Infrared Small Target Detection, High-Precision Localization and Segmentation: Using TDU Kernel

Chang DING [1,2], Siyang CHEN [3], Hai LIU [2,], Zhendong LUO [1], Jie ZHANG [2]*

[1] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China
[2] School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin 541004, China
[3] College of Communication and Information Technology, Xi'an University of Science and Technology, Xi'an 710699, China

dingchang@guet.edu.cn, 22407020230@stu.xust.edu.cn, sieeoe@cumt.edu.cn*, 22012302083@mails.guet.edu.cn, zhangjie@mails.guet.edu.cn

Submitted September 20, 2024 / Accepted October 29, 2024 / Online first November 12, 2024

Abstract. *Aiming at the challenge of infrared small target detection with different shape and size under the different scene, a novel algorithm architecture is proposed using the kernel of Target Detection Unit (TDU). The TDU incorporates the fractal geometry design and dual-scale structure, which can execute three main sub-tasks: preliminary target detection, target localization with high precision and target segmentation by pixel-level. First, the principle establishes a dual-scale target detection structure, selects the central point, decomposes the scale information and constructs the Integrated Local Contrast Saliency (ILCS) map, the target preliminary result is obtained by the visual attention mechanism of "top to bottom". Second, the principle adopts the scale-recursion algorithm by the mechanism of "bottom to up" to locate the target precisely from the preliminary result along with Area Optimal Recommend Mechanism (AORM) strategy. At last, the separated local histogram is used to segment the target by per-pixel with suitable threshold. From the experimental results, conducted across five different types of infrared-scenes including infrared sky scene, infrared maritime scene, backlight illuminance scene, infrared scene with interference and infrared scene with small & dim target, we observe the performance of high accuracy rate and remarkable robustness.*

Keywords

Infrared target detection, target detection unit, target localization with high-precision, target segmentation by per-pixel, area optimal recommend mechanism

1. Introduction

Infrared small target detection has extensive applications in infrared warning [1–3], precise guidance [4–6], and maritime surveillance [7]. Compared with typical target, infrared small target exhibits the following characteristics [8–10]: (1) Small size: high-distance infrared imaging makes the true target appear with small size. (2) Dim gray value: low Signal-to-Clutter Ratio (SCR) or submerge into the background easily. (3) Shapeless or low texture characteristic: infrared target can appear point-shape, line-shape or circle-shape characteristic, additionally the infrared target's inner region has less gray value change. (4) Easy to vary: external environmental change, illumination change or target motivation often happens.

Some theoretical frameworks have been proposed for detecting infrared small target, which can be broadly categorized into two main groups: (1) Feature extraction and processing. (2) Deep learning frameworks.

The first principle encompasses the methods such as feature multi-layer information processing [11], [12], background modeling [13], target feature information enhancement and fusion [14], [15], tensor form expansion of feature information [17], background filtering [18], diffusion of partial differential equations [19], background subtraction [20], and sparse representation and low-rank information extraction [21]. Researchers have developed numerous approaches within these above-mentioned frameworks, also proposing the models and measures. For example, Chen et al. [22] introduced the Local Contrast Method (LCM), which focuses on enhancing and fusing target features to describe the disparity between a point and its surrounding area. Tang et al. [23] further improved upon LCM by combining the background filtering, resulting in an efficient method for detecting suspicious target with low time complexity. Similar to LCM, Absolute Average Gray Difference (AAGD) [24] and Absolute Directional Mean Difference (ADMD) [25] algorithms have been proposed. AAGD establishes a mathematical model for suspected targets with high precision, particularly effective for detecting dim infrared target. On the other hand, ADMD generates a saliency map based on absolute directional mean difference, but selecting an appropriate threshold for target detection remains challenging. Zhang et al. [26] developed a mathematical model aiming at capturing

DOI: 10.13164/re.2024.0721

high-dimension features of suspected target within infrared image, producing notable efficiency, especially in detecting blurred infrared target. Xia et al. [27] also contributed a mathematical model proficient in capturing high-precision features, particularly effective in detecting the sporadic infrared target. Kong et al. [17] introduced the Nonconvex Tensor Fibered Rank Approximation (NTFRA) algorithm, combining tensor form expansion and sparse representation with low-rank information extraction to effectively suppress clutter. However, the algorithm exhibits high time complexity. Zhang et al. [28] utilized low-rank sparse tensor decomposition to detect the infrared small target, but its complexity may limit the adaptiveness of targets with different sizes. Tunc et al. [29] focuses on the changes in the Signal-to-Noise Ratio (SNR) before and after target detection and considers that extracting the saliency of the target is an effective method for detecting dim and small infrared targets.

Among the more general deep learning-based target recognition frameworks are the two classical models: one is Faster Region-based Convolutional Neural Network (Faster R-CNN) [30], the other one is You Only Look Once (YOLO) [31]. The main principle of Faster R-CNN lies in the convolutional operation between multiple regional candidate boxes and specific operators, involving a large number of samples and high-dimensional training characteristic data [30], [31]. The main advantage of the YOLO model for object detection lies in its refined region proposal mechanism. However, its limitation in small object detection primarily stems from the fact that training with a large number of samples can easily lead to overfitting. In fact, for small target characteristics, we can draw on YOLO's strategy for selecting region proposals. By extracting effective feature information and performing decomposition, fusion, and reconstruction across scales, small target recognition can be achieved through computation on a single frame image.

Our paper's workflow diagram can be shown in Fig. 1, it divides the entire task into three sequential subtasks: preliminary target detection, high-precision target localization, and target segmentation. There are two decision-making processes in both preliminary target detection and high-precision target localization stages. The integration of information between different scales and threshold filtering is used to decide the preliminary detection regions of the targets, and the high-precision localization results are determined by merged adjacent TDU regions and scale matching within the preliminary detection regions. Ultimately, by performing binary processing on local region based on the central coordinate of the target's high-precision localization, the segmentation result of the target on per-pixel basis can be obtained.

The main contribution of this paper is the decomposition, fusion and reconstruction of target information at different scales, achieving the segmentation of infrared small targets from coarse to fine.

2. The Preliminary Target Area Detection by TDU

2.1 TDU's Dual-Scale Geometry Design

As shown in Fig. 2, the outermost box, represented by a solid black line, indicates the overall outline of the entire TDU. The entire TDU contains a large target area (size: $w_1 \times w_1$, marked by blue color) and eight equal-sized background regions (size: $w_1 \times w_1$). Each large target area can be subdivided into nine small areas (size: $w_0 \times w_0$) with corresponding small backgrounds (size: $w_0 \times w_0$). Additionally, the total size of the TDU is $w_2 \times w_2$. Big-scale's Local Contrast Saliency ($LCS_{big\text{-}scale}$) is measured through evaluating the differences between the central target (T area) and its surrounding backgrounds (named from B_1 area to B_8 area). As for small-scale's LCS measurement ($LCS_{small\text{-}scale}$), each small target (named from t_1 area to t_9 area of size $w_0 \times w_0$) and its backgrounds (named from b_1^1 to b_1^8) are similarly assessed to determine their values.

The central point (x_0, y_0) of the TDU lies within the center of the T region and the t_5 region, covering the target's all possible locations in the entire image after traversal. According to SPIE standards for small target, which should be less than 0.12% with area proportion of the entire image area, so we can adjust the TDU's different size for different images. In this case, w_1 is defined as the largest integer that can be multiplied by three and needs to be less than or equal to $0.035 \times \sqrt{W \times H}$, w_0 is set as one-third of w_1, and w_2 is established as 3 times by w_1, which can format the fractal geometry's structure in TDU.

To sum up, TDU employs a fractal geometric design structure that includes both big-scale and small-scale level's feature data to measure target's gray value variation against background, enabling the target detection with different shape and size. Moreover, we have established a mathematical model for local contrast saliency based on two kinds of target contrast properties: positive and negative, represented by (1) and (2) respectively, where I represents the potential positions of the background at a large scale, with a total of 8 options. i indicates the possible scenarios for the target at a small scale, with a total of 9 options. j represents the scenarios that may occur whenever i is determined, with a total of 8 options. The property of target contrast are defined such that positive contrast occurs when the average grayscale value of the target area is greater than that of the background area, while negative contrast is when the average grayscale value of the target area is less than that of the background area (such as Fig. 11(a)).

$$\begin{aligned} \mathrm{LCS}_{\text{big-scale}}^{\text{positive}} &= \min_{I} \frac{\max^2(\mathrm{T})}{\mathrm{mean}(\mathrm{B}_I)}, \quad I = 1,2...8, \\ \mathrm{LCS}_{\text{small-scale}}^{\text{positive}} &= \min_{j} \frac{\max^2(\mathrm{t}_i)}{\mathrm{mean}(\mathrm{b}_i^j)}, \quad i = 1,2...9,\ j = 1,2...8, \end{aligned} \tag{1}$$

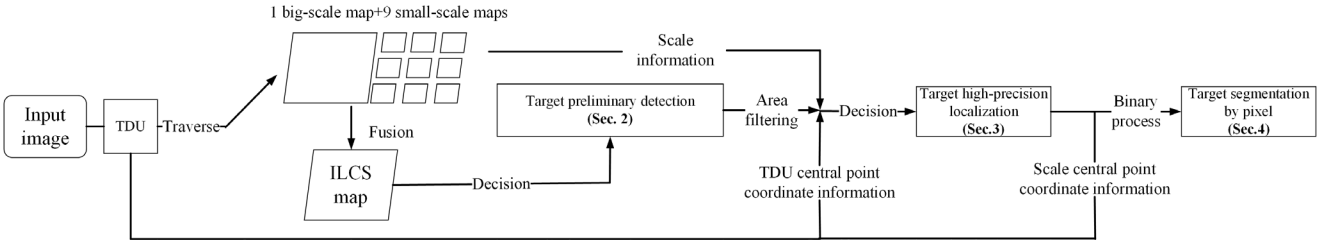


Fig. 1. Technical workflow diagram of the proposed principle.

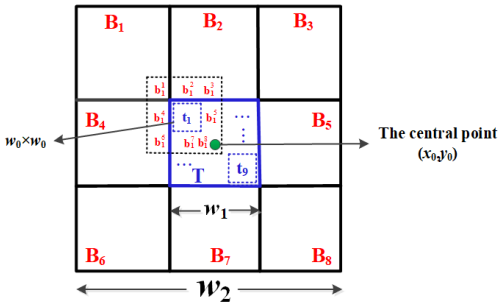


Fig. 2. The structure of TDU within the double scale.

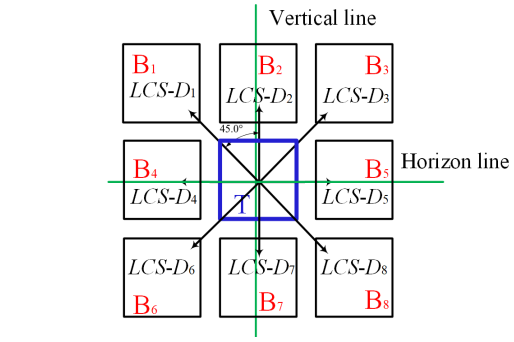
$$\mathrm{LCS}_{\text{big-scale}}^{\text{negative}} = \min_{I} \frac{\text{mean}^2(\mathrm{B}_I)}{\min(\mathrm{T})}, \quad I = 1,2...8,$$
$$\mathrm{LCS}_{\text{small-scale}}^{\text{negative}} = \min_{j} \frac{\text{mean}^2(\mathrm{b}_i^j)}{\min(\mathrm{t}_i)}, \; i = 1,2...9, \; j = 1,2...8 \tag{2}$$

where max(T) denotes the maximum grayscale value of the big scale target area, mean(B_I) represents the average grayscale value of the big scale background area, and min(T) indicates the minimum grayscale value of the big scale target area. As for the small scale, the variables and computing measurement for the small scale are designated with the same rationale.

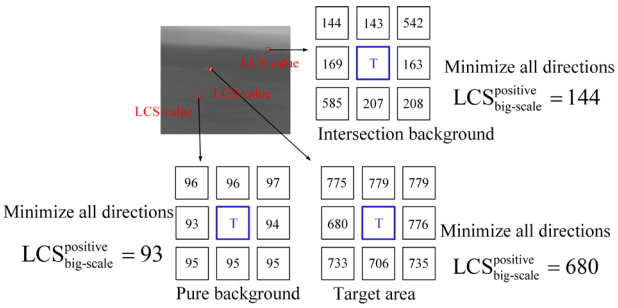
Figure 3 demonstrates the saliency detection capability for the positive contrast target and false target elimination ability of TDU with multiple directions under the guidance of (1). For instance, the LCS values in each direction of true target are high, thereby resulting in a high value of $\mathrm{LCS}_{\text{big-scale}}^{\text{positive}}$. On the other hand, the values in all directions for the pure background are low, hence weakening its influence. Simultaneously, the minimum value operation of (1) eliminates the easily confused target from the intersection background.

Equation (3) describes the generation of ILCS map, which is integrated by different scales within the same TDU. It can stimulate the visual attention mechanism of "top to bottom" [34] to find the preliminary target detection region. By incorporating multi-scale LCS information and considering the differences arising from various size and shape, the ILCS map formed through the maximum value operation of the TDU kernel ensures that target regions are detected without overlooking any genuine targets.

$$\mathrm{ILCS} = \max(\mathrm{LCS}_{\text{big-scale}}, \mathrm{LCS}_{\text{small-scale}}) \tag{3}$$



(a) big scale directions



(b) the multi-directional sensing capability of TDU

Fig. 3. The TDU perception ability of a multi-directional detector for different types of objects in an infrared image (taking each Directional-LCS value at the large scale as an example).

2.2 Scale Decomposed and Fusion

The schematic diagram of the traversal method is shown in Fig. 4. A TDU with dual-scale can produce nine small-scale LCS values and one big-scale LCS value, then the ILCS result combines the maximum values from both scales. The coordinate of the central point (x_0, y_0) traversed by TDU is crucial. Once determined, they can be used to deduce and obtain the coordinates of the central points at big scale and small scale. Figure 4's middle part also demonstrates the traversal path of the overlapping TDU detection units in the original infrared image, once the central point is moved, LCS maps at different scales will be formed, then leading to the integrated ILCS map. As shown in this diagram, moving TDU horizontally and vertically by a length of $\lfloor w_1/2 \rfloor$ within the image creates overlapping target search spaces, thereby ensuring a high detection rate for suspected targets. In fact, from an algorithmic perspective, $\lfloor w_1/2 \rfloor$, $\lfloor w_1/3 \rfloor$, and $\lfloor w_1/4 \rfloor$ can all be used to

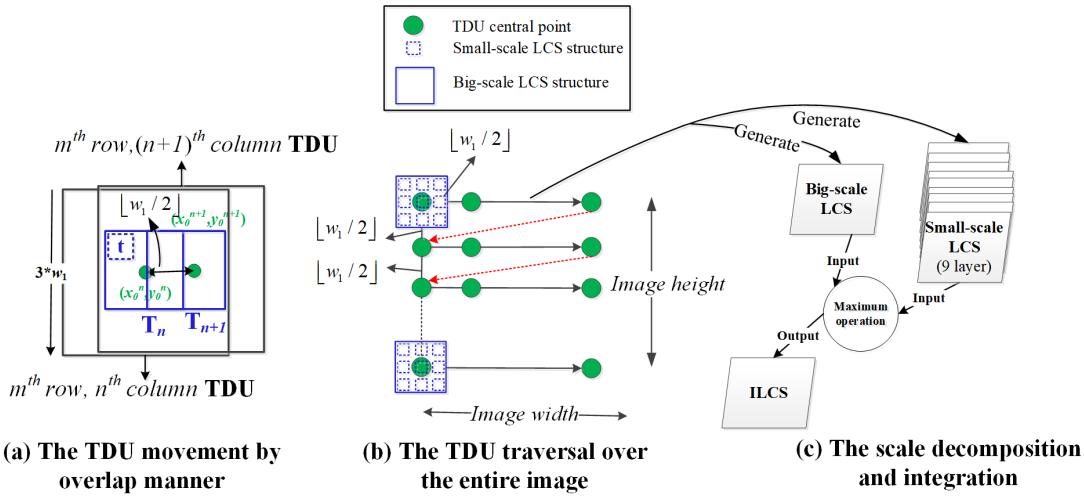


Fig. 4. TDU traversal produces LCS maps with different scale and ILCS map generation by "top to bottom" visual attention model.

partition the candidate box space. However, a smaller step length results in a denser partitioning of the space, leading to higher algorithm time complexity and storage requirements. Extensive testing experience indicates that $w_1/2$ is the optimal choice, as it does not result in the loss of target detection and high-precision localization.

2.3 Dual-Threshold Strategy for Target Preliminary Detection

The establishment of target threshold involves both global threshold and local threshold. Moreover, the selection and implementation of these thresholds stimulate the way the human eye focuses on the prominent target at both the overall and local level. When a certain region meets either the local threshold or global threshold, it can be recognized as a target. For both global threshold and local threshold, excessively high threshold may lead to missing the true target, on the contrary, excessively low threshold may lead to a large amount of background being falsely detected as targets. The global threshold method selects a threshold th_G based on the entire ILCS map, while the local threshold method determines a threshold th_L for each TDU central point in the green color region, as shown in Fig. 4(b), by establishing an 18×18 local area. The definition of global threshold and local threshold are shown in (4) and (5).

$$th_G = \mu_1 + k_1 \times \sigma_1, \tag{4}$$

$$th_L = \mu_2 + k_2 \times \sigma_2 \tag{5}$$

where th_G is a global threshold, and th_L is a local threshold. They are respectively derived from global ILCS statistical information and local ILCS statistical information. μ_1 represents the mean value of entire ILCS map, while μ_2 represents the mean value of local ILCS region. Similarly, σ_1 represents the standard deviation of entire ILCS map, and σ_2 represents the standard deviation of local ILCS region. k_1 and k_2 are proportional factors used to adjust the times of standard. In the decision-making process discovered in the preliminary target area, the default range for k_1 is from 2.5 to 4, and the default range for k_2 is from 3 to 9. When the special scene occurs, th_G is tuned first, followed by tuning th_L. On the side of numerical range, k_1 is suitable for fine-tuning within a small range, while k_2 is suitable for adjustments over a larger range. In general, for scenarios with a high target signal-to-noise ratio (SNR), both k_1 and k_2 can be set to relatively large values, while for scenarios with a low target SNR, k_1 and k_2 need to be set to smaller values.

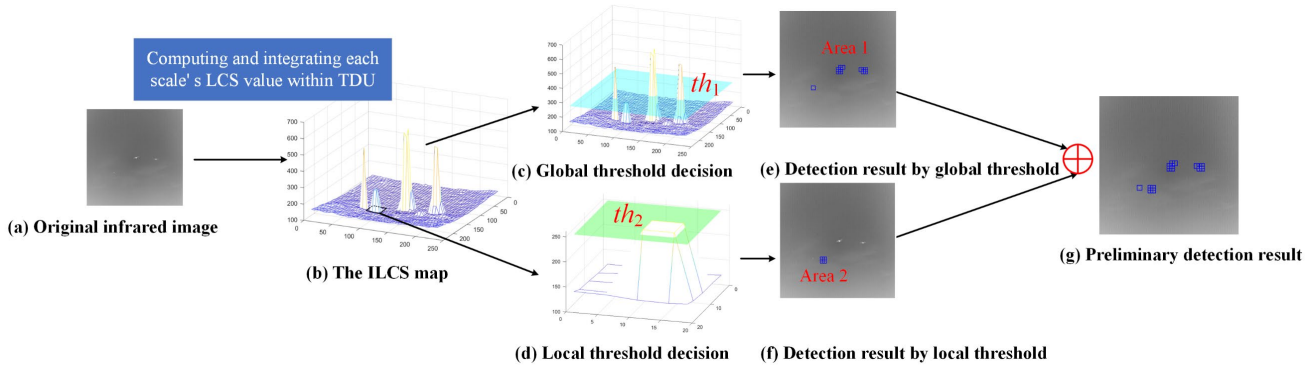


Fig. 5. The diagram of local and global threshold selection strategy.

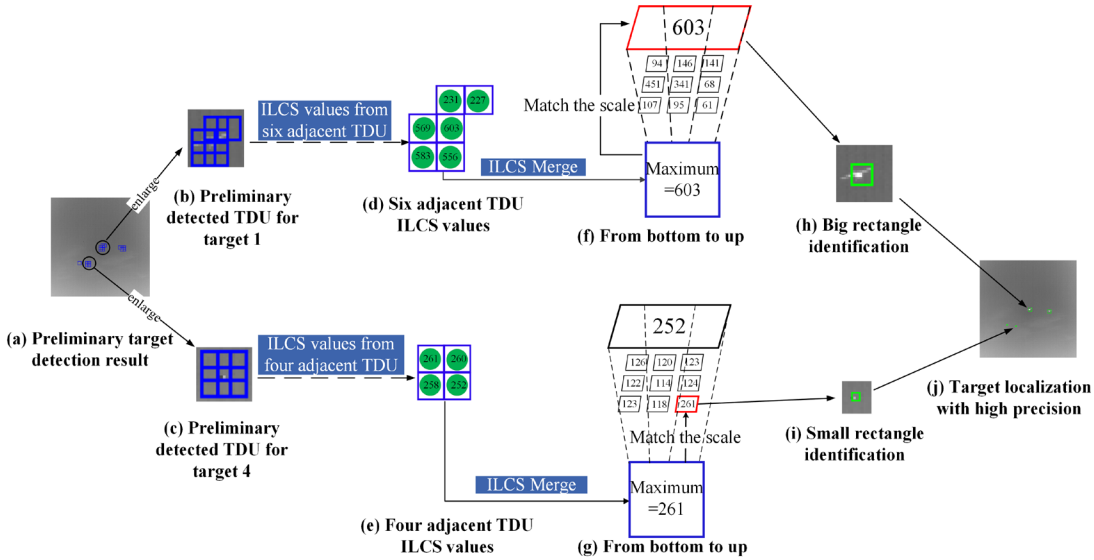


Fig. 6. The progress of target localization with high precision (from preliminary detection result to high-precision position result).

3. Target Localization with High Precision

3.1 Scale Reconstruction for Precise Localization

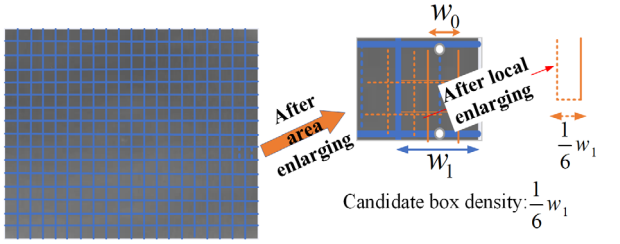
This sub-section discusses how to get the high-precision localization results from the preliminary detection area. It can be divided into two steps: the merging operation of the overlapping adjacent TDU areas based on the maximum criteria of ILCS and the scale-match operation within TDU searched from the ILCS value, which can simulate the "bottom-up" attention mechanism of the human eye attention [35]. The determination of adjacent TDU fusion regions is based on the L1 norm of the central points of adjacent TDUs as a threshold criterion. The mathematical description is shown in (6), typically sets the th_{dist} value to 2. $x^P_{\text{TDU}_i}$ and $y^P_{\text{TDU}_i}$ represent the horizontal and vertical coordinates of TDU central point respectively:

$$\left|x^P_{\text{TDU}_i} - x^P_{\text{TDU}_{i+1}}\right| + \left|y^P_{\text{TDU}i} - y^P_{\text{TDU}i+1}\right| \le th_{\text{dist}}. \qquad (6)$$

Figure 6 illustrates the "bottom-up" process of high-precision target localization mimicking human visual attention. Figure 6(d) and (e) depict the fusion process of ILCS values near TDU based on the maximum value principle, while Figure 6(f) and (g) show the scale matching process. The target lain above Fig. 6(a) is factually large, and is ultimately identified with a large-scale candidate box by this proposed algorithm. In the different case, the target lain below Fig. 6(a) is factually small, and is ultimately identified with a small-scale label by this proposed algorithm.

3.2 The Strategy of AORM

The layout of fixed candidate boxes is illustrated in Fig. 7(a), which are formed by uniform division of the image area from the width and height directions. This layout of TDU is highly advantageous for the central point location and describing target feature. Figure 7(b) shows the decomposition-scale candidate box generated by the fractal geometric scale and overlapping movement of TDU in the horizontal direction. The yellow solid line represents the result of scale decomposition applied to the blue solid line box, while the yellow dashed line demonstrates small-scale decomposition effects on the blue dashed line box after horizontal stepping. Taking as a size of 640 × 512 image as an example, the size of the fixed box is 21 × 21 pixels. After undergoing small-scale decomposition, this box can be derived into nine fractal small-scale boxes, each with size of 7 × 7 pixels. Through these analysis and calculation, total of 21,730 candidate boxes are generated, including 2,173 affiliated TDU kernels, 2,173 fixed candidate boxes (the central region of TDU) and 19557 (2,173 × 9) small-scale moving candidate boxes. After scale decomposition, adjacent candidate box generates a grid width of 4 pixels (w_1/6) from both horizontal side and vertical side. If the TDU is traversed with a smaller length, the overall layout of candidate boxes will be denser. This dense candidate box layout is crucial for the high-precision localization of infrared small targets of different sizes and arbitrary positions because it can provide a comprehensive and adequate searching space. In conclusion, the AORM (Area Optimal Recommend Mechanism) strategy



(a) Layout of fixed candidate box (b) Layout of scale-decomposition box

Fig. 7. The structure and layout of the candidate box.

is in the purpose of reducing the traversal time of each candidate box, the final traversal time can be effectively shortened from 2173 (fixed boxes number) × 10 (total number of scales) to "2173 + m × 10", m is total remaining number of TDU after merged. Typically, m variable varies between [2, 10], so "2173 + m × 10" is much smaller than 21730. This optimization is mainly based on the design of the TDU kernel among the global principle.

4. Target Segmentation by Per-pixel

This section discusses how to binarize the segmentation results based on high-precision localization's result. The key operation to binarization is the determination of local area, the generation of local histograms and threshold segmentation points through local area. The important factor of the local area decision is ($x^{H}_{\text{matched-scale}}$, $y^{H}_{\text{matched-scale}}$) and the suitable scale from above high-precision localization, the size of local area for target segmentation is chosen as $2w_0 \times 2w_0$ and $2w_1 \times 2w_1$ by small scale and big scale, respectively. By dividing the grayscale levels of the affiliated target region into local target (Category 0) and local background (Category 1), using the value th_3 as the optimal threshold, Category 0 comprises pixels with values higher than or equal to th_3, whereas Category 1 encompasses pixels with values lower than th_3.

After obtaining the local target region, maximize the between-class variation named by σ_B^2, the mathematical model is shown in (7):

$$\sigma_B^2 = \pi_0(\mu_0 - \mu_T)^2 + \pi_1(\mu_1 - \mu_T)^2 \qquad (7)$$

where π_0 represents the proportion of the number of pixels to the total number of pixels and μ_0 is the average grayscale value of all pixels in Category 0, while π_1 represents the proportion of the number of pixels to the total number of pixels and μ_1 is the average grayscale value of all pixels in Category 1. μ_T represents the average of the total grayscale value in the entire image.

Given that $\mu_T = \mu_0\pi_0 + \mu_1\pi_1$, the formula for the between-class variance can be simplified to (8). Thus, the best target segmentation results can be achieved by traversing th_3:

$$\sigma_B^2 = \pi_0\pi_1(\mu_1 - \mu_0)^2 . \qquad (8)$$

As observed from Fig. 8, the histogram approaches to a Gaussian distribution with one peak, that is to say the local target region contains one target (Category 0) and one background (Category 1). Figure 8(f) and (g) show the local histograms and optimal thresholds. Figure 8(h) and (i) correspond to the local region binarization results under the optimal threshold. From the binarized results, it can be seen that the threshold segmentation method is suitable to segment the target by pixel-level.

5. Experimental Results and Performance Assessment

To validate the effectiveness of the proposed method, we compared it with seven other state-of-the-art unsupervised infrared small target detection algorithms, namely LIG [26], AAGD [24], ADMD [25], LEF [27], NTFRA [17], GCMDO [23], and ANLPT [28]. The deep learning method is not compared here, because our proposed principle does not require data pre-training. We selected different infrared scenes to illustrate the proposed algorithm's performance in high-precision target localization and pixel-level target segmentation. These scenes include the infrared sky scene with irregular shaped targets [36], infrared maritime scene with irregular shaped targets (ours), infrared maritime target with negative contrast property (ours), point-shaped targets with inference (the upward datasets come from [36], and the downward datasets come from [37]), and infrared small & dim target scenes (the downward one comes from [38], and the upward one comes from ours). Our images are captured by an integrated maritime imaging facility that includes an infrared medium wave camera, a visible light camera and an infrared long wave camera. This equipment acquires a picture dataset of 640 × 512 pixels for a duration of two hours and has a maximum target capture distance of 20 km. Our data for

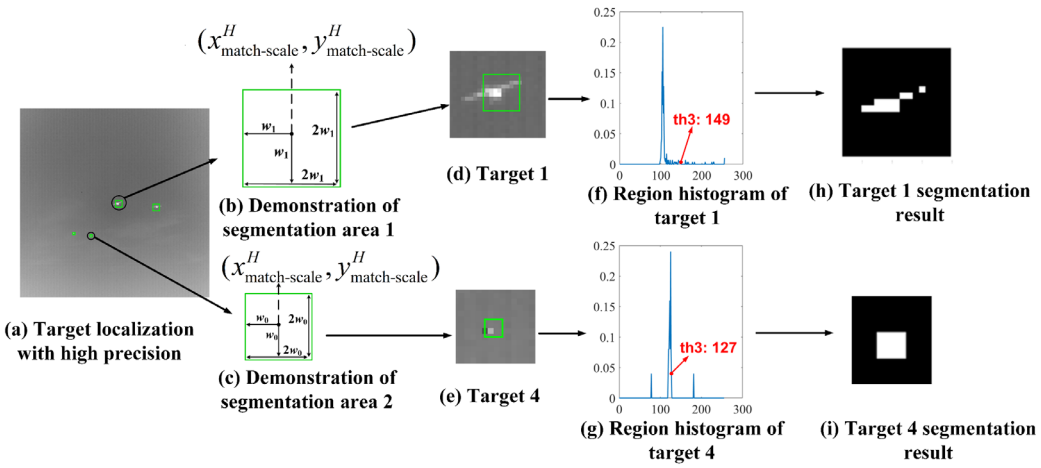


Fig. 8. Demonstration of local target histogram and local target segmentation by pixel-level.

this paper were collected at Marina Bay Plaza in Dalian on March 18, 2022. Subsequently, we chose two representative images from each scene to verify the algorithm's performance. All algorithms' experiments are implemented in MATLAB 2020a on a laptop with a 2.30-GHz Core i7-12700H CPU and 16.0-GB memory.

The blue rectangular region represents the true target region in the Ground truth, the small red rectangle represents the true target region localized accurately by each algorithm, the green rectangular region represents the target region localized falsely by each algorithm, meanwhile, the large green rectangle indicates that the representative region contains a large number of detected targets falsely.

5.1 Infrared Sky Scene with Irregular Shaped Targets

The detection results for infrared sky scenes with irregularly shaped targets are depicted in Fig. 9. This dataset comprises 998 images, characterized by the irregular shapes of the targets and the presence of numerous targets in each image. As illustrated in Fig. 9, most compared algorithms can detect the big scale targets successfully, while the LIG, AAGD, ADMD, and NTFRA algorithms fail to detect small-scale targets. Our proposed algorithm can detect and segment the double-scale targets successfully.

5.2 Infrared Maritime Scene with Irregular Shaped Targets

Figure 10 shows the detection results and segmentation results of infrared maritime scene with irregular shaped target, which is a dataset of 1106 images. The targets are easily submerged in a heavy noise and clutter background, from the results, LIG, ADMD, LEF, GCMDO algorithm detect a great number of false alarm regions as the true targets. Under the comparison, our proposed algorithm has the ability to suppress the noise or clutter.

5.3 Infrared Maritime Targets with Backlight Illumination

The detection results for this group are presented in Fig. 11. This dataset comprises 1000 images. Unlike other scenes, this backlight scenario poses a challenge for target detection due to the negative contrast property, where the central region has lower gray value than its surrounding. It is obvious that the compared detection algorithms are easy to detect the cluttered wave regions, whereas, our proposed algorithm can prevent this problem.

5.4 Point-Shaped Targets with Interference

The target detection and segmentation results for point-shaped targets with interference are illustrated in Fig. 12. This dataset comprises 2000 images, where targets appear as point-shaped with one or two pixel-size, and some clutter from background is easily confused with the true target in this scene. Due to the small size of the target, the red dashed rectangular box represents the magnified result of the target. However, the compared algorithms often misidentify the nature of snow interference as the true target. In this case, our proposed algorithm can suppress the background clutter through the parameter k_1 and k_2 setting with a relative higher value.

5.5 Infrared Small & Dim Targets Scene

The detection results for infrared small and dim target scenes are also displayed in Fig. 13. This dataset comprises 1500 images. The target appears extremely small, from the results, we can see only LIG algorithm and our proposed algorithm obtain the better result, other compared algorithms lose the true target and detects more false regions. It can be concluded that the two algorithms can detect the extremely dim target, which can be easily disappeared within the background.

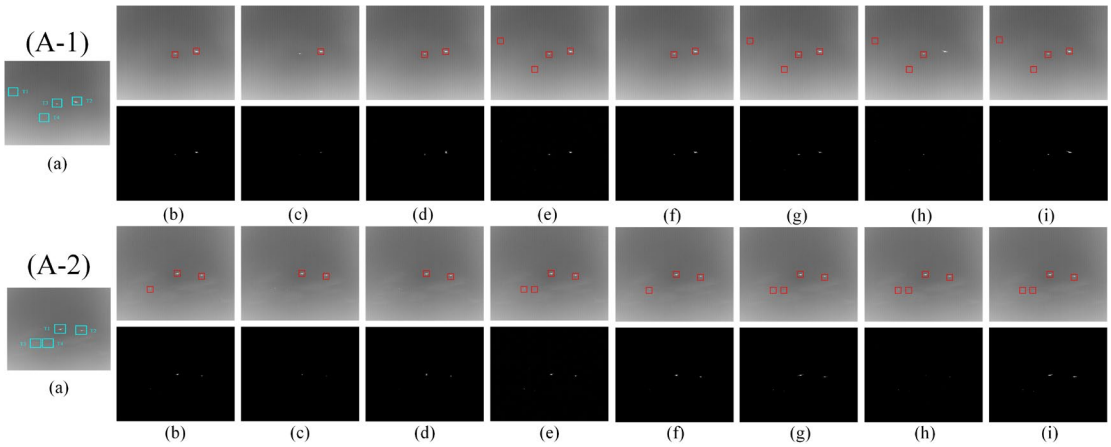


Fig. 9. The detection results and segment results of two representative images of infrared sky scene with irregular shaped target. (a) Original image; (b) LIG; (c) AAGD; (d) ADMD; (e) LEF; (f) NTFRA; (g) ANLTP; (h) GCMDO; (i) ours ($k_1 = 3$, $k_2 = 5$).

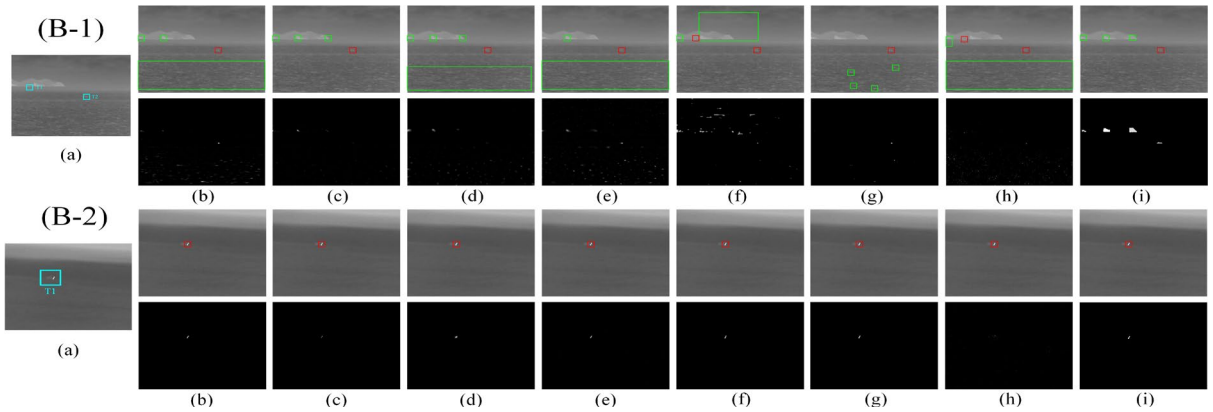


Fig. 10. The detection results and segment results of two representative images of infrared maritime scene with irregular shaped target. (a) Original image; (b) LIG; (c) AAGD; (d) ADMD; (e) LEF; (f) NTFRA; (g) ANLTP; (h) GCMDO; (i) ours ($k_1 = 3$, $k_2 = 5$).

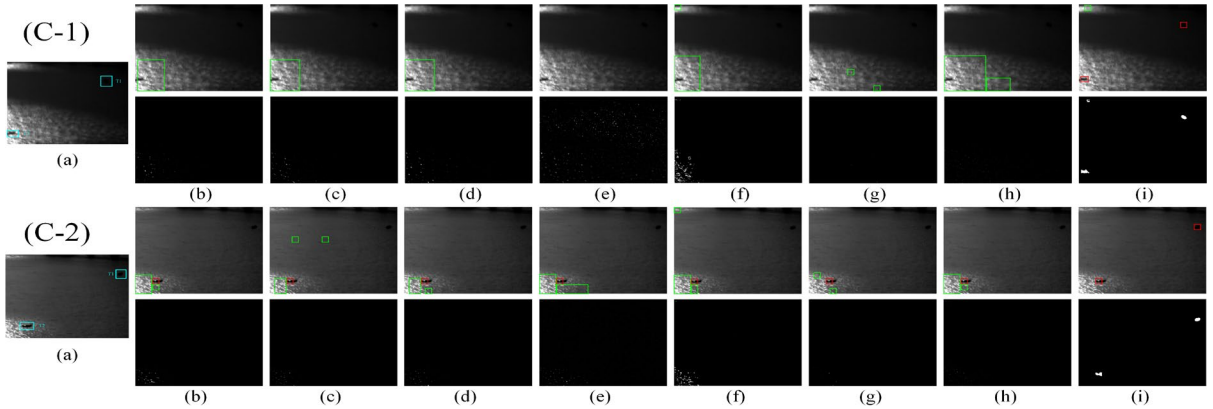


Fig. 11. The detection results and segment results of two representative images of infrared maritime target with backlight illumination. (a) Original image; (b) LIG; (c) AAGD; (d) ADMD; (e) LEF; (f) NTFRA; (g) ANLTP; (h) GCMDO; (i) ours ($k_1 = 3$, $k_2 = 5$).

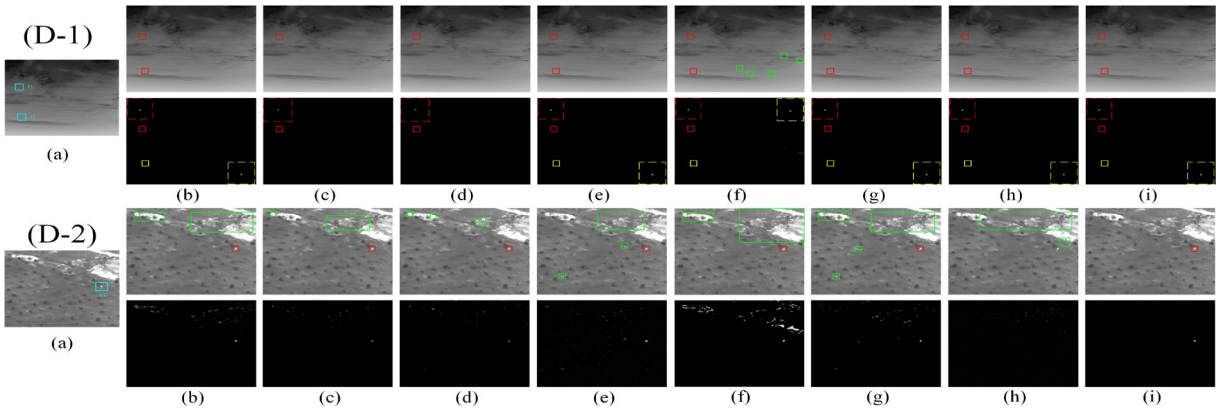


Fig. 12. The detection results and segment results of two representative images of point-shaped target with interference. (a) Original image; (b) LIG; (c) AAGD; (d) ADMD; (e) LEF; (f) NTFRA; (g) ANLTP; (h) GCMDO; (i) ours ($k_1 = 5$, $k_2 = 7$).

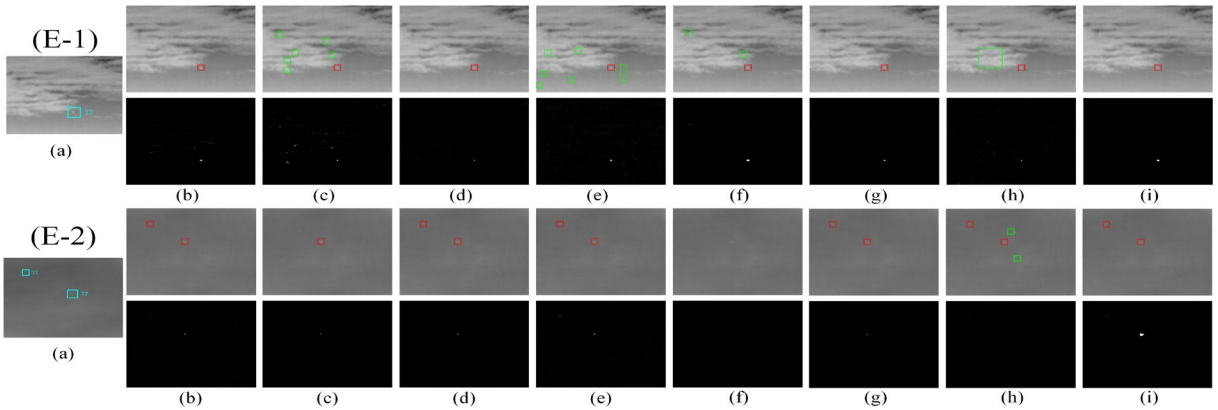


Fig. 13. The detection results and segment results of two representative images of infrared small & dim target. (a) Original image; (b) LIG; (c) AAGD; (d) ADMD; (e) LEF; (f) NTFRA; (g) ANLTP; (h) GCMDO; (i) ours ($k_1 = 3$, $k_2 = 5$).

5.6 Assessment for Infrared Target Detection and Segmentation

The detection algorithm's performance can be assessed using Detection Rate (DR) and False Alarm Rate (FAR) standard. DR reflects the algorithm's proficiency in accurately identifying targets, the larger DR means the algorithm has better performance, while lower FAR indicates that algorithm has better ability to decrease false alarm rates. The computing mathematical of DR and FAR are shown by (9).

$$\mathrm{DR}=\frac{N_{\mathrm{A}}}{N_{\mathrm{T}}},\quad \mathrm{FAR}=\frac{N_{\mathrm{F}}}{N_{\mathrm{T}}+N_{\mathrm{F}}} \tag{9}$$

where N_{A} represents the number of detected actual targets, N_{T} denotes the total number of true targets in the original images, and N_{F} is the number of detected false targets. To provide a comprehensive evaluation of the detection algorithm, the F_β score is also introduced, which combines DR and FAR. Precision Rate (PR) and Recall Rate (RR) are fundamental to calculating the F_β score. PR and RR are related to FAR and DR as shown in (10). Based on PR and RR, F_β score is calculated as shown in (11).

$$\begin{aligned}\mathrm{PR}&=\frac{N_{\mathrm{A}}}{N_{\mathrm{A}}+N_{\mathrm{F}}}=1-\mathrm{FAR},\\ \mathrm{RR}&=\frac{N_{\mathrm{A}}}{N_{\mathrm{T}}}=\mathrm{DR},\end{aligned} \tag{10}$$

$$F_\beta=\frac{(1+\beta^2)\times\mathrm{PR}\times\mathrm{RR}}{(\beta^2\times\mathrm{PR})+\mathrm{RR}}. \tag{11}$$

To analyze the performance of the algorithms under specific thresholds more precisely, first the detection results are obtained for all the compared algorithms and then noise suppression is performed by setting thresholds to remove low response regions. Specifically, the adaptive threshold T_{adp} is calculated according to (12).

$$T_{\mathrm{adp}}=\max[\max(R)\times 0.7,\sigma(R)+\mathrm{avg}(R)] \tag{12}$$

where max(R) represents the largest value of the detection result. $\sigma(R)$ and avg(R) represent the standard deviation and average value of the detection result, respectively. DR, FAR, and F_1 score are used as comprehensive metrics to evaluate the performance of the detection algorithm, providing a balanced measure of accuracy and robustness, as shown in Tab. 1. Notably, our algorithm consistently demonstrates exceptional performance across all datasets, achieving the highest F_β scores with corresponding high DR and negligible FAR. This underscores the robustness and effectiveness of our approach compared to the other algorithms evaluated. Specifically, our algorithm outperforms others by achieving F_β scores of 0.95, 0.92, 0.88, 0.95 and 0.95 across Dataset 1 to Dataset 5, respectively, with high DR ranging from 90.12% to 96.30% and low FAR. These results highlight the superior performance of our algorithm and its potential for real-world applications.

In the target segmentation assessment, Intersection over Union (IoU) serves as a crucial evaluation standard for measuring the overlap area between the target segmentation result and the ground truth. The maximum value of IoU is 1, the larger value indicates the larger overlap area between two sides. The computing model is shown as (13):

$$\mathrm{IoU}=\frac{I_{\mathrm{res}}\bigcap I_{\mathrm{GT}}}{I_{\mathrm{res}}\bigcup I_{\mathrm{GT}}} \tag{13}$$

where I_{res} represents the target segmentation result's coordinate set and I_{GT} represents the Ground Truth's.

The ROC curve shown in Fig. 14 reflects the changes in DR and FAR across different image data as the intrinsic parameters of the algorithm are adjusted. The larger the area formed with the horizontal axis, the better the performance of the algorithm. The numbers in the bottom right corner represent the average performance of each algorithm, showing that the proposed algorithm has a clear advantage.

Table 2 presents the IoU of five datasets under the comparison algorithms. It can be observed that our algorithm achieves the best IoU in all five datasets.

Table 3 presents the mean computational time of the five datasets tested with all algorithms. It can also be concluded that the proposed algorithm is relatively better than most other algorithms in real-time performance.

6. Conclusion

Through the above principles and experimental validation, it can be seen that this paper divides the target detection task into three stages. The core algorithm adopts a TDU kernel-based design, so that the TDU kernel achieves high-precision target localization and segmentation through parameter transmission and the transmission of subtask results between the kernel and the subtasks.

The advantage of the TDU kernel-based structure lies in the preliminary detection of the target area, and this principle does not rush to determine the true area, allowing overlapping detection results to ensure that no real targets are missed. In the task of high-precision target localization, it provides accurate localization of the true area. Meanwhile, the TDU dual-scale division facilitates the fine division of candidate boxes and the implementation of the AORM strategy. The final segmentation is centered on the high-precision localization coordinates from the second task. By selecting an appropriate threshold, binarization segmentation is performed, which can enhance the contour information and local representation of the small target.

Methods	Dataset 1			Dataset 2			Dataset 3			Dataset 4			Dataset 5		
	DR (%)	FAR (%)	F_β	DR (%)	FAR (%)	F_β	DR (%)	FAR (%)	F_β	DR (%)	FAR (%)	F_β	DR (%)	FAR (%)	F_β
LIG	68.23	0.55	0.81	84.26	10.15	0.87	47.86	28.66	0.57	91.56	20.12	0.85	95.26	5.12	**0.95**
AAGD	71.89	10.22	0.80	75.26	10.12	0.82	46.89	28.57	0.57	81.26	22.34	0.79	86.59	6.11	0.90
ADMD	70.82	14.00	0.78	76.76	11.23	0.82	39.66	33.57	0.50	83.66	5.26	0.89	87.22	4.21	0.91
LEF	80.04	0.79	0.89	86.66	19.22	0.84	49.26	39.64	0.54	90.56	4.96	0.93	94.88	7.36	0.94
NTFRA	76.00	0.19	0.86	91.12	31.06	0.78	79.12	42.12	0.67	87.88	28.92	0.79	97.10	12.55	0.92
ANLTP	86.04	0.88	0.92	81.68	10.66	0.85	45.16	11.27	0.60	84.78	20.83	0.82	86.27	4.77	0.91
GCMDO	85.08	0.69	0.92	79.91	19.12	0.80	41.23	20.12	0.54	60.22	19.68	0.69	87.12	5.60	0.91
Ours	**91.08**	**0.14**	**0.95**	**92.88**	**8.26**	**0.92**	**90.12**	**13.33**	**0.88**	**92.16**	**1.29**	**0.95**	**96.30**	**5.44**	**0.95**

Tab. 1. DR, FAR and F_β score for the five datasets tested by the compared algorithm.

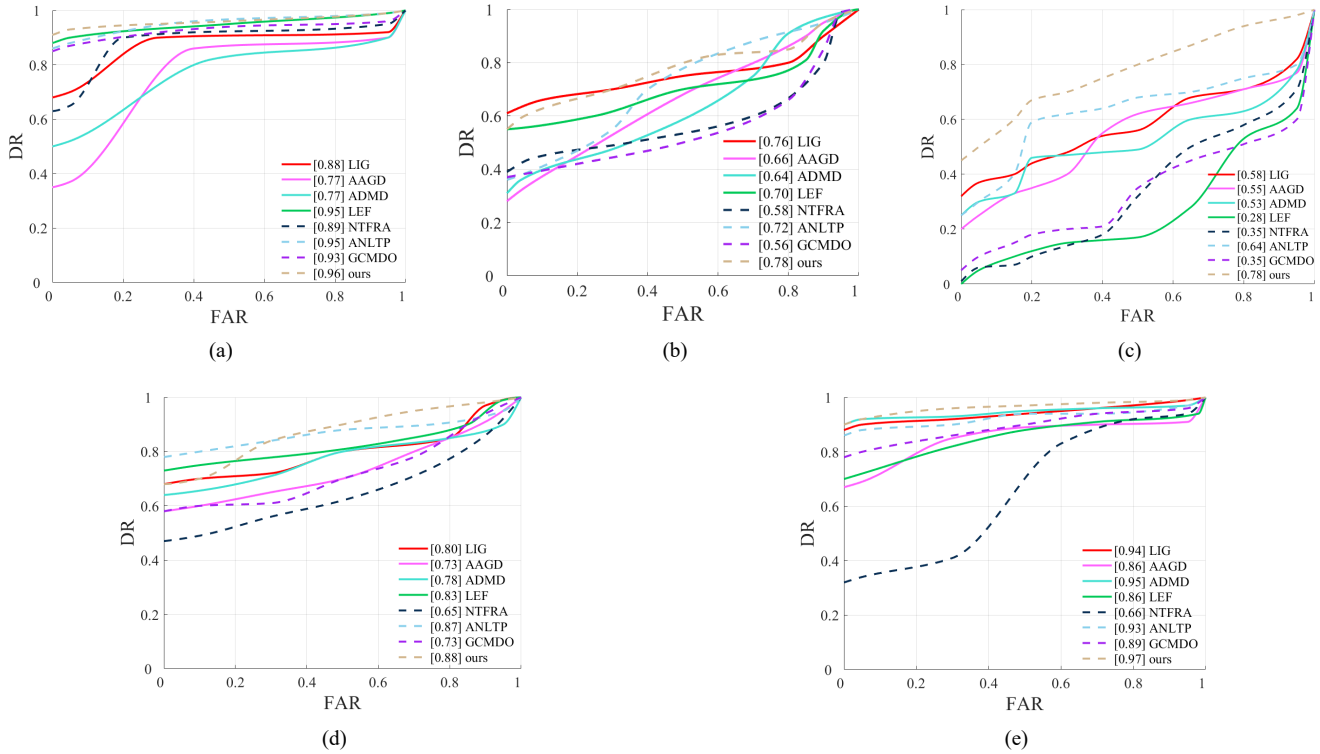


Fig. 14. ROC curves and AUCs of different algorithm's result in five scenes. (a) Dataset 1; (b) Dataset 2; (c) Dataset 3; (d) Dataset 4; (e) Dataset 5.

Methods	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
LIG	26.35	32.94	0.18	55.14	39.32
AAGD	7.25	12.30	0.26	34.45	17.74
ADMD	15.86	14.31	0.12	17.00	16.17
LEF	29.84	30.83	0.07	64.25	27.49
NTFRA	31.42	32.79	0.49	17.50	28.60
ANLTP	36.49	37.61	0.23	45.79	29.05
GCMDO	5.88	5.37	0.10	33.21	9.17
Ours	**49.95**	**52.78**	**47.63**	**91.16**	**66.45**

Tab. 2. IoU ($\times 10^{-2}$) for the five datasets tested by the compared algorithms.

Algorithm	LIG	AAGD	ADMD	LEF	NTFRA	ANLTP	GCMDO	Ours
Time(s)	1.0923	0.0886	0.086	2.1768	1.045	1.034	**0.0043**	0.165

Tab. 3. The mean computational time of the five datasets tested by the compared algorithms.

Acknowledgments

This work was supported by Guangxi Provincial Natural Science Foundation of China under Grant 2020GXNSFBA297077. We would like to thank the editor and reviewers for their efforts in the publication of this paper, and express our gratitude to Professor Haoyong Yu from the Advanced Robotics Center, National University of Singapore for providing the research environment.

References

[1] LI, Y., ZHANG, Y. Robust infrared small target detection using local steering kernel reconstruction. *Pattern Recognition*, 2018, vol. 77, p. 113–125. DOI: 10.1016/j.patcog.2017.12.012

[2] ZHAO, M., LI, W., LI, L., et al. Infrared small UAV target detection via isolation forest. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, vol. 61, p. 1–16. DOI: 10.1109/TGRS.2023.3321723

[3] CHEN, Y., WANG, H., PANG, Y., et al. An infrared small target detection method based on a weighted human visual comparison mechanism for safety monitoring. *Remote Sensing*, 2023, vol. 15, no. 11, p. 1–23. DOI: 10.3390/rs15112922

[4] BAI, Y., LI, R., GOU, S., et al. Cross-connected bidirectional pyramid network for infrared small-dim target detection. *IEEE Geoscience and Remote Sensing Letters*, 2022, vol. 19, p. 1–5. DOI: 10.1109/LGRS.2022.3145577

[5] LI, Z., LIAO, S., WU, M., et al. Small and dim infrared moving target detection based on spatial-temporal saliency. *Optik*, 2022, vol. 270, p. 1–12. DOI: 10.1016/j.ijleo.2022.170010

[6] LI, Z., LIAO, S., WU, M., et al. SFD-SR: Strengthened feature description and stabilized re-detection-based infrared small-dim target tracking algorithm. *Soft Computing*, 2024, vol. 28, no. 2, p. 963–979. DOI: 10.1007/s00500-023-09307-1

[7] SUN, J., WANG, J., HAO, Z., et al. AC-LSTM: Anomaly state perception of infrared point targets based on CNN+LSTM. *Remote Sensing*, 2022, vol. 14, no. 13, p. 1–19. DOI: 10.3390/rs14133221

[8] LI, B., XIAO, C., WANG, L., et al. Dense nested attention network for infrared small target detection. *IEEE Transactions on Image Processing*, 2022, vol. 32, p. 1745–1758. DOI: 10.1109/TIP.2022.3199107

[9] JIANG, K., WANG, Z., YI, P., et al. Deep distillation recursive network for remote sensing imagery super-resolution. *Remote Sensing*, 2018, vol. 10, p. 1–23. DOI: 10.3390/rs10111700

[10] LI, S., ZHANG, K., DUAN, P., et al. Hyperspectral anomaly detection with kernel isolation forest. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, vol. 58, no. 1, p. 319–329. DOI: 10.1109/TGRS.2019.2936308

[11] CUI, H., LI, L., LIU, X., et al. Infrared small target detection based on weighted three-layer window local contrast. *IEEE Geoscience and Remote Sensing Letters*, 2022, vol. 19, p. 1–5. DOI: 10.1109/LGRS.2021.3133649

[12] QIU, Z., MA, Y., FAN, F., et al. A pixel-level local contrast measure for infrared small target detection. *Defence Technology*, 2022, vol. 18, no. 9, p. 1589–1601. DOI: 10.1016/j.dt.2021.07.002

[13] LIU, Y., RAO, X., HU, J., et al. A weak target detection algorithm IAR-STFT based on correlated k-distribution sea clutter model. *Radioengineering*, 2023, vol. 32, no. 1, p. 33–43. DOI: 10.13164/re.2023.0033

[14] XU, Y., WAN, M., ZHANG, X., et al. Infrared small target detection based on local contrast-weighted multidirectional derivative. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, vol. 61, p. 1–16. DOI: 10.1109/TGRS.2023.3244784

[15] YANG, P., DONG, L., XU, W. Infrared small maritime target detection based on integrated target saliency measure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, vol. 14, p. 2369–2386. DOI: 10.1109/JSTARS.2021.3049847

[16] HU, X., WANG, X., YANG, X., et al. An infrared target intrusion detection method based on feature fusion and enhancement. *Defence Technology*, 2020, vol. 16, p. 737–746. DOI: 10.1016/j.dt.2019.10.005

[17] KONG, X., YANG, C., CAO, S., et al. Infrared small target detection via nonconvex tensor fibered rank approximation. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, vol. 60, p. 1–21. DOI: 10.1109/TGRS.2021.3068465

[18] CHAN, Y. T. Maritime filtering for images and videos. *Signal Processing: Image Communication*, 2021, vol. 99, p. 1–13. DOI: 10.1016/j.image.2021.116477

[19] FAN, X., QIN, W., LI, J., et al. Dim and small target detection based on spatio-temporal filtering and high-order energy estimation. *IEEE Photonics Journal*, 2023, vol. 15, no. 2, p. 1–20. DOI: 10.1109/JPHOT.2023.3242991

[20] CHAN, Y. T. Comprehensive comparative evaluation of background subtraction algorithms in open sea environments. *Computer Vision and Image Understanding*, 2021, vol. 202, p. 1 to 16. DOI: 10.1016/j.cviu.2020.103101

[21] LUO, Y., LI, X., CHEN, S. Feedback spatial-temporal infrared small target detection based on orthogonal subspace projection. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, vol. 62, p. 1–19. DOI: 10.1109/TGRS.2024.3368099

[22] CHEN, C., LI, H., WEI, Y., et al. A local contrast method for small infrared target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, vol. 52, no. 1, p. 574–581. DOI: 10.1109/TGRS.2013.2242477

[23] TANG, Y., XIONG, K., WANG, C. Fast infrared small target detection based on global contrast measure using dilate operation. *IEEE Geoscience and Remote Sensing Letters*, 2023, vol. 20, p. 1–5. DOI: 10.1109/LGRS.2023.3233958

[24] AGHAZIYARATI, S., MORADI, S., TALEBI, H. Small infrared target detection using absolute average difference weighted by cumulative directional derivatives. *Infrared Physics & Technology*, 2019, vol. 101, p. 78–87. DOI: 10.1016/j.infrared.2019.06.003

[25] MORADI, S., MOALLEM, P., SABAHI, M. F. Fast and robust small infrared target detection using absolute directional mean difference algorithm. *Signal Processing*, 2020, vol. 177, p. 1–9. DOI: 10.1016/j.sigpro.2020.107727

[26] ZHANG, H., ZHANG, L., YUAN, D., et al. Infrared small target detection based on local intensity and gradient properties. *Infrared Physics & Technology*, 2018, vol. 89, p. 88–96. DOI: 10.1016/j.infrared.2017.12.018

[27] XIA, C., LI, X., ZHAO, L., et al. Infrared small target detection based on multiscale local contrast measure using local energy factor. *IEEE Geoscience and Remote Sensing Letters*, 2020, vol. 17, no. 1, p. 157–161. DOI: 10.1109/LGRS.2019.2914432

[28] ZHANG, Z., DING, C., GAO, Z., et al. ANLPT: Self-adaptive and non-local patch-tensor model for infrared small target detection. *Remote Sensing*, 2023, vol. 15, no. 4, p. 1–24. DOI: 10.3390/rs15041021

[29] TUNC, S., ILGIN, H. A. Dim target detection in infrared images using saliency algorithms. *Radioengineering*, 2019, vol. 28, no. 3, p. 635–642. DOI: 10.13164/re.2019.0635

[30] HE, Z., ZHANG, L., GAO, X., et al. Multi-adversarial faster-RCNN with paradigm teacher for unrestricted object detection. *International Journal of Computer Vision*, 2023, vol. 131, no. 3, p. 680–700. DOI: 10.1007/s11263-022-01728-z

[31] WANG, C.-Y., BOCHKOVSKIY, A., LIAO, H.-Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* Vancouver (BC, Canada), 2023, p. 7464–7475. DOI: 10.1109/CVPR52729.2023.00721

[32] ZHANG, W., LI, Z., LI, G., et al. GACNet: Generate adversarial-driven cross-aware network for hyperspectral wheat variety identification. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, vol. 62, p. 1–14. DOI: 10.1109/TGRS.2023.3347745

[33] ZHANG, W., ZHAO, W. LI, J., et al. CVANet: Cascaded visual attention network for single image super-resolution. *Neural Networks*, 2024, vol. 170, p. 622–634. DOI: 10.1016/j.neunet.2023.11.049

[34] YANG, J., YANG, M. Top-down visual saliency via joint CRF and dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Learning*, 2016, vol. 39, no. 3, p. 576–588. DOI: 10.1109/TPAMI.2016.2547384

[35] SIRIS, A., JIAO, J., TAM, G. K. L., et al. Inferring attention shifts for salient instance ranking. *International Journal of Computer Vision*, 2024, vol. 132, p. 964–986. DOI: 10.1007/s11263-023-01906-7

[36] SUN, X., GUO, L., ZHANG, W., et al. Dataset for small infrared moving target detection under complex background. V3. *Science Data Bank*, 2024. [Online] Cited 2024-09-20. Available at: https://doi.org/10.11922/sciencedb.j00001.00231 DOI: 10.11922/sciencedb.j00001.00231

[37] HUI, B., SONG, Z., FAN, H., et al. A dataset for infrared image dim-small aircraft target detection and tracking underground / air background V1. *Science Data Bank*, 2019. [Online] Cited 2024-09-20. Available at: https://doi.org/10.11922/sciencedb.902. DOI: 10.11922/sciencedb.902

[38] SUN, Y., YANG, J., LI, M., et al. Infrared small target detection via spatial–temporal infrared patch-tensor model and weighted Schatten p-norm minimization. *Infrared Physics & Technology*, 2019, vol. 102, p. 1–10. DOI: 10.1016/j.infrared.2019.103050

About the Authors ...

Chang DING was born in 1988. He received the Ph.D. degree in Information and Communication Engineering from Dalian Maritime University, Dalian, China in 2018. He is currently post-doctoral from the School of Information and Control Engineering, China University of Mining and Technology, and a lecturer in the School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology. From February 2023 to February 2024, he worked as a visiting scholar in Advanced Robot Center, Department of Biomedical Engineering, National University of Singapore. He has published more than twenty papers in the international and national journals, his research interest includes infrared image processing, industrial detection and machine learning.

Hai LIU (corresponding author) was born in 1983. He received the B.S. and M.S. degrees in Optical Engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2003 and 2006, respectively, and the Ph.D. degree in Physical Electronics from the Wuhan National Laboratory for Optoelectronics, Wuhan, in 2009. He is a Professor with the China University of Mining and Technology, Xuzhou, China. His current research interests include integrated optoelectronic, computational electromagnetic, and photonics.