IRPruneDet: Efficient Infrared Small Target Detection via Wavelet Structure-Regularized Soft Channel Pruning

Mingjin Zhang[1], Handi Yang[1*], Jie Guo[1], Yunsong Li[1], Xinbo Gao[1], Jing Zhang[2*]
[1]Xidian University
[2]The University of Sydney
mjinzhang@xidian.edu.cn, 22011210777@stu.xidian.edu.cn, jguo@mail.xidian.edu.cn
ysli@mail.xidian.edu.cn, xbgao@mail.xidian.edu.cn, jing.zhang1@sydney.edu.au

Abstract

Infrared Small Target Detection (IRSTD) refers to detecting faint targets in infrared images, which has achieved notable progress with the advent of deep learning. However, the drive for improved detection accuracy has led to larger, intricate models with redundant parameters, causing storage and computation inefficiencies. In this pioneering study, we introduce the concept of utilizing network pruning to enhance the efficiency of IRSTD. Due to the challenge posed by low signal-to-noise ratios and the absence of detailed semantic information in infrared images, directly applying existing pruning techniques yields suboptimal performance. To address this, we propose a novel wavelet structure-regularized soft channel pruning method, giving rise to the efficient IRPruneDet model. Our approach involves representing the weight matrix in the wavelet domain and formulating a wavelet channel pruning strategy. We incorporate wavelet regularization to induce structural sparsity without incurring extra memory usage. Moreover, we design a soft channel reconstruction method that preserves important target information against premature pruning, thereby ensuring an optimal sparse structure while maintaining overall sparsity. Through extensive experiments on two widely-used benchmarks, our IRPruneDet method surpasses established techniques in both model complexity and accuracy. Specifically, when employing U-net as the baseline network, IRPruneDet achieves a 64.13% reduction in parameters and a 51.19% decrease in FLOPS, while improving IoU from 73.31% to 75.12% and nIoU from 70.92% to 74.30%. The code is available at https://github.com/hd0013/IRPruneDet.

Introduction

Single frame infrared small target (SIRST) detection plays an irreplaceable role in many practical applications, such as traffic management and maritime rescue (Cuccurullo et al. 2012; Law et al. 2016; Zhang and Tao 2020). When dealing with target detection tasks (Zou et al. 2023) in visible images, challenges arise under conditions of weak illumination and occlusion. In contrast, infrared images excel at capturing target information due to their penetrating infrared thermal radiation. Nevertheless, SIRST comes with stringent criteria (Chapple et al. 1999): target size below 0.15% of the total image, contrast ratio under 15%, and signal-to-noise ratio (SNR) below 1.5. Consequently, overcoming these obstacles involving small targets, noise, clutter, and object interference has sparked significant research interest in infrared small target detection (IRSTD) in recent years.

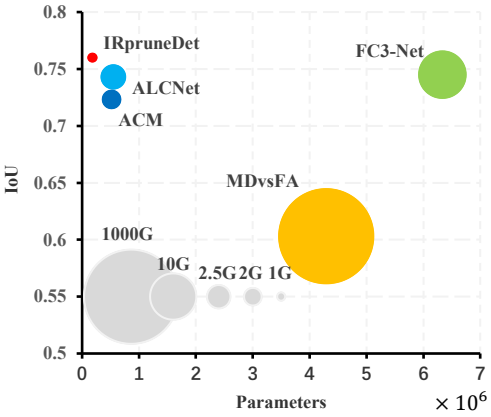

Figure 1: Comparison between the proposed IRPruneDet and other deep learning-based models on the NUAA-SIRST dataset. The area of the gray circles denotes the number of FLOPs. IRPruneDet achieves the highest IoU while maintaining the lowest parameters and FLOPs.

To cope with the above difficulties in IRSTD, traditional methods (Dai and Wu 2017; Han et al. 2020) usually employ filtering techniques to filter out background interference or image enhancement methods to enhance targets. However, these methods heavily rely on hyper-parameter tuning and exhibit certain limitations when confront with complex scenes characterized by variations in illumination, complex backgrounds, and target occlusion. In light of the rapid advancements in deep learning, an increasing number of deep convolutional neural network (CNN)-based models have demonstrated superior performance in IRSTD. For instance, the pioneering implementation of deep CNN for IRSTD can be attributed to the miss detection vs. false alarm (MDvsFA) model (Wang, Zhou, and Wang 2019). It employs two generative adversarial networks (GANs) (Goodfellow et al. 2020) to separately reduce MD and FA while re-

*Corresponding Author.

quiring a considerable number of computations (see Fig. 1). Dai *et al*. achieve significant advancements over MDvsFA by replacing GANs with a U-net in asymmetric context modulation (ACM) approach (Dai et al. 2021a). Furthermore, they propose an attentional local contrast network (ALCNet) (Dai et al. 2021b) to effectively combine discriminative and model-driven methods by increasing the network size. By propagating the target features to deeper layers of the network, Zhang *et al*. (Zhang et al. 2022b) present a feature compensation and cross-level correlation network (FC3-Net), which achieves superior detection performance while having much more parameters and computations than ALCNet and ACM. As the complexity of a model increases, the number of model parameters and computations grows significantly, leading to inefficiencies in storage, memory, and computation. Directly deploying them on platforms with limited resources is impractical. Therefore, there exists an imperative demand to explore a lightweight network architecture for efficient IRSTD.

Recently, model compressing methods have been proposed to devise lightweight networks for various tasks (Han et al. 2015; Rastegari et al. 2016; Denton et al. 2014; Hinton, Vinyals, and Dean 2015). Among them, structured pruning (He et al. 2019b) has garnered recognition for its ability to achieve practical storage space savings and inference acceleration on general-purpose hardware. This approach can prune redundant filters in convolutional layers. Nevertheless, the existing pruning methods encounter challenges that hinder their direct applicability to CNN-based IRSTD models. **(1)** The conventional criteria used to evaluate channel importance are not applicable to the IRSTD task. Currently, pruning methods predominantly rely on criteria that assess the informative importance within channels, assuming that channels with greater magnitude are more vital (Huang et al. 2021). However, due to the low SNR in infrared images, channels containing background noise and clutter often exhibit higher magnitudes. Consequently, relying on conventional criteria may erroneously prune important channels that have low magnitudes but contain crucial information about small targets, leading to the discarding of important channels. **(2)** During the iterative channel pruning process, certain channels that contain important information may be pruned prematurely and deactivated permanently. In the IRSTD task, the targets typically have small sizes, resulting in a limited number of channels carrying important information. The erroneous pruning of critical channels can lead to a drastic decrease in detection accuracy.

In this study, we introduce the concept of utilizing network pruning to enhance the efficiency of IRSTD for the first time. Specifically, we propose a novel wavelet structure-regularized soft channel pruning method, resulting in the efficient IRPruneDet model (see Fig. 2). Firstly, we design a wavelet channel pruning (WCP) strategy based on the wavelet-based sparse constraint. Through wavelet analysis of convolutional layers, weight matrices are decomposed into low and high-frequency components. By applying l_1-norm regularization to these wavelet coefficients, we assess channel importance according to their magnitude. To manage memory, we propose a memory-efficient wavelet-based pruning criterion within an energy minimization framework, treating the wavelet transform of weight matrices akin to their differential operators. Additionally, we avoid premature pruning of channels holding crucial SIRST information by implementing a soft channel reconstruction (SCR) method. This involves dynamically retaining parameters of convolutional layers with the highest detection accuracy during pruning. For soft channel reconstruction, we combine channel reconstruction with the pruning process, randomly interpolating between recovered pruned channel parameters and those associated with optimal performance. Our method assesses the importance of all channels to obtain desired network structure while adhering to sparsity constraints. Experiments on two widely-used benchmarks demonstrate that IRPruneDet outperforms existing methods in detection accuracy, while significantly reducing FLOPs and parameters.

In summary, the contribution of this study is three-fold. **(1)** We propose an efficient IRPruneDet model for IRSTD. To the best of our knowledge, IRPruneDet is the first attempt to design a lightweight network architecture tailored for the IRSTD task via network pruning. Using U-net18 as the baseline network, IRPruneDet reduces 64.13% parameters and 51.19% FLOPS while improving IoU from 73.31% to 75.12% and nIoU from 70.92% to 74.30% on the NUAA-SIRST dataset. **(2)** We design a WCP strategy by representing the weight matrix in the wavelet domain. To encourage structural sparsity without imposing additional memory requirements, we design and incorporate a novel wavelet regularization penalty into the network. **(3)** We develop an SCR method for the pruning process. It can mitigate the risk of prematurely and incorrectly pruning channels that carry critical information, thereby ensuring the preservation of important network features throughout the pruning procedure.

Related Work

Infrared Small Target Detection

The IRSTD algorithms can be categorized into traditional and deep learning-based methods. Traditional techniques focus on extracting distinctive features in infrared (IR) images. These methods encompass filter-based methods like the max-median filter (Deshpande et al. 1999) and top-hat filter (Bai and Zhou 2010), low-rank methods such as weighted strengthened local contrast measure (WSLCM) (Han et al. 2020) and tri-layer local contrast measure (TTLCM) (Chen et al. 2013), along with HVS-based methods such as infrared patch-image (IPI) (Gao et al. 2013), non-convex rank approximation minimization (NARM) (Zhang et al. 2018), and the partial sum of the tensor nuclear norm (PSTNN) (Zhang and Peng 2019). With the advent of deep learning, CNN-based techniques (Zhang et al. 2022a; Li et al. 2022a; McIntosh, Venkataramanan, and Mahalanobis 2020; Zhang et al. 2023) are introduced into the IRSTD task. For instance, MDvsFA (Wang, Zhou, and Wang 2019) eschews the traditional approach of relying on a single goal to jointly reduce MD and FA by decomposing it into two subtasks with two GANs (Goodfellow et al. 2020). To preserve feature information, Dai *et al*. (Dai et al. 2021a) propose an ACM model with global context feedback and a modulation path using

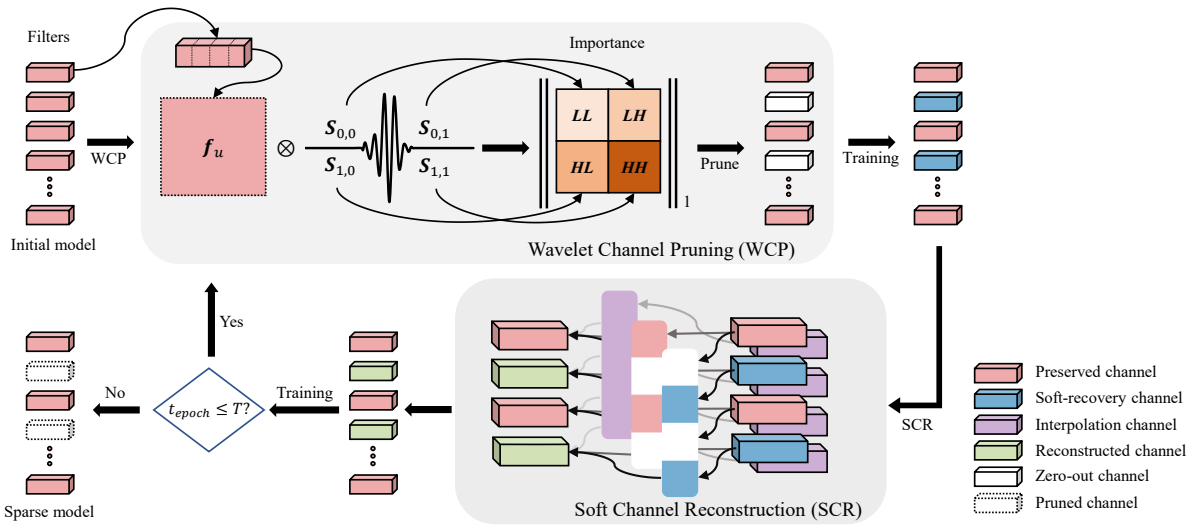

Figure 2: Illustration of the proposed IRPruneDet method. The pruning process of a specific convolutional layer is used to illustrate the dynamic iterative process of IRPruneDet, which includes wavelet channel pruning (WCP), training, soft channel reconstruction (SCR), and final hard channel pruning to obtain a sparse model. WCP assesses channel importance based on the $l1$-norm of the wavelet decomposition coefficients obtained by convolving the weight matrix with the Haar wavelet.

pointwise channel attention to exchange high-level semantics and low-level details. Furthermore, Dai *et al*. (Dai et al. 2021b) introduce a feature map cyclic shifting scheme and present an ALCNet with increased network size. Zhang *et al*. (Zhang et al. 2022b) develop an even larger network FC3-Net. However, these methods enhance small IR target detection by scaling up network size to increase model capacity and extract semantic features. This strategy often leads to increased model size, memory footprint, and computations.

Neural Network Pruning

Pruning (Han et al. 2015) removes unimportant structures in the network to produce a sparse and efficient model. Channel pruning (Li et al. 2016), a subset of this technique, falls into two categories based on channel status after pruning: hard and soft pruning. Hard pruning permanently deactivates channels identified by specific criteria (Sui et al. 2021; He et al. 2019b; Liu et al. 2017; Wang, Li, and Wang 2021; Tang et al. 2020; He et al. 2021). For instance, Li *et al*. (Li et al. 2016) introduce a pruning filter for efficient convNets (PEEC), which calculates channel importance via $l1$-norm (Li et al. 2016). HRank (Lin et al. 2020) suggests evaluating channel importance using the rank of convolutional layer weights. In contrast, soft pruning involves dynamic channel pruning without permanent discarding (He et al. 2019a; Guo et al. 2020; Lin et al. 2019; Ding et al. 2019; He et al. 2022). Channels' weights are approximated to 0, permitting their participation in future training and pruning iterations. For example, soft filter pruning (SFP) (He et al. 2018) generates masks based on channel norms over time, allowing updates in subsequent phases. Operation-aware soft channel pruning (SCP) (Kang and Han 2020) formulates discrete masks to differentiable forms for joint learning of model parameters and dynamic masks. To the best of our knowledge, no prior research has explored network pruning within the context of IRSTD. Our goal is to bridge this gap by utilizing the idea of network pruning to develop an efficient IRSTD model.

Methodology

Preliminaries

Given an infrared image $\boldsymbol{X}_{IR}$, the IRSTD problem based on deep learning can be formulated:

$$\boldsymbol{Y}_{IR} = f_{det}\left(\boldsymbol{X}_{IR}; \boldsymbol{\Theta}\right), \tag{1}$$

where f_{det} is a trainable deep neural network, $\boldsymbol{\Theta}$ represents the model parameters, and Y_{IR} denotes the segmentation mask of targets in the infrared image. Without loss of generality, it is assumed that there exist L layers of parameters, where the l_{th} convolution layer can be parameterized by $\left\{\boldsymbol{W}^{(l)} \in \mathbb{R}^{C^l \times C^{l-1} \times K \times K}, 1 \leq l \leq L\right\}$. Here, $\boldsymbol{W}^{(l)}$, C^l, C^{l-1}, K represent the learnable weight matrix (*i.e.*, model parameters), the number of output channels, the number of input channels, and the kernel size of the i_{th} convolutional layer, respectively. During the process of channel pruning, we can conceptualize the above model parameters $\boldsymbol{W}^{(l)}$ as a series of filters $\boldsymbol{F}^{(l)}$, which can be represented as a set of $\left\{\boldsymbol{F}_j^{(l)} \in \mathbb{R}^{C^{l-1} \times K \times K}, 1 \leq j \leq C^l, 1 \leq l \leq L\right\}$.

Wavelet Channel Pruning

In the IRSTD task, the detection accuracy is often reduced due to the loss of edge information of the targets. Accordingly, while pruning the CNN-based network, it's essential to preserve channels that contribute more to the edge information, ensuring that the network's performance remains in-

tact despite the reduction in model size. In the image processing domain, wavelet-based analysis (Mallat 1989) has been widely-used to decompose one image into wavelet coefficients containing low-frequency and high-frequency information. In the context of IRSTD, our investigation suggests that channels with higher high-frequency coefficients, as transformed by the wavelet framework, possess more edge information of the targets. Furthermore, the l_p-norm of the wavelet decomposition coefficients of a channel, computed for each convolutional layer, can effectively indicate the significance of the channel in terms of containing the edge information of the targets. To this end, we propose to regularize the l_1-norm of the wavelet decomposition coefficients, which can promote sparsity by pushing the weights of unimportant channels to zero during training. It can be formulated as follows:

$$\min_{\boldsymbol{F}} \frac{1}{N}\sum_{i=1}^{N} L\left(\boldsymbol{Y}_i, f\left(\boldsymbol{X}_i, \boldsymbol{F}\right)\right) + \lambda \sum_{i=1}^{N} \left\|\boldsymbol{H}\boldsymbol{F}\right\|_1, \quad (2)$$

where $L(\cdot)$ denotes the loss function. $f(\cdot)$, $\boldsymbol{X}_i$, $\boldsymbol{Y}_i$ are the prediction, input, and the ground truth label, respectively. $\boldsymbol{H}$ represents the two-dimensional discrete wavelet transform (DWT). λ is a hyper-parameters to balance the two loss terms. However, to apply the wavelet-based coefficient regularization penalty to the network, additional memory is required to store the results of wavelet transform for each filter $\boldsymbol{F}$ in the network. To address this issue, we propose to approximate the DWT as a differential operator.

Specifically, we adopt the two-dimensional Haar tight frame system (Chui 1992). When processing $\boldsymbol{F}$, particularly $\boldsymbol{F}_j^{(l)} \in \mathbb{R}^{C^{l-1}\times K\times K}$, we convert it to a 2D shape $\boldsymbol{f}_u \in \mathbb{R}^{M\times M}$ by tiling the $K \times K$ filters (*i.e.*, if $C^{l-1} = p \times p$, then $M = p * k$). The Haar basis is defined as: $\boldsymbol{S}_{0,0} = \frac{1}{4}[1,1;1,1]$, $\boldsymbol{S}_{0,1} = \frac{1}{4}[1,-1;1,-1]$, $\boldsymbol{S}_{1,0} = \frac{1}{4}[1,1;-1,-1]$, $\boldsymbol{S}_{1,1} = \frac{1}{4}[1,-1;-1,1]$. Denoting Ω as a domain in the two-dimensional real space $\mathbb{R}^2$, there exists $\boldsymbol{u} \in L_2(\Omega)$ (Li et al. 2022b), which is sufficiently smooth and related to $\boldsymbol{f}_u$, *i.e.*,

$$\left(\boldsymbol{f}_u\right)[i,j] = \boldsymbol{u}\left(x_i, y_j\right), s.t.\left(x_i, y_j\right) = (ih, jh), \quad and\ 0 \le i, j \le N, \quad (3)$$

where h is the reciprocal of N. Consequently, the regularization penalty term for a certain filter $\boldsymbol{F}$ in the network can be expressed as:

$$\left\|\boldsymbol{H}\boldsymbol{F}\right\|_1 = h^2 \sum_{i,j}((\frac{2}{h})^2(|(\boldsymbol{S}_{0,1}[-.] \otimes \boldsymbol{f}_u)[i,j]|^2 + |(\boldsymbol{S}_{1,0}[-.] \otimes \boldsymbol{f}_u)[i,j]|^2))^{1/2}. \quad (4)$$

The Taylor expansion of $\boldsymbol{u}$ at point (x_i, y_j) is given by:

$$\frac{2}{h}\left(\boldsymbol{S}_{0,1}[-.] \otimes \boldsymbol{f}_u\right)[i,j] = \frac{1}{2h}\left(\boldsymbol{u}\left(x_i, y_j\right) - \boldsymbol{u}\left(x_i - h, y_j\right)\right) + \frac{1}{2h}\left(\boldsymbol{u}\left(x_i, y_j - h\right) - \boldsymbol{u}\left(x_i - h, y_j - h\right)\right), \quad (5)$$

$$\frac{2}{h}\left(\boldsymbol{S}_{1,0}[-.] \otimes \boldsymbol{f}_u\right)[i,j] = \frac{1}{2h}\left(\boldsymbol{u}\left(x_i, y_j\right) - \boldsymbol{u}\left(x_i, y_j - h\right)\right) + \frac{1}{2h}\left(\boldsymbol{u}\left(x_i - h, y_j\right) - \boldsymbol{u}\left(x_i - h, y_j - h\right)\right). \quad (6)$$

In the case where h approaches 0, we can derive the following equation from the concept of partial derivatives:

$$\frac{2}{h}\left(\boldsymbol{S}_{0,1}[-.] \otimes \boldsymbol{f}_u\right)[i,j] + \frac{2}{h}\left(\boldsymbol{S}_{1,0}[-.] \otimes \boldsymbol{f}_u\right)[i,j] = \boldsymbol{u}_x\left(x_i, y_j\right) + \boldsymbol{u}_y\left(x_i, y_j\right). \quad (7)$$

Based on Eq. (4), we have:

$$\left\|\boldsymbol{H}\boldsymbol{F}\right\|_1 = \int_\Omega \sqrt{\boldsymbol{u}_x^2 + \boldsymbol{u}_y^2} dxdy. \quad (8)$$

Thus, we can combine the wavelet-based sparse regularization term in Eq. (2) with the gradient of the weight matrix itself, reducing the memory overhead while efficiently learning sparse network structures. Note that the existence of $\int_\Omega |\nabla \boldsymbol{u}|\, dxdy = \int_\Omega \sqrt{\boldsymbol{u}_x^2 + \boldsymbol{u}_y^2} dxdy$ is due to the sufficient smoothness of $\boldsymbol{u}$. Hence, we can establish the connection between WCR and the differential operator of weight matrices within the framework of energy functional minimization:

$$\min_{\boldsymbol{F}} \frac{1}{N}\sum_{i=1}^{N} L\left(\boldsymbol{Y}_i, f\left(\boldsymbol{X}_i, \boldsymbol{F}\right)\right) + \lambda \sum_{i=1}^{N} \left\|\boldsymbol{D}\boldsymbol{F}\right\|_1, \quad (9)$$

where $\boldsymbol{D}$ is the first-order differential operator of the channel weight matrix. Since we apply regularization to the channel weight matrix after wavelet decomposition using the gradient of the channel weight matrix itself, no extra memory footprint is needed to store the results of the wavelet transform. Meanwhile, wavelet analysis retains channels with crucial edge information of targets, striking a balance between detection accuracy and model compression efficiency.

Soft Channel Reconstruction

There is a challenge in existing iterative channel pruning approaches (Guo, Yao, and Chen 2016), *i.e.*, valuable information-rich channels can be prematurely discarded, resulting in reduced accuracy and generalization of the pruned model. This issue gains prominence when pruning channels in IRSTD models, given the small target size, which often implies only a few crucial channels hold vital information. Prematurely discarding these key channels during pruning substantially diminishes the pruned model's accuracy.

While dynamic soft channel pruning methods (He et al. 2018) maintain channel vitality by not entirely discarding pruned channels, they lack the assurance of finding the global optimal solution, leading to important channels becoming inactive. In the iterative pruning and training process of IRSTD, we observed that the channels retained as active after each pruning iteration can be further trained. These channels offer valuable information until the next pruning round. Unfortunately, due to short intervals between pruning rounds or insufficient activation, these channels are frequently pruned again, causing the iterative process to stagnate. To address this challenge, we propose the SCR method, which dynamically saves model parameters corresponding to the best detection accuracy throughout the pruning process. Prior to each pruning stage, SCR reconstructs previously pruned channels in a random manner, effectively reactivating them. The channels subjected to SCR can be expressed as follows:

$$\boldsymbol{F}_{SCR} = \alpha \boldsymbol{F}_{best} + (1-\alpha)\boldsymbol{F}_{SC}, \quad (10)$$

where $\boldsymbol{F}_{best}$ is the previously best channel parameters. $\boldsymbol{F}_{SC}$ represents pruned channel parameters. In SCR, the channel $\boldsymbol{F}_{SCR}$ after channel soft reconstruction can be regarded as cosine annealing interpolation between $\boldsymbol{F}_{best}$ and $\boldsymbol{F}_{SC}$. In this sense, α can be expressed as:

$$\alpha = \frac{1}{2}\left(1+\cos\left(\left(1-\frac{\Delta\beta_{SCR}}{\pi}\right)\frac{t_{epoch}\pi}{T}+\Delta\beta_{SCR}\right)\right), \quad (11)$$

where T denotes the total number of training iterations for SCR. $\Delta\beta_{SCR}$ represents the interpolation control factor of SCR. By changing the value of $\Delta\beta_{SCR}$, the initial value of α can be empirically controlled between 0 and 1, and it decays to 0 during the SCR training process. This explains that in the early stages of SCR, we believe that $\boldsymbol{F}_{best}$ needs to have a certain importance in channel recovery. As the model training and pruning progress, the sparse model gradually converges, and the current model parameters play an increasingly important role in the SCR process. Thus, the strong activation effect of SCR on pruned channels is most pronounced in the early stages of pruning, and its effectiveness gradually decreases as the sparse model converges.

Furthermore, it should be noted that the SCR strategy is not applied to every pruned channel, as doing so may result in certain channels with low scores being constantly suppressed under the WCP criterion. To promote the diversity of the channels, SCR randomly selects a portion of the pruned channels for channel reconstruction each time and also applies a cosine decay to the ratio of channels to be reconstructed, gradually reducing it as the model converges. The dynamic channel reconstruction rate in SCR can be represented as: $\beta = \frac{\beta_0}{2}\left(1+\cos\left(\frac{t_{epoch}\pi}{T}\right)\right)$, where β_0 is initial channel reconstruction rate. The scale of random channel reconstruction under a given global sparsity constraint can be controlled by adjusting β_0, so as to find the optimal sparse model by using SCR.

Experiments

Experimental Details

Dataset. We adopt the NUAA-SIRST (Dai et al. 2021a) and IRSTD-1k (Zhang et al. 2022c) datasets for evaluation. NUAA-SIRST consists of 427 infrared images with a total of 480 infrared targets. IRSTD-1k comprises 1,001 infrared images, encompassing various target categories. For each dataset, we divide IR images into three disjoint subsets: 50% for training, 30% for validation, and 20% for testing.

Evaluation Metrics. The evaluation metrics can be categorized into two types: objective detection accuracy-based metrics and model complexity-based metrics. The former includes pixel-level metrics such as Intersection over Union (IoU) and Normalized IoU (nIoU) (Dai et al. 2021a), and object-level metrics such as Probability of Detection (P_d) and False-Alarm Rate (F_a) (Li et al. 2022a). The latter consists of the number of FLOPs and model parameters.

Implementation Details. We resize the size of each IR image in NUAA-SIRST and IRSTD-1k datasets to 512×512, following the common practice (Zhang et al. 2022a; Dai et al. 2021b). For the pruning and training process, we utilize AdaGrad as the optimizer with a learning rate of 0.01. The training process lasts for 500 epochs with a weight decay of 10^{-4} and a batch size of 16. By default, we set $\Delta\beta_{SCR}$ to 0.5π and β_0 to 1. We apply IRPruneDet only to a U-Net18 baseline model and we compare it with representative CNN-based methods: MDvsFA (Wang, Zhou, and Wang 2019), ACM (Dai et al. 2021a), ALCNet (Dai et al. 2021b), and FC3-Net (Zhang et al. 2022b), and traditional methods: Top-Hat (Bai and Zhou 2010), Max-Median (Deshpande et al. 1999), WSLCM (Han et al. 2020), TLLCM (Chen et al. 2013), IPI, NRAM (Zhang et al. 2018), RIPT (Dai and Wu 2017), PSTNN (Zhang and Peng 2019), and MSLSTIPT (Sun, Yang, and An 2020).

Method	NUAA-SIRST				IRSTD-1k			
	Pixel-Level		Object-Level		Pixel-Level		Object-Level	
	IoU↑	nIoU↑	Pd↑	Fa↓	IoU↑	nIoU↑	Pd↑	Fa↓
Top-Hat	1.508	3.084	79.74	16456	10.06	7.438	75.11	1432
Max-Median	6.022	25.35	84.34	774.3	6.998	3.051	65.21	59.73
WSLCM	6.393	28.31	88.74	4462	3.452	0.678	72.44	6619
TLLCM	4.240	12.09	88.37	6243	3.311	0.784	77.39	6738
IPI	1.09	50.23	87.05	30467	27.92	20.46	81.37	16.18
NRAM	13.54	18.95	60.04	25.23	15.25	9.899	70.68	16.93
RIPT	16.79	20.65	69.76	59.33	14.11	8.093	77.55	28.31
PSTNN	30.30	33.67	72.80	48.99	24.57	17.93	71.99	35.26
MSLSTIPT	1.080	0.814	0.052	8.183	11.43	5.932	79.03	1524
IRPruneDet	**75.12**	**74.30**	**98.61**	**2.96**	**64.54**	**62.71**	**91.74**	**16.04**

Table 1: Comparison with traditional methods on NUAA-SIRST and IRSTD-1k in terms of $IoU(\%)$, $nIoU(\%)$, $P_d(\%)$, and $F_a(10^{-6})$.

Method	IoU↑	nIoU↑	Pd↑	Fa↓	FLOPs↓	Params↓
U-Net	73.31	70.92	96.15	39.87	1.922	0.5023
+$l1$-norm	67.18	67.84	86.55	27.34	0.957	0.1887
+WCP	74.25	73.59	96.37	8.94	0.938	0.1802
+$l1$-norm+SCR	73.51	72.01	93.06	16.45	0.957	0.1887
+WCP+SCR	75.12	74.30	98.61	2.96	0.938	0.1802

Table 2: Ablation study of IRPruneDet.

Ablation Study

To investigate the effectiveness of each component in IRPruneDet, we conduct ablation studies on the NUAA-SIRST dataset. Table 2 shows the results. **(1) Impact of WCP.** We apply the commonly used $l1$-norm criterion instead of the WCP strategy to measure the importance of channels. The comparative experiments demonstrate that pruning sparse models with the $l1$-norm criterion can obtain erroneous pruning of channels that encode low but important features, leading to a decrease in IoU and nIoU despite the compressed model size. **(2) Impact of SCR.** We control the usage of SCR with the same pruning criteria. From the exper-

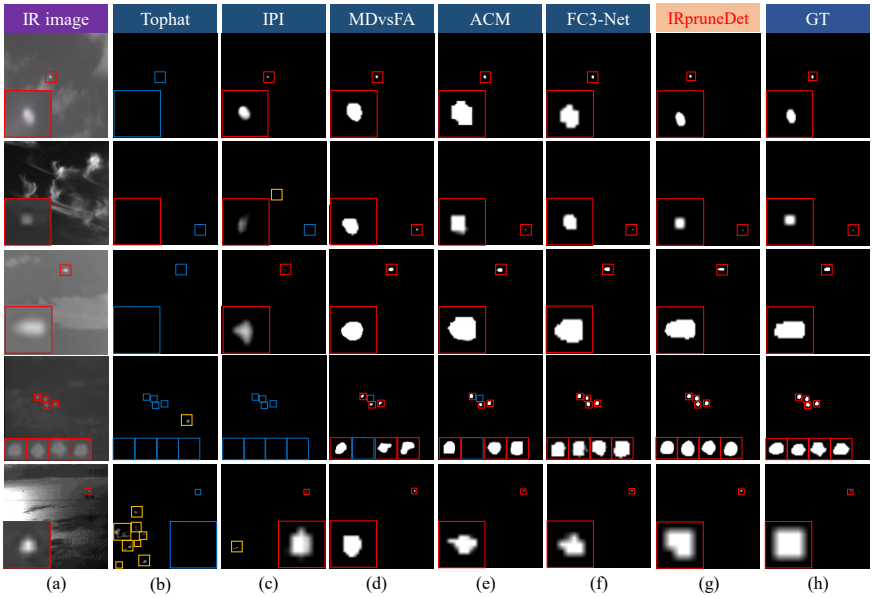

Figure 3: Visual results of different IRSTD methods. The boxes in red, yellow, and blue represent correct, missed, and false detections, respectively. Close-up views are shown in the bottom corners.

imental results, it is observed that SCR effectively recovers channels that were erroneously pruned early on. This prevents the erroneous pruning of channels containing important information, thereby enhancing the model's accuracy.

U-Net	IoU↑	nIoU↑	Pd↑	Fa↓	FLOPs↓	Params↓
+SFP($l1$-norm)	67.18	67.84	86.55	27.34	0.957	0.1887
+SFP($l2$-norm)	70.71	69.64	90.14	23.55	0.957	0.1887
+FPGM	71.54	70.80	91.47	33.64	1.022	0.1806
+ASFP($l1$-norm)	70.82	70.88	90.93	26.33	0.957	0.1887
+ASFP($l2$-norm)	71.86	71.9	91.83	21.42	0.957	0.1887
+IRPrune	75.12	74.30	98.61	2.96	0.938	0.1802

Table 3: Comparison with other pruning methods.

Comparisons with Other Pruning Methods

To our knowledge, no lightweight network architecture exists specifically designed for the IRSTD task. Thus, we compare IRPruneDet with other representative general pruning methods (He et al. 2018, 2019b,a). We conduct experiments under the constraint of an equal global sparsity level of 50% and evaluate the performance based on both target-level and pixel-level metrics. As shown in Table 3, the results demonstrate that the pruning technique utilized in developing IRPruneDet is more effective for the IRSTD task and outperforms other general pruning methods, validating its ability to compress model size while maintaining detection accuracy.

Quantitative Results

In Table 1 and Table 4, it can be observed that CNN-based IRSTD models outperform traditional algorithms in both pixel-level and object-level detection accuracy. However, CNN-based models have a high computational cost, *e.g.*, the FLOPs and parameters can reach up to 988.6G and 6.896M, respectively, resulting in significant storage and computational overhead. After applying the proposed pruning technique to the baseline model based on U-net18, we get a more efficient sparse network IRPruneDet. It owes to the WCP strategy for channel pruning and SCR for channel reconstruction, which effectively prevents erroneous channel pruning during the dynamic pruning process. In terms of both detection accuracy and model complexity, our IRPruneDet outperforms all other methods on the NUAA-SIRST dataset, achieving an impressive IoU of 75.12% with only 0.938G FLOPs and 0.1802M parameters.

Visual Results

In Figure 3, we present some visual object segmentation results of different IRSTD methods. IRPruneDet achieves superior target shapes and more accurate localization compared to other methods. For example, in the first, second, third, and fifth test images, our method produces masks that are closer to the ground truth images compared to other methods. In the first, fourth, and fifth test images, it can be observed that our method achieves accurate target localization and avoids false detections or missed detections even in complex backgrounds. Besides, IRPruneDet can effectively capture the edge information of small targets in infrared images, even in the presence of complex backgrounds, noise, and clutter interference. In Figure 4, we demonstrate the channel selection for the downsampling and upsampling convolutional layers during the pruning process. From the feature maps generated by different channels, we can observe that our method discards channels with excessive background noise and missing object edge information,

Method	NUAA-SIRST				IRSTD-1k				FLOPs↓	Params↓
	Pixel-Level		Object-Level		Pixel-Level		Object-Level			
	IoU↑	nIoU↑	Pd↑	Fa↓	IoU↑	nIoU↑	Pd↑	Fa↓		
MDvsFA	60.30	58.26	89.35	56.35	49.50	47.41	82.11	80.33	998.6	3.919
ACM	72.33	71.43	96.33	9.325	60.97	58.02	90.58	21.78	2.009	0.5198
ALCNet	74.31	73.12	97.34	20.21	62.05	59.58	92.19	31.56	2.127	0.5404
FC3-Net	74.75	73.81	98.13	3.21	**64.98**	**63.59**	**92.93**	**15.73**	10.57	6.896
IRPruneDet	**75.12**	**74.30**	**98.61**	**2.96**	64.54	62.71	91.74	16.04	**0.9380**	**0.1802**

Table 4: Comparison with CNN-based methods on NUAA-SIRST and IRSTD-1k in terms of $IoU(\%)$, $nIoU(\%)$, $P_d(\%)$, $F_a(10^{-6})$, $FLOPs(10^9)$, and number of parameters, *i.e.*, $Params(10^6)$.

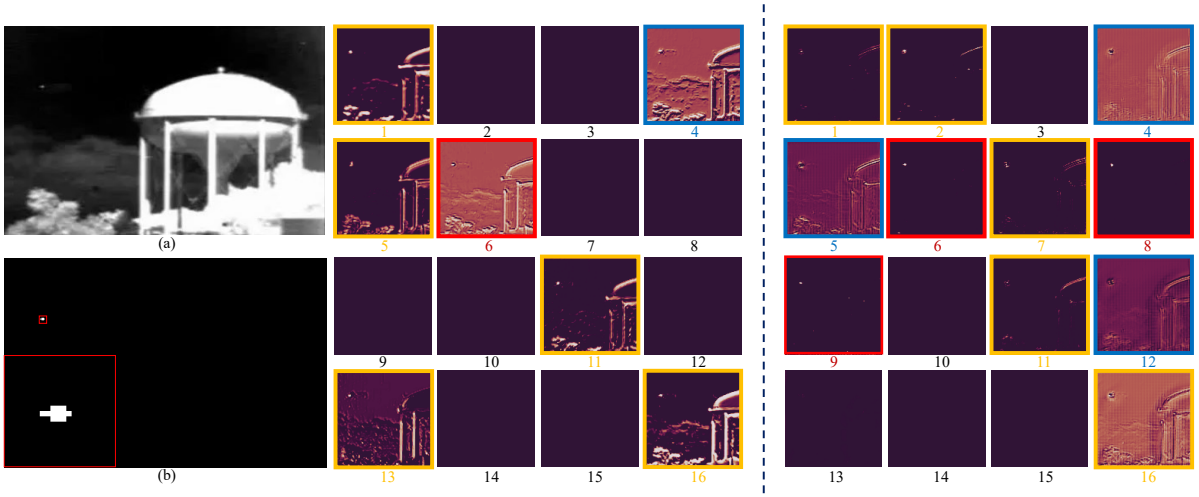

Figure 4: Above is a visualization of selected feature maps during pruning. The first conv layer of the downsampling process is shown on the left, and the last conv layer of the upsampling process is shown on the right. Blue boxes denote the channels selected by the norm-based pruning method (left: channels 4; right: channels 4, 5, 12). Red channels denote the channels selected by our WCP-based pruning method (left: channels 6; right: channels 6, 8, 9). Yellow boxes denote the channels retained by both pruning methods (left: channels 1, 5, 11, 13, 16; right: channels 1, 2, 7, 11, 16). (a) Input IR image. (b) Detection result.

such as channel 4 in the downsampling and upsampling processes. Furthermore, our method not only considers the responses in the channels but also emphasizes compelling target features, such as channels 6, 8, and 9 in the upsampling process. Although these channels may have small responses, they contain critical information about the IR targets.

Conclusion

In this paper, we introduce the idea of network pruning to the IRSTD task and develop a novel and efficient IRPruneDet model. IRPruneDet implements wavelet sparse regularization with the differential operator of the weight matrix, which efficiently discovers effective sparse structures in a pruning process without added memory usage. In addition, during the dynamic pruning process, it incorporates a soft recovery mechanism for pruned channels, preventing premature discarding of channels containing crucial features. Experiments on two public datasets validate that IRPruneDet significantly cuts FLOPs and parameters while maintaining or even improving detection accuracy. In future work, it is worth investigating the integration of the proposed method into alternative model architectures, such as vision transformers. Additionally, it would be valuable to explore more effective loss functions to enhance the preservation of useful edge information during the pruning process.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grants 62272363, 62036007, 62061047, 62176195, and U21A20514, the Young Elite Scientists Sponsorship Program by CAST under Grant 2021QNRC001, the Youth Talent Promotion Project of Shaanxi University Science and Technology Association under Grant 20200103, the Special Project on Technological Innovation and Application Development under Grant No.cstc2020jscx-dxwtB0032, the Chongqing Excellent Scientist Project under Grant No.cstc2021ycjh-bgzxm0339, and the Joint Laboratory for Innovation in Onboard Computing and Electronic Technology under Grant 2024KFKT001-1.

References

Bai, X.; and Zhou, F. 2010. Analysis of new top-hat transformation and the application for infrared dim small target detection. *Pattern Recognition*, 43(6): 2145–2156.

Chapple, P. B.; Bertilone, D. C.; Caprari, R. S.; Angeli, S.; and Newsam, G. N. 1999. Target detection in infrared and SAR terrain images using a non-Gaussian stochastic model. In *Targets and Backgrounds: Characterization and Representation V*, volume 3699, 122–132. SPIE.

Chen, C. P.; Li, H.; Wei, Y.; Xia, T.; and Tang, Y. Y. 2013. A local contrast method for small infrared target detection. *IEEE transactions on geoscience and remote sensing*, 52(1): 574–581.

Chui, C. K. 1992. *An introduction to wavelets*, volume 1. Academic press.

Cuccurullo, G.; Giordano, L.; Albanese, D.; Cinquanta, L.; and Di Matteo, M. 2012. Infrared thermography assisted control for apples microwave drying. *Journal of food engineering*, 112(4): 319–325.

Dai, Y.; and Wu, Y. 2017. Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection. *IEEE journal of selected topics in applied earth observations and remote sensing*, 10(8): 3752–3767.

Dai, Y.; Wu, Y.; Zhou, F.; and Barnard, K. 2021a. Asymmetric contextual modulation for infrared small target detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 950–959.

Dai, Y.; Wu, Y.; Zhou, F.; and Barnard, K. 2021b. Attentional local contrast networks for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 59(11): 9813–9824.

Denton, E. L.; Zaremba, W.; Bruna, J.; LeCun, Y.; and Fergus, R. 2014. Exploiting linear structure within convolutional networks for efficient evaluation. *Advances in neural information processing systems*, 27.

Deshpande, S. D.; Er, M. H.; Venkateswarlu, R.; and Chan, P. 1999. Max-mean and max-median filters for detection of small targets. In *Signal and Data Processing of Small Targets 1999*, volume 3809, 74–83. SPIE.

Ding, X.; Ding, G.; Guo, Y.; Han, J.; and Yan, C. 2019. Approximated oracle filter pruning for destructive cnn width optimization. In *International Conference on Machine Learning*, 1607–1616. PMLR.

Gao, C.; Meng, D.; Yang, Y.; Wang, Y.; Zhou, X.; and Hauptmann, A. G. 2013. Infrared patch-image model for small target detection in a single image. *IEEE transactions on image processing*, 22(12): 4996–5009.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.

Guo, S.; Wang, Y.; Li, Q.; and Yan, J. 2020. Dmcp: Differentiable markov channel pruning for neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1539–1547.

Guo, Y.; Yao, A.; and Chen, Y. 2016. Dynamic network surgery for efficient dnns. *Advances in neural information processing systems*, 29.

Han, J.; Moradi, S.; Faramarzi, I.; Zhang, H.; Zhao, Q.; Zhang, X.; and Li, N. 2020. Infrared small target detection based on the weighted strengthened local contrast measure. *IEEE Geoscience and Remote Sensing Letters*, 18(9): 1670–1674.

Han, S.; Pool, J.; Tran, J.; and Dally, W. 2015. Learning both weights and connections for efficient neural network. *Advances in neural information processing systems*, 28.

He, H.; Liu, J.; Pan, Z.; Cai, J.; Zhang, J.; Tao, D.; and Zhuang, B. 2021. Pruning self-attentions into convolutional layers in single path. *arXiv preprint arXiv:2111.11802*.

He, Y.; Dong, X.; Kang, G.; Fu, Y.; Yan, C.; and Yang, Y. 2019a. Asymptotic soft filter pruning for deep convolutional neural networks. *IEEE transactions on cybernetics*, 50(8): 3594–3604.

He, Y.; Kang, G.; Dong, X.; Fu, Y.; and Yang, Y. 2018. Soft filter pruning for accelerating deep convolutional neural networks. *arXiv preprint arXiv:1808.06866*.

He, Y.; Liu, P.; Wang, Z.; Hu, Z.; and Yang, Y. 2019b. Filter pruning via geometric median for deep convolutional neural networks acceleration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4340–4349.

He, Y.; Liu, P.; Zhu, L.; and Yang, Y. 2022. Filter pruning by switching to neighboring CNNs with good attributes. *IEEE Transactions on Neural Networks and Learning Systems*.

Hinton, G.; Vinyals, O.; and Dean, J. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Huang, Z.; Shao, W.; Wang, X.; Lin, L.; and Luo, P. 2021. Rethinking the pruning criteria for convolutional neural network. *Advances in Neural Information Processing Systems*, 34: 16305–16318.

Kang, M.; and Han, B. 2020. Operation-aware soft channel pruning using differentiable masks. In *International Conference on Machine Learning*, 5122–5131. PMLR.

Law, W.-C.; Xu, Z.; Yong, K.-T.; Liu, X.; Swihart, M. T.; Seshadri, M.; and Prasad, P. N. 2016. Manganese-doped near-infrared emitting nanocrystals for in vivo biomedical imaging. *Optics express*, 24(16): 17553–17561.

Li, B.; Xiao, C.; Wang, L.; Wang, Y.; Lin, Z.; Li, M.; An, W.; and Guo, Y. 2022a. Dense nested attention network for infrared small target detection. *IEEE Transactions on Image Processing*.

Li, H.; Kadav, A.; Durdanovic, I.; Samet, H.; and Graf, H. P. 2016. Pruning filters for efficient convnets. *arXiv preprint arXiv:1608.08710*.

Li, Z.; Meunier, D.; Mollenhauer, M.; and Gretton, A. 2022b. Optimal rates for regularized conditional mean embedding learning. *Advances in Neural Information Processing Systems*, 35: 4433–4445.

Lin, M.; Ji, R.; Wang, Y.; Zhang, Y.; Zhang, B.; Tian, Y.; and Shao, L. 2020. Hrank: Filter pruning using high-rank

feature map. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1529–1538.

Lin, S.; Ji, R.; Yan, C.; Zhang, B.; Cao, L.; Ye, Q.; Huang, F.; and Doermann, D. 2019. Towards optimal structured cnn pruning via generative adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2790–2799.

Liu, Z.; Li, J.; Shen, Z.; Huang, G.; Yan, S.; and Zhang, C. 2017. Learning efficient convolutional networks through network slimming. In *Proceedings of the IEEE international conference on computer vision*, 2736–2744.

Mallat, S. G. 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7): 674–693.

McIntosh, B.; Venkataramanan, S.; and Mahalanobis, A. 2020. Infrared target detection in cluttered environments by maximization of a target to clutter ratio (TCR) metric using a convolutional neural network. *IEEE Transactions on Aerospace and Electronic Systems*, 57(1): 485–496.

Rastegari, M.; Ordonez, V.; Redmon, J.; and Farhadi, A. 2016. Xnor-net: Imagenet classification using binary convolutional neural networks. In *proceedings of European Conference on ComputerVision*, 525–542.

Sui, Y.; Yin, M.; Xie, Y.; Phan, H.; Aliari Zonouz, S.; and Yuan, B. 2021. Chip: Channel independence-based pruning for compact neural networks. *Advances in Neural Information Processing Systems*, 34: 24604–24616.

Sun, Y.; Yang, J.; and An, W. 2020. Infrared dim and small target detection via multiple subspace learning and spatial-temporal patch-tensor model. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5): 3737–3752.

Tang, Y.; Wang, Y.; Xu, Y.; Tao, D.; Xu, C.; Xu, C.; and Xu, C. 2020. Scop: Scientific control for reliable neural network pruning. *Advances in Neural Information Processing Systems*, 33: 10936–10947.

Wang, H.; Zhou, L.; and Wang, L. 2019. Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8509–8518.

Wang, Z.; Li, C.; and Wang, X. 2021. Convolutional neural network pruning with structural redundancy reduction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14913–14922.

Zhang, J.; and Tao, D. 2020. Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things. *IEEE Internet of Things Journal*, 8(10): 7789–7817.

Zhang, L.; Peng, L.; Zhang, T.; Cao, S.; and Peng, Z. 2018. Infrared small target detection via non-convex rank approximation minimization joint l 2, 1 norm. *Remote Sensing*, 10(11): 1821.

Zhang, L.; and Peng, Z. 2019. Infrared small target detection based on partial sum of the tensor nuclear norm. *Remote Sensing*, 11(4): 382.

Zhang, M.; Bai, H.; Zhang, J.; Zhang, R.; Wang, C.; Guo, J.; and Gao, X. 2022a. RKformer: Runge-Kutta Transformer with Random-Connection Attention for Infrared Small Target Detection. In *Proceedings of the 30th ACM International Conference on Multimedia*, 1730–1738.

Zhang, M.; Yang, H.; Yue, K.; Zhang, X.; Zhu, Y.; and Li, Y. 2023. Thermodynamics-Inspired Multi-Feature Network for Infrared Small Target Detection. *Remote Sensing*, 15(19): 4716.

Zhang, M.; Yue, K.; Zhang, J.; Li, Y.; and Gao, X. 2022b. Exploring Feature Compensation and Cross-level Correlation for Infrared Small Target Detection. In *Proceedings of the 30th ACM International Conference on Multimedia*, 1857–1865.

Zhang, M.; Zhang, R.; Yang, Y.; Bai, H.; Zhang, J.; and Guo, J. 2022c. ISNET: Shape matters for infrared small target detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 877–886.

Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; and Ye, J. 2023. Object detection in 20 years: A survey. *Proceedings of the IEEE*.