RLPGB-Net: Reinforcement Learning of Feature Fusion and Global Context Boundary Attention for Infrared Dim Small Target Detection

Zhe Wang, Tao Zang, Zhiling Fu, Hai Yang, and Wenli Du

***Abstract*— In infrared scenes, humans can easily observe objects in the scene with their eyes, even dim ones. To make the robot have the same visual ability, this article proposes a pyramid-feature fusion target detection network, called RLPGB-Net, which combines reinforcement learning with aerial targets in the infrared scene. It makes use of the powerful decision-making ability of reinforcement learning to give corresponding weights to the extracted features and highlight the significant features of infrared dim small targets. In reinforcement learning, we use priori strategy guidance and long-term training methods to train weight-regulating agents. To eliminate the local influence on the detection results, such as bright interference points similar to the target, and to solve the problem of dim target detection effectively, the global context boundary attention (GB) module is introduced to eliminate the disadvantage of local comparison using the global characteristics of different dimensions. At the same time, it can prevent the edge information of the refined target from being submerged in the background. Experimental results on the SAITD and SIRST datasets show the effectiveness of the proposed method.**

***Index Terms*— Global context boundary attention (GB), infrared dim target, pyramid feature fusion, reinforcement learning.**

I. Introduction

INFRARED target detection technology is widely used in various fields, such as aerospace technology [1], air defense early warning [2], military [3], medical imaging [4], and target detection and tracking [5]. Because infrared imaging uses thermal radiation without lighting, it can overcome adverse factors affecting common imaging, such as smoke, fog, rain, haze, and other atmospheric conditions [5]. The ability to provide clear images even in bad weather is a key technology for applications in areas such as early warning systems, precision-guided weapons, and airspace surveillance systems. However, this remains a challenging problem: clutter noise in complex backgrounds results in a low signal–clutter ratio (SCR). Specifically, small infrared targets (because of the distance of the shooting) may be affected by thick clouds or object-like interferons [6], [7], [8], [9] and easily overwhelmed by the background. The small and dim target usually only occupies several pixels or even one pixel in the image, carries less feature information, and lacks texture or shape features [10].

In recent years, due to the necessity of real-time early warning, the single-frame detection task has attracted much attention. Especially in the military background, with the development of supersonic fighter aircraft and supersonic reconnaissance aircraft. The fast movement of the target causes the fast-changing background to be inconsistent with the target motion trajectory. This results in the difference between the sensor platform and the actual location of the target, thus significantly reducing the performance of the sequence detection method [11]. However, the single-frame detection algorithms excel in real-time performance. Nonetheless, there is room for improvement in terms of detection accuracy in the presence of complex background interferences. Therefore, the detection of small targets in the single-frame infrared images holds significant importance [12].

For infrared dim small target detection, the traditional methods include median subtraction filter [13], top cap filter [14], maximum mean/maximum median filter [15], multiscale fuzzy metric [16], multichannel kernel fuzzy correlation graph [17], and improved fuzzy C-means based on spatial information [18]. The above method performs well in the scene with high local contrast, abundant texture feature information, and no interference. However, idealized targets with global unique saliency, striation, or high contrast do not exist in real-world infrared images. Moreover, these methods require prior knowledge of background scenes, which does not apply to complex and changeable background scenes, and especially unpredictable military scenes, which lack robustness.

Recently, with the development of deep learning, many advanced models have been proposed by domestic and foreign scholars. To solve the problem of weak features, Li et al. [19] proposed a tripartite densely nested interaction module (DNIM) with a hierarchical connection channel and spatial attention module (CSAM), which realized progressive

Manuscript received 15 March 2023; revised 24 June 2023 and 24 July 2023; accepted 9 August 2023. Date of publication 14 August 2023; date of current version 23 August 2023. This work was supported in part by the Shanghai Science and Technology Program "Federated based cross-domain and cross-task incremental learning" under Grant 21511100800, in part by the Natural Science Foundation of China under Grant 62076094, in part by the Chinese Defense Program of Science and Technology under Grant 2021-JCJQ-JJ-0041, and in part by the China Aerospace Science and Technology Corporation Industry-University-Research Cooperation Foundation of the Eighth Research Institute under Grant SAST2021-007. *(Corresponding author: Zhe Wang.)*

Zhe Wang, Tao Zang, Zhiling Fu, and Wenli Du are with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China (e-mail: wangzhe@ecust.edu.cn).

Hai Yang is with the Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai 200237, China.

Digital Object Identifier 10.1109/TGRS.2023.3304755

feature interaction and adaptive feature enhancement. For the target detection network that has been relatively formed, McIntosh et al. [20] fine-tuned the existing universal target detection networks (such as Faster-region convolutional neural network (RCNN) [21] and Yolov3 [22]) for infrared small target detection. To select the correct low-level features according to the high-level semantics and improve the detection accuracy of small targets, Dai et al. proposed the first detection method based on segmentation and designed an asymmetric context module (ACM) [23] to replace the normal skip connection of U-NET [24]. But they suffer from the following shortcomings. First, in the process of feature fusion, the weight of information at different levels is the same, which cannot fully highlight the significant characteristics of infrared dim small targets, and it is easy to be disturbed by target-like interference information. Second, in the training process, the shallow spatial features and deep semantic features are not combined or only the adjacent features are combined, which leads to the loss of deep fuzzy small target detection.

To highlight the significant features of infrared dim small targets and enhance the robustness of the model, we propose a pyramid feature fusion network combined with reinforcement learning (RLFPN). When performing feature fusion, reinforcement learning is used to assign the corresponding weight to the features. Specifically, the module uses multiscale convolution checks to convolve the features extracted and uses reinforcement learning adaptive generating weights to weigh the features in the last convolution. Then map them in the layers of each stage of the backbone network, to obtain the context information of different stages. Then, the above features with the same resolution are fused in each stage. This not only makes up for the lack of texture and other information of the infrared dim target and highlights the significant features of the infrared dim target but also fully extracts the context information. Compared with random selection, this kind of guidance $\varepsilon - greedy$ policy can increase the proportion of positive rewards in the experience buffer, thus improving the performance of the agent. And the experience of each time step may be used for multiple parameter update processes in the future to improve data utilization.

To integrate global spatial features and semantic features while reducing interference from a background in target detection, we propose a global context boundary attention (GB) module. This module incorporates low-level information into high-level features and merges spatial and semantic information. The model can improve the understanding ability of semantic features and scene background, reduce the fuzziness between objects and background interference, and enhance the target boundary information. In addition, the module refines the spatial position of pixel-level targets to ensure effective information dissemination. This approach also reduces the number of parameters, simplifies operations, and improves training speed.

In this article, the RLFPN module and GB module are combined to form a reinforcement learning of feature fusion and global context boundary attention networks (RLPGB−Net) for infrared dim small targets' detection through a single image as shown in Fig. 1. In RLPGB−Net, based on the unique characteristics of different images, such as the presence of interference or severe background noise, the RLFPN module makes the model more focused on the feature layers that highlight target information. During training, we provide cumulative feedback rewards through intersection over union (IoU), enabling the model to autonomously learn and acquire optimal feature weights. Therefore, our approach demonstrates robust and stable performance in detecting infrared weak targets in all the scenarios. In addition, the GB module embeds low-level features into high-level features, enhancing the target's edge information and compensating for the lack of texture, shape, and other information of small infrared targets. The experimental results on the SAITD [25] and SIRST [23] datasets demonstrate that our approach outperforms the state-of-the-art (SOTA) methods, showcasing the effectiveness of our work. The contributions of this article are as follows.

1) A pyramid feature fusion module combined with reinforcement learning is proposed which converts the feature weight generation task into a Markov decision process (MDP). Through strategy guidance and a long-term training method, a predictive agent based on deep Q network (DQN) is trained. By iteratively adjusting feature weights, improve the detection accuracy of infrared dim small targets.
2) The introduced GB module imparts low-level features into high-level features to combine spatial information with semantic information and realize efficient transmission of information from low resolution to high resolution, to highlight the boundary information of dim targets and prevent targets from being submerged by the background.
3) According to the experiments on part of the SAITD and SIRST datasets, our proposed infrared dim target detection network achieves a detection accuracy of over 95% and an average accuracy of over 90% on all the selected sequences, exceeding the SOTA methods.

The rest of this article is organized as follows. Section II introduces some related work of this article. In Section III, we describe the detail of the RLFPN module, which includes the MDP, the algorithm of DQN, and GB module. In Section IV, some experiments are carried out to prove the effectiveness of the proposed infrared dim small target detection network. Section V draws conclusions.

II. Related Work

In general, infrared dim small target detection methods can be divided into traditional detection methods and deep learning detection methods.

A. Traditional Infrared Small Target Detection Algorithm

The model based on background consistency assumes that the background region of the infrared image has a strong correlation. When there is a target in the background, the correlation of the background will be destroyed. Two-dimensional least mean square (TDLMS) [26] adaptive filter was proposed to detect small targets. In the later development process, new methods based on TDLMS are proposed, such as 2-D block

diagonal LMS adaptive filtering [27] and TDLMS filtering based on neighborhood analysis [28]. Probabilistic PCA [29] made use of the local difference between the background and infrared dim small targets, combined the kernel-based nonparametric regression technology with background prediction and clutter removal technology, and proposed a new infrared dim small target detection technology based on empirical mode decomposition (EMD) and improved local entropy [30]. TopHat filter [14] and [31], a single-frame detection method focusing on background clutter suppression, proposed an improved top hat transform based on target contour structure elements according to the characteristics of the target region to enhance the detection of mid-infrared small targets in a single-frame image. The methods that extend from TDLMS are limited to local information, and the subsequent PCA method did not improve it further, but rather use the difference between local target information and background information for background suppression. Similarly, local contrast method (LCM) [32] also conducts target detection through local comparison and compares the difference between the eight adjacent pixels of the current pixel. Due to pixel-by-pixel calculation, the calculation speed is slow and takes a long time. Both improved LCM (ILCM) [33] and multi-scale patch-based contrast measure (MPCM) [34] take multiscale plaques as the difference comparison unit, pay attention to the multidirection difference between the current patch and the surrounding patch, and accelerate the calculation speed.

Different from the above method, the model based on low-rank and sparse decomposition focuses more on the attributes of the background and target and pays overall attention to the essence of the infrared image. Infrared patch image (IPI) [11] regards the background image patch as low-rank and the small target in the image to be detected as the outlier. This method converts the infrared small target detection process into an optimization problem of low-rank decomposition. On this basis, many improved algorithms have been proposed successively: Zhang et al. [35] proposed nonconvex rank approximation minimization (NRAM) suppressed background clutter with sparse strong edges by introducing $l_{2,1}$-norm. Similarly, to suppress sparse strong edges, Wang et al. [36] proposed the method of full variation regularization and achieved certain results. Dai et al. proposed that the nonnegative IPI model based on the partial sum of singular values (NIPPS) [37] stores target data related to large singular values and constructs the IPI model using partial singular values and minimization instead of input to the whole matrix. The method based on IPI necessitates iterative calculations during the solving process, preventing it from attaining real-time detection performance. Although it has good detection performance, it cannot meet the requirements of current practical applications. To improve processing efficiency, reweighted infrared patch tensor (RIPT) proposed by Dai and Wu [38] adopts local and nonlocal priors and improves the IPI model by converting infrared images into third-order tensors. However, the partial sum of the tensor nuclear norm (PSTNN) [39] proposes a new nonconvex low-rank constraint, which reduces the algorithm complexity and computation time through sparsity and efficient reweighting of high-tensor singular value decomposition. Nonconvex tensor fibered rank approximation (LogTFNN) [40] introduces nonconvex approximate tensor fiber rank and uses local structure tensor to extract prior information to suppress sparse interference such as strong edges and corner points. The model based on low-rank and sparse decomposition focuses more on the attributes of the background and target. However, when the target is dim and the shape and contour features are not obvious, the detection results of these methods will be greatly affected. Our method integrates the global feature information and enhances the intrinsic relationship between the target pixel and the background pixel. It not only reduces the interference of background to the target detection but also eliminates the interference information similar to the strong target, reducing the missed detection and false positive.

B. Infrared Small Target Detection Based on Deep Learning

In the development process of deep learning, the problems existing in traditional methods are gradually overcome. To solve the localization problem, attentional local contrast networks (ALC-NET) [41] transforms the traditional LCM module into a parameterless nonlinear feature refinement layer in the end-to-end network and proposes a deep infrared small target detection network that combines the identification network with the traditional model-driven one. Although the detection accuracy is improved, the essence is still based on the extension of the local comparison method. Our method is based on the global feature to reduce the influence of interference on detection accuracy. ACM proposed a method of asymmetric context modulation module based on the global feature information of images, and at the same time added a bottom-up modulation point-level channel attention to exchange high-level semantics and subtle low-level details into the network, eliminating the disadvantages of using local information for detection. Similarly, Zhang et al. [42] introduced the attention guided context block (AGCB) in attention-guided pyramid context networks (AGPCNet), which combines local semantics with global context, integrating contextual information from multiple scales. But in feature fusion, all the features have the same weight, and target features are easy to be submerged after feature fusion. We use reinforcement learning to generate weight adaptively to highlight important feature information. Kou et al. proposed infrared small target segmentation network (IRSTNet) [43], which uses a pixel clustering strategy to obtain candidate targets. Furthermore, it reduces errors caused by factors such as background occlusion and interference information by further narrowing down the range of candidate targets through threshold adjustment. To eliminate the influence of background and other interference information on detection results, interior attention-aware network (IAANet) [44] used the area suggestion network to obtain the coarse target region and filter out the background. Fan et al. [45] inject the potential target region based on the convolutional neural network into the classifier to eliminate the nontarget region and achieve the purpose of effectively suppressing complex background clutter. Yang et al. [46] combined local and nonlocal spatial information, the local heterogeneity of the target, and the

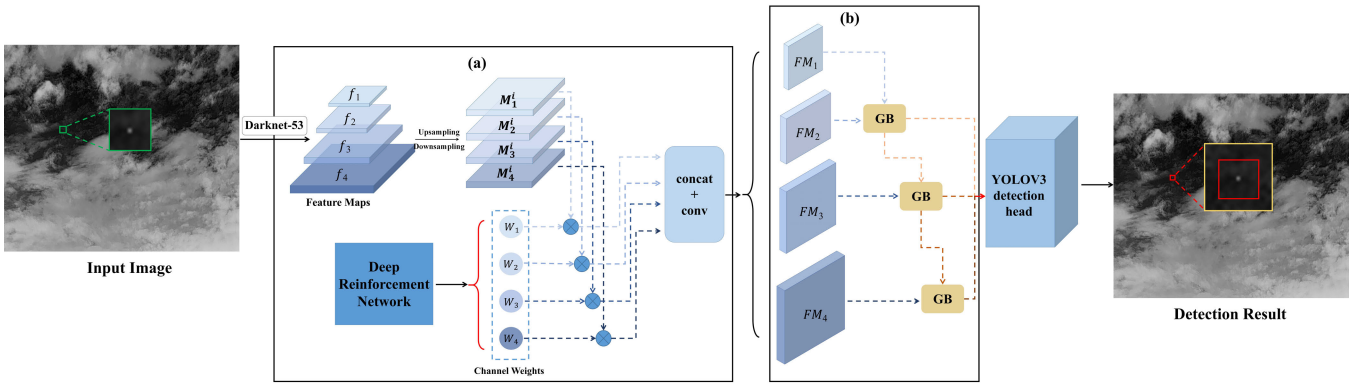

Fig. 1. Simple workflow of RLPGB-Net. It takes a single-frame image as input and first extracts features of four different dimensions through DarKnet-53. Through (a) RLFPN module, under the guidance of reinforcement learning, the original features are fused to obtain feature maps with four different resolutions. Then through (b) GB module combined with boundary refinement, the target edge is enhanced and the features of adjacent resolution are combined. Finally, the final detection result is output through the detection head of Yolov3.

nonlocal autocorrelation of the background with the sparsity of the target, and the real target is separated from the complex background clutter. Zhang et al. [47] proposed a method of calculating local intensity and gradient (LIG) from the original infrared image to enhance the target and suppress the background interference clutter. Infrared small target detector (ISTDet) [48] uses an image filtering module to obtain the confidence map to enhance the response of the infrared small target and suppress the response of the background. The above five methods [44], [45], [46], [47], [48] are to separate the background from the target to improve the detection accuracy. However, these methods will not be effective when the target is dim and close to the background or when there is a similar interference with the target. However, we make the model strengthen the learning target features and reduce the influence of background and distractors. Using global feature information, our method can solve the influence of background and similar object interference on detection results. When combining features of different dimensions, the features of infrared dim small targets (such as the lack of texture, shape, and other shallow features) are fully considered. The powerful decision-making ability of reinforcement learning enables the model to generate corresponding weights independently, highlighting the features that can improve the detection accuracy of infrared dim small targets.

C. Combined With Reinforcement Learning Target Detection Algorithm

Reinforcement learning has been introduced into object detection over the years, mainly used to search the target area, replace the classical detector, find the region of interest (ROI) in very large images to run the detector, etc. Liu et al. [49] replace the traditional attention mechanism by searching and locating subregions that may contain relevant objects through intelligent agents and scale all the generated boundary boxes according to the original resolution to clear subedges and overlapping parts. Yao et al. [50] proposed an object detection model of regional attention reinforcement learning, which uses the recursive network to extract historical information, fuses historical information with current relevant information, and pays attention to the fusion information of possible objects. Zhou et al. [51] proposed a new object detection framework based on reinforcement learning. By combining the action space of the reinforcement learning subject with the convolutional neural network, the detection model has the ability of regional selection and refinement to provide more accurate regional suggestions. Mathe et al. [52] proposed a model with a sequence of principles. By developing a search strategy for sequential search reinforcement learning, more areas of effective images were sampled to achieve the purpose of effectively detecting visual objects. Wang and Qin [53] proposed a small-moving target detection method under a pipeline framework based on reinforcement learning. Multiple separated target images of the same input image were confirmed by a voting mechanism, replacing the classical detector. A visual recognition model based on deep reinforcement learning is proposed [54], which consists of a sequence region suggestion network and a target detector. This is achieved by replacing the greedy RoI selection process with the sequential attention mechanism of deep reinforcement learning training to optimize the ROI and bring it closer to the final detection task. In the above methods, [49], [50], [51], and [53] use reinforcement learning to search the target region of the image, while [52] uses a voting mechanism to output the best detection results. The difference is that we use reinforcement learning adaptive weight generation to guide feature fusion. Compared with the above method, the advantages are as follows: 1) the model is easy to converge, saves computing time, and consumes less computing resources and 2) when the target is dim, it can reduce the influence of the dim target being submerged by the background and the interference of similar targets, which reduces the missed alarm rate and false alarm rate.

III. Method

In this section, we will describe the details of the RLFPN module and GB module.

A. Overview

Compared with other infrared target detection frameworks, we first used reinforcement learning to dynamically assign weight during feature fusion. Fig. 1 shows the RLPGB-Net,

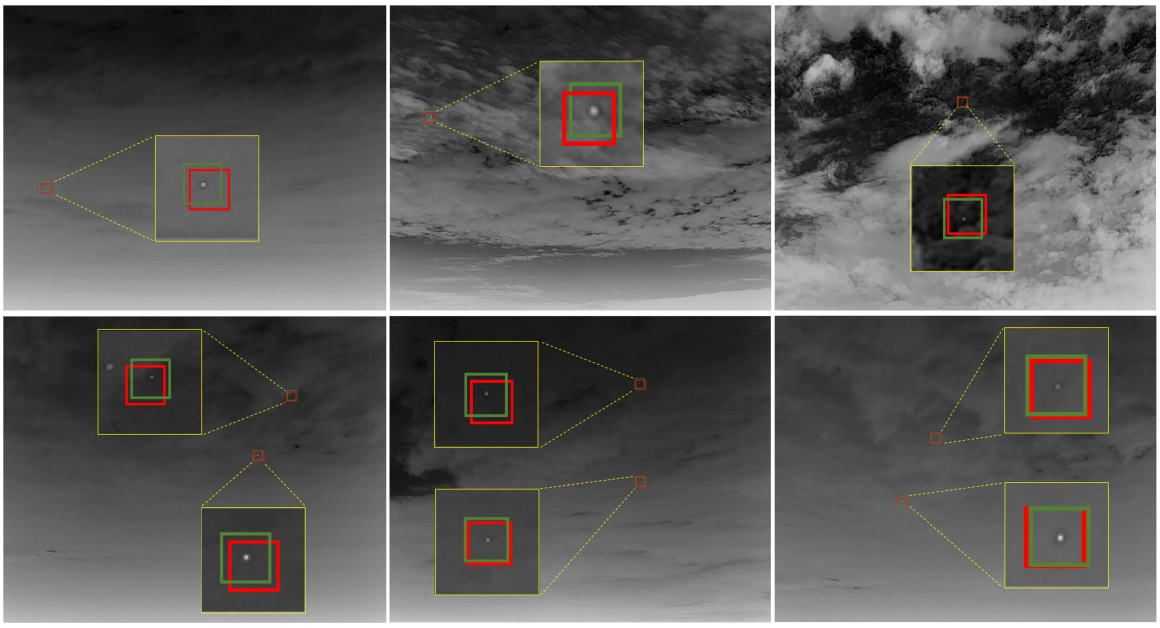

Fig. 2. Prediction box and real box of infrared small target. The green box is ground truth, and the red box is the result of the prediction. For better visualization, we expanded the display area.

and the most important architecture of this method is the RLFPN module and GB module. The entire detection process is as follows: first, a batch of original images is input into the network, and the DarKnet-53 feature extraction network uses to extract features and obtain the multilevel feature F (f_1, f_2, f_3, f_4). Next, multilayer feature F is input into the RLFPN module after upsampling and downsampling to the same dimension (M_1^i, M_2^i, M_3^i, M_4^i). Reinforcement learning is used to generate adaptive features of different dimensions and assign corresponding weights to highlight important features of the detected target. Through some intensive jumps, these features are connected and fused into four different resolutions ($\mathrm{FM}_1 - \mathrm{FM}_4$). Then, $\mathrm{FM}_1 - \mathrm{FM}_4$ propagate valid information from low resolution to high resolution through the GB module and refine the target boundary. Finally, the obtained feature mapping was input into the detection head of Yolov3 to obtain the final detection result.

B. Feature Fusion Pyramid Network With Reinforcement Learning

The feature pyramid networks (FPNs) adds the resized higher layer feature map to the current layer to obtain the new feature map. The aliasing effect produced by this feature mapping connection highlights the important features of the target and also enhances the features of the interference target, so it will harm the position and classification of the infrared dim small target. As Fig. 2 shows, the infrared dim target ground truth is marked with the green box. At the same center point, the red box is the result of the prediction. To improve the overlap between the predicted box and the real box, some researchers use extensive experiments and prior knowledge to adjust the weight of features. Some problems exist, such as weak generalization and strong subjectivity of the model, which may not be adjusted to the optimal weight. To solve these problems, we introduce reinforcement learning and propose the model of RLFPN, as shown in Fig. 3.

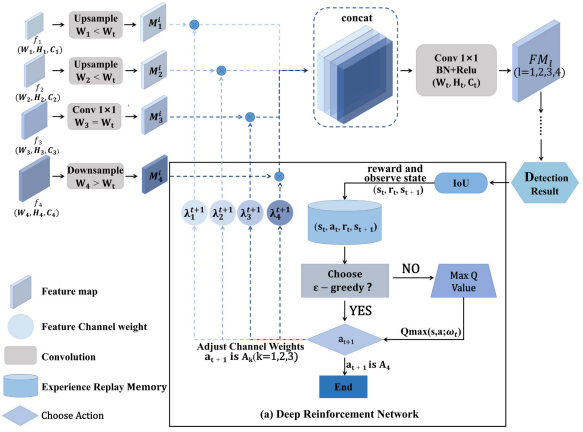

Fig. 3. Details about the RLFPN module. RLFPN module combines reinforcement learning and adopts multiscale features of the finite element method. Reinforcement learning adaptively assigns weights to different dimensional features and then adjusts the features to the same dimension as the target fusion resolution through scaling, downscaling, and transformation operations. Finally, the final fusion feature map $\mathrm{FM}_i\,(i = 1, 2, 3, 4)$ is output by the join and 1×1 convolution. (a) Deep reinforcement network.

In RLFPN, the outputs of each stage residual block in DarKnet-53 are used as original feature maps. For simplicity, these four features could be denoted by a feature set $F : F = (f_1, f_2, f_3, f_4)$. Then through convolution, resizing, giving weight, and concat, the original feature maps are converted into multichannel feature maps. For f_1, the module downscales it into four resolutions with a stack of convolution layers, and the output feature maps are as follows:

$$\begin{aligned} M_1^i &= \phi(f_1|W, b) \\ &= \sigma(\text{down-scale}(N_{k\times k} * f_1 + b))(i = 1, 2, 3, 4) \end{aligned} \tag{1}$$

where σ refers to the rectified linear unit (ReLU) activation, down-scale$(\cdot)$ signifies $N_{k\times k}$ (kernel size is $k \times k$, stride $s = k$) to downscale the feature map f_1, b is the bias, and $*$

TABLE I
DETAILS OF REINFORCEMENT LEARNING PYRAMID FEATURES' FUSION MODULE

Stage	FM_1	FM_2	FM_3	FM_4
Conv1_0	$1\times1, s=1$	$2\times2, s=2$	$4\times4, s=4$	$8\times8, s=8$
Conv2_1	deconv	$1\times1, s=1$	$2\times2, s=2$	$4\times4, s=4$
Conv3_2	deconv	deconv	$1\times1, s=1$	$2\times2, s=2$
Conv4_3	deconv	deconv	deconv	$1\times1, s=1$

is the convolution. For f_4, the model upsamples it into four resolutions, and the output feature maps M_4^i are as follows:

$$M_4^i = \psi(f_4|W,b) = \sigma(\text{upsample}(f_4;\varphi))(i=1,2,3,4) \quad (2)$$

where upsample $(\cdot;\varphi)$ refers to the deconvolution with parameters φ which are learned during training. For f_2 and f_3, the two feature maps need to use the combination of downscale and upsample to resize into four resolutions, and the output feature maps M_l^i are as follows:

$$M_l^i = \sigma(\phi(f_l)\&\psi(f_l))(i=1,2,3,4,l=2,3). \quad (3)$$

Here, $\phi(\cdot)$ and $\psi(\cdot)$ denote (2) and (3). The channel dimension of these resized feature maps $(M_1^i, M_2^i, M_3^i, M_4^i)$ is 128. Finally, the features with the same dimension in these output ones are fused to generate the final four fused feature maps. To be convenient, the four fused feature maps are named FM_1 $(n=0)$, FM_2 $(n=1)$, FM_4 $(n=2)$, and FM_8 $(n=3)$. The four fused features could be defined as

$$\text{FM}_i = \sigma\left(W_{1\times1} * \text{CAT}\left(\lambda_1^j M_1^j, \lambda_2^j M_2^j, \lambda_3^j M_3^j, \lambda_4^j M_4^j\right) + b\right) \times (j=1,2,3,4, i=j). \quad (4)$$

CAT is elementwise concatenated. The channel of all the fusion feature maps FM is 512. $W_{1\times1}$ is a convolution in which the kennel size is 1×1 to change the channel dimension of FM (512–128), and λ is the weight of the current feature, which is generated by adaptive learning of the reinforcement learning network. After the feature fusion was guided by reinforcement learning, the convolution layer, ReLU activation, and batch normalization were performed and these parameters are trainable (as shown in Table I).

This article uses DQN to generate feature weights, which uses experience playback to break the correlation between data. In the process of reinforcement learning, the agent saves the data to a database, extracts the data from the database through uniform random sampling, and then uses the extracted data to train the neural network to solve the problem of data association. The update of neural network weights is defined as

$$\omega_{i+1} = \omega_i + \alpha\left[r + \gamma_a^{\max} Q\left(s', a'; \omega_i^-\right) - Q(s,a;\omega_i)\right] \nabla Q(s,a;\omega_i) \quad (5)$$

where ω_i is the weight of DQN, and ω_{i+1} is the weight. The agent selects an action at state s; then, the state changes into s'. The learning rate is α, and γ represents the discount factor in discounted reward.

We use policy guidance instead of random action selection in the original algorithm to help the agent make appropriate actions in training. The original DQN selects random action with a probability in each $\varepsilon - greedy$ decision-making stage. However, we optimize this kind of selection in our method. Before the agent selects an action, we can calculate the IoU between the ground truth and the prediction. Based on the IoU difference, the most appropriate action is chosen for the agent to get the $+2$ or $+5$ encouragement. Compared with random selection, this kind of policy guidance increases the proportion of positive reward in the experience buffer, which improves the performance of the agent.

For the prediction task, we use long-term training that updates the weights until the end of an MDP. On one hand, this approach enables the agent to quickly accumulate experience. On the other hand, it accelerates the training of networks.

In the MDP of weight adjustment, the agent observes the IoU of the current prediction box and the ground-truth box, and it selects one action from our predefined action set [as shown in Fig. 3 (a)]. Then, the weight is adjusted by the value determined by the selected action. In addition, the agents will also be encouraged or punished according to whether the predicted result is closer to the ground truth. These decision processes and state transitions are continually cycled until the end of MDP. In the above MDP, there are three import elements: state set, predefined action set, and reward function which evaluate the selected action.

The action set consists of four actions: Action1, Action2, Action3, and Action4. When the agent selects Action1, the weight will increase $\mathscr{C}\%$. Correspondingly, when Action2 is selected, the weight will decrease $\mathscr{C}\%$. Action3 means the weight is unchanged, and Action4 marks the end of weight adjustment. After selecting Action4, we can get the best weight through the overall decision-making process. With these different values in action sets to adjust the weights, we can achieve higher detection accuracy and reduce the number of decision-making actions. This short-term MDP can facilitate the convergence of the training process.

The reward function encourages or punishes the agent through the selected action, and the value is obtained by [55]. The agent accumulates experience with the above reward, learns from it, and finally selects appropriate actions in each decision. We contrast the reward functions $R_{a1}(s_t, s_{t+1})$, $R_{a2}(s_t, s_{t+1})$, $R_{a3}(s_t, s_{t+1})$, and $R_{a4}(s_t, s_{t+1})$ for each action based on the difference between the ground-truth box ($\text{box}_{\text{truth}}$) and the next weighted prediction box (box_{pre}), which is generated by adjusting the weight corresponding to the current feature and the action selected by the agent. The initial weight of all the four features is defined as $\mathscr{A}\%$. The reward functions are defined as

$$R_{a1}(s_t, s_{t+1}) = \begin{cases} +2, & \text{if } \text{IoU}_{\text{now}} \le \text{IoU}_{\text{next}} \\ -2, & \text{otherwise} \end{cases}$$

$$R_{a2}(s_t, s_{t+1}) = \begin{cases} +2, & \text{if } \text{IoU}_{\text{now}} \le \text{IoU}_{\text{next}} \\ -2, & \text{otherwise} \end{cases}$$

$$R_{a3}(s_t, s_{t+1}) = \begin{cases} +2, & \text{if } \text{IoU}_{\text{now}} \le \text{IoU}_{\text{next}} \\ -2, & \text{otherwise} \end{cases}$$

$$R_{a4}(s_t, s_{t+1}) = \begin{cases} +5, & \text{if } \mathrm{IoU}_{\mathrm{next}} \geq \mathscr{B} \\ -5, & \text{otherwise} \end{cases} \tag{6}$$

where the state is s_t at time t, $\mathrm{IoU}_{\mathrm{now}}$ is based on the current ground-truth box and the prediction box, $\mathrm{IoU}_{\mathrm{next}}$ is the prediction box and the ground-truth box after the agent takes action, and the agent selects action a_k $(k = 1, 2, 3, 4)$, then the state changes into s_{t+1}. Based on the difference mentioned above between weight adjustment and the threshold we set, the agent gets encouragement $(+2, +5)$ or punishment $(-2, -5)$.

The reward functions $R_{a1}(s_t, s_{t+1})$, $R_{a2}(s_t, s_{t+1})$, and $R_{a3}(s_t, s_{t+1})$ evaluate whether the agent adjusts the weight reasonably. For example, when the difference between the ground-truth box and the prediction box and the calculated IoU is less than $\mathscr{B}$, it means that the agent needs to adjust the weight of the feature. At this time, if the agent selects action A1, the weight of at least one of the four chosen features will increase $\mathscr{C}\%$ and others will decrease. If $\mathrm{IoU}_{\mathrm{now}} \leq \mathrm{IoU}_{\mathrm{next}}$, then the agent can get $+2$ encouragement. If action A2 is selected, $\mathrm{IoU}_{\mathrm{now}} \geq \mathrm{IoU}_{\mathrm{next}}$, and the agent will be punished -2.

The reward function $R_{a4}(s_t, s_{t+1})$ affects the detection accuracy. When the end action A4 is selected, the entire MDP is over. If IoU $\geq \mathscr{B}$, the end encouragement is $+5$; otherwise, the punishment is -5. We will set parameters $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$ in the experimental section.

Algorithm 1 shows the training pseudocode of optimized DQN. We adopt the policy guidance and long-term training method to make the agent learn effectively. Subsequent experimental results demonstrate the advantages of the optimized algorithm.

Through this process, the model effectively obtains the information of multiscale context and realizes the overall perception of the target.

C. GB Module

If large convolution kernel bilinear upsampling is used in feature fusion, some details may be lost directly, and small infrared details and weak infrared details cannot be highlighted. At the same time, a lot of calculation leads to slow training speed, and even the target parameters cannot be trained. If bilinear upsampling with a small convolution kernel is used, the acceptance field can be reduced directly, but the global context information cannot be used effectively. Moreover, the detected targets are dark and easy to be swallowed by the background, so the edge refinement of small targets is conducive to improving the final detection accuracy. Therefore, we combined the GB [56] module. The proposed GB module is shown in Fig. 4, including two stages: The first stage, boundary refinement, is carried out on the low-dimensional feature map, and the output is as follows:

$$\tilde{B} = \sigma(W_{1\times1} * (S \oplus \mathbb{R}(S)) + b) \tag{7}$$

where $W_{1\times1}$ indicates trainable parameters, S and $\mathbb{R}(\cdot)$ signify the coarse score map and residual branch, respectively, $\oplus$ is the cross-channel concatenation, and b refers to the bias. Then, convolution is used to adjust the channel dimensions of high-resolution and low-resolution fusion feature maps, and

Algorithm 1 Algorithm1 Optimized Deep Q Learning

Input: Initialize replay memory D to capacity N, action-value function Q with random weights ω, target action-value function Q^- with weights ω^- which are duplicated from ω.
Output: the feature maps weights of trained DQN
Set with 100% in $\varepsilon - greedy$ policy
for RL counter = 1,R **do**
Initialize state (Infrared detection image m_1 + past action p_1)
Update with max($\varepsilon - 10\%$,0.1)
for plan counter = 1,P **do**
Duplication ω^- with ω each M action selection
if $random(0, 1) > \varepsilon$ **then**
Select action a_i corresponding to the maximum of the Q network output
else
Set network label y_i with r_i Calculate IoU
Perform policy guidance to get the action a_i
Adjust the corresponding weight according to the corresponding way of a_i to obtain new state s_{i+1}(new Infrared detection image m_{i+1}+ new past action p_{i+1})
Reward r_i storage (s_i, a_i, r_i, s_{i+1})
Set current state s_i with s_{i+1} in D.
end if
if $a_j == A_4$ **then**
Network label y_i with r_i
else
Network label y_i with $r_i + \gamma_{a'}^{max} Q(s', a'; \omega_i^-)$
end if
Calculate loss$(y_j - Q(s, a; \omega_i))^2$ and update the weights in Q with loss
Backpropagation
end for
end for

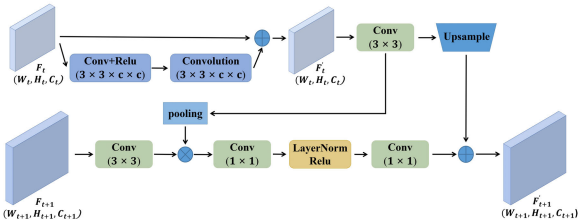

Fig. 4. Details of the GB module.

global context features are realized by global average pooling, which is then multiplied with high-resolution feature maps. The output is as follows:

$$f_1 = \sigma\left(C_{3\times3} * f^h + b\right) \otimes \mathcal{P}\left(\sigma\left(C_{3\times3} * \tilde{B} + b\right)\right) \tag{8}$$

where $\otimes$ and $*$ denote the elementwise multiplication and convolution, respectively, $\mathcal{P}$ denotes the global pooling operation, f^h and f^l represent the high-resolution and low-resolution fusion feature maps, and $C_{3\times3}$ indicates trainable parameters.

The second stage: the low-resolution fusion feature map is upsampled to make it have the same dimension as the

TABLE II
DETAILS OF TEN REAL IMAGE SEQUENCES. ALL IMAGES' SIZES ARE 640×512, TARGETS' TYPES ARE PLANE, AND TARGETS' SIZES ARE 3×3–6×6

	Frames	Target number	Target details	Background details
Sequence1	110	One	A long imaging distance with a single target keeping moving	Dim sky without cloud A similar target bright spot jamming
Sequence2	113	One	A long imaging distance with a single target keeping moving	Dim sky with scattered white clouds
Sequence3	114	One	A long imaging distance with a single target keeping moving	Dim sky with many heavy white clouds
Sequence4	121	Two	A long imaging distance with two targets keeping moving	Dim sky background without cloud
Sequence5	126	Two	A long imaging distance with two targets keeping moving	Dim sky with many heavy white clouds A similar target bright spot jamming
Sequence6	129	Two	A long imaging distance with two targets keeping moving	Dim sky with few white clouds A similar target bright spot jamming
Sequence7	131	Three	A long imaging distance with three targets keeping moving	Dim sky without cloud A similar target bright spot jamming
Sequence8	134	Three	A long imaging distance with three targets keeping moving	Dim sky with few white clouds
Sequence9	135	Three	A long imaging distance with three targets keeping moving	Dim sky has a few heavy white clouds and a few dull clouds
Sequence10	139	Three	A long imaging distance with three targets keeping moving	Dim sky with few white clouds

high-resolution fusion feature map, and then the output f_1 of the first stage is added. The output of the second stage f_2 is as follows:

$$f_2 = \sigma\left(\text{upsample}\left(f^l;\varphi\right)\right) \oplus (\sigma(\text{LN}(f_1 * C_{1\times1}) + b) * C_{1\times1})) \tag{9}$$

where upsample$(\cdot;\varphi)$ refers to the deconvolution with parameters φ which are learned during training, $C_{1\times1}$ indicates trainable parameters, LN refers to the layer norm, f_1 is the output of stage one, and $\oplus$ refers to elementwise addition.

In short, compared with the previous simple addition of coarse-resolution feature maps to the upsampled high-resolution feature maps, the GB module introduced in this article makes full use of the fusion feature maps of different resolutions, improves the efficiency of context acquisition, and achieves the corresponding pixel-level positioning of infrared dim small targets.

IV. EXPERIMENTS

A. Datasets

In this article, we select the public dataset SAITD and dataset SIRST.

1) SAITD Dataset: To compare with other methods, we need to divide the dataset into the training set and test set. First, we screened the sequences of empty targets in this dataset, excluded some sequences that did not exist in targets in the pictures, and finally got ten sequences. The details of the ten sequences are shown in Table II. The ten sequences are 3800 images in total. Among them, Fig. 5 shows some representative images from different sequences in the SAITD dataset. We randomly divided the selected sequences in a random way, fivefold cross-validation method, and selected two sequences each time as the test set, and the other eight sequences as the training set. Using sequence as the unit of division can eliminate the problem that the test result is too high due to the high similarity of adjacent pictures of sequence.

2) SIRST Dataset: The SIRST is a public, single-frame dataset. The dataset extracts the most representative images from hundreds of sequences and selects only one representative image from each infrared sequence to avoid the overlap between the training set and the test set. The dataset consists of 427 infrared images and 480 targets.

B. Evaluation Metrics and Setting of Parameters

1) Evaluation Metrics: In the final result of target detection, there are four types: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). TP is the detected target itself as a positive sample and predicted as a positive sample; TN is the detected target itself as a negative sample and predicted as a negative sample; FP is the detected target itself as a negative sample but predicted as a positive sample; and TP is the detected target itself as a positive sample but predicted as a negative sample.

In this article, we introduced common target detection indicators such as accuracy (Acc), precision (P), recall (R), average precision (AP), IoU, nIoU [41], and F1 Score (F1). Acc is the proportion of all the forecasts that are correct, and the calculation method is as follows:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{10}$$

The precision rate refers to the probability of correct detection among all the detected targets, which is calculated as follows:

$$P = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{11}$$

The recall rate refers to the probability of correct identification among all the positive samples and is calculated as follows:

$$R = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{12}$$

AP is the area enclosed under the precision–recall curve (PR curve). PR curve is the curve drawn by the detector after the values of precision and recall are obtained, the horizontal axis is recall, and the vertical axis is precision. The calculation formula is

$$P_{\text{smooth}}(r) = \max_{r' \geq r} P\left(r'\right) \tag{13}$$

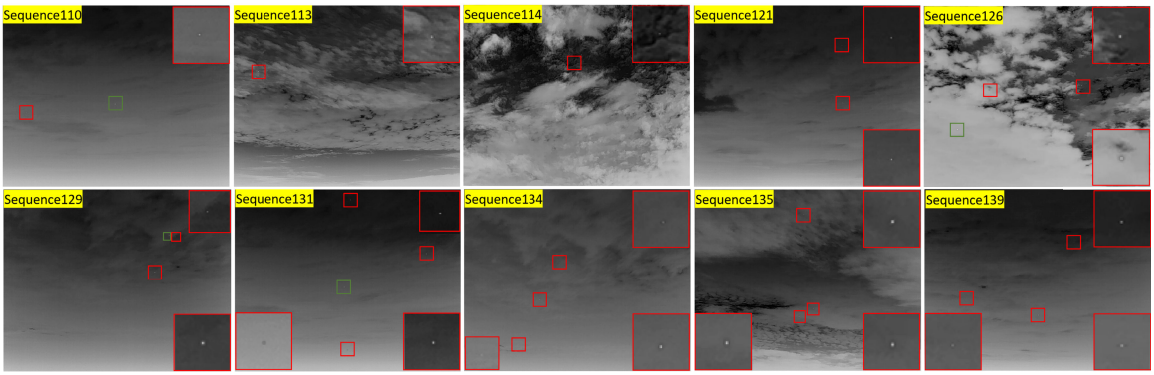

Fig. 5. Representative infrared images from different sequences in the SAITD dataset. For better visualization, the target area is enlarged. The correct targets and jamming objects are highlighted with red and green boxes, respectively.

$$\mathrm{AP} = \int_0^1 P(r)dr. \tag{14}$$

$P_{\mathrm{smooth}}(r)$ is the curve value corresponding to recall after smoothing, and $P(r')$ is the curve value corresponding to recall before smoothing. IoU calculates the ratio of the intersection and union of the border of the predicted target and the border of the real target label. The calculation method is as follows:

$$\mathrm{IoU} = \frac{\mathrm{intersection}}{\mathrm{union}}. \tag{15}$$

nIoU is a new evaluation index proposed by Dai et al. [41], which can more fairly compare the data-driven methods and traditional-model-driven methods. The calculation method is as follows:

$$\mathrm{nIoU} = \frac{1}{N}\sum_{i}^{N}\frac{\mathrm{TP}[i]}{T[i]+P[i]-\mathrm{TP}[i]} \tag{16}$$

where N is the total sample number. The F1 score is the harmonic average of accuracy and recall rate. For a single category of target detection, the formula is as follows:

$$F1 = 2 * \frac{\mathrm{Recall} * \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}}. \tag{17}$$

Receiver operating characteristic (ROC) curve is the curve drawn by the true positive rate (TPR) and false positive rate (FPR) after the detector detects the result. The horizontal axis is FPR, and the vertical axis is TPR. The formula for calculating TPR is

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}. \tag{18}$$

The formula for calculating FPR is

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}}. \tag{19}$$

We refer to the strategy in [41] and [57] and apply the same evaluation index to the comparison methods based on object detection and semantic segmentation below.

2) Parameters' Setting: Parameters $\mathscr{A}$ and $\mathscr{B}$ in MDP are set to 25 and 0.9, respectively. In the process of model training proposed in this article, we adopted Adam's optimization algorithm. In the first stage, we set 80 epochs of training, with a learning rate of 1e-3 and a batch size of 16. In the second stage, we set up 220 epochs of training, with a learning rate of 3e-3 and a batch size of 8. Through clustering, the Yolo anchors' boxes used in training are (10,13), (16,30), (33,23), (30,61), (62,45), (59,119), (116,90), (156,198), and (373,326). During the whole training process, the agent interacts with the dataset for 30 times each time. As shown in the Algorithm1, when a training dataset loop ends, the $\varepsilon - greedy$ policy is implemented. After the end of the tenth loop, ε is set to 0.1 and it remains the same, randomly selecting 32 units from the experience buffer at a time. In the training of DQN, we use mean squared error as the loss function. In the training of the detection and classification network, we adopt the original loss function of Yolov3. For comparison methods, Tophat [59], LCM, ILCM, FKRW [58], MPCM, IPI, NIPPS, RIPT, generalised structure tensor (GST) [60], edge and corner awareness-based spatial–temporal tensor (ECA-STT) [61], NRAM, NRAM, LogTFNN, PSTNN, U-NET, DNA-NET [19], Yolov4 [62], YoloX [63], AGPCNet, and lightweight infrared small target segmentation network (LW-IRSTNet) [64], the details of parameter setting are as shown in Table III.

C. Comparison With SOTA Methods

To demonstrate the validity of our method, we compared our model with the current SOTA methods. We selected 17 methods for comparison on the SAITD dataset, including tensor-based methods: GST, ECA-STT, LogTFNN, and PSTNN; consistency-based methods: Top-hat; human visual system-based methods: LCM, ILCM, and MPCM; IPI model-based methods: IPI, NRAM, and NIPPS; infrared patch-tensor model-based methods: RIPT; facet kernel and random-walker-based methods: FKRW; full-convolution-based methods: U-NET; and cnn-based methods: DNA-NET, Yolov4, YoloX, AGPCNet, and LW-IRSTNet. In the SIRST dataset, four more comparison methods (FPN [65], global attention upsampling (GAU) [66], ACM, and ALC-NET) were added to the 17 methods. As the SAITD dataset is more challenging, some comparison methods do not perform well in the results of this dataset, so the comparison methods of the two datasets are different.

First, we compared the proposed method with the above method on the SAITD dataset, and the results are shown in Table IV. The results of our proposed method are the best in all the indicators, with about 36% improvement in precision compared with the best traditional method, and about 4% improvement in precision compared with the best

TABLE III
DETAILED PARAMETER SETTINGS FOR COMPARISON METHODS

Method	Parameters
Tophat	Structure shape: disk, Structure size: 3× 3
LCM	Cell size: 3× 3
ILCM	Subblock size: 8 × 8, Moving step: 4
FKRW	$K = 4$, $p = 6$, $\beta = 200$, Window size= 11 × 11
MPCM	N = 3, 5, 7, 9, Mean filter size: 3× 3
IPI	Patch size: 50× 50, Sliding step: 10, $\lambda = 1/\sqrt{min(m,n)}$, $\varepsilon = 10^{-7}$
NIPPS	Patch size: 50× 50, Sliding step: 10, $\lambda = 1/\sqrt{min(m,n)}$, $\varepsilon = 10^{-7}$
RIPT	Patch size: 30× 30, Sliding step: 10, $\lambda = L/\sqrt{min(m,n)}$, $L = 1$, $h = 1$, $\varepsilon = 10^{-7}$
GST	$\sigma_1 = 0.6$, $\sigma_2 = 1.1$, $C_n = 1/255$, Filter size: 5 × 5
ECA-STT	$\beta = 0.1$, $t = 3$, $\lambda_1 = 0.009$, $\lambda_2 = 5.0/\sqrt{min(m,n) \times t}$, $\lambda_3 = 100$, $\varepsilon = 10^{-7}$
NRAM	Patch size: 30 × 30, Step:10, $\lambda = 1/\sqrt{min(m,n)}$, $\varepsilon = 10^{-7}$, $\mu = \sqrt{min(m,n)}$, $\gamma = 0.05$, $C = \sqrt{min(m,n)}/2.5$
LogTFNN	Patch size: 40 × 40, Step:40, $\lambda = 1/\sqrt{min(n_1,n_2) * n_3}$, $\beta = 0.01$, $\mu = 200$
PSTNN	Patch size: 40 × 40, Step:40, $\lambda = 0.7/\sqrt{min(n_1,n_2) * n_3}$, $\varepsilon = 10^{-7}$
U-NET	Epoch: 100, Batch size: 4, Learning rate: $1e^{-5}$
DNA-NET	Down-sampling layer:4, Batch size:16, Learning rate:0.05
Yolov4	Freeze epoch:50, Freeze batch size:6, UnFreeze epoch:500, UnFreeze batch size:2, Init learning rate:$1e^{-3}$, Min learning rate:$1e^{-5}$
YoloX	Freeze epoch:50, Freeze batch size:18, UnFreeze epoch:500, UnFreeze batch size:9, Init learning rate:$1e^{-3}$, Min learning rate:$1e^{-5}$
AGPCNet	SGD momentum: 0.9, SGD weight decay: $1e^{-4}$, Batch size: 8, Init learning rate: 0.05
LW-IRSTNet	SGD momentum: 0.9, SGD weight decay: $1e^{-4}$, Epoch: 100, Batch size: 64, Init learning rate: 0.05

TABLE IV
COMPARISON OF OUR MODEL WITH TENSOR-BASED METHODS: GST, ECA-STT, LOGTFNN, AND PSTNN; CONSISTENCY-BASED METHODS: TOP-HAT; HUMAN VISUAL SYSTEM-BASED METHODS: LCM, ILCM, AND MPCM; IPI MODEL-BASED METHODS: IPI, NRAM, AND NIPPS; INFRARED PATCH-TENSOR MODEL-BASED METHODS: RIPT; FACET KERNEL AND RANDOM-WALKER-BASED METHODS: FKRW; FULL-CONVOLUTION-BASED METHODS: U-NET; CNN-BASED METHODS: DNA-NET, YOLOV4, YOLOX, AGPCNET, AND LW-IRSTNET. ESPECIALLY, THE BEST RESULTS ARE SHOWN IN BOLD

Methods	P(%)	R(%)	Acc(%)	AP(%)	F1	IoU	nIoU
Tophat	57.31	67.11	44.75	54.42	0.618	0.514	0.580
LCM	39.39	30.27	20.65	34.08	0.342	0.328	0.376
ILCM	36.53	25.96	17.89	32.99	0.304	0.298	0.352
FKRW	41.38	32.88	22.43	36.90	0.366	0.351	0.382
MPCM	54.99	61.17	40.76	50.71	0.579	0.491	0.553
IPI	55.38	64.92	42.62	52.99	0.598	0.532	0.569
NIPPS	56.90	63.53	42.89	52.92	0.600	0.481	0.513
RIPT	37.36	31.98	20.81	50.71	0.345	0.343	0.384
GST	61.52	37.48	29.77	40.69	0.443	0.479	0.628
ECA-STT	58.85	65.10	43.07	56.33	0.577	0.472	0.586
NRAM	61.57	70.07	50.69	59.65	0.628	0.493	0.657
LogTFNN	53.49	54.55	35.90	52.28	0.522	0.431	0.516
PSTNN	57.39	61.97	41.78	55.60	0.573	0.464	0.577
U-NET	61.01	72.48	49.54	56.08	0.663	0.638	0.651
DNA-NET	88.09	84.77	80.06	85.49	0.901	0.876	0.891
Yolov4	93.37	89.03	80.31	89.05	0.924	0.839	0.849
YoloX	78.56	65.57	55.17	69.88	0.717	0.629	0.636
AGPCNet	83.74	78.65	65.89	79.36	0.811	0.791	0.804
LW-IRSTNet	79.54	76.88	59.20	71.51	0.769	0.701	0.713
RLPGB-Net(ours)	**97.42**	**90.13**	**80.75**	**89.89**	**0.934**	**0.906**	**0.922**

depth method. We selected some representative results for presentation, as shown in Fig. 6,; the results obtained by the traditional method are easy to lose dim infrared small targets, leading to missed detection. When there are heavy clouds in the background, it is easy to detect part of the clouds as small targets, causing the problem of wrong detection. U-NET is similar to the best detection results of traditional methods, but there are serious omissions. Dense nested attention network (DNA-Net), AGPCNet, LW-IRSTNet, Yolov4, and YoloX are much better than the traditional method, but they only reduce the influence of background interference and dim target on the final detection results to a certain extent. In addition, the above methods lack robustness for target detection in different scenarios. The detection accuracy can achieve good results in simple and noninterference scenarios, but the detection results are poor for targets with interference or complex background scenarios. This proves the effectiveness of our model compared with other baselines. Through reinforcement learning and the GB module, the influence of background and disturbance on the final detection results is reduced, which proves that our model is robust and applicable to complex and changeable military scenarios.

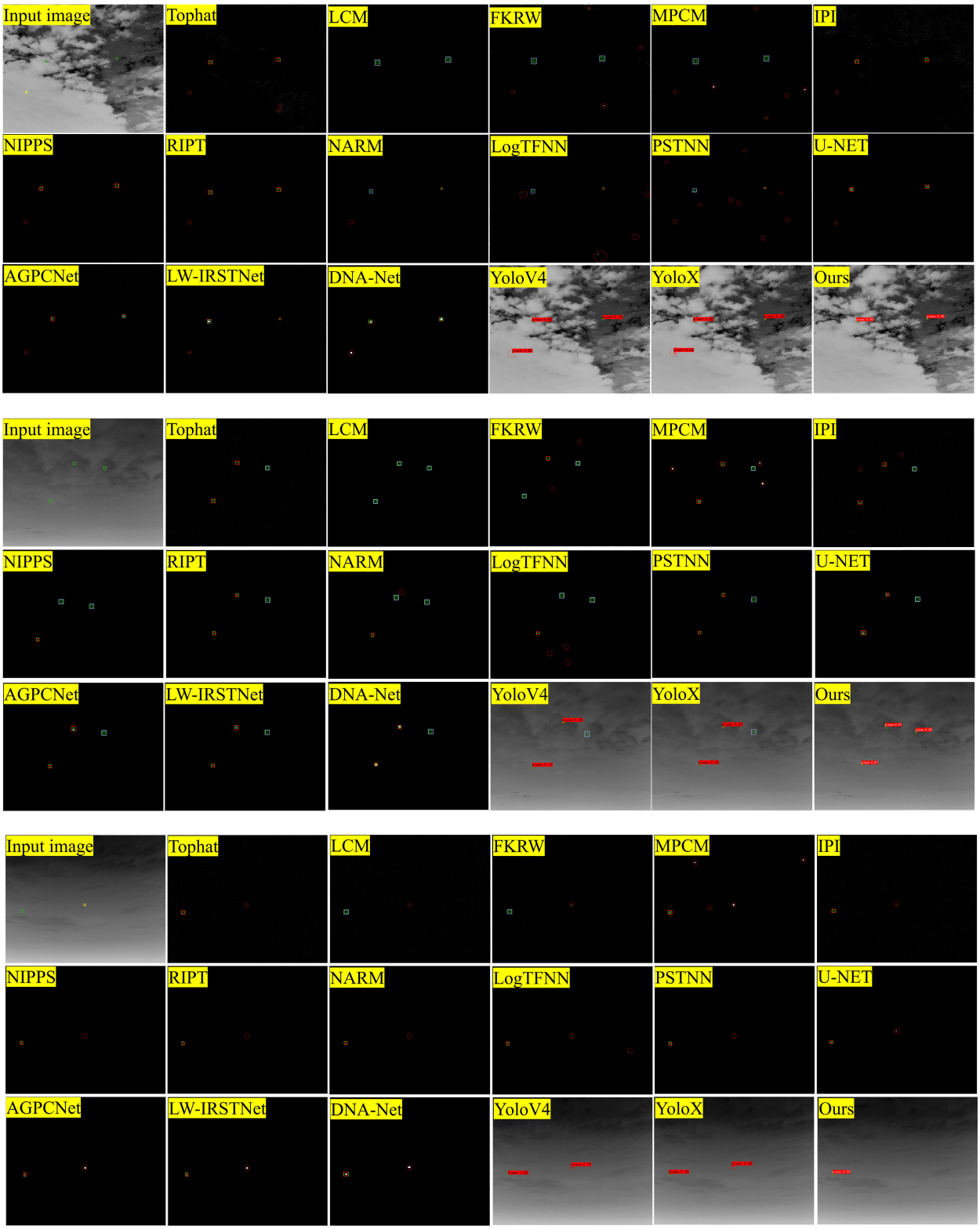

Fig. 6. Qualitative results obtained by different detection methods in the SAITD dataset. Detected targets, real targets, distractors, omissions, and false detection areas are highlighted with red solid wireframes, green solid wireframes, yellow solid wireframes, blue solid wireframes, and red dotted wire circles, respectively. The third row is the detection method of the YOLO series. The final detection graph has category and confidence score.

Compared with the comparison method, our method has almost no problems with missing and false detection. The results obtained by the traditional method are easy to lose relatively dim objects and be easily disturbed by the highlighted background and interferons, resulting in false detection. The U-NET, DNA-Net, AGPCNet, LW-IRSTNet, and the Yolo series method (YoloV4, YoloX) are much better than the traditional method and will not be interfered with by the

TABLE V

COMPARISON OF OUR METHOD WITH OTHER 23 METHODS ON THE SIRST DATASET, THE BEST RESULTS ARE SHOWN IN BOLD. AMONG THEM, THE METHODS WITH MARK † ARE THE RESULT OBTAINED BY OUR EXPERIMENT, AND THE RESULT WITHOUT MARK METHODS ARE OBTAINED FROM THE PUBLISHED ARTICLE [23]

Methods	Tophat	LCM	ILCM	FKRW	MPCM	IPI†
IoU	0.220	0.193	0.104	0.268	0.357	0.469
nIoU	0.352	0.207	0.123	0.339	0.445	0.597
Methods	NIPPS†	RIPT	PSTNN	FPN	GAU	U-NET†
IoU	0.478	0.146	0.605	0.720	0.701	0.733
nIoU	0.605	0.245	0.504	0.700	0.701	0.709
Methods	NRAM†	GST†	ECA-STT†	Yolov4†	LogTFNN†	DNA-NET†
IoU	0.304	0.451	0.371	0.749	0.451	0.769
nIoU	0.435	0.581	0.521	0.720	0.568	0.754
Methods	YoloX†	ACM	ALC-NET	AGPCNet†	LW-IRSTNet†	**ours**
IoU	0.708	0.743	0.757	0.776	0.754	**0.782**
nIoU	0.663	0.731	0.728	0.738	0.725	**0.765**

highlighted background, but will still be interfered with by the highlighted target, resulting in false detection. In addition, when detecting dim infrared small targets, it leads to missing detection. Our method can accurately detect bright targets and dim targets in the complex highlighted background with jamming objects and will not produce false detection of brighter jamming objects. This is because we make full use of the significant characteristics of infrared dim targets and make full use of global information, reducing the error caused by local contrast information.

To further demonstrate the effectiveness of our method, the proposed method was compared with 23 SOTA methods on the SIRST dataset. To highlight the effectiveness of our work, we compared the indicators in the comparison method, such as IoU and nIoU, and the results are shown in Table V. The results show that our model achieves the best performance among all the methods, demonstrating the effectiveness of the proposed architecture. Compared with the top three methods with the best results (AGPCNet, DNA-NET, and ALC-NET) of IoU and nIoU, our proposed method exceeds them in these two indexes, indicating that our proposed method has both a higher detection rate and a lower false positive rate and has achieved the best performance at present.

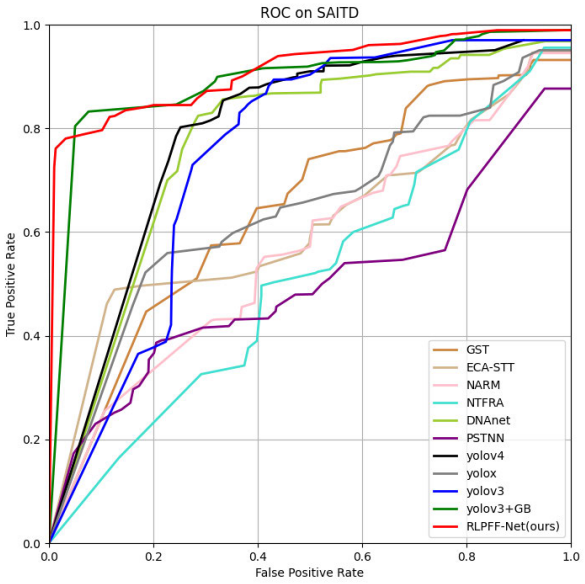

Fig. 7. ROC comparison of selected methods on dataset SAITD.

D. Ablation Study

1) Impact of GB: First, we add a GB module into the model network. It can be seen that the performance of the GB module is consistently and significantly better than that of the original model. The results are shown in Table VI that the global context attention module combined with boundary refinement is helpful to improve the accuracy of infrared dim small targets. This performance improvement can be explained in two ways. On one hand, boundary refinement makes the target boundary clear and enhances the boundary feature information, prevents the background from drowning the target, and improves detection accuracy. On the other hand, the global context information is used to enrich the infrared dim target feature information, and the high-level features of semantically accurate location are fused with the low-level features of detailed and lacking semantic information to produce a more precise representation.

2) Impact of Reinforcement Learning: Based on adding module GB, we added reinforcement learning. After adding reinforcement learning, we made corresponding parameter adjustments. We set the weight adjustment value of the agent to $\pm 1\%$, $\pm 3\%$, and $\pm 5\%$ each time. As can be seen from the results, $\pm 3\%$ we set has the best effect, and when it is set to $\pm 1\%$, the value of each adjustment is small. We set the agent to interact 20 times, and the optimal weight result is not reached during interaction. When it is set to $\pm 5\%$, the adjusted value is too large, leading to skipping the optimal weight and falling into the local optimal. The optimal result of adding reinforcement is much higher than that of only adding module GB, especially the accuracy is about three points higher, which indicates that the reinforcement learning module we added can make the model pay more attention to the characteristic information of the target. Therefore, we set the value of the adjusted weight hyperparameter $\mathscr{C}$ to 3. As shown in Table VI, compared with the baseline and the addition of module GB, finally, the reinforcement learning model is added to pay more

TABLE VI
ABLATION STUDY ON THE IMPACT OF THE GB MODULE AND PARAMETER ADJUSTMENT OF REINFORCEMENT LEARNING. THE BEST RESULTS ARE SHOWN IN BOLD

Methods	P(%)	R(%)	Acc(%)	AP(%)	F1	IoU	nIoU
Yolov3	88.54	88.91	76.43	84.78	0.881	0.842	0.863
Yolov3+GB	94.66	89.55	79.07	88.49	0.917	0.886	0.906
Yolov3+GB+RL($\pm 1\%$)	95.80	89.27	73.95	89.46	0.929	0.900	0.881
Yolov3+GB+RL($\pm 5\%$)	93.64	90.18	74.02	89.05	0.913	0.899	0.868
RLPGB-Net(ours)	**97.42**	**90.13**	**80.75**	**89.89**	**0.934**	**0.906**	**0.922**

attention to the target area, which shows that our work is very effective.

Finally, ROC curves of 11 methods (choosing the most representative of all the experiments) were compared, as shown in Fig. 7. It can be seen that our proposed RLPGB-Net has the best effect, followed by Yolov3+GB, which shows the effectiveness of our proposed pyramid fusion module combined with reinforcement learning module and the introduced GB module.

V. CONCLUSION

In this article, an infrared dim small target detection network combined with reinforcement learning is designed to detect remote dim small targets in the air. In our network, reinforcement learning gives corresponding weights to different levels of features, highlighting the significant features of infrared dim small targets. Therefore, high final detection accuracy means that our model has high accuracy, and each weight is automatically generated and adjusted by network learning, which has strong robustness. Second, we also introduce the GB module, which can make full use of the information of the context and reduce the error detection caused by local comparison and eliminate the problem that the target is too dim and drowned by the background. Even if the target is dim, it will be detected by our network, reducing the rate of missing detection. The experimental results show that compared with recent research methods, our model can obtain higher detection accuracy and achieve the most advanced performance in the task of infrared dim small target detection.

However, the experiments involved in the method proposed in this article are only infrared dim targets facing the air, without introducing other types of targets, such as ships in the sea and high-speed moving vehicles. So far, the diversification of detection targets cannot be satisfied. In future work, we will focus on the detection of diversified targets and realize correct classification, which can help our model realize the detection of all types of targets in infrared scenes and improve the generalization of the model's detection ability in different scenes.

REFERENCES

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

[2] R. C. Hall, *Missile Defense Alarm: The Genesis of Space-Based Infrared Early Warning*. Chantilly, VA, USA: National Reconnaissance Office, 1988.

[3] M. T. Eismann, C. R. Schwartz, J. N. Cederquist, J. A. Hackwell, and R. J. Huppi, "Comparison of infrared imaging hyperspectral sensors for military target detection applications," *Proc. SPIE*, vol. 2819, pp. 91–101, Nov. 1996.

[4] M. Malanowski and K. Kulpa, "Detection of moving targets with continuous-wave noise radar: Theory and measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3502–3509, Sep. 2012.

[5] Z. Wu, N. Fuller, D. Theriault, and M. Betke, "A thermal infrared video benchmark for visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 201–208.

[6] S. Kim and J. Lee, "Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track," *Pattern Recognit.*, vol. 45, no. 1, pp. 393–406, Jan. 2012.

[7] L. Deng, H. Zhu, C. Tao, and Y. Wei, "Infrared moving point target detection based on spatial–temporal local contrast filter," *Infr. Phys. Technol.*, vol. 76, pp. 168–173, May 2016.

[8] G.-Y. Wang, Z.-X. Chen, and Q.-L. Li, "A review of infrared weak and small targets detection under complicated background," *Infr. Technol.*, vol. 28, pp. 287–292, May 2006.

[9] H. Zhu, Y. Guan, L. Deng, Y. Li, and Y. Li, "Infrared moving point target detection based on an anisotropic spatial–temporal fourth-order diffusion filter," *Comput. Electr. Eng.*, vol. 68, pp. 550–556, May 2018.

[10] W. Zhang, M. Cong, and L. Wang, "Algorithms for optical weak small targets detection and tracking: Review," in *Proc. Int. Conf. Neural Netw. Signal Process.*, Dec. 2003, pp. 643–647.

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

[12] R. Kou et al., "Infrared small target segmentation networks: A survey," *Pattern Recognit.*, vol. 143, Nov. 2023, Art. no. 109788.

[13] J. Barnett, "Statistical analysis of median subtraction fifiltering with application to point target detection in infrared backgrounds," *Proc. SPIE*, vol. 1050, pp. 10–15, Jun. 1989.

[14] V. T. Tom, T. Peli, M. Leung, and J. E. Bondaryk, "Morphology-based algorithm for point target detection in infrared backgrounds," *Proc. SPIE*, vol. 1954, pp. 25–32, Oct. 1993.

[15] S. Deshpande, M. Er, and R. Venkateswarlu, "Max-mean and max-median fifilters for detection of small-targets," *Proc. SPIE*, vol. 3809, pp. 74–83, Oct. 1999.

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

[17] P. Chiranjeevi and S. Sengupta, "Detection of moving objects using multi-channel kernel fuzzy correlogram based background subtraction," *IEEE Trans. Cybern.*, vol. 44, no. 6, pp. 870–881, Jun. 2014.

[18] X. Bai, Z. Chen, Y. Zhang, Z. Liu, and Y. Lu, "Infrared ship target segmentation based on spatial information improved FCM," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3259–3271, Dec. 2016.

[19] B. Li et al., "Dense nested attention network for infrared small target detection," *IEEE Trans. Image Process.*, vol. 32, pp. 1745–1758, 2023.

[20] B. McIntosh, S. Venkataramanan, and A. Mahalanobis, "Infrared target detection in cluttered environments by maximization of a target to clutter ratio (TCR) metric using a convolutional neural network," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 1, pp. 485–496, Feb. 2021.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497*.

[22] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[23] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Asymmetric contextual modulation for infrared small target detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 949–958.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional net works for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent (MICCAI)*, Munich, Germany, 2015, pp. 234–241.

[25] X. Sun et al., "A dataset for small infrared moving target detection under clutter background," *China Sci. Data*, vol. 5, no. 6, p. 8, 2021.
[26] T. Soni, J. R. Zeidler, and W. H. Ku, "Performance evaluation of 2-D adaptive prediction filters for detection of small objects in image data," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 327–340, Jul. 1993.
[27] M. R. Azimi-Sadjadi and H. Pan, "Two-dimensional block diagonal LMS adaptive filtering," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2420–2429, Sep. 1994.
[28] Y. Cao, R. Liu, and J. Yang, "Small target detection using two-dimensional least mean square (TDLMS) filter based on neighborhood analysis," *Int. J. Infr. Millim. Waves*, vol. 29, no. 2, pp. 188–200, Feb. 2008.
[29] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, "Factor analysis, probabilistic principal component analysis, variational inference, and variational autoencoder: Tutorial and survey," 2021, *arXiv:2101.00734*.
[30] H. Deng, J. G. Liu, and Z. Chen, "Infrared small target detection based on modifified local entropy and EMD," *Chin. Opt. Lett.*, vol. 8, pp. 24–28, Jan. 2010.
[31] A. Toet and T. Wu, "Small maritime target detection through false color fusion," *Proc. SPIE*, vol. 6945, Apr. 2008, Art. no. 69453V.
[32] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 574–581, Jan. 2014.
[33] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2168–2172, Dec. 2014.
[34] Y. Wei, X. You, and H. Li, "Multiscale patch-based contrast measure for small infrared target detection," *Pattern Recognit.*, vol. 58, pp. 216–226, Oct. 2016.
[35] L. Zhang, L. Peng, T. Zhang, S. Cao, and Z. Peng, "Infrared small target detection via non-convex rank approximation minimization joint $\ell_{2,1}$ norm," *Remote Sens.*, vol. 10, no. 11, p. 1821, Nov. 2018.
[36] X. Wang, Z. Peng, D. Kong, P. Zhang, and Y. He, "Infrared dim target detection based on total variation regularization and principal component pursuit," *Image Vis. Comput.*, vol. 63, pp. 1–9, Jul. 2017.
[37] Y. Dai, Y. Wu, Y. Song, and J. Guo, "Non-negative infrared patch-image model: Robust target-background separation via partial sum minimization of singular values," *Infr. Phys. Technol.*, vol. 81, pp. 182–194, Mar. 2017.
[38] Y. Dai and Y. Wu, "Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3752–3767, Aug. 2017.
[39] L. Zhang and Z. Peng, "Infrared small target detection based on partial sum of the tensor nuclear norm," *Remote Sens.*, vol. 11, no. 4, p. 382, Feb. 2019.
[40] X. Kong, C. Yang, S. Cao, C. Li, and Z. Peng, "Infrared small target detection via nonconvex tensor fibered rank approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5000321.
[41] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Attentional local contrast networks for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9813–9824, Nov. 2021.
[42] T. Zhang, L. Li, S. Cao, T. Pu, and Z. Peng, "Attention-guided pyramid context networks for detecting infrared small target under complex background," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 4, pp. 4250–4261, Aug. 2023.
[43] R. Kou, C. Wang, Y. Yu, Z. Peng, F. Huang, and Q. Fu, "Infrared small target tracking algorithm via segmentation network and multi-strategy fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5612912.
[44] K. Wang, S. Du, C. Liu, and Z. Cao, "Interior attention-aware network for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5002013.
[45] M. Fan, S. Tian, K. Liu, J. Zhao, and Y. Li, "Infrared small target detection based on region proposal and CNN classifier," *Signal, Image Video Process.*, vol. 15, no. 8, pp. 1927–1936, Nov. 2021.
[46] P. Yang, L. Dong, and W. Xu, "Infrared small maritime target detection based on integrated target saliency measure," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2369–2386, 2021.
[47] H. Zhang, L. Zhang, D. Yuan, and H. Chen, "Infrared small target detection based on local intensity and gradient properties," *Infr. Phys. Technol.*, vol. 89, pp. 88–96, Mar. 2018.
[48] M. Ju, J. Luo, G. Liu, and H. Luo, "ISTDet: An efficient end-to-end neural network for infrared small target detection," *Infr. Phys. Technol.*, vol. 114, May 2021, Art. no. 103659.
[49] S. Liu, D. Huang, and Y. Wang, "Pay attention to them: Deep reinforcement learning-based cascade object detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 7, pp. 2544–2556, Jul. 2020.
[50] H. Yao, P. Dong, S. Cheng, and J. Yu, "Regional attention reinforcement learning for rapid object detection," *Comput. Electr. Eng.*, vol. 98, Mar. 2022, Art. no. 107747.
[51] M. Zhou et al., "ReinforceNet: A reinforcement learning embedded object detection framework with region selection network," *Neurocomputing*, vol. 443, pp. 369–379, Jul. 2021.
[52] S. Mathe, A. Pirinen, and C. Sminchisescu, "Reinforcement learning for visual object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2894–2902.
[53] C. Wang and S. Qin, "Approach for moving small target detection in infrared image sequence based on reinforcement learning," *J. Electron. Imag.*, vol. 25, no. 5, Oct. 2016, Art. no. 053032.
[54] A. Pirinen and C. Sminchisescu, "Deep reinforcement learning of region proposal networks for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6945–6954.
[55] K. Fu et al., "A ship rotation detection model in remote sensing images based on feature fusion pyramid network and deep reinforcement learning," *Remote Sens.*, vol. 10, no. 12, p. 1922, Nov. 2018.
[56] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters—Improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1743–1751.
[57] Y. Dai, X. Li, F. Zhou, Y. Qian, Y. Chen, and J. Yang, "One-stage cascade refinement networks for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5000917.
[58] Y. Qin, L. Bruzzone, C. Gao, and B. Li, "Infrared small target detection based on facet kernel and random Walker," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7104–7118, Sep. 2019.
[59] X. Bai and F. Zhou, "Analysis of new top-hat transformation and the application for infrared dim small target detection," *Pattern Recognit.*, vol. 43, no. 6, pp. 2145–2156, Jun. 2010.
[60] C. Q. Gao, J. W. Tian, and P. Wang, "Generalised structure tensor based infrared small target detection," *Electron. Lett.*, vol. 44, no. 23, pp. 1349–1351, 2008.
[61] P. Zhang, L. Zhang, X. Wang, F. Shen, T. Pu, and C. Fei, "Edge and corner awareness-based spatial–temporal tensor model for infrared small-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10708–10724, Dec. 2021.
[62] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[63] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.
[64] R. Kou, C. Wang, F. Huang, Y. Yu, Z. Peng, and Q. Fu, "LW-IRSTNet: Lightweight infrared small target segmentation network," *TechRxiv. Preprint*, Mar. 2023. [Online]. Available: https://www.researchgate.net/publication/369350779_LW-IRSTNet_Lightweight_Infrared_Small_Target_Segmentation_Network, doi: 10.36227/techrxiv.22280995.v3.
[65] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
[66] H. C. Li, P. F. Xiong, J. An, and L. X. Wang, "Pyramid attention network for semantic segmentation," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2018, pp. 1–13.

Zhe Wang received the B.Sc. and Ph.D. degrees from the Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2003 and 2008, respectively.

Currently, he is a Full Professor with the Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai, China. At present, he has more than 40 articles as the first or corresponding author published in some famous international journals including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING, *Pattern Recognition* etc. His research interests include feature extraction, kernel-based methods, image processing, and pattern recognition.

Tao Zang received the B.Sc. degree from the College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, China, in 2017, where he is currently pursuing the M.S. degree in computer science with the East China University of Science and Technology, Shanghai, China.

His research interests include pattern recognition, machine learning, and object detection.

Zhiling Fu is currently pursuing the Ph.D. degree in computer science with the East China University of Science and Technology, Shanghai, China.

His research interests include pattern recognition, machine learning, incremental learning, and representation learning.

Hai Yang received the B.Sc. degree in software engineering from Xi'an Jiaotong University, Xi'an, China, in 2008, and the Ph.D. degree in signal and information processing from Chinese Academy of Sciences University, Beijing, China, in 2013.

Currently, he is a Research Associate Professor with the Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai, China. At present, he has more than ten articles published in many famous international journals, including bioinformatics, nature neuroscience, etc. His research interests include artificial intelligence, machine learning, big data, and bioinformatics.

Wenli Du received the B.S. and M.S. degrees in chemical process control from the Dalian University of Technology, Dalian, China, in 1997 and 2000, respectively, and the Ph.D. degree in control theory and control engineering from the East China University of Science and Technology, Shanghai, China, in 2005.

She is currently a Professor and the Dean of the College of Information Science and Engineering and Vice Dean of the Key Laboratory of Advanced Control and Optimization for chemical process, Ministry of Education, East China University of Science and Technology, China. Her research interests include control theory and application, system modeling, advanced control, and process optimization.