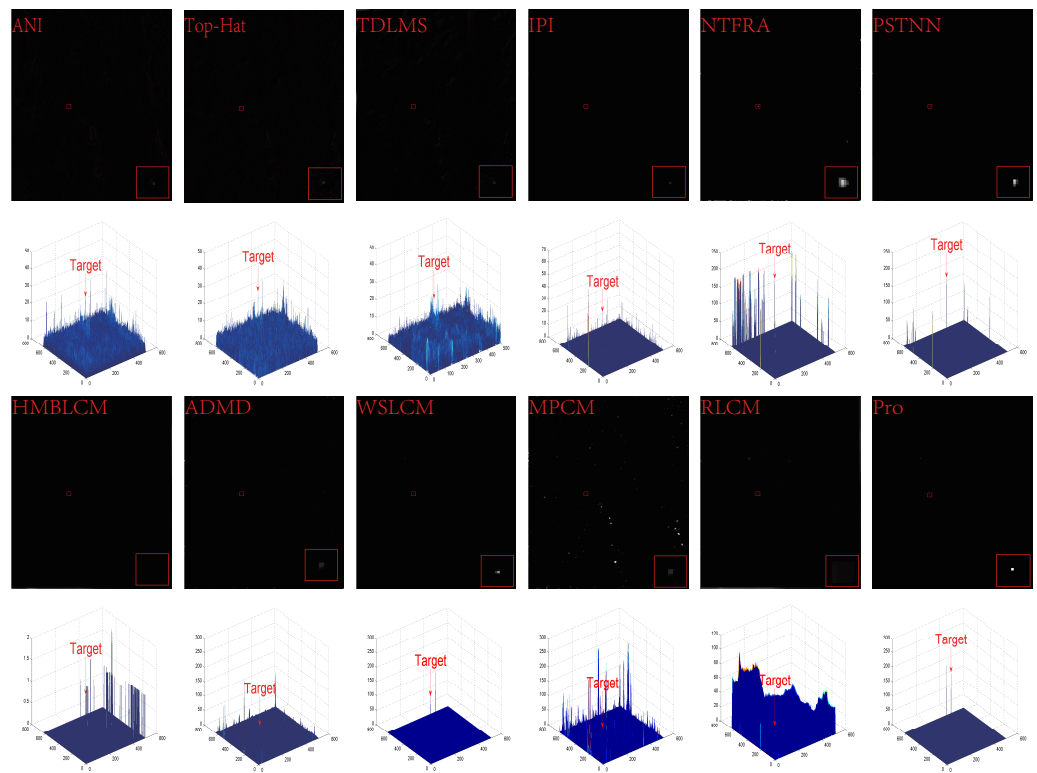


Figure 9. Background modeling results in scenario B.

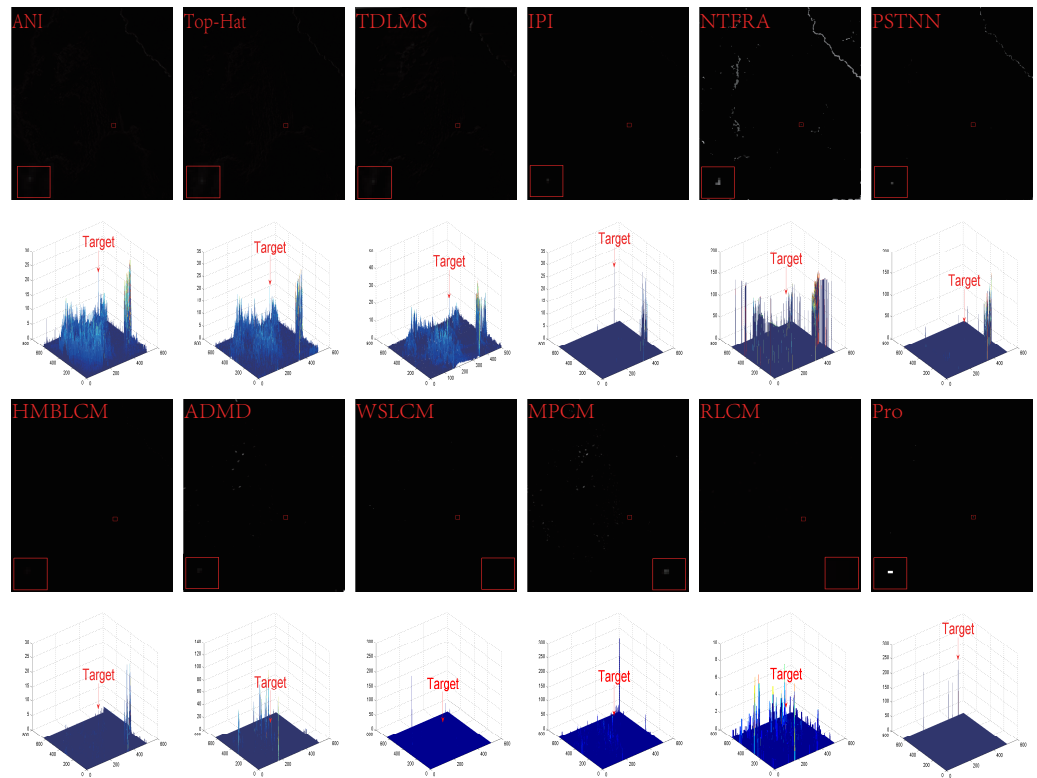
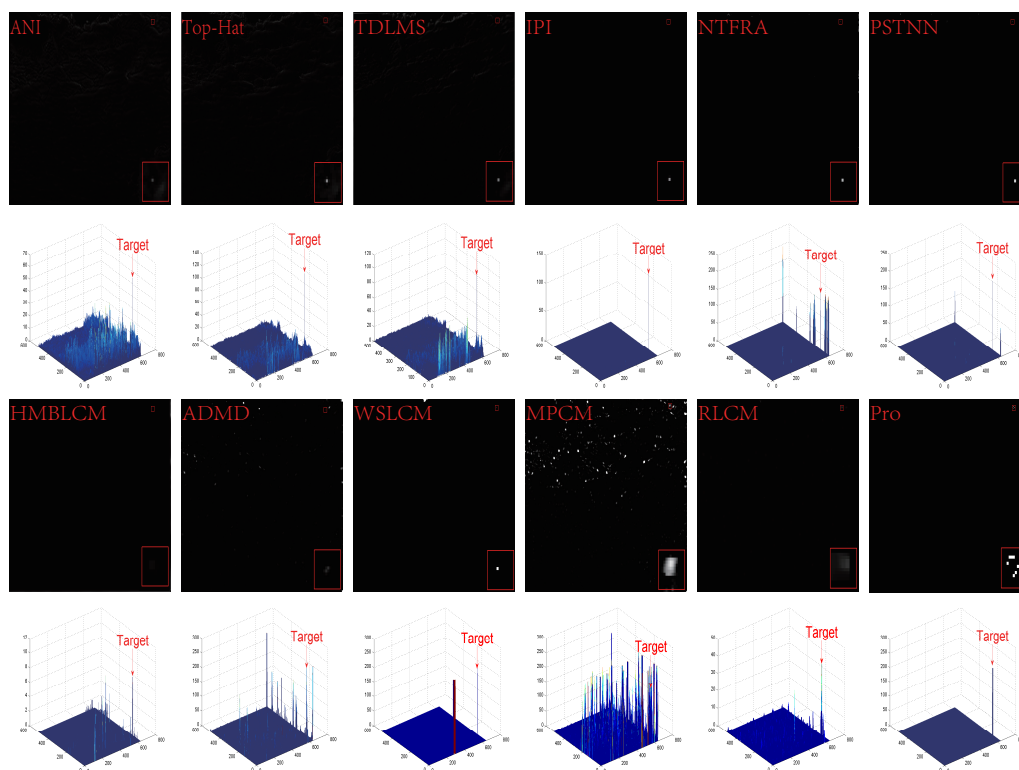


Figure 10. Background modeling results in scenario C.





**Figure 11.** Background modeling results in scenario D.

In the differential results of RLCM, WSLCM, and MPCM models, it can be observed that using local significance operations can effectively enhance the target signal. This provides a good theoretical basis for the research direction of visual saliency, such as the ADMD model and HMBLCM model. The former constructs a local inner and outer window discrimination mechanism to enhance the energy of the target, solving the difficulty of AAGD detection in strong edge scenes. The latter first proposes an improved IHBF model based on the HBF model to enhance the energy of the target. Then, on the basis of the LCM model, an adaptable MLCM model is proposed to combine with IHBF, making the target more energetic during background suppression processing. As shown in the differential and three-dimensional diagrams in Figures 8–11, the background modeling results of ADMD and HMBLCM still retain high false alarms and low target discrimination. It can be seen from the figure that in the four sequences, the background modeling method for local information calculation of area blocks proposed in this paper according to the background suppression idea of traditional gradient algorithm has a remarkable effect, which can be observed from the corresponding three-dimensional graph of the difference image, and it can completely suppress the edge contour of the multi-cloud layer in the scene, so that the protruding areas containing targets are enhanced and retained. It shows that the algorithm in this paper has certain feasibility in background suppression in the image with cloudy layer noise, and it can well retain the information of target points, effectively highlighting target points, which provides a basis for subsequent target point segmentation and extraction.

### 3.2. Analysis of Background Modeling Data of Each Algorithm

After background modeling, in order to accurately reflect the feasibility of the algorithm proposed in this paper, this section uses the signal gain (IC), background suppression factor (BSF), and background structure similarity (SSIM) defined above to show the performance of the algorithm in target detection of complex cloud background images. The larger

the index value, the more obvious the effect of the algorithm. The evaluation data of each algorithm in different scenarios are shown in the Table 2.

**Table 2.** The indexes acquired from different methods.

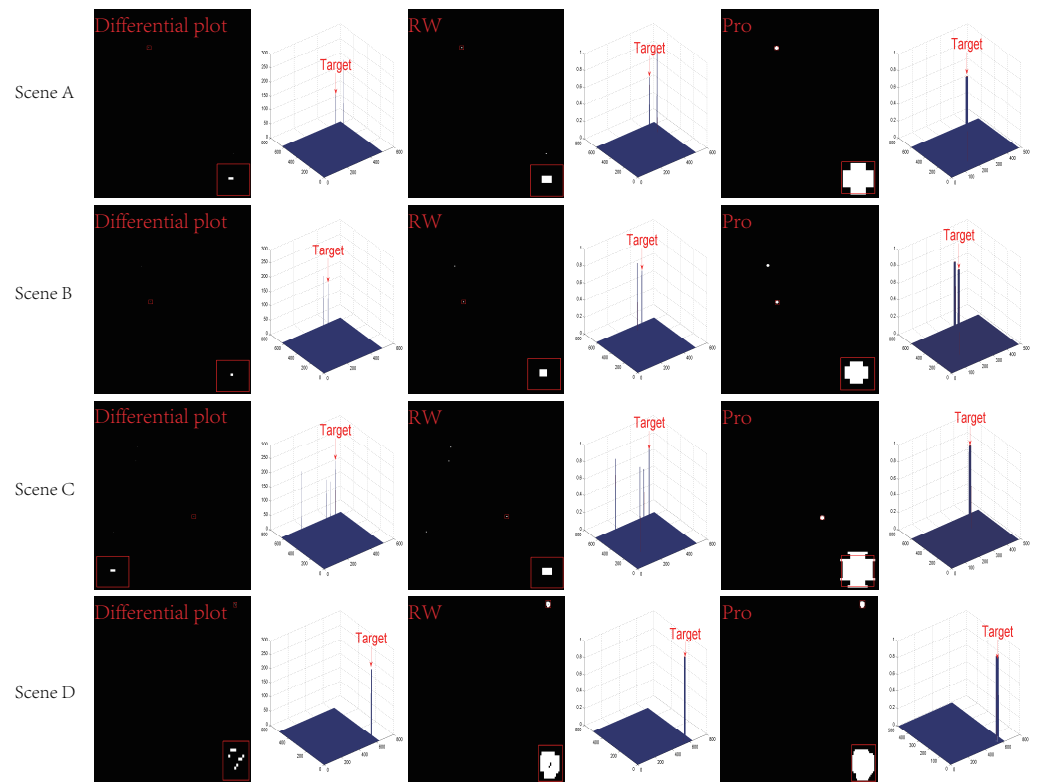
Frame	Scene A			Scene B			Scene C			Scene D		
Method\Index	IC	BSF	SSIM	IC	BSF	SSIM	IC	BSF	SSIM	IC	BSF	SSIM
ANI [54]	9.0203	91.3200	0.9939	25.8218	183.6795	0.9984	12.4826	169.9695	0.9982	1.1732	106.0274	0.9955
Top hat [11]	10.5927	94.4869	0.9802	107.1684	148.0800	0.9854	31.2586	204.7889	0.9913	0.3410	107.8749	0.9840
TDLMS [22]	45.1087	92.5729	0.9622	132.4805	123.1463	0.9872	33.1432	74.4022	0.9876	21.6857	93.8758	0.9630
IPI [37]	2.1981	128.8429	0.8830	0.4425	281.7852	0.8564	0.2368	184.9115	0.6771	0.7470	318.0104	0.8803
NTFRA [55]	12.5237	164.3357	0.9995	NaN	83.0662	0.9985	NaN	62.7185	0.9958	9.3106	139.4590	0.9992
PSTNN [8]	12.5237	181.5125	0.9999	NaN	350.7793	0.9999	NaN	452.772	0.7857	9.3106	423.6726	0.9979
HMBLCM [56]	10.9859	53.6475	0.9830	222.6875	66.5249	0.9892	66.5943	34.7095	0.9686	33.0000	69.8265	0.9899
ADMD [57]	10.9577	58.2625	0.9903	34.9862	240.4317	0.9993	22.6822	120.7594	0.9976	7.5804	63.8970	0.9944
WSLCM [34]	11.8859	54.8368	0.9998	222.6875	612.1966	0.9999	NaN	538.087	0.9999	51.0000	63.8913	1.0000
MPCM [33]	9.8687	19.5679	0.9673	222.6875	63.2017	0.9999	66.5943	123.2548	0.9972	51.0000	23.2017	0.9794
RLCM [32]	1.6760	53.3073	0.9830	49.5246	23.8160	0.9536	1.4083	15.3287	0.9161	11.4136	5.9400	0.9168
Proposed	43.5325	502.4472	0.9998	35.8208	482.0345	0.9999	56.1780	314.6325	0.9999	13.9145	193.5220	0.9997

As shown in the above table, the data obtained from differential image gain (IC), background suppression factor (BSF), and structural similarity (SSIM) calculations on four sequences of images can be observed to be outstanding in background modeling, and the algorithm proposed in this paper performs well in background estimation and reconstruction. Its background suppression ability reaches a maximum of 502.4472, and the reconstructed background has a similarity of over 99% with the original image. In the signal indicator IC, the maximum target signal gain can reach 56 dB, indicating that the model proposed in this article can meet the target detection requirements.

### 3.3. Comparison and Analysis of Difference Graph Segmentation Results

In summary, some noise points are still preserved in the differential image after background modeling in complex multi-cloud scenes. In order to enhance the visual discrimination of the target while removing noise, this section uses the multi-directional scale segmentation model proposed in this paper and the recurrent window segmentation algorithm (RW) [58] to segment and compare the difference results of the above four scenes on the basis of the same segmentation threshold to verify the effectiveness and feasibility of the algorithm proposed in this paper. The specific segmentation results are shown in the following Figure 12.

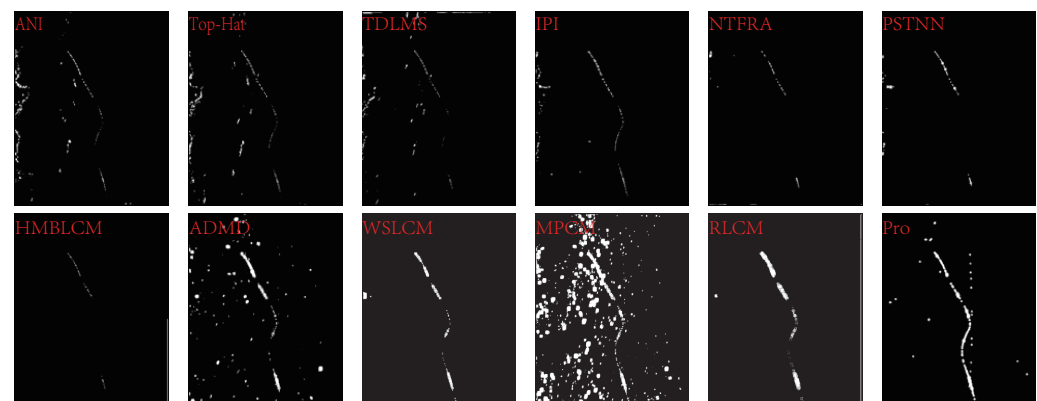
As shown in Figure 12, on the basis of the same segmentation threshold, the MDGSS model proposed in this paper has a significantly higher saliency of the target after segmenting the difference map compared to the traditional double window segmentation algorithm. The reason is that the algorithm in this article can control the corresponding segmentation scale for image segmentation, which better adapts to the uneven energy distribution during the target motion process. To some extent, it solves the problem of target confusion caused by residual noise segmentation in traditional dual window segmentation models that only use set windows and thresholds for image segmentation. The research idea of using the multi-scale characteristics of targets for segmentation and extraction is feasible. It not only enhances the saliency of the segmented targets but also improves the detection rate of the model and reduces the false alarm rate.



**Figure 12.** Comparison of segmentation results between the loop window segmentation algorithm and the multi-directional scale segmentation algorithm.

### 3.4. Analysis of Multiple Frame Trajectory Detection Results in Differential Images

After the above energy enhancement, background modeling, and multi-directional gradient segmentation, the feasibility and effectiveness of the proposed algorithm in detecting weak and small targets are demonstrated. On the basis of the above experiments, this section mainly conducts tracking research on the motion trajectories of weak and small targets in the sequence of multiple frames after background modeling. The specific effects are shown in the following Figures 13–16.



**Figure 13.** Detection trajectories of algorithms in scene A.

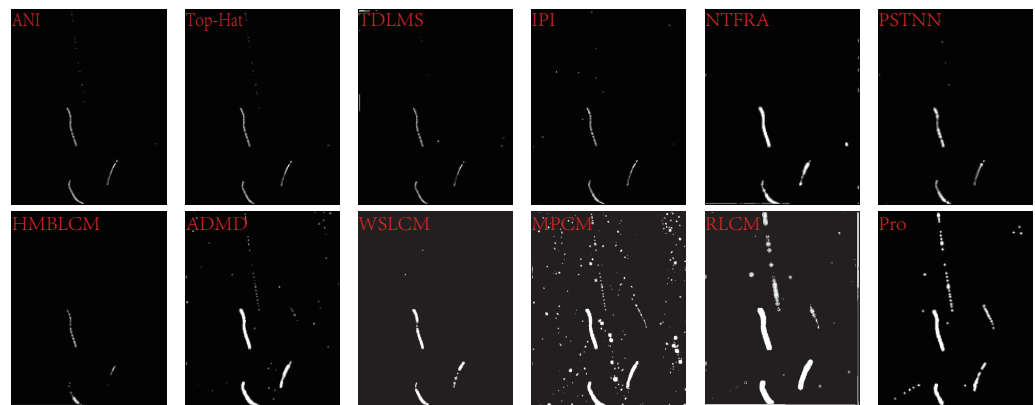


Figure 14. Detection trajectories of algorithms in scene B.

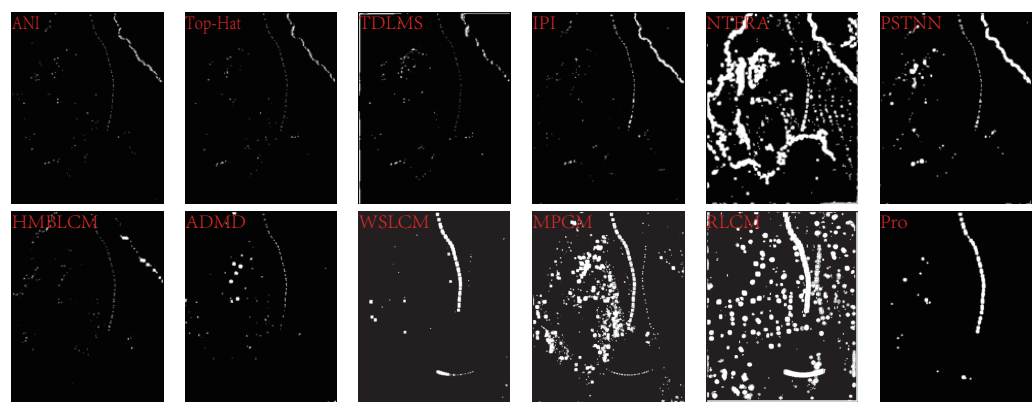


Figure 15. Detection trajectories of algorithms in scene C.

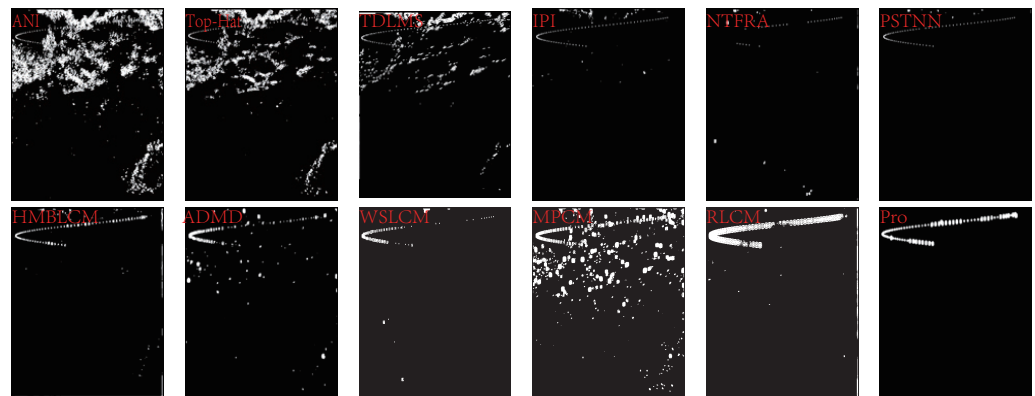


Figure 16. Detection trajectories of algorithms in scene D.

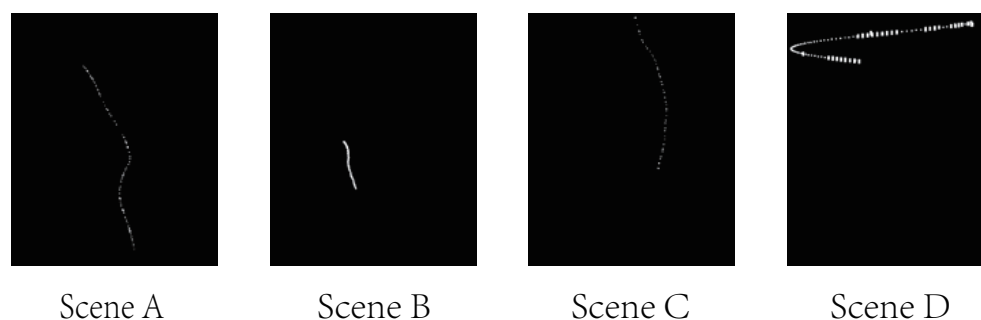
As shown in Figures 13–16, the trajectories are, respectively, ANI filtering, top hat filtering, TDLMS filtering, the IPI model, the NTFRA model, the PSTNN model, the HMBLCM model, and the ADMD model. The four sequences of target motion trajectories are output by the algorithm proposed in this paper. From the graph, it can be observed that the algorithm proposed in this paper has a significant detection effect in scenes with multiple cloud layers. It can not only output the running trajectory of the target more completely but also completely suppress information such as edge contours, which has good discrimination and achieves multi-frame detection of the target. Meanwhile, other algorithms have obvious differences in detection results when facing the cloudy scenes with complex backgrounds. ANI, TDLMS and top hat filter detection in traditional filtering can detect targets in four sequences, but many background edge contours are still preserved

in the following figure of the same threshold segmentation, which shows that the algorithm model using only filtering is not applicable in the cloudy scenes with complex backgrounds, and the background suppression is incomplete. There are many false alarm targets in the image. In the low-rank sparse recovery theory, the detection effects of the IPI model, NTFRA algorithm, and PSTNN model, which are compared in this paper in four scenarios, reflect the advantages of the low-rank theory in background modeling. As shown in the above figure, under the same segmentation condition, the cloud profile in the target track output by such an algorithm is less than that of the traditional filtering model, which shows that the low-rank theory is feasible in background modeling. However, in the face of a cloudy scene, when the target shuttles through the clouds, it may often be suppressed as a sparse component, while the cloud edge with a strong contour is retained, resulting in the phenomenon of high false alarm detection. Meanwhile, in the HMBLCM model, ADMD model, WSLCM model, MPCM model, and RLCM model of visual saliency, the local multi-directional grayscale aggregation model constructed in this paper performs well in preserving target signal processing. In the trajectory of sequence difference, it can be observed that the background modeling model constructed on the basis of energy aggregation in this paper has complete suppression of the background, and the target signal is clearly preserved. After the background modeling, the target's trajectory can be output with lower false alarm detection.

### 3.5. Analysis of Multi-Frame Energy-Sensing Detection Results

In the above figure, although the algorithm in this article has achieved multi-frame object detection in sequence images, there are still some false alarm target interferences that confuse the extraction of real targets in the figure. In order to extract the real target completely to output the real target track, this paper uses the gray energy perception of the target motion between frames to extract the target, avoiding the phenomenon of target loss caused by the pipeline limitation of traditional pipeline filtering. The specific detection results are as follows:

As shown in Figure 17, after using the multi-frame energy-sensing detection model (MFESD), the noise in scenes A, B, C and D in Figures 13–16 are removed, and the target trajectories in the four sequence scenes are separately extracted, resulting in obvious target motion trajectories. The overall detection rate of the sequence reaches over 95%, and the target motion trajectory output is complete, indicating that the multi-frame energy-sensing detection model (MFESD) proposed in this paper can effectively capture the temporal information of target motion, and to some extent, it has met the requirements of multi-frame target detection in the sequence.



**Figure 17.** Target detection trajectories in 4 scenarios of the energy perception detection model.

### 3.6. Algorithm ROC Analysis

Intending to further reflect the detection effect of the algorithm in this paper from a mathematical perspective, this section conducts statistical analysis on the detection rate and false alarm rate of the above algorithms, and it draws the corresponding receiver operating characteristic according to the data for corresponding analysis. Its mathematical calculation model is as follows [59]:

$$P_d = \frac{NTDT}{NT} \times 100\%$$

$$P_f = \frac{NFDT}{NP} \times 100\%$$
(29)

where  $P_d, P_f$  represents the detection rate and false alarm rate, respectively,  $NT$  represents the total number of images containing target points in the sequence image,  $NTDT$  represents the total number of images containing targets detected in the actual situation,  $NP$  represents the total number of all pixels in the calculated image, and  $NFDT$  represents the total number of targets that cannot be detected in the actual situation. The ROC curves of each algorithm in the four scenarios are as follows in Figure 18.

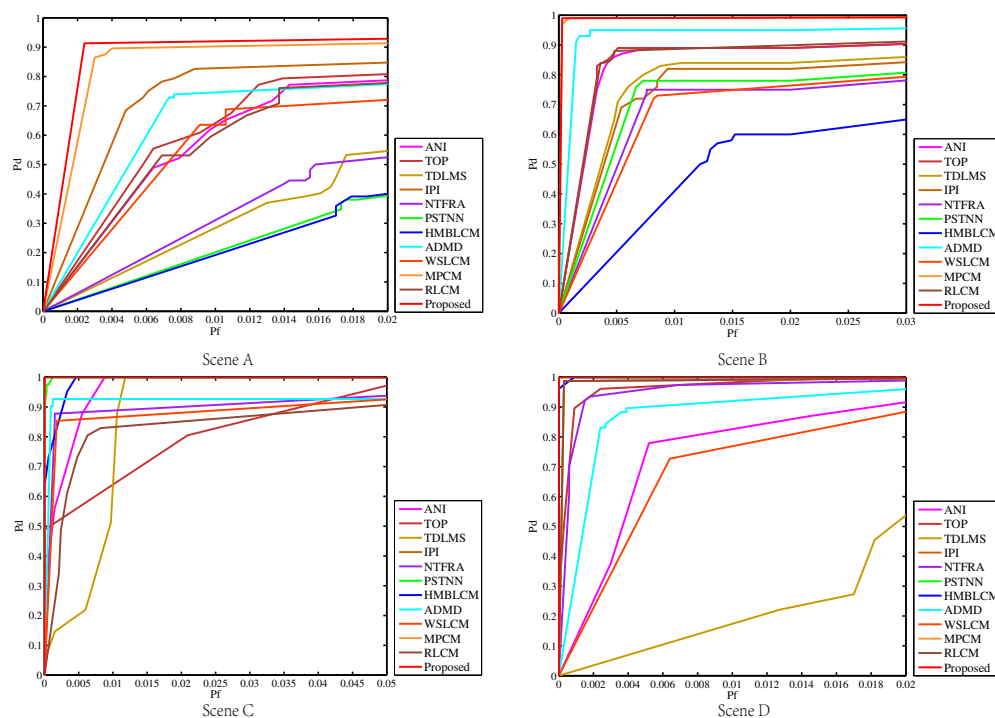


Figure 18. ROC curves in 4 scenarios of each algorithm.

In Figure 18, the detection rate of the algorithm proposed in this paper in the detection of small targets in the multi-cloud background has reached more than 90%, which shows that the algorithm in this paper has a good edge grinding effect on the background of the multi-cloud and multi-edge contour, and it can highlight the target while reducing the sharp edge contour, reduce the appearance of false alarm targets, and reduce the false alarm rate of detection. Its detection rate  $P_d$  in the four scenes has reached 91.3043%, 96%, 100%, and 100% respectively, and the maximum false alarm rate  $P_f$  is 0.0024% (no more than 0.01), indicating that the algorithm proposed in this paper has met the requirements of small target detection, and the detection effect is outstanding in the multi-cloud layer scene. It can be observed from the ROC diagrams of the four scenes that the detection rates of the traditional filtered ANI, TDLMS and top hat filter are quite different. The highest detection rates ( $P_d$ ) of the three algorithms in the three scenes are 77.1739%, 89%, 100%, and 87.0129% of the ANI algorithm, 53.2608%, 84%, 100%, and 45.4545% of the TDLMS, and 79.3478%, 89%, 100%, and 100% of the top hat filter, respectively. To some extent, this reflects the poor effect of background modeling in complex cloud scenes through filtering and the high target loss rate. It can be seen from the figure that the false alarm rate associated with the traditional algorithm is relatively high, and the  $P_f$  of the three algorithms is more than 0.01%, which also indicates that the traditional background modeling algorithm needs to be constantly improved to better adapt to complex scenes.



On the other hand, the low-rank sparse recovery theory also has poor performance in object detection in complex scenes with multiple cloud layers. In scenes A and B with sharp and prominent edge contours, such algorithms have lower detection rates and higher false alarms. When performing low-rank recovery, the strong edge contour is sparse processed as the target, while the real target is restored as the low rank part, resulting in the phenomenon of target detection loss. Among the visual saliency detection models, the highest detection rates for scene A and scene B are 89.58% and 99% of the MPCM model. In scenario C and scenario D, the detection rate of the visual saliency class model can reach 100%, reflecting the superiority of the model.

### 3.7. Comparison of Computational Model Complexity

In order to enable the algorithm in this article to adapt to practical engineering applications, this section conducts statistical analysis of the running time of the above models in four scenarios. The specific experimental data are shown as follows.

As shown in Table 3, ANI filtering, top hat filtering, TDLMS filtering, the NTFRA model, the HMBLCM model, the ADMD model, the WSLCM model, and the MPCM model have relatively short running times, with an average running time of no more than 2 s, and have great potential in practical applications. However, based on the above research, it has been found that the background modeling effect of these models is not ideal, and the detection false alarm rate is high. The PSTNN model, due to its excellent low-rank inversion operation, greatly reduces the operation time of the model while ensuring the detection rate. The mean operation time of the four scenarios is 1.2758 s. It can be clearly observed from the table that the detection time of the IPI model, the RLCM model, and the model proposed in this article is relatively long. The average running time of the algorithm in this article is 38.6120 s, indicating that while suppressing complex backgrounds, the optimization of the model's computational time still needs to be urgently addressed.

**Table 3.** Calculation of background modeling complexity in different scenarios of each model (unit: frames/second).

Model/Scene	Scene A	Scene B	Scene C	Scene D
	Time consumption			
ANI [54]	1.5682	1.3824	1.1728	1.1622
Top-Hat [11]	0.0806	2.3197	2.1920	2.168
TDLMS [22]	0.7323	3.1451	2.9622	2.9862
IPI [37]	20.1913	25.5687	21.5469	22.6102
NTFRA [55]	7.4413	7.7037	9.3783	9.4596
PSTNN [8]	1.3455	0.7817	1.5313	1.4449
HMBLCM [56]	0.0951	0.0621	0.0642	0.0617
ADMD [57]	0.1818	0.1235	0.1655	0.1220
WSLCM [34]	5.0788	6.6469	5.1151	5.2533
MPCM [33]	0.3727	0.3762	0.3716	0.3757
RLCM [32]	41.0290	52.0089	37.5410	41.5173
Pro	38.4084	40.0951	37.5508	38.3938

## 4. Discussion

To sum up, the feasibility and adaptability of the algorithm proposed in this paper for object detection in multi-cloud backgrounds can be demonstrated. After the above experiments, the detection model proposed in this paper can be discussed as follows:

- (1) Firstly, the energy enhancement model of multi-directional grayscale aggregation proposed in this article can effectively enhance the signal of the target based on the small correlation between the target and the background, and the enhancement of the target signal-to-noise ratio is evident. In addition, the model implements different enhancement strategies for different regions, significantly narrowing the grayscale difference of strong edge contours, making the background relatively flat and the

target singular and prominent in the form of regional blocks, improving the grayscale value and discrimination of the target in the original image. To some extent, it has changed the phenomenon of target detection failure caused by weak target signals in detection using only original imaging. However, this model still retains some of the noise, which affects the identification of the target during background modeling. In the later stage, a corresponding region interest extraction algorithm can be constructed based on the prominent characteristics of the target grayscale to obtain the target area, increasing the adaptability of the algorithm.

- (2) Based on the research work in (1), this paper proposes a background modeling model based on region blocks based on a gradient reciprocal to complete image background modeling. Transforming the traditional gradient reciprocal background modeling model processed by a single pixel into regional pixels for local imaging improves the utilization of target signals and the detection rate of algorithms. However, in the experiment, it was found that due to the complexity of local region operations, the modeling time for background modeling is longer than that of traditional gradient reciprocal background modeling models, and further optimization of the algorithm is still needed in the later stage.
- (3) In target extraction, this paper proposes a multi-directional and multi-scale segmentation model to segment candidate targets in the difference map to obtain real targets. After experiments, it was found that the effect was good because the segmentation model proposed in this article can adapt to the irregularity of energy distribution during target imaging based on the scale of the target imaging. In comparison with traditional double window segmentation models, it was found that the proposed segmentation model in this paper can improve the discrimination of the target while removing noise. The target area is significantly larger than the target area after double window segmentation, achieving the goal of target segmentation.
- (4) In order to fully output the motion trajectory of the sequence target, this paper proposes a multi-frame energy perception detection model to complete the detection of multiple frames. The experimental results show that the model effectively outputs the motion trajectory of the target, and the overall detection rate of the sequence image can reach over 90%. Introducing a pixel grayscale into multi-frame detection can reflect the significance of the local signal of the target, improve the accuracy of determining real targets between sequence frames, and achieve target detection and tracking. After research, it was found that the model is closely related to the background modeling algorithm. Due to the fact that the background modeling algorithm proposed in this article requires a lot of computation time on the selected sequence scenes, the algorithm takes a longer time for multi-frame detection. Therefore, future researchers can try to combine different background modeling methods to achieve the multi-frame detection of targets.

## 5. Conclusions

In order to improve the detection capability of dim target detection systems, this paper proposes a dim and small target detection based on the energy sensing of local multi-directional gradient information. After the above experimental verification, the model proposed in this paper is summarized as follows:

- (1) In background modeling for complex backgrounds with multiple cloud layers, the energy aggregation model EMDGA can effectively utilize local information of the image to aggregate and enhance the target signal, highlighting the target signal, and improving the target signal-to-noise ratio by an average of 3.12 dB. In addition, the model fuses sharp cloud edges with background information to achieve a preliminary smooth image effect, laying the foundation for the background modeling of subsequent images.
- (2) Based on the energy aggregation model EMDGA, the traditional filtering background modeling in multi-cloud scenarios is not effective. The LMDGA background modeling model constructed in this article has an SSIM index of over 99% for the structural

similarity between the reconstructed background and the original image, and it has an average background suppression factor BSF of 373.1591. The average target signal gain IC in the differential plot reaches 37.3615 dB, indicating that the constructed local region background modeling model has certain applicability and innovation in multi-cloud scenarios.

- (3) To address the issue of detection failure caused by the uncertainty of the threshold during the target extraction process, the MDGSS model proposed in the article can amplify the target signal while segmenting the image, improve the saliency of the target, and effectively eliminate the interference of noise in the differential image.
- (4) Considering the instability of the moving target signal in the time domain space, there is still some noise left in the binary image after segmentation, resulting in low target discrimination. The MFESD model constructed in this article is based on the local grayscale singularity characteristics of the target for research. Through the characteristics of grayscale accumulation and target motion, the real target is detected, and the target's trajectory is output. The overall detection rate of the sequence scene is over 95%.

**Author Contributions:** Conceptualization, X.F. and J.L.; methodology, J.L., X.F., L.M., L.F., L.Y. and Z.X.; software, J.L., X.F., L.M., L.F., L.Y. and Z.X.; validation, J.L., X.F., L.M., L.F., L.Y. and Z.X.; formal analysis, J.L., X.F., L.M., L.F., L.Y. and Z.X.; investigation, J.L., X.F., L.M., L.F., L.Y. and Z.X.; resources, J.L., X.F., L.M., L.F., L.Y. and Z.X.; data curation, J.L., X.F., L.M., L.F., L.Y. and Z.X.; writing—original draft preparation, J.L.; writing—review and editing, J.L., X.F., L.M., L.F., L.Y. and Z.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China under Grant 62261004 and 62001129.

**Data Availability Statement:** Not applicable.

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

## Abbreviations

The following abbreviations are used in this manuscript:

EMDGA	Energy Enhancement Model for Multi-Directional Gray Aggregation
LMDGR	Local Multi-Directional Gradient Reciprocal
MDGSS	Multi-Directional Gradient Scale Segmentation
MFESD	Multi-Frame Energy-Sensing Detection
ANI	Anisotropy
PSTNN	Partial Sum of Tensor Nuclear Norm
NTFRA	Non-Convex Tensor Fibered Rank Approximation
NRAM	Non-Convex Rank Approximation Minimization
HB-MLCM	High-Boost-Based Multi-Scale Local Contrast Measure
ADMD	Absolute Directional Mean Difference
WSLCM	Weighted Strengthened Local Contrast Measure
MPCM	Multi-Scale Patch-based Contrast Measure
RLCM	Relative Local Contrast Measure
RW	Recurrent Window
SSIM	Structural Similarity Image
SNR	Signal-to-Noise Ratio
IC	Contrast Gain

## References

1. Deng, L.; Zhang, J.; Xu, G.; Zhu, H. Infrared small target detection via adaptive M-estimator ring top-hat transformation. *Pattern Recognit.* **2021**, *112*, 107729. [\[CrossRef\]](#)
2. Ahmadi, K.; Salari, E. Small dim object tracking using frequency and spatial domain information. *Pattern Recognit.* **2016**, *58*, 227–234. [\[CrossRef\]](#)

3. Zhou, H.X.; Zhao, Y.; Qin, H.L.; Yin, S.M.; Liu, G.; Zhao, D.; Yan, X.; Rong, S.H. Infrared Dim and Small Target Detection Algorithm Based on Multi-scale Anisotropic Diffusion Equation. *Acta Photonica Sin.* **2015**, *44*, 910002. [[CrossRef](#)]
4. Qi, H.; Mo, B.; Liu, F.X.; He, Y.; Liu, S. Small infrared target detection utilizing Local Region Similarity Difference map. *Infrared Phys. Technol.* **2015**, *71*, 131–139. [[CrossRef](#)]
5. Nasiri, M.; Chehresa, S. Infrared small target enhancement based on variance difference. *Infrared Phys. Technol.* **2017**, *82*, 107–119. [[CrossRef](#)]
6. Li, Q.; Nie, J.; Qu, S. A small target detection algorithm in infrared image by combining multi-response fusion and local contrast enhancement. *Optik* **2021**, *241*, 166919. [[CrossRef](#)]
7. Liu, H.-K.; Zhang, L.; Huang, H. Small target detection in infrared videos based on spatio-temporal tensor model. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8689–8700. [[CrossRef](#)]
8. Zhang, L.; Peng, Z. Infrared small target detection based on partial sum of the tensor nuclear norm. *Remote Sens.* **2019**, *11*, 382. [[CrossRef](#)]
9. Lu, Y.; Huang, S.C.; Zhao, W. Sparse representation based infrared small target detection via an online-learned double sparse background dictionary. *Infrared Phys. Technol.* **2019**, *99*, 14–27. [[CrossRef](#)]
10. Lu, F.X.; Li, J.Y.; Chen, Q.; Chen, G.L.; Peng, R. Weak target detection based on Top hat transform of PM model. *Syst. Eng. Electron.* **2018**, *40*, 1417–1422.
11. Deng, L.Z.; Zhu, H.; Zhou, Q.; Li, Y.S. Adaptive top-hat filter based on quantum genetic algorithm for infrared small target detection. *Multimed. Tools Appl.* **2018**, *77*, 10539–10551. [[CrossRef](#)]
12. Fan, X.S. *Sequence Image Weak Target Detection and Tracking Algorithm Research*; University of Electronic Science and Technology of China, Chengdu, China, 2019; pp. 22–33.
13. Bai, X.Z.; Zhou, F.G. Analysis of new top-hat transformation and the application for infrared dim small target detection. *Pattern Recognit.* **2010**, *43*, 2145–2156. [[CrossRef](#)]
14. Wang, Y.H.; Liu, W.N. Dim target enhancement algorithm for low-contrast image based on anisotropic diffusion. *Opto-Electron. Eng.* **2008**, *35*, 15–19.
15. Perona, P.; Malik, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 629–639. [[CrossRef](#)]
16. Lin, Q.; Huang, S.C.; Wu, X.; Zhong, Y. Infrared small target detection based on nuclear anisotropic diffusion. *High Power Laser Part. Beams* **2015**, *27*, 93–98.
17. Zhu, H.; Guan, Y.; Deng, L.; Li, Y.; Li, Y. Infrared moving point target detection based on an anisotropic spatial-temporal fourth-order diffusion filter. *Comput. Electr. Eng.* **2018**, *68*, 550–556. [[CrossRef](#)]
18. Fan, X.S.; Xu, Z.Y.; Zhang, J.L.; Huang, Y.; Peng, Z.M. Dim small targets detection based on self-adaptive caliber temporal-spatial filtering. *Infrared Phys. Technol.* **2017**, *85*, 465–477. [[CrossRef](#)]
19. Hadhoud, M.M.; Thomas, D.W. The two-dimensional adaptive LMS (TDLMS) algorithm. *IEEE Trans. Circuits Syst.* **1988**, *35*, 485–489. [[CrossRef](#)]
20. Cao, Y.; Liu, R.M.; Yang, J. Small target detection using two-dimensional least mean square (TDLMS) filter based on neighborhood analysis. *Int. J. Infrared Millim. Waves* **2008**, *29*, 188–200. [[CrossRef](#)]
21. Li, H.; Wang, Q.; Wang, H.; Yang, W. Infrared small target detection using tensor based least mean square. *Comput. Electr. Eng.* **2021**, *91*, 106994. [[CrossRef](#)]
22. Wang, L.L.; Wang, M. Infrared small target detection based on TDLMS filter with neighborhood information. *Nat. Sci. Ed.* **2015**, *43*, 178–180.
23. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the IEEE Sixth International Conference on Computer Vision, Bombay, India, 7 January 1998; pp. 839–843.
24. Bae, T.-W. Small target detection using bilateral filter and temporal crossproduct in infrared images. *Infrared Phys. Technol.* **2011**, *54*, 403–411. [[CrossRef](#)]
25. Zeng, Y.Q.; Chen, Q. Dim and Small Target Background Suppression Based on Improved Bilateral Filtering for Single Infrared Image. *Infrared Technol.* **2011**, *33*, 537–540.
26. Li, Z.Z.; Dong, N.L. Adaptive filtering-based detection of weak targets in strong undulating backgrounds. *J. Instrum.* **2004**, *25*, 663–665.
27. Li, J.C.; Sheng, Z.K. Detection method for moving point targets in infrared undulating background. *Infrared Laser Eng.* **1997**, *26*, 8–13.
28. Zhang, X.; Ru, J.; Wu, C. A nonparametric regression-based multi-scale gradient correlation filtering method for infrared small target detection. *Electronics* **2023**, *12*, 1562. [[CrossRef](#)]
29. Chen, C.L.P.; Wei, Y.T.; Xia, T.; Tang, Y.Y. A local contrast method for small infrared target detection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 574–580. [[CrossRef](#)]
30. Han, J.H.; Ma, Y.; Zhou, B.; Fan, F.; Liang, K.; Fang, Y. A robust infrared small target detection algorithm based on human visual system. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 2168–2172.
31. Han, J.H.; Moradi, S.; Faramarzi, I.; Liu, C.; Zhang, H.; Zhao, Q. A local contrast method for infrared small-target detection utilizing a Tri-Layer window. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1822–1826. [[CrossRef](#)]
32. Han, J.H.; Liang, K.; Zhou, B.; Zhu, X.; Zhao, J.; Zhao, L. Infrared small target detection utilizing the multiscale relative local contrast measure. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 612–616. [[CrossRef](#)]

33. Wei, Y.T.; You, X.; Li, H. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recognit.* **2016**, *58*, 216–226. [[CrossRef](#)]
34. Han, J.H.; Moradi, S.; Faramarzi, I.; Zhang, H.; Zhao, Q.; Zhang, X.; Li, N. Infrared small target detection based on the weighted strengthened local contrast measure. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 1670–1674. [[CrossRef](#)]
35. Fan, J.L.; Gao, Y.M.; Wu, Z.H.; Li, L. Infrared dim small target detection technology based on RPCA. In Proceedings of the International Conference on Electronic Information Technology and Intellectualization, Guangzhou, China, 2–3 December 2017; pp. 748–752.
36. Wang, C.; Qin, S. Adaptive detection method of infrared small target based on target-background separation via robust principal component analysis. *Infrared Phys. Technol.* **2015**, *69*, 123–135. [[CrossRef](#)]
37. Gao, C.; Meng, D.; Yang, Y.; Wang, Y.; Zhou, X.; Hauptmann, A.G. Infrared patch-image model for small target detection in a single image. *IEEE Trans. Image Process.* **2013**, *22*, 4996–5009. [[CrossRef](#)]
38. Wang, X.; Peng, Z.; Kong, D.; Zhang, P.; He, Y. Infrared dim target detection based on total variation regularization and principal component pursuit. *Image Vis. Comput.* **2017**, *63*, 1–9. [[CrossRef](#)]
39. Xu, E.; Wu, A.; Li, J.; Chen, H.; Fan, X.; Huang, Q. Infrared Target Detection Based on Joint Spatio-Temporal Filtering and L1 Norm Regularization. *Sensors* **2022**, *22*, 6258. [[CrossRef](#)]
40. Hu, Y.; Ma, Y.; Pan, Z.; Liu, Y. Infrared Dim and Small Target Detection from Complex Scenes via Multi-Frame Spatial–Temporal Patch-Tensor Model. *Remote Sens.* **2022**, *14*, 2234. [[CrossRef](#)]
41. Sun, Y.; Yang, J.G.; An, W. Infrared Dim and Small Target Detection via Multiple Subspace Learning and Spatial-Temporal Patch-Tensor Model. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 3737–3752. [[CrossRef](#)]
42. Rawat, S.S.; Singh, S.; Alotaibi, Y.; Alghamdi, S.; Kumar, G. Infrared target-background separation based on weighted nuclear norm minimization and robust principal component analysis. *Mathematics* **2022**, *10*, 2829. [[CrossRef](#)]
43. Huang, Y.; Jin, L. Deep learning-based detection of weak targets in airspace. *Inf. Technol. Informatiz.* **2020**, *6*, 217–220.
44. Bai, J.J.; Zhang, H.Y.; Li, Z.J. The generalized detection method for the dim small targets by faster R-CNN integrated with GAN. In Proceedings of the IEEE 3rd International Conference on Communication and Information Systems, Singapore, 28–30 December 2018; pp. 1–5.
45. Zhao, M.; Cheng, L.; Yang, X.; Feng, P.; Liu, L.; Wu, N. TBC-Net: A realtime detector for infrared small target detection using semantic constraint. *arXiv* **2015**, arXiv:2001.05852.
46. Hou, Q.; Wang, Z.; Tan, F.; Zhao, Y.; Zheng, H.; Zhang, W. RISTDnet: Robust infrared small target detection network. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 7000805. [[CrossRef](#)]
47. Nian, B.; Jiang, B.; Shi, H.; Zhang, Y. Local Contrast Attention Guide Network for Detecting Infrared Small Targets. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5607513. [[CrossRef](#)]
48. Liu, S.; Wu, R.; Qu, J.; Li, Y. HPN-SOE: Infrared Small Target Detection and Identification Algorithm Based on Heterogeneous Parallel Networks with Similarity Object Enhancement. *IEEE Sensors J.* **2023**, *23*, 13797–13809. [[CrossRef](#)]
49. Xu, H.; Zhong, S.; Zhang, T.; Zou, X. Multi-scale Multi-level Residual Feature Fusion for Realtime Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5002116. [[CrossRef](#)]
50. Fan, X.S.; Xu, Z.Y. Dim and small target detection based on local energy aggregation degree of sequence images. *Int. J. Opt.* **2019**, *2019*, 9282141.
51. Fan, X.S.; Li, J.L.; Chen, H.J.; Min, L.; Li, F. Dim and small target detection based on improved hessian matrix and F-Norm collaborative filtering. *Remote Sens.* **2022**, *14*, 4490. [[CrossRef](#)]
52. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
53. Hu, X.; Wang, X.; Yang, X.; Wang, D.; Zhang, P.; Xiao, Y. An infrared target intrusion detection method based on feature fusion and enhancement. *Def. Technol.* **2020**, *16*, 737–746. [[CrossRef](#)]
54. Li, J.; Fan, X.; Chen, H.; Li, B.; Min, L.; Xu, Z. Dim and small target detection based on improved spatio-temporal filtering. *IEEE Photonics J.* **2022**, *14*, 7801211.
55. Kong, X.; Yang, C.; Cao, S.; Li, C.; Peng, Z. Infrared small target detection via nonconvex tensor fibered rank approximation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5000321. [[CrossRef](#)]
56. Shi, Y.; Wei, Y.; Yao, H.; Pan, D.; Xiao, G. High-boost-based multiscale local contrast measure for infrared small target detection. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 33–37. [[CrossRef](#)]
57. Moradi, S.; Moallem, P.; Sabahi, M.F. Fast and robust small infrared target detection using absolute directional mean difference algorithm. *arXiv* **2020**, arXiv:1810.03173.
58. Jiang, H.J.; Liu, W.; Liu, Z.H. Segmentation of small and weak multi-target image based on circular window. *Acta Photonica Sin.* **2007**, *36*, 2168–2171.
59. Qiu, Z.B.; Ma, Y.; Fan, F.; Huang, J.; Wu, M.-h.; Mei, X.-g. A pixel-level local contrast measure for infrared small target detection. *Def. Technol.* **2022**, *18*, 1589–1601. [[CrossRef](#)]

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